Notes 3: Dirichlet process mixture models – Applications

Outline

1. Summary and references
2. Dirichlet process mixture modeling for survival analysis
3. Bayesian semiparametric quantile regression
4. Curve fitting using Dirichlet process mixtures
5. Modeling for multivariate ordinal data
6. Nonparametric inference for Poisson processes
7. Modeling for stochastically ordered distributions
1. Summary and references

Dirichlet process (DP) mixture models, and their extensions, have largely dominated applied Bayesian nonparametric work, after the technology for their simulation-based model fitting was introduced. References categorized by methodological/application area include:

- Models for binary and ordinal data (Erkanli et al., 1993; Basu & Mukhopadhyay, 2000; Hoff, 2005; Das & Chattopadhyay, 2004; Kottas et al., 2005)

- Density estimation, mixture deconvolution, and curve fitting (West et al., 1994; Escobar & West, 1995; Cao & West, 1996; Gasparini, 1996; Müller et al., 1996; Ishwaran & James, 2002; Do et al., 2005; Leslie et al., 2007; Lijoi et al., 2007)

- Regression modeling with structured error distributions and/or regression functions (Brunner, 1995; Lavine & Mockus, 1995; Kottas & Gelfand, 2001b; Dunson, 2005; Kottas & Krnjajić, 2009)
Summary and references

- Survival/reliability regression models (Kuo & Mallick, 1997; Gelfand & Kottas, 2003; Merrick et al., 2003; Hanson, 2006; Argiento et al., 2009; De Iorio et al., 2009)

- Generalized linear, and linear mixed, models (Bush & MacEachern, 1996; Kleinman & Ibrahim, 1998; Mukhopadhyay & Gelfand, 1997; Müller & Rosner, 1997; Quintana, 1998)

- Errors-in-variables models (Müller & Roeder, 1997); Multiple comparisons problems (Gopalan & Berry, 1998); Analysis of selection models (Lee & Berger, 1999)

- Meta-analysis and nonparametric ANOVA models (Mallick & Walker, 1997; Tomlinson & Escobar, 1999; Burr et al., 2003; De Iorio et al., 2004; Müller et al., 2004; Müller et al., 2005)

- Time series/econometrics applications (Müller et al., 1997; Chib & Hamilton, 2002; Hirano, 2002; Hasegawa & Kozumi, 2003; Griffin & Steel, 2004)

- ROC data analysis (Erkanli et al., 2006; Hanson et al., 2008)
2. Dirichlet process mixture modeling for survival analysis

- Bayesian nonparametric work for survival analysis has focused on prior models for cumulative hazard or hazard functions (gamma processes, extended gamma processes, Beta processes), or survival functions (Dirichlet processes, Polya trees) (see, e.g., Walker et al., 1999; Ibrahim et al., 2001)

- Dirichlet process mixture model

\[ f(t; G) = \int k(t; \theta) dG(\theta), \quad t \in R^+, \quad G \sim DP(\alpha, G_0) \]

→ kernel \( k(t; \theta) \) that yields mixtures with flexible density (and hazard) shapes is needed

- Mixtures of Weibull or gamma distributions (Kottas, 2006b; Hanson, 2006)
- Weibull Dirichlet process mixture model

\[
    t_i \mid (\gamma_i, \lambda_i) \overset{\text{iid.}}{\sim} K(t \mid \gamma_i, \lambda_i) = 1 - \exp(-t^{\gamma_i}/\lambda_i), \quad i = 1, \ldots, n
\]

\[
    (\gamma_i, \lambda_i) \mid G \overset{\text{i.i.d.}}{\sim} G, \quad i = 1, \ldots, n
\]

\[
    G \mid \alpha, \phi, \psi \sim \text{DP}(\alpha, G_0); \quad G_0 = U(\gamma \mid 0, \phi)IG(\lambda \mid c, \psi)
\]

\[
    \alpha, \phi, \psi \sim p(\alpha)p(\phi)p(\psi)
\]

- full posterior inference for all functionals of interest in survival analysis, including non-linear functionals (e.g., hazard function, and median survival time) — uses sampling from the posterior of $G$

- also, the prior distribution of functionals can be sampled (quantifies prior to posterior learning)
Dirichlet process mixture modeling for survival analysis

Data Illustrations

- Simulated data \((n = 200)\) from a mixture of two Lognormal distributions \(0.8LN(0, 0.25) + 0.2LN(1.2, 0.02)\)
  \(\rightarrow\) bimodal density
  \(\rightarrow\) non-monotone hazard function with 3 change points in the interval \((0, 5)\)
  where essentially all the probability mass lies

- Remission times (in weeks) for leukemia patients (Lawless, 1982)
  \(\rightarrow\) comparison of two treatments, A and B, each with 20 patients (3 and 2
  right censored survival times, respectively)
  \(\rightarrow\) “no evidence of a difference in distributions” based on classical tests
  that rely on approximate normality and assume proportional hazard functions (Lawless, 1982)
Figure 1: Simulated data. Histograms of posterior draws for $\alpha$ (denoted by $v$ in the panels) and $n^*$, under three prior choices for $\alpha$. The prior densities for $\alpha$ are denoted by the solid lines.
Dirichlet process mixture modeling for survival analysis

Figure 2: Simulated data. Inference, under three prior choices for $\alpha$. The upper panels provide prior (dotted lines) and posterior (dashed lines) point and interval estimates for the survival function. The lower panels include the histogram of the data along with the posterior point estimate (dashed line) for the density function. In each graph, the solid line denotes the true curve.
Figure 3: Simulated data. Posterior inference for the hazard function. Under a \( \text{gamma}(2,0.1) \) prior for \( \alpha \), the left panel provides point and interval estimates (dashed lines). The right panel compares point estimates under three priors for \( \alpha \), \( \text{gamma}(2,0.9) \) (smaller dashed line), \( \text{gamma}(2,0.1) \) (dashed line) and \( \text{gamma}(3,0.05) \) (dotted line). In each graph, the solid line denotes the true hazard function.
Figure 4: Data on remission times for leukemia patients. (a) Posterior point and interval estimates of the survival function for treatment A, under three different priors for $\alpha$. Under the gamma(2,0.9) prior for $\alpha$, Figures 4(b), 4(c) and 4(d) compare the survival functions (point and interval estimates), density functions and hazard functions (point estimates), respectively, for treatments A (solid lines) and B (dashed lines).
Dirichlet process mixture modeling for survival analysis

Figure 5: Data on remission times for leukemia patients. (a) Posterior point estimate (solid line) and 95% interval estimates (dashed lines) for $p(\lambda_B(t_0)/\lambda_A(t_0) \mid \text{data})$ (ratio of hazard functions). (b) Posterior point estimate (solid line), 80% interval estimates (dotted lines) and 95% interval estimates (dashed lines) for $p(F_B(t_0) - F_A(t_0) \mid \text{data})$ (difference of survival functions). (c) Histogram of draws from $p(\text{median}(A) - \text{median}(B) \mid \text{data})$ (difference of median survival times).
3. Bayesian semiparametric quantile regression

