Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 75 (2005) 49–57
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Pairwise likelihood approach to grouped continuous model and its extension

A.R. de Leon*

Department of Mathematics & Statistics, University of Calgary, Calgary, AB, Canada T2N 1N4

Received 28 January 2003; received in revised form 8 April 2005; accepted 21 May 2005
Available online 6 July 2005

Abstract

A pseudo-likelihood estimation method for the grouped continuous model and its extension to mixed ordinal and continuous data is proposed as an alternative to maximum likelihood estimation. The method, based on the pairwise likelihood approach, advocates simply pooling marginal pairwise likelihoods to approximate the full likelihood. In addition to being consistent and asymptotically normally distributed, maximum pairwise likelihood estimates are computationally simple to obtain. Simulations show that the estimates are quite accurate, yielding minimal bias and small root mean-squared errors. The methodology is illustrated using real-data examples.

Keywords: Composite likelihood; Mixed data; Multivariate ordinal data; Partition maximum likelihood; Polychoric and polyserial correlations

1. Introduction

A common approach to handling ordinal data with naturally ordered levels is to assume that the ordinal variables are coarsely measured versions of unobservable continuous variables called latent variables, and are obtained by partitioning or thresholding the space of the latent variables into non-overlapping intervals. This paper is concerned with the grouped continuous model

*Tel.: +1 403 220 6782; fax: +1 403 282 5150.
E-mail address: adeleon@math.ucalgary.ca.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.05.017

(GCM), a model for multivariate ordinal data that assumes a multivariate normal distribution for the latent variables, leading to a probit model for the ordinal variables (Anderson and Pemberton, 1985). It relies on so-called polychoric correlations to model the covariance structure of the data, in contrast to log-linear models that rely on odds ratios or Pearson's correlations as measures of ordinal data association. Unlike Pearson's correlation, polychoric correlation does not restrict the correlation parameter space. This is because polychoric correlations are nothing but the usual pairwise correlations between the continuous latent variables. Moreover, the number of polychoric correlations does not increase with the number of levels of the ordinal data, a common problem with using odds ratios.

Another advantage of GCM is that it can easily be extended to mixed data with ordinal and continuous variables. The conditional GCM was introduced by Anderson and Pemberton (1985) in the context of regression analysis of multivariate ordinal outcomes, where the continuous variables are treated as covariates. It was later studied by Poon and Lee (1987) as a model for mixed data with continuous and ordinal outcomes.

Because maximum likelihood estimation (MLE) of polychoric correlations and thresholds in the GCM can be computationally demanding in practice, a number of alternative approaches have been proposed in the literature (see, e.g., Bedrick et al., 2000; Poon and Lee, 1987). Although these methods are less computationally demanding than MLE, estimation of the parameters is done separately for several models with common parameters. Because they are non-simultaneous, they yield multiple sets of estimates with no clear prescription for combining them to obtain the final estimates.

Section 2 outlines an alternative estimation method for the GCM based on the pairwise likelihood approach (Kuk and Nott, 2000). The method derives from Lindsay's (1988) composite likelihood approach, recent examples of its applications have been discussed by Parner (2001), Nott and Rydén (1999), and Heagerty and Lele (1998). A small simulation study on the efficiency and bias of maximum pairwise likelihood estimates is reported in Section 3. The corresponding development for the conditional GCM is given in Section 4. A brief summary concludes the paper in Section 5.

