

# Nonparametric Bayesian Statistics

Debdeep Pati  
Florida State University

October 23, 2014

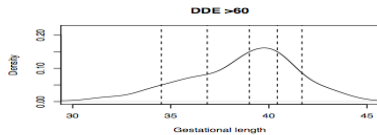
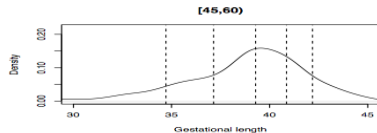
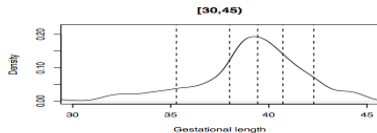
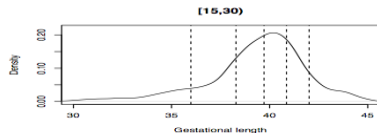
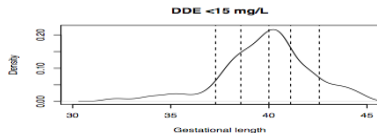
- ▶ Interest in relationship between a predictor  $x$  and response  $y$  adjusting for covariates  $z$ .
- ▶ In epidemiology & toxicology studies,  $x =$  dose of a potentially adverse exposure
- ▶ Biologically reasonable to assume that response distribution is stochastically non-decreasing with  $x$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

- ▶ 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



- ▶ 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  $P_1 \lesssim P_2$ .
- ▶ Bayesian estimation considered by Gelfand & Kottas (01) - used products of independent DP components
- ▶ Our Goals:
  1. Flexible Prior on set  $\mathcal{C}_E$  of stochastically ordered distributions
  2. Efficient MCMC approach for computation
  3. Methods for hypothesis testing

- ▶ Goal is to choose a prior for  $(P_1, \dots, P_K) \in C_E$
- ▶  $C_E =$  weakly closed convex set with extreme points  $\{(\delta_s) : s = (s_1, \dots, s_K) \in S_K\}$
- ▶  $S_K = \{(s_1, \dots, s_K) \in \mathcal{X}^K : s_i \leq s_j \forall (i, j) \in E\}$ , 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  $(P_1, \dots, P_K) \in C_E$ :

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

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:

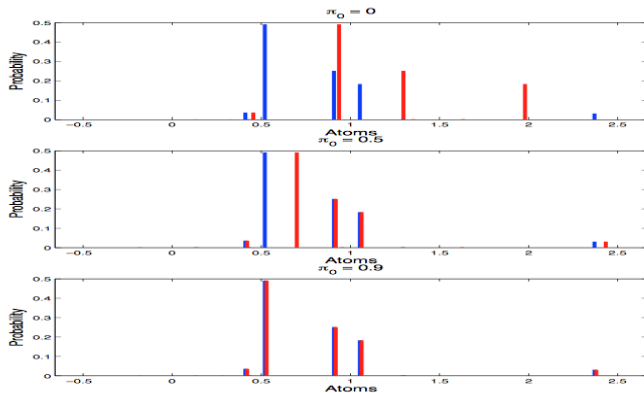
$$P_1 = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_{h1}}, \quad P_2 = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_{h2}}$$

- ▶  $(\Theta_{h1}, \Theta_{h2}) \sim Q_0$ , 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:

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

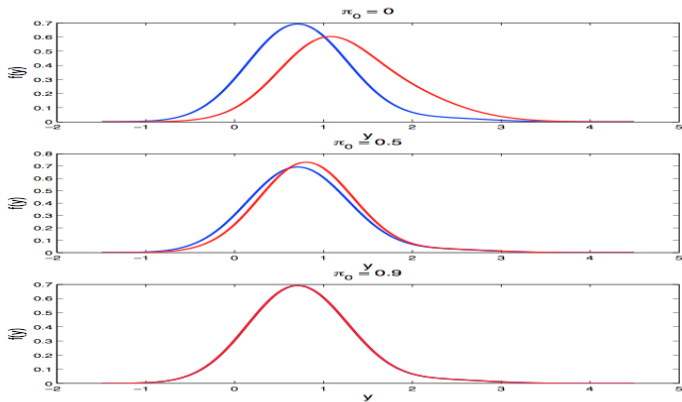
where  $f_1$  a density on  $\mathbb{R}$  and  $f_2$  is a density supported on  $\mathbb{R}^+$

# One draw from the rDDP



- ▶ We can use  $(P_1, \dots, P_K)$  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  $C_E \rightarrow \mathcal{L}_E$ , with  $(P_1, \dots, P_K) \in C_E \& (g_1, \dots, g_K) \in \mathcal{L}_E$ .
- ▶ 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:  $H_0 : d_{12} \leq \epsilon, H_1 : d_{12} > \epsilon$ .

