AVOIDING BOUNDARY ESTIMATES IN LATENT CLASS ANALYSIS BY BAYESIAN POSTERIOR MODE ESTIMATION

Francisca Galindo Garre* and Jeroen K. Vermunt**

In maximum likelihood estimation of latent class models, it often occurs that one or more of the parameter estimates are on the boundary of the parameter space; that is, that estimated probabilities equal 0 (or 1) or, equivalently, that logit coefficients equal minus (or plus) infinity. This not only causes numerical problems in the computation of the variance-covariance matrix, it also makes the reported confidence intervals and significance tests for the parameters concerned meaningless. Boundary estimates can, however, easily be prevented by the use of prior distributions for the model parameters, yielding a Bayesian procedure called posterior mode or maximum a posteriori estimation. This approach is implemented in, for example, the Latent GOLD software packages for latent class analysis (Vermunt & Magidson, 2005). Little is, however, known about the quality of posterior mode estimates of the parameters of latent class models, nor about their sensitivity for the choice of the prior distribution. In this paper, we compare the quality of various types of posterior mode point and interval estimates for the parameters of latent class models with both the classical maximum likelihood estimates and the bootstrap estimates proposed by De Menezes (1999). Our simulation study shows that parameter estimates and standard errors obtained by the Bayesian approach are more reliable than the corresponding parameter estimates and standard errors obtained by maximum likelihood and parametric bootstrapping.

1. Introduction

Latent class (LC) analysis has become one of the standard methods for analyzing data in social and behavioral research. As is well known, it often occurs that maximum likelihood (ML) estimates of the parameters of LC models lie on the boundary of the parameter space; i.e., that estimated model probabilities equal 0 (or 1) or, equivalently, that estimated log-linear parameter equal plus (or minus) infinity. Though such boundary estimates are most likely to occur when the sample size is small compared to number of unknown parameters, they may also occur in other situations. Occurrence of boundary estimates not only leads to numerical problems in the computation of the parameters' asymptotic variance-covariance matrix, it also causes confidence intervals and significance tests to become meaningless.

To overcome the problems associated with the asymptotic standard errors in the case a LC model contains boundary estimates, De Menezes (1999) proposed using a parametric bootstrap approach. She showed that this computationally intensive method yields more reliable point and interval estimates for the class-specific conditional response probabilities than the ML method. However, bootstrap replicate samples may also contain boundary

estimates, which is not a problem when one is only interested in the model probabilities, but is a big problem if one is also interested in the logit parameters defining the class-specific response probabilities: A single bootstrap replicate with a boundary estimate of say minus infinity yields a point estimate of minus infinity for the parameter concerned with an asymptotic standard error of infinity. This may even occur if the ML solution does not contain boundary estimates, which shows that the bootstrap method does not seem to be very useful for inference concerning the logit parameters of a LC model.

An alternative method for dealing with the boundary problems associated with ML estimation involves making use of prior information on the model parameters. This yields a Bayesian estimation method called posterior mode or maximum a posteriori estimation, which was used in the context of LC analysis by, for example, Maris (1999) and which is implemented in the Latent GOLD program for LC analysis (Vermunt & Magidson, 2000, 2005). An important difference with the bootstrap method is that this Bayesian estimation method *prevents* boundaries from occurring by introducing external a priori information about the possible (non-boundary) values of the parameters. Another difference is that Bayesian posterior mode estimation does not take more computation time than ML estimation. In fact, it requires only a minor modification of the standard ML estimation algorithms. Note that posterior mode estimation is a computationally non-intensive variant of a computationally intensive full Bayesian analysis using Markov Chain Monte Carlo (MCMC) methods (Gelman, Carlin, Stern & Rubin, 2003). Moreover, Galindo, Vermunt & Bergsma (2004) showed for log-linear analysis of sparse contingency tables – which is also a situation in which boundaries are very likely to occur – that bias and root median squared errors of posterior mode estimates may even be smaller than those of posterior mean estimates obtained via MCMC, indicating that the former may be a good alternative to a full Bayesian analysis.

An important issue when applying Bayesian methods is the choice of the prior distribution. When no previous knowledge about the possible values of the unknown model parameters is available, it is common to take a noninformative prior, which should yield point estimates similar to the ones obtained by ML while avoiding boundary estimates. In a simulation study on the performance of various noninformative prior distributions in logit modeling with sparse contingency tables, Galindo Garre, Vermunt & Bergsma (2004) found that any of the investigated prior distributions studied yielded better point and interval estimates than standard ML. Further research is needed, however, to explore whether these results can be generalized to more complex categorical data models, such as the LC models.

The aim of the present paper is to determine whether a Bayesian approach yields better point and interval estimates for parameters of LC models than asymptotic and bootstrap procedures. The adopted Bayesian approach consists of maximum a posteriori estimation using the noninformative priors that performed best in the simulation study conducted by Galindo et al. (2004). Section 2 describes the asymptotic method based on ML theory as well as the bootstrap method for obtaining standard errors. Section 3 describes the adopted Bayesian approach in combination with various types of noninformative priors. Section 4 presents the results of a small simulation study. The paper ends with a

discussion.

2. Estimation of parameters and standard errors of latent class models

This section illustrates the numerical problems occurring in the estimation of standard errors when one or more parameter estimates of a LC model are on the boundary of the parameter space. After describing the basic LC model, we show how asymptotic and bootstrap standard errors can be obtained. These ML and bootstrap methods are illustrated with a small empirical example.

The aim of a basic LC analysis is to explain the associations among a set of K categorical response variables $(Y_1, .., Y_k, ..., Y_K)$ from the existence of a categorical latent variable X, with the intension to find meaningful latent classes. The number of categories of X (number of latent classes) is denoted by T and the number of categories of response variable Y_k by I[1]. Let $P(X = x)$ be the overall probability of belonging to latent class x and $P(Y_k = y_k | X = x)$ the conditional probability of providing response y_k to item Y_k given that one belongs to latent class x, where $1 \leq x \leq T$ and $1 \leq y_k \leq I$. In the basic LC model, we assume that for each possible response combination $(y_1, .., y_k, ..., y_K)$ and each category x of X

$$P(y_1, .., y_k, ..., y_K | X = x) = \prod_{k=1}^{K} P(Y_k = y_k | X = x).$$

In other words, responses are assumed to be mutually independent within latent classes, which is known as the local independence assumption (Goodman, 1974a, 1974b). The unconditional probability of having a certain response sequence is obtained as follows:

$$P(y_1, .., y_k, ..., y_K) = \sum_{x=1}^{T} P(X = x) \cdot P(y_1, .., y_k, ..., y_K | X = x).$$