2. A model for multivariate ordinal data

Suppose $\mathbf{z} = (Z_1, \ldots, Z_Q)^\top$ is a vector of ordinal variables such that Z_q has $L_q + 1$ levels $a_q^1 < \cdots < a_q^{L_q+1}$, $q = 1, \ldots, Q$. Underlying $\mathbf{z}$ is $\mathbf{y}^* = (Y_1^*, \ldots, Y_Q^*)^\top$, a vector of continuous latent variables whose relationship with $\mathbf{z}$ is defined by the threshold model $Z_q = a_q^{\ell_q} \Longleftrightarrow \alpha_q^{\ell_q - 1} < Y_q^* \leqslant \alpha_q^{\ell_q}$, where Y_q^* is the qth element of $\mathbf{y}^*$ and $\alpha_q^0 = -\infty, \alpha_q^1, \ldots, \alpha_q^{L_q}, \alpha_q^{L_q+1} = +\infty$ are the unknown thresholds. Without loss of generality, it is assumed that $a_q^{\ell_q} = \ell_q$, $\ell_q = 1, \ldots, L_q + 1$. Assuming that $\mathbf{y}^* \sim \mathcal{N}_Q(\mathbf{0}, \mathbf{R})$, with $\mathbf{R}$ a correlation matrix, the vector $\mathbf{z}$ is said to be distributed according to the GCM if and only if, for some $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_Q)^\top$, $\Pr(\mathbf{z} = \boldsymbol{\ell}) = \int_{\mathscr{S}} \phi_Q(v_1, \ldots, v_Q \mid \mathbf{R}) \mathrm{d}v_1 \cdots \mathrm{d}v_Q$, where $\mathscr{S} = \{(v_1, \ldots, v_Q) : \alpha_q^{\ell_q - 1} < v_q \leqslant \alpha_q^{\ell_q}, q = 1, \ldots, Q\}$ and $\phi_Q(\cdot \mid \mathbf{R})$ is the Q-dimensional normal distribution function with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}$. The parameters of the model are represented by $\boldsymbol{\theta}_{P\times 1}^\top = (\boldsymbol{\alpha}^\top, \{\mathrm{vech}(\mathbf{R})\}^\top)$, with $\boldsymbol{\alpha}_{L\times 1} = (\alpha_q^{\ell_q}, \ell_q = 1, \ldots, L_q; q = 1, \ldots, Q)$, $L = \sum_{q=1}^{Q} L_q$, $\mathrm{vech}(\mathbf{R})$ is the $Q(Q-1)/2 \times 1$ vector containing the upper triangle of $\mathbf{R}$, and

$P = Q(Q-1)/2 + L$. Note that the thresholds account for the ordinal information in the data while the polychoric correlations represent the associations between the ordinal variables.

2.1. Estimation of $\boldsymbol{\theta}$

Let $\mathbf{z}_1, \ldots, \mathbf{z}_N$ be a random sample from the GCM with parameter $\boldsymbol{\theta}$. As the full likelihood involves evaluation of high dimensional normal integrals, obtaining the MLE can be computationally demanding in practice. A natural alternative approach, motivated by the current interest in estimating equations, is to work with pseudo-likelihoods instead. The approach adopted in this paper employs pairwise likelihoods in constructing a pseudo-likelihood function from which an estimating function is constructed.

Let $\ell^p_{iqq'} = \log \Pr(Z_{iq} = \ell_q, Z_{iq'} = \ell_{q'})$ for the ith observation. Then $\ell^p_{iqq'} = \log[\Phi^{\ell_q,\ell_{q'}}_{q,q'} - \Phi^{\ell_q,\ell_{q'}-1}_{q,q'} - \Phi^{\ell_q-1,\ell_{q'}}_{q,q'} + \Phi^{\ell_q-1,\ell_{q'}-1}_{q,q'}]$, where $\Phi^{\ell_q,\ell_{q'}}_{q,q'} = \Phi_2(\alpha^{\ell_q}_q, \alpha^{\ell_{q'}}_{q'} | r_{qq'})$ is the standard bivariate normal distribution function with correlation $r_{qq'}$. An overall pairwise log-likelihood function for $\boldsymbol{\theta}$ may then be constructed as $\ell^p(\boldsymbol{\theta}) = \sum_{i=1}^N \sum_{q<q'} \ell^p_{iqq'}$. Observe that this is similar to the clustered data case considered by Kuk and Nott (2000) where a cluster here corresponds to a subject. Hence, unlike in the usual clustered data case where the cluster sizes may vary across clusters, cluster sizes in this case are all the same. This, however, is not the case for data with missing observations.

Analogous to MLE, the pairwise score vector $\mathbf{s}^p(\boldsymbol{\theta}) = \partial\ell^p(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ can be similarly defined. The maximum pairwise likelihood (MPL) estimate $\widehat{\boldsymbol{\theta}}^p$ of $\boldsymbol{\theta}$ is defined as the maximizer of $\ell^p(\boldsymbol{\theta})$. It can be obtained by solving the pairwise score equation $\mathbf{s}^p(\boldsymbol{\theta}) = \mathbf{0}$ via a modified Fisher scoring algorithm. Under certain regularity conditions (Heagerty and Lele, 1998), $\widehat{\boldsymbol{\theta}}^p$ is consistent and asymptotically $\mathcal{N}_P(\boldsymbol{\theta}, \mathbf{J}_P^{-1}\mathbf{K}_P\mathbf{J}_P^{-1})$, where $\mathbf{J}_P = \mathrm{E}[-\partial\mathbf{s}^p(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top]$ and $\mathbf{K}_P = \sum_{i=1}^N \mathrm{E}[(\sum_{q<q'} \partial\ell^p_{iqq'}/\partial\boldsymbol{\theta})(\sum_{q<q'} \partial\ell^p_{iqq'}/\partial\boldsymbol{\theta})^\top]$.

