Nonparametric Bayesian Statistics

Debdeep Pati
Florida State University

October 21, 2014

Incorporating constraints in NP Bayes

- Interest in relationship between a predictor x and response y adjusting for covariates z.
- In epidemiology & toxicology studies, $x =$ dose of a potentially adverse exposure
- Biologically reasonable to assume that response distribution is stochastically non-decreasing with x for any fixed z

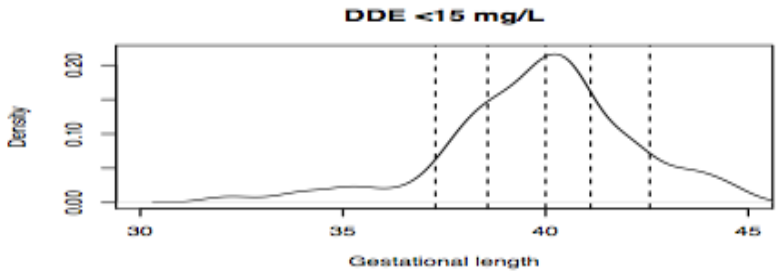
Isotonic Regression

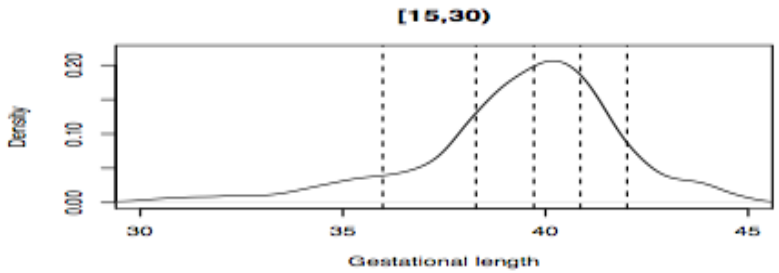
- In order to incorporate non-decreasing constraint, one may consider an isotonic regression model
- Ramsay (1988) propose monotone regression splines - flexible estimation of smoothly increasing regression curve
- In toxicology & epidemiology, interest in inferences on flat vs increasing regions - flat corresponds to no effect of predictor

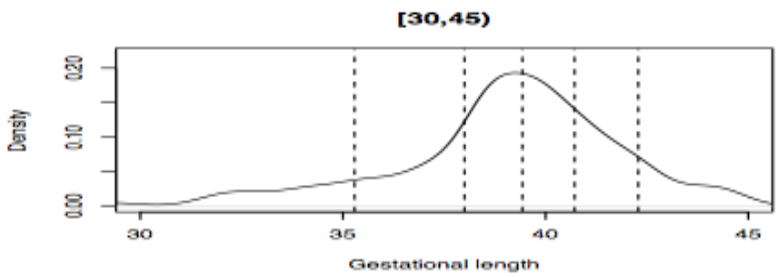
Isotonic Mean Regression $\rightarrow$ density regression

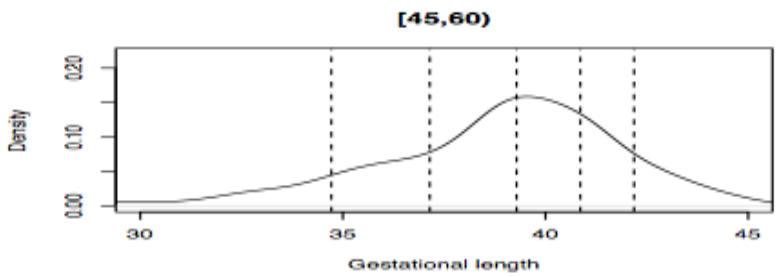
- Above methods focus on estimation of a non-decreasing function, $f(x)$, subject to $f(x) \leq f(x')$, for all $x < x'$.
- Typically, $f(x) =$ mean regression function, with residual distribution constant with x
- Casady & Cryer (1976) - isotonic quantile estimator
- Our focus: allow conditional distribution, $f(y|x)$, to change flexibly with x subject to non-increasing stochastic order