The probabilities defining the LC model – $P(X = x)$ and $P(Y_k = y_k | X = x)$ – can also be expressed as logit models (Haberman, 1979; Heinen, 1996, pp.51–53, Vermunt, 1997); that is,

$$P(X = x) = \frac{\exp(\gamma_{x0}^X)}{\sum_{s=1}^{T} \exp(\gamma_{s0}^X)},$$

$$P(Y_k = y_k | X = x) = \frac{\exp(\beta_{y_k0}^{Y_k} + \beta_{y_kx}^{Y_kX})}{\sum_{r=1}^{I} \exp(\beta_{r0}^{Y_k} + \beta_{rx}^{Y_kX})}.$$

Here γ_{x0}^X are the logit parameters defining the class sizes, $\beta_{y_k0}^{Y_k}$ are the constants in the model for item Y_k and $\beta_{y_kx}^{Y_kX}$ are the association parameters describing the relationship between X and Y_k.[2] This parametrization is useful because it is the logit coefficients that

---

[1] In this paper, for simplicity of exposition, we assume that the number of categories is the same for all items. The LC model can, of course, also be used when this is not the case.

[2] Note that the standard identifying – effect or dummy coding – constraints are impose on these logit coefficients.

are asymptotically normally distributed rather than the probabilities themselves. Moreover, the logit parametrization makes it possible to imposed all kind of interesting model restrictions, leading, for example, to discretized variants of IRT models, as well as makes it possible to expand the LC model in various ways, for example, by including covariates affecting class membership.

Let $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ be a column vector containing all unknown logit parameters, where $\boldsymbol{\gamma}$ denotes the subvector of parameters corresponding to the latent class probabilities and $\boldsymbol{\beta}$ denotes the subvector of logit parameters corresponding to the conditional response probabilities. Note that the total number of free logit parameters equals $(T-1)+K \cdot T \cdot (I-1)$. Let $\mathbf{Z}$ be a design matrix of the form

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_0 & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_1 & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_k & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{Z}_K \end{pmatrix},$$

where $\mathbf{Z}_0$ is a $T \times (T-1)$ submatrix whose elements connect the γ parameters to the response probabilities and where, for $1 \leq k \leq K$, $\mathbf{Z}_k$ is a $(T \cdot I) \times [T \cdot (I-1)]$ submatrix whose elements connect the right β parameters to the conditional response probabilities corresponding to item Y_k. Using these submatrices, the logit of the subvectors of model probabilities can alternatively be defined as follows:

$$\text{logit } \boldsymbol{\pi}^X = \mathbf{Z}_0\, \boldsymbol{\gamma} = \mathbf{Z}_0\, \boldsymbol{\theta}_0,$$
$$\text{logit } \boldsymbol{\pi}^{Y_k/X} = \mathbf{Z}_k\, \boldsymbol{\beta}_k = \mathbf{Z}_k\, \boldsymbol{\theta}_k.$$

where $\boldsymbol{\pi}^X$ contains the T $P(X=x)$ terms and $\boldsymbol{\pi}^{Y_k/X}$ the $T \cdot I$ $P(Y_k=y_k|X=x)$ terms.

Parameter estimates of LC models are usually obtained by ML. Let $n(y_1,..,y_K)$ be the observed frequency for response pattern $(y_1,..,y_K)$ and $\mathbf{n}$ be the full vector of observed frequencies. Summing over all the response sequences, the multinomial log-likelihood, $\log p(\mathbf{n}|\boldsymbol{\theta})$, is

$$\log p(\mathbf{n}|\boldsymbol{\theta}) = \sum n(y_1,..,y_K) \log P(y_1,...,y_k,...,y_K) + \text{Constant} \tag{1}$$

The most popular algorithms for maximizing equation (1) under a LC model are the Expectation-Maximization (EM) (Dempster, Laird & Rubin, 1977), Fisher scoring (Haberman, 1979), and Newton-Raphson algorithms.

2.1 Asymptotic standard errors

Once the ML estimates of the parameters have been computed, standard errors are needed to assess their accuracy and to test their significance. Asymptotic standard errors for $\widehat{\boldsymbol{\theta}}$ estimates can directly be obtained as the square root of the main diagonal

elements of the inverse of the Fisher information matrix, $i(\boldsymbol{\theta})$. Because $P(X = x)$ and $P(Y_k = y_k|X = x)$ are functions of the θ parameters, their asymptotic standard errors (ASE) can be obtained by the delta method; that is, by

$$ASE[P(X = x)] = \sqrt{\frac{\partial P(X = x)}{\partial \boldsymbol{\gamma}} i(\boldsymbol{\theta})^{-1} \frac{\partial P(X = x)}{\partial \boldsymbol{\gamma}}'} ,$$

$$ASE[P(Y_k = y_k|X = x)] = \sqrt{\frac{\partial P(Y_k = y_k|X = x)}{\partial \boldsymbol{\beta}} i(\boldsymbol{\theta})^{-1} \frac{\partial P(Y_k = y_k|X = x)}{\partial \boldsymbol{\beta}}'} .$$

The Fisher information matrix – i.e., the expected information matrix – is defined as minus the expected value of the matrix of second-order derivatives to all model parameters: $i(\boldsymbol{\theta}) = E(-\frac{\partial^2 \log p(\mathbf{n}|\theta)}{\partial \theta \partial \theta})$. For any categorical data model, an element (ℓ, m) of the information matrix is obtained as follows:

$$i(\theta_\ell, \theta_m) = -N \sum \frac{1}{P(y_1, .., y_k, ..., y_K)} \frac{\partial P(y_1, .., y_k, ..., y_K)}{\partial \theta_\ell} \frac{\partial P(y_1, .., y_k, ..., y_K)}{\partial \theta_m}, \quad (2)$$

where θ_ℓ and θ_m are the logit parameters corresponding with rows ℓ and m of the design matrix $\mathbf{Z}$ and N is the total sample size. Denoting the probability of belonging to latent class x and response sequence $(y_1, ..., y_K)$ by $P(x, y_1, ..., y_K)$, the necessary partial derivative for θ_ℓ can be obtained by

$$\frac{\partial P(y_1, .., y_k, ..., y_K)}{\partial \theta_\ell} = \sum_{x=1}^{T} P(x, y_1, ..., y_K) \left[z_{x\ell} - \sum_{r=1}^{T} z_{r\ell} P(X = x) \right]$$

for the parameters corresponding to the latent class probabilities, and by

$$\frac{\partial P(y_1, .., y_k, ..., y_K)}{\partial \theta_\ell} = \sum_{x=1}^{T} P(x, y_1, ..., y_K) \left[z_{ky_k x\ell} - \sum_{s=1}^{I} z_{ksx\ell} P(Y_k = y_k|X = x) \right] ,$$

for the parameters corresponding to the conditional responses probabilities (for a detailed description, see Bartholomew & Knott, 1999, p. 40, and De Menezes, 1999). In presence of parameter estimates on the boundary of the parameter space, the matrix $i(\boldsymbol{\theta})$ is not full rank, which means that its (standard) inverse is not defined. As a consequence, neither interval estimates nor significance tests can be obtained by the usual procedure.[3]

