STAT 222: Bayesian Nonparametric Methods (Spring 2020)

Homework set on DP mixture models: variational inference,
and semiparametric regression for count responses
(due Friday June 5)

1. Assume the (truncated) location normal DP mixture model:

$$
\begin{aligned}
y_i \mid \boldsymbol{Z}, L_i, \phi &\overset{ind.}{\sim} \mathrm{N}(y_i \mid Z_{L_i}, \phi^{-1}) && i = 1, \ldots, n \\
L_i \mid \boldsymbol{V} &\overset{i.i.d.}{\sim} \prod_{\ell=1}^{N} (p_\ell(\boldsymbol{V}))^{1(L_i=\ell)} && i = 1, \ldots, n \\
Z_\ell &\overset{i.i.d.}{\sim} \mathrm{N}(Z_\ell \mid m, s^2) && \ell = 1, \ldots, N \\
V_\ell &\overset{i.i.d.}{\sim} \mathrm{Beta}(V_\ell \mid 1, \alpha) && \ell = 1, \ldots, N-1 \\
\phi &\sim \mathrm{ga}(\phi \mid a_\phi, b_\phi)
\end{aligned}
$$

where the variance of the normal mixture kernel is ϕ^{-1}, $\mathrm{ga}(a, b)$ denotes the gamma distribution with mean a/b, and

$$
p_1(\boldsymbol{V}) = V_1; \qquad p_\ell(\boldsymbol{V}) = V_\ell \prod_{r=1}^{\ell-1} (1 - V_r), \;\; \ell = 2, \ldots, N-1; \qquad p_N(\boldsymbol{V}) = \prod_{r=1}^{N-1} (1 - V_r).
$$

The parameters of the DP centering distribution, (m, s^2), and the DP precision parameter, α, are fixed, such that the full parameter vector is $\boldsymbol{\theta} = (\boldsymbol{Z}, \boldsymbol{V}, \boldsymbol{L}, \phi)$, with $\boldsymbol{Z} = (Z_1, ..., Z_N)$, $\boldsymbol{V} = (V_1, ..., V_{N-1})$, and $\boldsymbol{L} = (L_1, ..., L_n)$.

Consider the synthetic data from problem 2 of homework 3. Implement a mean-field variational method, using the variational approximation:

$$
q_{\boldsymbol{\eta}}(\boldsymbol{\theta}) = q_{\boldsymbol{\beta}}(\phi) \prod_{\ell=1}^{N-1} q_{\boldsymbol{\gamma}_\ell}(V_\ell) \prod_{\ell=1}^{N} q_{\boldsymbol{\xi}_\ell}(Z_\ell) \prod_{i=1}^{n} q_{\boldsymbol{\pi}_i}(L_i)
$$

where $q_{\boldsymbol{\beta}}(\phi) = \mathrm{ga}(\phi \mid \beta_1, \beta_2)$, $q_{\boldsymbol{\gamma}_\ell}(V_\ell) = \mathrm{Beta}(V_\ell \mid \gamma_{\ell 1}, \gamma_{\ell 2})$, $q_{\boldsymbol{\xi}_\ell}(Z_\ell) = \mathrm{N}(Z_\ell \mid \xi_{\ell 1}, \xi_{\ell 2})$, and $q_{\boldsymbol{\pi}_i}(L_i) = \prod_{\ell=1}^{N} \pi_{i\ell}^{1(L_i=\ell)}$. Refer to the second set of course notes for the variational algorithm.

You can use an informative prior for ϕ, and *optimal* values for (m, s^2) and α, based on your results from homework 3. For inference, focus on estimation of the posterior predictive density. Discuss sensitivity of the results to the initial values for the variational parameters, $\boldsymbol{\eta} = (\beta_1, \beta_2, \{(\gamma_{\ell 1}, \gamma_{\ell 2})\}_{\ell=1}^{N-1}, \{(\xi_{\ell 1}, \xi_{\ell 2})\}_{\ell=1}^{N}, \{(\pi_{i1}, ..., \pi_{iN})\}_{i=1}^{n})$.

2. Consider data 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:

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

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 DP mixture model. Use a model comparison criterion for more formal comparison of the three models.