AMS 241: Bayesian Nonparametric Methods (Fall 2015)

Homework set on Dirichlet process priors
(due Tuesday October 20)

1. Assume a Dirichlet process (DP) prior, $\mathrm{DP}(\alpha, G_0)$, for distributions G on $\mathcal{X}$. Show that for any (measurable) disjoint subsets B_1 and B_2 of $\mathcal{X}$, $\mathrm{Corr}(G(B_1), G(B_2))$ is negative. Is the negative correlation for random probabilities induced by the DP prior a restriction? Discuss.

2. Simulation of Dirichlet process prior realizations
Consider a $\mathrm{DP}(\alpha, G_0)$ prior over the space of distributions (equivalently c.d.f.s) G on $\mathbb{R}$, with $G_0 = \mathrm{N}(0,1)$. Use both Ferguson's original definition and Sethuraman's constructive definition to generate (multiple) prior realizations from the $\mathrm{DP}(\alpha, \mathrm{N}(0,1))$ for fixed α with values ranging from *small* to *large*. In addition to prior c.d.f. realizations, obtain, for each value of α, the corresponding prior distribution for the mean functional $\mu(G) = \int t \mathrm{d}G(t)$ and for the variance functional $\sigma^2(G) = \int t^2 \mathrm{d}G(t) - \{\int t \mathrm{d}G(t)\}^2$. (Note that, because G_0 has finite first and second moments, both of the random variables $\mu(G)$ and $\sigma^2(G)$ take finite values almost surely; see Section 4 in Ferguson, 1973.)

Finally, consider simulation under a mixture of DPs (MDP) prior, which extends the DP above by adding a gamma prior for α. Then, the MDP prior for G is defined such that, given α, $G \mid \alpha \sim \mathrm{DP}(\alpha, \mathrm{N}(0,1))$. To simulate from the MDP, one can use either of the DP definitions given draws for α from its prior. You can work with 2-3 different gamma priors for α.

3. Posterior inference for one-sample problems using DP priors
Consider data $= \{y_1, ..., y_n\}$, and the following DP-based nonparametric model:

$$y_i \mid G \overset{\text{i.i.d.}}{\sim} G, \;\; i = 1, ..., n; \quad G \sim \mathrm{DP}(\alpha, G_0)$$

with $G_0 = \mathrm{N}(m, s^2)$ for fixed m, s^2, and α. The objective here is to use simulated data to study posterior inference results for G under different prior choices for α and G_0, different underlying distributions that generate the data, and different sample sizes. In particular, consider:

- two data generating distributions: a $\mathrm{N}(0,1)$ distribution, and the mixture of normal distributions, $0.5\mathrm{N}(-2.5, 0.5^2) + 0.3\mathrm{N}(0.5, 0.7^2) + 0.2\mathrm{N}(1.5, 2^2)$, which yields a bimodal c.d.f. with *heavy* right tail;
- sample sizes $n = 20$, $n = 200$, and $n = 2000$.

Discuss prior specification for the DP prior parameters m, s^2, and α. For each of the 6 data sets corresponding to the combinations above, obtain posterior point and interval estimates for the c.d.f. G and discuss how well the model fits the data. Perform a prior sensitivity analysis to study the effect of m, s^2, and α on the posterior estimates for G.

4. Posterior inference for count data using MDP priors
Consider again modeling a single distribution F, here for count responses, that is, the support for F is $\{0, 1, 2, ...\}$. The model for the data $= \{y_1, ..., y_n\}$ is given by

$$y_i \mid F \overset{\text{i.i.d.}}{\sim} F, \;\; i = 1, ..., n; \quad F \mid \alpha, \lambda \sim \mathrm{DP}(\alpha, F_0(\cdot) = \mathrm{Poisson}(\cdot \mid \lambda))$$

that is, we now have a DP prior for F, given random precision parameter α, and random mean λ for the centering Poisson distribution. Moreover, assume independent gamma priors for α and λ. Again, use simulated data under two different scenarios for the true data generating distribution:

- Poisson distribution with mean 5.
- Mixture of two Poisson distributions with means 3 and 11, and corresponding mixture weights given by 0.7 and 0.3.

For both cases, work with a sample of size $n = 300$ for the simulated data. Discuss specification for the prior hyperparameters of α and λ. Develop a posterior simulation method to explore the posterior distribution for α, and to estimate the posterior predictive distribution, $\Pr(Y = y \mid \text{data})$, $y = 0, 1, 2, ...$