2.2 Bootstrap standard errors

De Menezes (1999) proposed obtaining estimates for the standard errors in a LC analysis by a parametric bootstrap, which implies that their values are approximated by a Monte Carlo simulation (Efron & Tibshirani, 1993). More specifically, B replication samples of the same size as the original sample are generated from the population defined by

---

[3] Standard errors for the non-boundary parameters are usually obtained by taking the generalized inverse of the information matrix. These standard errors are valid only under the conditional that the boundary estimates can be seen as a priori model constraints.

the ML solution obtained for the model of interest. For each of these replication samples we re-estimate the model of interest using ML. The bootstrap parameter values and their standard errors are respectively the means and standard deviations of the B sets of parameter estimates. Using an extended simulation study, De Menezes (1999) showed that, in general, the parametric bootstrap procedure yields accurate point estimates and standard errors for the conditional response probabilities defining a LC model. She, however, did not investigate the performance of the method for logit parameters.

2.3 Empirical example

We will now illustrate the impact of the occurrence of boundary estimates using an empirical example. Table 1 reports women's answers to five opinion items about gender roles in a Dutch survey taken from Heinen (1996, pp.44–49). Heinen found that while the LC model with two classes had to be rejected $\left(G^2 = 53.78, \chi^2 = 54.30, d.f. = 20\right)$, the LC model with three latent classes provided a good fit $\left(G^2 = 17.53, \chi^2 = 19.31, d.f. = 14\right)$. However, some of the conditional probabilities – $P(Y_2 = 1|X = 1)$, $P(Y_3 = 1|X = 3)$, and $P(Y_5 = 1|X = 3)$ – took values of zero or one, resulting in numerical problems when calculating the corresponding standard errors. Table 2 gives ML and bootstrap parameter estimates and standard errors for the model probabilities. As can be seen, none the bootstrap estimates is on the boundary of the parameter space and reasonable standard errors are obtained for all parameters, also for the ones for which ML yielded a boundary estimate.

Table 3 reports the same information for the logit parameters. As can be seen, both ML and bootstrap methods produce estimates close to plus or minus infinity for $\beta_{11}^{Y_2 X}$, $\beta_1^{Y_3}$, $\beta_{11}^{Y_3 X}$, $\beta_{12}^{Y_3 X}$, $\beta_1^{Y_5}$, $\beta_{11}^{Y_5 X}$, and $\beta_{12}^{Y_5 X}$. Although bootstrap standard errors exist for these parameters, their values are extremely large, which shows that the parametric bootstrap does not always yield reasonable standard errors for logit coefficients. The reason for this is that some replicate samples contain (close to) boundary estimates.

The cause of the poor performance of the parametric bootstrap with the logit parameterization is that the range of values for logit parameters is between $[-\infty, +\infty]$, whereas conditional probabilities are constrained to lie in the interval $[0, 1]$. Extreme parameter estimates in the replication samples will thus affect the means and variances more seriously for logit parameters than for probabilities.

Table 1: Women's answers to five gender role opinion items (observed frequencies).

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
11111	23	12111	3	21111	4	22111	2
11112	0	12112	0	21112	1	22112	1
11121	49	12121	45	21121	13	22121	45
11122	14	12122	29	21122	10	22122	47
11211	0	12211	2	21211	0	22211	0
11212	0	12212	0	21212	1	22212	1
11221	4	12221	15	21221	7	22221	42
11222	3	12222	51	21222	3	22222	177

Table 2: Estimates of the conditional probabilities for Heinen's data.

	ML		Bootstrap		Jeffreys		Normal		Dirichlet		LG	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$P(X=1)$	0.46	0.06	0.49	0.04	0.49	0.03	0.50	0.09	0.55	0.05	0.46	0.08
$P(X=2)$	0.43	0.05	0.40	0.04	0.38	0.03	0.35	0.07	0.25	0.06	0.42	0.06
$P(X=3)$	0.11	0.02	0.11	0.03	0.13	0.01	0.15	0.06	0.20	0.04	0.12	0.03
$P(Y_1=1|X=1)$	0.19	0.03	0.22	0.03	0.20	0.02	0.20	0.03	0.23	0.03	0.19	0.03
$P(Y_1=1|X=2)$	0.50	0.05	0.48	0.04	0.51	0.04	0.48	0.09	0.49	0.08	0.49	0.07
$P(Y_1=1|X=3)$	0.91	0.06	0.89	0.07	0.90	0.01	0.87	0.08	0.80	0.06	0.90	0.07
$P(Y_2=1|X=1)$	0.00	—	0.05	0.01	0.01	0.01	0.02	0.02	0.04	0.02	0.00	0.02
$P(Y_2=1|X=2)$	0.28	0.06	0.19	0.05	0.28	0.04	0.26	0.11	0.26	0.09	0.27	0.07
$P(Y_2=1|X=3)$	0.93	0.08	0.92	0.01	0.90	0.01	0.83	0.11	0.74	0.08	0.90	0.09
$P(Y_3=1|X=1)$	0.10	0.06	0.30	0.02	0.12	0.04	0.12	0.08	0.20	0.04	0.10	0.07
$P(Y_3=1|X=2)$	0.76	0.05	0.78	0.02	0.79	0.06	0.79	0.11	0.79	0.09	0.76	0.08
$P(Y_3=1|X=3)$	1.00	—	0.99	0.00	0.98	0.00	0.95	0.04	0.90	0.04	0.99	0.03
$P(Y_4=1|X=1)$	0.00	0.01	0.01	0.02	0.01	0.00	0.01	0.01	0.02	0.01	0.00	0.01
$P(Y_4=1|X=2)$	0.04	0.02	0.03	0.00	0.03	0.01	0.03	0.03	0.07	0.04	0.04	0.02
$P(Y_4=1|X=3)$	0.41	0.08	0.37	0.10	0.41	0.04	0.34	0.10	0.29	0.06	0.39	0.09
$P(Y_5=1|X=1)$	0.14	0.04	0.24	0.02	0.15	0.03	0.16	0.04	0.19	0.03	0.14	0.04
$P(Y_5=1|X=2)$	0.59	0.06	0.74	0.03	0.61	0.05	0.60	0.11	0.65	0.09	0.59	0.08
$P(Y_5=1|X=3)$	1.00	—	0.99	0.03	0.96	0.01	0.92	0.06	0.83	0.05	0.98	0.04

Table 3: Estimates of the logit parameters for Heinen's data.

