

Multivariate Bernoulli Models and Generation Techniques

by Junchi Guo

Bachelor of Science and Art, June 20 2009, Nankai University

A Thesis submitted to

The Faculty of

Columbian College of Arts and Sciences

of the George Washington University

in partial fulfillment of the requirements

for the degree of Master of Science

May 15, 2011

Thesis directed by

Reza Modarres

Professor of Statistics

Dedication

To my mother, Sun Xiqin, who taught me honesty and kindness. To my father, Guo Jiancheng, who taught me diligence and gratification.

Acknowledgments

I would like to acknowledge my academic advisor, Professor Reza Modarres, for his time and insightful advice on this thesis. I would also like to express my gratitude for his help and guidance throughout my masters degree. Thanks also due to Professor Hosam M. Mahmoud for serving on my committee and his comprehensive feedback that improved the thesis. Special thanks to Professor Nozer Singpurwalla, who was a source of inspiration and taught me to appreciate the beauty of statistics. Finally, I would like to thank my family and friends who are always there to offer motivation and support.

Abstract

Multivariate Bernoulli Models and Generation Techniques

The purpose of this thesis is to explore different modeling strategies for generation of high dimensional Bernoulli vectors. We discuss the basic properties of the Multivariate Bernoulli (MB) distribution and examine several existing models. Some methods in the literature, including fully specified, latent variable, mixture and conditional mean models are discussed. Mixture models allow for the indirect full specification of a MB distribution and yield closed form solutions for the probability mass function once a suitable prior joint distribution for success probabilities is available.

We discuss a MB distribution based on the Dirichlet prior, obtain its joint pmf and describe a method of bit-flipping that can accommodate positive and negative correlations. The Dirichlet distribution provides a MB distribution whose parameter space is defined on a simplex. We present parameter implantation as a variant of mixture method that specifies the success probabilities as a given function of independent latent variables.

We also propose a generalized normal mixture model whose higher order moments are not only functions of success probabilities and second order moments, but further controlled by latent variables. Several examples are provided. While the methods of Emrich and Piedmonte (1991) and Modarres (2010) are most flexible mixture models provide more efficient methods, especially for large dimensions.

Contents

Dedication	ii
Acknowledgments	iii
Abstract	iv
1 Introduction	1
2 Methods	3
2.1 Fully specified models	3
2.1.1 Bahadur (1961)	3
2.1.2 Teugels (1990)	4
2.1.3 Kang and Jung (2001)	5
2.2 Latent Variable Models	6
2.2.1 Emrich and Piedmonte (1991)	7
2.2.2 Modarres (2010)	8
2.3 Mixture models	11
2.3.1 Dirichlet Mixture	12
2.3.2 Parameters Implantation	14
2.3.3 Generalized Normal Mixture	17
2.3.4 Mixture Models (Modarres, 2010)	19
2.4 Conditional Mean Models	20
2.4.1 Qaqish (2003)	20
2.4.2 Farrell and Sutradhar (2006)	22
2.4.3 Modarres (2010)	23
3 Conclusion	26

1 Introduction

An attribute that exists in two states is often recorded as a binary random variable whose values of 0 and 1 correspond to the presence or absence of that attribute. As a result, Bernoulli random variables form a cornerstone of statistical modeling. In applications, however, the Bernoulli random variables are usually correlated. Analysis of MB observations not only requires a better understanding of the MB distribution, it also calls for the generation of MB vectors to model the underlying process in computer simulations.

There are numerous examples of the use of Bernoulli vectors in applications, including networks (computer, social, financial, among others) where each node is represented by a Bernoulli random variable, medical research (Rubin and Wu (1997)), social science where Bernoulli observations can represent a collection of multiple social measurements obtained on a binary scale and longitudinal studies where the presence or absence of a disease is recorded over time.

Suppose Y_1, \dots, Y_d are Bernoulli random variables with given success probabilities $P(Y_i = 1) = p_i$ for $i = 1, \dots, d$ where $0 < p_i = 1 - q_i < 1$. The random vector $\mathbf{Y} = (Y_1, \dots, Y_d)'$ has a d dimensional Bernoulli distribution with mean vector \mathbf{P} and probability mass function $P(\mathbf{Y}) = P(Y_1 = y_1, \dots, Y_d = y_d)$. The distribution is defined on $[0, 1]^d$. A MB distribution is denoted with $\text{MB}(\mathbf{P}, \mathbf{\Lambda})$ where \mathbf{P} is the mean vector and $\mathbf{\Lambda}$ is the covariance matrix $\mathbf{\Lambda} = E(\mathbf{Y}\mathbf{Y}') - \mathbf{P}\mathbf{P}'$. To obtain samples from a MB distribution one is often provided with the mean vector \mathbf{P} and the pairwise covariances or the probabilities of joint occurrences $P(Y_i = 1, Y_j = 1)$ for $i < j = 1, \dots, d$.

We discuss four models to construct MB distributions: 1) fully specified models, 2) latent variables models, 3) mixture models, and 4) conditional mean models. Fully specified models ask for the realization of all 2^d possible cell probabilities. Such models are theoretically important and provide insight to the structure and the relationships between all parameters of a MB distribution. However, as d increases, they

pose formidable computational problems. We discuss the work of Bahadur(1961), Teugels(1990) and Kang and Jung(2001) in this category.

Latent variables models use the techniques of discretization of a continuous d variate vector to define a MB vector. The multivariate normal distribution is a good choice because it embeds a wide range of dependence structures. Emrich and Piedmonte (1991) and Modarres (2010) are the methods we discuss.

Mixture models use a prior distribution for \mathbf{P} so that $Cov(Y_i, Y_j) = Cov(P_i, P_j)$ for $i \neq j$ and $E(Y_i) = E(P_i)$, for $i = 1, 2, \dots, d$. As a result, there is no need to fully specify the MB distribution. In fact, the full distribution is indirectly specified once a suitable prior \mathbf{P} is available. This technique provides for a closed form solution, which makes generation of MB vectors straightforward. We propose three such mixtures in this paper: 1) Dirichlet Mixture, 2) Parameters Implantation and 3) Generalized Normal Mixture.

Conditional mean models specify $E(Y_i|Y_1, Y_2, \dots, Y_{i-1})$, for $i = 1, 2, \dots, d$ as a linear or logistic function of Y_1, Y_2, \dots, Y_{i-1} . In such models $P(Y_i)$ is often a function of the previous trials. We discuss the work of Qaqish (2003), Farrell and Sutradhar (2006) and Modarres (2010) in this category.

2 Methods

2.1 Fully specified models

2.1.1 Bahadur (1961)

Bahadur (1961) proposes a representation for the joint pmf of a MB distribution as a product of two other functions, one of which represents the pmf under independence of the variables while the other represents the dependence structure of the variables. To fully specify a d dimensional MB distribution it requires $2^d - 1$ parameters, which corresponds to the number of cells in a 2^d multinomial distribution.

Let $Z_i = (Y_i - p_i)/\sqrt{p_i(1 - p_i)}$ and $r_{ij} = E(Z_i Z_j)$. Let $r_{ijk} = E(Z_i Z_j Z_k)$ and $r_{12\dots d} = E(Z_1 Z_2 \dots Z_d)$, for $i = 1, 2, \dots, d$. The joint mass function is expressed as $P(y_1, y_2, \dots, y_d) = g(y)f(y)$, where $g(y) = \prod_{i=1}^d p_i^{y_i} (1 - p_i)^{1 - y_i}$, and the polynomial $f(y)$ is,

$$f(y) = 1 + \sum_{i < j} r_{ij} z_i z_j + \sum_{i < j < k} r_{ijk} z_i z_j z_k + \dots + r_{12\dots d} z_1 z_2 \dots z_d.$$

The $2^d - 1$ parameters are $p_i, r_{ij}, r_{ijk}, \dots, r_{12\dots d}$, for $i = 1, 2, \dots, d, \quad , i < j < k$, etc.

There are too many parameters involved in this model for large d , subsequently, Bahadur's representation is computationally inefficient. An intuitive idea is to truncate the model by omitting the third and higher-order terms in order to obtain an approximation $g(y)(1 + \sum_{i < j} r_{ij} z_i z_j)$. The second-order approximation is denote with $p_{[2]}$. Unfortunately, the second-order or higher-order approximations do not guarantee that the resulting expressions are probability distribution functions. Bahadur studies this restriction for $p_{[2]}$. Let R denote the $d \times d$ matrix of second-order correlations r_{ij} , where $r_{ii} = 1$. The approximation is non-negative only when the smallest characteristic root of R is no less than $1 - 2/\sum_{i=1}^d \beta_i$, where $\beta_i = \max(p_i/q_i, q_i/p_i)$.

Furthermore, if we assume $p_i = \alpha$, and $r_{ij} = r$, for $i \neq j$ where $t = \sum_{i=1}^d y_i$, then

one can show $p_{[2]}$ is a distribution function if and only if

$$-\frac{2}{d(d-1)}\min\left(\frac{\alpha}{1-\alpha}, \frac{1-\alpha}{\alpha}\right) \leq \frac{2\alpha(1-\alpha)}{(n-1)\alpha(1-\alpha) + \frac{1}{4} - \alpha_0},$$

where $\alpha_0 = \min_t [t - (n-1)\alpha - 1/\alpha]^2$.

