Bayesian Statistics

Debdeep Pati
Florida State University

October 7, 2014

GP-LVM for density estimation (Kundu & Dunson, 2013, Biometrika)

- Consider the latent variable model,

$$y_i = \mu(\eta_i) + \epsilon_i, \quad \epsilon_i \sim \mathsf{N}(0, \sigma^2), \ (i = 1, \ldots, n)$$
$$\mu \sim GP(0, c), \quad \sigma \sim \mathsf{IG}(a, b), \quad \eta_i \sim \mathsf{U}(0, 1),$$

- η_i's are subject specific latent variables, $\mu \in C[0, 1]$ is a *transfer function* relating the latent variables to the observed variables.
- The density of y conditional on the transfer function μ and scale σ is obtained on marginalizing out the latent variable as

$$f(y; \mu, \sigma) \stackrel{\text{def}}{=} f_{\mu,\sigma}(y) = \int_0^1 \phi_\sigma\{y - \mu(x)\}dx \approx \frac{1}{T}\sum_{j=1}^{T} \phi_\sigma\left\{y - \mu\left(\frac{j}{T}\right)\right\}.$$

Flexibility of GP-LVM

- Let $F_0(t) = \int_{-\infty}^{t} f_0(x)dx$.
- Letting $\mu_0(x) = F_0^{-1}(x)$, one obtains

$$f_{\mu_0,\sigma}(y) = \int_0^1 \phi_\sigma(y - F_0^{-1}(x))dx = \int_{-\infty}^{\infty} \phi_\sigma(y - t)f_0(t)dt,$$

  where the second equality follows from the change of variable theorem.
- $\|\phi_\sigma * f_0 - f_0\| \to 0$
- Posterior computation: i) $\mu \mid \mathbf{y}, \eta, \sigma$ is Gaussian ii) $\eta \mid \mathbf{y}, \mu, \sigma$ Metropolis Hastings iii) $\sigma \mid \mathbf{y}, \eta, \sigma$ is Inverse Gamma

Some applications of the Dirichlet process Mixture models

Hierarchical Modeling

- θ_i = random effects specific to "subject" i
- Hierarchical models let $\theta_i \sim P$
- P = random effects distribution
- Choice of P critical in controlling borrowing of information

Some Classical Applications

- Meta Analysis: combine data from multiple studies to make overall conclusion (e.g., drug is effective)
- Multi-level Designs: subjects are nested in schools, regions or study centers
- Longitudinal Data: data collected for subject over time - important to accommodate within-subject dependence

Some Emerging Applications

- Joint modeling of data from different domains
  1. Images and captions
  2. Diagnostic images or functional predictors & health responses
  3. Multiple types of omics data (sequence & expression)
- Multi-task learning: borrow strength across tasks
  1. Multiple images, music pieces, security videos
  2. Compressive sensing
  3. User preferences in different domains (film, books, etc)

Application 1 - Multinational Bioassay

- Increasing concern about adverse effects of environmental estrogens on human development
- Rodent uterotrophic bioassay: system for identifying suspected agonists or antagonists of estrogen.
- OECD study: collected data from 19 laboratories to investigate consistency of effects of known agonist (EE) & antagonist (ZM)
- y_{ij} = uterus weight for rat j in lab i
- $\mathbf{x}_{ij}$=lab indicator, dose of EE, dose of ZM

Some Comments

- Can potentially fit normal random effects model,

$$\begin{aligned} y_{ij} &= \mathbf{x}_{ij}'\mathbf{b}_i + \epsilon_{ij}, \epsilon_{ij} \sim \mathrm{N}(0, \sigma^2) \\ \boldsymbol{\theta}_i &\sim \mathrm{N}_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}) \end{aligned}$$

- Normal distribution has light tails & does not allow outlying labs or clusters of labs
- Conclusions may be sensitive to violations of normality
- Appealing to have a more flexible approach available