The pairwise likelihood approach to the GCM is attractive because it allows for a log-likelihood involving high dimensional normal integrals to be approximated by a sum of bivariate normal integrals, which can be easily evaluated. Although the pairwise likelihood approach is very similar to Poon and Lee (1987) partition maximum likelihood (PML) and Bedrick et al.'s (2000) pairwise PML methods, it is more conceptually appealing than these partition methods because it entails maximizing a single objective function, the pairwise log-likelihood function, to obtain a single set of parameter estimates. Hence, there is no need to average several estimates as is done in the partition methods. Because a full likelihood approach raises questions about the robustness of the likelihood specification quite apart from any computational difficulties there might be, working with a modification of the likelihood function, such as the pairwise likelihood, should be appealing in practice.

2.2. Example: coloring of birds

Anderson and Pemberton (1985) considered data from an ornithological study on the coloring of "first-year" blackbirds. The colors of the lower mandible (Z_1), the upper mandible (Z_2) and orbital ring (Z_3) of 90 birds were recorded as ordinal variables, ranging from all black to all

Table 1
Maximum pairwise likelihood (MPL) estimates and standard errors for the bird data

Parameter	Estimate	SE	*P*-value
Polychoric correlations			
r_{12}	0.933	0.062	<0.001
r_{13}	0.827	0.132	<0.001
r_{23}	0.853	0.189	<0.001
Thresholds			
α_1^1	0.429	0.137	0.002
α_1^2	0.837	0.136	<0.001
α_2^1	0.728	0.169	<0.001
α_2^2	0.968	0.236	<0.001
α_3^1	−0.056	0.122	<0.001
α_3^2	1.291	0.354	<0.001

yellow. As in Anderson and Pemberton (1985), Z_q ($q = 1, 2, 3$) is assumed to have three levels 1, 2 and 3, corresponding to the colorings "mostly black", "intermediate", and "mostly yellow", respectively. These data are modelled using the GCM with $Q = 3$ and $L_1 = L_2 = L_3 = 2$. The model parameters are then estimated by the pairwise likelihood method, and these are shown in Table 1 with their corresponding standard errors and p-values. The p-values are obtained using the asymptotic normality of the estimates. These suggest a strong degree of dependence between the three color levels.

3. Simulation study

To assess the performance of the MPL estimates, a series of simulation experiments were conducted using the GCM with $Q = 3$. Random samples were generated from a 3-dimensional multivariate normal latent distribution with correlation matrix $\mathbf{R}$, and the data $\mathbf{y}_1^*, \ldots, \mathbf{y}_N^*$ were then transformed into $\mathbf{z}_1, \ldots, \mathbf{z}_N$ using two sets of pre-assigned thresholds. These are, for case (I), $\alpha_1^1 = 0, \alpha_2^1 = -0.4, \alpha_2^2 = 0.4, \alpha_3^1 = -0.6, \alpha_3^2 = 0$, and $\alpha_3^3 = 0.6$, and for case (II), $\alpha_1^1 = 0.5, \alpha_2^1 = -0.75, \alpha_2^2 = 0.1, \alpha_3^1 = -0.25, \alpha_3^2 = 0.3$, and $\alpha_3^3 = 1$. The following correlation matrices $\mathbf{R}_{\mathrm{I}}$ and $\mathbf{R}_{\mathrm{II}}$ for cases (I) and (II), respectively, were assumed:

$$\boldsymbol{R}_{\mathrm{I}} = \begin{pmatrix} 1 & 0.5 & 0.5 \\ & 1 & 0.5 \\ & & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{R}_{\mathrm{II}} = \begin{pmatrix} 1 & 0.8 & 0.3 \\ & 1 & 0.4 \\ & & 1 \end{pmatrix}.$$

Case (I) corresponds to symmetric marginal distributions for Z_1, Z_2, and Z_3, while case (II) implies that their marginal distributions are skewed. These cases are similar to those considered by Poon and Lee (1987).

