Cuts for Multinomial Distribution

EDAWRD K. IP AND YUCHUNG J. WANG

Abstract

In this note, we provide results for parameter separation that are useful for conditional inference for multivariate categorical responses. Specifically, difference and addition operators are proposed to parameterize a multinonial distribution, and based upon such parameterization, we present conditions for obtaining cuts. Cuts are statistics ancillary to the parameter of interest and are general tools for facilitating a strong form of separation of inference.

Key Words: multivariate categorical responses; odds ratio; canonical parameter; mean parameter; Hamadan product; separation of inference.

1. Introduction

For a multidimensional parameter, sometimes inference concerns with only a proper subset of it. A common approach is to first separate the parameter space into mutually exclusive parts- the portion of interest and the portion that is not, and then draw inference conditioned on a statistic that is non-informative about the portion of interest. Not only the conditional inference operates on a smaller dimensional space, the conditional variance is also smaller than the unconditional variance, see Barndorff-Neilson (1978). There are two types of parameter separation. The weak form of parameter separation is orthogonality. Cox & Reid (1987) show that when two complementary portions of a vector parameter are orthogonal, the profile likelihood conditioned on ancillary statistics is asymptotically as efficient as the full likelihood. The strong form of separation is likelihood-independence, where the likelihood can be factorized. A cut of statistics is one of the ways that facilitates such a factorization. A cut is formally defined as follows:

Suppose the unknown parameter can be written as vector $\theta^T = (\theta^{(1)}, \theta^{(2)})$, where $\theta^{(1)}$ is the portion of interest. If

a) parameter $\theta^{(1)}$ and $\theta^{(2)}$ are variation independent, i.e., range of θ= range of $\theta^{(1)}\times$ range of $\theta^{(2)}$;

b) the minimal sufficient statistic is $S^T = (S^{(1)}, S^{(2)})$;

c) the conditional distribution of $S^{(1)}$ given $S^{(2)}$ depends only on $\theta^{(1)}$; and

d) the marginal distribution of $S^{(2)}$ depends only on $\theta^{(2)}$ and not on $\theta^{(1)}$,

then the statistic $S^{(2)}$ is said to be a *cut* ancillary for $\theta^{(1)}$ and sufficient for $\theta^{(2)}$. This definition is due to Barndorff-Nielson & Cox (1994, p.38). Cuts are useful in the modeling of multivariate categorical responses and in finding the proper ancillary statistic for efficient conditional inference.

A common assumption for multivariate categorical responses is the multinomial distribution, of which the canonical parameter is a vector of logarithms of cell probabilities when parameterized into an exponential family. It is well-known that canonical parameter is invariant under linear transformations. Wang (1986) proposed an order-dependent linear transformation such that the vector of logarithms of logits and logarithms of conditional odds ratios emerges as the canonical parameter. Wang's linear transformation, though quite general, is hard to construct when the dimension of responses becomes large. In this paper, we extend the results in Wang (1986) and propose an operator to streamline the expressions for odds and odds ratios for multivariate categorical data. Based on the reparameterized multinomial distribution, we propose a theory to find cuts.

2. Reparameterization of multinomial distribution

Throughout this paper, we assume that there are J categorical response variables $Y_j, 1 \leq j \leq J$ and each response has K_j consecutive categories. Hence, the total number of cells is $K_1 \times \cdots \times K_J = K$. Let the observed cell counts of $Y = (Y_1^T, \cdots, Y_J^T)^T$ be $y = (y_1^T, \cdots, y_J^T)^T$, where each Y_i or y_i is a vector of length K_i and let π be the vector of K cell probabilities arranged in the same lexicographical order as Y satisfying $\sum \pi_{k_1,\cdots,k_J} = 1$. Specifically, the first index in $(k_1, \cdots, k_J)$ changes fastest and the last index slowest. Then the multinomial distribution of the responses has density function

