

Existence and Uniqueness Conditions for the Maximum Likelihood Solution In Regression Models For Correlated Bernoulli Random Variables

¹David Todem and ²KyungMann Kim

¹Department of Epidemiology, Michigan State University, East Lansing, U.S.A

²Department of Biostatistics and Medical informatics, University of Wisconsin-Madison, Madison, U.S.A

Abstract: We give sufficient and necessary conditions for the existence of the maximum likelihood estimate in a class of multivariate regression models for correlated Bernoulli random variables. The models use the concept of threshold crossing technique of an underlying multivariate latent variable with univariate components formulated as a linear regression model. However, in place of their Gaussian assumptions, any specified distribution with a strictly increasing cumulative distribution function is allowed for error terms. A well known member of this class of models is the multivariate probit model. We show that our results are a generalization of the concepts of separation and overlap of Albert and Anderson for the study of the existence of maximum likelihood estimate in generalized linear models. Implications of our findings are illustrated through some hypothetical examples

Key words: Correlated Bernoulli data; Existence/uniqueness conditions; Latent variables; Linear programming; Maximum likelihood estimate; Overlap/separation conditions; Threshold values.

INTRODUCTION

The question of the existence of the maximum likelihood estimate (MLE) for correlated binary regression models arises primarily because of the iterative nature of the estimation process. Typically, when the score equations do not have closed form solutions, resources and time can be saved if a prior study on the existence of a finite solution is undertaken. Without a proper precaution, estimates from a Newton-Raphson or any other optimization algorithm output may look fine even though, analytically, solutions to the equations are infinite. Although the identifiability issues are common to all models for which parameters are obtained using numerical techniques, the extent of the difficulties increase with the complexity and the high dimensionality of the models under investigation. A good example of a high dimensional model is the multivariate regression models for correlated Bernoulli random variables. This family of models, besides location parameters which account for the effects the regressors have on univariate probabilities, have association parameters which correct these probabilities for potential correlation among the Bernoulli random variables.

Multivariate regression models for correlated Bernoulli random variables are widely used in almost all areas of research (see for example the work of Muthén^[8]; Muthén and Muthén^[9] imple-

mented in the statistical software package Mplus; and Todem, Kim and Lesaffre^[15]).

A well known member of this class of models is the multivariate probit regression model routinely used in medical research. General areas of application include situations where realizations of multiple binary random variables are recorded on the same experimental unit. For instance, in medical decision making, one is often confronted with the problem that patients can suffer from more than one disease. It is then important not only to know the risk factors for each disease, but also to have an individual prediction rule to classify patients in the different possible combination of diseases. The same is true in clinical trials, where the profile of a new drug is usually evaluated from the joint distribution of a response vector comprising information on efficacy and safety parameters. Although separate models can be fitted to each Bernoulli random variable, such an approach can fail to borrow strength across the correlated random variables. By exploiting the correlation structure with a multivariate model, efficiency and power could be greatly increased (O'Brien^[10]).

Considerable attention has been given to the existence and uniqueness of maximum likelihood estimate (MLE) in univariate probit and logit models (Silvapulle^[12]; and Albert and Anderson^[2]). For univariate survival data, similar work has been done for some special models such as the

Weibull distribution tampered failure rate model (see, for example, the work of Wang et al.^[16]).

In the context of correlated responses, Demidenko and Massam^[5]; and Birkes *et al.*^[3] have studied the conditions for the existence of MLE in the class of variance component models. These authors have shown that the MLE exists if and only if the outcome vector does not belong to the space spanned by columns of the fixed effects and random effects. This approach was limited to the identity link and assumes a normal response. From a Bayesian perspective, an important piece of work is that of Sun *et al.*^[14] who established a necessary and sufficient condition for the propriety of the posterior distribution in hierarchical linear mixed effects models for a collection of improper priors. Although most of these previous work have focused on univariate outcomes, a number of authors have also investigated a family of multivariate models for correlated binary/ordinal data such as the multi-group logistic regression and the multivariate probit models (see, for example, Albert and Anderson^[1]; and Lesaffre and Kaufmann^[6]). Specifically, Lesaffre and Kaufmann^[6] have studied the existence and uniqueness of the MLE for the joint estimation of the location parameters as well as the polychoric correlation coefficient in a multivariate probit model. Although the behavior of the maximum likelihood estimators in this restricted class of models for correlated data is well understood, there seems to be a gap in the literature for models that allow any family of link functions. Specifically, no theoretical work on the joint estimation of the location and association parameters has been conducted for multivariate binary regression models with link functions beyond the probit.

This paper proposes a simple proof for the existence and uniqueness conditions of the maximum likelihood estimate in regression models for correlated Bernoulli data. Our proof relies on the concept of data separation, which is fundamental to this research area. As described by Lesaffre and Kaufmann^[6], our results are derived in a stepwise fashion where association component is first profiled out from the likelihood. The joint MLE results are later derived using continuity arguments of the profile likelihood and other known functional analysis results. The model is

described in section 2; section 3 discusses existence results of the maximum likelihood estimate and gives some implications of the results in a hypothetical example..