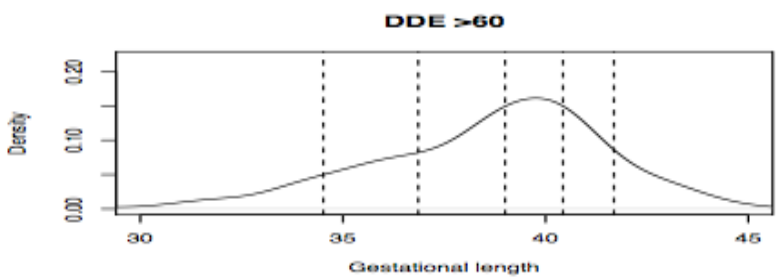
Application - DDE & Preterm Birth











Application - DDE & Preterm Birth

- Interest - how distribution of a continuous health response changes with a discrete / continuous predictor
- y_i = gestational age at delivery, x_i =level of DDE in maternal serum, z_i =potential confounders
- As exposure increases, distribution of gestational age at delivery stochastic non-increasing
- How to nonparametrically model such changes and conduct hypothesis tests

Modeling of Stochastic Ordering

- Initially consider case with two unknown distributions, P_1 and P_2 defined on $\mathcal{X}$ with $P_1 \precsim P_2$.
- Bayesian estimation considered by Gelfand & Kottas (01) - used products of independent DP components
- Our Goals:
  1. Flexible Prior on set C_E of stochastically ordered distributions
  2. Efficient MCMC approach for computation
  3. Methods for hypothesis testing

Hoff (2003, Biometrika) Formulation

- Goal is to choose a prior for $(P_1, \ldots, P_K) \in C_E$
- C_E = weakly closed convex set with extreme points $\{(\delta_{s_1}, \ldots, \delta_{s_K}) : (s_1, \ldots, s_K) \in S_K\}$
- $S_K = \{(s_1, \ldots, s_K) \in \mathcal{X}^K : s_i \leq s_j \forall (i,j) \in E\}$, with E pre-specified matrix defining ordering
- Main Result: DP prior on S_K induces prior on C_E

Restricted Dependent DP (Dunson & Peddada, 07)

- Hoff (03) does not allow continuous distributions, computation is difficult & no allowance for hypothesis testing
- Applying Sethuraman (94) & Hoff (03), we define restricted dependent DP (rDDP) priors for $(P_1, \ldots, P_K) \in C_E$:

$$P_k(\cdot) = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_{hk}}, \quad \boldsymbol{\Theta}_h = (\Theta_{h1}, \ldots, \Theta_{hK}) \sim Q_0$$

  with Q_0 a base measure on S_K & $\{\pi_h\}$ = typical DP weights
- rDDP modifies DDP (MacEachern, 99) to use restricted Q_0

Two Group Example

- $(P_1, P_2) \sim rDDP(\alpha Q_0)$ implies:

