A note on the identifiability of certain latent class models

Diana Cole

IMSAS, University of Kent, Canterbury, UK

Abstract

Wiering (2005, Statistics and Probability Letters, 75, 211-218) provides conditions for the identifiability of a class of latent models. Here we derive an alternative more general method of proving this result, which is based on standard identifiability methods involving forming Jacobians.

Key words: Computer algebra, identifiability, Latent class model, Maple, symbolic algebra

1 Introduction

Wieringen (2005) examines the identifiability of a certain class of latent models using a set of results that are specific to this type of problem. Here we show how a general symbolic method, similar in concept to that used by Goodman (1974) for latent class models, can be used to determine the identifiability of this class of latent models. A parameterisation is identifiable if it is one-to-one. Identifiability can be further classified as locally and globally identifiable, where global refers to the identifiability holding for the whole parameter domain and local identifiability refers to the identifiability holding for a neighbourhood of the parameter domain.

The latent class of models of Wieringen (2005) is defined by the unconditional probability of manifest random variable $\mathbf{X}_i = [X_{i1}, \ldots, X_{in}]$ occurring:

$$P(\mathbf{X}_i = \mathbf{x}_i; \Psi) = \bar{\theta} \prod_{j=1}^{m} \begin{pmatrix} l_j \\ x_{ij} \end{pmatrix} \bar{\pi}_{j,0}^{l_j - x_{ij}} \pi_{j,0}^{x_{ij}} + \theta \prod_{j=1}^{m} \begin{pmatrix} l_j \\ x_{ij} \end{pmatrix} \bar{\pi}_{j,1}^{l_j - x_{ij}} \pi_{j,1}^{x_{ij}} \tag{1}$$

Email address: d.j.cole@kent.ac.uk (Diana Cole).

with $\bar{z} = 1-z$ and where the parameters are $\Psi = [\theta, \pi_{1,1}, \ldots, \pi_{m,1}, \pi_{1,0}, \ldots, \pi_{m,0}]$. Wieringen (2005) show that the class of latent model is globally identifiable if

$$-1 + \prod_{j=1}^{m}(l_j + 1) \geq 2m + 1, \tag{2}$$

with the condition $\pi_{j,0} \neq \pi_{j,1}$ for any j. Here we provide an alternative proof.

The alternative approach involves comparing the rank of a Jacobian matrix. This was first considered for latent models in Goodman (1974) although the origins of this method can be attributed to Rothenberg (1971). Here an extended version of this approach is used, where a Jacobian matrix is formed by differenting an exhaustive summary of the model with respect to the parameters, with an exhaustive summary defined as a vector that uniquely defines the model. Regardless of the exhaustive summary used, the model will be non-identifiable if the rank of the Jacobian is less than the number of parameters (Cole & Morgan, 2009a). An exhaustive summary for this class of latent models are the $P(\mathbf{X}_i = \mathbf{x}_i; \Psi)$ given by equation (1). This exhaustive summary consists of $\prod_{j=1}^{m}(l_j+1)$ elements. Due to the constraint $\sum P(\mathbf{X}_i = \mathbf{x}_i; \Psi) = 1$, a simpler exhaustive summary can be created by removing one of the $P(\mathbf{X}_i = \mathbf{x}_i; \Psi)$. As there are $2m+1$ parameters and only $\prod_{j=1}^{m}(l_j+1)-1$ elements in this simpler exhaustive summary, it is obvious that a model would be non-identifiable if inequality (2) does not hold (Wieringen, 2005, Remark 5). The alternative proof that the class of latent models is actually globally identifiable if inequality (2) holds is given in Section 2. The symbolic algebra involved in the proof is executed using the symbolic algebra package Maple; the Maple worksheet is available as supplementary material.

2 Extended Jacobian Method

Proving global identifiability is achieved by first finding a simpler exhaustive summary for specific values of m and l_i using reparameterisation of the original model (Cole & Morgan, 2009a) and then generalising these results for any m and l_i using the extension theorem (Catchpole & Morgan, 1997, Theorem 6). Consider the case when $m = 3$ and $l_j = 1$ for all j. This particular case is considered as a model for a naive Bayesian network (Whiley, 1999, Chapter3). Using reparameterisation Cole & Morgan (2009a) show that a simpler exhaustive summary is

$$\mathbf{s} = \left[\theta, \bar{\theta}\pi_{1,0}, \theta\pi_{1,1}, \bar{\theta}\pi_{1,0}\pi_{2,0}, \theta\pi_{1,1}\pi_{2,1}, \bar{\theta}\pi_{1,0}\pi_{2,0}\pi_{3,0}, \theta\pi_{1,1}\pi_{2,1}\pi_{3,1}\right]^T. \tag{3}$$

