STAT 222: Bayesian Nonparametric Methods (Spring 2020)

Homework set on Dirichlet processes: prior properties and simulation
(due Wednesday April 22)

1. Assume that random vector $(Y_1, ..., Y_k)$ follows a Dirichlet$(k; a_1, ..., a_k)$ distribution, such that $(Y_1, ..., Y_{k-1})$ has density

$$\frac{\Gamma(a_1 + ... + a_k)}{\Gamma(a_1) \times ... \times \Gamma(a_k)}\, y_1^{a_1-1} \times ... \times y_{k-1}^{a_{k-1}-1} \times (1 - y_1 - ... - y_{k-1})^{a_k-1}$$

where $y_i > 0$, for $i = 1, ..., k-1$, and $\sum_{i=1}^{k-1} y_i < 1$. Consider a partition $I_1, ..., I_M$ of $\{1, ..., k\}$ (so that the I_j, $j = 1, ..., M$, are pairwise disjoint, and their union is $\{1, ..., k\}$), and let $U_j = \sum_{i \in I_j} Y_i$, for $j = 1, ..., M$. Show that:

(1) $(U_1, ..., U_M)$ follows a Dirichlet$(M; b_1, ..., b_M)$ distribution, where $b_j = \sum_{i \in I_j} a_i$, for $j = 1, ..., M$.

(2) $(Y_i/U_j : i \in I_j) \overset{ind.}{\sim}$ Dirichlet$(M_j; a_i, i \in I_j)$, for $j = 1, ..., M$, where M_j is the size of I_j.

(3) $(U_1, ..., U_M)$ and $(Y_i/U_j : i \in I_j)$, $j = 1, ..., M$, are independent.

(**Hint.** Use the definition of the Dirichlet distribution in terms of independent gamma random variables: the Dirichlet$(k; a_1, ..., a_k)$ distributed random vector $(Y_1, ..., Y_k)$ can be constructed as $(Y_1, ..., Y_k) = (Z_1/Z, ..., Z_k/Z)$, where $Z_i \overset{ind.}{\sim}$ gamma$(a_i, 1)$, for $i = 1, ..., k$, and $Z = \sum_{i=1}^{k} Z_i$. Moreover, $(Z_1/Z, ..., Z_k/Z)$ is independent of Z.)

For problems 2 and 3, consider a Dirichlet process (DP) prior, DP(α, P_0), for distributions (probability measures) P on sample space $\mathcal{X}$, where $\alpha > 0$ is the total mass (precision) parameter, and P_0 is a specified distribution on $\mathcal{X}$.

2. Show that for any (measurable) disjoint subsets B_1 and B_2 of $\mathcal{X}$, Corr$(P(B_1), P(B_2))$ is negative. Is the negative correlation for random probabilities induced by the DP prior a restriction? Discuss.

3. Let B be a specific (measurable) subset of $\mathcal{X}$ such that $P_0(B) > 0$, and denote by P_B the corresponding conditional distribution, that is, $P_B(A) = P(A \mid B) = P(A \cap B)/P(B)$.

(a) Prove that if $P \sim$ DP(α, P_0), then $P_B \sim$ DP$(\alpha P_0(B), P_0(A \mid B))$, that is, a DP with precision parameter $\alpha P_0(B)$ and centering distribution $P_0(A \mid B) = P_0(A \cap B)/P_0(B)$.

(b) Consider finite measurable partitions $A_1, ..., A_r$ and $C_1, ..., C_s$ of B and B^c, respectively. Show that $P(B)$, $P(B^c)$, $\{P_B(A_1), ..., P_B(A_r)\}$ and $\{P_{B^c}(C_1), ..., P_{B^c}(C_s)\}$ are independent.

(**Hint.** Use problem 1. Note that, given the definition of the conditional distribution P_B, it suffices to consider partitions of B in order to verify the DP definition for part (a).)

4. 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 $G_0 = \mathrm{N}(0,1)$.

(a) Use both Ferguson's original definition and Sethuraman's constructive definition to generate (multiple) prior c.d.f. realizations from the $\mathrm{DP}(\alpha, \mathrm{N}(0,1))$, for different values of α ranging from *small* to *large*.

(b) 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.)

(c) Consider also simulation under a mixture of DPs (MDP) prior, which extends the DP above by adding a prior for α. Therefore, the MDP prior for G is defined such that, $G \mid \alpha \sim \mathrm{DP}(\alpha, \mathrm{N}(0,1))$, with a prior assigned to precision parameter α. To simulate from the MDP, you can use either of the DP definitions given draws for α from its prior. You can work with a gamma prior for α and 2-3 different choices for the gamma prior parameters.