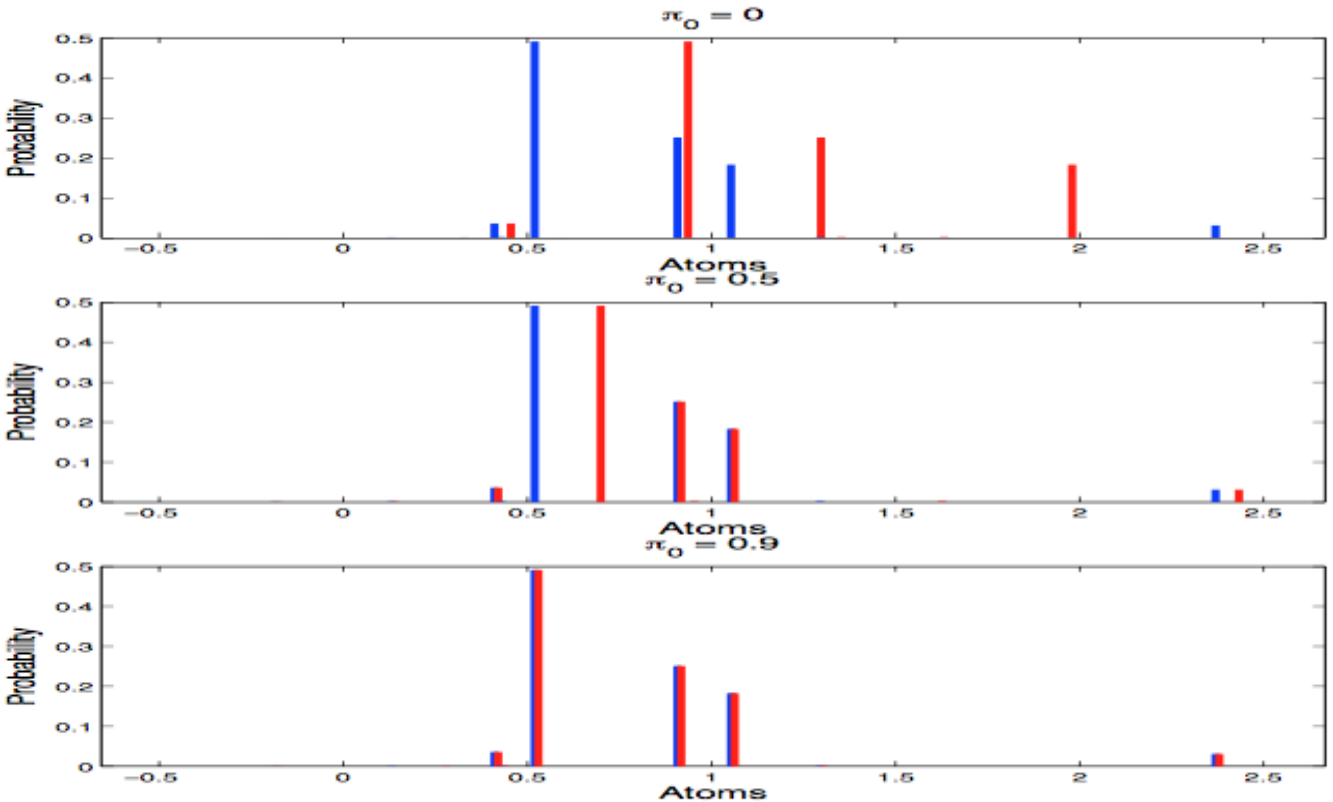
$$P_1 = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_{h1}}, \quad P_2 = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_{h2}}$$

- $(\Theta_{h1}, \Theta_{h2}) \sim Q_0$, with $Q_0(\Theta_{h1} \leq \Theta_{h2}) = 1$
- For example, $Q_0 =$ truncated bivariate normal
- To limit bias & facilitate testing, allow probability mass on the boundary by choosing Q_0 to correspond to:

$$f(\Theta_1, \Theta_2) = f_1(\Theta_1)\{\pi_0 \delta_0(\Theta_2 - \Theta_1) + (1 - \pi_0) f_2(\Theta_2 - \Theta_1).$$

where f_1 a density on $\mathbb{R}$ and f_2 is a density supported on $\mathbb{R}^+$