	ML		Bootstrap		Jeffreys		Normal		Dirichlet		LG	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$\beta_1^{X_1}$	1.43	0.30	1.51	0.32	1.35	0.09	1.21	0.52	1.02	0.25	1.35	0.41
$\beta_2^{X_2}$	1.35	0.24	1.29	0.32	1.10	0.14	0.87	0.44	0.23	0.39	1.25	0.28
$\beta_1^{Y_1}$	2.36	0.78	3.77	5.40	2.15	0.10	1.86	0.68	1.38	0.39	2.22	0.74
$\beta_{11}^{Y_1X}$	−3.80	0.78	−4.72	5.46	−3.55	0.17	−3.22	0.67	−2.60	0.41	−3.65	0.73
$\beta_{12}^{Y_1X}$	−2.38	0.80	−4.30	5.40	−2.13	0.20	−1.92	0.66	−1.40	0.57	−2.26	0.74
$\beta_1^{Y_2}$	2.61	1.22	5.81	7.43	2.21	0.16	1.61	0.79	1.06	0.44	2.25	1.00
$\beta_{11}^{Y_2X}$	−33.78	—	−16.86	12.90	−6.42	0.77	−5.74	1.21	−4.12	0.56	−8.02	4.97
$\beta_{12}^{Y_2X}$	−3.55	1.22	−13.61	12.45	−3.18	0.28	−2.66	0.77	−2.11	0.69	−3.23	0.96
$\beta_1^{Y_3}$	22.06	—	21.91	7.58	3.67	0.20	2.90	0.88	2.20	0.48	4.88	3.47
$\beta_{11}^{Y_3X}$	−24.23	—	−23.17	7.88	−5.68	0.36	−4.87	1.06	−3.61	0.52	−7.05	3.41
$\beta_{12}^{Y_3X}$	−20.90	—	−23.13	8.09	−2.33	0.44	−1.55	1.23	−0.87	0.86	−3.74	3.52
$\beta_1^{Y_4}$	−0.34	0.35	−0.49	1.01	−0.34	0.18	−0.67	0.43	−0.91	0.28	−0.43	0.37
$\beta_{11}^{Y_4X}$	−5.74	2.24	−9.96	10.05	−4.65	0.71	−4.41	1.11	−2.87	0.55	−5.50	2.01
$\beta_{12}^{Y_4X}$	−2.82	0.51	−7.33	8.88	−3.01	0.42	−2.78	0.84	−1.68	0.73	−2.80	0.57
$\beta_1^{Y_5}$	22.50	—	18.99	9.20	3.20	0.25	2.46	0.82	1.58	0.38	4.02	2.32
$\beta_{11}^{Y_5X}$	−24.32	—	−19.98	9.59	−4.92	0.29	−4.13	0.82	−3.04	0.42	−5.83	2.27
$\beta_{12}^{Y_5X}$	−22.13	—	−19.47	9.23	−2.73	0.35	−2.05	0.89	−0.95	0.64	−3.66	2.29

* A dummy coding is used and the parameters fixed to zero are not reported

3. Estimation of parameters and standard errors of latent class model using Bayesian methods

In contrast to ML and bootstrap methods, Bayesian methods assume that parameters $\boldsymbol{\theta}$ are random variables rather than taking fixed values that have to be estimated. As a consequence, probability models are used to fit the data. The main advantage of these models is that the information coming from the current sample may be combined with available knowledge on the parameters' values summarized in a prior distribution $p(\boldsymbol{\theta})$. The posterior distribution $p(\boldsymbol{\theta}|\mathbf{n})$, representing the updated parameters' distribution given the data $\mathbf{n}$, can be determined by applying the Bayes' rule,

$$p(\boldsymbol{\theta}|\mathbf{n}) = \frac{p(\mathbf{n}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{\int p(\mathbf{n}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto p(\mathbf{n}|\boldsymbol{\theta})\,p(\boldsymbol{\theta}),$$

where, in our case, $p(\mathbf{n}|\boldsymbol{\theta})$ is the exponential of the log-likelihood function defined in equation (1).

By using a prior distribution, parameter estimates can be smoothed so that they stay within the parameter space. When no previous knowledge about the parameters is available, it is most common to use a noninformative prior, in which case the contribution of the data to the posterior distribution will be dominant. Below, three types of prior distributions for the Bayesian estimation of LC models are presented: Jeffreys' prior, normal priors, and Dirichlet priors.

3.1 Jeffreys' prior

A commonly used prior in Bayesian analysis is the Jeffreys' prior (Jeffreys, 1961). This prior is obtained by applying Jeffreys' rule, which involves taking the prior density to be proportional to the square root of the determinant of the Fisher information matrix $i(\boldsymbol{\theta})$; that is,

$$p(\boldsymbol{\theta}) \propto |i(\boldsymbol{\theta})|^{\frac{1}{2}},$$

with $|\cdot|$ denoting the determinant. In the case of the LC model and taking a logit formulation, the elements of $i(\boldsymbol{\theta})$ can be obtained with equation (2).

An important property of Jeffreys' prior is its invariance under scale transformations of the parameters. This means, for example, that it does not make any difference whether the prior is specified for the logit parameters or for the probabilities defining the LC model, or whether we use dummy or effect coding for the logit parameters. As Gill (2002) pointed out, other relevant properties of the Jeffreys' prior are that it mediates as little subjective information into the posterior distribution as possible since it comes from a mechanical process in which the researcher does not need to determine the prior's hyperparameters. Firth (1993) showed for various generalized linear models that parameter estimates obtained by using Jeffreys' prior present less bias than the corresponding ML estimates. Similar results were obtained in a simulation study conducted by Galindo et al. (2004) for logit models. However, more research is needed to determine whether the Jeffreys' prior also has good frequentist properties under more complex models, such as LC models.

