Bootstrap methods for measuring classification uncertainty in latent class analysis

José G. Dias[1] and Jeroen K. Vermunt[2]

[1] ISCTE – Higher Institute of Social Sciences and Business Studies, Edifício ISCTE, Av. das Forças Armadas, 1649–026 Lisboa, Portugal
jose.dias@iscte.pt

[2] Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
J.K.Vermunt@uvt.nl

This paper addresses the issue of classification uncertainty in latent class analysis. It proposes a new bootstrap-based approach for quantifying the level of classification uncertainty at both the individual and the aggregate level. The procedure is illustrated by means of two applications.

1 Introduction

Model-based clustering by latent class (LC) models can be formulated as follows. Let $\mathbf{y}$ denote a J-dimensional observation and $D = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$ a sample of size n. Each data point is assumed to be a realization of the random variable $\mathbf{Y}$ coming from an S-component mixture probability density function (p.d.f.)

$$f(\mathbf{y}_i; \boldsymbol{\varphi}) = \sum_{s=1}^{S} \pi_s f_s(\mathbf{y}_i; \boldsymbol{\theta}_s), \quad (1)$$

where π_s are positive mixing proportions that sum to one, $\boldsymbol{\theta}_s$ are the parameters defining the conditional distribution $f_s(\mathbf{y}_i; \boldsymbol{\theta}_s)$ for component s, and $\boldsymbol{\varphi} = \{\pi_1, ..., \pi_{S-1}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_S\}$. Note that $\pi_S = 1 - \sum_{s=1}^{S-1} \pi_s$. The log-likelihood function for a LC model – assuming i.i.d. observations – has the form $\ell(\boldsymbol{\varphi}; \mathbf{y}) = \sum_{i=1}^{n} \log f(\mathbf{y}_i; \boldsymbol{\varphi})$, which is straightforward to maximize (yielding the MLE - maximum likelihood estimator) by the EM algorithm [DLR77].

Our results concern standard LC models; that is, mixtures of independent multinomial distributions [Clo95, VM03]. For nominal data, let Y_j have L_j categories, *i.e.*, $y_{ij} \in \{1, ..., L_j\}$. The standard LC model with S latent classes is obtained by defining the conditional density as $f_s(\mathbf{y}_i; \theta_s) = \prod_{j=1}^{J} \prod_{l=1}^{L_j} \theta_{sjl}^{I(y_{ij}=l)}$, where θ_{sjl} denotes the probability that an observation belonging to latent class s gives response l on variable j, and where $I(y_{ij} = l)$

is an indicator function taking the value 1 if the condition $y_{ij} = l$ is true and 0 otherwise. It should be noted that $\sum_{l=1}^{L_j} \theta_{sjl} = 1$. McHugh [McH56] and Goodman [Goo74] give sufficient conditions for the identifiability of the LC model. All the models used in this paper are identified and are straightforwardly estimated by means of the EM algorithm [Eve84]. For our analysis, we used programs written in MATLAB [MAT02].

From the parameters of the LC model one can derive the posterior probability that an observation belongs to a certain class or cluster conditional on its response pattern. Bayes' theorem gives the estimated *a posteriori* (MLE) probability that observation i was generated by component s:

$$\hat{\alpha}_{is} = \frac{\hat{\pi}_s f_s(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_s)}{\sum_{v=1}^{S} \hat{\pi}_v f_v(\mathbf{y}_i; \hat{\boldsymbol{\theta}}_v)}. \tag{2}$$

It should be noted that these probabilities can not only be used for classification purposes, but also for profiling classes. More specifically, Magidson and Vermunt [MV01] showed that one can investigate the relationship between covariates and classes by comparing (and plotting) the average $\hat{\alpha}_{is}$ across subgroups defined by covariate categories.

