AMS 241: Bayesian Nonparametric Methods (Fall 2015)

Homework set on Dirichlet process mixture models
(due Tuesday November 17)

1. Consider the location normal Dirichlet process (DP) mixture model

$$f(\cdot \mid G, \phi) = \int k_N(\cdot \mid \theta, \phi)\,\mathrm{d}G(\theta), \qquad G \mid \alpha, \mu, \tau^2 \sim \mathrm{DP}(\alpha, G_0 = \mathrm{N}(\mu, \tau^2)),$$

where $k_N(\cdot \mid \theta, \phi)$ denotes the density function of a normal distribution with mean θ and variance ϕ. Assume an inv-gamma(a_ϕ, b_ϕ) prior for ϕ, a gamma(a_α, b_α) prior for α, and take $\mathrm{N}(a_\mu, b_\mu)$ and inv-gamma(a_{τ^2}, b_{τ^2}) priors for the mean, μ, and variance, τ^2, respectively, of the normal centering distribution G_0. (Here, inv-gamma(a, b) denotes the inverse gamma distribution with mean $b/(a-1)$, provided $a > 1$, and gamma(a, b) denotes the gamma distribution with mean a/b.) Therefore, the hierarchical version of this semiparametric DP mixture model is given by

$$\begin{array}{ccc} y_i \mid \theta_i, \phi & \overset{ind.}{\sim} & k_N(y_i \mid \theta_i, \phi), \quad i = 1, ..., n \\ \theta_i \mid G & \overset{i.i.d.}{\sim} & G, \quad i = 1, ..., n \\ G \mid \alpha, \mu, \tau^2 & \sim & \mathrm{DP}(\alpha, G_0 = \mathrm{N}(\mu, \tau^2)) \\ \alpha, \mu, \tau^2, \phi & \sim & p(\alpha)p(\mu)p(\tau^2)p(\phi), \end{array}$$

with the (independent) priors $p(\alpha)$, $p(\mu)$, $p(\tau^2)$, $p(\phi)$ for α, μ, τ^2, ϕ given above.

To study inference under this model, consider a simulated data set (available from the course webpage, `https://ams241-fall15-01.courses.soe.ucsc.edu/node/7`) of size $n = 250$, generated from the mixture $0.2\mathrm{N}(-5, 1) + 0.5\mathrm{N}(0, 1) + 0.3\mathrm{N}(3.5, 1)$.

(1) Obtain the required expressions for the Pólya urn based Gibbs sampler, which can be used for posterior simulation from $p(\theta_1, ..., \theta_n, \alpha, \phi, \mu, \tau^2 \mid \text{data})$, where data $= \{y_i : i = 1, ..., n\}$.

(2) Discuss specification of the prior hyperparameters for ϕ, μ, and τ^2. Study sensitivity of posterior inference for ϕ, μ, and τ^2 to the prior choice. In addition to the posterior distributions for ϕ, μ, τ^2, examine sensitivity of posterior predictive inference (see (5) below).

(3) Obtain the posterior distributions for α and n^* under different prior choices for α (and hence for n^*) suggesting, a priori, an increasing number of distinct components for the mixture. For example, you can consider $a_\alpha = 2$, $b_\alpha = 15$ ($\mathrm{E}(n^*) \approx 1$), $a_\alpha = 2$, $b_\alpha = 4$ ($\mathrm{E}(n^*) \approx 3$), $a_\alpha = 2$, $b_\alpha = 0.9$ ($\mathrm{E}(n^*) \approx 10$) and $a_\alpha = 2$, $b_\alpha = 0.1$ ($\mathrm{E}(n^*) \approx 48$). Discuss prior sensitivity analysis results for α and n^*, as well as for posterior predictive inference (again, see (5) below).