MODEL UNDER INVESTIGATION

Consider a random vector $Y = (Y_1, Y_2, \dots, Y_g)'$ $g \geq 2$, recorded for each experimental unit, and taking values in the product space $\otimes_{\ell=1}^g \{0, 1\}$, where 0 refers to failure and 1 to success. In addition, a non random design component x_ℓ , assumed to be a potential predictor of Y_ℓ , $\ell = 1, 2, \dots, g$, is also recorded for each experimental unit $k = 1, \dots, N$, where N is the number of independent realizations of the random vector Y . Multivariate regression models make explicit assumptions on the distribution of the vector Y . They are based on the threshold crossing technique of a multivariate latent variable model $W = (W_1, W_2, \dots, W_g)'$. In turn, each univariate component of this latent vector is modeled as a linear regression model. However, in place of their Gaussian assumptions, the error vectors are allowed to have any specified strictly increasing cumulative distribution function (cdf). Specifically, for an experimental unit $k = 1, 2, \dots, N$, and for a realized value $y_{\ell k}$ of the random variable $Y_{\ell k}$, we have the following model,

$$p(Y_{\ell k} = y_{\ell k}, \ell = \ell_1, \dots, \ell_{g'}) = p(s_{\ell k}(W_{\ell k} - a_\ell) \leq 0, \ell = \ell_1, \dots, \ell_{g'}), \quad (1)$$

for any ordered series $\ell_1 < \ell_2 < \dots < \ell_{g'}$ with any dimension g' such that $1 \leq g' \leq g$. The quantities a_ℓ , $\ell = 1, 2, \dots, g$ are threshold values and $s_{\ell k} = (-1)^{y_{\ell k}}$. In particular, if all events are successes, we have, $p(Y_{1k} = 1, \dots, Y_{gk} = 1) = p(a_1 \leq W_{1k}, \dots, a_g \leq W_{gk})$.

As stated above, each univariate random variable W_ℓ , $\ell = 1, 2, \dots, g$, of the this latent model is modeled as a linear regression. Specifically, we assume that,

$$W_{\ell k} = x'_{\ell k} \beta_\ell + \varepsilon_{\ell k}, \ell = 1, 2, \dots, g; k = 1, 2, \dots, N, \quad (2)$$

where the random vector $\varepsilon_k = (\varepsilon_{1k}, \varepsilon_{2k}, \dots, \varepsilon_{gk})'$ is the residual vector. We assume that the random vector ε_k can take any a parametric distribution with a strictly increasing cdf, a zero mean and a variance-covariance matrix with entries $\text{cov}(\varepsilon_{\ell k}, \varepsilon_{\ell' k}) = \rho_{\ell \ell'}, \ell, \ell' = 1, 2, \dots, g$. Since

the observed responses are dichotomized versions of the latent responses, the probability distribution of the observed responses contains no information on the scale of the latent responses. Therefore, the individual variances of the random error terms, $\text{var}(\varepsilon_{\ell k})$, $\ell = 1, 2, \dots, g$, are not estimable and therefore fixed to unity (Todem *et al.*^[15]). The parameter β_ℓ is the slope parameter associated to the design matrix $x_{\ell k}$, $\ell = 1, 2, \dots, g$. Without any loss of generality, the threshold values a_ℓ , $\ell = 1, 2, \dots, g$, are typically set to zero if intercept terms are included in the latent regression models. If ε_k is generated from a multivariate normal distribution, the model reduces to the classical multivariate probit model.

An alternative to this class of parametric models is the widely known generalized estimating equations (GEE) model of Liang and Zeger^[7], which does not require a full specification of the association among the univariate components of Y . It provides consistent and asymptotically normal estimates of the marginal covariate effects even when the association structure is misspecified. The GEE model has several advantages over a fully parametric model in that it is computationally tractable in applications where the parametric approaches are computationally very demanding, if not impossible. Although the GEE does not require a full specification of the data distribution, this assumption is unavoidable if a profile analysis is required.

The contribution of subject k to the likelihood function is given by,

$$p(Y_k = y_k) = \Psi_{H_k}(-s_{1k}e_{1k}, \dots, -s_{gk}e_{gk}) \quad (3)$$

where $y_k = \{y_{1k}, y_{2k}, \dots, y_{gk}\}$ is the observed vector, $\beta = (\beta'_1, \dots, \beta'_g)'$ and ρ is a vector of dimension $g(g-1)/2$ with entries being the polychoric correlation coefficients $\rho_{\ell\ell'}$, $\ell' < \ell$. The quantity $\Psi_{H_k}(\cdot)$, is the distribution function of the random vector $(s_{1k}\varepsilon_{1k}, \dots, s_{gk}\varepsilon_{gk})$ and $e_{\ell k} = x'_{\ell k}\beta_\ell$ is the expectation of the random variable $W_{\ell k}$, $\ell = 1, 2, \dots, g$. The matrix H_k is the correlation matrix of the random vector $(s_{1k}\varepsilon_{1k}, \dots, s_{gk}\varepsilon_{gk})$ with below diagonal entries $s_{\ell k}s_{\ell' k}\rho_{\ell\ell'}$, $\ell' < \ell$. If ε_k has a symmetric distribution, it is clear that,

$$(s_{1k}\varepsilon_{1k}, \dots, s_{gk}\varepsilon_{gk}) \stackrel{d}{\sim} (\varepsilon_{1k}, \dots, \varepsilon_{gk}), \quad (4)$$

where “ $\stackrel{d}{\sim}$ ” stands for equality in distribution.

