Banff Workshop

May, 2013

Nonparametric mixtures based on Skew-Normal Distributions An application to density estimation

Rosangela H. Loschi (UFMG)

joint work with

Caroline C. Vieira (UFES)

Denise Duarte (UFMG)

Financial support:CNPq, CAPES, FAPEMIG

Motivation and Goals

To build more flexible models for density estimation.

- We consider Bayesian nonparametric models
  - $*$ Dirichlet Process mixture (DPM) of continuous distributions (Ferguson, 1973).
- We extend Escobar and West(1995)'s model
  - $*$ By mixing more flexible parametric distributions;
  - $*$ By using a more flexible mixing measure.
- We consider the skew-normal family of distributions (Azzalini, 1985).
- We analyze the eruption duration time of Old Faithful Geyser data set.

Dirichlet Process Mixture of Distributions

A random mixture of distribution is defined as

$$F_G(y) = \int f(y|\theta) dG(\theta)$$

– $f(y|\theta)$, $\theta \in \Theta$, is the sample distribution

– $\theta \sim G$ and G is a random measure on Θ.

The DPM model is represented hierarchically as

$$Y_i|\theta_i \overset{ind}{\sim} f(\cdot|\theta_i), \quad \theta_i|G \overset{iid}{\sim} G, \quad i = 1, \ldots, n, \quad G|\alpha, G_0 \sim DP(\alpha, G_0),$$

– $G_0 = E(G)$ is the center measure over Θ

– $\alpha \in \mathbb{R}^+$ controls the concentration of the prior for G about G_0.

– α is the precision parameter.

- Other approaches: Polya tree (Lavine, 1994), Bernestein Polynomials (Petrone, 1999), for discrete distributions (Canale and Dunson, 2011).
- Discrete mixtures of skewed distributions (Liu et al, 2007, Cabral et al (2008).
- References on Bayesian Non-parametric: Müller and Quintana (2004), Walker (2005), Dey, Müller and Sinha (1998).

Dirichlet Process Mixture of Distributions

The idea behind DPM of distributions is that of clustering the θs.

- If $G \sim DP(\alpha, G_0)$, there is a positive probability of identical θ_i's which is due to the discreteness of G.
- Thus, for $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ we have

$$\theta_i \mid \boldsymbol{\theta}_{(-i)} \sim \frac{\alpha}{\alpha + n - 1} G_0(\theta_i) + \frac{1}{\alpha + n - 1} \sum_{j=1, j \neq i}^{n} \delta_{\theta_j}(\theta_i),$$

  where $\delta_{\theta_j}(\theta_i)$ is a unit point mass at $\theta_i = \theta_j$.

For details see Ferguson(1973) and Antoniak(1974).

Escobar & West model for density estimation (DPMN-N)

They consider

- a DPM of normal distribution;
- A conjugate prior distribution as the center measure G_0,

that is

$$Y_j|\mu_j, \sigma_j^2 \overset{ind}{\sim} N(\mu_j, \sigma_j^2), \quad (\mu_j, \sigma_j^2) \overset{iid}{\sim} G, \; j = 1, \dots, n, \quad G|\alpha, G_0 \sim DP(\alpha, G_0),$$

and the center measure G_0 for $\boldsymbol{\theta}_j = (\mu_j, \sigma_j^2)$ is

$$\mu_j|\sigma_j^2 \overset{ind}{\sim} N(m, \tau\sigma_j^2), \quad \sigma_j^2 \overset{iid}{\sim} IG(s/2, S/2),$$

where m is a real number, $\tau > 0$ and s and S are non negative numbers.

DPM of Skew-Normal (DPMSN)

We consider

- a DPM of skew-normal distribution (Azzalini, 1985), that is

$$Y_j|\mu_j, \sigma_j^2, \lambda_j \overset{ind}{\sim} SN(\mu_j, \sigma_j^2, \lambda_j),$$

$$(\mu_j, \sigma_j^2, \lambda_j) \overset{iid}{\sim} G,\ j = 1, \ldots, n, \quad G|\alpha, G_0 \sim DP(\alpha, G_0),$$

- A conjugate prior distribution as the center measure G_0, that is, the center measure G_0 for $\boldsymbol{\theta}_j = (\mu_j, \sigma_j^2, \lambda_j)$ is

$$\mu_j|\sigma_j^2 \overset{ind}{\sim} N(m, \tau\sigma_j^2), \quad \sigma_j^2 \overset{iid}{\sim} IG(s/2, S/2), \quad \lambda_j \overset{iid}{\sim} N(\varepsilon, \zeta^2),$$

where m and ε are real numbers, $\tau > 0$ and s, S and ζ are non negative numbers.