$$h(y) \exp(y^T \log \pi). \tag{1}$$

The linear transformation proposed by Wang (1986) is to construct a $K \times K$ square and invertible matrix A and change the log-likelihood of (1) into

$$\ell(\theta) = s^T \theta - \kappa(\theta) + \log h([A^T]^{-1} s), \tag{2}$$

where $\theta = A^{-1} \log \pi$, $s^T = y^T A$ and $\kappa(\theta) = \log(\sum_{k=1}^K \exp(A_k \theta))$ with A_k being the k^{th} column vector of A. It can be easily proved that $\kappa(\theta)$ is the cummulant generating function. To illustrate the notation for θ, we adopt Whitaker's (1990) difference and identity operators for a function $g(k_1, \cdots, k_J)$ of J categorical variables. Whittaker (1990, p.35) defined a difference operator $\bigtriangledown_j g(k_1, \cdots, k_J) = g(k_1, \cdots, k_j, \cdots, k_J) - g(k_1, \cdots, k_j + 1, \cdots, k_J)$, and the identity operator $Ig(k_1, \cdots, k_J) = g(k_1, \cdots, k_J)$. Also, define $\bigtriangledown_j \bigtriangledown_i g = \bigtriangledown_j(\bigtriangledown_i g)$, and $\bigtriangledown_j(g_1 \pm g_2) = \bigtriangledown_j g_1 \pm \bigtriangledown_j g_2$. It is straightforward to verify that the operators satisfy the commutative law: $\bigtriangledown_j \bigtriangledown_i g = \bigtriangledown_i \bigtriangledown_j g$ and $I \bigtriangledown_i g = \bigtriangledown_i I g = \bigtriangledown_i g$. For simplicity, write the $\bigtriangledown_i \bigtriangledown_j g$ as $\bigtriangledown_{ij} g$. Now consider the function $g(k_1, \cdots, k_J) = \log \pi_{k_1, \cdots, k_J}$. For example, in a 2×3 categorical responses, let $g(i_1, i_2)$ be $\log \pi_{i_1 i_2}$. Then, $\bigtriangledown_1 \log \pi_{13} = \log(\pi_{13}/\pi_{23})$ is the adjacent-categories logit (Agresti 1990, p.318); and $\bigtriangledown_{12} \log \pi_{11} = \log(\pi_{11}\pi_{22}/\pi_{12}\pi_{21})$ is the local log odds ratio.

Example 1. Consider a 2×3 table with cell probabilities $\pi = (\pi_{11}, \pi_{21}, \cdots, \pi_{23})^T$ and cell counts $y = (y_{11}, \cdots, y_{23})^T$. Wang (1986) used the following matrix:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

Let $P = (p_{ij})$ and $Q = (q_{ij})$ be two matrices of dimension $a \times b$ and $m \times n$, respectively. Define a $am \times bn$ matrix

$$P \otimes Q = \begin{pmatrix} q_{11}P & \cdots & q_{1n}P \\ \cdots & \cdots & \cdots \\ q_{m1}P & \cdots & q_{mn}P \end{pmatrix}.$$

It is easy to see that $A = B_1 \otimes B_2$, where

$$B_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Because the inverse matrix of a $\otimes$-product of matrices is the $\otimes$-product of the inverse matrices, we have $A^{-1} = B_1^{-1} \otimes B_2^{-1}$, where

$$B_1^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad B_2^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Each row, except the last row, of B_1^{-1} and B_2^{-1} represents a difference operator, so we may replace B_1^{-1} and B_2^{-1} by the vectors $(\bigtriangledown_1, I)^T$ and $(\bigtriangledown_2, \bigtriangledown_2, I)^T$, respectively. Then the new canonical parameter becomes $\theta^T = (\bigtriangledown_{12} \log \pi_{11}, \bigtriangledown_2 \log \pi_{21}, \bigtriangledown_{12} \log \pi_{12}, \bigtriangledown_2 \log \pi_{22}, \bigtriangledown_1 \log \pi_{13}, I \log \pi_{23})$. They are the odds ratios and adjacent logits, which are the conventional measures of dependence among categorical responses. The corresponding sufficient statistics s^T is $(y_{11}, y_{+1}, y_{11} + y_{12}, y_{+1} + y_{+2}, y_{1+}, y_{++})$, where the plus sign indicates a summation over the index at position of the subscript.