MLE EXISTENCE AND UNIQUENESS RESULTS

We now examine the conditions for the existence and uniqueness of the MLE of parameters from this marginal distribution. We first review the concepts of data configuration that are fundamental in this area.

Brief review of data configuration

The concepts of separation and overlap for the regression parameters have been central in the work on the existence of MLE in the independence probit and logit models. Albert and Anderson^[1] first used these concepts to represent three separate data configurations in a univariate logit model. Lesaffre and Kaufmann^[6] generalized these concepts to establish existence conditions of the MLE in a fixed effects multivariate probit model. Speckman *et al.*^[13] also extended these notions to study necessary and sufficient conditions for the existence of the MLE for a wide class of discrete (or multinomial) choice models. Basically, it is known that some degree of overlap in the data is necessary to guarantee the existence of MLE for the location parameters. For example, in fitting logistic models for quantal bioassay, there should be an overlap in the doses that produces response and no response. We extend these concepts to the context of multivariate data for which we consider the following arbitrary sets:

$$U_\ell = \{s_{\ell k}x_{\ell k}, k = 1, \dots, N\}, \ell = 1, 2, \dots, g. \quad (5)$$

The set U_ℓ , $\ell = 1, 2, \dots, g$, has a *complete separation*, or the sample has a complete separation with respect to the regression parameter, if there is a vector $\beta_\ell \neq 0$ of \mathcal{R}^{p_ℓ} such that,

$$u'\beta_\ell < 0; \text{ for all } u \in U_\ell. \quad (6)$$

In other words, complete separation means that it is possible to classify the outcomes using a single rule. The set U_ℓ , $\ell = 1, 2, \dots, g$, has a *quasi-complete separation* if there is a vector $\beta_\ell \neq 0$ of \mathcal{R}^{p_ℓ} such that,

$$u'\beta_\ell \leq 0; \text{ for all } u \in U_\ell. \quad (7)$$

Finally, U_ℓ , $\ell = 1, 2, \dots, g$, is *overlapped* if for any vector $\beta_\ell \neq 0$ of \mathcal{R}^{p_ℓ} , we have,

$$u'\beta_\ell > 0; \text{ for some } u \in U_\ell. \quad (8)$$

Existence and uniqueness results for fixed ρ

We now state in the following theorem, the conditions for the existence and uniqueness of the MLE for the location parameters β_ℓ , $\ell = 1, \dots, g$, when the association component ρ is held fixed.

Theorem 1 *We assume that the association component represented by ρ with $|\rho_{\ell\ell'}| \neq 1, \ell, \ell' (\ell' < \ell) = 1, 2, \dots, g$, is fixed. The MLE of β exists and is unique if and only if all $\mathcal{U}_\ell, \ell = 1, 2, \dots, g$, are overlapped.*

To establish the proof of this theorem, we first define some key concepts in the theory of convex sets and review some known results that have been beautifully summarized by Speckman *et al.*^[13]. A subset \mathcal{C} of \mathcal{R}^p is a cone, if for all points $u \in \mathcal{C}$ and $\lambda \geq 0$, we have $\lambda u \in \mathcal{C}$. A cone $\mathcal{C} \subset \mathcal{R}^p$, is called a convex cone if it is closed under addition, i.e. for all $\{u_1, \dots, u_n\} \subset \mathcal{C}$, the vector $\sum_{i=1}^n u_i \in \mathcal{C}$. Elements of the form $\sum_{i=1}^n \lambda_i u_i$ where $\{u_1, \dots, u_n\} \subset \mathcal{C}$ and $\{\lambda_1, \dots, \lambda_n\}$ is a set of positive real numbers, are conical combinations. The conical hull of \mathcal{C} , which we denote by $Hull(\mathcal{C})$, is the smallest cone containing all convex combinations of points in \mathcal{C} . It can easily be shown that the conical hull of a set \mathcal{C} is given by,

$$Hull(\mathcal{C}) = \left\{ \sum_{i=1}^j \lambda_i u_i : u_i \in \mathcal{C} \text{ and } \lambda_i \geq 0, i = 1, \dots, j \right\}.$$

We now introduce a quasi-norm function on \mathcal{R}^p associated with a set \mathcal{C} when its conical hull is the p -dimensional space \mathcal{R}^p . For any subset \mathcal{C} of \mathcal{R}^p and any $\beta \in \mathcal{R}^p$ define,

$$\|\beta\|_{\mathcal{C}} = \max_{u \in \mathcal{C}} (u' \beta). \tag{9}$$

This quasi-norm clearly satisfies the usual triangular inequality, that is, for $\beta, \beta_* \in \mathcal{R}^p$, $\|\beta + \beta_*\|_{\mathcal{C}} \leq \|\beta\|_{\mathcal{C}} + \|\beta_*\|_{\mathcal{C}}$. However, for this function to be a norm, we need the following conditions to hold as well,

- C1. $\|\beta\|_{\mathcal{C}} \geq 0$ for all $\beta \in \mathcal{R}^p$
- C2. $\|\beta\|_{\mathcal{C}} = 0$ if and only if $\beta = 0$
- C3. $\|\alpha\beta\|_{\mathcal{C}} = |\alpha| \|\beta\|_{\mathcal{C}}$ for all $\alpha \in \mathcal{R}$.