3.2 Univariate normal priors

An alternative to using a prior based on a statistical rule, is to assume a normal prior distribution for the θ parameters. When no information about the dependence between parameters is available, it is convenient to adopt a set of univariate normal priors.

Normal distributions have repeatedly been used as prior distributions for logit models and also for latent trait models. For instance, the example "LSAT: latent variable models for item response data" that appears in the manual of the WinBUGS computer program (Gilks, Thomas & Spiegelhalter, 1994) makes use of normal priors. Congdon (2001) suggested that noninformative priors may be approximated in WinBUGS by taking univariate normal distributions with means equal to zero and large variances. The effect of using normal priors with means of 0 is that parameter estimates will be smoothed toward zero. The prior variances determine the amount of prior information that is added, implying that the smoothing effect can be decreased by using larger variances.

3.3 Dirichlet priors

In contrast to the normal priors presented above which are defined for the logit parameters, the Dirichlet prior is defined for the vectors of latent class probabilities and conditional response probabilities, denoted as $\boldsymbol{\pi}^X$ and $\boldsymbol{\pi}^{Y_k/X}$ respectively. More specifically, the Dirichlet prior for the latent class probabilities equals

$$p(\boldsymbol{\pi}^X) \propto \prod_{x=1}^{T} P(X=x)^{\alpha_x - 1}, \tag{3}$$

and for the conditional probabilities corresponding to item Y_k

$$p(\boldsymbol{\pi}^{Y_k|X}) \propto \prod_{x=1}^{T} \prod_{y_k=1}^{I} P(Y_k = y_k|X=x)^{\alpha_{y_k x} - 1}. \tag{4}$$

Here, the α terms are the hyperparameters defining the Dirichlet. These hyperparameters can be interpreted as a number of observations in agreement with a certain model that is added to the data. If there is no previous information about the hyperparameters, the α terms may be chosen to be constant, in which case the added observations are in agreement with a model in which the cases are uniformly distributed over the cells of the frequency table. For example, suppose that one decides to add L observations to the data. If a constant value is chosen for the α_x and $\alpha_{y_k x}$ parameters, respectively, then each $\alpha_x = 1 + \frac{L}{T}$, with T denoting the number of latent classes, and each $\alpha_{y_k x} = 1 + \frac{L}{T \cdot I}$, with I denoting the number of categories for Y_k.

It is sometimes desirable to use α parameters that are in agreement with a more realistic model. For example, the Latent GOLD computer program (Vermunt & Magidson, 2000, 2005) uses Dirichlet priors for the conditional probabilities which preserve the observed univariate marginal distributions of the Y_k's; that is, which is in agreement with the independence model. More precisely, if we add L observations to the data, $\alpha_x = 1 + \frac{L}{T}$

and $\alpha_{y_k x} = 1 + \frac{n(y_k) \cdot L}{N \cdot T}$, where $n(y_k)$ is the observed marginal frequency for $Y_k = y_k$. A similar type of Dirichlet prior was proposed by Clogg et al. (1991) and Schafer (1997) for log-linear and logit models.

3.4 Estimators and algorithms

Two types of estimation methods can be used within a Bayesian context. The more standard approach focuses on obtaining information on the entire posterior distribution of the parameters by sampling from this distribution using Markov Chain Monte Carlo techniques. The point estimator usually used under this perspective is the mean of the posterior distribution, called mean a posteriori or expected a posteriori estimator. The second method, which is more similar to classical approach, is oriented to find a single optimum estimate; that is, the estimated values corresponding to the maximum of the posterior distribution. This point estimator is called the posterior mode or maximum a posteriori (MAP) estimator. In the example and in the simulation study of the next section, we worked with this MAP estimator. The two main advantages of this estimator are: (1) it is computationally less intensive than the expected a posteriori estimator, and (2) it can be obtained with minor adaptations of ML algorithms. As already mentioned, Galindo et al. (2004) found that the MAP estimators is more reliable than the EAP estimator in logit modeling with sparse tables.

To obtain the MAP estimates, one can, for example, use and EM algorithm. The algorithm that we used was a hybrid: we started with a number of EM iterations and switched to Newton-Raphson when EM was close to reach the convergence criterion.[4]

As shown by Gelman et al. (2003, pp.101–108), the posterior density can usually be approximated by a normal distribution centered at the mode; that is,

$$p(\boldsymbol{\theta}|\mathbf{y}) \approx N\left(\boldsymbol{\theta}, [I(\boldsymbol{\theta})]^{-1}\right),$$

where $I(\boldsymbol{\theta})$ is the observed information matrix, which is defined as minus the matrix of second-order derivatives towards all model parameters, $I(\boldsymbol{\theta}) = -\frac{\partial^2 \log p(\mathbf{n}|\theta)}{\partial\theta\partial\theta}$. Gelman et al. (2003) demonstrated that, given a large enough sample size, the standard asymptotic theory underlying ML can also be applied to the MAP estimator. As a result, one can, for example, obtain 95% confidence intervals by taking MAP point estimates plus and minus two standard errors, with the standard errors estimated from the diagonal elements of the inverse of $I(\boldsymbol{\theta})$ evaluated at the MAP estimates of $\boldsymbol{\theta}$. Note that contrary to the ML case, here, numerical problems are avoided because of the prior constrains the parameter estimates lie within the parameter space.

3.5 Empirical example (continued)

To illustrate Bayesian posterior mode estimation of LC model probabilities and logit

---

[4] In M step of the EM algorithm, we used analytical derivatives of the complete data log-likelihood and the log priors, except for the derivatives of the Jeffreys' prior, which were obtained numerically. In the Newton-Raphson procedure, we used numerical derivatives of the log-posterior.

parameters, we reestimated the three-class model with the data from Table 1 using this method. Table 2 and 3 report MAP parameter estimates and their respective standard errors obtained with the following priors: Jeffreys' prior, Dirichlet priors with constant hyperparameters corresponding to adding a single case to the analysis, Dirichlet priors with hyperparameters that preserve the observed distributions of the items which again corresponds to adding a single case to the analysis (this prior is labeled LG because it is the one used by Latent GOLD), and normal priors with zero means and variances equal to 10. The chosen values of hyperparameters for the Dirichlet and normal distributions correspond to best performing values according to the study by Galindo et al. (2004) for the case of logit modeling.

