An EM Algorithm for a Doubly Smoothed MLE in Normal Mixture Models

Byungtae Seo1,a

1Department of Statistics, Sungkyunkwan University

Abstract

It is well known that the maximum likelihood estimator (MLE) in normal mixture models with unequal variances does not fall in the interior of the parameter space. Recently, a doubly smoothed maximum likelihood estimator (DS-MLE) (Seo and Lindsay, 2010) was proposed as a general alternative to the ordinary maximum likelihood estimator. Although this method gives a natural modification to the ordinary MLE, its computation is cumbersome due to intractable integrations. In this paper, we derive an EM algorithm for the DS-MLE under normal mixture models and propose a fast computational tool using a local quadratic approximation. The accuracy and speed of the proposed method is then presented via some numerical studies.

Keywords: EM algorithm, normal mixture, doubly-smoothed MLE, quadratic approximation.

1. Introduction

Although normal mixture models play a central role in the mixture literature, it is well known that the maximum likelihood approach fails to produce a consistent estimator due to an unbounded likelihood. This type of failure is also common when we use a mixture of location-scale family of distributions. To resolve this problem, a constrained maximum likelihood estimator (MLE) (Hathaway, 1985; Tanaka and Takemura, 2006) uses a constraint on the scale parameters to compactify the parameter space. A penalized MLE proposed by Ciuperca et al. (2003) and Chen et al. (2008) adds some penalty functions to the ordinary likelihood so that the likelihood does not explode when one of the scale parameters goes to zero. The penalized MLE can also be considered as a Bayesian estimator (Fraley and Raftery, 2007) with an inverse Gamma or Wishart prior for scale parameters. The penalized MLE and constrained MLE can be obtained using slight modification of the EM algorithm for normal mixture models (Hathaway, 1986; Ingrassia and Rocci, 2007; Ciuperca et al., 2003; Chen et al., 2008).

As an alternative, Seo and Lindsay (2010) suggested the doubly-smoothed MLE (DS-MLE) which uses smoothing techniques for both the model and data with a fixed kernel and bandwidth. In some sense, the DS-MLE is considered a smoothed MLE because the DS-MLE tends to the ordinary MLE as the bandwidth goes to zero. However, unlike other estimators based on smoothing techniques, the DS-MLE is generally consistent even with a fixed bandwidth. Hence the choice of a bandwidth is less sensitive to the quality of estimators than other smoothing based estimators. If we use the DS-MLE to mixture models, the degeneracy problem is naturally removed, because the smoothed mixture likelihood is bounded for any fixed positive bandwidth.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0014607).

1Assistant Professor, Department of Statistics, Sungkyunkwan University, 53 Myeongryun-dong 3-ga, Chongno-gu, Seoul 110-745, Korea. E-mail: seobt@skku.edu
Although the DS-MLE has many good theoretical properties, its computation requires several numerical integrations. Seo and Lindsay (2010) proposed a simple computational strategy using the Monte Carlo method and applied to the EM algorithm for normal mixtures. However, the EM algorithm under their framework has not been studied thoroughly and their method could be inaccurate in addition, it requires a large amount of computing effort.

In this paper, we derive an EM algorithm to obtain the DS-MLE in normal mixture models and give a fast computing method using a quadratic approximation. Although we focus on the EM algorithm of the DS-MLE for normal mixture models, our work can be easily extended to the DS-MLE for other types of mixture models. This paper is organized as follows: In Section 2, we give a brief review for the unbounded mixture likelihood and the DS-MLE. Then we derive an EM algorithm for the DS-MLE in normal mixtures in Section 3, and propose a fast computing method in Section 4. Some numerical examples to show the accuracy and the speed of the proposed method is given in Section 5. We then give brief concluding remarks in Section 6.

2. Unbounded Likelihood and DS-MLE

In this section, we briefly review the unboundedness of the mixture likelihood and the DS-MLE for normal mixture models. To explain the unboundedness of the mixture likelihood, let us consider the J-component univariate normal mixture model

\[ f(x; \theta) = \sum_{j=1}^{J} p_j N(x; \mu_j, \sigma_j^2), \]  

where \( \theta = (\mu_1, \ldots, \mu_J, \sigma_1^2, \ldots, \sigma_J^2, p_1, \ldots, p_J) \in \mathbb{R}^J \times \mathbb{R}^J \times (0,1)^J \) with the constraint \( \sum_{j=1}^{J} p_j = 1 \), and \( N(x; \mu, \sigma^2) \) stands for the normal density with mean \( \mu \) and variance \( \sigma^2 \). Suppose \( X_1, \ldots, X_n \) is a random sample from (2.1), then it is easy to show that the likelihood of \( \theta \) is unbounded (Kiefer and Wolfowitz, 1956). For a simple example, let us consider the likelihood function for \( J = 2 \)

\[ L(\theta) = \prod_{i=1}^{n} \left[ \frac{p_1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}\right) + \frac{p_2}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(x_i - \mu_2)^2}{2\sigma_2^2}\right) \right]. \]  

In \( L(\theta) \), if we fix \( \mu_1 = x_1 \) (or any observation) and let \( \sigma_1^2 \) go to zero, then (2.2) diverges regardless of a given sample. Indeed, we can observe many infinite spikes at \( \sigma_j^2 = 0, \ j = 1, 2 \). Consequently, the ordinary MLE always occurs on the boundary of the parameter space and this degenerate MLE is neither meaningful nor consistent.

As a tool to regularize this likelihood, Seo and Lindsay (2010) proposed the doubly-smoothed (DS) log likelihood that smooths both the model and data. They constructed two smoothed densities based on a given model and data. The smoothed model density is given by

\[ f_h(t; \theta) = \int f(x; \theta) K_h(x, t) dx, \]  

where \( K_h(x, t) \) is a kernel density with a bandwidth \( h \). The smoothed empirical density is then constructed as

\[ \hat{f}_h(t) = \frac{1}{n} \sum_{i=1}^{n} K_h(x_i, t), \]