We should note that for $\alpha > 0$, the condition C3 is always satisfied. Now we establish a connection between this quasi-norm and the data configuration in the following lemma.

Lemma 1 *The following conditions are equivalent.*

- (a) $\mathcal{U}_\ell, \ell = 1, 2, \dots, g$, are all overlapped.
- (b) The conical hull of \mathcal{U}_ℓ is \mathcal{R}^{p_ℓ} , $\ell = 1, 2, \dots, g$.
- (c) Conditions C1 and C2 hold.

Proof of Lemma 1

The proof of this theorem is given by Speckman *et al.*^[13] in the context of multinomial choice models. We will show that the three conditions are equivalent for the first random variable Y_1 and the corresponding conical set \mathcal{U}_1 . Similar results can be established for the other random variables Y_2, \dots, Y_g .

(a) \Rightarrow (b) Assume that (a) is true but suppose that (b) is false, i.e. $Hull(\mathcal{U}_1) \neq \mathcal{R}^{p_1}$. Then there is a nonzero $\beta_1^* \in Hull(\mathcal{U}_1)^*$, the dual cone of $Hull(\mathcal{U}_1)$, defined as,

$$Hull(\mathcal{U}_1)^* = \{\beta_1 : u' \beta_1 \leq 0 \text{ for all } u \in \mathcal{U}_1\}.$$

Since $u' \beta_1^* \leq 0$ for all $u \in \mathcal{U}_1$, the quasi-complete separation of \mathcal{U}_1 with respect to the first outcome then follows. This therefore contradicts the condition in (a) which assumes that \mathcal{U}_1 is overlapped.

(b) \Rightarrow (c) Assume that $Hull(\mathcal{U}_1) = \mathcal{R}^{p_1}$. Hence, for all $\beta_1 \in \mathcal{R}^{p_1}$, we can then find a sequence of parameters $\lambda_j > 0$ and a sequence of elements $u_j \in \mathcal{U}_1, j = 1, \dots, i$, such that $\beta_1 = \sum_{j=1}^i \lambda_j u_j$. Therefore we get,

$$0 \leq \beta_1' \beta_1 = \sum_{j=1}^i \lambda_j u_j' \beta_1 \leq \sum_{j=1}^i \lambda_j \|\beta_1\|_{\mathcal{U}_1} = \|\beta_1\|_{\mathcal{U}_1} \sum_{j=1}^i \lambda_j. \tag{10}$$

Hence, to show that $\|\beta_1\|_{\mathcal{U}_1} \geq 0$, suppose $\beta_1 \neq 0$ but $\|\beta_1\|_{\mathcal{U}_1} < 0$. Since all $\lambda_j > 0, j = 1, \dots, i$, the right hand side of (10) is strictly negative, and thus leads to $\beta_1' \beta_1 < 0$, which is impossible. Therefore condition C1 holds. Similarly, if $\|\beta_1\|_{\mathcal{U}_1} = 0$ then $\beta_1' \beta_1 = 0$ by (10). Therefore $\beta_1 = 0$ and condition C2 holds. It is also easy to show that if $\beta_1 = 0$, then $\|\beta_1\|_{\mathcal{U}_1} = 0$.

(c) \Rightarrow (a) If conditions C1 and C2 hold, therefore for any nonzero $\beta_1 \in \mathcal{R}^{p_1}$, we get

$$\|\beta_1\|_{\mathcal{U}_1} = \max_{u \in \mathcal{U}_1} (u' \beta_1) > 0.$$

Thus, for any nonzero $\beta_1 \in \mathcal{R}^{p_1}$ there exists a $u \in \mathcal{U}_1$ such that $u'\beta_1 > 0$, which establishes that \mathcal{U}_1 is overlapped.

We will use the results of this lemma to establish a connection between the quasi-norm defined earlier and the Euclidean norm under some conditions. This is presented in the following lemma.

Lemma 2 *If any of the conditions stated in Lemma 1 holds, then there are constants $C_\ell > 0$ such that,*

$$C_\ell \|\beta\| \leq \|\beta\|_{\mathcal{U}_\ell} \text{ for all } \beta \in \mathcal{R}^{p_\ell}, \ell = 1, 2, \dots, g$$

where $\|\beta\| = \sqrt{\beta'\beta}$ is the Euclidean norm of β . The connection between the two norms is made to define compact sets on \mathcal{R}^{p_ℓ} which are easily constructed using spheres with the Euclidean norm.