# Multiple Groups & Computation

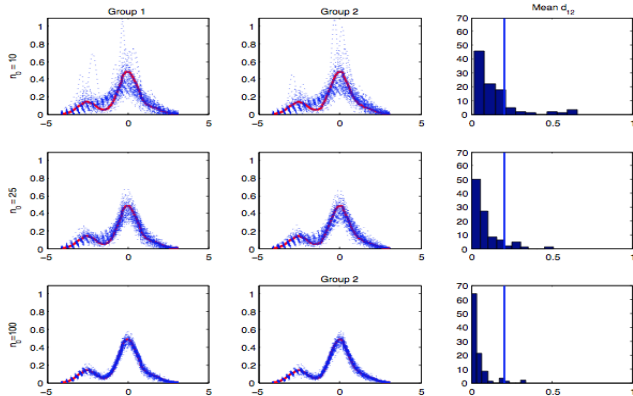
- ▶ Extensions to multiple groups & censored data are trivial
- ▶ Posterior computation can proceed via a highly-efficient & simple blocked Gibbs sampler (Ishwaran & James, 01)
- ▶ We ran a simulation study to assess frequentist operating characteristics

## Simulation study ( $K = 2$ )

- ▶ Case 1: Both Group 1 and Group 2 generated from  $f(y) = 0.2N(y; -2.5, \tau^{-1}) + 0.7N(y; 0, \tau^{-1}) + 0.1N(y; 1.5, \tau^{-1})$ ,  $\tau = 3$
- ▶ Case 2: Change the component-specific means  $(-2.4, 0.4, 2.2)$ .
- ▶ Simulated 100 datasets under three sample sizes, 10, 25 and 100.
- ▶ Fix  $\epsilon$ , choose  $\pi_0$  such that  $P(d_{12} < \epsilon) = 0.5$ .



# Simulation results (Case 1)



# Simulation results (Case 2)

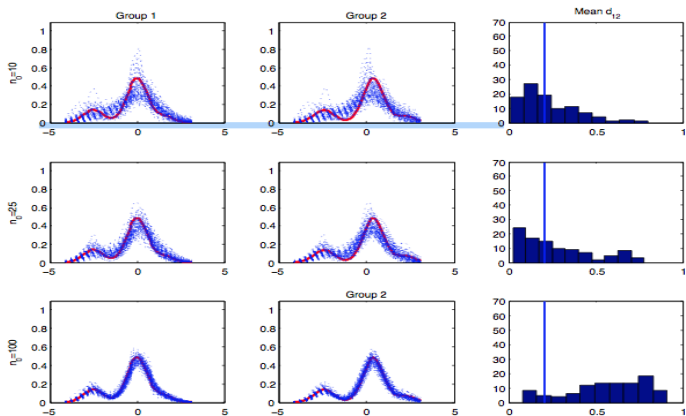


Table 1. *Simulation study. Summary statistics of  $d_{12}$  across the simulations for each sample size. Results shown are means with 95% empirical confidence limits in parentheses*

Case	Common sample size	$d_{12}$	$\text{pr}(d_{12} < \epsilon \mid \text{data})$		
			$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0.1$
1	10	0.140	0.67	0.72	0.75
		(0.028, 0.614)	(0.17, 0.86)	(0.18, 0.91)	(0.22, 0.93)
1	25	0.085	0.72	0.78	0.81
		(0.011, 0.299)	(0.19, 0.89)	(0.22, 0.94)	(0.28, 0.96)
1	100	0.044	0.79	0.85	0.89
		(0.003, 0.215)	(0.28, 0.94)	(0.30, 0.99)	(0.42, 1.00)
2	10	0.233	0.55	0.60	0.63
		(0.029, 0.660)	(0.10, 0.82)	(0.10, 0.91)	(0.10, 0.93)
2	25	0.278	0.46	0.51	0.54
		(0.032, 0.765)	(0.02, 0.84)	(0.02, 0.87)	(0.06, 0.91)
2	100	0.560	0.13	0.15	0.18
		(0.091, 0.884)	(0.00, 0.74)	(0.00, 0.79)	(0.00, 0.83)