Bahadur (1961) also discusses a symmetric distribution where $p_i = \alpha$, $r_{ij} = r_{(2)}$, $r_{ijk} = r_{(3)}$, and similarly for higher-order moments. This model only requires d parameters $\alpha, r_{(2)}, r_{(3)}, \dots, r_{(d)} = r_{12\dots d}$. The pmf $P(y_1, y_2, \dots, y_d)$ is symmetric if and only if the conditional distribution $p(y|t)$ is uniform, where $t = \sum_{i=1}^d y_i$. In this case, the pmf depends on Y only through t . That is, $q(t) = \frac{n!}{t!(n-t)!}p(x)$ where $q(t)$ is the pmf of t . If we assume $r_{ijk} = r_{(3)}$, $r_{ijkl} = r_{(4)}$ and similarly for higher-order moments, then there are d p'_i 's, $d(d-1)/2$ r'_{ij} 's and $d-2$ parameters $r_{(3)}, r_{(4)}, \dots, r_{(d)}$ for a total of $2d - 2 + d(d-1)/2$ parameters.

While this model is computationally inefficient, it can accommodate unpatterned correlation structures. The symmetric model, which corresponds to exchangeability, is also of practical interest.

2.1.2 Teugels (1990)

Using the concept of Kronecker product from matrix calculus, Teugels(1990) proposes a MB distribution and studies its properties. Kronecker product provides for a convenient representation for this distribution. Consider the binary expansion of an integer k , $1 \leq k \leq 2^d$, $k = 1 + \sum_{i=1}^d k_i 2^{i-1}$, where $k_i \in \{0, 1\}$. By this expansion, a one to one correspondence between an integer and a vector is established as $k \leftrightarrow (k_1, k_2, \dots, k_d)$.

The Kronecker product of two matrices A ($m \times n$) and B ($p \times q$) is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \cdots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

In particular, a $2^d \times 1$ vector with the k th element as $\prod_{i=1}^d b_i^{k_i} a_i^{1-k_i}$ can be expressed

using the Kronecker product as

$$\begin{bmatrix} a_d \\ b_d \end{bmatrix} \otimes \begin{bmatrix} a_{d-1} \\ b_{d-1} \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}.$$

Let $p_k^{(d)} = P(Y_1 = k_1, Y_2 = k_2, \dots, Y_d = k_d)$, and $\bar{Y}_i = 1 - Y_i$ for $i = 1, 2, \dots, d$. This probability can also be expressed as $p_k^{(d)} = E[\prod_{i=1}^d Y_i^{k_i} \bar{Y}_i^{1-k_i}]$, since the MB distribution's cell probabilities can be rewritten as an expectation of a product. Let $a_i = \bar{Y}_i$, and $b_i = Y_i$, so that we can write $p_k^{(d)}$ as the expectation of the k th element of a $2^d \times 1$ vector as

$$\begin{bmatrix} \bar{Y}_d \\ Y_d \end{bmatrix} \otimes \begin{bmatrix} \bar{Y}_{d-1} \\ Y_{d-1} \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} \bar{Y}_1 \\ Y_1 \end{bmatrix}.$$

Since the vector $p^{(d)} = [p_1^{(d)}, p_2^{(d)}, \dots, p_{2^d}^{(d)}]^T$ consists of all the cell probabilities we obtain

$$p^{(d)} = E \left[\begin{bmatrix} \bar{Y}_d \\ Y_d \end{bmatrix} \otimes \begin{bmatrix} \bar{Y}_{d-1} \\ Y_{d-1} \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} \bar{Y}_1 \\ Y_1 \end{bmatrix} \right].$$

Teugels (1990) discusses two other equivalent ways of parameterization. Instead of using $p_k^{(d)}$, one can either use the vector of ordinary moments $\mu^{(d)} = [\mu_1^{(d)}, \mu_2^{(d)}, \dots, \mu_{2^d}^{(d)}]^T$ or the vector of central moments $\sigma^{(d)} = [\sigma_1^{(d)}, \sigma_2^{(d)}, \dots, \sigma_{2^d}^{(d)}]^T$, where $\mu_k^{(d)} = E \left[\prod_{i=1}^d Y_i^{k_i} \right]$, and $\sigma_k^{(d)} = E \left[\prod_{i=1}^d (Y_i - p_i)^{k_i} \right]$, for $k = 1, 2, \dots, 2^d - 1$. Note that

$$p^{(d)} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{\otimes n} \mu^{(d)}$$

and

$$p^{(d)} = \begin{bmatrix} q_d & -1 \\ p_d & 1 \end{bmatrix} \otimes \begin{bmatrix} q_{d-1} & -1 \\ p_{d-1} & 1 \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} q_1 & -1 \\ p_1 & 1 \end{bmatrix} \sigma^{(d)}.$$

The representations provide a compact method of specifying the MB distribution.

2.1.3 Kang and Jung (2001)

Kang and Jung (2001) propose a simple method for generating correlated binary data with a fully specified MB distribution. The idea is to consider the d -variate MB

as a multinomial distribution with 2^d possible outcomes, the probabilities of which correspond to the 2^d cell probabilities of the given MB distribution.

The cell probabilities are $p_k^{(d)} = p(Y_1 = k_1, Y_2 = k_2, \dots, Y_d = k_d)$, $k = 1, 2, \dots, 2^d$, where (k_1, k_2, \dots, k_d) is the binary expansion of integer k . The sequence k is generated if $z_{k-1} \leq U < z_k$, where $U \sim U(0, 1)$, $z_0 = 0$, $z_k = z_{k-1} + p_k^{(d)}$, $z_{-1} = 0$, $z_{2^d} = 1$, for $k = 1, 2, \dots, 2^d$. The algorithm divides interval $(0, 1)$ into 2^d subintervals. A numerical inversion of the cumulative distribution function is then used to generate a MB random vector.

Example 1 : When $d = 2$,

$$(Y_1, Y_2) = \begin{cases} (0, 0) & \text{if } 0 < U < p_0^{(2)} \\ (1, 0) & \text{if } p_0^{(2)} \leq U < p_0^{(2)} + p_1^{(2)} \\ (0, 1) & \text{if } p_0^{(2)} + p_1^{(2)} \leq U < p_0^{(2)} + p_1^{(2)} + p_2^{(2)} \\ (1, 1) & \text{if } p_0^{(2)} + p_1^{(2)} + p_2^{(2)} \leq U < 1 \end{cases}$$

The exchangeable model, which is the same as the symmetric model of Bahadur, is discussed. Under this assumption, the conditional distribution $p(y|t)$ is uniform, where $t = \sum_{i=1}^d y_i$. Hence, the symmetry of the model provides for considerable simplification and only requires d parameters since there are only d possible values for T . Using uniformity, we calculate all the cell probabilities by dividing $p(y|t)$ by $d!/t!(d-t)!$, for $t = 0, 1, \dots, d$. As a special case of exchangeable Bernoulli, beta-binomial is suggested, but it only permits positive correlations.

Although the approach of Kang and Jung allows for unpatterned correlation, and the Bernoulli data can be easily generated once the fully specified model is given, the method is very inefficient when d is large.

2.2 Latent Variable Models

The genesis of the latent variable models is the discretization of a latent response from a chosen continuous multivariate distribution. The discretization depends on the desired correlations and marginal probabilities of the MB distribution. Let $Z =$

(Z_1, Z_2, \dots, Z_d) be a continuous d -variate vector, with $Z \sim F$, and $Z_i \sim F_0$, for $i = 1, 2, \dots, d$. The corresponding MB variables are defined as $Y_i = I(Z_i < a_i)$, for $i = 1, 2, \dots, d$, where a_i serves as the cutoff point and satisfies $F_0(a_i) = p_i$. The probability mass function of Y is

$$P(Y_i = y_i, i = 1, 2, \dots, d) = P[(-1)^{1-y_i} Z_i < (-1)^{1-y_i} a_i, i = 1, 2, \dots, d]$$

To use the model that one specifies $p_i = E(Y_i)$ and $P(Y_i = 1, Y_j = 1) = E(Y_i Y_j)$. The higher order moments are borrowed from the latent variables. For bivariate case, there are many candidates for the model of the latent variables, especially by making use of the copula, $F = C(F_0, F_0)$, where C is a copula. But, for higher dimensions, models that provide computational efficiency and a wide range of dependence are hard to find.

2.2.1 Emrich and Piedmonte (1991)

Emrich and Piedmonte (1991) study a latent variable model that uses a multivariate normal distribution to generate MB responses. Let Φ be the cumulative distribution function for a standard bivariate normal random vector with correlation coefficient ρ

$$\Phi[x_1, x_2; \rho] = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(z_1, z_2; \rho) dz_1 dz_2$$

where $f(z_1, z_2; \rho)$ is the pdf of the standard bivariate normal distribution. Let $z(p_i) = \Phi^{-1}(p_i)$ be the p_i th percentile of a standard normal random variable. Given the marginal distributions p_1, p_2, \dots, p_d and correlations δ_{ij} for $i \neq j$, the latent variables model produces

$$P(Y_i = 1, Y_j = 1) = \Phi[z(p_i), z(p_j); \rho_{ij}] = \int_{-\infty}^{z(p_i)} \int_{-\infty}^{z(p_j)} f(z_1, z_2; \rho_{ij}) dz_1 dz_2 \quad i \neq j.$$

Since $\Phi[x_1, x_2; \rho]$ is strictly increasing in ρ , one can use the bisection method to solve for the unique root of ρ_{ij} . Note that the desired marginal distributions are ensured by the parameterization of $z(p_i)$, $i = 1, 2, \dots, d$. However, one needs to solve $d(d-1)/2$

non-linear equations to obtain all the ρ_{ij} since

$$\begin{aligned}
\text{corr}(Y_i, Y_j) &= \text{cov}(Y_i, Y_j) / (p_i q_i p_j q_j)^{1/2} = \{p(Y_i = 1, Y_j = 1) - p_i p_j\} / (p_i q_i p_j q_j)^{1/2} \\
&= \{P[Z_i < z(p_i), Z_j < z(p_j)] - p_i p_j\} / (p_i q_i p_j q_j)^{1/2} \\
&= \{\Phi[z(p_i), z(p_j); \rho_{ij}] - p_i p_j\} / (p_i q_i p_j q_j)^{1/2} \\
&= \delta_{ij} \quad i \neq j
\end{aligned}$$

where $\Lambda = \text{Cov}(Y) = (\delta_{ij})$ is given.