- In regression settings, the covariates may have effect not only on the location of response distribution but also on its shape. Quantile regression quantifies relationship between a set of quantiles of response distribution and covariates, and thus, provides a more complete explanation of the response distribution in terms of available covariates (Applications: econometrics, medicine, social sciences, and educational studies)

- Response observations $y_i$, with covariate vectors $x_i$. Additive quantile regression formulation: $y_i = x_i' \beta + \varepsilon_i, i = 1, ..., n$

  $\rightarrow \varepsilon_i$ i.i.d. from an error distribution with $p$-th quantile equal to 0, i.e.,
  $\int_{-\infty}^{0} f_p(\varepsilon) d\varepsilon = p$

- **Objective:** develop flexible nonparametric prior models for the random error density $f_p(\cdot)$ (Kottas & Krnjajić, 2009)
Bayesian semiparametric quantile regression

- **Classical nonparametric approaches:** point estimates for quantile regression coefficients $\beta$ through optimization, $\min \sum_{i=1}^{n} \rho_p(y_i - x_i' \beta)$, where $\rho_p(u) = up - u1_{(u<0)}$ is the *check function* — the least absolute deviations criterion, $\min \sum_{i=1}^{n} |y_i - x_i' \beta|$, is obtained for $p = 0.5$

(Any inference beyond point estimation relies on asymptotics or resampling techniques)

- **Parametric modeling:** specifies parametrically the error distribution → e.g., asymmetric Laplace (AL) distribution (model $M_0$):

\[
\varepsilon_i \overset{iid}{\sim} k_p^{AL}(\varepsilon; \sigma) = \frac{p(1-p)}{\sigma} \exp\left\{ -\rho_p\left( \frac{\varepsilon}{\sigma} \right) \right\}
\]

with $\int_{-\infty}^{0} k_p^{AL}(\varepsilon; \sigma)d\varepsilon = p$

- Limitation: one parameter $p$ determines both quantile and skewness ($p > 0.5$ left skewed, $p = 0.5$ symmetric, $p < 0.5$ right skewed) — for example, the error distribution is symmetric in the median regression case
DP mixture models for the quantile regression error density

- **Model** $\mathcal{M}_1$: general scale mixture of asymmetric Laplace densities

\[ f_p^1(\varepsilon; G) = \int k_p^{AL}(\varepsilon; \sigma) dG(\sigma), \quad G \sim \text{DP}(\alpha, G_0) \]

\[ \rightarrow \text{captures more flexible tail behavior (mixing preserves quantiles, } \int_{-\infty}^{0} f_p^1(\varepsilon; G) d\varepsilon = p) \]

- $\mathcal{M}_1$ extends $\mathcal{M}_0$ with regard to tail behavior, but the skewness of the mixture $f_p^1(\cdot; G)$ suffers the same limitation as the kernel $k_p^{AL}(\cdot; \sigma)$
Bayesian semiparametric quantile regression

- In going beyond model $\mathcal{M}_1$, the key result is a representation theorem for non-increasing densities on $\mathbb{R}^+$: For any non-increasing density $f(\cdot)$ on $\mathbb{R}^+$ there exists a distribution function $G$, with support on $\mathbb{R}^+$, such that
  \[ f(t; G) = \int \theta^{-1} 1_{[0, \theta)}(t) dG(\theta) \]

- This result leads to a mixture representation for any unimodal density on the real line with $p$-th quantile (and mode) equal to zero,
  \[ \int \int k_p(\varepsilon; \sigma_1, \sigma_2) dG_1(\sigma_1) dG_2(\sigma_2), \]
  with $G_1$ and $G_2$ supported by $\mathbb{R}^+$, and
  \[ k_p(\varepsilon; \sigma_1, \sigma_2) = \frac{p}{\sigma_1} 1_{(-\sigma_1, 0)}(\varepsilon) + \frac{(1-p)}{\sigma_2} 1_{[0, \sigma_2)}(\varepsilon), \]
  with $0 < p < 1$, $\sigma_r > 0$, $r = 1, 2$.
Bayesian semiparametric quantile regression

- Assuming independent DP priors for $G_1$ and $G_2$, we obtain model $\mathcal{M}_2$:

$$f^2_p(\varepsilon; G_1, G_2) = \int \int k_p(\varepsilon; \sigma_1, \sigma_2) dG_1(\sigma_1) dG_2(\sigma_2), \ G_r \sim \text{DP}(\alpha_r, G_{r0}), \ r = 1, 2$$

→ model $\mathcal{M}_2$ can capture general forms of skewness and tail behavior

- The full hierarchical model $\mathcal{M}_2$:

$$Y_i \mid \beta, \sigma_{1i}, \sigma_{2i} \sim k_p(y_i - x_i'\beta; \sigma_{1i}, \sigma_{2i}), \ i = 1, \ldots, n$$

$$\sigma_{ri} \mid G_r \sim G_r, \ r = 1, 2, \ i = 1, \ldots, n$$

$$G_r \mid \alpha_r, d_r \sim \text{DP}(\alpha_r, G_{r0} = \text{IGamma}(c_r, d_r)), \ r = 1, 2$$

- Posterior inference under all models is obtained using straightforward extensions of methods for DP mixture models
Bayesian semiparametric quantile regression

Data Illustrations

- Simulated data \((n = 250\) in each case) from distributions with a specific quantile fixed at 0 (no covariates) and with varying shapes
  - three standard Laplace distributions \((\sigma = 1)\) for three values of \(p\) \((p = 0.5, 0.9,\) and 0.1\)
  - a standard normal distribution, and two mixtures of normals, one with 0.6-th quantile at zero and another with median zero (the components for both normal mixtures are chosen so that the resulting mixture densities are right skewed with non-standard tail behavior)

- Small cell lung cancer data: survival times in days for 121 patients with small cell lung cancer; 23 survival times are right censored
  - each patient was randomly assigned to one of two treatments A and B, achieving 62 and 59 patients, respectively (treatment indicator is the covariate)
Figure 6: Simulation study. Prior and posterior predictive densities (denoted by dotted and dashed lines respectively) under model $M_2$. The solid lines denote the true densities; the histograms of the data are also included.
Bayesian semiparametric quantile regression

Figure 7: Small cell lung cancer data. The top row includes posterior predictive densities (left panel) and survival functions (right panel) under treatments A and B. The middle row displays the posteriors for the 25th percentile survival time (left panel) and for the median survival time (right panel). The bottom row shows the posteriors for the 75th and the 90th percentile survival times (left and right panels, respectively). All the results are based on model $M_2$. In each panel the solid and dashed lines represent treatments A and B, respectively.
Figure 8: Small cell lung cancer data. The top panel displays posterior predictive densities for treatment A under model $\mathcal{M}_0$ (solid line), model $\mathcal{M}_2$ (dashed line), and a parametric Weibull model (dotted line). The bottom panels include CPO plots for the uncensored and censored data (left and right panel, respectively). In both cases, “o” denotes points under model $\mathcal{M}_0$, “+” under model $\mathcal{M}_2$, and “w” under the Weibull model.
Bayesian semiparametric quantile regression

Quantile regression with dependent error densities

- **Motivation**: Under the previous setting, the distribution of $\varepsilon_i$ is the same for all $x_i$, and thus, the distribution of $y_i$ changes with $x_i$ only through the $p$-th quantile $x'_i/\beta$

- To model nonparametrically error distribution that changes with covariates, we need a prior model for $f_{p,x}(\cdot) = \{f_{p,x}(\cdot) : x \in X\}$, where $X$ is the covariate space and $\forall x, \int_{-\infty}^{0} f_{p,x}(\varepsilon) d\varepsilon = p$.

