The label switching problem in mixture models

G. Gholami\textsuperscript{a,}\textsuperscript{*}, A. Etemadi\textsuperscript{b}, H. Rasi\textsuperscript{c}
\textsuperscript{a}Department of Mathematics, Urmia University, Urmia, Iran.
\textsuperscript{b}Department of Statistics, University of Mazandaran, Babolsar, Iran.
\textsuperscript{c}Department of Mathematics, Tabriz University, Tabriz, Iran.

\textsuperscript{*}Corresponding author; Department of Mathematics, Urmia University, Urmia, Iran.

\textbf{A R T I C L E I N F O}

Article history
Received 11 June 2014
Accepted 22 July 2014
Available online 25 July 2014

Keywords,
Mixture models
Label switching
Identifiability
Artificial constraints

\textbf{A B S T R A C T}

Mixture models are fascinating objects in that, while based on elementary distributions, they offer a much wider range of modeling possibilities than their components. They also need both highly complex computational challenges and delicate inferential derivations. In Bayesian framework these kinds of models do not admit an analytical solution and one should content him/her by an approximative solution. In this work, we introduce definition of identifiability in statistical model. We focus on definition of identifiability of mixtures of models from Bayesian point of view. This issue is called label-switching problem in Bayesian literatures. We will study a method to identify the mixtures parameter by using MCMC output.

© 2014 Sjournals. All rights reserved.

1. Introduction

Mixture models were introduced by Pearson in 1894. These models are widely used in astronomy, economics and etc. Mixture models have provided appropriate and flexible family for estimating and modeling data.

These models have been estimated by using frequentist and Bayesian methodologies. Nevertheless, the Bayesian viewpoint to estimate these models provides a very powerful framework for analysis of results of scientific experiments. In this approach, most of our interested quantities appear in integral form which does not admit any analytical solutions; therefore they should be approximated by numerical methods. One of the major
problems in estimating the mixture models is that they are invariant under permutation of their component indices. In other word, components of mixture model are exchangeable. This phenomenon results to non-identifying its components. Non-Identifiability of the components in the mixture models is called label switching problem. There are many methods for dealing with this problem in the literature of mixture models [3]. One of the most widely used is to impose artificial constraints on the parameter space. We will study this method in details.

The structure of the paper is as follows. In the next section we introduce identifiability problem in statistical models, and in particular, in mixture models. Section 3 deals with label-switching problem. The common strategy of solving label-switching by imposing artificial constraints on the model parameters will introduce also. We will demonstrate this problem by an example. Then by using this strategy we will solve the problem.

2. Identifiability

Let $\sim f(x|\theta)$, where $\theta$ could be a vector. One of the main objectives of statistical inference is to estimate $\theta$ according to the observed values. Before an attempt to estimate the parameter a question should be answered. Regarding the structure of the parameter, dose it estimable to clarify, let $X_1 \sim N(\mu_1, \sigma^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$. Also consider $Y = X_1 + X_2$, then $\sim N(\mu_1 + \mu_2, 2\sigma^2)$. Based on a sample from,

$$Y_1, ..., Y_n \sim N(\mu_1 + \mu_2, 2\sigma^2),$$

One can estimate $\mu_1 + \mu_2$. But one cannot performed any inference about $\mu_1$ or $\mu_2$. In other word, $\mu_1 + \mu_2$ is estimable but $\mu_1$ and $\mu_2$ are not.

It is one of the basic and main issues in all statistical methods and data analysis. In most application, the structure of the models is so complicated that cannot be distinguished easily whether, due to the structure, interested parameters are estimable or not these usually need accurate studies and in most cases, one should consider some conditions that are called estimability constraints [1].

Generally, estimability of the parameters is known as identifiability of the model as well [5]. If the model is not identifiable then its parameter should not be estimable too. Identifiability will be defined officially as follow.

A family of distributions, $f(x|\theta)$, are identifiable if and only if:

$$f(x|\theta_1) = f(x|\theta_2) \Leftrightarrow \theta_1 = \theta_2 \quad , \forall \theta_1, \theta_2 \in \Theta,$$

Where $f(x|\theta)$ is the likelihood of the data, $x$.