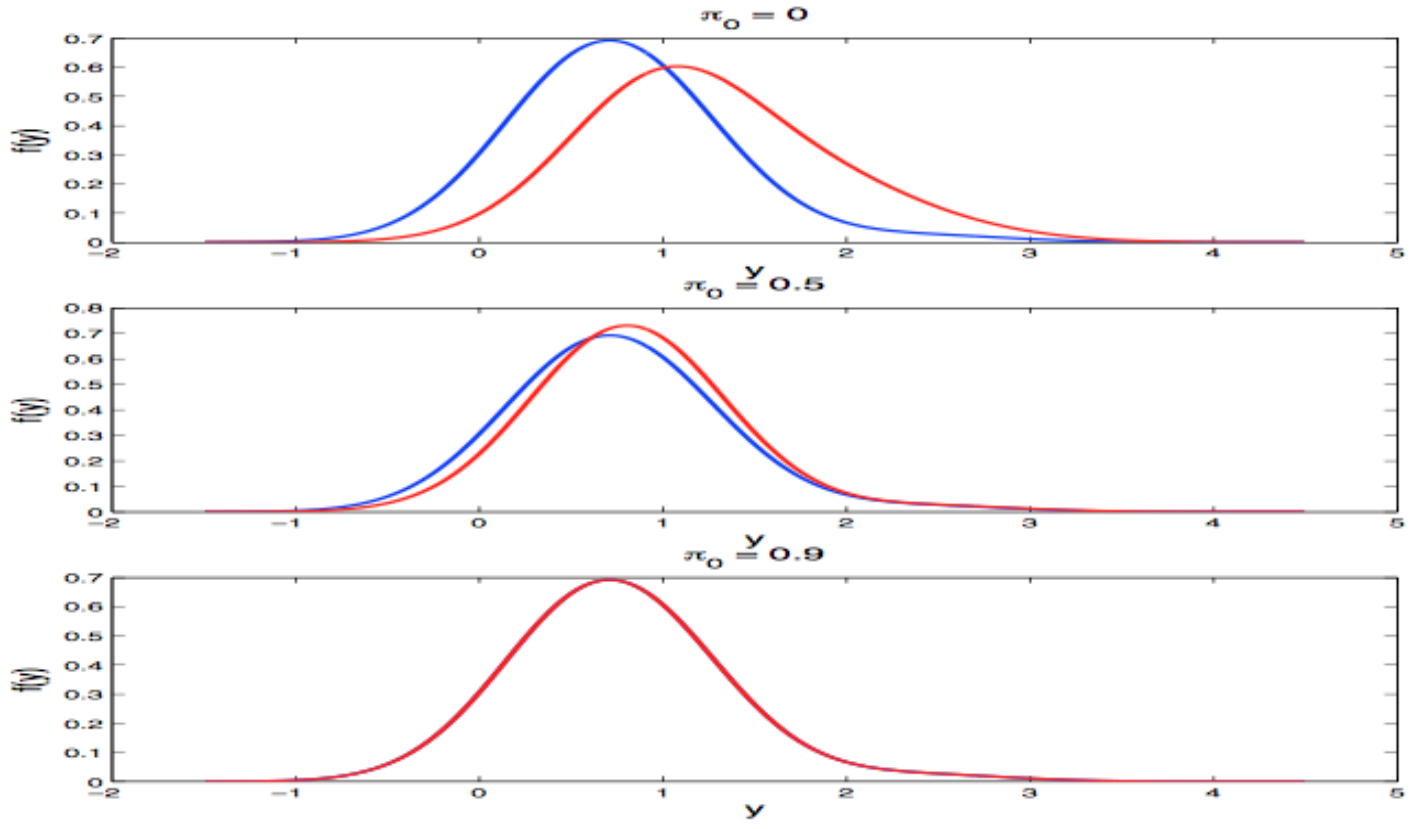
One draw from the rDDP



rDDP Mixtures

- We can use $(P_1, \ldots, P_K)$ as a collection of mixture distributions to obtain a class of rDDPM models
- Let $g_k(y) = \int K(y, \mu) dP_k(\mu)$ denote the density in group k, with $K(\cdot)$ a kernel satisfying monotone stochastic order
- Integral operator induces mapping from $C_E \to \mathcal{L}_E$, with $(P_1, \ldots, P_K) \in C_E \& (g_1, \ldots, g_K) \in \mathcal{L}_E$.
- For normal $K(\cdot)$, $\mathcal{L}_E$ contains all $K \times 1$ collections of densities satisfying partial ordering E in its closure

One draw from the rDDPM



Hypothesis Testing - 2 Group Case

- Hypothesis testing of equalities in distributions, g_1, g_2, against stochastically ordered alternatives
- Differences in g_1, g_2 controlled through differences in mixture distributions P_1, P_2
- In two group case, focus on interval null based on TV distance

$$d_{12} = \max_{B \in \mathcal{B}} |P_1(B) - P_2(B)|$$

- With $\beta_h = \Theta_{2h} - \Theta_{1h}$,
  $d_{12} = \sum_{h=1}^{\infty} \pi_h 1_{(\beta_h > 0)} \sim \text{Beta}(\alpha(1-\pi_0), \alpha\pi_0)$
- Hypotheses: $H_0 : d_{12} \leq \epsilon, H_1 : d_{12} > \epsilon$.