Let $\Sigma = (\rho_{ij})$ and note that a MB vector can be generated from a multivariate normal vector $N(0, \Sigma)$ with $Y_i = y_i$ if $(-1)^{1-y_i} Z_i < (-1)^{1-y_i} z(p_i)$, $i = 1, 2, \dots, d$.

However, even when the Λ is positive definite one cannot assure that the obtained Σ is positive definite. The algorithm becomes inefficient for large d as the bisection method is required to find ρ_{ij} .

2.2.2 Modarres (2010)

Emrich and Piedmonte (1991) study a method that transforms multivariate normal random vectors to obtain MB vectors. The method only assumes δ_{jk} 's are known, but does not guarantee that the generated vector has a positive definite correlation matrix. Bounds on the components p_i to ensure compatibility with Λ are complicated and involve probabilities that are not uniquely determined by the correlations and univariate means (see Chaganty and Joe, 2006).

Let $z(p_j) = \Phi^{-1}(p_j)$ represents the p_j -th percentile of a standard normal random variable. Given \mathbf{P} and Δ we first need to solve $d(d-1)/2$ non-linear equations

$$\Phi(z(p_j), z(p_k); \rho_{jk}) = \delta_{jk} (p_j q_j p_k q_k)^{1/2} + p_j p_k$$

for ρ_{jk} , for $j = 1, \dots, d$ and $k = j+1, \dots, d$. These equations relate the correlation δ_{ij} of a bivariate Bernoulli distribution to the correlation of a latent bivariate normal distribution. Since the function is monotone it has a unique zero in the interval $[-1, 1]$. Method of bisection is suggested to solve each equation.

While the solution of the non-linear equations for ρ_{jk} always exist, the corresponding covariance matrix $\Sigma = (\rho_{jk})$ is not guaranteed to be positive definite. One

possibility is to obtain the smallest eigen value of Σ and reject the generated sample if it is not positive. The rejection method may prove to be computationally costly as the efficiency of the method depends on the probability that the smallest eigen value is positive. However, the proposed Plackett latent model makes this approach feasible as it does not require solving non-linear equations.

Alternatively, to address a similar issue in financial applications, Higham (2002) considers an iterative scheme to find the closest correlation matrix. For arbitrary symmetric matrix Σ , the distance measure is defined as a weighted Frobenius norm $\gamma(\Omega) = \min\{\|\Sigma - \Omega\| : \Omega \text{ is a correlation matrix}\}$. The matrix that achieves this minimum is obtained using the alternating projection method as described in Higham (2002). This algorithm can be used in case the given correlation matrix Δ or the correlation matrix of the latent variables $\Sigma = (\rho_{jk})$ is not positive definite. It is of interest to study the effects of this approximation on the covariance matrix of the generated Bernoulli vectors.

It is well understood that correlation coefficient δ_{jk} of a bivariate Bernoulli distribution is confined to a subinterval of $(-1, +1)$. Clearly $\lambda_{jk} \geq -p_j p_k$ since $P(Y_j = 1, Y_k = 1) \geq 0$. The former inequality coupled with $P(Y_j = 1, Y_k = 1) \leq \min(p_j, p_k)$ implies $\lambda_{jk} = \min(p_j q_k, p_k q_j)$ and $\delta_{jk} \leq b_{jk} = \min((p_j q_k / p_k q_j)^{1/2}, (p_k q_j / p_j q_k)^{1/2})$. Similarly, $a_{jk} = \max(-(p_j p_k / q_j q_k)^{1/2}, -(q_j q_k / p_j p_k)^{1/2}) \leq \delta_{jk}$ so that $a_{jk} \leq \delta_{jk} \leq b_{jk}$. It should be pointed out that these are necessary, but not sufficient conditions for $\Delta = (\delta_{jk})$ to be positive definite when $d > 2$.

In fact, if we use the bounds $a_{jk} \leq \delta_{jk} \leq b_{jk}$ to set $\delta_{jk} = b_{jk} - \epsilon$ where ϵ is a small positive constant, then Δ remains positive definite. However, if we use the bounds to set $\delta_{jk} = a_{jk} + \epsilon$, then Δ remains negative definite until ϵ is increased to $1/2$ for some \mathbf{P} . Hence, Δ is not positive definite when there are large negative correlations even when $a_{jk} \leq \delta_{jk} \leq b_{jk}$.

The given \mathbf{P} and \mathbf{A} provide $d + d(d-1)/2$ parameters of the total $2^d - 1$ parameters needed to fully specify $P(\mathbf{Y})$. The dependence vector of \mathbf{Y} also includes $\sum_{k=3}^d \binom{d}{k}$ other parameters that represent k -way mixed moments. These moments are obtained indirectly from the latent vector.

To increase the efficiency of the model by Emrich&Piedmonte(1991) discussed previously, Modarres(2010) uses the bivariate Plackett distribution to approximate the bivariate normal distribution so that the computations are feasible. Because we do not have to solve the $d(d-1)/2$ non-linear equations to find $\Sigma = (\rho_{ij})$.

Plackett (1965) gives a copula

$$h(u, v) = \frac{\psi[(\psi - 1)(u + v - 2uv) + 1]}{\{[1 + (u + v)(\psi - 1)]^2 - 4\psi(\psi - 1)uv\}^{3/2}}$$

where $0 < u < 1$, $0 < v < 1$ and $\psi > 0$. The independence copula is obtained when $\psi = 1$. The Frechet lower bound is achieved when ψ approaches 0 at $U = 1 - V$, the upper bound is achieved when ψ tends to be infinity at $U = V$

Let $H(x, y)$ denote the corresponding bivariate distribution function with marginal distribution functions $F(x)$ and $G(y)$. Plackett distribution has constant odds ratio for all $(x, y) \in R^2$ where

$$\psi = \frac{H(x, y)(1 - F(x) - G(y) + H(x, y))}{(F(x) - H(x, y))(G(y) - H(x, y))}.$$

The corresponding distribution function for a given odds ration is

$$H(x, y; \psi) = \frac{S(x, y) - (S^2(x, y) - 4\psi(\psi - 1)F(x)G(y))^{1/2}}{2(\psi - 1)}$$

where $S(x, y) = 1 + (F(x) + G(y))(\psi - 1)$.

Let $F = G = \Phi$, which is the $N(0, 1)$ distribution function, so that H has normal marginal distributions. Plackett (1965) indicated that a bivariate normal distribution with correlation ρ is well-approximated with $H(x, y; \psi)$, often with more than 3 digits of accuracy, that is $\Phi(x, y; \rho) \approx H(x, y; \psi)$. By solving the equation $\Phi(0, 0; \rho) = H(0, 0; \psi)$, we have $\rho = \cos(\frac{\pi}{1+\sqrt{\psi}})$

We solve

$$H_{ij} = H(z(p_i), z(p_j); \phi_{ij}) = \delta_{ij}(p_i q_i p_j q_j)^{1/2} + p_j p_k$$

from which ϕ_{ij} can be easily computed as

$$\phi_{ij} = \frac{H_{ij}(1 - p_i - p_j + H_{ij})}{(p_i - H_{ij})(p_j - H_{ij})}.$$

Using $\rho_{ji} = \cos(\frac{\pi}{1+\sqrt{\psi_{ij}}})$, we obtain the corresponding ρ_{ji} .

2.3 Mixture models

Our motivation for the use of mixture models to define and generate MB vectors is to assess a network's integrity to perform its function. The existing paradigm in performance assessment aims to evaluate $P(\Psi(\mathbf{Y}) = 1|\mathbf{p})$ where Ψ is a measure of network's reliability (Singpurwalla, 2006). Given the success probability of each node in a network, the goal is to calculate the network's reliability. Let p_i denote the reliability of node i with distribution $f(p_i)$ and $P(Y_i = 1) = \int_0^1 p_i f(p_i) dp_i = E_f(p_i)$. Classical network performance assessment focuses on p_i , assuming it is known and evaluates the reliability of the network $P(\Psi(X) = 1|P)$.

Arguing that one is also interested in system's survivability or $P(\Psi(\mathbf{Y}) = 1)$, Singpurwalla (2006) calls for a paradigm shift from \mathbf{p} to $E_f(\mathbf{p})$ where $f(p_1, \dots, p_d)$ is the joint distribution of p_i 's on the d -dimensional hypercube. Dependence can arise at the level of the Y_i 's and/or p_i 's and is hierarchically argued. Nodes Y_i 's are independent given the p_i 's and p_i 's are themselves dependent.

