Statistics and Probability Letters 117 (2016) 93–99

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A mean-constrained finite mixture of normals model

Junshu Bao [a], Timothy E. Hanson [b,*]

[a] Department of Mathematics and Computer Science, Duquesne University, Pittsburgh, PA 15282, Unites States

[b] Department of Statistics, University of South Carolina, Columbia, SC 29208, United States

ARTICLE INFO

Article history:
Received 18 October 2015
Received in revised form 25 April 2016
Accepted 15 May 2016
Available online 22 May 2016

Keywords:
Linear mixed model
Stick-breaking prior

ABSTRACT

A simple constructive approach to imposing a mean constraint in a finite mixture of multivariate Gaussian densities is proposed. All parameters in the model except for one have closed-form full conditional distributions and are fit through a simple Markov chain Monte Carlo algorithm. For illustration, the mean-constrained finite mixture is implemented in a linear mixed model. Simulations reveal that the mean-constrained model is able to precisely estimate the regression coefficients and mean-constrained random effects distribution simultaneously. An analysis of the Framingham cholesterol data shows that, with relatively simple structure, the model has competitive predictive power with earlier approaches.

1. Introduction

This paper proposes a mean-constrained finite mixture of multivariate Gaussian densities. The model proposed is constructive, i.e. the mean-constraint is built into the model, versus approaches which impose mean constraints after fitting (e.g. Li et al., 2010; Jara et al., 2009; Yang and Dunson, 2010) or approaches that impose the constraints during nonparametric estimation (e.g. Hall and Presnell, 1999; Eloyan and Ghosh, 2011; Laurence et al., 2014). Mean constraints are necessary in many inferential situations, including generalized linear mixed models (e.g. Jara et al., 2009), structural equation models (e.g. Yang and Dunson, 2010), and in the modeling of extreme value distributions (e.g. Boldi and Davison, 2007), to name a few.

Linear mixed models (LMM) are widely applied in the analysis of longitudinal and other types of repeated measures data. An open question in LMM is how to best model the random effects. Classical approaches assume that the random effects follow a mean-zero Gaussian distribution (Laird and Ware, 1982). However, it has been found that this assumption is often violated, affecting prediction for subjects not in the data set (Claeskens and Hart, 2009). To relax this assumption, novel approaches have been proposed to model the random effects more flexibly including the Dirichlet process prior (Kleinman and Ibrahim, 1998), Hermite expansions (Zhang and Davidian, 2001), penalized Gaussians over a grid (Ghidey et al., 2004), and mixtures of multivariate Polya trees (Jara et al., 2009). For identifiability, ideally a random effects distribution should be centered at zero, enhancing interpretation of fixed effects in terms of population averages. This is a simple constraint for parametric approaches but becomes challenging for semi- and non-parametric approaches. In Section 2, we introduce the mean-constrained finite mixture (MCFM) for multivariate density estimation and, for illustration, show how the MCFM can be used in mixed models. In Section 3, we evaluate the performance of the proposed MCFM model using a simple simulation study. Section 4 describes one application of the model on the Framingham cholesterol data and Section 5 concludes the paper.

* Corresponding author.
E-mail addresses: baoj@duq.edu (J. Bao), hansont@stat.sc.edu (T.E. Hanson).

http://dx.doi.org/10.1016/j.spl.2016.05.009
0167-7152/

2. Model

Consider the continuous mixture model density

$$p(y) = \int_a^b p(y|\mu(t), \sigma(t))w(t)dt = E\{p(y|\mu(T), \sigma(T))\},$$

where $T \sim w(t)$ is a known density with support (a, b) and $E(Y|\mu, \sigma) = \int_{-\infty}^{\infty} yp(y|\mu, \sigma)dy = \mu$. Let $Z(t)$ be a random, differentiable process over $t \in (a, b)$ such that $Z(a) = Z(b)$ and let $\mu(t)$ be determined from $Z'(t) = \mu(t)w(t)$. Then

$$\begin{aligned} E(Y|Z) &= E_{T|Z}E_{Y|T,Z}(Y|T, Z) \\ &= E_{T|Z}\left[\frac{Z'(T)}{w(T)}\right] = \int_a^b \tfrac{Z'(t)}{w(t)} w(t)dt \\ &= Z(b) - Z(a) = 0, \end{aligned}$$

for all Z. This model, which forces Y to have mean-zero, is adapted to the finite mixture model case where instead $t \in \{0, 1, \dots, J\}$, integration is replaced by summing (or premultiplying by a vector of ones), and differentiation is replaced by differencing (or premultiplying by any matrix orthogonal to a vector of ones). Since the idea immediately generalizes to multiple dimensions, we develop the approach for multivariate data from the outset.