For each case, samples of sizes $N = 50$ and 100 were generated and the thresholds $\alpha_1^1, \alpha_2^1, \alpha_2^2, \alpha_3^1, \alpha_3^2$, and α_3^3, and the polychoric correlations r_{12}, r_{13}, and r_{23}, were estimated using the MPL method. This was replicated a total of $R = 100$ times, and the mean of the MPL estimates calculated. As a measure of the accuracy of the estimates, the approach of Lee and Poon (1986) was adopted and the root mean-squared error $\mathrm{RMSE} = \sqrt{\sum_{r=1}^{R}(\widehat{\boldsymbol{\theta}}_r^p - \boldsymbol{\theta})^2/R}$ was calculated, where $\widehat{\boldsymbol{\theta}}_r^p$ is the MPL estimate of $\boldsymbol{\theta}$ for the rth replicate, $r = 1, \ldots, R$.

Table 2 reports the simulation results for case (I). Those for case (II) are displayed in Table 3. The MPL estimates yielded generally small bias, which decreased with increasing sample size. The bias of the estimates in case (I) was, in general, smaller than the bias of those in case (II), indicating that MPL estimation for the GCM performs better when the threshold model results in symmetric, rather than skewed, distributions for the ordinal variables. Likewise, the RMSEs were generally small. As expected, increasing the sample size decreased the RMSE. From Table 3, it appears that RMSE is smaller for large than for small polychoric correlations, which was similarly noted by Lee and Poon (1986). Furthermore, the estimates for case (I), which gives a symmetric distribution for the ordinal vector, yielded generally smaller RMSEs than those for the skewed ordinal distribution in case (II).

Table 2
Maximum pairwise likelihood (MPL) estimates based on 100 random samples of size $N = 50$ and 100 from the grouped continuous model with $Q = 3$ and parameters given by Case (I)

Parameter	True value	N	Ave. MPL estimate	Bias	Relative bias (%)	RMSE
Polychoric correlations						
r_{12}	0.5	50	0.5246	−0.0246	−4.9121	0.1656
		100	0.4939	0.0061	1.2141	0.109
r_{13}	0.5	50	0.5263	−0.0263	−5.2681	0.1308
		100	0.5004	−0.0004	−0.0781	0.1057
r_{23}	0.5	50	0.5133	−0.0133	−2.6528	0.1356
		100	0.4979	0.002	0.4087	0.1046
Thresholds						
α_1^1	0	50	−0.0007	0.0007	—	0.1819
		100	0.0081	−0.0081	—	0.1189
α_2^1	−0.4	50	−0.3998	−0.0002	0.0524	0.1847
		100	−0.4048	0.0048	−1.1946	0.1322
α_2^2	0.4	50	0.4576	−0.0575	−14.3895	0.2106
		100	0.4107	−0.0107	−2.6712	0.1401
α_3^1	−0.6	50	−0.6207	0.0207	−3.45	0.1961
		100	−0.5905	−0.0095	1.5833	0.1316
α_3^2	0	50	0.1816	−0.1816	—	0.2126
		100	−0.0521	0.0521	—	0.1227
α_3^3	0.6	50	0.5891	0.0109	1.8167	0.2144
		100	0.6011	−0.0011	−0.1833	0.144

Note: Relative bias = (bias/true) × 100.

Table 3
Maximum pairwise likelihood (MPL) estimates based on 100 random samples of size $N = 50$ and 100 from the grouped continuous model with $Q = 3$ and parameters given by Case (II)