For the conditional probabilities, Table 2 shows that MAP estimates and the corresponding standard errors are very similar to those obtained using parametric bootstrapping. Regarding the logit parameters, Table 3 shows that, contrary to the bootstrap, the Bayesian methods overcome the presence of logit parameters close to the boundary because the prior distributions smooth the parameters to be within the parameter space. Comparison of the standard errors obtained with the Bayesian approaches to the ones of the parametric bootstrap shows that the former produce smaller and more reasonable standard errors than the latter. However, a simulation study is needed to determine whether these results can be generalized to other data sets.

4. A simulation experiment

This section presents the results from a small simulation study that was conducted to determine whether Bayesian methods produce more reliable point estimates and standard errors than the parametric bootstrap and the ML approach, as well as to determine which prior distribution yields the best point and interval estimators. Samples were generated from a LC model for binary items. The population parameters were chosen to be close to the ML estimates obtained with the empirical example described above. We varied the following design factors:

1. the number of latent classes: either $T = 3$ with unconditional latent class probabilities equal to $P(X = 1) = 0.46$, $P(X = 2) = 0.43$ and $P(X = 3) = 0.11$, or $T = 4$ with class sizes equal to $P(X = 1) = 0.40$, $P(X = 2) = 0.30$, $P(X = 3) = 0.20$, and $P(X = 4) = 0.10$;
2. the sample size: either $N = 100$ or $N = 1000$;
3. the number of items: either $K = 5$ or $K = 9$.

By crossing these three factors, we obtained a design with eight configurations. The population model contained conditional response probabilities of various sizes, but we made sure that several of them were (somewhat) close to the boundary. More specifically, as extreme values we used response probabilities equal to 0.01, 0.05, and 0.10, each of which corresponds to one or more large negative or positive logit parameters.[5] Note that

---

[5] In the four-class nine-item design, the population values for the probabilities of giving the 1 re-

logit parameters are a function of various probabilities and vice versa, so there is no simple relationship between a single probability and a single logit coefficient.

For each of the eight condition mentioned above, we simulated 1000 data sets. For each of these data sets, the parameters of the LC model were estimated by ML and MAP, where MAP estimates were obtained under the same four prior distributions as we used in our empirical example. For each data set, we also executed a parametric bootstrap based on 400 replication samples.

The parameter point estimates are evaluated using the median of the ML and MAP estimates and the root median square errors (RMdSE); that is, the square root of the median of $(\widehat{\theta} - \theta)^2$. Although mean square errors are more common statistics for measuring the quality of point estimates, we used median square errors instead to avoid the effect that a few extreme values may have on the mean. It should be noted that for many simulated samples the ML estimates will be equal to plus (or minus) infinity. A single occurrence of infinity will give a mean of infinity and a root mean square error of infinity, which shows that these measures are not very informative for the performance of ML and that median-based measures are better suited for the comparison of Bayesian, bootstrap, and ML estimates. To evaluated the accuracy of the SE estimates, we report their medians, as well as the coverage probabilities and the median widths of the confidence intervals. By definition, a 95% confidence interval should have a coverage probability of at least 0.95. However, even if the true coverage probability equals 95%, the coverage probabilities coming from the simulation experiment will not be exactly equal to 0.95 because of Monte Carlo error. This error tends to zero when the number of replications tends to infinity. Since we worked with 1000 replications, the Monte Carlo standard error was equal to $\sqrt{\frac{0.95 \cdot 0.05}{1000}} = 0.007$, which means that coverage probabilities between 0.936 and 0.964 are in agreement with the nominal level of 95%.

The confidence intervals were obtained using the well-known normal approximation

$$[\widehat{\theta} - z_{\alpha/2}\widehat{\sigma}(\widehat{\theta}), \widehat{\theta} + z_{\alpha/2}\widehat{\sigma}(\widehat{\theta})] \tag{5}$$

where $\widehat{\theta}$ represents the ML or MAP estimate and $\widehat{\sigma}(\widehat{\theta})$ the estimated asymptotic standard error, which is obtained by taking the square root of the corresponding diagonal elements of estimated variance-covariance matrix. Note that this normal approximation cannot be used for obtaining confidence intervals of latent and conditional probabilities unless a logit transformation is applied to these parameters.

Rather than showing what happens with each of the model parameters, for simplicity of exposition, we will concentrate on results obtained for two typical $\beta_{y_k x}^{Y_k X}$ parameters, one that is non-extreme and another that is extreme. That gives us the possibility to demonstrate the impact of the various methods for the two most interesting cases; namely, the situation in which a boundary estimate is unlikely to occur and a situation in which a

sponse were equated to 0.2, 0.01, 0.1, 0.01, 0.15, 0.01, 0.01, 0.4, and 0.15 for class 1; 0.5, 0.3, 0.75, 0.05, 0.6, 0.3, 0.75, 0.05, and 0.6 for class 2; 0.9, 0.95, 0.99, 0.4, 0.99, 0.75, 0.9, 0.2, and 0.9 for class 3; and 0.4, 0.8, 0.7, 0.6, 0.5, 0.9, 0.6, 0.7, and 0.99 for class 4. The population values for the other (smaller) designs were obtained from these by omitting the ones for last four items in the five-item designs and for the last class in the three-class designs.

Table 4: Estimates for $\beta_{21}^{Y_1 X} = -2.20$ in the three-class model.

	N = 100					N = 1000				
	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths
K = 5										
Boot.	−6.03	15.15	7.87	0.69	∞	−3.14	3.39	0.99	0.99	13.28
ML	−7.35	1.07	7.79	0.55	∞	−2.76	0.61	0.72	0.99	∞
Jeff.	−2.31	1.37	1.03	0.82	5.37	−2.11	0.55	0.41	0.87	2.16
Norm.	−1.40	1.53	1.11	0.84	6.02	−1.91	0.50	0.41	0.81	1.94
Dir.	−1.11	1.53	1.34	0.72	6.02	−1.67	0.49	0.53	0.80	1.93
LG	−2.08	1.50	1.21	0.91	5.86	−2.06	0.60	0.43	0.92	2.34
K = 9										
Boot.	−9.09	10.78	8.77	0.62	∞	−3.30	1.60	1.10	0.95	6.27
ML	−3.13	1.08	1.73	0.78	∞	−2.25	0.48	0.26	1.00	∞
Jeff.	−1.83	1.03	0.97	0.81	4.05	−2.16	0.43	0.34	0.94	1.68
Norm.	−1.16	0.93	1.15	0.69	3.65	−2.11	0.42	0.27	0.96	1.66
Dir.	−2.67	1.01	1.05	0.81	3.97	−2.26	0.44	0.26	0.97	1.74
LG	−1.99	1.08	1.14	0.81	4.23	−2.26	0.47	0.29	0.97	1.83

Table 5: Estimates for $\beta_{11}^{Y_1 X} = -0.98$ in the four-class model.