- For example, under model $M_2$, to allow $f^2_p(\varepsilon; G_1, G_2)$ to change with $x$, the mixing distributions $G_1, G_2$ need to change with $x$ — we need to replace $G_r$ with a stochastic process $G_{r,X}$ over $X$

- Dependent DP (DDP) priors can be used for $G_{r,X}$ (refer to the course topic on dependent nonparametric priors) — briefly, the idea is to use the constructive definition of the DP where now the point masses are i.i.d. realizations from a base stochastic process (say a Gaussian process working with $\log(\sigma_{r_i})$)

- A critical advantage of the DDP model is its flexibility in capturing different shapes for different covariate values (both observed and unobserved covariate values)
Bayesian semiparametric quantile regression

Figure 9: Simulation experiment for the DDP quantile regression model. Posterior predictive densities under the DDP model (dashed lines) at the five observed covariate values (overlaid on histograms of the corresponding response observations), and at two new covariate values, $x = 10$ and $x = 95$ (overlaid on corresponding true densities denoted by solid lines).
4. Curve fitting using Dirichlet process mixtures

- Two dominant trends in the Bayesian regression literature: seek increasingly flexible regression function models, and accompany these models with more comprehensive uncertainty quantification.

- Typically, Bayesian nonparametric modeling focuses on either the regression function or the error distribution.

- Bayesian nonparametric extension of *implied conditional regression* (West et al., 1994; Müller et al., 1996)
  - use flexible nonparametric mixture model for the joint distribution of response and covariates
  - inference for the conditional response distribution given covariates

- Both the response distribution and, implicitly, the regression relationship are modeled nonparametrically, thus providing a flexible framework for the general regression problem.
Curve fitting using Dirichlet process mixtures

- Focus on univariate continuous response $y$ (though extensions for categorical and/or multivariate responses also possible)

- DP mixture model for the joint density $f(y, x)$ of the response $y$ and the vector of covariates $x$:

  $$f(y, x) \equiv f(y, x; G) = \int k(y, x; \theta)dG(\theta), \quad G \sim \text{DP}(\alpha, G_0(\psi))$$

- For the mixture kernel $k(y, x; \theta)$ use:
  - multivariate normal for (real-valued) continuous response and covariates
  - mixed continuous/discrete distribution to incorporate both categorical and continuous covariates
  - kernel component for $y$ supported by $\mathbb{R}^+$ for problems in survival/reliability analysis
Curve fitting using Dirichlet process mixtures

- Again, introduce latent mixing parameters \( \theta = \{ \theta_i : i = 1, ..., n \} \) for each response/covariate observation \( (y_i, x_i), i = 1, ..., n \) — full posterior:

  \[
p(G, \theta, \alpha, \psi \mid \text{data}) = p(G \mid \theta, \alpha, \psi)p(\theta, \alpha, \psi \mid \text{data})
\]

- \( p(\theta, \alpha, \psi \mid \text{data}) \) is the posterior of the finite-dimensional parameter vector that results by marginalizing \( G \) over its DP prior
  → MCMC posterior simulation to sample from this marginal posterior

- \( p(G \mid \theta, \alpha, \psi) \) is a DP with precision parameter \( \alpha + n \) and mean
  \[
  (\alpha + n)^{-1} \left\{ \alpha G_0(\cdot; \psi) + \sum_{j=1}^{n^*} n_j \delta_{\theta_j^*}(\cdot) \right\},
  \]
  where \( n^* \) is the number of distinct \( \theta_i \), and \( n_j \) is the size of the \( j \)-th distinct component
  → sample from the posterior distribution of \( G \) using one of the techniques discussed earlier (see Notes 2)

- Alternatively, \( G \) can be truncated from the outset resulting in a finite mixture model that can be fitted with blocked Gibbs sampling
Curve fitting using Dirichlet process mixtures

- For any grid of values \((y_0, x_0)\), obtain posterior samples for:
  - joint density \(f(y_0, x_0; G)\), marginal density \(f(x_0; G)\), and therefore, conditional density \(f(y_0 | x_0; G)\)

  \(\rightarrow\) conditional expectation \(E(y | x_0; G)\), which, estimated over grid in \(x\), provides inference for the regression relationship

  \(\rightarrow\) conditioning in \(f(y_0 | x_0; G)\) and/or \(E(y | x_0; G)\) may involve only a portion of vector \(x\)

- **Key features** of the modeling approach:
  - full and exact nonparametric inference (no need for asymptotics)
  - model for both non-linear regression curves and non-standard shapes for the conditional response density

  \(\rightarrow\) model does not rely on additive regression formulations; it can uncover interactions between covariates that might influence the regression relationship
Curve fitting using Dirichlet process mixtures

Data Example

- Simulated data set with a continuous response $y$, one continuous covariate $x_c$, and one binary categorical covariate $x_d$
  - $x_{ci} \mid x_{ci} \text{ ind. } N(0, 1)$
  - $x_{di} \mid x_{ci} \text{ ind. } \text{Bernoulli(probit}(x_{ci}))$
  - $y_i \mid x_{ci}, x_{di} \text{ ind. } N(h(x_{ci}), \sigma_{x_{di}}), \text{ with } \sigma_0 = 0.25, \sigma_1 = 0.5, \text{ and }$
    $$h(x_c) = 0.4x_c + 0.5\sin(2.7x_c) + 1.1(1 + x_c^2)^{-1}$$
  - two sample sizes: $n = 200$ and $n = 2000$

- DP mixture model with a mixed normal/Bernoulli kernel:
  $$f(y, x_c, x_d; G) = \int N_2(y, x_c; \mu, \Sigma) \pi^{x_d}(1 - \pi)^{1-x_d} \ dG(\mu, \Sigma, \pi),$$

  with $G \sim \text{DP}(\alpha, G_0(\mu, \Sigma, \pi) = N_2(\mu; m, V) \times \text{IWish}(\Sigma; \nu, S) \times \text{Beta}(\pi; a, b))$
Figure 10: Posterior point and 90% interval estimates (dashed and dotted lines) for conditional response expectation $E(y \mid x_c, x_d = 0; G)$ (left panels), $E(y \mid x_c, x_d = 1; G)$ (middle panels), and $E(y \mid x_c; G)$ (right panels). The corresponding data is plotted in grey for the sample of size $n = 200$ (top panels) and $n = 2000$ (bottom panels). The solid line denotes the true regression curve.
Model-based nonparametric approach to **quantile regression**
(Taddy & Kottas, 2010)

\[ f(y, x; G) = \int N_{M+1}(y, x; \mu, \Sigma)dG(\mu, \Sigma), \quad G \sim \text{DP}(\alpha, G_0) \]

with \( G_0(\mu, \Sigma) = N_{M+1}(\mu; m, V) \times \text{IWish}(\Sigma; \nu, S) \)