The full conditional distributions

Under these assumptions, it follows that

(i) the fcd of μ_i^*, given σ_i^{*2}, λ_i^* and $\mathbf{y}_{(i)}$, is

$$f(\mu_i^* | \sigma_i^{*2}, \lambda_i^*, \mathbf{y}_{(i)}) \propto \phi(\mu_i^*; M^*, V^*) \Phi_{n_i}(\mu_i^* \boldsymbol{\lambda}_i^* + \mathbf{y}_i^*); \qquad (1)$$

(ii) the fcd of σ_i^{*2}, given μ_i^*, λ_i^* and $\mathbf{y}_{(i)}$, is

$$f(\sigma_i^{*2} | \mu_i^*, \lambda_i^*, \mathbf{y}_{(i)}) \propto \Phi_{n_i}(\lambda_i^* \mathbf{Z}_i) \, IG\left(\sigma_i^{*2}; \frac{n_i + s + 1}{2}, \right. \qquad (2)$$

$$\left. \frac{1}{2}\left[S + \frac{1}{\tau}(\mu_i^* - m)^2 + \sum_{k=1}^{n_i}(y_k - \overline{y}_{(i)})^2 + n_i(\mu_i^* - \overline{y}_{(i)})^2\right]\right);$$

(iii) the fcd of λ_i^*, given μ_i^*, σ_i^{*2} and $\mathbf{y}_{(i)}$, is

$$f(\lambda_i^* | \mu_i^*, \sigma_i^{*2}, \mathbf{y}_{(i)}) \propto \phi(\lambda_i^*; \varepsilon, \zeta^2) \, \Phi_{n_i}(\lambda_i^* \mathbf{Z}_i), \qquad (3)$$

where

$$
\begin{aligned}
M^* &= (m + \tau n_i \bar{y}_{(i)})(1 + \tau n_i)^{-1}, \\
V^* &= \tau \sigma_i^{*2}(1 + \tau n_i)^{-1}, \\
\boldsymbol{\lambda}_i^* &= \frac{-\lambda_i^*}{\sigma_i^*} \mathbf{1}_{n_i}, \\
\mathbf{y}_i^* &= \frac{\lambda_i^*}{\sigma_i^*} \mathbf{y}_{(i)}, \\
\mathbf{Z}_i &= (\mathbf{y}_{(i)} - \mu_i^* \mathbf{1}_{n_i}) \frac{1}{\sigma_i^*}.
\end{aligned}
$$

Stochastic Representations

Let W and $\mathbf{U}_i$ be two real random variables.

(i) Assume that, given $\mathbf{y}_{(i)}$, σ_i^{*2} and λ_i^*, W and $\mathbf{U}_i$ are independent with $W \sim N(M^*, V^*(1 + V^* \boldsymbol{\lambda}^* \boldsymbol{\lambda}_i^{*T})^{-1})$ and $\mathbf{U}_i \sim LTN_{n_i}(\mathbf{0}, \boldsymbol{I}_{n_i} + V^* \boldsymbol{\lambda}_i^{*T} \boldsymbol{\lambda}_i^*; -M^* \boldsymbol{\lambda}_i^* - \mathbf{y}_i^*)$. Then,

$$\mu_i^* | \sigma_i^{*2}, \lambda_i^*, \mathbf{y}_{(i)} \stackrel{d}{=} W + [V^* \boldsymbol{\lambda}_i^* \mathbf{U}_i^T][1 + V^* \boldsymbol{\lambda}_i^* \boldsymbol{\lambda}_i^{*T}]^{-1}.$$