	N = 100					N = 1000				
	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths
K = 5										
Boot.	−13.59	25.75	19.91	0.49	∞	−0.15	4.78	1.17	0.96	18.74
ML	−0.27	1.17	12.92	0.54	∞	−0.70	0.50	0.43	0.98	∞
Jeff.	−1.24	1.42	2.41	0.56	5.56	−0.19	0.65	0.80	0.88	2.56
Norm.	−1.16	1.34	0.70	0.96	5.26	−0.71	0.51	0.38	0.97	2.01
Dir.	−0.73	1.69	1.01	0.98	6.61	−0.52	0.59	0.49	0.95	2.33
LG	−0.49	2.63	1.68	0.98	10.33	−0.63	0.58	0.44	0.93	2.28
K = 9										
Boot.	−0.12	12.14	5.17	0.74	∞	−0.74	0.53	0.33	0.95	2.07
ML	−0.66	0.94	0.98	0.84	∞	−0.84	0.32	0.24	1.00	∞
Jeff.	−0.35	1.03	1.00	0.89	4.06	−0.90	0.33	0.18	0.97	1.29
Norm.	−0.79	0.91	0.70	0.91	3.59	−0.86	0.32	0.22	0.94	1.27
Dir.	−0.79	0.80	0.79	0.86	3.14	−0.72	0.33	0.27	0.90	1.30
LG	−0.65	1.04	1.03	0.93	4.07	−0.90	0.33	0.19	0.96	1.28

boundary is very likely to occur. More specifically, Tables 4 and 5 report the results for the non-extreme population parameter $\beta_{21}^{Y_1 X}$ for the three-class model and $\beta_{11}^{Y_1 X}$ for the four-class model, whereas Tables 6 and 7 gives the same information for the extreme population parameter $\beta_{11}^{Y_3 X}$ for the three-class model and $\beta_{11}^{Y_2 X}$ for the four-class model.

As can be seen from these four tables, each of methods investigated in this study - bootstrap, ML, and Bayesian methods - provides more accurate point and interval estimates in the non-extreme parameter with large sample case, which is, of course, the easiest

situation to deal with. Overall the Bayesian estimators turn out to perform better than parametric bootstrap and ML (see, for example, the lower RMdSE values), with really large differences for the most difficult situation (extreme parameter with small sample). The fact that coverage probabilities for the Bayesian estimators are sometimes lower than 0.95, especially for the smaller sample size and the extreme parameter, indicates that the normal approximation used for constructing the confidence interval is failing. It should be noted that though in some cases parametric bootstrap and ML yield coverage probabilities that are closer to the nominal level than MAP, in these situations the corresponding median widths are much larger for the MAP estimates.

As far as the point estimates and standard errors for the non-extreme parameters are concerned, Tables 4 and 5 illustrate that the parametric bootstrap may yield more extreme point estimates and standard errors than ML and MAP even when the sample size is large. This is because bootstrap estimates and standard errors are the means and standard deviations of the ML estimates for the replicated samples. Note that even if only a small portion of the bootstrap replicates yield an extreme estimate for the parameter concerned, the bootstrap mean and standard deviation will be severely affected. When comparing the various prior distributions with one another, one can see that in the three-class model the normal prior produces median estimates that are lower than the population values and large RMdSEs, which indicates that it tends to smooth parameter estimates somewhat too much toward zero. The Jeffreys' prior achieves accurate point estimates in the three-class model but not in the four-class model. As far as the interval estimates is concerned, Tables 4 and 5 illustrate that the coverage probabilities are often under the nominal level. Overall the LG prior seems to perform slightly better than the other priors in the non-extreme parameters case.

For the (more interesting) extreme parameter situation, we see much larger differences in performance among the four priors (see Tables 6 and 7). The point estimates are more accurate with the LG prior (MAP medians close to the population parameters and small RMdSE's) than with the other ones investigated. The Jeffreys' and the Normal priors perform very well when the sample size is 1000, but they too strongly smooth the parameter estimates toward zero if the sample size is 100. The Dirichlet prior also smooths the parameter estimates too much toward zero in the model with 5 items, but not in the model with 9 items. Note that the same amount of information (one observation) is added in both the models with 5 and 9 items, but this observation is spread out over the 32 cells of the contingency table is the first case and over 512 cells in the second case. According to Table 7, the Dirichlet prior performs very badly for the four-class model with 9 items, but despite of that still much better than ML.

The information on the interval estimates that is reported in Tables 6 and 7 (median widths and coverage probabilities) shows that the Bayesian methods perform much better than ML and bootstrap for the extreme parameters. It should, however, be noted that for the bootstrap this is partially the result of the procedure we use to compute confidence intervals. If we would compute confidence intervals based on bootstrap percentiles, the coverage probabilities can be expected to be larger than with the normal approximation. We, however, decided to use the normal approximation because we were more interested

Table 6: Estimates for $\beta_{11}^{Y_3 X} = -6.79$ in the three-class model.

| | $N = 100$ | | | | | $N = 1000$ | | | | |
	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths
$K=5$										
Boot.	-27.04	16.31	22.40	0.42	∞	-15.81	12.27	9.00	0.75	∞
ML	-32.35	2.85	25.54	0.30	∞	-6.71	1.39	4.44	0.62	∞
Jeff.	-4.23	1.45	2.15	0.66	5.72	-5.52	0.96	1.27	0.70	3.75
Norm.	-3.53	1.16	3.26	0.20	4.55	-5.02	0.81	1.77	0.40	3.19
Dir.	-3.74	1.52	3.05	0.46	5.97	-4.66	0.68	2.13	0.72	2.67
LG	-5.05	2.49	1.75	0.84	9.75	-6.37	1.86	0.87	0.92	7.27
$K=9$										
Boot.	-26.59	12.92	21.65	0.39	∞	-10.62	9.67	3.88	0.99	37.90
ML	-24.07	1.48	17.00	0.54	∞	-7.18	2.24	0.79	0.99	∞
Jeff.	-4.56	1.32	2.24	0.55	5.18	-6.03	0.87	0.77	0.79	3.42
Norm.	-3.84	1.01	2.95	0.20	3.95	-5.79	0.74	1.00	0.73	2.90
Dir.	-8.99	3.17	4.59	0.53	12.42	-7.10	1.79	0.70	0.89	7.02
LG	-5.21	2.32	1.72	0.71	9.11	-6.82	1.72	0.59	0.95	6.74

