Equivalence of Simultaneous ML and Two-step ML in Estimating Polychoric Correlation Coefficient

Yuchung J. Wang*
Rutgers University
Camden, N. J. USA

* This research was sponsored by the National Science Council, Taiwan, Republic of China. Request for reprints should be sent to Yuchung J. Wang, Department of Mathematical Sciences, Rutgers University, Camden, New Jersey 08102 USA

ABSTRACT

When qualitative data are tabulated in a two-way contingency table with ordered rows and columns, a possible latent model is the bivariate normal distribution. The maximum likelihood (ML) estimation of the bivariate normal correlation coefficient presents a challenging computational problem, see Olsson (1979), Olsson, Drasgow and Dorans (1982) and Lee and Poon (1986). Their Newton-type algorithms usually involve the first and the second order derivatives of the likelihood function.

Here we propose an algorithm based on integration instead of derivatives. We show that simultaneous ML is equivalent to a two-step ML. In the first step, the ML points of polytomy are estimated by matching the expected with the observed. Then, the likelihood becomes a function of the polychoric correlation. We show that the above likelihood is strongly unimodal. The second optimization step is based on Kiefer's (1953) search algorithm for a continuous unimodal function. The rate of convergence is $2(0.618)^n$. We consider this approach optimal because of the two-step decomposition and the exponential rate of convergence.

The same algorithm can be used to compute other minimum-distance estimates (MDE) for polychoric correlation. Several MDE, including the minimum Hellinger distance estimate, are compared. The two-step procedure can also be used to compute the ML estimate of the correlation parameter for Gumbel's bivariate exponential distribution and Morgenstern's family of bivariate distributions.

Some Key Words: Ordered Contingency tables; bivariate normal distribution; normal correlation; Kullback-Lieber information distance; Kiefer's search algorithm; maximum likelihood estimates; minimum-distance estimates and Morgenstern's family of distributions.

1. Introduction

In behavioral and social sciences, bivariate qualitative data are often recorded via a two-way contingency table with r ordered rows and c ordered columns. Let n_{ij} be the observed cell frequency of ith row and jth column, where $1 \leq i \leq r$ and $1 \leq j \leq c$. The correlation between row and column classifications is called the *polychoric correlation* . It is almost always to assume that a pair of unobservable bivariate normal (latent) variables give rise to $\{n_{ij} : 1 \leq i \leq r \text{ and } 1 \leq j \leq c\}$ and their correlation is the polychoric correlation. Martinson and Hamdan (1971) proposed an algorithm to compute the maximum likelihood estimate (MLE) of the underlying normal correlation coefficient ρ. To apply their algorithm, it is necessary to compute the derivatives of the tetrochoric series for cell probabilities. Martinson and Hamdan also discussed minimum-distance estimates (MDE) for ρ and compared them with the MLE. Note that MLE is a MDE if Kullback-Lieber information is used as the distance.

Olsson (1979) pointed out two possible approaches to estimate ρ: (1) the method of simultaneous ML estimates ρ and the marginal thresholds simultaneously; and (2) two-step sequential approach which first computes the thresholds and then estimate ρ given the thresholds. He claimed that the latter approach has the advantage of greater ease in numerical computation, although the former is more correct. We prove that two-step ML is equivalent to simultaneous ML. Based on a parameterization of multinomial likelihood due to Wang (1987b), it is shown that marginal matching always reduces the Kullback-Lieber information. Conditioned on a given value of ρ, the first ML step matches the expected marginal distributions with the observed marginal distributions. The second ML step finds the value of ρ which minimizes Kullback-Lieber information.

The degree of freedom (number of independent parameters) is used to illuminate

the decomposition of the two- step approach. For an r by c ordered contingency table, the total degree of freedom is $rc-1$. The bivariate normal assumption reduces the degree of freedom to $r+c-1$, of which $r+c-2$ parameters pertain to the marginal distributions and the remaining one degree of freedom pertains to ρ. Hence, after the first ML step, the degree of freedom is reduced to one. Based on results of Burridge (1982), strong unimodality of the discrete likelihood is obtained. Kiefer (1953) proposed a sequential search algorithm for finding the maximum or minimum of a continuous unimodal function over a closed interval. This algorithm is employed to find the ML estimate of ρ.