Whereas the $\hat{\alpha}_{is}$ values define a soft partitioning/clustering of the data set, that is $\sum_{s=1}^{S} \hat{\alpha}_{is} = 1$ and $\hat{\alpha}_{is} \in [0, 1]$; the next step will usually be the transforming of the resulting soft partition into a hard partition by applying the optimal Bayes rule. Obtaining such a hard partition may be a goal on its own – for example, if the LC model is used for a diagnostic instrument – but it may also serve as input for a subsequent analysis – for example, an analysis aimed at profiling clusters. Let c_i represent the true cluster membership (the missing data) of observation i. Alternatively, the missing data for case i (c_i) can be represented by a set of dummy variables $z_{is} = I(c_i = s)$, where $I(.)$ is an indicator function – $I(A) = 1$, if condition A is true and zero otherwise; *i.e.*, $z_{is} = 1$, if case i belongs to class s and 0 otherwise. Then, the optimal Bayes rule assigning observation i to the class with maximum *a posteriori* probability can be defined as follows:

$$\hat{c}_i = \arg\max_s \hat{\alpha}_{is}, i = 1, ..., n, \tag{3}$$

which is equivalent to

$$\hat{z}_{is} = I\left(\max_{s'} \hat{\alpha}_{is'} = \hat{\alpha}_{is}\right), \ s, s' = 1, ..., S, \ i = 1, ..., n. \tag{4}$$

Therefore, $\hat{z}_{is}$ defines a hard partition, because $\sum_{s=1}^{S} \hat{z}_{is} = 1$ and $\hat{z}_{is} \in \{0, 1\}$.

In this paper, we address the following question: How should we measure the level of uncertainty in the mapping from the [0, 1] soft partition to the {0,1} hard partition obtained by applying the optimal Bayes rule? Note that it is assumed here that the number of labels/classes is known (S is assumed to be fixed); *i.e.*, in determining the level of uncertainty, we do not take into account the uncertainty connected to selection of the model with the correct number of clusters.

The remaining of the paper is organized as follows. In the next section, we describe measures of classification uncertainty at the aggregate and at the individual level. Then, we discuss the proposed bootstrap method for estimating uncertainty and illustrate this procedure by means of two empirical applications. The paper ends with a short discussion.

2 Measures of classification uncertainty

It is natural to assume that the model-based clustering procedure is providing a classification of observations into clusters with a small uncertainty when $\max_s \hat{\alpha}_{is}$ is close to one for most of the observations; conversely, the uncertainty of classification can be assumed to be high when the posterior probabilities are very similar across classes – or when $\max_s \hat{\alpha}_{is}$ is far below 1. At the individual level, classification uncertainty can be measured by the posterior probabilities $\hat{\alpha}_{is}$. Another measure of individual uncertainty in classifying observation i into the class with the largest posterior probability is given by

$$e_i = 1 - \max_s \alpha_{is}. \tag{5}$$

If the observation provides clear information, then $e_i \approx 0$ [DHS01]. The definition of e_i is supported by decision theory as follows. Let s' and s indicate the predicted classification (the decision) and the true state of nature of observation i, respectively. Then, the decision is correct if $s' = s$ and in error otherwise. The loss function of interest is the so-called zero-one loss function, which is defined as follows:

$$L(c_i = s'|c_i = s) = \begin{cases} 0, \; s' = s \\ 1, \; s' \neq s \end{cases} \tag{6}$$

for s', $s = 1,...,S$. The conditional risk associated with this loss function is [DHS01, p. 27]

$$R(c_i = s'|\mathbf{y}_i) = 1 - p(c_i = s|\mathbf{y}_i). \tag{7}$$

Therefore, under zero-one loss, the misclassification risk is minimized if and only if observation i is assigned to the component s for which $p(c_i = s|\mathbf{y}_i)$ is the largest (equation 3) and e_i is the misclassification risk for LC models.