Suppose the Bernoulli parameter is random with some density $f(p)$, and given p , the d Bernoulli random variables are conditionally independent and identically distributed. The probability mass function is given as

$$p(y_1, y_2, \dots, y_d) = \int_0^1 p^t (1-p)^{d-t} f(p) dp$$

where $t = \sum_{i=1}^d y_i$. This is in fact the exchangeable case.

The more general mixture models are generated by introducing a vector $P = (P_1, P_2, \dots, P_d)$ rather than a single P , and accordingly, let $f(p) = f(p_1, p_2, \dots, p_d)$ be the density of P . The d Bernoulli random variables are still conditionally independent, but not identically distributed, it has the pmf

$$p(y_1, y_2, \dots, y_d) = \int_0^1 \dots \int_0^1 \prod_{i=1}^d p_i^{y_i} (1-p_i)^{1-y_i} f(p_1, \dots, p_d) dp_1 \dots dp_d.$$

The marginal distributions are

$$\begin{aligned} p(y_j) &= \int_0^1 \dots \int_0^1 \sum_{y_1=0}^1 \dots \sum_{y_{j-1}=0}^1 \sum_{y_{j+1}=0}^1 \dots \sum_{y_d=0}^1 \prod_{i=1}^d p_i^{y_i} (1-p_i)^{1-y_i} f(p_1, \dots, p_d) dp_1 \dots dp_d \\ &= E[P_j^{y_j} (1-P_j)^{1-y_j}]. \end{aligned}$$

The following properties hold for this model: $E(Y_i) = E(P_i)$, $Var(Y_i) = E(P_i)[1 - E(P_i)]$, for $i = 1, 2, \dots, d$, $Cov(Y_i, Y_j) = Cov(P_i, P_j)$, for $i \neq j$.

2.3.1 Dirichlet Mixture

A d -dimensional Dirichlet distribution $\mathcal{D}(\theta_1, \dots, \theta_d, \theta_{d+1})$ for $P = (P_1, P_2, \dots, P_d)$ has the probability density function

$$f(p) = f(p_1, p_2, \dots, p_d; \theta_1, \dots, \theta_d, \theta_{d+1}) = \frac{\Gamma(\sum_{i=1}^{d+1} \theta_i)}{\prod_{i=1}^{d+1} \Gamma(\theta_i)} \left(\prod_{i=1}^d p_i^{\theta_i-1} \right) (1 - \sum_{i=1}^d p_i)^{\theta_{d+1}-1}$$

where $p_i \geq 0$, $i = 1, 2, \dots, d$, $\sum_{i=1}^d p_i \leq 1$, and the $d+1$ parameters $\theta_1, \theta_2, \dots, \theta_d, \theta_{d+1}$ are all real and positive. It is straightforward to show

$$\begin{aligned} P(Y_j = 1) = E(P_j) &= \frac{\Gamma(\sum_{i=1}^{d+1} \theta_i)}{\prod_{i=1}^{d+1} \Gamma(\theta_i)} \left[\frac{\Gamma(\sum_{i=1}^{d+1} \theta_i + 1)}{\Gamma(\theta_j + 1) \prod_{i \neq j} \Gamma(\theta_i)} \right]^{-1} = \frac{\Gamma(\sum_{i=1}^{d+1} \theta_i) \Gamma(\theta_j + 1)}{\Gamma(\theta_j) \Gamma(\sum_{i=1}^{d+1} \theta_i + 1)} \\ &= \frac{\theta_j}{\sum_{i=1}^{d+1} \theta_i}, \quad j = 1, 2, \dots, d. \end{aligned}$$

Similarly, we have

$$P(Y_m = 1, Y_n = 1) = E(P_m P_n) = \frac{\theta_m \theta_n}{(\sum_{i=1}^{d+1} \theta_i)(\sum_{i=1}^{d+1} \theta_i + 1)}, \quad m \neq n,$$

and

$$P(Y_1 = 1, Y_2 = 1, \dots, Y_d = 1) = E(P_1 P_2 \dots P_d) = \frac{\prod_{i=1}^d \theta_i}{\prod_{j=0}^{d-1} \sum_{i=1}^{d+1} (\theta_i + j)}.$$

Other ordinary moments can be obtained in a similar way, so all the cell probabilities can be solved by the relationship given by Teugels(1990).

We can also obtain a closed-form expression for the pmf as follows. Let $x_i = 2y_i - 1$ and note that the product $\prod_{i=1}^d p_i^{y_i} (1 - p_i)^{1-y_i}$ can be expanded as

$$\begin{aligned} \prod_{i=1}^d p_i^{y_i} (1 - p_i)^{1-y_i} &= \prod_{i=1}^d (1 - y_i) + p_1 x_1 \prod_{i=2}^d (1 - y_i) + \dots + p_d x_d \prod_{i=1}^{d-1} (1 - y_i) \\ &\quad + p_1 p_2 x_1 x_2 \prod_{i=3}^d (1 - y_i) + \dots + p_{d-1} p_d x_{d-1} x_d \prod_{i=1}^{d-2} (1 - y_i) \\ &\quad + \dots + p_1 p_2 \dots p_d \prod_{i=1}^d x_i. \end{aligned}$$

The coefficient of $p_{s_1}p_{s_2}\dots p_{s_l}$ is $(2y_{s_1}-1)(2y_{s_2}-1)\dots(2y_{s_l}-1)(1-y_{r_1})(1-y_{r_2})\dots(1-y_{r_{m-l}})$. There are a total of 2^d such coefficients. Let $\mathbf{C}(y)$ be a 2^d dimensional vector of these coefficients, arranged corresponding to the elements of the following vector π

$$\pi = (1, p_1, p_2, \dots, p_d, p_1p_2, p_1p_3, \dots, p_{d-1}p_d, \dots, p_1p_2\dots p_d).$$

Substituting the expansion

$$\prod_{i=1}^d p_i^{y_i}(1-p_i)^{1-y_i} = \mathbf{C}(y)' \pi$$

in the mixture model produces

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_d = y_d) = \int \mathbf{C}(y)' \pi f(p) dp = \mathbf{C}(y)' E(\pi)$$

where

$$E(\pi) = (1, E(P_1), \dots, E(P_d), E(P_1P_2), \dots, E(P_{d-1}P_d), \dots, E(P_1P_2\dots P_d)).$$

This expression gives a clear picture of all mixture models.

Note that

$$Cov(Y_m, Y_n) = \frac{\theta_m \theta_n}{(\sum_{i=1}^{d+1} \theta_i)(\sum_{i=1}^{d+1} \theta_i + 1)} - \frac{\theta_m}{\sum_{i=1}^{d+1} \theta_i} \frac{\theta_n}{\sum_{i=1}^{d+1} \theta_i} < 0$$

which shows that only negative correlations can be achieved under this model. In order to obtain positive correlations, one can, instead of defining $P(Y_n = 1) = E(P_n)$, define $P(Y_n = 1) = E(1 - P_n)$. Hence,

$$Cov(Y_m, Y_n) = E(P_m(1 - P_n)) - E(P_m)E(1 - P_n) = E(P_m)E(P_n) - E(P_mP_n)$$

which is clearly positive.

Using this technique, we cannot convert all $\binom{d}{2}$ correlations into positive. If one Y_i is changed, then we get $d - 1$ correlations converted to positive and if we change Y_i and Y_j , the correlation between these two variables remains negative. In general, $i(d - i)$ positive correlations can be converted out of the $d(d - 1)/2$ original ones as long as $d - i > 0$. Another clear restriction is that the Dirichlet distribution requires $\sum_{i=1}^d p_i \leq 1$.

2.3.2 Parameters Implantation

The difficulty with the mixture approach is the selection of a suitable prior distribution $f(p) = f(p_1, p_2, \dots, p_d)$ and the subsequent d -dimensional integration in order to obtain the pmf of MB distribution. Analytical evaluation is not possible and numerical integration is costly.

To describe the technique of parameter implantation, let $P_i = g_i(\sum_{j=1}^s \alpha_{ij} Z_j)$, where g_i are functions in $[0,1]$ and α_{ij} are $d \times s$ implanted parameters to adjust the model, $i = 1, \dots, d$. There are also s independent random variables Z_j with specified distributions. The advantage of this model is that its construction calls for independent Z 's while P 's would be dependent (note that s does not have to equal d). Hence, we can simply use multivariate distribution of independent variables Z 's

$$f(z_1, z_2, \dots, z_s) = \prod_{i=1}^s f(z_i).$$

Define

$$P(Y_1 = y_1, \dots, Y_d = y_d) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^d (p_i)^{y_i} (1 - p_i)^{1-y_i} f(z_1, \dots, z_s) dz_1 \dots dz_s$$

where $p_i = g_i(\sum_{j=1}^s \alpha_{ij} z_j)$. Since $f(z_1, z_2, \dots, z_s) = \prod_{i=1}^s f(z_i)$, z 's are averaged out, and α 's are left to adjust the model. Note that, given z 's, P 's are functions of α 's only.