Table 7: Estimates for $\beta_{11}^{Y_2 X} = -5.98$ in the four-class model.

| | $N = 100$ | | | | | $N = 1000$ | | | | |
	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths	Medians	Median SE	RMdSE	Coverage Probabilities	Median Widths
$K=5$										
Boot.	-32.99	21.79	28.78	0.32	∞	-16.36	13.69	10.31	0.76	∞
ML	-32.92	121.92	27.48	0.35	∞	-11.81	2.57	5.76	0.67	∞
Jeff.	-4.96	1.68	1.90	0.64	6.58	-5.41	1.05	0.65	0.90	4.10
Norm.	-3.47	1.69	2.51	0.75	6.63	-5.38	1.14	0.63	0.92	4.45
Dir.	-4.40	2.05	1.59	0.86	8.04	-5.67	1.50	0.70	0.90	5.88
LG	-5.28	3.36	1.71	0.88	13.17	-6.51	2.50	1.07	0.87	9.79
$K=9$										
Boot.	-27.72	19.64	28.34	0.33	∞	-7.86	7.31	2.37	0.92	28.65
ML	-30.83	138.09	24.93	0.43	∞	-6.11	0.79	0.59	0.95	∞
Jeff.	-4.85	1.59	1.30	0.82	6.23	-5.90	0.72	0.44	0.96	2.82
Norm.	-4.21	1.44	1.77	0.76	5.66	-5.74	0.69	0.47	0.92	2.69
Dir.	-4.46	0.70	1.87	0.38	6.76	-4.92	0.52	1.06	0.46	2.03
LG	-6.05	3.34	1.54	0.88	13.10	-6.14	0.80	0.52	0.96	3.14

in the quality of the standard errors than in the accuracy of the interval estimates. As was already mentioned, the interval estimates obtained with the Bayesian methods have sometimes coverage probabilities lower that the nominal level.

Comparison of the median of the estimated standard errors for ML with the ones for the various Bayesian methods shows that the LG estimates are very close to the ML estimates in those situations in which ML provides reasonable standard errors whereas the other priors often provide smaller medians than ML. This is an indication that other

priors tend to smooth the parameter estimates too strongly toward zero. Overall the LG prior seems to performs best of the investigated prior distributions: It has (close to) the lowest RMdSE and (among) the best coverage rate in all situations.

5. Discussion

We investigated the posterior mode estimation of parameters and standard errors of LC models for several types of noninformative prior distributions. By means of a simulation study, we have shown that this Bayesian method provides much more accurate point and interval estimates than the traditional ML and the parametric bootstrap approach when the true parameter is close to the boundary.

Comparing of the proposed noninformative priors showed that the prior that we labeled "LG" yields accurate parameter estimates and standard errors under more conditions than any of the other investigated priors. An additional advantage is that this prior requires only minor modifications of the standard ML estimation algorithms. Jeffreys' prior also performed well under most conditions, but has the disadvantage that its use requires computation of derivatives of the determinant of the Fisher information matrix, which is not straightforward and computationally demanding. The quality of the estimates of the Dirichlet priors with constant hyperparameters and of the normal priors varied considerably across conditions. Our conclusion would thus be that, among the priors studied, the most reasonable one seems to be the LG prior, which is a Dirichlet prior whose hyperparameters preserve the univariate marginal distributions of the Y_k's. This prior is used in the LC analysis program Latent GOLD (Vermunt & Magidson, 2000, 2005).

Interval estimates turned out to be of low quality for all the investigated procedures in various of the investigated conditions as could be seen from the too low coverage probabilities despite of large median widths. More specifically, it seems as if the normal approximation is not appropriate when there are parameters near the boundary, even if (a small amount of) prior information is used to keep the point estimates away from the boundary. However, in the context of the parametric bootstrap, there are more reliable methods for obtaining interval estimates than with the normal approximation (Davison & Hinkley, 1997). Also within a Bayesian framework alternative intervals estimates can be constructed; that is, by reconstruction of the full posterior distribution rather than "just" finding its maximum and relying on the normal approximation at its maximum (Gelman et al., 2003).

References

Bartholomew, D.J., & Knott, M. (1999). *Latent Variable Models and Factor Analysis*. London: Arnold.

Clogg, C.C., & Eliason, S.R. (1987). Some common problems in log-linear analysis. *Sociological Methods and Research*, **16**, 8–44.

Clogg, C.C., Rubin, D.B., Schenker, N., Schultz, B., & Widman, L. (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association*, **86**, 68–78.

Congdon, P. (2001). *Bayesian Statistical Modelling.* Chichester: Wiley.

Davison, A.C., & Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

De Menezes, L.M. (1999). On fitting latent class models for binary data: The estimation of standard errors. *British Journal of Mathematical and Statistical Psychology*, **52**, 149–168.

Efron, B., & Tibshirani, R.J. (1993). *An introduction to the bootstrap.* London: Chapman & Hall.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27-38.

Galindo, F., Vermunt, J.K., & Bergsma, W.P. (2004). Bayesian posterior estimation of logit parameters with small samples. *Sociological methods and Research*, **33**, 88–117.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2003). *Bayesian Data Analysis.* London: Chapman & Hall.

Gilks, W.R., Thomas, A., & Spiegelhalter, D. (1994). A language and program for complex Bayesian modelling. *The Statistician*, **43**, 169–177.

Gill, J. (2002). *Bayesian Methods: A Social and Behavioral Sciences Approach.* New York: Chapman & Hall/CRC.

Goodman, L.A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part-I a modified latent structure approach. *American Journal of Sociology*, **79**, 1179–1259.

Goodman, L.A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, **61**, 215–231.

Haberman, S.J. (1979). *Analysis of Qualitative Data: New Developments.* NY: Academic Press.

Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences.* Newbury Park, CA: Sage Publications.

Jeffreys, H. (1961). *Theory of Probability*, (3rd ed.) Data. Oxford: Oxford University Press.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika,* **64**, 187–212.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data.* London: Chapman & Hall.

Vermunt, J.K. (1997). *Log-linear Models for Event Histories.* Thousand Oakes: Sage Publications.

Vermunt, J.K., & Magidson, J. (2000). *Latent GOLD 2.0 User's Guide.* Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K., & Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced.* Belmont, MA: Statistical Innovations Inc.