ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# Likelihood inference in some finite mixture models☆

Xiaohong Chen[a], Maria Ponomareva[b], Elie Tamer[c,*]

[a] Department of Economics, Yale University, United States

[b] Department of Economics, Northern Illinois University, United States

[c] Department of Economics, Northwestern University, United States

ARTICLE INFO

*Article history:*
Available online xxxx

*JEL classification:*
C1

*Keywords:*
Finite mixtures
Parametric bootstrap
Profiled likelihood ratio statistic

ABSTRACT

Parametric mixture models are commonly used in applied work, especially empirical economics, where these models are often employed to learn for example about the proportions of various types in a given population. This paper examines the inference question on the proportions (mixing probability) in a simple mixture model in the presence of nuisance parameters when sample size is large. It is well known that likelihood inference in mixture models is complicated due to (1) lack of point identification, and (2) parameters (for example, mixing probabilities) whose true value may lie on the boundary of the parameter space. These issues cause the profiled likelihood ratio (PLR) statistic to admit asymptotic limits that differ discontinuously depending on how the true density of the data approaches the regions of singularities where there is lack of point identification. This lack of uniformity in the asymptotic distribution suggests that confidence intervals based on pointwise asymptotic approximations might lead to faulty inferences. This paper examines this problem in details in a finite mixture model and provides possible fixes based on the parametric bootstrap. We examine the performance of this parametric bootstrap in Monte Carlo experiments and apply it to data from Beauty Contest experiments. We also examine small sample inferences and projection methods.



## 1. Introduction

This paper studies the question of inference, mainly testing and confidence regions, on the mixing probability in the following finite mixture model with two components where the density of the observed data is:

$$p(\cdot;\theta,\delta)=\delta f_\theta+(1-\delta)f_0 \quad \text{with } f_\theta=f(\cdot,\theta) \quad \text{and} \quad f_0=f(\cdot,\theta_0). \tag{1.1}$$

The mixing probability $\delta$ takes values in the closed interval [0, 1]. We observe a sample of $n$ independent random draws $\{X_i, i = 1,\dots,n\}$ from the density $p(.;\theta,\delta)$, and are interested in inference on $\delta$ in the presence of a nuisance parameter $\theta\in\Theta$, a *compact* subset of $\mathbb{R}^k$. Also, here we assume that $\theta_0$ and the form of $f(.;.)$ are known. The model above is a member of a class of parametric finite mixture models, and in this paper we focus on complications that arise mainly due to the possibility that the true parameters in (1.1) are in the *singularity region* where $\delta*\|\theta-\theta_0\|=0$. The singularity region leads to two problems: lack of identification (if $\theta=\theta_0$, the model has no information about $\delta$ and if $\delta=0$, the model has no information about $\theta$) and parameters lying on the boundary of the parameter space (when $\delta=0$). Those two issues create problems for inference based on the maximum likelihood estimators of $\delta$ and $\theta$, since the maximum likelihood estimators of the parameters in the singularity region are no longer necessarily consistent, and the asymptotic distribution of the likelihood ratio statistic is no longer standard.

Allowing for cases in which the true parameters can lie in this singularity region is key in mixture models as it is related to learning the number of mixture components in a population, which in many cases is the main object of interest in applications. Each point $(\delta,\theta)$ in the singularity region, plotted in Fig. 1, leads to *the same* density for the observed data, i.e., $p(\cdot)=f_0(\cdot)$. We use a profile likelihood ratio statistic to construct confidence region for $\delta$ while treating $\theta$ as a nuisance parameter. *The main objective of this paper is to examine the asymptotic behavior of this profiled likelihood ratio statistic when the true model lies close to the singularity region.*

☆ We thank Don Andrews, Ivan Canay, Xu Cheng, Ulrich Mueller and participants at June 2012 Cowles Summer Conference, Princeton Conference on Non-Standard Econometrics, SEA conference, Harvard/MIT, LSE and OSU seminars for comments. We thank Norm Swanson and referee comments that helped improve the paper. We also thank Rosemarie Nagel for providing us with the data.

* Corresponding author. Tel.: +1 847 491 8218.
*E-mail addresses:* xiaohong.chen@yale.edu (X. Chen), mponomareva@niu.edu (M. Ponomareva), tamer@northwestern.edu (E. Tamer).

http://dx.doi.org/10.1016/j.jeconom.2014.04.010
0304-4076/

Please cite this article in press as: Chen, X., et al., Likelihood inference in some finite mixture models. Journal of Econometrics (2014), http://dx.doi.org/10.1016/j.jeconom.2014.04.010

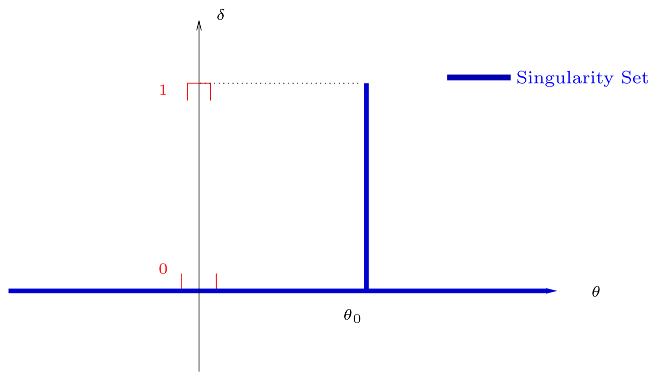


**Fig. 1.** Singularity set in the main model (1.1): when true $\delta$ is equal to zero, then the implied density of the data is $f_0$ no matter what the value of $\theta$ is, i.e., when true $\delta$ is zero, the parameter $\theta$ is not identified at all. Also, when true $\theta$ is equal to $\theta_0$, the implied density of the data is also $f_0$ and $\delta$ is completely not identified.