Proof of Lemma 2

We also present the proof for the first random variable Y_1 and its corresponding space \mathcal{U}_1 as that of the other random variables can be similarly derived. The proof, which we include here for completeness, is essentially the same as that of equivalence of norms in a finite-dimensional space. Suppose that the conclusion of the lemma does not hold. We can then find a sequence β_{1n} such that

$$\frac{\|\beta_{1n}\|_{\mathcal{U}_1}}{\|\beta_{1n}\|} \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Since condition C3 holds for positive α , we get $\frac{1}{\|\beta_{1n}\|} \times \beta_{1n} \|_{\mathcal{U}_1} = \frac{\|\beta_{1n}\|_{\mathcal{U}_1}}{\|\beta_{1n}\|}$. Hence without a loss of generality we assume $\|\beta_{1n}\| = 1$ for all n . The unit sphere in \mathcal{R}^{p_1} is compact, so for any β_1 , there is a convergent subsequence $\beta_{1n_j} \longrightarrow \beta_1$ with $\|\beta_1\| = 1$. Since the quasi-norm $\|\cdot\|_{\mathcal{U}_1}$ is continuous, we get $\|\beta_{1n_j}\|_{\mathcal{U}_1} \longrightarrow \|\beta_1\|_{\mathcal{U}_1}$. In addition, since $\|\beta_{1n_j}\|_{\mathcal{U}_1} \longrightarrow 0$ as a subsequence of β_{1n} , we get from the uniqueness of the limits that $\|\beta_1\|_{\mathcal{U}_1} = 0$ for some nonzero β_1 , contradicting condition C2 of Lemma 1.

Proof of Theorem 1

Necessity: To show this, we assume that the MLE is finite and then we attempt to show that all $\mathcal{U}_\ell, \ell = 1, 2, \dots, g$ are each overlapped. First, let us assume for example that there exists at least one of the sets, say \mathcal{U}_1 , which does not have

an overlap. Therefore for this set, we have quasi-complete (which includes complete) separation and there exists a nonzero β_{1*} such that,

$$s_{1k}x'_{1k}\beta_{1*} \leq 0 \text{ for all } k.$$

(i) If \mathcal{U}_1 is of full rank, therefore there exists k^* such that, $s_{1k^*}x'_{1k^*}\beta_{1*} < 0$. Hence, for any finite β , the profile likelihood function, $L(\beta | \rho_0) = \prod_{k=1}^N P(Y_k = y_k)$ at fixed $\rho = \rho_0$, is respectively given and bounded as follows,

$$\begin{aligned} L(\beta | \rho_0) &= \prod_{k=1}^N \Psi_{H_k}(-s_{\ell k}x'_{\ell k}\beta_\ell, \ell = 1, 2, \dots, g) \\ &< \prod_{k=1}^N \Psi_{H_k}(-s_{1k}x'_{1k}(\beta_1 + \beta_{1*}), -s_{\ell k}x'_{\ell k}\beta_\ell, \ell \geq 2) \\ &= L(\beta_1 + \beta_{1*}, \beta_{2:g} | \rho_0), \end{aligned}$$

where $\beta_{2:g}$ is the collection of all vectors $\beta_\ell, \ell = 2, 3, \dots, g$. The strict inequality holds because of the strict inequality in the quasi-complete separation equation for k^* and because $\Psi_{H_k}(\cdot)$ is strictly increasing in its coordinates. Thus no finite β_1 can maximize the profiled likelihood $L(\beta_1, \beta_{2:g} | \rho_0)$ for all $\beta_{2:g}$.

(ii) If the equality in the quasi-complete separation inequality holds for all k , therefore \mathcal{U}_1 is not of full rank and the likelihood function is given by,

$$\begin{aligned} L(\beta | \rho_0) &= L(\beta_1 + \beta_{1*}, \beta_{2:g} | \rho_0) \\ &= L(\beta_1 + \gamma\beta_{1*}, \beta_{2:g} | \rho_0) \quad \forall \gamma \in \mathcal{R}. \end{aligned}$$

By taking γ large enough, this finding contradicts the assumption that the MLE is finite.

Sufficiency: We assume that all $\mathcal{U}_\ell, \ell = 1, 2, \dots, g$, are each overlapped and we need to show that the MLE of β is finite. If the likelihood function $L(\beta | \rho_0) \equiv 0$, the existence of the MLE for β is obvious. Now assume that $L(\beta | \rho_0) \not\equiv 0$. The likelihood function for any finite vector β , is bounded as follows,

$$L(\beta | \rho_0) \leq \min_k \Psi_{H_k}(-s_{\ell k}x'_{\ell k}\beta_\ell, \ell = 1, 2, \dots, g)$$

where the inequality is due to the fact that $\Psi_{H_k}(\cdot)$ is bounded by 1. We consider the quantity function $\min_k \Psi_{H_k}(\cdot)$. If the distribution of ε_k is symmetric, the cdf $\Psi_{H_k}(\cdot)$ does not depend on the sign of the product $s_{\ell k}s_{\ell' k}, \ell \neq \ell'$, according to (4). Therefore, $\min_k \Psi_{H_k}(\cdot) = \Psi_{\rho_0}(\cdot)$, where $\Psi_{\rho_0}(\cdot)$ is the cdf of ε_k . Now assume that the