Example 2. For a $2\times2\times2$ response variables, let $y = (y_{111}, y_{211}, y_{121}, y_{221}, y_{112}, y_{212}, y_{122}, y_{222})^T$ and π be arranged in the same order. The linear transformation is represented by the following matrices:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here, $\theta = A^{-1} \log \pi = (\bigtriangledown_{123} \log \pi_{111}, \bigtriangledown_{23} \log \pi_{211}, \bigtriangledown_{13} \log \pi_{121}, \bigtriangledown_3 \log \pi_{221}, \bigtriangledown_{12} \log \pi_{112}, \bigtriangledown_2 \log \pi_{212}, \bigtriangledown_1 \log \pi_{122}, I \log \pi_{222})^T$ and $s = A^T y = (y_{111}, y_{+11}, y_{1+1}, y_{++1}, y_{11+}, y_{+1+}, y_{1++}, y_{+++})^T$.

The difference operator is a discrete version of the partial derivative. The derivation of association functions from partial derivatives for continuous random variables has been studied by Holland and Wang (1987). Let $f(x, y)$ denote the bivariate density between variables x and y. They define the local dependence function of $f(x, y)$ as $\gamma(x, y) = \partial^2 \log f/\partial y \partial x$. An interpretation of $\gamma(x, y)$ as the local Pearson correlation is provided by Jones (1996). This notion of using derivatives to derive association parameters leads to an elegant expression for A^{-1} in terms of difference operators.

Let $\bigtriangledown^j = (\bigtriangledown_j, \cdots, I)$ be a K_j-vector of $(K_j - 1)$ $\bigtriangledown_j$ and one I, $j = 1, \cdots, J$. The vector $\bigtriangledown = \bigtriangledown^J \otimes \cdots \otimes \bigtriangledown^1$ is therefore of length $K = K_1 \times \cdots \times K_J$. Define a Hamadan product $\cdot$ between a vector of operators $\bigtriangledown = (\bigtriangledown_1, \cdots, \bigtriangledown_J)^T$ and a vector of functions $g = (g_1, \cdots, g_J)^T$ as $\bigtriangledown \cdot g = (\bigtriangledown_1 g_1, \cdots, \bigtriangledown_J g_J)^T$.

Lemma 2.1. $\theta = A^{-1} \log \pi = \bigtriangledown \cdot \log \pi$.

In general, the canonical parameter that the difference operator $\bigtriangledown$ derives consists of the conditional adjacent-categories logits, the conditional log odds ratios and the conditional log ratios of odds ratios, and so on. These parameters have the same interpretations as the model parameters in a saturated log-linear model. To mimic the difference operator, we define the following addition operator:

Definition 2.1. The *j^{th}-coordinate addition operator* $\triangle_j$ is defined as

$$\triangle_j g(x_1, \cdots, x_j, \cdots, x_J) = \sum_{x_t=1}^{x_j} g(x_1, \cdots, x_t, \cdots, x_J)$$

where x_t is in the jth position, and that $\triangle_i \triangle_j g = \triangle_i(\triangle_j g)$, and $\triangle_i(g_1 \pm g_2) = \triangle_i g_1 \pm \triangle_i g_2$. Write $\triangle_i \triangle_j g$ as $\triangle_{ij} g$. Further, let $\triangle = \triangle^J \otimes \triangle^{J-1} \cdots \otimes \triangle^1$ where $\triangle^j = (I, \triangle_j, \cdots, \triangle_j)$ is a vector of I and $(K_j - 1)$ $\triangle_j$.

Lemma 2.2. $s = A^T y = \triangle \cdot y$.