The intrinsic robustness of minimum Hellinger distance estimators have generated considerable interest in its estimation, see Donoho and Liu (1988), Tamura and Boos (1986), and Eslinger and Woodward (1991). The above two-step minimization can be used to compute a variety of MDE of ρ satisfying the marginal constraints. Replacing likelihood by information distance makes it possible to compute MLE and MDE via the same computer program.

Sometimes the context of the data or the instrument used to collect psychometrical data deems that the marginal distributions of the latent variables follow distributions other than normal. For example, past test item response patterns might prefer univariate Cauchy to univariate normal. The same two-step procedure is applicable as long as the latent distribution reduces the number of parameters from $rc-1$ to $r+c-1$ and the marginals are l-independent of the correlation. The concept of l-independent was first defined by Barndorff-Nielsen (1978, P.26). In section 4, we describe how to modify our two-step algorithm for computing MLE and MDE when the latent variables belong to Morgenstern's family of distributions.

In Section 2, we will give a discussion about the estimation problem and its two-step decomposition. It includes an introduction to Kiefer's algorithm. In section 3,

several examples are presented. In section 4, the sufficient conditions for the two-step decomposition are discussed.

2. Equivalence of Simultaneous ML and Two-step ML

For an $r \times c$ ordered cross-classified contingency table, let n_{ij} denote the observed frequency in the ith row and the jth column and let N_{ij} denote the corresponding expected frequency under some model. Let $\phi(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ denote the density of a bivariate normal distribution with means μ_1 and μ_2, standard deviations σ_1 and σ_2, and correlation coefficient ρ. Assuming that N_{ij} are the results of categorizing $\phi(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, then N_{ij} is set to equal to $n_{++}p_{ij}$ where ,

$$p_{ij} = \int_{a_{i-1}}^{a_i} \int_{b_{j-1}}^{b_j} \phi(x, y; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) dx dy, \tag{2.1}$$

and $n_{++} = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}$ is considered fixed. Here, the two sequences $\{-\infty = a_0 < a_1 < \cdots < a_r = \infty\}$ and $\{-\infty = b_0 < b_1 < \cdots < b_{c-1} < b_c = \infty\}$ are, respectively, the thresholds of categorization or *points of polytomy* for the x-margin and the y-margin. The integral in (2.1) can also be written as

$$p_{ij} = \int_{\alpha_{i-1}}^{\alpha_i} \int_{\beta_{j-1}}^{\beta_j} \phi(x, y; 0, 0, 1, 1, \rho) dx dy \tag{2.2}$$

by transforming a_i and b_j into $\alpha_i = (a_i - \mu_1)/\sigma_1$ and $\beta_j = (b_j - \mu_2)/\sigma_2$, respectively. Equation (2.2) implies, without loss of generality, that we can assume that the latent variables follow standard bivariate normal distribution with correlation coefficient ρ. The bivariate normal model reduces the degree of freedom from $rc - 1$ to $r + c - 1$, of which $(r - 1)$ are for $\{\alpha_i\}, (c - 1)$ are for $\{\beta_j\}$, and 1 is for ρ. Since the marginal distributions of $\phi(x, y; 0, 0, 1, 1, \rho)$ are univariate standard normal distributions, we have

$$p_{i+} \quad = \quad \sum_{j=1}^{c} p_{ij} = \int_{\alpha_{i-1}}^{\alpha_i} (1/\sqrt{2\pi}) e^{-x^2/2} dx, \tag{2.3}$$

$$p_{+j} = \sum_{i=1}^{r} p_{ij} = \int_{\beta_{j-1}}^{\beta_j} (1/\sqrt{2\pi}) e^{-y^2/2} dy. \tag{2.4}$$

From a contingency-table perspective, the expected frequencies $\{N_{ij}\}$ are uniquely determined by the following sets of parameters:

a) the expected row marginals: $\{N_{i+} = \sum_{j=1}^{c} N_{ij} = n_{++} \sum_{j=1}^{c} p_{ij} = n_{++} p_{i+}, \quad \text{for } 1 \leq i \leq (r-1)\}$;

b) the expected column marginals: $\{N_{+j} = \sum_{i=1}^{r} N_{ij} = n_{++} \sum_{i=1}^{r} p_{ij} = n_{++} p_{+j}, \quad \text{for } 1 \leq j \leq (c-1)\}$; and

c) the expected local cross-product ratios (lcpr): $\{\gamma_{ij} = N_{ij} N_{i+1\ j+1} / N_{i+1\ j} N_{ij+1}, \text{ for } 1 \leq i \leq (r-1) \quad \text{and} \quad 1 \leq j \leq (c-1)\}$.

