Geometry of Local Mixture Models

Vahed Maroufy
joint work with
Marriott, P and Small, C. G

Department of Statistics and Actuarial Science University of Waterloo

April 6th, 2011
Warwick University

Outline

Definition

- A continuous mixture models

$$f_M(x) = \int f(x;\theta)\, dQ(\theta)$$

  is called local mixture model if Q is a mixing distribution with *"small" variation*.

- For instance, $f(x;\theta) = \phi(x;\theta,1)$ and $\theta \sim N(\theta_0, \epsilon)$

Measurement Error Model

Consider a simple linear regression of Y against ρ

$$Y = \alpha + \beta\rho + \epsilon, \qquad X = \rho + \eta$$

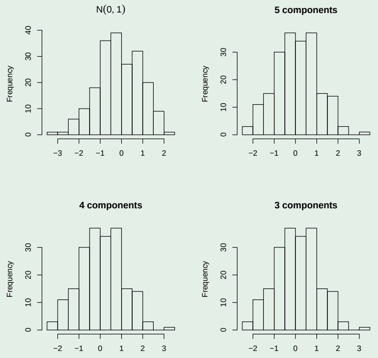
1- $\epsilon \sim \mathcal{N}(0, \sigma^2)$
2- $\eta \sim Q$ is independent of ρ and ϵ

$$\begin{aligned} f_M(x,y) &= \int f(y|\rho,\eta,x,\sigma^2)\, f_1(x|\eta)\, dQ(\eta) \\ &= \int \tilde{f}(x,y|\eta)\, dQ(\eta) \end{aligned}$$

Small Mixing

How hard is to model small mixing?

- $\mathcal{N}(0,1)$
- 5 components, $\mu = 0$, $\sigma_1^2 = 1.116$
- 4 components, $\mu = 0$, $\sigma_2^2 = 1.106$
- 3 components, $\mu = 0$, $\sigma_3^2 = 1.150$



Laplace expansion

Under the assumption required for **Laplace expansion**

(Marriott (2002) and Anaya-Izquierdo & Marriott (2007)),

$$f_M(x; Q) = f(x; \theta_0) + \sum_{j=1}^{k} \lambda_j f^{(j)}(x; \theta_0) + R(x; \theta_0, \epsilon)$$

where

- $f^{(j)} = \frac{\partial^j f}{\partial \theta^j}$ and Q is postulated to be a **dispersion model** with shape and dispersion parameters (θ_0, ϵ).
- $\lambda_j := \lambda_j(\theta_0, \epsilon)$, and $R = O(\epsilon^{\lfloor \frac{k+1}{2} \rfloor})$.
- if $f(x; \theta)$ and $f^{(j)}(x; \theta)$ are bounded, the approximation **is uniform in** x.

Recap: Affine Space and Convex Hull

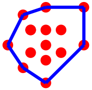
- Space $\langle \mathcal{X}, \mathcal{V}, + \rangle$ is called **affine space** if

$$\mathcal{X} = \left\{ f(x) | f \in L^2(\nu), \ \int f(x)\, d\nu = 1 \right\}$$

and

$$\mathcal{V} = \left\{ f(x) | f \in L^2(\nu), \ \int f(x)\, d\nu = 0 \right\}$$

- **Convex hull** of a set of points is the smallest convex set containing all the points.



Affine Property and Identifiability

- Family of local mixture models is an **affine space** (under some regularity conditions)

$$g_1(x;\theta_0,\lambda) = f(x;\theta_0) + \sum_{j=1}^{k} \lambda_j f^{(j)}(x;\theta_0), \qquad \lambda \in \Lambda(\theta_0) \tag{1}$$

(locally non-identifiable)

$$g_2(x;\theta_0,\lambda) = f(x;\theta_0) + \sum_{j=2}^{k} \lambda_j f^{(j)}(x;\theta_0), \qquad \lambda \in \Lambda(\theta_0) \tag{2}$$

(identifiable)

- the boundary of $\Lambda(\theta_0)$, called **hard boundary**, guaranties positivity.
- this models may not behave similar to **genuine mixture** models.

True LMM

local mixture $g(x; \theta_0, \lambda)$, of order k, is called "**true**" local mixture models if it can locally mimic the behavior of an actual mixture model.

That is; iff there is a Q such that, $g(x; \theta_0, \lambda)$ and

$$\int f(x; \theta) \, dQ(\theta)$$

share the same k first moments.

Anaya and Marriott (2007)

Let $g(x;\mu,\lambda)$, be an order k LMM of natural exponential family with $\mu = E(X)$

- g is **identifiable** in all parameters and the parametrization (μ,λ) is **orthogonal** at $\lambda = 0$
- g is *"true"* LMM if $(\mu_g^1, \cdots, \mu_g^k) \in Co\left(\left\{(\mu_f^1, \cdots, \mu_f^k), \mu \in M\right\}\right)$
- The log likelihood function of g is a concave function of λ at a fixed μ_0
- $\Lambda(\mu)$, the hard boundary, is convex or empty.

Example (frailty models)

- In survival analyses for some cancer clinical trials **mixture cure models** are used rather than traditional survival models.

  $$S_{pop}(t) = (1-\pi) + \pi\, S_0(t),$$ Berkson and Gage (1952)

  π is an **uncured rate** and $S_0(t)$ is a **survival function of the latency distribution**.