This definition implies that there should be a one-to-one map between parameters and probability distribution. In other words, distinct values of parameters should correspond to distinct probability distributions. Identifiability is a property of the model (not the estimate of the parameter) [3], so solving identifiability problems involves changing the model.

Example 1. We would like to assess with identifiability of the family of normal distributions,

$$\mathcal{P} = \{f(x|\theta) = \frac{1}{\sqrt{2\pi\sigma}} \exp \{ -\frac{1}{2\sigma^2}(x - \mu)^2 \} , \theta = (\mu, \sigma) ; \mu \in \mathbb{R}, \sigma > 0 \}.$$ 

$$f(x|\theta_1) = f(x|\theta_2) \Leftrightarrow \frac{1}{\sqrt{2\pi\sigma_1}} \exp \{ -\frac{1}{2\sigma_1^2}(x - \mu_1)^2 \} = \frac{1}{\sqrt{2\pi\sigma_2}} \exp \{ -\frac{1}{2\sigma_2^2}(x - \mu_2)^2 \} \Leftrightarrow \frac{1}{\sigma_1^2} (x - \mu_1)^2 + \ln \sigma_1^2 = \frac{1}{\sigma_2^2} (x - \mu_2)^2 + \ln \sigma_2^2 \Leftrightarrow x^2 \left( \frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right) - 2x \left( \frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2} \right) + \left( \frac{\mu_1^2}{\sigma_1^4} - \frac{\mu_2^2}{\sigma_2^4} + \ln \sigma_1^2 - \ln \sigma_2^2 \right) = 0$$

To hold this equality for more than 2 values of $x$ we should have:

$$\frac{\mu_1}{\sigma_1^2} = \frac{\mu_2}{\sigma_2^2} \Rightarrow \sigma_1^2 = \sigma_2^2 \Rightarrow \sigma_1 = \pm \sigma_2.$$
Then there is no condition for \( \in \mathbb{R} \), but \( \sigma \) should be positive or negative. Since the scale parameter is restricted to be greater than zero \( \sigma > 0 \) then the model is identifiable.

3. Identifiability of mixture model

A. Finite mixture models

Let \( f(x|\eta) \) be a family of distributions indexed by \( \eta \). A finite mixture of these distributions is a convex combination

\[
f(x|\theta) = \sum_{j=1}^{k} p_j f(x|\eta_j), \quad k > 1, \quad p_j > 0,
\]

\[
\sum_{j=1}^{k} p_j = 1,
\]

Where \( = (k, \eta_1, ..., \eta_k, p_1, ..., p_k) \).

These distributions arise naturally where a population contains two or more sub-populations. Indeed, \( p_j \) is the weight \( j^{th} \) sub-population which is usually unknown. Since these models have more complicated structure than a simple probability model, the identifiability in these models is very important and challenging.

B. Identifiability in finite mixture models

A family of finite mixture distributions indexed by \( \theta \) is identifiable [6] if the equality of the corresponding mixture densities

\[
\sum_{j=1}^{k} p_j f(x|\eta_j) = \sum_{j=1}^{k^*} p_j f(x|\eta_j^*),
\]

Implies that \( k = k^* \) and there exist a permutation \( \nu \) of 1, ..., \( k \) such that \( (p_j, \eta_j) = (p_{\nu(j)}, \eta_{\nu(j)}) \) for \( = 1, ..., k \).

4. The label-switching problem

For any permutation \( \nu \) of 1, 2, ..., \( k \), we define the corresponding permutation of the parameter vector \( \theta \) by

\[
\nu(\theta) = ((\eta_{\nu(1)}, ..., \eta_{\nu(k)}), (p_{\nu(1)}, ..., p_{\nu(k)})).
\]

Label switching problem has its root in the fact that its likelihood [6],

\[
L(\theta|x) = \prod_{i=1}^{n} [p_1 f(x|\eta_1) + \cdots + p_k f(x|\eta_k)],
\]

is the same for all permutations of \( \theta \), where \( x = (x_1, ..., x_n) \) is random sample of \( f(x|\theta) \). In the context of Bayesian inference, posterior will replace the likelihood. The following example clarifies the problem.