The proofs Lemma 2.1 and Lemma 2.2 are straightforward and the details can be found in Ip *et al.* (2003).

Example 1(continued). For 2×3 table, the sufficient statistic is $s^T = (Iy_{11}, \triangle_1 y_{21}, \triangle_2 y_{12}, \triangle_{12}\, y_{22}, \triangle_2 y_{13}, \triangle_{12} y_{33})$.

3. Cuts for multinomial distribution

In the previous section, we show that A^{-1} is equivalent to a vector of difference operators. Accordingly, $\exp(y^T \log \pi)$ is parameterized into $\exp((\triangle \cdot y)^T(\bigtriangledown \cdot \log \pi))$. While Wang (1986) illustrated that different A matrices may be of interest, in this note we call $\bigtriangledown \cdot \log \pi$ the standard canonical parameter due to its equivalence to log-linear models.

Barndorff-Nielsen (1978, p.208) showed that, for a two-dimensional multinomial distribution, either $\{y_{i+}\}$ or $\{y_{+j}\}$ is a cut but $\{y_{i+}, y_{+j}\}$ is not a cut. No explicit example of cut for table of higher dimension was given. Hence, we present a theory for deriving cuts in higher dimensional tables. Before we can discuss our results it is necessary to introduce the notions of mean parameter and mixed parameterization.

Definition 3.1. For any distribution of the exponential family, the expectation of the sufficient statistic S is called the *mean parameter*, i.e. $\mu(\theta) = E_\theta S$. For a partition $\theta^T = ({\theta^{(1)}}^T, {\theta^{(2)}}^T)$, the sufficient statistic can be correspondingly partitioned into $S^T = ({S^{(1)}}^T, {S^{(2)}}^T)$. Hence, the mean parameter is also partitioned into $\mu^{(1)} = E_\theta S^{(1)}$, $\mu^{(2)} = E_\theta S^{(2)}$. It is known that $(\mu^{(1)}, \theta^{(2)})$ or $(\mu^{(2)}, \theta^{(1)})$ uniquely parameterizes the distribution and each of them is called a *mixed parameter*.

Arnold & Gokhale (1994) proved that two two-dimensional discrete distributions have the same local odds ratios if and only if they have identical uniform marginal representations. This result can be generalized and its proof is a direct consequence of mixed parameterization. Let A and B be two J-dimensional multinomial distributions of the same size whose mixed parameters are $(\mu_A^{(1)}, \theta_A^{(2)})$ and $(\mu_B^{(1)}, \theta_B^{(2)})$, respectively, where $\mu_A^{(1)}$ and $\mu_B^{(1)}$ are the J one-dimensional marginal densities of A and B, respectively. Also, let $\mu_U^{(1)}$ represent the J uniform one-dimensional distributions. Then it is immediate the two J-dimensional distributions characterized by $(\mu_U^{(1)}, \theta_A^{(2)})$ and $(\mu_U^{(1)}, \theta_B^{(2)})$ are identical if and only if $\theta_A^{(2)} = \theta_B^{(2)}$. Moreover, the result holds for any J one-dimensional distributions, i.e., there is no need to restrict $\mu^{(1)}$ to J uniform marginals.

Example 1(continued). For a 2×3 table, partition s into $s^{(1)} = \{y_{+1}, y_{+1}+y_{+2}, y_{1+}, y_{++}\}^T$ and $s^{(2)} = \{y_{11}, y_{11}+y_{12}\}^T$. Then $\mu^{(1)} = \{\pi_{+1}, \pi_{+1}+\pi_{+2}, \pi_{1+}, 1\}^T$, the marginal probabilities, and $\theta^{(2)} = \{\bigtriangledown_1 \bigtriangledown_2 \log \pi_{11}, \bigtriangledown_1 \bigtriangledown_2 \log \pi_{12}\}^T$, the local log odds ratios, form a mixed parameter. This form of mixed parameter is used by Goodman (1979) to model $r \times c$ contingency tables of ordered variables.