An aggregate measure of classification uncertainty is the entropy. For LC models, the entropy is obtained by

$$EN(\boldsymbol{\alpha}) = -\sum_{i=1}^{n}\sum_{s=1}^{S} \alpha_{is} \log \alpha_{is}. \tag{8}$$

Its normalized version has been used as a model selection criterion indicating the level of separation of components [CS96]. The relative entropy that scales the entropy to the interval [0,1] is defined as [WK00]

$$E = 1 - EN(\boldsymbol{\alpha})/(n \log S). \tag{9}$$

For well-separated latent classes, $E \approx 1$; for ill-separated latent classes, $E \approx 0$. This provides a method for assessing the "fuzzyness" of the partition of the data under the hypothesized model. The ML estimates of e_i, $EN(\boldsymbol{\alpha})$, and E – $\hat{e}_i$, $EN(\hat{\boldsymbol{\alpha}})$, and $\hat{E}$ – can be obtained using the MLE $\hat{\alpha}_{is}$ of α_{is} instead of α_{is} in equations (5), (8), and (9).

3 The bootstrap method

The bootstrap is a computer intensive resampling technique introduced by Efron [Efr79] for determining, among other things, standard errors, biases, and confidence intervals in situations where theoretical statistics are difficult to obtain. The bootstrap technique is easily stated. Suppose we have a random sample D from an unknown probability distribution F and we wish to estimate the parameter $\boldsymbol{\varphi} = t(F)$. Let $S(D, F)$ be a statistic. Whereas for theoretical statistical inference, the underlying sampling distribution of $S(D, F)$ has to be known, the bootstrap method approximates F by some estimate $\hat{F}$ based on D. This gives a sampling distribution based on $S(D^*, \hat{F})$, where the bootstrap sample $D^* = \{\mathbf{y}_1^*, \mathbf{y}_2^*, ..., \mathbf{y}_n^*\}$ is a random sample of size n drawn from $\hat{F}$, and $\hat{\boldsymbol{\varphi}}^* = S(D^*, \hat{F})$ is a bootstrap replication of $\hat{\boldsymbol{\varphi}}$, the ML estimator of $\boldsymbol{\varphi}$. The bootstrap performs a Monte Carlo evaluation of the properties of $\hat{\boldsymbol{\varphi}}$ using repeated sampling, say B times, from $\hat{F}$ to approximate the sampling distribution of $\hat{\boldsymbol{\varphi}}$. The B samples are obtained using the following cycle:

1. Draw a bootstrap sample $D^{(*b)} = \{\mathbf{y}_i^{(*b)}\}$, $i = 1, ..., n$, with $\mathbf{y}_i^{(*b)} \sim \hat{F}$;
2. Estimate $\hat{\boldsymbol{\varphi}}^{(*b)} = S(D^{(*b)}, \hat{F})$.

The quality of the approximation depends on the value of B and on the similarity between $\hat{F}$ and F. For an overview of the bootstrap methodology, we refer to Efron and Tibshirani [ET93].

Here, we propose using the bootstrap technique as a tool for better understanding the aggregate- and individual-level classification uncertainty measures presented in the previous section. We not only obtain bias-corrected point estimates for these measures, but also standard errors and confidence intervals. In other words, we get an indication about the sampling variability of the encountered values for the various measures of classification uncertainty.

Given a fixed S, e_i, $EN(\boldsymbol{\alpha})$, and E can be bootstrapped. For each bootstrap sample $D^{(*b)}$ and parameter estimate $\hat{\boldsymbol{\varphi}}^{(*b)}$, $\hat{\alpha}_{is}^{(*b)}$ are obtained by equation (2). By plugging in $\hat{\alpha}_{is}^{(*b)}$ in equations (5), (8), and (9), we obtain the bootstrap distribution of e_i, $EN(\boldsymbol{\alpha})$, and E, respectively. Graphical and summary descriptive measures of these distributions can be displayed.

4 Bootstrapping LC models

4.1 Number of bootstrap samples

Efron and Tibshirani [ET93, p. 13] suggested using a B value between 50 to 200 when the bootstrap is used for the computation of standard errors. For example, van der Heijden *et al.* [HHD97] and Albanese and Knott [AK94] used 50 and 100 replications, respectively. For confidence intervals, on the other hand, a much larger B value of at least 1000 is required [Efr87]. In all the analyses reported below, we worked with $B = 5000$. This value gives stable and smooth bootstrap results.