(ii) Assume that, given $\mathbf{y}_{(i)}$, σ_i^{*2} and μ_i^*, W and $\mathbf{U}_i$ are independent with $W \sim N(\varepsilon, \zeta^2/(1 + \zeta^2 \mathbf{Z}_i \mathbf{Z}_i^T))$ and $\mathbf{U}_i \sim LTN_{n_i}(\mathbf{0}, \boldsymbol{I}_{n_i} + \zeta^2 \mathbf{Z}_i^T \mathbf{Z}_i; -\varepsilon \mathbf{Z}_i)$. Then,

$$\lambda_i^* | \mu_i^*, \sigma_i^{*2}, \mathbf{y}_{(i)} \stackrel{d}{=} W + [\zeta^2 \mathbf{Z}_i \mathbf{U}_i^T][1 + \zeta^2 \mathbf{Z}_i \mathbf{Z}_i^T]^{-1}$$

.

- This result extends some previous ones by Arellano-Valle et al.(2012).

DPM of normal with Skewed G_0 (DPMN-SN)

We consider

- a DPM of normal distribution that is

$$Y_j|\mu_j,\sigma_j^2 \overset{ind}{\sim} N(\mu_j,\sigma_j^2), \quad (\mu_j,\sigma_j^2) \overset{iid}{\sim} G,\ j=1,\ldots,n, \quad G|\alpha,G_0 \sim DP(\alpha,G_0),$$

- The center measure G_0 for $\boldsymbol{\theta}_j = (\mu_j,\sigma_j^2)$ is

$$\mu_j|\sigma_j^2,\lambda \overset{ind}{\sim} SN(m,\tau\sigma_j^2,\lambda), \quad \sigma_j^2 \overset{iid}{\sim} IG(s/2,S/2), \quad \lambda \sim N(\varepsilon,\zeta^2),$$

where m and ε are real numbers, $\tau > 0$ and s, S and ζ are non negative numbers.

- If we assume $\lambda \sim N(0,\zeta^2) \Rightarrow \mu_j|\sigma_j^2 \overset{ind}{\sim} N(m,\tau\sigma_j^2)$
- We have Escobar and West's model (DPMN-N).

The full conditional distributions

Under this assumptions, for all $i = 1, \ldots, k$, it follows that

(i) the fcd of μ_i^*, given σ_i^{*2}, λ and $\mathbf{y}_{(i)}$, is

$$f(\mu_i^* | \sigma_i^{*2}, \lambda, \mathbf{y}_{(i)}) \propto \phi(\mu_i^*; M_i^*, V_i^*)\, \Phi(\lambda_i^*(\mu_i^* - m)); \qquad (4)$$

(ii) the fcd of σ_i^{*2}, given μ_i^*, λ and $\mathbf{y}_{(i)}$, is

$$\begin{aligned} f(\sigma_i^{*2} | \mu_i^*, \lambda, \mathbf{y}_{(i)}) \quad \propto \quad & \Phi\left(\lambda Z_i^*\right) IG\left(\sigma_i^{*2}; \frac{n_i + s + 1}{2}, \right. \qquad (5) \\ & \left. \frac{1}{2}\left[S + \frac{(\mu_i^* - m)^2}{\tau} + \sum_{k=1}^{n_i} (y_k - \overline{y}_{(i)})^2 + n_i(\mu_i^* - \overline{y}_{(i)})^2\right]\right); \end{aligned}$$

(iii) the fcd of λ, given μ_i^*, σ_i^{*2} and $\mathbf{y}_{(i)}$, is

$$f(\lambda|\mu_i^*, \sigma_i^{*2}, \mathbf{y}_{(i)}) \propto \phi(\lambda; \varepsilon, \zeta^2)\, \Phi_k\left(\lambda(\boldsymbol{\mu}^* - m\mathbf{1}_k); (\tau\sigma_i^{*2})^{-1/2}\mathbf{I}_k\right), \tag{6}$$

where $\boldsymbol{\mu}^*$ is the vector formed by the k different components of the vector $\boldsymbol{\mu}$, $M_i^* = (m + \tau n_i \overline{y}_{(i)})(1 + \tau n_i)^{-1}$, $V_i^* = \tau\sigma_i^{*2}(1 + \tau n_i)^{-1}$, $\lambda_i^* = -\lambda(\tau\sigma_i^{*2})^{-1/2}$, and $\mathbf{Z}_i^* = (\mu_i^* - m)(\tau\sigma_i^{*2})^{-1/2}$.