Parameter	True value	N	Ave. MPL estimate	Bias	Relative bias (%)	RMSE
Polychoric correlations						
r_{12}	0.8	50	0.8115	−0.0115	−1.4332	0.0997
		100	0.8041	−0.0041	−0.5129	0.0699
r_{13}	0.3	50	0.3211	−0.0211	−7.0308	0.1913
		100	0.2987	0.0013	0.4333	0.14
r_{23}	0.4	50	0.3872	0.0128	3.2	0.1536
		100	0.4116	−0.0116	−2.9034	0.1041
Thresholds						
α_1^1	0.5	50	0.4485	0.0115	2.3011	0.1839
		100	0.5014	−0.0014	−0.2865	0.1116
α_2^1	−0.75	50	−0.8166	0.0666	−8.8815	0.1664
		100	−0.7716	0.0216	−2.8891	0.1237
α_2^2	0.1	50	0.0875	0.0125	12.5227	0.1778
		100	0.0977	0.0023	2.3114	0.1252
α_3^1	−0.25	50	−0.2656	0.0156	−6.2352	0.1834
		100	−0.2647	0.0147	−5.8952	0.1186
α_3^2	0.3	50	0.272	0.0279	9.3307	0.1711
		100	0.2657	0.0343	11.4311	0.1273
α_3^3	1	50	1.0039	−0.0039	−0.3923	0.2267
		100	0.9908	0.0092	0.9219	0.1423

4. Extension to mixed continuous and ordinal data

Cox (1974) presents a study on a flock of 25 mating-age Romney ewes, the objective of which was to determine the correlation between the weights at mating of the ewes and the number of lambs born. The data consist of the number of lambs born and the standardized weight at mating for each of the 25 ewes. These data are used to illustrate the conditional GCM (CGCM), an extension of the GCM to situations involving data consisting of both continuous and ordinal outcomes.

Consider a vector $\mathbf{y} = (Y_1, \ldots, Y_C)^\top$ of continuous variables in addition to $\mathbf{z}$. As in the GCM, a latent vector $\mathbf{y}^* \sim \mathcal{N}_Q(\mathbf{0}, \mathbf{R}^*)$ is assumed for $\mathbf{z}$, such that $\mathbf{y}$ and $\mathbf{y}^*$ are jointly normally distributed with $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu}$, $\mathrm{var}(\mathbf{y}) = \boldsymbol{\Sigma}$, and $\mathrm{cov}(\mathbf{y}, \mathbf{y}^*) = \boldsymbol{\Sigma}_{\mathrm{yy}^*}$. If $\boldsymbol{\Sigma}$ is the correlation matrix of $\mathbf{y}$, $\boldsymbol{\Sigma}_{\mathrm{yy}^*}$ becomes the matrix containing the polyserial correlations of $\mathbf{y}$ and $\mathbf{y}^*$. Conditional on $\mathbf{y}$, $\mathbf{y}^*$ is multivariate normal with mean $\boldsymbol{\Sigma}_{\mathrm{yy}^*}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ and covariance matrix $\mathbf{R}^* - \boldsymbol{\Sigma}_{\mathrm{yy}^*}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}_{\mathrm{yy}^*} = \mathbf{DRD}$, where $\mathbf{D}$ is the diagonal matrix of conditional standard deviations and $\mathbf{R}$ is the conditional polychoric correlation matrix of $\mathbf{z}$. The conditionally standardized latent vector $\mathbf{D}^{-1}\mathbf{y}^* - \mathbf{B}(\mathbf{y} - \boldsymbol{\mu})$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}$, where $\mathbf{B} = \mathbf{D}^{-1}\boldsymbol{\Sigma}_{\mathrm{yy}^*}^\top \boldsymbol{\Sigma}^{-1}$. Assuming the same threshold model in Section 2 for $\mathbf{y}^*$ and $\mathbf{z}$, the thresholds $\alpha_q^{\ell_q}$ are similarly standardized as $v_q^{\ell_q} = \gamma_q^{\ell_q} - \boldsymbol{\beta}_q^\top \mathbf{y}$, where $\gamma_q^{\ell_q} = \alpha_q^{\ell_q}/d_q + \boldsymbol{\beta}_q^\top \boldsymbol{\mu}$, and $\boldsymbol{\beta}_q^\top$ is the qth row of $\mathbf{B}$, with $\gamma_q^0 = -\infty$ and $\gamma_q^{L_q+1} = +\infty$. For some $\ell = (\ell_1, \ldots, \ell_Q)^\top$, the conditional distribution $[\mathbf{z} \mid \mathbf{y}]$ of $\mathbf{z}$ given $\mathbf{y}$ is $[\mathbf{z} = \ell \mid \mathbf{y}] = \int_{\mathscr{S}_\mathbf{y}} \phi_Q(v_1, \ldots, v_Q \mid \mathbf{R}) \mathrm{d}v_1 \cdots \mathrm{d}v_Q$, where $\mathscr{S}_\mathbf{y} = \{(v_1, \ldots, v_Q) : v_q^{\ell_q - 1} < v_q \leqslant v_q^{\ell_q}, q = 1, \ldots, Q\}$. Then $\mathbf{y}$ and $\mathbf{z}$ are said to be jointly distributed according to the CGCM if and only if $\mathbf{y} \sim \mathscr{N}_C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $[\mathbf{z} \mid \mathbf{y}]$ is given as above.