4.2 Parametric versus nonparametric bootstrap

There are two types of bootstrap procedures that differ in the way F is approximated. The parametric bootstrap assumes a parametric form for F and estimates the unknown parameters by their sample quantities ($\hat{F}_{par}$). That is, one draws B samples of size n from the parametric estimate of the function F – the function defined by the MLEs of the unknown model parameters. In the nonparametric bootstrap procedure, the approximation of F ($\hat{F}_{nonpar}$) is obtained by its nonparametric maximum likelihood estimate; that is, by the empirical distribution function which puts equal mass $1/n$ at each observation. In that procedure, sampling from $\hat{F}$ means sampling with replacement from the data D.

It has been argued that the parametric bootstrap is better for categorical data whenever the frequency table to be analyzed is sparse [LPP96, Dav97]. In the nonparametric bootstrap, because the sampling is from the empirical distribution, a data pattern that is not observed in the sample has probability zero of being selected into the bootstrap samples and, consequently, $\hat{F}_{nonpar}$ may be too far from the true distribution F. The same problem can, however, also occur in the parametric bootstrap, namely, when certain parameter estimates are on the boundary of the parameter space [HHD97]. In such a case, the resampling will not show any variability within the component concerned, although zero estimated cell frequencies are very rare. On the other hand, Albanese and Knott [AK94] obtained similar results with the parametric and nonparametric bootstrap for latent trait models estimated for binary responses. In our analysis, we compare results from the nonparametric (NP) and parametric (PAR) versions of the bootstrap technique.

4.3 Starting values

For estimating the parameters of the LC model for each resample b ($\hat{\boldsymbol{\varphi}}^{(*b)}$), one needs to use an iterative algorithm. The EM algorithm is an elegant alternative, but its success in converging to the global maximum depends on various factors, such as the quality of the starting values [MK97]. Because the original sample D and the replicated sample $D^{(*b)}$ may not differ too much, McLachlan and Peel [MP00] suggested using the MLE of $\boldsymbol{\varphi}$ from D as a starting value. Dias [Dia05] showed that for the LC model this strategy performs well in comparison with starting the EM algorithm 10 times with random values for the parameters $\boldsymbol{\varphi}$. Therefore, in our analysis, within the bootstrap procedure, the EM algorithm is started from the MLE for sample D.

4.4 Label-switching problem

As is well-known, the likelihood function of mixture models is invariant under permutations of the S latent classes; *i.e.*, any rearrangement of the latent class indices yields the same likelihood value. In bootstrap analysis, as well as in Bayesian analysis by Markov chain Monte Carlo (MCMC) techniques, a permutation of the latent classes may occur, resulting in a distortion of the distribution of interest [Ste97, Ste00, DW04]. Dias [Dia05] showed that for the computation of standard errors and confidence intervals of LC parameters by bootstrap methods, the label-switching problem can have severe impact on the estimation if not handled properly. However, the statistics we use here to measure classification uncertainty at individual (equation 5) and aggregate (equations 8 and 9) levels are invariant to the label switching of the latent classes.

5 Applications

5.1 Stouffer-Toby dataset

This first example illustrates the measurement of classification uncertainty in LC modeling using the classical Stouffer-Toby (ST) data set (Table 4 in [ST51, p. 406]), which has been used by various other authors [Goo74, AK94].[3] It contains the information for 216 respondents with respect to whether they tend towards particularistic or universalistic values when confronted with four different role conflict situations. In our analysis, we assume that $S = 2$. Dias [Dia05] showed with this data set that the label-switching problem for the latent class model can have a severe impact at the parameter level.

Table 1 reports the obtained values for the measures of individual and aggregate uncertainty. The encountered value for the relative entropy $\hat{E}$ (0.72) indicates that the level of separation of components is moderately high, with a 95% nonparametric bootstrap confidence interval of (0.59, 0.87). Note that

[3] Our estimates are slightly different from results reported in [Goo74], because the original dataset of Souffer and Toby is slightly different from the dataset utilized later by [Goo74, p. 216]. However, given the purposes of this application the difference is irrelevant.

the percentile method takes a direct 95% bootstrap confidence interval using the empirical 2.5% and 97.5% quantiles of the bootstrap replicates. From the individual uncertainty indicator $\hat{e}_i$, we conclude that the most problematic patterns to be classified are the $(2, 1, 2, 2)$, $(2, 2, 2, 1)$, and $(2, 2, 1, 2)$ responses. For these three patterns and for $(1, 2, 2, 2)$ we find the largest differences between the bootstrap estimate and MLE (the largest biases) for e_i, as well as upper 95% confidence interval limits near to 0.5, which is the maximum value for e_i with $S = 2$.