For fixed n_{++}, equations (2.3) and (2.4) imply that parameters $\{N_{i+}\}$ and $\{N_{+j}\}$ determine the points of polytomy, $\{\alpha_i\}$ and $\{\beta_j\}$, and vice versa. Moreover, the determination of the points of polytomy is independent of the correlation.

Let $D(n_{ij}, N_{ij})$ denote some kind of distance between the observed data $\{n_{ij}\}$ and the expected distribution $\{N_{ij}\}$. For ML estimation, $D(n_{ij}, N_{ij})$ is the Kullback-Lieber information distance:

$$\begin{aligned} G^2(n_{ij}, N_{ij}) &= \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} ln(n_{ij}/N_{ij}) \\ &= \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} ln(n_{ij}) - \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} ln(N_{ij}). \end{aligned} \tag{2.5}$$

ML is equivalent to minimum G^2 because the second term of G^2 is the log likelihood and the first term depends on the data only. Moreover, using information distance instead of likelihood, makes it possible to compute the ML estimate and the MDE in one computer program.

The objective function G^2 is an explicit function of $\{N_{ij}\}$ but an implicit function of $\{\alpha_i\}, \{\beta_j\}$ and ρ. Let N_{ij} be expressed as $N_{ij}(\{\alpha_k\}, \{\beta_l\}, \rho) = n_{++} p_{ij}(\{\alpha_k\}, \{\beta_l\}, \rho)$

and $G^2(n_{ij}, N_{ij})$ as $G^2(\{\alpha_k\}, \{\beta_l\}, \rho)$. Simultaneous ML is to find $\{\hat{\alpha}_k\}, \{\hat{\beta}_l\}$ and $\hat{\rho}$ such that

$$G^2(\{\hat{\alpha}_k\}, \{\hat{\beta}_l\}, \hat{\rho}) = \min_{\{\alpha_k\}, \{\beta_l\}, \rho} G^2(\{\alpha_k\}, \{\beta_l\}, \rho), \tag{2.6}$$

where $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_r = \infty, -\infty = \beta_0 < \beta_1 < \cdots < \beta_c = \infty$ and $-1 \leq \rho \leq 1$.

To articulate our two-step approach, the following sets are defined:

$$\begin{aligned} \Omega &= \{N_{ij} : N_{ij} > 0, N_{ij} = n_{++} p_{ij} \quad \text{and} \quad p_{ij} \quad \text{satisfying} \quad (2.2)\}, \\ A &= \{N_{ij} : N_{i+} = n_{i+} \quad \text{and} \quad N_{+j} = n_{+j}\}, \quad \text{and} \end{aligned}$$

Let $\{N_{ij}^{\#}(\rho)\}$ be obtained from $\{N_{ij}(\{\alpha_k\}, \{\beta_l\}, \rho)\}$ by selecting $\{\alpha_k^{\#}\}$ and $\{\beta_l^{\#}\}$ such that

$$\int_{\alpha_{i-1}^{\#}}^{\alpha_i^{\#}} (1/\sqrt{2\pi}) e^{-x^2/2} dx = n_{i+}/n_{++} \quad \text{and} \quad \int_{\beta_{j-1}^{\#}}^{\beta_j^{\#}} (1/\sqrt{2\pi}) e^{-y^2/2} dy = n_{+j}/n_{++}. \tag{2.7}$$

Theorem 2.1 For any $\{N_{ij}\}$ in Ω, conditioned on $\{\gamma_{ij}\}$, $G^2(n_{ij}, N_{ij})$ is minimizes by matching the expected marginals with the observed marginals, i.e.

$$G^2(n_{ij}, N_{ij}) \geq G^2(n_{ij}, N_{ij}^{\#}) \tag{2.8}$$

when $N_{ij}^{\#} N_{i+1\ j+1}^{\#} / N_{i+1\ j}^{\#} N_{ij+1}^{\#} = N_{ij} N_{i+1\ j+1} / N_{i+1\ j} N_{ij+1}$. The above inequality is strict unless $N_{ij} = N_{ij}^{\#}$ for all i and j.