# DNA Repair Studies Application

- ▶ Interest in identifying genes predictive of DNA repair rates for individuals in Environmental Genome Project
- ▶ Frequency of strand breaks on individual cell level measured at baseline, after induced damage & after repair
- ▶ Initially study assessed impact of dose of  $H_2O_2$  & repair time using samples of 100 cells from single subject

# DNA Repair Studies Application

- ▶ Batches of cells were exposed to 0, 5, 20, 50 or  $100\mu$  mol of  $H_2O_2$  (hydrogen peroxide)
- ▶ DNA damage was then measured in individual cells after allowing a repair time of 0, 60 or 90 min.
- ▶  $x_i$  for cell  $i$ : Olive tail moment, surrogate of the frequency of DNA strand-breaks obtained using the comet assay.
- ▶ Let  $a_i \in \{1, \dots, K\}$  be a group index denoting the level of  $H_2O_2$  and repair time for cell  $i$ .
- ▶ The value of  $a_i$  for each dose  $\times$  repair time value is shown in Fig. 3.
- ▶ The total sample size is 1400, with 100 observation per group except for groups 9 and 13, which had 50.

# DNA Repair Studies Application

- ▶ Among cells with zero repair time, DNA damage should be nondecreasing with the dose of  $H_2O_2$ .
- ▶ In addition, within a given dose level, DNA damage should be non-increasing with repair time.
- ▶ we make the ordering assumption illustrated in Fig. 3 using a directed graph, with arrows pointing towards stochastically larger groups.
- ▶ We wish to assess whether or not DNA damage continues to increase at higher levels of  $H_2O_2$  exposure, and investigate whether or not damage is significantly reduced across each increment of the repair time.

# Directed graph indicating stochastic order

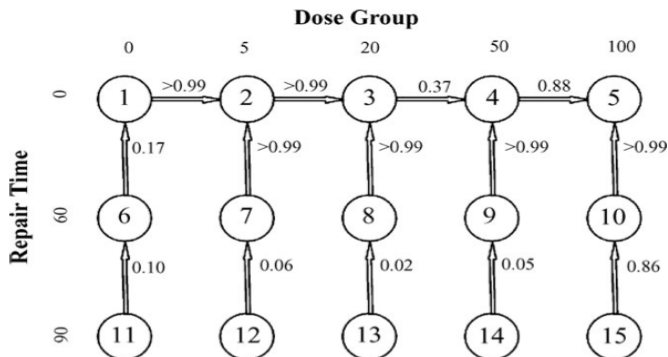
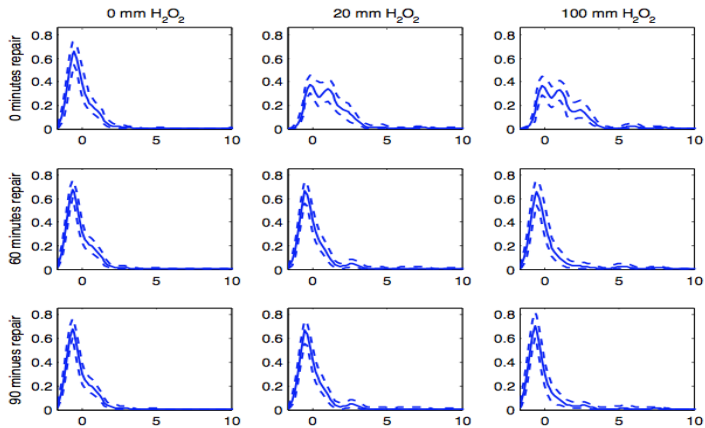


Fig. 3. Genotoxicity application. Directed graph illustrating order restriction. Arrows point towards stochastically larger groups. Posterior probabilities of  $H_{1k}$  are shown.

# Results for Genotoxicity Application





- ▶ Bayesian method for testing differences among groups against stochastically ordered alternatives
- ▶ Also allows estimation of densities (or pmfs) in each group
- ▶ Covariates can be included & approach can be embedded in hierarchical model (allowing censoring, missing data, multivariate observations of latent variable, etc)
- ▶ Allow stochastic ordering with continuous predictors?