Example 2 : Let $s = 2$, $f(z_1, z_2) = \lambda_1 e^{-\lambda_1 z_1} \lambda_2 e^{-\lambda_2 z_2}$, $z_1 \geq 0$ and $z_2 \geq 0$. That is, the components of Z are independent, each with an exponential distribution. Let $p_i = e^{-\sum_{j=1}^s \alpha_{ij} z_j} = e^{-(\alpha_{i1} z_1 + \alpha_{i2} z_2)}$, $i = 1, 2, \dots, d$, where α 's ≥ 0 . Substituting back

in the model, we obtain

$$\begin{aligned}
P(Y_1 = 1, \dots, Y_d = 1) &= \int_0^\infty \int_0^\infty e^{-\sum_{i=1}^d \sum_{j=1}^d \alpha_{ij} z_j} \lambda_1 \lambda_2 e^{-\lambda_1 z_1} e^{-\lambda_2 z_2} dz_1 dz_2 \\
&= \lambda_1 \lambda_2 \int_0^\infty e^{-(\sum_{i=1}^d \alpha_{i1} + \lambda_1) z_1} dz_1 \int_0^\infty e^{-(\sum_{i=1}^d \alpha_{i2} + \lambda_2) z_2} dz_2 \\
&= \frac{\lambda_1 \lambda_2}{(\sum_{i=1}^d \alpha_{i1} + \lambda_1)(\sum_{i=1}^d \alpha_{i2} + \lambda_2)} \\
&= \frac{1}{(\sum_{i=1}^d \frac{\alpha_{i1}}{\lambda_1} + 1)(\sum_{i=1}^d \frac{\alpha_{i2}}{\lambda_2} + 1)}, \\
P(Y_i = 1, Y_j = 1) &= \frac{\lambda_1 \lambda_2}{(\alpha_{i1} + \alpha_{j1} + \lambda_1)(\alpha_{i2} + \alpha_{j2} + \lambda_2)} \\
&= \frac{1}{(\frac{\alpha_{i1} + \alpha_{j1}}{\lambda_1} + 1)(\frac{\alpha_{i2} + \alpha_{j2}}{\lambda_2} + 1)}, \quad i \neq j \\
P(Y_i = 1) &= \frac{\lambda_1 \lambda_2}{(\alpha_{i1} + \lambda_1)(\alpha_{i2} + \lambda_2)} \\
&= \frac{1}{(\frac{\alpha_{i1}}{\lambda_1} + 1)(\frac{\alpha_{i2}}{\lambda_2} + 1)}, \quad i = 1, 2, \dots, d.
\end{aligned}$$

An equivalent parameterization is to divide the original α_{i1} ($i = 1, 2, \dots, d$) by λ_1 , α_{i2} ($i = 1, 2, \dots, d$) by λ_2 . Hence, let $\lambda_i = 1$, i.e., we can simply let $f(z_1, z_2) = e^{-z_1} e^{-z_2}$ and obtain

$$\begin{aligned}
P(Y_1 = 1, \dots, Y_d = 1) &= \frac{1}{(\sum_{i=1}^d \alpha_{i1} + 1)(\sum_{i=1}^d \alpha_{i2} + 1)}, \\
P(Y_i = 1, Y_j = 1) &= \frac{1}{(\alpha_{i1} + \alpha_{j1} + 1)(\alpha_{i2} + \alpha_{j2} + 1)}, \quad i \neq j \\
P(Y_i = 1) &= \frac{1}{(\alpha_{i1} + 1)(\alpha_{i2} + 1)}, \quad i = 1, 2, \dots, d.
\end{aligned}$$

Since $(\alpha_{i1} + \alpha_{j1} + 1) < \alpha_{i1} \alpha_{j1} + \alpha_{i1} + \alpha_{j1} + 1 = (\alpha_{i1} + 1)(\alpha_{j1} + 1)$ we have $(\alpha_{i1} + \alpha_{j1} + 1)(\alpha_{i2} + \alpha_{j2} + 1) < (\alpha_{i1} + 1)(\alpha_{j1} + 1)(\alpha_{i2} + 1)(\alpha_{j2} + 1) = (\alpha_{i1} + 1)(\alpha_{i2} + 1)(\alpha_{j1} + 1)(\alpha_{j2} + 1)$. Hence, $E(Y_i Y_j) > E(Y_i)E(Y_j)$ and the correlation is always positive.

Suppose $\text{Corr}(Y_i, Y_j) = \rho$ for $i \neq j$, $E(Y_i) = p$ for $i = 1, 2, \dots, d$, i.e., exchangeable model. $P(Y_i = 1) = \frac{1}{(\alpha_{i1} + 1)(\alpha_{i2} + 1)} = p$ for $i = 1, 2, \dots, d$, $P(Y_i = 1, Y_j = 1) = \frac{1}{(\alpha_{i1} + \alpha_{j1} + 1)(\alpha_{i2} + \alpha_{j2} + 1)} = \rho p q + p^2$ for $i \neq j$. To satisfy the above condition, we need

$\alpha_{i1} = \alpha_1, \alpha_{i2} = \alpha_2$ for $i = 1, 2, \dots, d$, then we have the equations

$$\frac{1}{(\alpha_1 + 1)(\alpha_2 + 1)} = p, \quad \frac{1}{(2\alpha_1 + 1)(2\alpha_2 + 1)} = \rho pq + p^2$$

Solving the equations we have

$$\alpha_1 = \frac{\Psi + \sqrt{\Psi^2 - 4\Phi}}{2}, \quad \alpha_2 = \frac{\Psi - \sqrt{\Psi^2 - 4\Phi}}{2}$$

where $\Psi = \frac{2}{p} - \frac{1}{2(\rho pq + p^2)} - \frac{3}{2}$ and $\Phi = \frac{1}{2} \left(\frac{1}{\rho pq + p^2} - \frac{2}{p} + 1 \right)$

Example 3 : Let $s = 2$, $f(z_1, z_2) = 1$, $0 \leq z_1 \leq 1$ and $0 \leq z_2 \leq 1$. That is, the components of Z are independent uniform(0, 1) variables. Let $p_i = \frac{\alpha_{i1}z_1 + \alpha_{i2}z_2}{C_i}$, $i = 1, 2, \dots, d$, where α 's ≥ 0 and $C_i = \sqrt{2\alpha_{i1}^2 + 2\alpha_{i2}^2}$. Substituting back in the model, we obtain for $i \neq j$,

$$\begin{aligned} P(Y_i = 1, Y_j = 1) &= \int_0^1 \int_0^1 \frac{(\alpha_{j1}z_1 + \alpha_{j2}z_2)}{C_j} dz_1 dz_2 \\ &= \frac{1}{C_i C_j} \left(\frac{\alpha_{i1}\alpha_{j1} + \alpha_{i2}\alpha_{j2}}{3} + \frac{\alpha_{i1}\alpha_{j2} + \alpha_{i2}\alpha_{j1}}{4} \right), \\ P(Y_i = 1) &= \int_0^1 \int_0^1 \frac{(\alpha_{i1}z_1 + \alpha_{i2}z_2)}{C_i} dz_1 dz_2 = \frac{\alpha_{i1} + \alpha_{i2}}{2C_i}, \quad i = 1, 2, \dots, d. \end{aligned}$$

Notice that in this case

$$E(Y_i)E(Y_j) = \frac{1}{C_i C_j} \left(\frac{\alpha_{i1}\alpha_{j1} + \alpha_{i2}\alpha_{j2}}{4} + \frac{\alpha_{i1}\alpha_{j2} + \alpha_{i2}\alpha_{j1}}{4} \right) < E(Y_i Y_j)$$

So the correlation is always positive. Restriction also exists for the margin, since $p(Y_i = 1) = \frac{\alpha_{i1} + \alpha_{i2}}{2C_i} \leq \frac{1}{2}$.

It is not surprising that both examples can only offer positive correlations, since every Y_i ($i = 1, 2, \dots, d$) shares the same prior variance from Z 's. Negative correlations can be obtained using the sign flipping. For instance, in example 2, define $p_i = 1 - e^{-\sum_{j=1}^s \alpha_{ij} z_j}$.

2.3.3 Generalized Normal Mixture

Based on the pdf of multivariate normal distribution, we discuss a new mixture model. Let $P_i = \exp\{-\frac{1}{2}Z'\Sigma_i^{-1}Z\}$, $i = 1, 2, \dots, d$, where Σ_i is $s \times s$ positive definite, and $Z = (Z_1, Z_2, \dots, Z_s) \sim N(0, \Psi)$. Let $f(z)$ denote the pdf of the s -dimensional vector Z . Clearly, P_i 's are in the interval $[0,1]$ and we obtain the joint pmf $P(Y_1 = y_1, \dots, Y_d = y_d)$ as

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^d p_i^{y_i} (1 - p_i)^{1-y_i} f(z_1, \dots, z_s) dz_1 \dots dz_s$$

which provides a closed form for the pmf of MB distribution since $P(Y_1 = 1, \dots, Y_d = 1)$ equals

$$\frac{1}{(2\pi)^{s/2} \sqrt{|\Psi|}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}z'(\Sigma_1^{-1} + \dots + \Sigma_d^{-1} + \Psi^{-1})z\} dz_1 \dots dz_s.$$

In order to evaluate the d -fold integration, let $A = \Sigma_1^{-1} + \dots + \Sigma_d^{-1} + \Psi^{-1}$, which is positive definite. We obtain

$$P(Y_1 = 1, \dots, Y_d = 1) = \frac{1}{\sqrt{|\Psi||A|}}$$

since $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (2\pi)^{-s/2} \sqrt{|A|} \exp\{-\frac{1}{2}z'Az\} dz_1 \dots dz_s = 1$.

