# AMS 241: Bayesian Nonparametric Methods (Winter 2009)

Homework set on Dirichlet process mixture models
(due Tuesday February 24)

**Density estimation with location normal Dirichlet process mixtures**

Consider the Dirichlet process (DP) mixture model

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

where $K_N(\cdot; \theta, \phi)$ ($k_N(\cdot; \theta, \phi)$) denotes the distribution function (density function) of a normal distribution with mean $\theta$ and variance $\phi$. Assume an inv-gamma$(a_\phi, b_\phi)$ prior for $\phi$, a gamma$(a_\alpha, b_\alpha)$ prior for $\alpha$, and take $\mathrm{N}(a_\mu, b_\mu)$ and inv-gamma$(a_{\tau^2}, b_{\tau^2})$ priors for the mean, $\mu$, and variance, $\tau^2$, respectively, of the normal base 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

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

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

To study inference under this model, consider a simulated data set (available from the course webpage, http://www.soe.ucsc.edu/classes/ams241/Winter09), of size $n = 250$, from a mixture of three normals, 0.2 N(-5,1) + 0.5 N(0,1) + 0.3 N(3.5,1).

(1) Obtain all the required expressions for the Pólya urn based Gibbs sampler, which can be used to draw 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 $\phi$, $\mu$, and $\tau^2$. Study sensitivity of posterior inference for $\phi$, $\mu$, and $\tau^2$ to the prior choice. In addition to the posteriors for $\phi$, $\mu$, $\tau^2$, examine sensitivity of posterior predictive inference (see (5) below).

(3) Obtain the posteriors for $\alpha$ and $n^*$ under different prior choices for $\alpha$ (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 of posterior results for $\alpha$ and $n^*$, as well as of posterior predictive inference (again, see (5) below).

(4) Illustrate the *clustering* induced by this DP mixture model using the posteriors for the $\theta_i$, $i = 1, ..., n$. For example, you can plot, for each $i = 1, ..., n$, the median and two quantiles from $p(\theta_i \mid \text{data})$. You can also report $n^*$ and $\theta_j^*$, $n_j$, $j = 1, ..., n^*$, from a few specific posterior draws (say 5 to 10 draws). Finally, you can show

$$p(\theta_0 \mid \text{data}) = \int p(\theta_0 \mid \theta_1, ..., \theta_n, \alpha, \mu, \tau^2) p(\theta_1, ..., \theta_n, \alpha, \mu, \tau^2 \mid \text{data}),$$

the posterior predictive density for $\theta_0$ (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 distribution $p(y_0)$.