STAT 222: Bayesian Nonparametric Methods (Spring 2020)

Homework set on Dirichlet process mixture models
(due Monday May 18)

1. Assume that $\theta_i \mid G \overset{i.i.d.}{\sim} G$, for $i = 1, ..., n$, with $G \mid \alpha, \psi \sim \mathrm{DP}(\alpha, G_0(\psi))$, where G_0 is a continuous distribution (i.e., it has no atoms) with density g_0. Then, marginalizing G over its DP prior, the joint distribution for the θ_i can be written as

$$p(\theta_1, ..., \theta_n \mid \alpha, \psi) = \sum_{\boldsymbol{\pi} \in \mathcal{P}_n} p(\boldsymbol{\pi} \mid \alpha) \left\{ \prod_{j=1}^{n^*} g_0(\theta_{e_{j,1}} \mid \psi) \prod\nolimits_{i=2}^{n_j} \delta_{\theta_{e_{j,1}}}(\theta_{e_{j,i}}) \right\}$$

where $\mathcal{P}_n$ denotes the set of all partitions $\boldsymbol{\pi} = \{s_j : j = 1, ..., n^*\}$ of $\{1, ..., n\}$, and $p(\boldsymbol{\pi} \mid \alpha)$ the DP-implied prior probability for partition $\boldsymbol{\pi}$. Here, n^* is the number of cells of partition $\boldsymbol{\pi}$, n_j is the number of elements in cell s_j, and $e_{j,1} < ... < e_{j,n_j}$ are the elements of cell s_j. Prove by induction that

$$p(\boldsymbol{\pi} \mid \alpha) = \frac{\alpha^{n^*}}{\alpha^{(n)}} \prod_{j=1}^{n^*} (n_j - 1)!$$

where $\alpha^{(n)} = \prod_{m=1}^{n} (\alpha + m - 1) = \Gamma(\alpha + n)/\Gamma(\alpha)$ is the ascending factorial.

2. Consider the location normal 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 the 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

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

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 in "hwk3-data.txt" (the file is available from Files on canvas). 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) Derive the expressions for the posterior full conditionals of 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 $= \{y_i : i = 1, ..., n\}$. Develop your own code to implement the Gibbs sampler.

(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: $a_\alpha = 2$, $b_\alpha = 15$ ($\mathrm{E}(n^*) \approx 1$); $a_\alpha = 2$, $b_\alpha = 4$ ($\mathrm{E}(n^*) \approx 3$); $a_\alpha = 2$, $b_\alpha = 0.9$ ($\mathrm{E}(n^*) \approx 10$); 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.

3. 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 2. 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 2.