Justification for Hypothesis Formulation

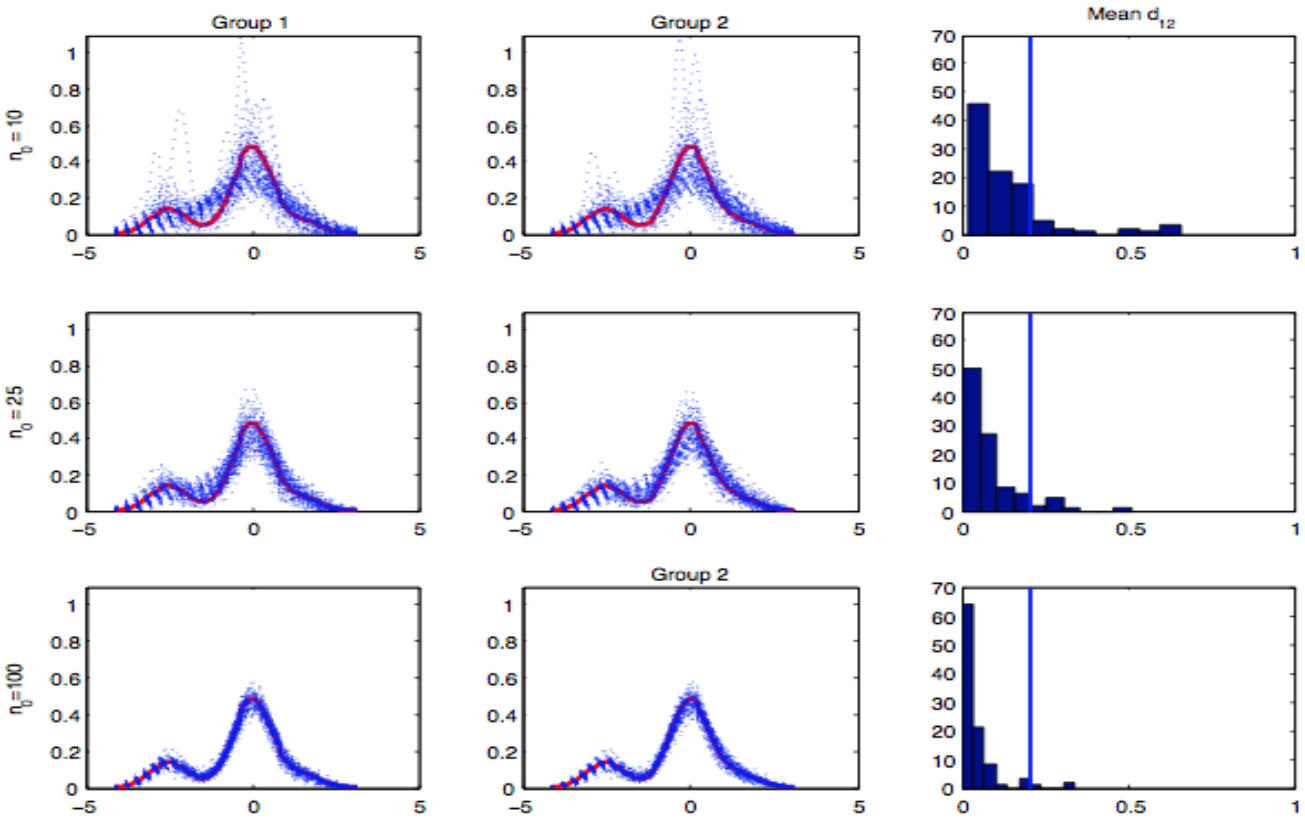
- Basing hypothesis tests on d_{12} appealing for simple prior elicitation & posterior computation
- Can choose π_0 to assign 0.5 probability to the null hypothesis.
- *Theorem 1.* Let $G_k(B) = \int_B g_k(y)dy, k = 1, 2$ with $g_k(y) = \int K(y, s)dP_k(s)$. Then, $H_0 : d_{12} \leq \epsilon$ implies

$$\max_{y \in \mathcal{Y}} |G_2(y, \infty) - G_1(y, \infty)| \leq \epsilon$$

Multiple Groups & Computation

- Extensions to multiple groups & censored data are trivial
- Posterior computation can proceed via a highly-efficient & simple blocked Gibbs sampler (Ishwaran & James, 01)
- We ran a simulation study to assess frequentist operating characteristics

Simulation results (2 groups, null hypothesis true)



Simulation results (2 groups, alternative hypothesis true)