Consider p-dimensional data arising from a finite mixture model with J components. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_J)$, $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_J)$, and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)'$ be component means, covariance matrices, and weights respectively. Assume for now that $\boldsymbol{\pi}$ is given. Then

$$\mathbf{y}_1, \dots, \mathbf{y}_n|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \overset{i.i.d.}{\sim} G = \sum_{j=1}^{J} \pi_j N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j).$$

The mean-zero constraint

$$E(\mathbf{y}) = \sum_{j=1}^{J} \pi_j \boldsymbol{\mu}_j = \mathbf{0}_p, \tag{1}$$

which can be written as

$$[\boldsymbol{\pi}' \otimes \mathbf{I}_p]\text{Vec}(\boldsymbol{\mu}) = \mathbf{0}_p,$$

forces $\text{Vec}(\boldsymbol{\mu}) = (\boldsymbol{\mu}_1', \dots, \boldsymbol{\mu}_J')'$ to live in a $(J-1) \times p$-dimensional hyperplane in $\mathbb{R}^{J\times p}$. Let $\boldsymbol{\theta}_j = \pi_j\boldsymbol{\mu}_j$ and $\boldsymbol{\theta} = (\pi_1\boldsymbol{\mu}_1, \dots, \pi_J\boldsymbol{\mu}_J)'$. The constraint (1) is satisfied when $\mathbf{1}_J'\boldsymbol{\theta} = \mathbf{0}$ where $\mathbf{1}_J$ is a vector of J ones. Define

$$\boldsymbol{\theta} = \mathbf{M}_{J\times(J-1)}\mathbf{Z}_{(J-1)\times p} = \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_J \end{bmatrix} \begin{bmatrix} \mathbf{z}_1' \\ \vdots \\ \mathbf{z}_{J-1}' \end{bmatrix} \tag{2}$$

where

$$\mathbf{z}_1, \dots, \mathbf{z}_{J-1}|\boldsymbol{\Omega} \overset{ind.}{\sim} N_p(\mathbf{0}, \boldsymbol{\Omega})$$

and $\mathbf{m}_j$'s are $1 \times (J-1)$ vectors. Note that from Eq. (2), we have $\boldsymbol{\theta}_j = (\mathbf{m}_j\mathbf{Z})'$. The columns of matrix $\mathbf{M}$, of dimension $J \times (J-1)$, span the space orthogonal to the vector of all ones $\mathbf{1}_J$, $\mathcal{C}(\mathbf{1}_J)$. That is, $\mathcal{C}(\mathbf{M}) = \mathcal{C}(\mathbf{1}_J)^{\perp}$, e.g. $m_{jj} = -1$, $m_{j+1,j} = 1$ for $j = 1, \dots, J-1$, and zeros elsewhere. Note then $\mathbf{1}_J'\boldsymbol{\theta} = \mathbf{0}_{1\times p}$ a.s., i.e. $E(\mathbf{y}) = \mathbf{0}$. To make this discrete mixture completely analogous to the continuous mixture that motivated it, let $\mathbf{M}$ be the difference matrix above but with two additional columns: on the very left a column of zeros except the first element is 1, and on the right a column of zeros except the last element is -1; then augment the $\{\mathbf{z}_j\}_{j=1}^{J-1}$ with any $\mathbf{z}_0 = \mathbf{z}_J$.

Let $s_i = j$ if $\mathbf{y}_i$ comes from component j and define $\mathbf{s} = (s_1, \dots, s_n)'$. The data model conditional on the $\mathbf{s} = (s_1, \dots, s_n)'$ is

$$\mathbf{y}_i|\mathbf{Z}, \boldsymbol{\Sigma}, \mathbf{s} \overset{ind.}{\sim} N_p\left(\pi_{s_i}^{-1}(\mathbf{m}_{s_i}\mathbf{Z})', \boldsymbol{\Sigma}_{s_i}\right), \qquad P(s_i = j) = \pi_j.$$

The full conditional for $\mathbf{Z}$ is proportional to

$$p(\mathbf{Z}|\text{else}) \propto \prod_{j=1}^{J-1} \exp\left\{-0.5\mathbf{z}_j'\boldsymbol{\Omega}^{-1}\mathbf{z}_j\right\} \prod_{i=1}^{n} \exp\left\{-0.5\left[\mathbf{y}_i - \pi_{s_i}^{-1}(\mathbf{m}_{s_i}\mathbf{Z})'\right]' \boldsymbol{\Sigma}_{s_i}^{-1}\left[\mathbf{y}_i - \pi_{s_i}^{-1}(\mathbf{m}_{s_i}\mathbf{Z})'\right]\right\}.$$