Table 1. Classification uncertainty (ST dataset)

Patterns	MLE	Bias		Percentile method	
		NP	PAR	NP	PAR
Individual ($\hat{e}_i$)					
(1,1,1,1)	0.000	0.000	0.000	(0.000, 0.002)	(0.000, 0.001)
(1,1,1,2)	0.001	0.004	0.003	(0.000, 0.039)	(0.000, 0.029)
(1,1,2,1)	0.001	0.003	0.002	(0.000, 0.028)	(0.000, 0.024)
(1,1,2,2)	0.017	0.034	0.033	(0.000, 0.353)	(0.000, 0.345)
(1,2,1,1)	0.001	0.003	0.002	(0.000, 0.026)	(0.000, 0.018)
(1,2,1,2)	0.013	0.032	0.029	(0.000, 0.336)	(0.000, 0.311)
(1,2,2,1)	0.018	0.030	0.027	(0.000, 0.293)	(0.000, 0.287)
(1,2,2,2)	0.287	−0.117	−0.121	(0.000, 0.482)	(0.000, 0.485)
(2,1,1,1)	0.002	0.002	0.001	(0.000, 0.025)	(0.000, 0.017)
(2,1,1,2)	0.031	0.029	0.024	(0.000, 0.345)	(0.000, 0.304)
(2,1,2,1)	0.045	0.024	0.014	(0.000, 0.318)	(0.000, 0.236)
(2,1,2,2)	0.495	−0.235	−0.220	(0.000, 0.489)	(0.000, 0.490)
(2,2,1,1)	0.033	0.018	0.013	(0.000, 0.231)	(0.000, 0.194)
(2,2,1,2)	0.425	−0.170	−0.164	(0.000, 0.490)	(0.000, 0.491)
(2,2,2,1)	0.483	−0.137	−0.129	(0.016, 0.493)	(0.036, 0.495)
(2,2,2,2)	0.041	−0.003	−0.002	(0.008, 0.086)	(0.011, 0.082)
Aggregate					
Entropy ($\hat{EN}$)	42.051	−3.121	−3.666	(18.862, 60.777)	(19.817, 58.363)
Rel. entropy ($\hat{E}$)	0.719	0.021	0.025	(0.594, 0.874)	(0.610, 0.868)

5.2 Political dataset

This second example applies the procedure to a dataset with 1156 observations and 5 binary variables (System responsiveness: 1 - Low, 2 - High; Ideological level: 1 - Nonideologues, 2 - Ideologues; Repression potential: 1 - High, 2 - Low; Protest approval: 1- Low, 2 - High; Convential participation: 1 - Low, 2 - High). These dataset has been used by others [Hag93, VM03]. Based on BIC [Sch78] and AIC3 [Boz93], we picked a three-class solution ($S = 3$).

The relative entropy $\hat{E} = 0.637$ indicates that the level of separation of components is moderate, with a 95% parametric confidence interval of (0.50, 0.80). From the individual uncertainty indicator e_i, we conclude that as far as classification performance is concerned patterns $(1, 2, 1, 1, 1)$, $(2, 1, 1, 2, 1)$, $(1, 1, 2, 1, 2)$, and $(2, 2, 1, 1, 1)$ are the most problematic ones. For these patterns, the confidence interval reaches 0.5, which indicates that there is a uncertainty in the classification (the maximum is 0.67 here). It should be noticed that some patterns such as $(1, 1, 1, 2, 2)$ with small uncertainty ($\hat{e} = 0.19$) has

(0.00, 0.60) as nonparametric 95% CI, which means that a small change in the data set may have a huge impact on the classification of this pattern. We noticed that patterns $(1,2,1,1,1)$, $(1,1,2,1,1)$, and $(1,2,2,1,1)$ present larger biases (differences between the ML estimate and the bootstrap estimates), as well as upper 95% confidence interval limits above 0.5. On the other hand, patterns such as $(1,2,2,2,2)$ have a clear and certain classification. Overall, the results obtained with the parametric and the nonparametric bootstrap are very similar.