# Stochastic Ordering with Continuous Predictors (Wang & Dunson, 07)

- ▶ Generalize estimation & testing methods to continuous predictors
- ▶ Let  $\mathcal{P}_{\mathcal{X}} = \{(P_x, x \in \mathcal{X}) : P_x \in \mathcal{P}, x \in \mathcal{X}\}$ ,  $\mathcal{P}$  = set of probability measures on  $\mathcal{B}(\mathcal{U})$ ,  $\mathcal{U}, \mathcal{X}$  = subsets of  $\mathbb{R}$
- ▶ Need to define a prior for  $\mathcal{P}_{\mathcal{X}}$  with support on

$$\mathcal{C}_{\mathcal{X}} = \{(P_x, x \in \mathcal{X}) \in \mathcal{P}_{\mathcal{X}} : P_x \preceq P_{x'}, x \leq x', x, x' \in \mathcal{X}\}.$$

- ▶ How to accomplish, while having simple computation & hypothesis testing?

# Mixing over Extreme Points

- ▶ Lemma 1 -  $C_{\mathcal{X}}$  = weakly closed convex set  $\rightarrow$  any elements of  $C_{\mathcal{X}}$  expressed as mixture over the extreme points
- ▶ Lemma 2 - the extreme points of  $C_{\mathcal{X}}$  are  $ex\{C_{\mathcal{X}}\} = \{(\delta_{s(x)}, x \in \mathcal{X}) : s \in \mathcal{S}_{\mathcal{X}}\}$ , where  $\mathcal{S}_{\mathcal{X}}$  = space of non-decreasing functions
- ▶ Theorem 1: For any  $P \in C_{\mathcal{X}}$  there exists a mixture measure  $Q$  such that  $(P_x, x \in \mathcal{X}) = \int_{\mathcal{S}_{\mathcal{X}}} (\delta_{s(x)} : x \in \mathcal{X}) dQ(s)$ .
- ▶ Prior for  $Q$  on  $\mathcal{S}_{\mathcal{X}}$  induces prior for  $P$  on  $C_{\mathcal{X}}$ !

- ▶ Motivated by this theory, we choose the rDDP prior:

$$P_x = \sum_{h=1}^{\infty} \pi_h \delta_{\Theta_h(x)},$$

where  $\theta_h \sim Q_0$  and  $Q_0$  has support on  $\mathcal{S}_X$ .

- ▶ The “atoms” in the Dirichlet process correspond to non-decreasing stochastic processes
- ▶ Monotone splines (Ramsay, 1988) for the functional atoms
- ▶ Letting  $f(y|x) = \int N(y; \mu, \sigma^2) dP_x(\mu)$ , obtain a method for isotonic density regression

- ▶ We consider the following local null and alternative:

$$H_0(x, x') : d_{TV}(x, x') < \epsilon \quad \text{vs.} \quad H_1(x, x') : d_{TV}(x, x') \geq \epsilon,$$

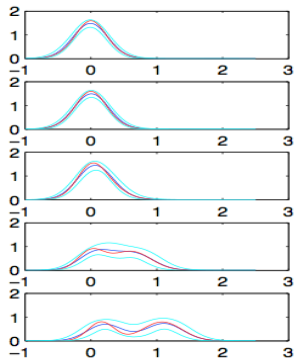
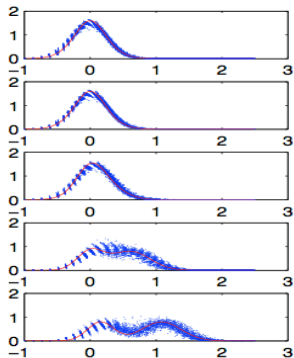
$$\text{with } d_{TV}(x, x') = \sup_{y \in \mathbb{R}} |F_x(y) - F_{x'}(y)|$$

- ▶ Local null  $H_0(x, x') \rightarrow$  effectively no change between  $x$  and  $x'$
- ▶ Global null formulated as intersections of local nulls
- ▶  $H_0(0, x) = \bigcap_{x' < x} H_0(x, x')$ .

# Simulation Study

- ▶ To assess performance, we ran a simulation study with 200 datasets have size  $n = 500$
- ▶ The response variable  $y$  was generated from a mixture of two normals, with no association for  $x \in [0, 0.3]$  & a non-linear association in  $x \in [0.3, 0.7]$
- ▶ Two normal components had different polynomial regressions in the mean

# Simulation Study Results - Conditional Densities



# Simulation Study Results - local testing