- For any grid of values \((y_0, x_0)\), obtain posterior samples for:
  - conditional density \( f(y_0 \mid x_0; G) \) and conditional cdf \( F(y_0 \mid x_0; G) \)
  - conditional quantile regression \( q_p(x_0; G) \), for any \( 0 < p < 1 \)

- Key features: modeling framework enables simultaneous inference for more than one quantile regression; model allows flexible response distributions **and** non-linear quantile regression relationships
Data Example

- Moral hazard data on the relationship between shareholder concentration and several indices for managerial moral hazard in the form of expenditure with scope for private benefit (Yafeh & Yoshua, 2003)

→ data set includes a variety of variables describing 185 Japanese industrial chemical firms listed on the Tokyo stock exchange

→ response $y$: index $MH5$, consisting of general sales and administrative expenses deflated by sales

→ four-dimensional covariate vector $x$: Leverage (ratio of debt to total assets); log($Assets$); Age of the firm; and $TOPTEN$ (the percent of ownership held by the ten largest shareholders)
Curve fitting using Dirichlet process mixtures

Marginal Average Medians with 90% CI

Figure 11: Posterior mean and 90% interval estimates for median regression for \( MH5 \) conditional on each individual covariate. Data scatterplots are shown in grey.
Curve fitting using Dirichlet process mixtures

Marginal Average 90th Percentiles with 90% CI

Figure 12: Posterior mean and 90% interval estimates for 90th percentile regression for MH5 conditional on each individual covariate. Data scatterplots are shown in grey.
Figure 13: Posterior estimates of median surfaces (left column) and 90th percentile surfaces (right column) for MH5 conditional on Leverage and TOPTEN. The posterior mean is shown on the top row and the posterior interquartile range on the bottom.
Figure 14: Posterior mean and 90% interval estimates for response densities $f(y \mid \mathbf{x}_0; G)$ conditional on four combinations of values $\mathbf{x}_0$ for the covariate vector ($TOPTEN$, $Leverage$, $Age$, log($Assets$))
Further Applications

- Nonparametric classification (Shahbaba & Neal, 2009; Dunson & Bhat-tacharya, 2010); nonparametric functional data analysis (Rodriguez et al., 2009)

- Semiparametric switching regression (Taddy & Kottas, 2009a)

- Inference for stock-recruitment relationships (Fronczyk et al., 2009)

- Nonparametric binary regression, with an application to inference for fitness functions in natural selection studies (with Maria DeYoreo)

- Nonparametric regression for censored survival data (with Valerie Poynor)

- Emulation and calibration for stochastic computer simulators (Ph.D. thesis work for Marian Farah)
5. Modeling for multivariate ordinal data

- Values of $k$ ordinal categorical variables $V_1, \ldots, V_k$ recorded for $n$ subjects
  \[ C_j \geq 2: \text{number of categories for the } j\text{-th variable, } j = 1, \ldots, k \]
  \[ n_{\ell_1 \ldots \ell_k}: \text{number of observations with } V = (V_1, \ldots, V_k) = (\ell_1, \ldots, \ell_k) \]
  \[ p_{\ell_1 \ldots \ell_k} = \Pr(V_1 = \ell_1, \ldots, V_k = \ell_k) \text{ is the classification probability for} \]
  \[ \text{the } (\ell_1, \ldots, \ell_k) \text{ cell} \]

- The data can be summarized in a $k$-dimensional contingency table with
  \[ C = \prod_{j=1}^{k} C_j \text{ cells, with frequencies}\{n_{\ell_1 \ldots \ell_k}\} \text{ constrained by}\]
  \[ \sum_{\ell_1 \ldots \ell_k} n_{\ell_1 \ldots \ell_k} = n \]
Modeling for multivariate ordinal data

- A possible modeling strategy (alternative to log-linear models) involves the introduction of $k$ continuous latent variables $Z = (Z_1, \ldots, Z_k)$ whose joint distribution yields the classification probabilities for the table cells, i.e.,

$$p_{\ell_1 \ldots \ell_k} = \Pr \left( \bigcap_{j=1}^{k} \{ \gamma_{j,\ell_{j}-1} < Z_j \leq \gamma_{j,\ell_j} \} \right)$$

for cutoff points $-\infty = \gamma_{j,0} < \gamma_{j,1} < \cdots < \gamma_{j,c_j-1} < \gamma_{j,c_j} = \infty$, for each $j = 1, \ldots, k$ (e.g., Johnson and Albert, 1999)

- Common distributional assumption: $Z \sim N_k(0, S)$ (probit model)

$\rightarrow \rho_{st} = \text{Corr}(Z_s, Z_t) = 0, s \neq t$, implies independence of the corresponding categorical variables

$\rightarrow$ coefficients $\rho_{st}, s \neq t$: polychoric correlation coefficients (traditionally used in the social sciences as a measure of association)
Modeling for multivariate ordinal data

- Richer modeling and inference based on normal DP mixtures for the latent variables \( Z_i \) associated with data vectors \( V_i, i = 1, \ldots, n \) (Kottas et al., 2005)

- Model \( Z_i \mid G \) i.i.d. \( f \), with \( f(\cdot; G) = \int N_k(\cdot; m, S)dG(m, S) \), where

\[
G \mid \alpha, \lambda, \Sigma, D \sim \text{DP}(\alpha, G_0(m, S) = N_k(m|\lambda, \Sigma)\text{IWarn}_k(S|\nu, D))
\]

- Advantages of the DP mixture modeling approach:
  \( \to \) can accommodate essentially any pattern in \( k \)-dimensional contingency tables
  \( \to \) allows local dependence structure to vary accross the contingency table
  \( \to \) implementation does not require random cutoffs (so the complex updating mechanisms for cutoffs are not needed)
Data Example

- A Data Set of *Interrater Agreement*: data on the extent of scleral extension (extent to which a tumor has invaded the sclera or “white of the eye”) as coded by two raters for each of \( n = 885 \) eyes

- The coding scheme uses five categories: 1 for “none or innermost layers”; 2 for “within sclera, but does not extend to scleral surface”; 3 for “extends to scleral surface”; 4 for “extrascleral extension without transection”; and 5 for “extrascleral extension with presumed residual tumor in the orbit”

- Results under the DP mixture model (and, for comparison, using also a probit model)

- The \((0.25, 0.5, 0.75)\) posterior percentiles for \( n^* \) are \((6, 7, 8) \) – in fact, \( \Pr(n^* \geq 4 \mid \text{data}) = 1 \)
Table 1: For the interrater agreement data, observed cell relative frequencies (in bold) and posterior summaries for table cell probabilities (posterior mean and 95% central posterior intervals). Rows correspond to rater A and columns to rater B.