Similarly, for any set $H = \{h_1, h_2, \dots, h_l\}$ containing l elements out of $\{1, 2, \dots, d\}$, we have

$$P[\cap_{i \in H} (Y_i = 1)] = \frac{1}{(2\pi)^{s/2} \sqrt{|\Psi|}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}z'(\Sigma_{h_1}^{-1} + \dots + \Sigma_{h_l}^{-1} + \Psi^{-1})z\} dz_1 \dots dz_s.$$

Let $A_H = \Sigma_{h_1}^{-1} + \dots + \Sigma_{h_l}^{-1} + \Psi^{-1}$. We obtain

$$P[\cap_{i \in H} (Y_i = 1)] = \frac{1}{\sqrt{|\Psi||A_H|}}.$$

Example 4 : Let $\Psi = \text{diag}(\theta_1, \theta_2, \dots, \theta_s)$, $\Sigma_i = \text{diag}(\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{is})$, $i = 1, 2, \dots, d$ where, to be positive definite, all diagonal elements are positive. We have

$$P(Y_1 = 1, Y_2 = 1, \dots, Y_d = 1) = \frac{1}{\sqrt{|\Psi||\Psi^{-1} + \Sigma_1^{-1} + \dots + \Sigma_d^{-1}|}} = \frac{1}{\sqrt{\prod_{j=1}^s \theta_j (\frac{1}{\theta_j} + \sum_{i=1}^d \frac{1}{\sigma_{ij}})}}.$$

$$P(Y_1 = 1, Y_2 = 1) = \frac{1}{\sqrt{\prod_{j=1}^s \theta_j (\frac{1}{\theta_j} + \sum_{i=1}^2 \frac{1}{\sigma_{ij}})}} = \rho \sqrt{p_1 q_1 p_2 q_2} + p_1 p_2,$$

$$P(Y_1 = 1) = \frac{1}{\sqrt{|\Psi| |\Psi^{-1} + \Sigma_1^{-1}|}} = \frac{1}{\sqrt{\prod_{j=1}^s \theta_j (\frac{1}{\theta_j} + \frac{1}{\sigma_{1j}})}}.$$

It follows from $P(Y_1 = 1, Y_2 = 1, \dots, Y_d = 1)$ that if the j th column ($j = 1, 2, \dots, s$) of Σ is divided by θ_j , then the new Σ values θ_j/σ_{ij} together with $\Psi = I$ are equivalent with the original given conditions. Hence, we can always assume $\Psi = I$ when both Σ 's and Ψ are diagonal.

Similar to the model in the previous section, this method only offers positive correlation.

Suppose $\Psi = (1 - \rho)I + \rho 11'$ is the constant correlation matrix with correlation ρ where $-\frac{1}{d-1} < \rho < 1$; that is,

$$\Psi = \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho & \rho \\ \rho & 1 & \rho & \cdots & \rho & \rho \\ \rho & \rho & 1 & \cdots & \rho & \rho \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho & \rho & \rho & \cdots & 1 & \rho \\ \rho & \rho & \rho & \cdots & \rho & 1 \end{bmatrix}.$$

The inverse is $\Psi^{-1} = \frac{1}{1-\rho}(I - \frac{\rho}{1+(d-1)\rho}11')$ with diagonal elements $t = \psi_{ii}^{(-1)} = \frac{1}{1-\rho}(1 - \frac{\rho}{1+(s-1)\rho})$ and off-diagonal elements $w = \psi_{ij}^{(-1)} = -\frac{1}{1-\rho} \frac{\rho}{1+(s-1)\rho}$. One can show $A = \Sigma_1^{-1} + \dots + \Sigma_d^{-1} + \Psi^{-1}$ with elements

$$A = \begin{bmatrix} \Delta_1 & w & w & \cdots & w & w \\ w & \Delta_2 & w & \cdots & w & w \\ w & w & \Delta_3 & \cdots & w & w \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ w & w & w & \cdots & \Delta_{s-1} & w \\ w & w & w & \cdots & w & \Delta_s \end{bmatrix}$$

where $\Delta_j = t + \sum_{i=1}^d \frac{1}{\sigma_{ij}}$, and $|A| = \prod_{j=3}^s (\Delta_j - w) [\Delta_1(\Delta_2 - w) - w(w - \Delta_1)] - (s - 2)w \prod_{j=1}^{s-1} (\Delta_j - w)$. Hence, one can calculate $P(Y_1 = 1, Y_2 = 1, \dots, Y_d = 1)$ by the formula

$$P(Y_1 = 1, \dots, Y_d = 1) = \frac{1}{\sqrt{|\Psi||A|}}$$

Other moments can be calculated in a similar manner. The Mixture Model offers relationships among the 2^d cell probabilities (or some functions of the 2^d parameters). For the traditional mixture models, all the third and higher-order moments are functions of the margins and second-order moments. However, the dimension of parameter space that we introduce in this method determines how many moments are sufficient to fix the higher order under a generalized Normal mixture model.

2.3.4 Mixture Models (Modarres, 2010)

Suppose $f_i(P_i|P_1, \dots, P_{i-1}) \sim \text{Beta}(\alpha_i + S_{i-1}, \beta_i - S_{i-1})$ where $\alpha_i > -S_{i-1}$ and $\beta_i > S_{i-1} = \sum_{j=1}^{i-1} y_j$. We obtain

$$P(\mathbf{Y} = \mathbf{y}) = \int_0^1 \dots \int_0^1 \prod_{i=1}^d P_i^{y_i} (1 - P_i)^{1-y_i} C_i P_i^{\alpha_i + S_{i-1} - 1} (1 - P_i)^{\beta_i - S_{i-1} - 1} dP_i$$

where $C_i = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i + S_{i-1})\Gamma(\beta_i - S_{i-1})}$. That is,

$$P(Y_i = y_i, i = 1, \dots, d) = \prod_{i=1}^d \left(\frac{\alpha_i + S_{i-1}}{\alpha_i + \beta_i} \right)^{y_i} \left(\frac{\beta_i - S_{i-1}}{\alpha_i + \beta_i} \right)^{1-y_i}.$$

The joint distribution of (Y_1, \dots, Y_d) is the product of d correlated binary random variables whose success probabilities $p_i = E\left(\frac{\alpha_i + S_{i-1}}{\alpha_i + \beta_i}\right)$ depend on past history of the trials. The moments of \mathbf{Y} are given by

$$P(Y_i = 1, i = 1, \dots, d) = \prod_{i=1}^d \frac{\alpha_i + i - 1}{\alpha_i + \beta_i}.$$

It follows from the joint density that

$$P(Y_i = 1|y_1, \dots, y_{i-1}) = E(Y_i|y_1, \dots, y_{i-1}) = \frac{\alpha_i + S_{i-1}}{\alpha_i + \beta_i}$$

and

$$P(Y_i = y_i, \dots, Y_j = y_j | Y_1 = y_1, \dots, Y_{i-1} = y_{i-1}) = \prod_{k=i}^j \left(\frac{\alpha_k + S_{k-1}}{\alpha_k + \beta_k} \right)^{y_k} \left(\frac{\beta_k - S_{k-1}}{\alpha_k + \beta_k} \right)^{1-y_k}.$$

The former expression is useful for obtaining $E(Y_i)$ and the latter expression can be used to obtain mixed moments. One can show $E(Y_i) = E(E(Y_i | y_1, \dots, y_{i-1})) = p_i = E\left(\frac{\alpha_i + S_{i-1}}{\alpha_i + \beta_i}\right)$. Hence, $E(Y_i) = p_i = \frac{1}{\alpha_i + \beta_i}(\alpha_i + \sum_{j=1}^{i-1} p_j)$ and $\text{Var}(Y_i) = p_i(1 - p_i) = \frac{(\alpha_i + \sum_{j=1}^{i-1} p_j)(\beta_i - \sum_{j=1}^{i-1} p_j)}{(\alpha_i + \beta_i)^2}$.

There are not enough constraints to solve for (α_i, β_i) in terms of the given p_i 's and the pairwise correlations. One can fix either α_i or β_i to obtain unique solutions. For example, $\beta_i = d$ yields $\alpha_i = (dp_i - \sum_{j=1}^{i-1} p_j)/(1 - p_i)$. Similarly, one can use $\beta_i = i$

Suppose we partition $\mathbf{Y} = (Y_1, \dots, Y_i)'$ as $\mathbf{Y}^{(1)} = Y_i$ and $\mathbf{Y}^{(2)} = (Y_1, \dots, Y_{i-1})'$. It is not difficult to show $E(\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)} = \mathbf{y}^{(2)}) = p_i + \frac{1}{\alpha_i + \beta_i} \sum_{j=1}^{i-1} (y_j - p_j)$. This shows that the conditional mean is linear and establishes a connection between the mixture model and the conditional mean model. Using the same technique as in the previous section to find the covariance matrix, we can show

$$\text{Cov}(Y_k, Y_j) = \frac{1}{\alpha_j + \beta_j} \sum_{i=1}^{j-1} \text{Cov}(Y_k, Y_i)$$

for $k = 1, \dots, j - 1$. This identity is useful for finding the form of Δ . For example,

$$\Delta = \begin{pmatrix} p_1(1 - p_1) & \frac{p_1(1-p_1)}{\alpha_2 + \beta_2} & \frac{p_1(1-p_1)(\alpha_2 + \beta_2 + 1)}{(\alpha_2 + \beta_2)(\alpha_3 + \beta_3)} \\ & p_2(1 - p_2) & \frac{p_1(1-p_1)}{(\alpha_2 + \beta_2)(\alpha_3 + \beta_3)} + \frac{p_2(1-p_2)}{\alpha_3 + \beta_3} \\ & & p_3(1 - p_3) \end{pmatrix}.$$