- This model with a frailty term in latency components

  $$S_{pop}(t) = (1-\pi) + \pi\, L_\nu(H_0(t)),$$ Price and Manatunga (2001)

  where, $L_\nu(s) = \int e^{s\nu}\, dF(\nu)$, $V \sim F(\nu)$ is frailty, and $H_0(t)$ is the **baseline cumulative hazard function**

LMM2 and LMM4

Suppose $f(x;\mu)$ is the density of $N(\mu,1)$ then

$$g_3(x;\mu,\lambda_2) = f(x;\mu) + \lambda_2 f^{(2)}(x;\mu), \qquad 0 \leq \lambda_2 \leq 1 \tag{3}$$

$$g_4(x;\mu,\lambda_2) = f(x;\mu) + \lambda_2 f^{(2)}(x;\mu) + \lambda_3 f^{(3)}(x;\mu) + \lambda_4 f^{(4)}(x;\mu) \tag{4}$$

where the hard boundary conditions for g_4 are equivalent with the positivity conditions of a **quartic** polynomial.

The central moments of LMM4 and λ are related through

$$\begin{cases} \mu_{g_4}^{(2)} = 1 + 2\lambda_2 \\ \mu_{g_4}^{(3)} = 6\lambda_3 \\ \mu_{g_4}^{(4)} = 3 + 12\lambda_2 + 24\lambda_4 \end{cases} \tag{5}$$

MLE for μ of LMM4

$$g_4(x;\mu,\lambda_2) = \phi(x;\mu,1) + \lambda_2\,\phi^{(2)}(x;\mu,1) + \lambda_3\,\phi^{(3)}(x;\mu,1) + \lambda_4\,\phi^{(4)}(x;\mu,1)$$

$$\begin{cases} a = \lambda_4, \quad b = \frac{\lambda_3}{4} \\ d = -\frac{3\lambda_3}{4}, \quad c = \frac{\lambda_2}{6} - \lambda_4 \\ e = 3\lambda_4 - \lambda_2 + 1 \end{cases}$$

$$\begin{cases} H = ac - b^2 \\ I = ae - 4bd + 3c^2 \\ J = ace + 2bcd - ad^2 - c^3 - eb^2 \end{cases}$$

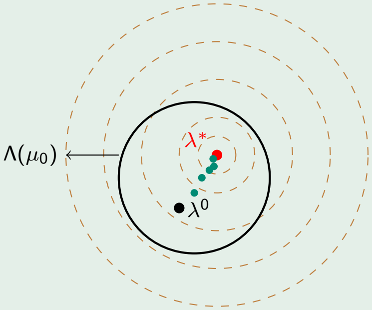
$$\begin{cases} I > 0 \\ I\sqrt{I} + 3\sqrt{3}J > 0 \\ H + a\sqrt{\frac{I}{12}} > 0 \\ e > 0, \quad a > 0 \end{cases} \tag{6}$$

(Barnard, S. and Child, J. M. (1936))

For LMM4 of normal family

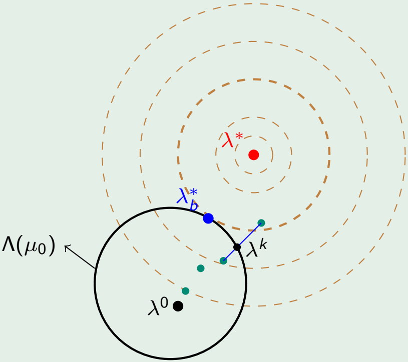
The goal is to find some constrained optimization algorithms which exploits the **concavity** of log likelihood function, as a function of $\lambda = (\lambda_2, \lambda_3, \lambda_4)$, and **convexity** of $\Lambda(\mu_0)$.

- $\lambda^* \in \Lambda(\mu)$
- An ordinary gradient algorithm (Newton-Raphson) is applied

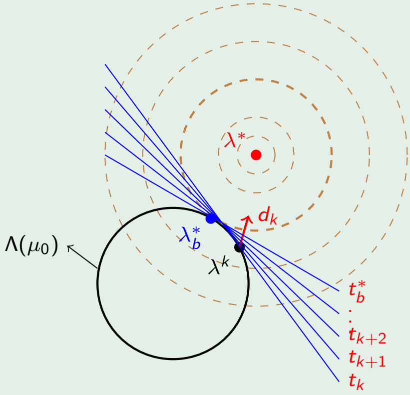


- $\lambda^{(k+1)} = \lambda^{(k)} + H_k^{-1}\, d_k$

- $\lambda^* \notin \Lambda(\mu)$
- λ_b^* is maximum on the boundary



- t_j's are the planes constructing the hard boundaries
- d_j is projected on t_j, which is $P_{t_j} d_j$
- $\lambda^{(k+1)} = \lambda^{(k)} + H_k^{-1} \, P_t \, d_k$
- $||P_{t_b^*} d_b^*|| = 0$ is the optimality condition



Summary

- Introduced local mixture models
- Remarked the nice geometry and fruitful properties
- Taking advantage of the remarkable properties, a gradient based algorithm was introduced for constrained optimization of log likelihood function of LMM4.

Thank You!