Table 2. Classification uncertainty (Political dataset)

Patterns	MLE	Bias		Percentile method	
		NP	PAR	NP	PAR
Individual ($\hat{e}_i$)					
(1,1,1,1,1)	0.098	0.010	0.004	(0.008, 0.332)	(0.010, 0.241)
(1,1,1,1,2)	0.347	−0.056	−0.075	(0.018, 0.521)	(0.008, 0.519)
(1,1,1,2,1)	0.141	−0.007	0.003	(0.036, 0.248)	(0.070, 0.241)
(1,1,1,2,2)	0.192	0.041	0.075	(0.000, 0.601)	(0.000, 0.584)
(1,1,2,1,1)	0.045	0.009	0.003	(0.008, 0.150)	(0.013, 0.103)
(1,1,2,1,2)	0.452	−0.105	−0.104	(0.000, 0.580)	(0.000, 0.587)
(1,1,2,2,1)	0.269	0.003	0.000	(0.033, 0.483)	(0.145, 0.421)
(1,1,2,2,2)	0.096	0.017	0.006	(0.000, 0.391)	(0.000, 0.261)
(1,2,1,1,1)	0.511	−0.171	−0.193	(0.000, 0.587)	(0.000, 0.584)
(1,2,1,1,2)	0.274	−0.058	−0.078	(0.000, 0.457)	(0.000, 0.414)
(1,2,1,2,1)	0.139	0.061	0.080	(0.000, 0.504)	(0.000, 0.523)
(1,2,1,2,2)	0.006	0.056	0.091	(0.000, 0.452)	(0.000, 0.457)
(1,2,2,1,1)	0.358	−0.102	−0.093	(0.000, 0.540)	(0.000, 0.541)
(1,2,2,1,2)	0.021	0.104	0.101	(0.000, 0.480)	(0.000, 0.475)
(1,2,2,2,1)	0.067	0.034	0.023	(0.000, 0.425)	(0.000, 0.306)
(1,2,2,2,2)	0.003	0.018	0.009	(0.000, 0.183)	(0.000, 0.092)
(2,1,1,1,1)	0.321	−0.036	−0.017	(0.032, 0.516)	(0.040, 0.513)
(2,1,1,1,2)	0.231	−0.033	−0.065	(0.011, 0.408)	(0.005, 0.314)
(2,1,1,2,1)	0.449	−0.066	−0.033	(0.139, 0.511)	(0.242, 0.569)
(2,1,1,2,2)	0.046	0.050	0.107	(0.000, 0.474)	(0.000, 0.496)
(2,1,2,1,1)	0.190	0.007	0.006	(0.033, 0.425)	(0.055, 0.400)
(2,1,2,1,2)	0.142	0.091	0.078	(0.000, 0.537)	(0.000, 0.535)
(2,1,2,2,1)	0.353	−0.033	0.005	(0.093, 0.492)	(0.189, 0.495)
(2,1,2,2,2)	0.021	0.025	0.012	(0.000, 0.278)	(0.000, 0.149)
(2,2,1,1,1)	0.406	−0.077	−0.095	(0.000, 0.592)	(0.000, 0.591)
(2,2,1,1,2)	0.299	−0.088	−0.093	(0.000, 0.460)	(0.000, 0.437)
(2,2,1,2,1)	0.031	0.068	0.077	(0.000, 0.467)	(0.000, 0.447)
(2,2,1,2,2)	0.001	0.044	0.082	(0.000, 0.414)	(0.000, 0.439)
(2,2,2,1,1)	0.101	0.105	0.099	(0.000, 0.547)	(0.000, 0.539)
(2,2,2,1,2)	0.005	0.083	0.094	(0.000, 0.463)	(0.000, 0.464)
(2,2,2,2,1)	0.014	0.029	0.012	(0.000, 0.307)	(0.000, 0.106)
(2,2,2,2,2)	0.001	0.020	0.008	(0.000, 0.275)	(0.000, 0.081)
Aggregate					
Entropy ($\hat{EN}$)	461.65	−19.6	−40.32	(241.46, 618.89)	(256.53, 632.34)
Rel. entropy ($\hat{E}$)	0.637	0.015	0.003	(0.513, 0.810)	(0.502, 0.798)