2.4 Conditional Mean Models

2.4.1 Qaqish (2003)

Qaqish (2003) proposes a conditional mean model in which a process is built up in the sense that the success probability of Y_i is conditional on the performance of previous trails $(Y_1, Y_2, \dots, Y_{i-1})$, $i = 2, 3, \dots, d$. Let $X_i = (Y_1, Y_2, \dots, Y_{i-1})$. Qaqish's model is

$$\lambda_i = E(Y_i | X_i = x_i) = p_i + \sum_{j=1}^{i-1} b_{ij}(y_j - p_j), \quad i = 2, 3, \dots, d$$

where $b_i = (b_{i1}, b_{i2}, \dots, b_{i,i-1})$ is given in closed form as $b_i = \text{cov}(X_i)^{-1} \text{cov}(X_i, Y_i)$, which is $\Sigma_{22}^{-1} \Sigma_{21}$ if Y is so partitioned. Subsequently, given the covariance matrix and b , the vector p and all cell probabilities are determined. Note that λ_i is a function of X_i , since it is the conditional mean, each λ_i corresponds to 2^{i-1} values as the configuration of x_i varies. As probabilities, the conditional means must be in the unit interval.

The detection of reproducibility proceeded as follows. For given Σ , λ_i is maximized over 2^{i-1} possible configurations of x_i by taking $y_j = 1$ if $b_{ij} > 0$, and $y_j = 0$ if $b_{ij} < 0$, $j = 1, 2, \dots, i-1$. Reversing the choice produces the minimum λ_i , and a lemma is given as

$$\begin{aligned} \max \lambda_i(x_i; \Sigma) &= p_i + \sum_{+} b_{ij}(1 - p_j) - \sum_{-} b_{ij}p_j \\ \min \lambda_i(x_i; \Sigma) &= p_i + \sum_{-} b_{ij}(1 - p_j) - \sum_{+} b_{ij}p_j \end{aligned}$$

where \sum_{+} and \sum_{-} denote the summation over $\{j : 1 \leq j < i, b_{ij} > 0\}$ and $\{j : 1 \leq j < i, b_{ij} < 0\}$ respectively. Based on this lemma, in order to make sure the conditional linear model is reproducible, one only needs to satisfy $\min \lambda_i(x_i; \Sigma) \geq 0$ and $\max \lambda_i(x_i; \Sigma) \leq 1$, $i = 2, 3, \dots, d$. This lemma makes it a lot easier for checking reproducibility. It is important to note that before checking reproducibility, one needs to make sure that the given Σ satisfies the conditions of being the covariance matrix of a MB distribution.

For patterned correlation matrices let $r_{ij} = \text{corr}(Y_i, Y_j)$ for $i \neq j$, and $v_{ii} = \text{var}(Y_i)$, for $i = 1, 2, \dots, d$. When all $r_{ij} = r$, b_{ij} can be obtained as

$$b_{ij} = \frac{r}{1 + (i-2)r} \left(\frac{v_{ii}}{v_{jj}} \right), \quad j = 1, 2, \dots, i-1.$$

Moreover, for equal means, $p_1 = p_2 = \dots = p_d$, the conditional means can be simplified to

$$\lambda_i = \frac{(1-r)p_1 + r \sum_{j=1}^{i-1} y_j}{1 + (i-2)r}, \quad i = 2, \dots, d.$$

In the case of AR(2) correlations, in which $r_{ij} = r^{|i-j|}$ for $i \neq j$

$$\lambda_i = p_i + r(y_{i-1} - p_{i-1}) \left(\frac{v_{ii}}{v_{i-1,i-1}} \right)^{(1/2)}, \quad i = 2, \dots, d.$$

In the case of MA(1) correlations, in which $r_{ij} = rI(|i - j| = 1)$ for $i \neq j$

$$b_{ij} = \left(\frac{v_{ii}}{v_{jj}}\right)^{(1/2)} \frac{a^j - a^{-j}}{a^{-i} - a^i}, \quad j = 1, 2, \dots, i - 1$$

where $a = [(1 - 4r^2)^{1/2} - 1]/(2r)$.

To generate a vector of observations from a MB distribution under the conditional mean model one first generates a Bernoulli random variable Y_1 with probability p_1 and subsequent values of Y_i are obtained iteratively from the model.

This model is computationally efficient for large dimensions. However, one must ensure that the conditional probabilities obtained from the model remain in the unit interval.

2.4.2 Farrell and Sutradhar (2006)

To ensure that the conditional means in Qaqish (2003) remain in the unit interval, Farrell and Sutradhar (2006) propose to use a logistic link function to model the conditional probability of Y_i . However, the resulting model is more complex to analyze and suitable for small d .

Using Qaqish (2003)'s notation, let $X_i = (Y_1, Y_2, \dots, Y_{i-1})$. The covariates are represented as $Z_i = (Z_1, Z_2, \dots, Z_p)$, the regression parameters as $\beta = (\beta_1, \beta_2, \dots, \beta_p)$, and the regression parameters for X_i are represented as $b = (b_{i-1}, b_{i-2}, \dots, b_1)$, then the non-linear model given by:

$$\lambda_i = E(Y_i | X_i = x_i) = \exp(z'_i \beta + x'_i b) / [1 + \exp(z'_i \beta + x'_i b)], \quad i = 2, 3, \dots, d.$$

Clearly, λ_i is in the unit interval.

Consider a simple autoregressive model

$$\lambda_i = E(Y_i | Y_i = y_{i-1}) = \exp(z'_i \beta + y_{i-1} b_1) / [1 + \exp(z'_i \beta + y_{i-1} b_1)], \quad i = 2, 3, \dots, d$$

Let $P(Y_1 = 1) = \exp(z'_1 \beta) / [1 + \exp(z'_1 \beta)]$. Farrell and Sutradhar(2006) show $p_i = P(Y_i = 1) = P(Y_i = 1 | Y_{i-1} = 1)P(Y_{i-1} = 1) + P(Y_i = 1 | Y_{i-1} = 0)P(Y_{i-1} = 0)$.

Denote $p_{i1} = P(Y_i = 1|Y_{i-1} = 1)$ and $p_{i0} = P(Y_i = 1|Y_{i-1} = 0)$. For $i > j$ we have

$$\text{cov}(Y_j, Y_i) = p_j(1 - p_j) \prod_{k=j+1}^i (p_{k1} - p_{k0}).$$

It then easily follows that

$$\text{corr}(Y_j, Y_i) = \sqrt{p_j(1 - p_j)/p_i(1 - p_i)} \prod_{k=j+1}^i (p_{k1} - p_{k0}).$$

Note that $\text{corr}(Y_j, Y_i)$ can range from -1 to 1 since p_{k0} and p_{k1} are within $(0, 1)$.

2.4.3 Modarres (2010)

Drezner and Farum (1993) introduced a Bernoulli process $\{Y_j; j \geq 1\}$ in which the random variables Y_j are correlated as the success probability of a trial conditional on the previous trials depends on the total number of successes so far. That is,

$$P(Y_{j+1} = 1|F_j) = (1 - \theta_j)p + \frac{\theta_j}{j}S_j$$

where $0 \leq \theta_j \leq 1$ are dependence parameters, $S_j = \sum_{i=1}^j Y_i$ is the total number of successes so far and F_j is the σ -field generated by (Y_1, \dots, Y_j) . If the initial trial Y_1 has a Bernoulli distribution with parameter p , it follows that Y_1, \dots, Y_d are identically distributed Bernoulli random variables with mean $E(Y_{j+1}) = (1 - \theta_j)p + \frac{\theta_j}{j}(jp) = p$, and $\text{Var}(Y_{j+1}) = p(1 - p)$.

Modarres (2010) notes that $P(Y_{j+1} = 1|F_j) = (1 - \theta_j)p + \theta_j\hat{p}$ where $\hat{p} = \frac{1}{j} \sum_{i=1}^j Y_i$ is an estimate p based on the first j trials. That is, the conditional probability of success on the $(j + 1)$ -th trial is the weighted average of the unconditional probability of success p and this estimate. The dependence parameter θ_j determines the mixture proportion with large values of θ_j placing more weight on \hat{p} . When $\lambda_{jk} = \text{Corr}(Y_j, Y_k)$ is constant we have $\theta_j = 1$. When $\theta_j = 0$ for all values of j we obtain the Bernoulli process. An intuitive appeal of the model lies in its "success breathes success" disposition. When $\theta_j > 0$, the $(j + 1)$ variate is expected to have a larger success probability than p provided that the average probability of success for the previous j variates

is greater than p . James, James and Qi (2008) study limit theorems of this class of distributions.

Modarres (2010) generalize the correlated Bernoulli process of Drezner and Farum (1993) by changing the common p to p_i . Suppose

$$P(Y_{j+1} = 1|F_j) = p_{j+1} + \frac{\theta_j}{j} \sum_{i=1}^j (Y_i - p_i)$$

for $i = 1, \dots, j+1$. If $p_i = p$, then the model reduces to the correlated Bernoulli process described by Drezner and Farum (1993); that is, $P(Y_{j+1} = 1|F_j) = p + \frac{\theta_j}{j}(-jp + \sum_{i=1}^j Y_i)$ which simplifies to $(1 - \theta_j)p + \frac{\theta_j}{j}S_j$.