Next consider changing $l_1 = 1$ to l_1. The exhaustive summary consisting of the $P(\mathbf{X}_i = \mathbf{x}_i; \Psi)$ given by equation (1) can be split into two parts. First consider the exhaustive summary terms derived from $x_{i1} = l_j$ and $x_{i1} = 0$. These terms are reparameterised in terms of s_i from equation (3). This reparameterised part of the exhaustive summary is differentiated with respect to the 7 s_i and the resulting derivative matrix is found to have rank 7. The rest of the exhaustive summary terms form the second part, as know new s_i parameters are added this is a trivial application of the extension theorem (Catchpole & Morgan, 1997, remark 7) and therefore the extended model also has full rank 7 (see Maple worksheet for details). Similar results would hold for changing a different l_j or for changing two or more l_j. Therefore (3) is a exhaustive summary for any l_j.

As m increases the exhaustive summary can be extended to

$$\mathbf{s}^e = \begin{bmatrix} \theta \\ (1-\theta)\pi_{1,0} \\ \theta\pi_{1,1} \\ (1-\theta)\pi_{1,0}\pi_{2,0} \\ \theta\pi_{1,1}\pi_{2,1} \\ \vdots \\ (1-\theta)\pi_{1,0}\pi_{2,0}\dots\pi_{m,0} \\ \theta\pi_{1,1}\pi_{2,1}\dots\pi_{m,1} \end{bmatrix}. \tag{4}$$

as long as $\pi_{i,0} \neq \pi_{i,1}$ for all i. This can also be shown to be true using the extension theorem. (The details for which are given in the Maple worksheet).

The exhaustive summary given by equation (4) can then be differentiated with respect to the original parameters to form a Jacobian matrix. If the rank of this Jacobian matrix is less than the number of parameters then the latent model is non-identifiable. This is confirmed to be true for $m \geq 3$ in the Maple worksheet. Theorem 7 of Cole & Morgan (2009a) states a model is globally identifiable if there is a unique solution to $\mathbf{s}^e = \mathbf{k}$ (where $\mathbf{s}^e$ is an exhaustive summary the same length as the number of parameters). Here the unique solution is of the form $\theta = k_1$, $\pi_{1,0} = \frac{k_2}{1-k_1}$, $\pi_{1,1} = \frac{k_3}{k_1}, \dots, \pi_{m,1} = \frac{k_{2m+1}}{k_{2m-1}}$. Therefore this latent model class is always globally identifiable for $m \geq 3$. The cases $m = 2$ and $m = 1$ are show in a similar vein in the Maple worksheet.

3 Discussion

The concept of being able to use an exhaustive summary as the basis for determining identifiability is very general and is applicable to any parametric model. The long-established Jacobian method for testing for local identifiability certainly more useable for simpler exhaustive summaries, such as those that result from the reparameterisation method. Global identifiability can also be determined easily from any simple exhaustive summary that is the same length as the number of parameters. Here we have shown how the identifiability of a particular class of latent models can be determined using this general method. It can be seen that the application of the extension theorem is very useful for generalising identifiability results.

This Jacobian method also has the advantage that it is easy to generalise to other families of latent class models such as those with covariates. Examples of latent class models with covariates are given in Huang & Bandeen-Roche (2004) and Forcina (2008). The extension of this Jacobian method to include covariates is examined in Cole & Morgan (2009b).

References

Catchpole, E. A. and Morgan, B. J. T., Detecting parameter redundancy. Biometrika, 1997, 84, 187-196.

Cole, D.J., & Morgan, B.J.T, Determining the Parametric Structure of Non-Linear Models, 2009a. University of Kent Technical Report. UKC/IMS/09/005 www.kent.ac.uk/ims/personal/djc24/para.pdf

Cole, D.J., & Morgan, B.J.T, Parameter Redundancy with Covariates. 2009b University of Kent Technical Report. UKC/IMS/07 www.kent.ac.uk/ims/personal/djc24/cova.pdf

Forcina, A. Identifiability of extended latent class models with individual covariates, 2008, Computational Statistics and Data Analysis, 52, 52635268

Huang G., Bandeen-Roche, K., Building an identifiable latent class model with covariat effects on underlying and measuring variables. Psychometrika, 69, 5-32

Goodman, L. A., Exploratory latent structure analysis using both identifiable and unidentifiable models, 1974, Biometrika, 61, 215-231.

Rothenberg, T. J., Identification in parametric models, 1971, Econometrica, 39, 577-591.

Whiley, M. Aspects of the Interface Between Statistics and Neural Networks, 1999, Ph.D Thesis University of Glasgow, Scotland.

Wieringen, W. N., On identifiability of certain latent class models, 2005, Statistics & Probability Letters, 75, 211-218.