Following the notations in (2), let the component in s be denoted by $s_k, k = 1, \cdots, K$ and define $s_i \wedge s_j$as the sum of y_l that is common to both s_i and s_j. Let R denote a subset of $\{s_k, k = 1, \cdots, K\}$, R is said to be closed under $\wedge$ if and only if for all $s_i, s_j \in R$, we have $s_i \wedge s_j \in R$.

We first state and prove the following Lemma.

Lemma 3.1. *Let $(s^{(1)}, s^{(2)})$ be a partition of s and $\mu^{(2)} = Es^{(2)}$. If $s^{(2)}$ is closed under $\wedge$, then cov$(s_i^{(2)}, s_k^{(2)})$ depends only on $\mu^{(2)}$.*

Proof. Let $\mu_k^{(2)} = A_k^T\pi$, where A_k is the k-th column of A.

$$\begin{aligned}
\mathrm{cov}(s_i^{(2)}, s_k^{(2)}) &= \mathrm{cov}(A_i^T y, A_k^T y) = A_i^T \mathrm{cov}(y, y) A_k \\
&= A_i^T(\mathrm{diag}\pi - \pi\pi^T)A_k = A_i^T(\mathrm{diag}\pi)A_k - A_i^T\pi\pi^T A_k \\
&= \sum_l a_{il}\pi_l a_{kl} - (\pi^T A_i)^T(\pi^T A_k),
\end{aligned}$$

where a_{kl} denotes the l-th element of A_k, $l = 1, \cdots, K$. But $\sum a_{il}\pi_l a_{kl}$ equals the sum of elements of π common to $s_i^{(2)}$ and $s_k^{(2)}$, so does $(\pi^T A_i)^T(\pi^T A_k)$, which only selects π_j^2 common to both A_i and A_k. The fact that $\mathrm{cov}(s_i^{(2)}, s_k^{(2)})$ depends on $\mu^{(2)}$ follows directly from the given that $s^{(2)}$ is closed under $\wedge$.

Theorem 3.1. *If $s^{(2)}$ is closed under $\wedge$, then $s^{(2)}$ is a cut.*

Proof. Theorem 3.1 of Barndorff-Nielsen and Koudou (1995) states several equivalent conditions of a statistic $s^{(2)}$ being a cut. Among them is the condition that

$V_{22}(\mu_1, \mu_2) = \text{cov}(s^{(2)}, s^{(2)})$ does not depend on $\mu^{(1)}$. This, together with lemma 3.1, furnishes the proof.

Example 1.(continued). Let $\theta^{(1)} = \bigtriangledown_2 \log \pi_{21}$ and $\theta^{(2)} = \theta \backslash \theta^{(1)}$. Then $s^{(2)} = \{y_{11}, y_{11} + y_{12}, y_{+1} + y_{+2}, y_{1+}, y_{++}\}^T$ is close under $\wedge$. Hence it is cut for $\theta^{(1)}$. For $\theta^{(1)} = (\bigtriangledown_{12} \log \pi_{11}, \bigtriangledown_{12} \log \pi_{12})$, the corresponding $s^{(2)}$ is not a cut. But for $(\bigtriangledown_{12} \log \pi_{11}, \bigtriangledown_2 \log \pi_{21})$, the corresponding $s^{(2)}$ is a cut.

Corollary 3.1. *If $s^{(1)}$ contains all J one-way cumulative marginal sums, then $s^{(2)}$ is a cut.*

Proof. It suffices to show that $s^{(2)}$ is closed under $\wedge$. But any component in $s^{(2)}$ has the form $\triangle_{J \cdots 1} y_{k_1 \cdots k_J}$, where $1 \leq k_j \leq K_j - 1$. The intersection of any two components of $s^{(2)}$ always results in a statistic of the form $\triangle_{J \cdots 1} y_{k_1 \cdots k_J}$, where the resulting k_j, $1 \leq k_j \leq K_j - 1$, is the smaller of the two k_j's. That is, their intersection stays in $s^{(2)}$.