Example 2. Assume that a population consists of two sub-populations such as girls and boys. Let the distribution of the population be

\[
f(x|\theta) = p_1 N(x; \mu_1, 1) + p_2 N(x; \mu_2, 1),
\]

Where \( \theta = ((\mu_1, \mu_2), (p_1, p_2)) \) and \( N(x; \mu, \sigma^2) \) stands for a normal density. Furthermore suppose that indices 1 and 2 in parameters stand for the boys and girls respectively.

Based on a sample from this population, suppose that the estimate of \( \theta \) be
Then the distribution is:

\[
 f(x; \theta) = \frac{1}{2} N(x; 2, 1) + \frac{1}{2} N(x; 3, 1) = \frac{1}{2} N(x; 3, 1) + \frac{1}{2} N(x; 2, 1).
\]

Then one cannot identify the components. In other words, he/she cannot correspond the values of estimator for each sub-population.

As we see above the parameters, \( \mu_j \)s cannot be identified marginally. It means \( \mu_j \) can be either 2 or 3. If we have two components, both \((\mu_1, p)\) and \((\mu_2, 1 - p)\) are exchangeable. Exchangeability of components leads to non-identifiability of the model. In mixture model this is called label switching. There exists at least 3 methods to deal with this problem in the literature:

- artificial Identifiability Constraints,
- relabeling Algorithms and
- Label Invariant Loss Functions.

We introduce the artificial identifiability constraints method in following.

A. Artificial identifiability constraints

In this method, one imposes a constraint on the parameters such that only one of \(k!\) permutations satisfies the constraint. For instance, if parameter is \(= (\mu_1, ..., \mu_k)\), constraints on parameter can be considered as \(\mu_1 < \cdots < \mu_k\). Such a limiting structure will remove the symmetry in the likelihood. This constraint could be considered as a prior information about the parameters [1].

If we adapt to appropriate priors for the parameters, then we can perform a posterior analysis. Symmetry in the posterior will result to equality of all marginal posteriors of parameters. It means that all the marginal posteriors are invariant under switching of labels of components. Also it leads to multi modal marginal distributions of parameters. Then removing symmetry from posterior will solve label - switching problem [1]. Note that this is a fully artificial constraint and has no effect on the conclusion.

If some MCMC algorithm were done to approximate the posterior of parameters, these constraints can also be performed after implementation of MCMC algorithm [5]. It means after running the algorithm MCMC required constraint can be performed about every single samples. A major issue in this method is how to choose the constraints. Particularly in multivariate distribution, finding the appropriate constraint is not an easy task [3].

Example 3 Let \(X_1, ..., X_n \sim i.i.d. \frac{1}{2} N(x; \mu_1, 1) + \frac{1}{2} N(x; \mu_2, 1)\) and \(\mu_j \sim N(\delta, \lambda), j = 1, 2\) where \(\delta = 0\) and \(\lambda = 0.1\). By considering \(n = 500\) and \(\mu_1 = 0\) and \(\mu_2 = 1\), Gibbs algorithm will be performed \(10^6\) times and considering its last 10000 repetition, the following output is designed.

Figure the top of left figure 1, will show the direction of Markov chain on the parameter which is moved on both exponential. Above graph on the right, show the direction of chain for each of the means. This figure indicates that \(\mu_1\) and \(\mu_2\), are mixture. Two graphs in the bottom show the related histogram of \(\mu_1\) and \(\mu_2\) and density function. It indicates that marginal distribution of \(\mu_1\) and \(\mu_2\) is two - exponential and they are almost equal to each other and by repeating they will be totally the same.

Figure on the top left figure 2, will show the direction of Markov chain on the parameter which is moved on one exponential. Above graph on the right show the direction of chain for each of the means. Two graphs in the bottom show the related histogram of \(\mu_1\) and \(\mu_2\) and density function. It indicates that marginal distribution of \(\mu_1\) and \(\mu_2\) is one exponential.

5. Conclusion

In this paper we show that label switching problem has accrued in mixture models frequently and illustrated with example and finally introduce existence methods to deal with this problem and explore Artificial Identifiability Constraints method for solving label switching problem with example.
References