Proof: Wang (1987b) proposed an order-dependent parameterization for the multinomial likelihood. For a two-way table with ordered rows and columns, the likelihood is expressed as parameters of cumulative marginals and logarithms of lcpr. From Example 3.4 of Wang (1987b, P.204), the Kullback-Lieber distance can be written as follows:

$$
\begin{aligned}
G^2(n_{ij}, N_{ij}) &= \sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} ln(n_{ij}) - n_{++} ln(n_{++}) + n_{++} a(\theta) - n_{++}\theta_{11} \\
&- \sum_{j=2}^{c}(\sum_{l=j}^{c} n_{+l})\theta_{1j} - \sum_{i=2}^{r}(\sum_{k=i}^{r} n_{k+})\theta_{i1} - \sum_{i=2}^{r}\sum_{j=2}^{c}(\sum_{k=i}^{r}\sum_{l=j}^{c} n_{kl})\theta_{ij}, \quad (2.9)
\end{aligned}
$$

where $\theta_{i+1j+1} = ln(\gamma_{ij})$ for $1 \leq i \leq (r-1)$ and $1 \leq j \leq (c-1)$ and $\theta = (\theta_{11}, \cdots, \theta_{rc})$ is the natural parameter of the exponential family. Moreover, the moment-generating function and the cell probabilities are as follows:

$$
a(\theta) = ln(\sum_{i=1}^{r}\sum_{j=1}^{c} exp(\sum_{k=1}^{i}\sum_{l=1}^{j}\theta_{kl})) \text{ and} \quad (2.10)
$$

$$
p_{ij} = exp(\sum_{k=1}^{i}\sum_{l=1}^{j}\theta_{kl}) / \sum_{i=1}^{r}\sum_{j=1}^{c} exp(\sum_{k=1}^{i}\sum_{l=1}^{j}\theta_{kl}) \quad (2.11)
$$

When γ_{ij} are fixed, distance G^2 is minimized by taking the derivative with respect to the rest of the parameters. Since (2.9) belongs to the exponential family, the derivatives are the likelihood equations which equate the expected with the observed, i.e.

$$
\begin{aligned}
E(\theta_{1j}) &= \partial a(\theta)/\partial\theta_{1j} = \sum_{l=j}^{c} n_{+l}/n_{++}, \text{ and} \\
E(\theta_{i1}) &= \partial a(\theta)/\partial\theta_{i1} = \sum_{k=i}^{r} n_{k+}/n_{++},
\end{aligned}
$$

for $2 \leq j \leq c$ and $2 \leq i \leq r$. From (2.10) and (2.11), we can show that $\partial a(\theta)/\partial\theta_{1j} = \sum_{l=j}^{c} p_{+l}$ and $\partial a(\theta)/\partial\theta_{i1} = \sum_{k=i}^{c} p_{k+}$. The likelihood equations become

$$
\begin{aligned}
\sum_{l=j}^{c} p_{+l} &= \sum_{l=j}^{c} n_{+l}/n_{++}, \text{ and} \\
\sum_{k=i}^{r} p_{k+} &= \sum_{k=i}^{r} n_{k+}/n_{++},
\end{aligned}
$$

which are equivalent to (2.7). Hence, the theorem is proved.

Simultaneous ML is to search the minimum of G^2 in Ω but equation (2.8) suggests that it is sufficient to limit the search to $N_{ij}^{\#}(\rho)$ in A. Let $f(\rho)$ be $G^2(n_{ij}, N_{ij}^{\#})$. The second step is to minimize $f(\rho)$ with respect to ρ.

Since Kiefer's search algorithm is applicable to unimodal function, we need to prove the following:

Theorem 2.2 The likelihood $\sum n_{ij} ln N_{ij}^{\#}$ is strongly unimodal and function $f(\rho)$ is strictly convex.