The pointwise asymptotic distribution of this profiled likelihood ratio statistic (or PLR) has a well known limit distribution even when true $(\delta, \theta)$ belong to the singularity region (see, for example, Liu and Shao (2003)). We complement these results by showing that the limit distribution of this PLR statistic is a discontinuous function of true $\delta$ when this true $\delta$ is in a close neighborhood and drifting towards the singularity region at a given rate. This discontinuity in the asymptotic distribution of the PLR statistic – or lack of uniformity – when the true parameters are sufficiently close to the singularity region can cause misleading inferences (such as undersized test) when these pointwise limit distributions are used with finite samples. We first examine the nature of this discontinuity by deriving the asymptotic distribution of the PLR statistic under drifting – towards the singularity region – sequences of true parameters, and second we propose an approach to inference using a parametric bootstrap procedure. The reason to consider the bootstrap in this setup is that the asymptotic limits involve supremum over complicated stochastic process that are hard to deal with. In addition, the parametric bootstrap seems like a natural approach to use in the setup above. We evaluate these issues using some simulations and an empirical application with experimental data.

There are a few well known issues that one must confront in the above setup. The first is the lack of identification in the region of singularity. Our approach delivers confidence sets that maintain coverage whether or not the model is identified. These sets are constructed by inverting the PLR statistic. In the case where the true model belongs to the singularity region, the confidence sets will be the whole parameter space. Another non-standard issue is that in the singularity region some parameters lie on the boundary of the parameter space, which also creates discontinuities in the limiting distribution of the PLR statistic. Due to these two problems, we take an approach of drifting sequences of true parameters (local to the singularity region approach) to derive the limiting distribution of the PLR statistic. In particular, if these sequences stay away from the singularity set, then the PLR statistic has regular $\chi^2$ limits. As these sequences are allowed to approach the singularity set, the PLR statistic admits varying limits depending on the rate at which the true parameter sequence approaches the singularity and the location of the limit point in the singularity set. Critical sequences are the ones where $\delta$ approaches the singularity region at the rate of square root of the sample size. We propose a parametric bootstrap as the method to construct valid confidence set for $\delta$ (or joint confidence sets for $(\delta, \theta)$). In most cases, the parametric bootstrap mimics the small sample distribution and is shown to be consistent (even in some cases when the limiting mixing probability $\delta$ lies on the boundary of the parameter space). The parametric bootstrap is a particularly attractive approach to inference here since using the asymptotic distribution directly may be computationally difficult. In critical cases, we show how the parametric bootstrap can be adjusted in such a way to guarantee the correct uniform coverage.

So, the empirical takeaway from the paper is that when doing inference on the mixing probabilities in the presence of other parameters, a theoretically attractive approach is to build confidence regions by inverting a (profiled) likelihood ratio statistic. Getting critical values is complicated, but the parametric bootstrap seems to do a reasonable job in approximating the small sample distribution.

Although the model we focus on in this paper is simple, it is a prototypical case that highlights the statistical problems that arise when analyzing more complicated models (with more than 2 components and/or vector $\theta$) and so we consider this model in details and discuss extending our methods to more complicated cases (with vector $\theta$'s and larger number of mixtures) in the Appendix.

### *1.1. Motivation, examples, and literature*

Mixture models are important modeling tools in all areas of applied statistics. See for example McLachlan and Peel (2000). In empirical economics, finite mixtures are used to introduce unobserved heterogeneity. In a nutshell, suppose that an individual or a datum can be one of $K$ types, and each type $k \in \{1, \ldots, K\}$ leads to "behavior" with a density $f_k$. Then, since we do not observe individuals' types, the likelihood of the observed data is a mixture over these densities with the proportion of types being a main object of interest. An important example of this setup from the econometrics literature is Keane and Wolpin (1997).

In addition, mixture models can arise when analyzing some class of games with multiple Nash equilibria. For example, one equilibrium can involve pure strategies and one in mixed strategies.[1] The observed data are proportions of various outcomes where here a given outcome can be observed if (1) it is the pure strategy equilibrium, or (2) if it is on the support of the mixed strategy equilibrium. So, the predicted proportions will be a mixture where the mixing weights are the selection probabilities. See for example Berry and Tamer (2006).

In statistics, there is a large and ongoing literature on inference in finite mixture models using the likelihood ratio statistic. Most results in this literature focus on deriving the limit distribution when the true parameter is fixed. These results can allow for lack of identification. See, for example, Liu and Shao (2003), Dacunha-Castelle and Gassiat (1999), Azaïs et al. (2006), Chernoff and Lander (1995), Chen et al. (2004) and others. Pointwise asymptotic distribution of LR statistic under a fixed null hypothesis for finite mixture models and closely related regime switching models have also been studied in econometrics; see, e.g., Cho and White (2007).

In econometrics, the literature on uniform approximation and confidence intervals is motivated by situations where the pointwise asymptotic distribution of a test statistic has discontinuities in its limit distribution. See, e.g., Mikusheva (2007), Romano and Shaikh (2012), Andrews and Cheng (2005, 2011), Andrews et al. (2011) and references cited therein. Our paper's approach to finite mixtures is motivated by this literature. In particular, Andrews and Cheng (2005, 2011) provide methods for building valid confidence intervals in moment based and likelihood setups in which some parameters can be non-point identified but they assume the true parameters belong to the interior of the parameter space. We follow their approach in that we consider all possible sequences that approach the region of singularity. A key difference

[1] One such game is a $2 \times 2$ entry game in which for some values of the payoffs, there are three equilibria, 2 in pure strategies and one in mixed strategies.