Note that $P(Y_{j+1} = 1) = E(Y_{j+1}) = E(E(Y_{j+1}|F_j)) = p_{j+1}$. Similarly, $E(Y_{j+1}^2) = p_{j+1}$, so that $\text{Var}(Y_{j+1}) = p_{j+1}(1 - p_{j+1})$. Modarres (2010) obtain the moments of \mathbf{Y} as

$$P(Y_1 = 1, Y_2 = 1, \dots, Y_j = 1) = \prod_{i=1}^j \left\{ p_i + \frac{\theta_{i-1}}{i-1} \left(j-1 - \sum_{k=1}^{i-1} p_k \right) \right\}$$

where the term involving θ_{i-1} is treated as zero when $i = 1$. The distribution of \mathbf{Y} is defined as

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_j = y_j) = \prod_{i=1}^j \alpha_i^{y_i} (1 - \alpha_i)^{1-y_i}$$

where $\alpha_i = p_i + \frac{\theta_{i-1}}{i-1} \sum_{k=1}^{i-1} (y_i - p_k)$. Note that S_j or the total number of successes so far is the sum of j correlated Bernoulli random variables, each with probability of success p_i .

Suppose we partition $\mathbf{Y} = (Y_1, \dots, Y_{j+1})'$ as $\mathbf{Y}^{(1)} = Y_{j+1}$ and $\mathbf{Y}^{(2)} = (Y_1, \dots, Y_j)'$. We know

$$E(Y_{j+1}|Y_1 = y_1, \dots, Y_j = y_j) = p_{j+1} + \mathbf{B}(y_1 - p_1, \dots, y_j - p_j)'$$

where \mathbf{P} is partitioned similarly and $\mathbf{B} = \mathbf{\Lambda}_{12}\mathbf{\Lambda}_{22}^{-1}$. In this case, $\mathbf{\Lambda}_{22}$ is the $j \times j$ covariance matrix of (Y_1, \dots, Y_j) and $\mathbf{\Lambda}_{12}$ is the $1 \times j$ vector of covariances between Y_{j+1} and (Y_1, \dots, Y_j) .

Comparing the two representations of $E(Y_{j+1}|Y_1 = y_1, \dots, Y_j = y_j)$ one can identify $\mathbf{B} = (\theta_j/j)\mathbf{1}'$ or $\mathbf{\Lambda}_{12} = (\theta_j/j)\mathbf{1}'\mathbf{\Lambda}_{22}$. The right hand side of the later expression

can be expanded and equated to the left hand side, term by term, to show

$$\text{Cov}(Y_k, Y_j) = \frac{\theta_{j-1}}{j-1} \sum_{i=1}^{j-1} \text{Cov}(Y_k, Y_i)$$

for $k = 1, \dots, j-1$. We can use this identity to show

$$\theta_{j-1} = (j-1) \frac{\text{Cov}(Y_k, Y_j)}{\sum_{i=1}^{j-1} \text{Cov}(Y_i, Y_k)}$$

which is clearly in the interval $(0, 1)$ and reduces to $\theta_j = 1$ for constant correlations.

Example 5 *This identity is very useful in obtaining the form of Δ for this model.*

For example,

$$\Delta = \begin{pmatrix} \delta_{11} & \theta_1 \delta_{11} & \frac{1}{2} \theta_2 (1 + \theta_1) \delta_{11} \\ \theta_1 \delta_{11} & \delta_{22} & \frac{1}{2} \theta_2 (\delta_{22} + \theta_1 \delta_{11}) \\ \frac{1}{2} \theta_2 (1 + \theta_1) \delta_{11} & \frac{1}{2} \theta_2 (\delta_{22} + \theta_1 \delta_{11}) & \delta_{33} \end{pmatrix}$$

when $d = 3$.

More generally, the covariance matrix $\Delta = (\delta_{rc})$ has elements $\delta_{rr} = p_r(1 - p_r)$ and

$$\delta_{rc} = \frac{\theta_{c-1}}{c-1} (\delta_{r1} + \dots + \delta_{r(c-1)})$$

for $r = 1, \dots, j$ and $c = r+1, \dots, j$. Clearly, $\text{Cov}(Y_r, Y_c)$ is $\frac{\theta_{c-1}}{c-1}$ times the sum all the previous covariances on that row. Similarly, the correlation matrix $\Lambda = (\lambda_{rc})$ has elements $\lambda_{rr} = 1$ and

$$\lambda_{rc} = \frac{\theta_{c-1}}{(c-1) \sqrt{p_c(1-p_c)}} \sum_{i=1}^{c-1} \lambda_{ri} \sqrt{p_i(1-p_i)}$$

for $r = 1, \dots, j$ and $c = r+1, \dots, j$. Modarres (2010) uses this identity to show

$$P(Y_k = 1, Y_j = 1) = p_k \left\{ p_j + \frac{\theta_{j-1}}{j-1} \sum_{i=1}^{j-1} (p(Y_i = 1 | Y_k = 1) - p_i) \right\}.$$

Furthermore, $\text{Var}(Y_{j+1} | Y_1 = y_1, \dots, Y_j = y_j) = \Lambda_{11} - \Lambda_{12} \Lambda_{22}^{-1} \Lambda_{21}$, and $\mathbf{B} = \Lambda_{12} \Lambda_{22}^{-1} = (\theta_j/j) \mathbf{1}'$ so that

$$\text{Var}(Y_j | Y_1 = y_1, \dots, Y_{j-1} = y_{j-1}) = \delta_{jj} - \frac{\theta_{j-1}}{j-1} \sum_{i=1}^{j-1} \delta_{ij}.$$

3 Conclusion

The fully specified models proposed by Bahadur (1961), Teugels (1990) and Kang and Jung (2001) call for the realization of all 2^d possible cell probabilities. Full specification of the cell probabilities is only feasible when d is small and can pose formidable computational problems for large d . However, such models are theoretically important.

Emrich and Piedmonte (1991)'s latent variable model is a very practical choice, especially when dealing with unpatterned correlations and high dimensions, and Modarres (2010) gives a good solution for the calculation problem. This model also provides a way to generate Bernoulli index from continuous situation. Conditional mean models provide a Bernoulli process. They are applicable in longitudinal studies.

The traditional Mixture models use the moments of the selected prior distribution to adjust the moments of the MB distribution. The difficulty with this method is the selection of a suitable prior distribution with enough flexibility that accommodates different correlation structures and marginal distributions. The Dirichlet Mixture offers flexible MB distributions. The method of bit-flipping can be used to obtain positive and negative correlations. The method of parameter implantation is proposed to side-step the issue of prior selection. One can set the dimension s of the implanted parameters to strike a balance between flexibility and efficiency. Generalized Normal Mixture model makes use of the special integration property of the pdf of multivariate normal distribution.

The mixture models we present here can be summarized as

$$p_1(\mathbf{y}|\theta_1, \theta_2) = \int_{\alpha} p_2(\mathbf{y}|\theta_1, \alpha)df(\alpha|\theta_2)$$

where θ_1 and θ_2 are the parameters of the resulting pmf, α represents the terms that are averaged out. In the case of traditional mixture, there is no θ_1 and Y 's are independent given α . While under the other two cases, Y 's are dependent given α and θ_1 .

References

- [1] Bahadur, R. R. (1961). *A Representation of the Joint Distribution of the Responses to n Dichotomous Items*. In *Studies in Item Analysis and Prediction*, Ed. H. Solomon, 158–168. Stanford University Press.
- [2] Chaganty, N. R., Joe, H. (2006). *Range of correlation matrices for dependent Bernoulli random variables*. *Biometrika*, 93(10), 197–206.
- [3] Emrich, L. J., Piedmonte, M. R. (1991). *A method for generating high-dimensional multivariate binary variates*. *American Statistician*, 45, 302–304.
- [4] Farrell, P. J. and Sutradhar, B. C. (2006) *A non-linear conditional probability model for generating correlated binary data*. *Statistics and Probability Letters*, 76, 353–361.
- [5] James, B., James, K., Qi, Y. (2008). *Limit Theorems for correlated Bernoulli random variables*. *Statistics and Probability Letters*, 78, 2339–2345.
- [6] Kang, S. H. and Jung, S. H. (2001). *Generating Correlated Binary with Complete Specification of the Joint Distribution*. *Biometrical Journal*, 43, 263–269.
- [7] Higham, N. J. (2002). *Computing the nearest correlation matrix a problem from finance*. *IMA Journal of Numerical Analysis*, 22(3), 329–343.
- [8] Modarres, R. (2010). *High Dimensional Generation of Bernoulli Random Vectors*. Under Review.
- [9] Plackett, R. L. (1965) *A Class of Bivariate Distributions A Class of Bivariate Distributions*. *Journal of the American Statistical Association*, Vol. 60, No. 310, 516–522.
- [10] Qaqish, B. (2003). *A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations*. *Biometrika*, 90, 455–463.

- [11] Teugels, J. L. (1990). *Some Presentations of the Multivariate Bernoulli and Binomial Distributions*. *Journal of Multivariate Analysis*, 32, 256–268.
- [12] Donald B. Rubin and Ying Nian Wu. (1997). *Modeling Schizophrenic Behavior Using General Mixture Components*. *Biostatistics*, 53, 243-261.