Proof: Since the bivariate normal is assumed to have unit variances, the strict unimodality of $\sum n_{ij} ln N_{ij}^{\#}$ is stated as a direct consequence of Prekopa (1973) theorem in Burridge (1982, P.149). Prekopa's theorem states that when the region of integration can be indexed by a parameter, say ρ, then the integral, say $N_{ij}^{\#}(\rho)$, is strongly unimodal with respect to that parameter if and only if the density function is strongly unimodal.

The conditional distribution of y given x is a univariate normal with mean ρx and variance $(1-\rho^2)$. This representation suggests that $N_{ij}^{\#}$, the integral of $\phi(x, y; 0, 0, 1, 1, \rho)$ over $(\alpha_{i-1}^{\#}, \alpha_i^{\#}) \times (\beta_{j-1}^{\#}, \beta_j^{\#})$ can be transformed into an integral of $\phi(x, y; 0, 0, 1, 1, 0)$ over the parallelogram cornered by $(\alpha_{i-1}^{\#}, (\beta_{j-1}^{\#} - \rho\alpha_{i-1}^{\#})/\sqrt{(1-\rho^2)}), (\alpha_{i-1}^{\#}, (\beta_j^{\#} - \rho\alpha_{i-1}^{\#})/\sqrt{(1-\rho^2)}), (\alpha_i^{\#}, (\beta_{j-1}^{\#} - \rho\alpha_i^{\#})/\sqrt{(1-\rho^2)}), (\alpha_i^{\#}, (\beta_j^{\#} - \rho\alpha_i^{\#})/\sqrt{(1-\rho^2)})$. Because $\phi(x, y; 0, 0, 1, 1, 0)$ is strongly unimodal, the integral $N_{ij}^{\#}$ and, consequently, the likelihood itself are strongly unimodal in ρ. Hence, $f(\rho)$ is strictly convex.

Burridge's theorem is applicable only to a discretized likelihood whose partitions are fixed. Hence, his results can not be used to prove the strong unimodality for simultaneous ML because the partition grids are to be estimated. The two-step sequential decomposition makes the proof of unimodality feasible.

Kiefer's algorithm is a version of the golden search. Let $g(x)$ be a unimodal continuous mapping from unit interval, $I = [0, 1]$, and the goal is to find the minimum

for $g(x)$. Let $x_2 = 1 - x_1 = -\frac{1}{2} + \sqrt{5}/2 \cong 0.618$. These two numbers: $x_2 = 0.618$ and $x_1 = 0.382$ in I are the initial search points. If $g(x_1) \leq g(x_2)$ then define $v(x) = x/0.618, y = v(x)$ and $g^*(y) = g(v^{-1}(y))$ for $y \in I$. Putting $y_2 = v(x_1)$ and $y_1 = 0.382$, the algorithm repeats itself using y as the variable and $g^*(y)$ as the function to be minimized over I. If $g(x_1) > g(x_2)$ then define $v(x) = (x - 0.382)/0.618, y = v(x)$ and $g^*(y) = g(v^{-1}(y))$ for $y \in I$. Letting $y_1 = v(x_2)$ and $y_2 = 0.618$ the algorithm repeat itself using y as the new variable and $g^*(y)$ as the function to be minimized over I. In essence, Kiefer's algorithm is to initially select $x_1 = 0.382$ and $x_2 = 0.618$ and compare $g(x_1)$ and $g(x_2)$. For minimization purpose, when $g(x_1) \leq g(x_2)$ then interval $(x_2, 1)$ is discarded and interval $[0, x_2]$ is scaled to $[0, 1]$. In this "new" $[0, 1]$, the previous x_1 becomes $x_2(0.618)$ and a new $x_1(0.382)$ is selected for the next round of comparison. When $g(x_1) > g(x_2)$, interval $(0, x_1)$ is discarded and $[x_1, 1]$ is scaled to $[0, 1]$ by first subtracting x_1 and then dividing by $(1 - x_1)$. In the "new" $[0, 1]$, the previous x_2 becomes $x_1(0.382)$ and a new $x_2(0.618)$ is selected for the next round of iteration. Because a segment of length 0.382 is discarded after each iteration, the length of the interval containing the minimum becomes $(0.618)^n$ after n iterations. Coordinate x in I can be transformed to coordinate y in $[-1, 1]$ via transformation $y = 2x - 1$. In this case, the rate of convergence is $2(0.618)^n$ because the length of the interval $[-1, 1]$ is 2.