distribution of ε_k is not symmetric. The magnitude of $\Psi_{H_k}(\cdot)$ as a function of index k , then depends on the sign of quantity $s_{\ell k} s_{\ell' k}$ relative to that of $\rho_{0\ell\ell'}$. Moreover, the function $\Psi_{H_k}(\cdot)$ is continuous on the set $S = \{s_{\ell k} s_{\ell' k}, \ell \neq \ell'; k = 1, 2, \dots, N\}$ which is closed and bounded, so compact due to the fact that $|s_{\ell k} s_{\ell' k}| = 1$, for all $\ell \neq \ell'$. The function $\Psi_{H_k}(\cdot)$, therefore, has a maximum value on S which we denote by $\Psi_{max}(\cdot) = \max_k(\Psi_{H_k}(\cdot))$. The function $\min_k \Psi_{H_k}(\cdot)$ is then bounded by $\Psi_{max}(\cdot)$. We also note that the maximum function $\Psi_{max}(\cdot)$ is a strictly monotone function due to the fact that $\Psi_{H_k}(\cdot)$ is strictly monotone function and indices k belong to a finite set. An upper bound of the likelihood function is then given by,

$$L(\beta | \rho_0) \leq \underline{\Psi} \left(\min_k (-s_{\ell k} x'_{\ell k} \beta_{\ell}), \ell \geq 1 \right) \quad (11)$$

where the function $\underline{\Psi}(\cdot)$ is defined as,

$$\underline{\Psi}(\cdot) = \begin{cases} \Psi_{\rho_0}(\cdot) & \text{if } \varepsilon_k \text{ is symmetric} \\ \Psi_{max}(\cdot) & \text{if otherwise.} \end{cases}$$

Furthermore, since,

$$\min_k \{-s_{\ell k} x'_{\ell k} \beta_{\ell}\} = -\max_k \{s_{\ell k} x'_{\ell k} \beta_{\ell}\} = -\|\beta_{\ell}\|_{\mathcal{U}_{\ell}},$$

for $\ell = 1, 2, \dots, g$, we then have,

$$L(\beta | \rho_0) \leq \underline{\Psi}(-\|\beta_{\ell}\|_{\mathcal{U}_{\ell}}, \ell = 1, 2, \dots, g).$$

From the assumption of the theorem and using results from Lemma 1, there exist constants $C_{\ell} > 0$, $\ell = 1, 2, \dots, g$, such that,

$$L(\beta | \rho_0) \leq \underline{\Psi}(-C_{\ell}\|\beta_{\ell}\|, \ell = 1, 2, \dots, g) \quad (12)$$

Since $L(\beta | \rho_0)$ is not trivial, there exists β_* such that $L(\beta_* | \rho_0) > 0$. The quantity on the right of (12) is bounded by 0 and 1 and takes values throughout this range, we can then find g constants D_1, D_2, \dots and D_g such that for all $\|\beta_{\ell}\| > D_{\ell}, \ell = 1, 2, \dots, g$, we have,

$$\begin{aligned} L(\beta | \rho_0) &\leq \underline{\Psi}(-C_{\ell}D_{\ell}, \ell = 1, 2, \dots, g) \\ &< L(\beta^* | \rho_0). \end{aligned}$$

Therefore we get,

$$\sup_{\beta_{\ell} \in \mathcal{R}^{p_{\ell}}, \ell=1, \dots, g} L(\beta | \rho_0) = \sup_{\|\beta_{\ell}\| \leq D_{\ell}, \ell=1, \dots, g} L(\beta | \rho_0).$$

The function $L(\beta | \rho_0)$ is continuous in β and the spheres $\|\beta_{\ell}\| \leq D_{\ell}, \ell = 1, 2, \dots, g$, are all

bounded and closed (compact), the maximum then exists and any MLE is finite.

Since the distribution function of ε_k is strictly log-concave on \mathcal{R}^g , then strict log-concavity of the probability function $\Psi_{H_k}(\cdot)$ applies based on Prékopa^[11]. Thus uniqueness results from Theorem 1 immediately follow.

MLE existence results for joint estimation of β and ρ

When we let both the location parameter vector β and the polychoric correlation coefficient vector ρ vary, things become more complicated. It can easily be shown that, the probability $\Psi_{H_k}(\cdot)$ is not strictly log concave as a function of the vector ρ , which comprises unique elements of the matrix H_k (see, for example, Lesaffre and Kaufmann^[6]). We therefore state below a partial result pertaining to the existence theorem for a joint maximum likelihood estimation of β and ρ .