<table>
<thead>
<tr>
<th></th>
<th>.3288</th>
<th>.0836</th>
<th>.0011</th>
<th>.0011</th>
<th>.0011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.3264</td>
<td>.0872</td>
<td>.0013</td>
<td>.0020</td>
<td>.0008</td>
</tr>
<tr>
<td>(.2940, .3586)</td>
<td>(.0696, .1062)</td>
<td>(.0002, .0041)</td>
<td>(.0003, .0055)</td>
<td>(.0, .0033)</td>
<td></td>
</tr>
<tr>
<td>.2102</td>
<td>.2136</td>
<td>.2893</td>
<td>.2817</td>
<td>.0079</td>
<td>.0079</td>
</tr>
<tr>
<td>(.1867, .2404)</td>
<td>(.2524, .3112)</td>
<td>(.0033, .0146)</td>
<td>(.0022, .0143)</td>
<td>(.0006, .0074)</td>
<td></td>
</tr>
<tr>
<td>.0023</td>
<td>.0021</td>
<td>.0045</td>
<td>.0060</td>
<td>.0 .0016</td>
<td>.0023</td>
</tr>
<tr>
<td>(.0004, .0059)</td>
<td>(.0021, .0118)</td>
<td>(.0004, .0037)</td>
<td>(.0004, .0059)</td>
<td>(.0, .0030)</td>
<td></td>
</tr>
<tr>
<td>.0034</td>
<td>.0043</td>
<td>.0113</td>
<td>.0101</td>
<td>.0011</td>
<td>.0158</td>
</tr>
<tr>
<td>(.0124, .0094)</td>
<td>(.0041, .0185)</td>
<td>(.0004, .0058)</td>
<td>(.0069, .0238)</td>
<td>(.0006, .0066)</td>
<td></td>
</tr>
<tr>
<td>.0011</td>
<td>.0013</td>
<td>.0079</td>
<td>.0071</td>
<td>.0011</td>
<td>.0090</td>
</tr>
<tr>
<td>(.0001, .0044)</td>
<td>(.0026, .0140)</td>
<td>(.0003, .0054)</td>
<td>(.0033, .0159)</td>
<td>(.0011, .0090)</td>
<td></td>
</tr>
</tbody>
</table>
Modeling for multivariate ordinal data

- Posterior predictive distributions \( p(Z_0 | \text{data}) \) (Figure 15) – DP mixture version is based on the posterior predictive distribution for corresponding mixing parameter \((m_0, S_0)\)

- Inference for the association between the ordinal variables:
  \( \rightarrow \) e.g., Figure 15 shows the posterior density for \( \rho_0 \), the correlation coefficient implied by \( S_0 \)
  \( \rightarrow \) the probit model underestimates the association of the ordinal variables (as measured by \( \rho_0 \)), since it fails to recognize the clustering structure suggested by the data (as the DP mixture model reveals)

- Inference for log-odds ratios, \( \psi_{ij} = \log p_{i,j} + \log p_{i+1,j+1} - \log p_{i,j+1} - \log p_{i+1,j} \) (Figure 16)

- Utility of mixture modeling for this data example – one of the clusters dominates the others, but identifying the other three is important; one of them corresponds to agreement for large values in the coding scheme; the other two indicate regions of the table where the two raters tend to agree less strongly
Modeling for multivariate ordinal data

Figure 15: For the interrater agreement data, draws from $p(Z_0|\text{data})$ and $p(\rho_0|\text{data})$ under the DP mixture model (panels (a) and (c), respectively) and the probit model (panels (b) and (d), respectively).
Figure 16: For the interrater agreement data, posteriors for four log-odds ratios under the DP mixture model (solid lines) and the probit model (dashed lines). The circles denote the corresponding empirical log-odds ratios.
6. Nonparametric inference for Poisson processes

- Point processes are stochastic process models for events that occur separated in time or space

- Applications of point process modeling in traffic engineering, software reliability, neurophysiology, weather modeling, forestry, ...

- Poisson processes, along with their extensions (Poisson cluster processes, marked Poisson processes, etc.), play an important role in the theory and applications of point processes (e.g., Kingman, 1993; Guttorp, 1995; Moller & Waagepetersen, 2004)

- Existing Bayesian nonparametric work based on gamma processes, weighted gamma processes, and Lévy processes (e.g., Lo & Weng, 1989; Kuo & Ghosh, 1997; Wolpert & Ickstadt, 1998; Gutiérrez-Peña & Nieto-Barajas, 2003; Ishwaran & James, 2004)
Nonparametric inference for Poisson processes

- For a point process over time, let \( N(t) \) be the number of event occurrences in the time interval \((0, t]\) – the point process \( \mathcal{N} = \{N(t) : t \geq 0\} \) is a non-homogeneous Poisson process (NHPP) if:
  - for any \( t > s \geq 0 \), \( N(t) - N(s) \) follows a Poisson distribution with mean \( \Lambda(t) - \Lambda(s) \), and
  - \( \mathcal{N} \) has independent increments, i.e., for any \( 0 \leq t_1 < t_2 \leq t_3 < t_4 \), \( N(t_2) - N(t_1) \) and \( N(t_4) - N(t_3) \) are independent random variables.

- \( \Lambda \) is the mean measure (or cumulative intensity function) of the NHPP.

- For any \( t \in \mathbb{R}^+ \), \( \Lambda(t) = \int_0^t \lambda(u)du \), where \( \lambda \) is the NHPP intensity function.
  - \( \lambda \) is a non-negative and locally integrable function (i.e., \( \int_B \lambda(u)du < \infty \), for all bounded \( B \subset \mathbb{R}^+ \)).

- So, from a modeling perspective, of interest for a NHPP is its intensity function.
Nonparametric inference for Poisson processes

- Consider a NHPP observed over the time interval \((0, T]\) with events that occur at times \(0 < t_1 < t_2 < ... < t_n \leq T\)
- The likelihood for the NHPP intensity function \(\lambda\) is proportional to
\[
\exp\left\{ - \int_0^T \lambda(u) du \right\} \prod_{i=1}^{n} \lambda(t_i)
\]

- **Key observation:** \(f(t) = \frac{\lambda(t)}{\gamma}\), where \(\gamma = \int_0^T \lambda(u) du\), is a density function on \((0, T)\)
- Hence, \((f, \gamma)\) provides an equivalent representation for \(\lambda\), and so a nonparametric prior model for \(f\), with a parametric prior for \(\gamma\), will induce a semiparametric prior for \(\lambda\) — in fact, since \(\gamma\) only scales \(\lambda\), it is \(f\) that determines the shape of the intensity function \(\lambda\)
Nonparametric inference for Poisson processes

- **Beta DP mixture model** for $f$

  $$f(t) \equiv f(t; G) = \int \text{be}(t; \mu, \tau) dG(\mu, \tau), \quad G \sim \text{DP}(\alpha, G_0)$$

  where $\text{be}(t; \mu, \tau)$ is the Beta density on $(0, T)$ with mean $\mu \in (0, T)$ and scale parameter $\tau > 0$, and $G_0(\mu, \tau) = \text{Unif}(\mu; 0, T) \text{ inv-gamma} (\tau; c, \beta)$ with fixed shape parameter $c$ and random scale $\beta$ (Kottas, 2006a)