The operating sequence is summarized as follows. Equation (2.7) is solved to find the points of polytomy $\{\alpha_i^{\#}\}$ and $\{\beta_j^{\#}\}$. For a fixed ρ, numerical integrations are performed to find $\{p_{ij}(\rho)\}$, and consequently $\{N_{ij}^{\#}(\rho)\}$, via equation (2.2). Then, $f(\rho)$ is computed and minimized using Kiefer's algorithm.

To consider other MDE, some commonly used distance functions are as follows:

a) the minimum Chi-square estimator (MCS) is obtained by minimizing the Pear-

son's X^2 defined as

$$X^2(n_{ij}; N_{ij}) = \sum_{i=1}^{r} \sum_{j=1}^{c} (n_{ij} - N_{ij})^2 / N_{ij};$$

b) the modified minimum chi-square estimator (MMCS) is obtained by minimizing the Neyman-modified NM^2 defined as

$$NM^2(n_{ij}, N_{ij}) = \sum_{i=1}^{r} \sum_{j=1}^{c} (n_{ij} - N_{ij})^2 / n_{ij}; \text{ and}$$

c) the minimum Hellinger distance estimator (MHD) is obtained by minimizing the Hellinger distance H^2 defined as

$$H^2(n_{ij}, N_{ij}) = 2 - 2 \sum \sum (n_{ij} N_{ij})^{1/2} / n_{++}.$$

For distance functions other than G^2, there is no formal proof that inequality similar to (2.8) holds. However, the approach described above can still be used to find the MDE of ρ constrained by (2.7). Let $f(\rho)$ be $D(n_{ij}; N_{ij}(\{\alpha_k^{\#}\}, \{\beta_l^{\#}\}, \rho))$, where $D(n_{ij}; N_{ij})$ represents one of the above distances. The MDE ρ_D satisfies the following equation:

$$f(\rho_D) = \min_{-1 \le \rho \le 1} D(n_{ij}; N_{ij}(\{\alpha_k^{\#}\}, \{\beta_l^{\#}\}, \rho)).$$

3. Examples

Table 3.1, Table 3.2 and Table 3.3 are analyzed by Martinson and Hamdan (1971). Kiefer's search algorithm is used to find the minimum for $G^2(n_{ij}, N_{ij}^{\#}(\rho))$, $X^2(n_{ij}, N_{ij}^{\#}(\rho))$, $NM^2(n_{ij}, N_{ij}^{\#}(\rho))$ and $H^2(n_{ij}, N_{ij}^{\#}(\rho))$, where $N_{ij}^{\#}(\rho) = N_{ij}(\{\alpha_k^{\#}\}, \{\beta_l^{\#}\}, \rho)$ and $\{\alpha_k^{\#}\}$ and $\{\beta_l^{\#}\}$ are computed via (2.7). The results of four MDE and its corresponding minimum distance are listed in Table 3.4. The distances for the independence model ($\rho = 0$) are also listed for comparison. In all four instances, the extra

degree of freedom due to ρ significantly improved the fit. The bivariate normal model provides a near perfect fit for the 2×3 Table 3.2 and the 3×3 Table 3.3.

Compared with the results reported in Martinson and Hamdan (1971), there are slight differences. The minimum obtained via Kiefer's algorithm is always smaller than the distances evaluated at Martinson and Hamdan's estimates. Kiefer's algorithm not only provides a more accurate approximation but also converges at exponential rate.

Martinson and Hamdan(1971), Olsson (1979) and Olsson, Drasgow and Dorans (1982) among others employed a Newton-Raphson algorithm to find MLE for ρ. Their approaches require the approximation of the derivative $dN_{ij}/d\rho$ and $d^2N_{ij}/d\rho^2$ numerically. Our approach requires numerical integration to evaluate $\{N_{ij}^{\#}\}$. There are many computational algorithms available to calculate the bivariate normal integrals. A recent article by Terza and Welland (1991) compared these algorithms. We used a program published by Dunnett (1989). In general, numerical integration achieves better accuracy than numerical differentiation. Also, because numerical integrations compute the expected cell frequencies $\{N_{ij}\}$, which gives us an opportunity to examine the normal model microscopically (cell-by-cell). Derivative-based algorithms computes only the estimates, not the cell estimates. Consequently, those algorithms can not report the minimal values of the distance functions.