Stochastic Representations

Let W, U_i and $\mathbf{U}$ be real random quantities.

(i) Assume that, given $\mathbf{y}_{(i)}$, σ_i^{*2} and λ, W and U_i are independent with $W \sim N(M^*, V_i^*[1 + V_i^*(\lambda_i^*)^2]^{-1})$ and $U_i \sim LTN_1(0, 1 + V_i^*(\lambda_i^*)^2; -\lambda_i^*(M^* - m))$. Then,

$$\mu_i^* | \sigma_i^{*2}, \lambda, \mathbf{y}_{(i)} \stackrel{d}{=} W + [V_i^* \lambda_i^* U_i][1 + V_i^*(\lambda_i^*)^2]^{-1}.$$

(ii) Assume that, given $\mathbf{y}_{(i)}$, σ_i^{*2} and μ_i^*, W and U_i are independent with $W \sim N(\varepsilon, \zeta^2[1 + \zeta^2 \mathbf{Z}_\mu \mathbf{Z}_\mu^T]^{-1})$ and $\mathbf{U} \sim LTN_k(\mathbf{0}, \boldsymbol{I}_k + \zeta^2 \mathbf{Z}_\mu^T \mathbf{Z}_\mu; -\varepsilon \mathbf{Z}_\mu;)$, where $\mathbf{Z}_\mu = [\boldsymbol{\mu}^* - m\mathbf{1}_k][\tau(\sigma_i)^2]^{-1/2}$. Then,

$$\lambda | \mu_i^*, \sigma_i^{*2}, \mathbf{y}_{(i)} \stackrel{d}{=} W + [\zeta^2 \mathbf{Z}_\mu \mathbf{U}^T][1 + \zeta^2 \mathbf{Z}_\mu \mathbf{Z}_\mu^T]^{-1}$$

.

Simulated Study: Model comparison

Prior Specifications

- $\mu_j \mid \sigma_j^2 \sim N(\bar{y}, 5\sigma_j^2)$
- $\sigma_j^2 \sim IG(3, 8) \rightarrow E(\sigma_j^2) = 4$ and $V(\sigma_j^2) = 16$
- $\lambda_j \sim N(0, 100) \leftarrow$ DPMSN
- $\lambda \sim N(100, 1) \leftarrow$ DPMN-SN

For the MCMC

- Final sample size+10,000
- Burn-in = 190,000
- lag=1

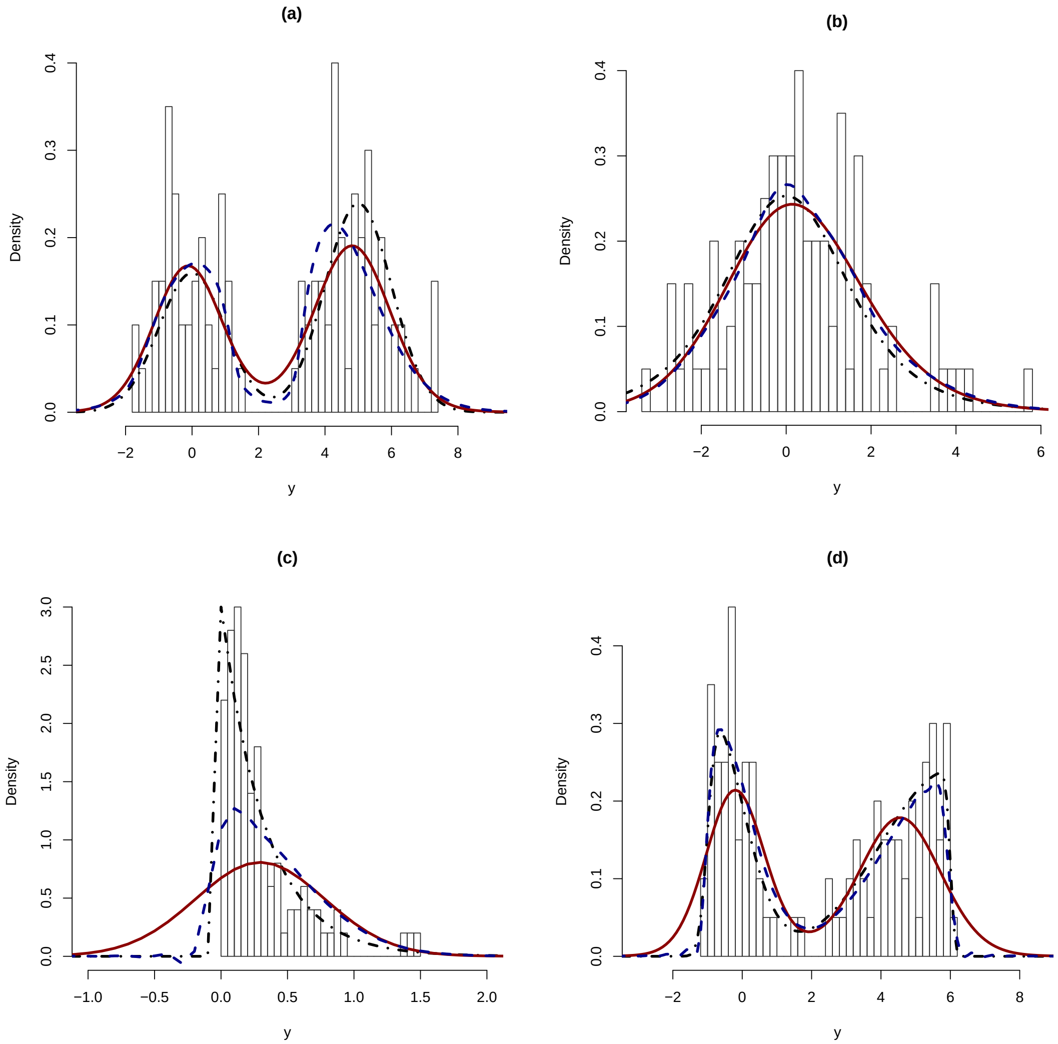
Data are generated of both symmetric and asymmetric distributions.