- Flexible density shapes through mixing of Betas (e.g., Diaconis & Ylvisaker, 1985) – Beta mixture model avoids edge effects (the main drawback of a normal DP mixture model in this setting)

- Full Bayesian model:

  $$\exp(-\gamma) \gamma^n \left\{ \prod_{i=1}^{n} \int \text{be}(t_i; \mu_i, \tau_i) dG(\mu_i, \tau_i) \right\} p(\gamma)p(G \mid \alpha, \beta)p(\alpha)p(\beta)$$

  $\rightarrow$ DP prior structure $p(G \mid \alpha, \beta)p(\alpha)p(\beta)$ for $G$ and its hyperparameters

  $\rightarrow$ reference prior for $\gamma$, $p(\gamma) \propto \gamma^{-1}$
Letting $\theta = \{(\mu_i, \tau_i) : i = 1, ..., n\}$, we have

$$p(\gamma, G, \theta, \alpha, \beta \mid data) = p(\gamma \mid data)p(G \mid \theta, \alpha, \beta)p(\theta, \alpha, \beta \mid data)$$

$\rightarrow p(\gamma \mid data)$ is a gamma($n, 1$) distribution

$\rightarrow$ MCMC with Metropolis steps to sample from $p(\theta, \alpha, \beta \mid data)$

$\rightarrow p(G \mid \theta, \alpha, \beta)$ is a DP with updated parameters (can be sampled using the methods discussed in Notes 2)

• Full posterior inference for $\lambda$, $\Lambda$, and any other functional of the NHPP

• Extensions to inference for spatial NHPP intensities, using DP mixtures with bivariate Beta kernels (Kottas & Sansó, 2007)
Data Illustrations

- Example for temporal NHPPs: data on the times of 191 explosions in mines, leading to coal-mining disasters with 10 or more men killed, over a time period of 40,550 days, from 15 March 1851 to 22 March 1962

- Prior specification for $\text{DP}(\alpha, G_0(\mu, \tau|\beta) = \text{Unif}(\mu|0, T)\text{IG}(\tau|2, \beta))$
  $\rightarrow \text{gamma}(a_\alpha, b_\alpha)$ prior for $\alpha$ – recall the role of $\alpha$ in controlling the number $n^*$ of distinct components in the DP mixture model
  $\rightarrow \text{exponential}$ prior for $\beta$ – its mean can be specified using a prior guess at the range, $R$, of the event times $t_i$ (e.g., $R = T$ is a natural default choice)

- Inference for the NHPP intensity under three prior choices: priors for $\beta$ and $\alpha$ based on $R = T$, $E(n^*) \approx 7$; $R = T$, $E(n^*) \approx 15$; and $R = 1.5T$, $E(n^*) \approx 7$

- Examples for spatial NHPPs: two forestry data sets
Nonparametric inference for Poisson processes

Figure 17: Coal-mining disasters data. Posterior point and 95% pointwise interval estimates for the intensity function under three prior settings. The observed times of the 191 explosions in mines are plotted on the horizontal axis.
Figure 18: Redwood seedlings data. Contour plots of posterior mean intensity estimates under two different priors for $\alpha$. The dots indicate the locations of the redwood seedlings.
Figure 19: Maple data. Panels (a) and (b) include the posterior mean intensity estimate (contour plot and perspective plot, respectively). Panels (c) and (d) plot contour plots for the posterior median and posterior interquartile range intensity estimates, respectively. The dots denote the locations of the maple trees.
Application to neuronal data analysis

- One of the key techniques in neuroscience involves recording of electrical activity of neurons in laboratory animals.

- The technique studies action potentials (spikes) generated by the neuron and measured using an electrode inserted into the animal’s brain.

- The firing times (times at which spikes occur) are recorded to provide the neuronal data.

- In this context, the focus of statistical modeling approaches is on the temporal evolution of the neuronal firing activity.

- Bayesian nonparametric modeling for neuronal data arising as firing times from a single neuron under two distinct experimental conditions (Kottas & Behseta, 2010).
Motivating neurophysiological study: neurons recorded from the primary motor cortex area (M1) of a Macaque monkey’s brain while performing the sequential task of reaching a series of illuminating targets on a touch-sensitive screen

The animal was trained to respond to the visual stimuli under two conditions
→ repeating condition: a sequence of targets would appear on the screen in a repeating order
→ random condition: the stimuli were sent in a pseudo-random order

Two neurons (29 and 32):
→ for neuron 29, 21 and 32 trials were recorded under the random and repeating mode, respectively, resulting in a total of 108 firing times for the random condition and 224 for the repeating condition
→ in both conditions, 20 trials were recorded for neuron 32, achieving a total of 52 and 102 firing times under the random and repeating condition
Figure 20: Raster and PSTH plots for the firing times under neurons 32 and 29 (top and bottom row, respectively). The left panels correspond to the random condition, and the right panels to the repeating condition.
Nonparametric inference for Poisson processes

- Data from each neuron can be represented in their most general form through vectors \( \{y_{ij}^{(\ell)} : i = 1, \ldots, N^{(\ell)}; j = 1, \ldots, n_i^{(\ell)} \} \), where \( y_{ij}^{(\ell)} \) is the \( j \)-th firing time in the \( i \)-th trial under condition \( \ell = 1, 2 \).

- Given our inferential objective of comparison of firing intensities under the two conditions, it suffices to consider modeling for the firing times aggregated over all trials — for condition \( \ell = 1, 2 \), the data vector

\[
t^{(\ell)} = \{t_k^{(\ell)} : k = 1, \ldots, K^{(\ell)} \}
\]

where \( K^{(\ell)} = \sum_{i=1}^{N^{(\ell)}} n_i^{(\ell)} \) is the total number of firing times from all trials, and \( t_k^{(\ell)} \) is the \( k \)-th spike time in the aggregated set of firing times.