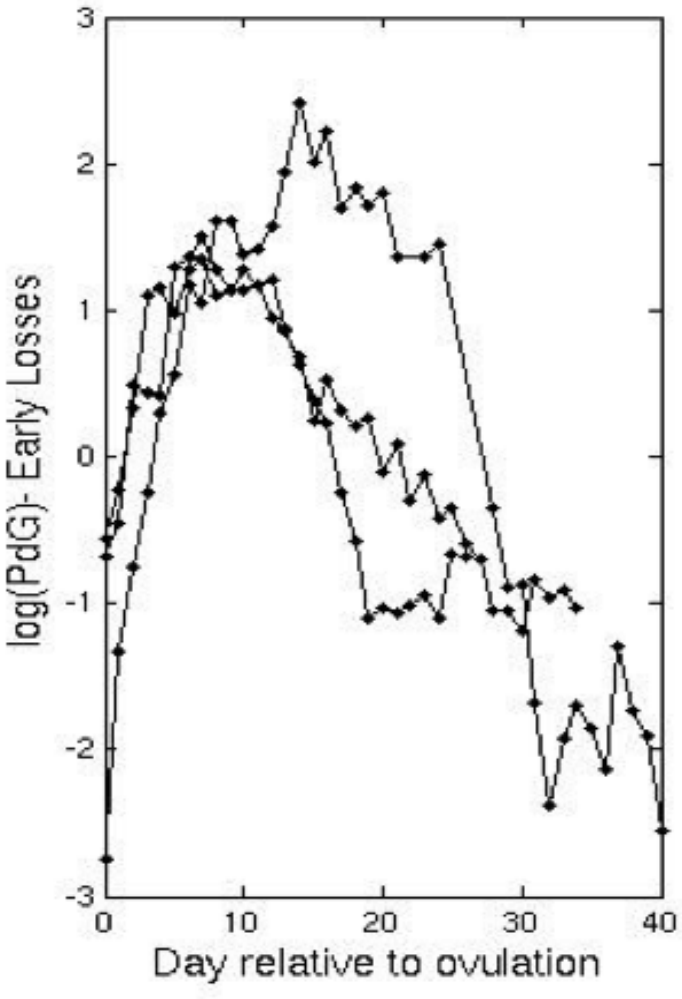
DPM for the random effects

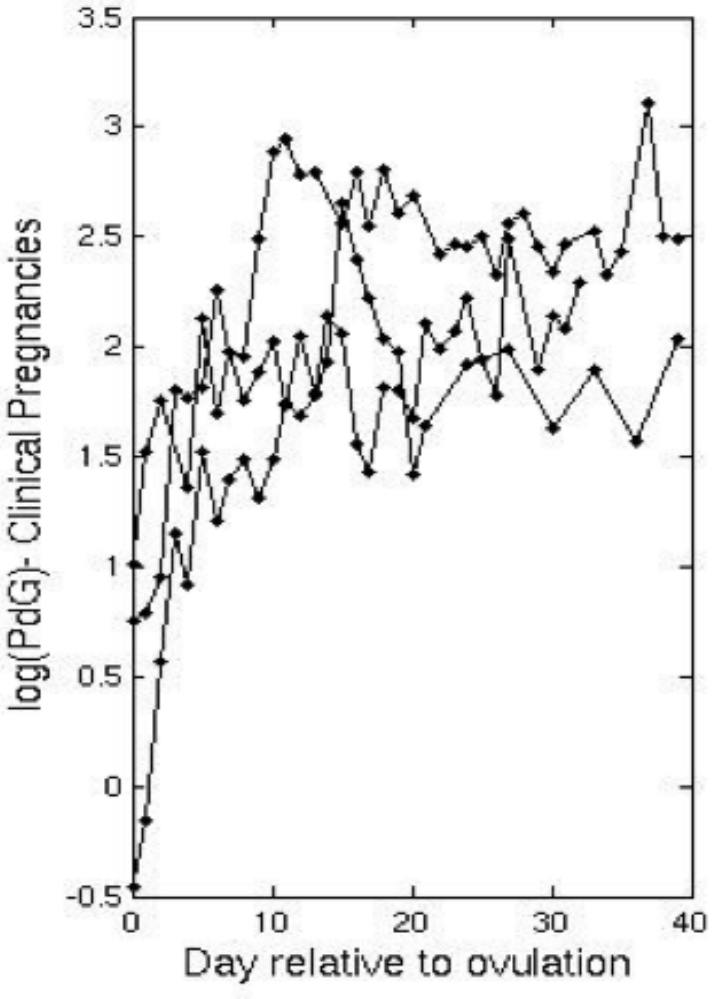
- A simple case corresponds to the linear mixed effects model