DPMN-N *versus* DPMSN

Tabela 1: Integrated mean square error

Scenario	Model	DPMSN	DPMN-N
1	$Y \sim 0.4N(0,1) + 0.6N(5,1)$	3.93×10^{-4}	1.69×10^{-4}
2	$Y \sim t_5(0, 1.5)$	5.65×10^{-5}	5.73×10^{-5}
3	$Y \sim Exp(3)$	4.63×10^{-2}	9.50×10^{-2}
4	$Y \sim 0.6SN(-1, 1, 5)$ $+0.4SN(6, 2, -20)$	0.29×10^{-3}	1.20×10^{-3}

Generated data, the true density (dotdashed (black) line) and estimates using the DPMSN (dashed (blue) line) and DPMN-N (solid (red) line), Scenarios 1 to 4 in (a), (b), (c) and (d).



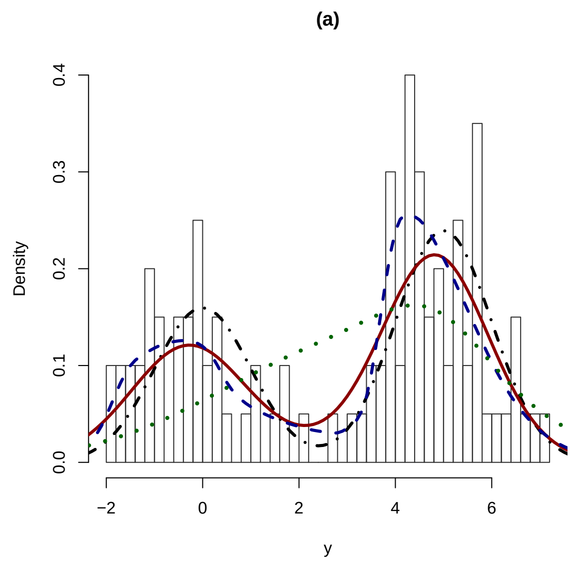
DPMN-N *versus* DPMSN

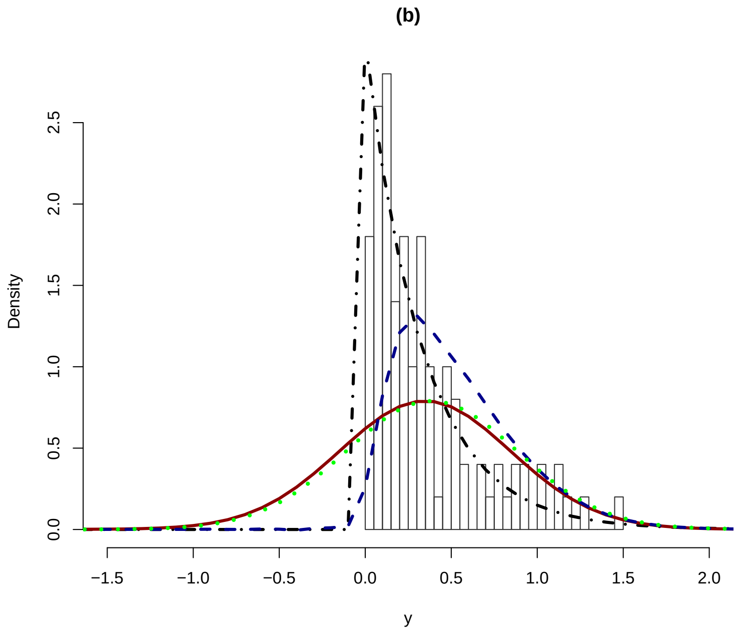
In summary,

- Scenario 1: DPMN-N and DPMSN provide similar estimates for the first component of the mixture. DPMN-N works better to identify the second mode location and DPMSN do better to estimate its height.
- Scenario 2: DPMN-N and DPMSN are comparable.
- Scenarios 3 and 4: DPMSN has better performance. DPMN-N fails in estimating the tails of the distributions and the location and heights of the modes.

DPMN-N *versus* DPMN-SN *versus* DPMSN

Generated data, the true density (dotdashed (black) line) and the estimates using the DPMSN (dashed (blue) line), DPMN-SN (dotted (red) line) and DPMN-N (solid (green) line) for Scenarios 1 (a) and 3 (b).