- NHPP a plausible model for the underlying point process of aggregated firing times (pooled point patterns across a large number of replicated trials follow approximately a NHPP model).
Nonparametric inference for Poisson processes

- NHPP with intensity function $\lambda^{(\ell)}(\cdot)$ under condition $\ell = 1, 2$
- Beta DP mixture model for firing intensities

$$
\lambda^{(\ell)}(t) \equiv \lambda^{(\ell)}(t; \gamma^{(\ell)}, G^{(\ell)}) = \gamma^{(\ell)} f^{(\ell)}(t; G^{(\ell)}) = \int be(t; \mu, \tau) \, dG^{(\ell)}(\mu, \tau)
$$
with (independent) DP priors for mixing distributions $G^{(1)}$ and $G^{(2)}$

- Direct comparison of intensities $\lambda^{(1)}(\cdot)$ and $\lambda^{(2)}(\cdot)$ is hindered by their different scales, $\gamma^{(1)}$ and $\gamma^{(2)}$
- Work instead with densities $f^{(1)}(\cdot)$ and $f^{(2)}(\cdot)$
  $\rightarrow$ posterior point and interval estimates for function $f^{(1)}(\cdot) - f^{(2)}(\cdot)$
  $\rightarrow$ entire posterior distribution $p(f^{(1)}(t_0; G^{(1)}) - f^{(2)}(t_0; G^{(2)}) \mid \text{data})$ for specific points $t_0$ in the experimental time interval

- Inference for local and global differences in the neuronal firing intensities
Figure 21: Neuron 32. Posterior mean and 95% interval estimates for the intensity functions. Posterior mean and 95% interval band for the difference of densities between the random and repeating condition.
Nonparametric inference for Poisson processes

Figure 22: Neuron 29. Posterior mean and 95% interval estimates for the intensity functions. Posterior mean and 95% interval band for the difference of densities between the random and repeating condition.
Nonparametric inference for Poisson processes

Figure 23: Neuron 29. Posterior distributions for the difference of density functions between the random and repeating condition at 12 time points.
Nonparametric inference for Poisson processes

Related recent and current work

- Dynamic modeling for spatial NHPPs (Taddy, 2010)

- Nonparametric modeling and inference for marked Poisson processes (Taddy and Kottas, 2009b)

- Analysis of extremes of a process observed over time using the threshold approach — based on the bivariate point pattern comprising the times of exceedances and the excess values
  - DP mixture modeling for the corresponding NHPP intensity using more structured kernels
  - application to inference for environmental extremes using spatial DP mixture models for temporal NHPP intensities (Ph.D. thesis work for Ziwei Wang)
7. Modeling for stochastically ordered distributions

- *Probability order* restrictions often appropriate/desirable when comparing two or more populations — different types of probability orders

- $\mathbb{R}$-valued random variables $Y_1, Y_2$ with respective distribution functions $F_1, F_2$, density functions $f_1, f_2$, and hazard functions $h_1, h_2$

- **Stochastic order**: $Y_1 \leq_{st} Y_2$ (or $F_1 \leq_{st} F_2$) if, by definition,

  \[ F_1(u) \geq F_2(u), \quad \forall u \in \mathbb{R} \quad \Leftrightarrow \quad \Pr(Y_1 > u) \leq \Pr(Y_2 > u), \quad \forall u \in \mathbb{R} \]

  → characterization: $Y_1 \leq_{st} Y_2$ if-f there exist r.v.s $Y_1'$ and $Y_2'$, defined on the same probability space, such that $Y_1$ and $Y_2$ have the same distribution with $Y_1'$ and $Y_2'$, and $\Pr(Y_1' \leq Y_2') = 1$
Modeling for stochastically ordered distributions

- **Hazard rate order**: For $\mathbb{R}^+$-valued r.v.s $Y_1, Y_2, Y_1 \leq_{hr} Y_2$ if, by definition, the function $(1 - F_1(t))/(1 - F_2(t))$ decreases in $t$
  → equivalent definition for continuous r.v.s $Y_1, Y_2$:
  $Y_1 \leq_{hr} Y_2$ if-f $h_1(t) \geq h_2(t), \forall t \in \mathbb{R}^+$
  → stronger restriction than stochastic order ($Y_1 \leq_{hr} Y_2$ implies $Y_1 \leq_{st} Y_2$)

- **Likelihood ratio order**: $Y_1 \leq_{lr} Y_2$ if, by definition, the function $f_1(u)/f_2(u)$ decreases over the union of the supports of $Y_1$ and $Y_2$
  → stronger restriction than hazard rate order ($Y_1 \leq_{lr} Y_2$ implies $Y_1 \leq_{hr} Y_2$)

- **Stochastic precedence order**: $Y_1 \leq_{sp} Y_2$ if, by definition,
  $\Pr(Y_1 \leq Y_2) \geq 0.5$
  → weaker restriction than stochastic order ($Y_1 \leq_{st} Y_2$ implies $Y_1 \leq_{sp} Y_2$)
Modeling for stochastically ordered distributions

- Substantial literature on properties of distributions ordered according to one of these orders as well as on several other probability orders (Shaked & Shanthikumar, 1994) — also, extensive literature on classical estimation (typically, maximum likelihood estimation) and distribution-free testing for stochastic order, hazard rate order, and likelihood ratio order.

- Arguments for forcing order restriction in the model:
  → order constraint of interest may not hold for the empirical distribution functions (especially for small or moderate sample sizes)
  → incorporating the order restriction can improve predictive accuracy
  → Bayesian framework attractive, since any order restriction in the prior model for the distributions is preserved to the posterior analysis

- **Bayesian nonparametric work:** stochastic and partial stochastic orders (Arjas & Gasbarra, 1996; Gelfand & Kottas, 2001; Hoff, 2003; Karabatsos & Walker, 2007; Dunson & Peddada, 2008) — variability order (Kottas & Gelfand, 2001a) — stochastic precedence order (Chen & Dunson, 2004; Kottas, 2010)
Modeling for stochastically ordered distributions

A mixture modeling approach for stochastic order

- Focusing on two stochastically ordered distribution functions $F_1$ and $F_2$ (corresponding to distributions supported on $\mathbb{R}$), we seek nonparametric prior models over the space

$$\mathcal{P} = \{(F_1, F_2) : F_1 \leq_{st} F_2\}$$

- Constructive approach to building the restriction $F_1(u) \geq F_2(u)$, $u \in \mathbb{R}$, through latent distribution functions $G_1$ and $G_2$ (on $\mathbb{R}$) such that

$$F_1(u) = G_1(u), \quad F_2(u) = G_1(u)G_2(u)$$

(Note: with $\theta \sim G_1$ and independently $\delta \sim G_2$, $F_1$ and $F_2$ are the distributions of $\theta$ and $\max\{\theta, \delta\}$, respectively)

- Work with (independent) nonparametric priors for $G_1$ and $G_2$ to induce a prior over $\mathcal{P}' = \{(F_1, F_2) : F_1 = G_1, \ F_2 = G_1G_2\}$, and hence over $\mathcal{P}$
Modeling for stochastically ordered distributions

- How about using DP priors for $G_1$ and $G_2$
  → discreteness? simulation-based model fitting?

- Introduce DP mixing to overcome both difficulties

- **Key result:** for a parametric family of distributions $K(\cdot; \theta)$, $\theta \in (\underline{\theta}, \overline{\theta})$, strictly decreasing in $\theta$, and $H_1$, $H_2$ two distribution functions on $(\underline{\theta}, \overline{\theta})$ with $H_1 \leq_{st} H_2$, defining

$$F(\cdot; H_i) = \int_{\underline{\theta}}^{\overline{\theta}} K(\cdot; \theta) \, dH_i(\theta), \ i = 1, 2$$

we have $F(\cdot; H_1) \leq_{st} F(\cdot; H_2)$

→ result valid, e.g., for normal kernels with mixing on the mean

→ add a dispersion parameter $\sigma^2$ to the model, to conclude that

$F(\cdot; H_1, \sigma^2) \leq_{st} F(\cdot; H_2, \sigma^2)$ (semiparametric specification)
Setting $H_1 = G_1$ and $H_2 = G_1 G_2$, we obtain the stochastically ordered DP mixture of normals model:

$$F_1(\cdot) \equiv F(\cdot; G_1, \sigma^2) = \int N(\cdot; \theta, \sigma^2) \, dG_1(\theta)$$

$$F_2(\cdot) \equiv F(\cdot; G_1, G_2, \sigma^2) = \int \int N(\cdot; \max\{\theta, \delta\}, \sigma^2) \, dG_1(\theta) dG_2(\delta)$$

with independent $\text{DP}(\alpha_\ell, N(\mu_\ell, \tau_\ell^2))$ priors for $G_\ell$, $\ell = 1, 2$.

