STAT 222: Bayesian Nonparametric Methods (Spring 2020)

Homework set on Dirichlet processes: posterior inference
(due Thursday April 30)

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

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

with $G_0 = \text{N}(m, s^2)$ for fixed m, s^2, and α. Work with 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 $\text{N}(0, 1)$ distribution, and the mixture of normal distributions, $0.5\text{N}(-2.5, 0.5^2) + 0.3\text{N}(0.5, 0.7^2) + 0.2\text{N}(1.5, 2^2)$, which yields a bimodal distribution 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 six 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.

2. 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

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

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 joint posterior distribution for F and (α, λ). Obtain point estimates for the underlying data generating probability mass functions through the posterior predictive distribution, $\Pr(Y = y \mid \text{data})$, for values of y in the effective support of each of the two data generating distributions.