(4) Illustrate the *clustering* induced by this DP mixture model using the posterior samples for the θ_i. For example, you can plot the median and two quantiles from $p(\theta_i \mid \text{data})$, for $i = 1, ..., n$. You can also obtain $p(\theta_0 \mid \text{data}) = \int p(\theta_0 \mid \theta_1, ..., \theta_n, \alpha, \mu, \tau^2) p(\theta_1, ..., \theta_n, \alpha, \mu, \tau^2 \mid \text{data})$, that is, the posterior predictive density for θ_0 (associated with a *new* observation y_0).

(5) Obtain the posterior predictive density $p(y_0 \mid \text{data})$ and use it to study how successful the model is in capturing the distributional shape suggested by the data. Compare also with the prior predictive density $p(y_0)$.

2. Consider the more general location-scale normal DP mixture model

$$f(\cdot \mid G) = \int k_N(\cdot \mid \theta, \phi)\, \mathrm{d}G(\theta, \phi), \qquad G \mid \alpha, \boldsymbol{\psi} \sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\psi})),$$

with the conjugate normal/inverse-gamma specification for the centering distribution

$$G_0(\theta, \phi \mid \boldsymbol{\psi}) = \mathrm{N}(\theta \mid \mu, \phi/\kappa) \times \text{inv-gamma}(\phi \mid c, \beta)$$

for fixed c and random $\boldsymbol{\psi} = (\mu, \kappa, \beta)$.

Use the function `DPdensity` from the `DPpackage` to fit this model to the same data set with problem 1. Discuss prior specification for the hyperparameters μ, κ and β. Use appropriate types of inference to compare the performance of the location-scale normal DP mixture above with the location normal DP mixture model from problem 1.

3. Consider the data set on the incidence of faults in the manufacturing of rolls of fabric:

`http://www.stat.columbia.edu/~gelman/book/data/fabric.asc`

where the first column contains the length of each roll, which is the covariate with values x_i, and the second column contains the number of faults, which is the response with values y_i, for $i = 1, ..., n$, with $n = 32$.

A Poisson regression is a possible model for such data, where the y_i are assumed to arise independently, given parameters $\theta > 0$ and $\beta \in \mathbb{R}$, from Poisson distributions with means $\mathrm{E}(y_i \mid \beta, \theta) = \theta \exp(\beta x_i)$, such that $\log(\theta)$ is the intercept and β is the slope of a linear regression function under a logarithmic transformation of the Poisson means. The Bayesian model is completed with priors for θ and β.

The Poisson regression can be extended in a hierarchical fashion to allow for over-dispersion relative to the Poisson response distribution. In particular, the response distribution can be extended to a negative Binomial under the following hierarchical structure:

$$\begin{aligned} y_i \mid \theta_i, \beta &\overset{ind.}{\sim} \mathrm{Poisson}(y_i \mid \theta_i \exp(\beta x_i)), \quad i = 1, ..., n \\ \theta_i \mid \mu, \zeta &\overset{i.i.d.}{\sim} \mathrm{gamma}(\zeta, \zeta\mu^{-1}), \quad i = 1, ..., n \end{aligned}$$

such that the mean of the gamma distribution for the θ_i is μ and the variance is μ^2/ζ. Under this hierarchical model, $\mathrm{E}(y_i \mid \beta, \mu, \zeta) = \mu \exp(\beta x_i)$ and $\mathrm{Var}(y_i \mid \beta, \mu, \zeta) > \mu \exp(\beta x_i)$, thus achieving over-dispersion relative to the Poisson regression model. In this case, the Bayesian model is completed with priors for β, μ and ζ.

Develop a semiparametric DP mixture regression model for the count responses y_i, which includes as limiting cases both of the parametric regression models discussed above. Discuss prior specification for your DP mixture model, and implement it for the specific data set (you can use any MCMC algorithm you wish, but you should write your own code). Compare the inference for the mean regression function arising from the two parametric models and from the semiparametric DP-based extension. Use a model comparison criterion for more formal comparison of the three models.