STA 941: Bayesian Nonparametrics
HW Set 2

1. Suppose $P \sim \mathrm{DP}(a, \pi)$ and $A \subset \Theta = \mathrm{supp}(\pi)$. Let $P^* = P|_A$, the restriction of P to A. Show that $P^* \sim \mathrm{DP}(a\pi(A), \pi|_A)$. [Hint: Use Ferguson's definition]

2. Suppose $Z_1, \ldots, Z_n | P \overset{\text{IID}}{\sim} P$, $P \sim \mathrm{DP}(a, \pi)$ with π being non-atomic [i.e., π has a pdf w.r.t the Lebesgue measure on $\Theta = \mathrm{supp}(\pi)$]. Let K_n denote the number of unique observations among $Z_1, \ldots, Z_n$. Show that

$$\Pr(K_n = k|a) = C_n(k) a^k \frac{n!\Gamma(a)}{\Gamma(n+a)}, \quad k \in \{1, \ldots, n\},$$

where $C_n(k) = P(K_n = k|a = 1)$. Bonus points for deriving the expression for $C_n(k)$. [Hint: Use recursion on k and n, noticing that $K_n = k$ can happen only when K_{n-1} is either $k-1$ or k.]

3. Assume the same setting as in #3 and define $D_i = I(Z_i \notin \{Z_1, \ldots, Z_{i-1}\})$, $i = 1, \ldots, n$ with $D_1 \equiv 1$. Then $K_n = D_1 + \cdots + D_n$.

   (a) Calculate $\mathbb{E}D_i$ and $\mathbb{V}\mathrm{ar}D_i$ and show that $\mathbb{E}K_n \asymp \mathbb{V}\mathrm{ar}K_n \asymp a\log(1 + n/a)$ as $n \to \infty$ [and $a > 0$ is fixed].

   (b) Show that $\{K_n - a\log(1 + n/a)\}/\sqrt{a\log(1 + n/a)} \xrightarrow{d} N(0, 1)$.

   (c) Take $\pi = Ga(2, 2)$. Evaluate the 95% prior range for K_n for each combination of $n \in \{100, 200, 500\}$ and $a \in \{1/5, 1, 5\}$ by both the asymptotic result from part (b) as well as direct simulation.

4. Using the identity $\Gamma(a)/\Gamma(a + n) = B(a, n)/\Gamma(n)$, rework the calculations in Section 6 of Escobar and West (1995) to derive a simpler auxiliary parameter sampling scheme for the precision parameter in a DP mixture model. In particular, equation (13) of the paper can be reduced to making a draw from a single Gamma distribution.

5. Reanalyze the Traffic data from R package MASS, but now with a Dirichlet process mixture of Poisson model. Take the base measure to be a Gamma distribution with equal, and small, rate and shape parameters. Implement 3 different Markov chain samplers, as given by Algorithms 2, 3 and 8 of Neal (2000), augmented with a learning of the precision parameter under a unit Exponential prior. Make a comparison between the three samplers as in Table 1 of Neal's paper. For each sampler, make posterior predictive draws for daily accident counts under each speed limit condition, and, report both a 95% posterior predictive interval for the count reduction and the posterior probability that the count reduction is positive. Also comment on how many mixture components were needed to explain the accident count distribution under each speed limit condition.

References

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the american statistical association 90*(430), 577–588.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics 9*(2), 249–265.