STA 3000, Spring 2014, Assignment #4. Due in class on April 4.

These questions are to be answered by each student individually. Any discussions you have with other people about these questions should concern general issues only, and should not result in your taking away written or electronically-recorded notes.

Consider a Polya tree process that defines a distribution over distributions on the interval $(0, 1)$, using an infinite set of parameters θ_0, θ_1, θ_{00}, θ_{01}, θ_{10}, θ_{11}, θ_{000}, etc. Given values for these parameters, the distribution of a data point, X, is defined by

$$P(X = 0.d_1 d_2 d_3 \ldots \mid \theta) \;=\; \theta_{d_1}\, \theta_{d_1 d_2}\, \theta_{d_1 d_2 d_3} \, \cdots$$

where d_1, d_2, d_3, etc. are the digits in the binary representation of the value of X (eg, $0.11000\ldots$ is $1/2 + 1/4 = 3/4$).

We fix $\theta_{d_1 \ldots d_k 0} = 1 - \theta_{d_1 \ldots d_k 1}$, and give independent Beta$(\alpha_{k+1}, \alpha_{k+1})$ priors to each $\theta_{d_1 \ldots d_k 1}$, where $\alpha_1, \alpha_2, \ldots$ are fixed positive real constants. (For example, the prior for θ_1 is Beta(α_1, α_1), and the prior for θ_{011} is Beta(α_3, α_3).)

A) Prove that if $\alpha_k = c\, 2^{-k}$ for some positive c, then this Polya tree process is also a Dirichlet process, and find the parameters of this Dirichlet process.

B) Suppose that X_1, X_2, $\ldots$ are modelled as being i.i.d. given θ, with θ and each X_i given θ having the Polya tree distribution defined above, so that the marginal distribution of X_1, X_2, $\ldots$ is exchangeable. Consider the following specifications of the α_k parameters of the Polya tree (with c being a positive real):

   1) $\alpha_k = c/k^2$.
   2) $\alpha_k = c$.
   3) $\alpha_k = c\, 2^k$.

   For each of these three specifications, determine if you can whether $P(X_1 = X_2)$ is non-zero, and whether $\lim_{i \to \infty} P(|X_2 - x_1| < 2^{-i} \mid X_1 = x_1) \;/\; 2^{-i+1}$ is zero, infinity, or something finite and non-zero. Discuss whatever implications your results have regarding whether a distribution on $(0, 1)$ drawn from one of these Polya tree distributions is continuous or absolutely continuous (with respect to Lesbegue measure).