When the fits are nearly perfect, all four estimates are identical. When discrepancy among different MDE appears, MLE and MCS are usually the smallest estimates (conservative) and MMCS is the largest estimate, whereas MHD is in the middle. This seems to indicate the robustness of MHD. One reason to compute and examine several different MDE's is to check the validity of the latent model. Agresti (1990, p.472) pointed out that "when the model holds, different estimation methods yield nearly identical estimates when the sample size is large". Heuristically, the bivariate latent

model holds for Table 3.3 and 2×3 Table 3.2.

4. Discussions and extensions:

There have been many arguments comparing simultaneous optimization with two-step optimization in ML estimation of ρ. Olsson's (1979) numerical results indicated that two-step ML provided slightly more accurate estimates than simultaneous MLE, though the differences were negligible. Lee and Poon(1986) discussed the advantages and disadvantages of these two approaches, but gave no conclusive arguments. In this paper, the equivalence between them is formally established. This equivalence is derived from three facts: First, without loss of generality, the means and standard deviations of the bivariate normal can be assumed to be zero and one, respectively. Second, matching the expected marginals with the observed marginals always reduces the information distance. Third, the G^2-optimal points of polytomy conditioned on ρ are independent of ρ, which is due to the l-independence between ρ and the marginal distributions for the standard bivariate normal, see Goodman (1981) and Wang (1987a).

We consider our algorithm optimal because of the two-step decomposition and the exponential rate of convergence. After 30 iterations, Kiefer's algorithm computes ρ with a maximum error of $0.00237(2(0.618)^{30})$. This kind of predetermined precision is not available for Newton-type approximations. For distance functions other than the information distance, the same algorithm can be used to compute MDE of polychoric correlation constrained by (2.7). Equating the expected with the observed marginals may not be unreasonable when the focus is the correlation. The results reported in Section 3 are computed by the same computer program incorporating different distance functions.

Two-step ML is applicable to other latent models whenever the latent model

makes the determination of the points of polytomy independent of the polychoric correlation. This l-independence property is not unique to the bivariate normal distribution. Gumbel's bivariate exponential has its marginal exponential distributions l-independent of the correlation. Morgenstern's family is another rich group of bivariate distributions with this property. For any given pair of univariate densities $f_1(x)$ and $f_2(y)$, Morgenstern's family has its joint density defined by

$$h(x, y; \alpha) = f_1(x) f_2(y)[1 + 2\alpha\{2F_1(x) - 1\}\{2F_2(y) - 1\}], \tag{4.1}$$

where $F_1(x)$ and $F_2(y)$ are the distribution functions of $f_1(x)$ and $f_2(y)$, respectively. Parameter α is the association parameter. Because $f_1(x)$ and $f_2(y)$ are l-independent of α, the same procedure can be used to compute MLE and MDE for α. The only modification is to replace integration of $\phi(x, y; 0, 0, 1, 1, \rho)$ by integration of $h(x, y; \alpha)$ in (4.1) and to determine points of polytomy using $F_1(x)$ and $F_2(y)$ instead of the standard normal distribution. This opens the possibility to consider other latent models for the two-dimensional discrete data, such as Gumbel's bivariate exponential and the bivariate Cauchy. When the psychometrical instrument deems normal marginals inappropriate, model (4.1) provides an alternative for the polychoric correlation.

Acknowledgment: The author benefited from helpful discussions with Dr. George Wong of Strang Cancer Prevention Center. Theorem 2.2 was suggested by a referee.