Theorem 2 *If the estimated vector of polychoric correlation coefficient is not on the boundary of its parameter space, i.e. $|\hat{\rho}_{\ell\ell'}| \neq 1$, for all $\ell' < \ell$, then the joint MLE of β and ρ exists if all $\mathcal{U}_{\ell}, \ell = 1, 2, \dots, g$, are overlapped.*

Proof of Theorem 2

Let $L(\beta, \rho)$ be the likelihood function for the location parameter vector β and association vector ρ . Before we proceed with the proof of the theorem, let us first investigate the behavior of the profile likelihood when β is held fixed at its MLE $\hat{\beta}(\rho)$ for fixed ρ . We denote the corresponding profile likelihood of ρ as $L(\rho | \hat{\beta}(\rho)) = \sup_{\beta} L(\beta, \rho)$. The set $\mathcal{N} = \{\rho = (\rho)_{\ell\ell'} : |\rho_{\ell\ell'}| \leq 1, \ell \neq \ell'\}$ is compact and moreover the profile likelihood function $\sup_{\beta} L(\beta, \rho)$ is continuous in $\rho \in \mathcal{N}$. Hence, the existence of MLE of ρ from the profile likelihood then follows. For the joint estimation, we use a standard continuity argument which states that if $\log\{L(\beta, \rho)\} > -\infty$ is a continuous function in β and ρ and $\log\{L(\beta, \rho)\} \leq h(\beta)$ for all vector ρ , with $h(\beta) \rightarrow -\infty$ as $\|\beta_{\ell}\| \rightarrow \infty$ for $\ell = 1, 2, \dots, g$, then $\log L(\beta, \rho)$ attains its maximum. To complete the proof, we then need to construct a dominating function $h(\cdot)$. For this, consider the function defined as,

$$h(\beta) = \log \{ \underline{\Psi}(-C_{\ell}\|\beta_{\ell}\|, \ell = 1, 2, \dots, g) \}.$$

which is an increasing function on the support of β . As all $\|\beta_\ell\| \rightarrow \infty$, $\ell = 1, 2, \dots, g$, the function $\Psi(-C_\ell\|\beta_\ell\|, \ell = 1, 2, \dots, g) \rightarrow 0$. Therefore, $h(\beta) \rightarrow -\infty$ as $\|\beta_\ell\| \rightarrow \infty$ for $\ell = 1, 2, \dots, g$. The condition that none of the sets $\mathcal{U}_\ell, \ell = 1, 2, \dots, g$ is quasi-separated ensures that the inequality in (12) holds. Therefore, we have $\log\{L(\beta, \rho)\} \leq h(\beta)$ for all ρ , which completes the proof. Necessarily, the estimated ρ should not be on the boundary of \mathcal{N} , otherwise the multivariate model under investigation becomes questionable and should be reduced to a lower dimension.

Example 1 We assume a no intercept model on both univariate margins of a bivariate latent variable formulated as follows:

$$W_{\ell k} = \beta_\ell x_{\ell k} + \varepsilon_{\ell k}, \ell = 1, 2.$$

Here we assume that $x_{\ell k}$ takes values in the set $\mathcal{O}_k = \{1, 2, 3, 4\}$, for all $k \in \{1, \dots, N\}$.

(a) Assume that for all $x_{1k} \in \mathcal{O}_k$ and all $k \in \{1, \dots, N\}$, we have $s_{1k} = -1$. It is straightforward to show that the condition specified in Theorem 1 is not fulfilled for the first component, that is \mathcal{U}_1 is not marginally overlapped. First, note that, the half line $(-\infty, 0)$ is the solution to the inequalities,

$$s_{1k}x_{\ell k}\beta_1 > 0, \text{ for } x_{\ell k} \in \mathcal{O}_k, k \in \{1, \dots, N\}.$$

Therefore, for $\beta_1 > 0$ there are no indices k so that $s_{1k}x_{\ell k}\beta_1 > 0$. The most extreme case of separation is given in the figure 1. Here it is assumed that β_1 and β_2 are all two-dimensional vectors so that the linear combinations of covariates are straight lines in \mathcal{R}^2 . All four quadrants, excluding the boundaries represented by $x'_{1k}\beta_1 = 0$ and $x'_{2k}\beta_2 = 0$, are the solutions to the inequalities $s_{1k}x'_{1k}\beta_1 > 0$ and $s_{2k}x'_{2k}\beta_2 > 0$ where $s_{\ell k} \in \{-1, 1\}$, $\ell = 1, 2$, is the transformed outcome. For example, the quadrant with outcomes $(1, 1)$, or the $(+, +)$ half space, coincides with the solution of inequalities $x'_{1k}\beta_1 > 0$ and $x'_{2k}\beta_2 > 0$. By including the regression lines $x'_{1k}\beta_1 = 0$ and $x'_{2k}\beta_2 = 0$ in these inequalities, we have a quasi-complete separation in the data.