The parameters of the model are represented by $\boldsymbol{\theta}_{P\times 1}^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$, where $\boldsymbol{\theta}_1^\top = (\boldsymbol{\mu}^\top, \{\mathrm{vech}(\boldsymbol{\Sigma})\}^\top)$, $\boldsymbol{\theta}_2^\top = (\boldsymbol{\gamma}^\top, \{\mathrm{vech}(\mathbf{R})\}^\top, \boldsymbol{\beta}^\top)$, $\boldsymbol{\gamma}^\top = (\gamma_q^{\ell_q}, \ell_q = 1, \ldots, L_q; q = 1, \ldots, Q)$, $\boldsymbol{\beta} = \mathrm{vec}(\mathbf{B})$ is the vector obtained by stacking the rows of $\mathbf{B}$, and $P = C + C(C-1)/2 + Q(Q-1)/2 + CQ + L$. The regression parameters $\boldsymbol{\beta}_q$, $q = 1, \ldots, Q$, represent the polyserial correlations between $\mathbf{y}$ and $\mathbf{z}$. If $\boldsymbol{\beta}_q = \mathbf{0}\ \forall q$, then $\mathbf{y}$ and $\mathbf{z}$ are independent and separate analyses suffice. In this case, the conditional probit model above reduces to the GCM.

4.1. Estimation of $\boldsymbol{\theta}$

For a sample $(\mathbf{y}_i^\top, \mathbf{z}_i^\top)^\top$, $i = 1, \ldots, N$, the pairwise log-likelihood function can be defined as $\ell^p(\boldsymbol{\theta}) = \ell_1^p(\boldsymbol{\theta}_1) + \sum_{i=1}^N \sum_{q<q'} \ell_{iqq'}^p = \ell_1(\boldsymbol{\theta}_1) + \ell_2^p(\boldsymbol{\theta}_2)$, where $\ell_1(\boldsymbol{\theta}_1)$ is the usual multivariate normal log-likelihood function. The usual MLEs $\bar{\mathbf{y}}$ and $\mathrm{vech}(\mathbf{S})$ are obtained as $\widehat{\boldsymbol{\theta}}_1^p$ while $\widehat{\boldsymbol{\theta}}_2^p$ entails an iterative method as in Section 2. Under certain regularity conditions (see, e.g., Heagerty and Lele, 1998), $\widehat{\boldsymbol{\theta}}^p = (\{\widehat{\boldsymbol{\theta}}_1^p\}^\top, \{\widehat{\boldsymbol{\theta}}_2^p\}^\top)^\top$ is consistent and asymptotically $\mathscr{N}_P(\boldsymbol{\theta}, \mathbf{V})$, where $\mathbf{V} = \mathrm{diag}(\mathbf{I}_{P_1}^{-1}(\boldsymbol{\theta}_1), \mathbf{J}_{P_2}^{-1}\mathbf{K}_{P_2}\mathbf{J}_{P_2}^{-1})$, $\mathbf{K}_{P_2} = \mathrm{E}[(\sum_{q<q'} \partial \ell_{iqq'}^p / \partial \boldsymbol{\theta}_2)(\sum_{q<q'} \partial \ell_{iqq'}^p / \partial \boldsymbol{\theta}_2)^\top]$, $\mathbf{J}_{P_2} = \mathrm{E}[-\partial^2 \ell_2^p(\boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top]$, $\mathbf{I}_{P_1}(\boldsymbol{\theta}_1) = \mathrm{E}[-\partial \ell_1(\boldsymbol{\theta}_1)/\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top]$, $P_1 = C + C(C-1)/2$, and $P_2 = P - P_1$. Note that the asymptotic covariance between $\widehat{\boldsymbol{\theta}}_1^p$ and $\widehat{\boldsymbol{\theta}}_2^p$ is zero because the pairwise likelihood function is expressed as the sum of separate functions of $\boldsymbol{\theta}_1^p$ and $\boldsymbol{\theta}_2^p$.

