AMS 241: Bayesian Nonparametric Methods (Spring 2018)

Homework set on Dirichlet process mixture models
(due Tuesday May 22)

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

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

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 the data available from

`https://ams241-spring18-01.courses.soe.ucsc.edu/homework-assignments`

This is a synthetic data set based on $n = 250$ responses 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 $p(\boldsymbol{\theta}, \alpha, \phi, \mu, \tau^2 \mid \text{data})$, where $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)$, and 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 corresponding to an increasing number of distinct mixture components. 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 $p(\theta_0 \mid \text{data}) = \int p(\theta_0 \mid \boldsymbol{\theta}, \alpha, \mu, \tau^2) p(\boldsymbol{\theta}, \alpha, \mu, \tau^2 \mid \text{data})$, 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.

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

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

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.