(b) However, if we modify the data so that for some k_1 and k_2 , we have $s_{1k_1} = -1$ and $s_{1k_2} = 1$, the condition specified in Theorem 1 is satisfied. Indeed for $\beta_1 < 0$, by taking $k = k_1$, we have $s_{1k_1}x_{1k_1}\beta_1 > 0$, and for $\beta_1 > 0$, by taking $k = k_2$

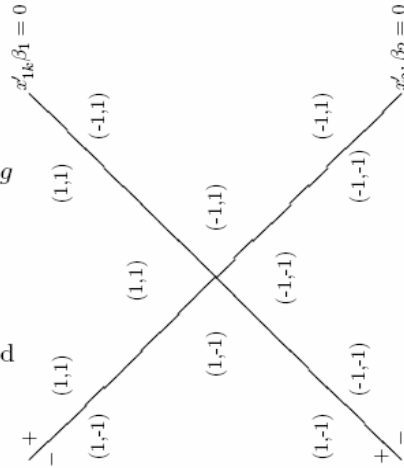


Figure 1: Schematic representation of complete separation (excluding the regression lines) and quasi-complete separation (including the regression lines)

we have $s_{1k_2}x_{1k_2}\beta_1 > 0$. Similar findings can be obtained for the second component. Therefore, the existence of the MLE for the parameters β_1 and β_2 is guaranteed only if there exist two distinct realizations (success and failure) for each corresponding outcome.

Quasi-complete separation detection and computational difficulties

We now discuss a practical method for checking the conditions for the existence of the MLE for the location parameters. Of equal importance is early detection of possible divergence problems in determining whether the vector of polychoric correlation coefficients ρ lies on the boundary. For location parameters, the existence and uniqueness conditions of the joint probability model can be related to the univariate model. It is well known that the univariate MLE of β_ℓ consistently estimates the true location parameter of the joint model. As reported by Lesaffre and Kaufmann^[6] in the multivariate probit model, the univariate solution $\tilde{\beta}_\ell$ lie close to the multivariate solution $\hat{\beta}_\ell, \ell = 1, 2, \dots, g$, in practice. This result is important for early detection of the divergence of the location parameters in the multivariate model. An informal check consists of

looking at the estimates and standard errors of $\beta_\ell, \ell = 1, 2, \dots, g$ from the univariate models. If one or more Wald statistics for the corresponding parameters converge to zero or the standard errors blow up to infinity, computational difficulties will arise in the multivariate context.

An alternative approach consists of using the linear programming strategy. This approach relies heavily on the convex theory concepts for which the concept of full-dimensional set is sometimes used interchangeably with that of quasi-complete separation. A dual cone in \mathcal{R}^p is full-dimensional if it has non-empty interior. It can be shown that the dual cone $\mathcal{C} = \{x \in \mathcal{R}^p : x'\beta \leq 0\}$ is full-dimensional if and only if the system of equations $x'\beta < 0$ has a solution. By Farkas' 1902 lemma, it follows that \mathcal{C} is full-dimensional if and only if there does not exist a nonnegative and nonzero $y \in \mathcal{R}^p$ such that $y'\beta = 0$. Hence, our conditions reduces to checking the feasibility of a system of linear equations, which is a standard problem in linear programming and which can be solved using commercial software such as CPLEX (CPLEX Optimization^[4]).

REFERENCES

1. Albert, A and J. A. Anderson, 1984. On the existence maximum likelihood estimates in logistic regression models. *Biometrika*, 71:1–10.
2. Albert, A. and J.A. Anderson, 1981. Probit and logistic discriminant functions. *Communications in Statistics*, A 10:641–657.
3. Birkes, D. and S. S. Wulff, 2003. Existence of maximum likelihood estimates in normal variance-components models. *Journal of Statistical Planning and Inference*, 113(1):35–47.
4. CPLEX Optimization, 1992. *CPLEX*. Incline Village, Nevada: CPLEX Optimization Inc..
5. Demidenko, E. and H. Massam, 1999. Existence of the maximum likelihood estimate in variance components models. *Sankhyā*, 61:431–443.
6. Lesaffre, E. and H. Kaufmann, 1992. Existence and uniqueness of the maximum likelihood estimator for a multivariate probit model. *Journal of the American Statistical Association*, 87:805–811.
7. Liang, K. Y. and S. L. Zeger, 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22.
8. Muthén, B., 1979. A structural probit model with latent variables. *Journal of the American Statistical Association*, 74:807–811.
9. Muthén, B. and L. Muthén, 1998-2004. *Mplus [Computer software]*. Muthén & Muthén, Los Angeles.
10. O'Brien, P. C., 1984. Procedures for comparing multiples endpoints. *Biometrics*, 40:1079–1087.
11. Prékopa, A., 1973. On logarithmic concave measures and functions. *Acta Scientiarum mathematicarum (Szeged)*, 34:355–343.
12. Silvapulle, M.J., 1981. On the existence of maximum likelihood estimates for binomial response models. *Journal of the Royal Statistical Society*, B 43:310–313.
13. Speckman, P. and J. Lee, and D. Sun, 2000. Existence of the mle and propriety of posteriors for a general multinomial choice model. Technical report, Department of Statistics, University of Missouri-Columbia.
14. Sun, D. and R. K. Tsutakawa, and Z. He, 2001. Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica*, 11:77–95.
15. Todem, D. and K. Kim, and E. Lesaffre, 2006. Latent-variable models for longitudinal data with bivariate ordinal outcomes. *Statistics in Medicine*, in press.
16. Wang, R. and H. L. Fei, 2003. Uniqueness of the maximum likelihood estimate of the Weibull distribution tampered failure rate model. *Communications in Statistics: Theory and Methods*, 32(12): 2321–2338.