4.2. Example: mating of ewes

The model described above is now applied to the mating data of Cox (1974). With Y_i denoting the standardized weight and Z_i the number of lambs born for ewe $i = 1, \ldots, 25$, a CGCM with $C = Q = 1$ is adopted for the mixed data, where Z_i is taken to be ordinal with levels $\ell = 0, 1, 2$, as in Cox (1974). The parameter vector is then $\boldsymbol{\theta}^\top = (\mu, \sigma^2, \gamma^1, \gamma^2, \beta)$, where γ^1 and γ^2 are the standardized thresholds and $\beta = \rho/(\sigma\sqrt{1-\rho^2})$, with ρ the polyserial correlation between Y and Z.

The parameter $\boldsymbol{\theta}$ is estimated via the pairwise likelihood method. Estimates and their standard errors in Table 4 correspond to the usual MLEs. Note that these are very close, if not identical, to those obtained by Cox (1974). The estimated polyserial correlation between the number of lambs born and the standardized weight at mating is given by $\widehat{\rho} = \mathrm{sgn}(\widehat{\beta})\widehat{\sigma}/\sqrt{1 + \widehat{\beta}^2 \widehat{\sigma}^2} = 0.2112$, which agrees with the estimate reported by Cox (1974).

Table 4
Maximum pairwise likelihood (MPL) estimates and standard errors for the mating data

Parameter	Estimate	SE	*P*-value
Standardized weight (Y)			
μ	90.68	2.362	<0.001
σ	11.808	1.67	<0.001
Number of lambs (Z)			
γ^1	0.466	0.152	0.001
γ^2	2.525	0.177	<0.001
β	0.018	0.002	<0.001

5. Discussion

An alternative method to maximum likelihood estimation was proposed for the GCM and its extension, the conditional GCM. By working with pairwise likelihoods instead of the full likelihood of the model, high dimensional numerical integration is avoided. The pairwise likelihood method is thus computationally simple, and properties such as consistency and asymptotic normality of the estimators readily follow from standard theory. The proposed method is also appealing in practice where there might be concerns about the robustness of the likelihood specification insofar as it involves high order integrals.

Moreover, the pairwise likelihood method provides a viable alternative to the PML methods of Poon and Lee (1987) and Bedrick et al. (2000). Unlike the latter, the former simultaneously estimates the parameters yielding a single set of estimates. Thus, the problem of having to deal with several estimates required in PML methods is avoided. This is accomplished by specifying a single objective function, the pairwise log-likelihood function, which is maximized to obtain the estimates, resulting in possibly fewer convergence problems.

Acknowledgements

The author is grateful to Prof. K.C. Carrière of the University of Alberta for helpful discussions and to G. Singh and R. Wang for computational assistance. He also thanks a referee for very helpful comments. This research is supported by the Alberta Heritage Foundation for Medical Research and the Natural Sciences and Engineering Research Council of Canada.

References

Anderson, J.A., Pemberton, J.D., 1985. The grouped continuous model for multivariate ordered categorical variables and covariate adjustment. Biometrics 41, 875–885.

Bedrick, E.J., Lapidus, J., Powell, J.F., 2000. Estimating the Mahalanobis distance from mixed continuous and discrete data. Biometrics 56, 394–401.

Cox, N.R., 1974. Estimation of the correlation between a continuous and a discrete variable. Biometrics 30, 171–178.

Heagerty, P.J., Lele, S.R., 1998. A composite likelihood approach to binary spatial data. J. Am. Statist. Assoc. 93, 1099–1111.

Kuk, A.Y.C., Nott, D.J., 2000. A pairwise likelihood approach to analyzing correlated binary data. Statist. Probab. Lett. 47, 329–335.

Lee, S.-Y., Poon, W.-Y., 1986. Maximum likelihood estimation of polyserial correlations. Psychometrika 51, 113–121.

Lindsay, B.G., 1988. Composite likelihood methods. Contemporary Math. 80, 221–239.

Nott, D.J., Rydén, T., 1999. Pairwise likelihood methods for inference in image models. Biometrika 86, 661–676.

Parner, E.T., 2001. A composite likelihood approach to multivariate survival data. Scand. J. Statist. 28, 295–302.

Poon, W.-Y., Lee, S.-Y., 1987. Maximum likelihood estimation of multivariate polyserial and polychoric correlation coefficients. Psychometrika 52, 409–430 (correction in 53, p. 301).