Table 3.1

First and second lambing records

	1952			
1953	No lambs	1 lamb	2 lambs	Totals
No lambs	58	52	1	111
1 lamb	26	58	3	87
2 lambs	8	12	9	29
Totals	92	122	13	227

Table 3.2

Pearson's table for the brother-sister head length (millimeters) reduced to 3×3 and 2×3 tables

	Brother			
Sister	163.5-175.5	177.5-189.5	191.5-205.5	Totals
159.5-171.5	43	65	2	110
173.5-183.5	53	425	50	528
185.5-197.5	8	97	52	157
Totals	104	587	104	795
159.5-177.5	77	265	21	363
179.5-197.5	27	322	83	432
Totals	104	587	104	795

Table 3.3

Pearson's table for brother-sister head breadth

(millimeters) reduced to 3×3 tables

	Brother			
Sister	122.5-136.5	138.5-146.5	148.5-160.5	Totals
121.5-133.5	40.5	58	9	107.5
135.5-145.5	52.5	340.5	143.5	536.5
147.5-159.5	1	36.5	77.5	115
Totals	94	435	230	759

Table 3.4

Comparisons of minimum-distance estimates of the bivariate normal correlation

	Table 3.1	Table 3.2		Tables 3.3
Size of the table	3×3	3×3	2×3	3×3
Estimate(minimum distance)				
MLE (G^2)	0.425(11.56)	0.498(7.84)	0.425(0.085)	0.565(2.53)
MCS (X^2)	0.431(11.79)	0.481(9.13)	0.425(0.073)	0.557(2.717)
MMCS (NM^2)	0.476(12.88)	0.553(5.50)	0.425(0.074)	0.576(2.097)
MHD (H^2)	0.446(0.0129)	0.523(0.0023)	0.425(0.00004)	0.565(0.000)
Distance for the independence hypothesis ($\rho = 0$)				
G^2	35.59	117.22	64.15	144.30
X^2	49.64	141.90	61.00	159.14
NM^2	48.35	157.27	85.50	156.10
H^2	0.0378	0.03582	0.02116	0.0493

Reference

Agresti, A. (1990) it Analysis of Categorical Data. J. Wiely & Sons, New York.

Barndorff-Nielsen O. (1978) *Information and exponential families.* John Wiley & Sons, Ltd.

Burridge, J. (1982) Some unimodality properties of likelihoods derived from grouped data. *Biometrika*, 69,145-51.

Donoho, D.L. and Liu, R.C. (1988) The automatic robustness of minimum distance functionls. *The Annals of statistics*, 16, 552-586.

Dunnett, C.W. (1989) Multivariate normal Probability integrals with product correlation structure. *Applied Statistics* 38, 564-571.

Eslinger, P.W. and Woodward, W.A. (1991) Minimum Hellinger distance estimation for normal models. *J. Statist. Comput. Simul.*., 39, 95-114.

Kiefer, J. (1953) Sequential minimax search for a maximum. *Proc. Amer. Math. Soc.* 4, 502-506.

Goodman, L.A. (1981) Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, 68, 347-355.

Lee, S.Y. and Poon, W.Y. (1986) Maximum likelihood estimation of polyserial correlations. *Psychomatrika*, 51, 113-121.

Martinson, E.O. and Hamdan, M.A. (1971) Maximum likelihood and some other asymptotically efficient estimators of correlation in two way contingency tables. *J. Statist. Comput. Simul.* 1, 45-54.

Olsson, U. (1979) Maximum likelihood estimation of the polychoric correlation coefficient. *Psychomatrika*, 44, 443-460.

Olsson, U. Drasgow, F. and Dorans, N.J. (1982) The polyserial correlation coefficient. *Psychomatrika*, 47, 337-347.

Plackett, R.L. (1981) *The Analysis of Categorical Data.* Second Edition, Charles Griffin & Company, London.

Prekopa, A. (1973). On logarithmic concave measures and functions. *Acta. Sci. Math.*, 34, 335-343.

Tamura, R.N. and Boos, D.D. (1986) Minimum Hellinger distance estimation for multivariate location and covariance. *JASA*, 81, 223-229.

Terza, J.V. and Welland, U. (1991) A comparison of bivariate normal algorithms. *J. Statist. Comput. Simul.* 39, 115-127.

Wang, Y.J. (1987a) Probability integral for bivariate normal distributions: A contingency-table approach. *Biometrika*, 74, 190-195.

Wang, Y. J. (1987b) Order-dependent parameterization of multinomial distributions. *Scand. J. Statist.*, 13, 199-205.