AMS 241: Bayesian Nonparametric Methods (Fall 2010)

Homework set on Dirichlet process mixture models
(due Tuesday November 9)

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

\[ F(\cdot; G, \phi) = \int K_N(\cdot; \theta, \phi) dG(\theta), \quad G | \alpha, \mu, \tau^2 \sim \text{DP}(\alpha, G_0 = N(\mu, \tau^2)), \]

where \( K_N(\cdot; \theta, \phi) \) denotes the distribution 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 \( N(\mu, \tau^2) \) 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 \( ab \).) Therefore, the hierarchical version of this semiparametric DP mixture model is given by

\[
\begin{align*}
\theta_i | G & \sim i.i.d. \kappa_N(y_i; \theta_i, \phi), \quad i = 1, ..., n \\
G | \alpha, \mu, \tau^2 & \sim \text{DP}(\alpha, G_0 = N(\mu, \tau^2)) \\
\alpha, \mu, \tau^2, \phi & \sim p(\alpha)p(\mu)p(\tau^2)p(\phi),
\end{align*}
\]

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/Fall10), 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 \( p(\theta_1, ..., \theta_n, \alpha, \mu, \tau^2 | \text{data}) \), 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 \( a_\alpha = 2, b_\alpha = 15 \) (\( E(n^*) \approx 1 \)), \( a_\alpha = 2, b_\alpha = 4 \) (\( E(n^*) \approx 3 \)), \( a_\alpha = 2, b_\alpha = 0.9 \) (\( E(n^*) \approx 10 \)) and \( a_\alpha = 2, b_\alpha = 0.1 \) (\( E(n^*) \approx 48 \)). 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 \( p(\theta_i | \text{data}) \). You can also obtain

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

the posterior predictive density for \( \theta_0 \) (associated with a new observation \( y_0 \)).

(5) Obtain the posterior predictive density \( p(y_0 | \text{data}) \) 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. (OPTIONAL)

Consider the more general location-scale normal DP mixture model

\[ F(\cdot;G) = \int K_N(\cdot;\theta, \phi)dG(\theta, \phi), \quad G \mid \alpha, \psi \sim \text{DP}(\alpha, G_0(\psi)), \]

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

\[ G_0(\theta, \phi; \psi) = N(\theta; \mu, \phi/\kappa) \times \text{inv-gamma}(\phi; c, \beta) \]

for fixed \( c \) and random \( \psi = (\mu, \kappa, \beta) \).

Use the function \texttt{DPdensity} from the \texttt{DPpackage} to fit this model to the same data set with problem 1. Discuss prior specification for hyperparameters \( \mu, \kappa \) and \( \beta \). 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.