Beyond the summary analysis provided above, it is possible to explore further the richness of the bootstrap results. For example, Figure 1 plots the bootstrap distribution for the relative entropy obtained with the nonparametric and parametric procedures. Given the large number of bootstrap resamples, the plots have a nice smooth shape.

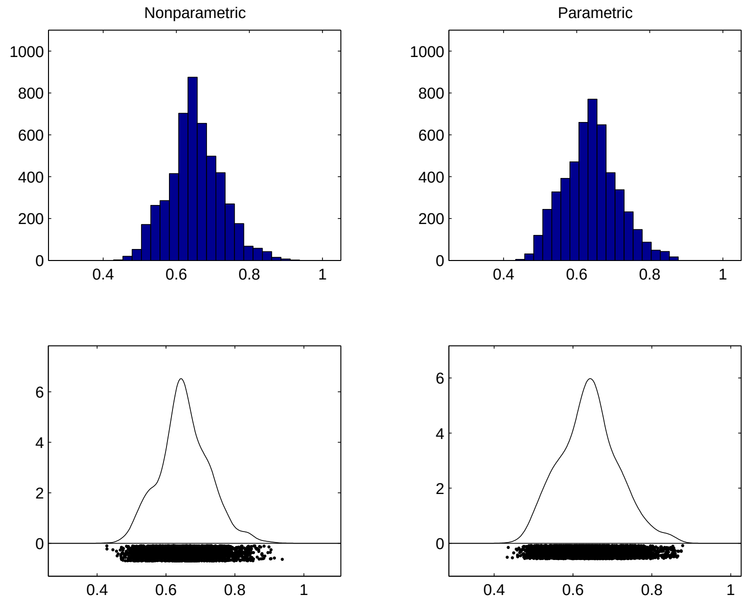

Fig. 1. Nonparametric and parametric distributions of the relative entropy

6 Discussion

We dealt with the measurement of classification uncertainty at the individual and aggregate level in the context of LC modeling. We proposed using parametric and nonparametric bootstrap techniques to determine the bias and sampling fluctuation of measures of classification uncertainty. At the aggregate level, we used the entropy, which is related to the level of separation of the latent classes. As an individual measure of classification uncertainty, we applied $e_i = 1 - \max_s \alpha_{is}$. The exact implementation of bootstrap procedures in LC modeling is not at all straightforward since issues such as label switching and local maxima have to be taken into account. An important advantage of the proposed approach is, however, that it is not affected by the label-switching problem. Results obtained with the parametric and nonparametric bootstrap were discussed for two examples. It should be noticed that this paper introduces a methodology for measuring the uncertainty at individual and global levels. In particular, this approach allows the identification of problematic observations with higher classification uncertainty level. In this case, we know for each observation the risk of the traditional optimum Bayes classification rule (equation 3).

Whereas our research focused on LC models, which are mixtures of conditionally independent multinomial distributions, future research could be aimed at extending our findings and proposals to other types of finite mixture models, such as mixture regression models, or even to more general models with latent discrete variables for longitudinal or multilevel data structures.