$$\begin{aligned} y_{ij} &= \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}_i + \epsilon_{ij}, \epsilon_{ij} \sim \mathrm{N}(0, \sigma^2) \\ \mathbf{b}_i &\sim P, \quad P \sim DP(\alpha P_0) \end{aligned}$$

- DP prior on P, the distribution of the random effects
- Useful semiparametric model for longitudinal & correlated data
- Bush & MacEachern (1996), Müller & Rosner (1997), Kleinman & Ibrahim (1998), Ishwaran & Takahara (2002), etc

Application 2 - longitudinal data (log PdG trajectories (Bigelow & Dunson, JASA, 08))





Functional Dirichlet process

- Interest in estimating a collection of functions, $\{f_i\}_{i=1}^n$.
- Longitudinal trajectories for different individuals
- We will focus on the following model:

$$\begin{aligned} y_{ij} &= f_i(t_{ij}) + \epsilon_{ij}, \quad \epsilon_{ij} \sim t_\nu(\sigma^2) \\ f_i(t) &= \sum_{j=1}^{p} \beta_{ij} b_j(t) = \mathbf{b}(t)'\boldsymbol{\beta}_i \\ \boldsymbol{\beta}_i &\sim P \end{aligned}$$

$\mathbf{b} = \{bj\}$=basis functions, $\boldsymbol{\beta}_i$ =basis coefficients

Comments on Functional Data Model

- Subject-specific basis coefficients, β_i , allow variability in the functional trajectories for different individuals
- Heterogeneity among subjects controlled by the random effects distribution, P
- Number of basis functions, p, is not small ($p \geq 20$)

Clustering the curves

- Characterize variability in growth curves & cluster subjects having similar trajectories
- Can be accomplished using DPM linear mixed model with

$$f_i(t_{ij}) = \sum_{l=1}^{p} \beta_{il} b_l(t_{ij})$$

$$\boldsymbol{\beta}_i \sim P = \sum_{h=1}^{\infty} \pi_h \delta_{\boldsymbol{\beta}_h^*}$$

Comments- Functional Dirichlet Process

- Recalling the DP stick-breaking property (Sethuraman, 1994):

$$\beta_i \sim P = \sum_{h=1}^{\infty} V_h \prod_{l<h}(1 - V_l)\delta_{\beta_h^*}, V_h \overset{i.i.d}{\sim} \text{Beta}(1, \alpha), \beta_h^* \sim P_0$$

- Hence, the n subjects are grouped into $k \leq n$ clusters
- Subjects in cluster l all have $\beta_i = \beta_l^*$.
- Provides a semiparametric Bayes version of latent trajectory class or growth mixture models.
- Avoids fixing the number of clusters in advance

Comments- Functional Dirichlet Process

- The curve in cluster l is $f(t) = \mathbf{b}(t)'\boldsymbol{\beta}_l^*$.
- The number of functional clusters in n growth curves is treated as unknown
- Gibbs samplers are straightforward to generalize
- Number of clusters and configuration of subjects into clusters varies across the MCMC iterations
- Problem: label switching!

Label Switching

- Problem arises because the labels on the cluster-specific parameters are ambiguous, so vary in meaning across the iterations
- Not meaningful to calculate posterior summaries of ??h across the iterations
- Strategies:
  1. Relabeling algorithms that align the clusters after running MCMC (Stephens, 00);
  2. Define clusters as individuals that are grouped together with high posterior probability
  3. Estimate optimal clustering (Dahl, 06; Lau & Green, 97)
  4. Ignore problem & avoid cluster-specific inferences