This corollary justifies modeling the J one-dimensional distributions conditioned on the dependence structure represented by odds ratios, see Liang & Zeger (1986).

Corollary 3.2. *If $s^{(1)}$ contains a subset of all J one-way cumulative marginal sums, then $s^{(1)}$ is a cut if and only if it only contains cumulative marginal sums of one dimension.*

Proof. The "if" part is straightforward as the cumulative one-dimensional marginal sums are closed under $\wedge$. For the "only if" part, we suppose that $s^{(1)}$ contains cumulative marginal means of two dimensions. That is, $s^{(1)}$ contains components of the form $\triangle_{J \cdots 1} y_{k_1 \cdots k_J}$ where at least one k_j is K_j. But any intersection from the two dimensions must be of the form $\triangle_{J \cdots 1} y_{k_1 \cdots k_J}$, where $1 \leq k_j \leq K_j - 1$. Such elements belong to $s^{(2)}$. From the proof of Lemma 3.1, it can be seen that this implies that $\text{cov}(s^{(1)}, s^{(1)})$ depends on $\mu^{(2)}$. By Theorem 3.1 of Barndorff-Nielsen and Koudou (1995), $s^{(1)}$ is not a cut.

4. Conclusion

In order to deliver cuts of common interest, multinomial likelihood is parameterized into its standard form. When the parameter of interest is expressed in terms of a partition of the standard canonical parameters, Theorem 3.1 of this paper provides a useful tool to find a cut. If a cut exists, inference conditioned on cut is, in general, more efficient than the inference conditioned on the sampling total. Related use of cuts can be found in Barndorff-Nielsen & Koudou (1995).

References

Agresti, A. (1990), *Categorical Data Analysis*, Wiley, New York.

Arnold, B.C. and Gokhale, D.V. (1994), On Uniform Marginal Representation of Contingency Tables. *Statistics and Probability Letters* 21,311-16.

Barndorff-Nielsen, O.E. (1978), *Information and Exponential Families in Statistical Theory*, John Wiley, New York.

Barndorff-Nielsen, O.E., and Cox, D.R. (1994), *Inference and Asymptotics*, Chapman and Hall, London.

Barndorff-Nielsen, O.E., and Koudou, A. E. (1995), Cuts in Natural Exponential Families, *Theory of Probability and Its Applications* 40, 361-372.

Cox, D.R., and Reid, N. (1987), Parameter Orthogonality and Approximate Conditional Inference (with discussion), *Journal of the Royal Statistical Society*, Ser. B 49, 1-39.

Ip, E.H., Wang, Y., de Boeck, P. and Meulders, M. (2003). Locally Dependent Latent Trait Model for Polytomous Response. Technical Report, Marshall School of Business, University of Southern California. (A version of it is submitted and is currently under review).

Goodman, L.A. (1979), Simple Models for the Analysis of Association in Cross-classifications Having Ordered Categories, *Journal of the American Statistical Association* 74, 537-552.

Holland, P., and Wang, Y. (1987), Dependence Functions for Continuous Bivariate Densities, *Communications in Statistics – Theory and Methods* 16, 863-876.

Jones, M. C. (1996), The Local Dependence Function, *Biometrika* 83, 899-904.

Liang, K.Y., and Zeger, S.L. (1986), Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika* 73, 13-22.

Wang, Y., (1986), Ordered-dependent Parameterization of Multinomial Distributions, *Scandinavian Journal of Statistics*, 13, 199-205.

Whittaker, J. (1990), *Graphical Models in Applied Mathematical Multivariate Statistics*, Wiley, West Sussex, England.

Addresses:

Edward H. Ip

Biostatistics, Department of Public Health Sciences

Wake Forest University School of Medicine

Winston-Salem, NC 27157, USA

Email: eip@wfubmc.edu

Yuchung J. Wang

Department of Mathematical Sciences

Rutgers University

Camden, NJ 08102 USA

Email:yuwang@crab.rutgers.edu