References

[AK94] Albanese, M.T., Knott, M.: Bootstrapping latent variable models for binary response. British Journal of Mathematical and Statistical Psychology, **47**, 235 – 246 (1994)

[Boz93] Bozdogan, H.: Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In: Opitz, O., Lausen, B., Klar, R. (eds) Information and Classification, Concepts, Methods and Applications. Springer, Berlin, 40-54 (1993)

[CS96] Celeux, G., Soromenho, G.: An entropy criterion for assessing the number of clusters in a mixture model. Journal of Classification, **13**, 195 – 212 (1996)

[Clo95] Clogg, C.C.: Latent class models. In: Arminger, G., Clogg, C.C., Sobel, M.E. (eds) Handbook of Statistical Modeling for the Social and Behavioral Sciences. Plenum, New York, 311 – 359 (1995)

[DLR77] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B, **39**, 1 – 38 (1977)

[Dia05] Dias, J.G.: Bootstrapping latent class models. In: Weihs, C. and Gaul, W. (eds) Classification - The Ubiquitous Challenge. Springer, Berlin, 121 – 128 (2005)

[DW04] Dias J.G., Wedel, M.: An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. Statistics and Computing, **14**, 323 – 332 (2004)

[DHS01] Duda, R.O., Hart, P.O., Stork, D.G.: Pattern Classification (2nd ed). Wiley, New York (2001)

[Efr79] Efron, B.: Bootstrap methods: another look at the jackknife. The Annals of Statistics, **7**, 1 – 26 (1979)

[Efr87] Efron, B.: Better bootstrap confidence intervals (with discussion). Journal of the American Statistical Association, **82**, 171 – 200 (1987)

[ET93] Efron, B., Tibshirani, R.: An Introduction to the Bootstrap. Chapman & Hall, London (1993).

[Eve84] Everitt, B.S.: A note on parameter estimation for Lazarsfeld's latent class model using the EM algorithm. Multivariate Behavioral Research, **19**, 79 – 89 (1984)

[Goo74] Goodman, L.A.: Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, **61**, 215 – 231 (1974)

[Hag93] Hagenaars, J.A.: Loglinear models with latent variables. Sage, Newbury Park (1993)

[LPP96] Langeheine, R., Pannekoek, J., van de Pol, F.: Bootstrapping goodness-of-fit measures in categorical data analysis. Sociological Methods and Research, **24**, 492 – 516 (1996)

[MV01] Magidson, J., Vermunt, J.K.: Latent class factor and cluster models, biplots and related graphical displays. Sociological Methodology, **31**, 223 – 264 (2001)

[MAT02] MathWorks: MATLAB 6.5. The MathWorks, Natick, MA (2002)
[McH56] McHugh, R.B.: Efficient estimation and local identification in latent class analysis. Psychometrika, **21**, 331 – 347 (1956)
[MK97] McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions. Wiley, New York (1997)
[MP00] McLachlan, G.J., Peel, D.: Finite Mixture Models. Wiley, New York (2000)
[Sch78] Schwarz, G.: Estimating the dimension of a model. Annals of Statistics, **6**, 461 – 464 (1978)
[Ste97] Stephens, M.: Discussion on 'On Bayesian analysis of mixtures with an unknown number of components (with discussion)', Journal of Royal Statistical Society B, **59**, 768 – 769 (1997)
[Ste00] Stephens, M.: Dealing with label switching in mixture models. Journal of the Royal Statistical Society B, **62**, 795 – 809 (2000)
[ST51] Stouffer, S.A., Toby, J.: Role conflict and personality. American Journal of Sociology, **56**, 395 – 406 (1951)
[HHD97] van der Heijden, P., 't Hart, H., Dessens, J.: A parametric bootstrap procedure to perform statistical tests in a LCA of anti-social behaviour. In: Rost, J., Langeheine, R. (eds) Applications of Latent Trait and Latent Class Models in the Social Sciences. Waxmann, New York, 196 – 208 (1997)
[VM03] Vermunt, J.K., Magidson, J.: Latent class models for classification. Computational Statistics & Data Analysis, **41**, 531 – 537 (2003)
[Dav97] von Davier, M.: Bootstrapping goodness-of-fit statistics for sparse categorical data - Results of a Monte Carlo study. Methods of Psychological Research Online, **2**, 29 – 48 (1997)
[WK00] Wedel, M., Kamakura, W.A.: Market Segmentation. Conceptual and Methodological Foundations (2nd ed). International Series in Quantitative Marketing, Kluwer Academic Publishers, Boston (2000)