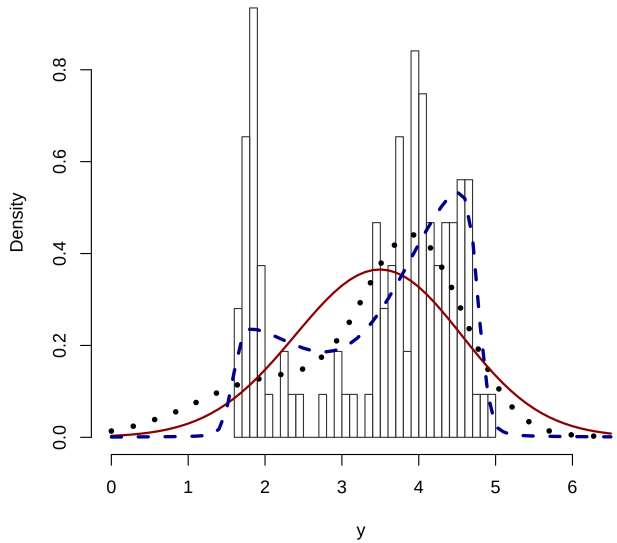
DPMN-N *versus* DPMN-SN *versus* DPMSN

In summary,

- Scenario 1: DPMN-SN capture the asymmetry of the distribution but not its bimodality.
- Scenario 3: DPMN-N and DPMN-SN works in a similar way. That is expected only if $\lambda \sim N(0, V)$.

Old Faithful Geyser data set

Data: the eruption duration time of the geyser (107 observations)



Eruption duration time data and the density estimates: DPMSN (dashed (blue) line) and DPMN-N using MacEachern and Müller's (solid (black) line) and Escobar and West's (dotted (red) line) algorithms.

Final Comments

In summary,

- DPMSN was better to estimate asymmetric distribution;
- DPMSN was comparable to DPMN-N in scenarios that favors the DPMN-N;
- Computational procedures are expensive, mainly, under skewness;
- We can use the stochastic representation or the Metropolis-Hastings algorithm to sample from the f.c.d..

Challenges...

- To estimate multivariate distributions;
- To improve the performance of the computational procedures
  - $*$ Particle learning filtering (Is it possible?)
  - $*$ INLA $\leftarrow$ under Gaussian structures use to works well.
- Is the estimated curve a good approximation for the true one?
  $\leftarrow$ Bayesian hypothesis tests.

Thank you!!