- Consider data $\{y_{1i} : i = 1, \ldots, n_1; y_{2j} : j = 1, \ldots, n_2\}$ where the $y_{1i}$ (given $G_1, \sigma^2$) are ind. from $F_1$ and the $y_{2j}$ (given $G_1, G_2, \sigma^2$) are ind. from $F_2(\cdot)$.
Modeling for stochastically ordered distributions

- Hierarchical formulation of the model:

  \[ y_{1i} \mid \theta_i, \sigma^2 \sim \text{ind.} \ N(\theta_i, \sigma^2), \ i = 1, \ldots, n_1 \]

  \[ y_{2j} \mid \theta_{n_1+j}, \delta_j, \sigma^2 \sim \text{ind.} \ N(\max\{\theta_{n_1+j}, \delta_j\}, \sigma^2), \ j = 1, \ldots, n_2 \]

  \[ \theta_i \mid G_1 \sim \text{i.i.d.} \ G_1, \ i = 1, \ldots, n_1 + n_2 \]

  \[ \delta_j \mid G_2 \sim \text{i.i.d.} \ G_2, \ j = 1, \ldots, n_2 \]

  \[ G_1, G_2 \mid \mu_1, \tau^2_1, \mu_2, \tau^2_2 \sim \text{DP(}\alpha_1, N(\mu_1, \tau^2_1)) \times \text{DP(}\alpha_2, N(\mu_2, \tau^2_2)) \]

- Through the introduction of the additional mixing parameters \( \theta_{n_1+j}, \ j = 1, \ldots, n_2 \), the first stage conditionally independent specification is retained after marginalizing \( G_1 \) and \( G_2 \) over their DP priors

- **Posterior inference**: simulation from the marginal posterior

  \[ p(\theta, \delta, \sigma^2, \psi \mid \text{data}), \text{ where } \theta = \{\theta_i : i = 1, \ldots, n_1 + n_2\} \text{ and } \delta = \{\delta_j : j = 1, \ldots, n_2\}, \text{ enables estimation of posterior predictive densities} \]
Modeling for stochastically ordered distributions

- More general inference requires the posteriors of $G_1$ and $G_2$:

$$p(G_1, G_2, \theta, \delta, \sigma^2, \psi \mid \text{data}) = p(G_1 \mid \theta, \mu_1, \tau_1^2)p(G_2 \mid \delta, \mu_2, \tau_2^2)p(\theta, \delta, \sigma^2, \psi \mid \text{data})$$

→ where $p(G_1 \mid \theta, \mu_1, \tau_1^2)$ denotes a DP distribution with precision parameter $\alpha_1 + n_1 + n_2$ and base distribution

$$\frac{\alpha_1}{\alpha_1 + n_1 + n_2} \text{N}(\cdot; \mu_1, \tau_1^2) + \frac{1}{\alpha_1 + n_1 + n_2} \sum_{i=1}^{n_1+n_2} \delta_{\theta_i}(\cdot)$$

with analogous expressions for $p(G_2 \mid \delta, \mu_2, \tau_2^2)$

- Sample from these two DPs using the DP stick-breaking representation with a truncation approximation

- Posterior samples for $G_1, G_2$ yield, for any set of grid points $u$, samples from the posterior of $F_1(u; G_1, \sigma^2)$ and $F_2(u; G_1, G_2, \sigma^2)$ (analogously for the mixture densities $f_1(u; G_1, \sigma^2)$ and $f_2(u; G_1, G_2, \sigma^2)$)
A commonly encountered task in epidemiologic research (both human and veterinary) involves the characterization of the discriminatory ability of a continuous diagnostic test.

In particular, *serologic scores* measure the concentration of antigen-specific antibodies in serum.

Commonly used continuous diagnostic measures result in an optical density value or a serum-to-positive ratio for an enzyme linked immunosorbent assay (ELISA) serological test — a relatively large serologic score is suggestive of disease or infection presence.

Data illustrations with commercially available ELISAs designed to detect antibodies to Johne’s disease in dairy cows — Johne’s disease is endemic throughout the US affecting multiple species of animals.
Modeling for stochastically ordered distributions

- *Gold-standard* data setting: disease (infection) status is assumed known

- $F_1$ and $F_2$ are the distribution functions associated with serologic scores for the noninfected and infected populations, respectively — typically, $F_1$ and $F_2$ are modeled independently

- Incorporate stochastic order constraint $F_1 \leq_{st} F_2$ (Hanson et al., 2008)

- Biologically such a constraint is essentially always appropriate because serologic values for infected individuals tend to be larger than serologic values for noninfected individuals (provided the diagnostic test has reasonable discriminatory ability)

- Employ the stochastically ordered DP mixture model $F_1(\cdot) = F(\cdot; G_1, \sigma^2); F_2(\cdot) = F(\cdot; G_1, G_2, \sigma^2)$
Modeling for stochastically ordered distributions

- Receiver operating characteristic (ROC) curve provides a commonly used graphical measure of the accuracy of the diagnostic test.

- A cutoff value $z$ can be used to dichotomize the serologic data into test positive (serologic score $> z$) or test negative (serologic score $< z$) categories.

- ROC curve plots all possible pairs of true positive probability of infection $(1 - F_2(z))$ versus false positive probability $(1 - F_1(z))$ across all cutoff values $z$.

- $\text{ROC}(u) = 1 - F_2(F_1^{-1}(1 - u)), \; u \in (0, 1)$.

- Area under the curve, $\text{AUC} = \int_0^1 \text{ROC}(u) \, du$ (probability that a randomly selected infected individual has a serologic score that is greater than that for a randomly selected noninfected individual).

- Posterior inference for $\text{ROC}(\cdot)$ and AUC through the posteriors of $F(\cdot; G_1, \sigma^2)$ and $F(\cdot; G_1, G_2, \sigma^2)$.
Figure 24: HerdChek ELISA test. Serologic scores for $n_1 = 393$ noninfected and $n_2 = 66$ infected cows.
Modeling for stochastically ordered distributions

![Graphs showing distributions and cdfs for non-diseased and diseased groups.](image)

**Figure 25:** Institut Pourquier ELISA test. Scores for $n_1 = 345$ noninfected and $n_2 = 258$ infected cows.
References


References


References


References


References


References


References


References


