STAT 222 Homework #1 **Mary Silva**
UCSC 22 April 2020

1. (1) $(Y_1, ..., Y_k) \sim Dir(a_1, ..., a_k)$. Consider a partition $I_1, ..., I_M$ of $\{1, ..., k\}$. Let $U_j = \sum_{u \in I_j} Y_i$ for $j = 1, ..., M$. Using the definition of the Dirichlet as independent gamma random variables, $(Y_1, ..., Y_k)$ can be constructed as $\left(\frac{Z_i}{Z}, ..., \frac{Z_k}{Z}\right)$ where $Z_i \overset{ind}{\sim} \Gamma(a_i, 1)$ and $Z = \sum_{i=1}^{k} Z_i$. Then

$$\begin{aligned}
(U_1, ..., U_m) &= \left(\sum_{i \in I_1} Y_i, \sum_{i \in I_1} Y_i, ..., \sum_{i \in I_M} Y_i\right) \\
&= \frac{1}{\sum_{i=1}^{k} Z_i}\left(\sum_{i \in I_1} Z_i, \sum_{i \in I_2} Z_i, ... \sum_{i \in I_M} Z_i\right) \\
&\overset{d}{=} \frac{1}{\Gamma(\sum_{i=1}^{k} a_i, 1)}\left(\Gamma\left(\sum_{i \in I_1} a_i, 1\right), \Gamma\left(\sum_{i \in I_2} a_i, 1\right), ...\Gamma\left(\sum_{i \in I_M} a_i, 1\right)\right) \\
&\overset{d}{=} Dir\left(\sum_{i \in I_1} a_i, \sum_{i \in I_2} a_i, ..., \sum_{i \in I_M} a_i\right) \\
&\overset{d}{=} Dir(b_1, ..., b_M)
\end{aligned}$$

Therefore, $(U_1, ..., U_M) \sim Dir(b_1, ..., b_m)$ where $b_j = \sum_{i \in I_j} a_i$.

(2) Since we just showed $U_j \sim Dir(b_j, ..., b_m)$, then $Y_i/U_j \overset{ind}{\sim} Dir(a_i, i \in I_J)$

(3) I never got around to these.

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.

Solution: Define $B_1 \cup B_2 \cup B_3 = X$, where $B_3 = (B_1 \cup B_2)^c$. Without loss of generality, B_1 and B_2 are any measurable subset of the sample space. In order to compute the correlation, it is sufficient to show that the covariance is negative because the variance is just a positive scalar factor.

Given that P is a Dirichlet process

$$P(B_1), P(B_2), P(B_3) \sim Dir(\alpha P_0(B_1), \alpha P_0(B_2), \alpha P_0(B_3))$$

where $\alpha > 0$ and $P_0(\cdot)$ is a probability measure.

The covariance is computed as

$$Cov(P(B_1), P(B_2)) = E[P(B_1)P(B_2)] - E[P(B_1)]E[P(B_2)]$$

Using the definition of the Dirichlet distribution in terms of independent gamma random variables, $Z_i \sim Gamma(\alpha P_0(B_i), 1)$ and $Z \sim Gamma(\sum_{j=1}^{k} \alpha P_0(B_i), 1)$. Then the expected value is defined as

$$E[P(B_i)] = \frac{E(Z_i)}{E(Z)} = \frac{\alpha P_0(B_i)}{\alpha P_0(B_1) + \alpha P_0(B_2) + \alpha P_0(B_3)} = \frac{\alpha P_0(B_i)}{\alpha P_0(X)} = P_0(B_i)$$

We can derive the multivariate moment of the Dirichlet random variables. For simplicity of notation, let $U_i = P(B_i)$ and $v_i = P_0(B_i)$. Then we have

$$\begin{aligned}
E[P(B_1)P(B_2)] &= E(U_1 U_2) \\
&= \int u_1 u_2 \frac{\Gamma(\alpha v_1 + \alpha v_2 + \alpha v_3)}{\Gamma(\alpha v_1)\Gamma(\alpha v_2)\Gamma(\alpha v_3)} u_1^{\alpha v_1 - 1} u_2^{\alpha v_2 - 1} u_3^{\alpha v_3 - 1} \, d\mathbf{u} \\
&= \frac{\Gamma(\alpha v_1 + 1)\Gamma(\alpha v_2 + 1)}{\Gamma(\alpha v_1 + \alpha v_2 + \alpha v_3 + 2)} \frac{\Gamma(\alpha v_1 + \alpha v_2 + \alpha v_3)}{\Gamma(\alpha v_1)\Gamma(\alpha v_2)} \times \\
&\times \int \frac{\Gamma(\alpha v_1 + \alpha v_2 + \alpha v_3 + 2)}{\Gamma(\alpha v_1 + 1)\Gamma(\alpha v_2 + 1)\Gamma(\alpha v_3)} u_1^{(\alpha v_1 + 1) - 1} u_2^{(\alpha v_2 + 1) - 1} u_3^{\alpha v_3 - 1} \, d\mathbf{u} \\
&= \frac{\Gamma(\alpha v_1 + 1)\Gamma(\alpha v_2 + 1)}{\Gamma(\alpha v_1 + \alpha v_2 + \alpha v_3 + 2)} \frac{\Gamma(\alpha v_1 + \alpha v_2 + \alpha v_3)}{\Gamma(\alpha v_1)\Gamma(\alpha v_2)} \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha + 2)} \frac{\alpha v_1 \Gamma(\alpha v_1) \cdot \alpha v_2 \Gamma(\alpha v_2)}{\Gamma(\alpha v_1)\Gamma(\alpha v_2)} \\
&= \frac{\alpha v_1 v_2}{(\alpha + 1)} \\
&= \frac{\alpha P_0(B_1) P_0(B_2)}{(\alpha + 1)}
\end{aligned}$$

Then

$$\begin{aligned}
Cov(P(B_1), P(B_2)) &= \frac{\alpha}{\alpha + 1} P_0(B_1) P_0(B_2) - \alpha P_0(B_1) \alpha P_0(B_2) \\
&= \left(\frac{\alpha}{\alpha + 1} - 1 \right) P_0(B_1) P_0(B_2) < 0
\end{aligned}$$

Since $\frac{\alpha}{\alpha+1} < 1$ given that $\alpha > 0$ and $P_0(\cdot)$ is a probability measure.

4. **Simulation of Dirichlet process prior realizations** Consider a $DP(\alpha, G_0)$ prior over the space of distributions (equivalently c.d.f's) G on $\mathbb{R}$, with $G_0 = N(0, 1)$.

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

   I choose α to be 0.1, 1, 10 and 100. DP's are drawn from Ferguson's definition as well as from the constructive definition (both are comparable). Changing α changes the shapes of each realization of the prior distributions. As α increase, we see less discreteness and the distributions are more concentrated around the theoretical N(0,1) $-$ which represented by the black line. This is due to the fact that α represents how confident we are that the prior distribution is true. For each simulation using the DP I am generating 10 CDF, each represented with a different color.

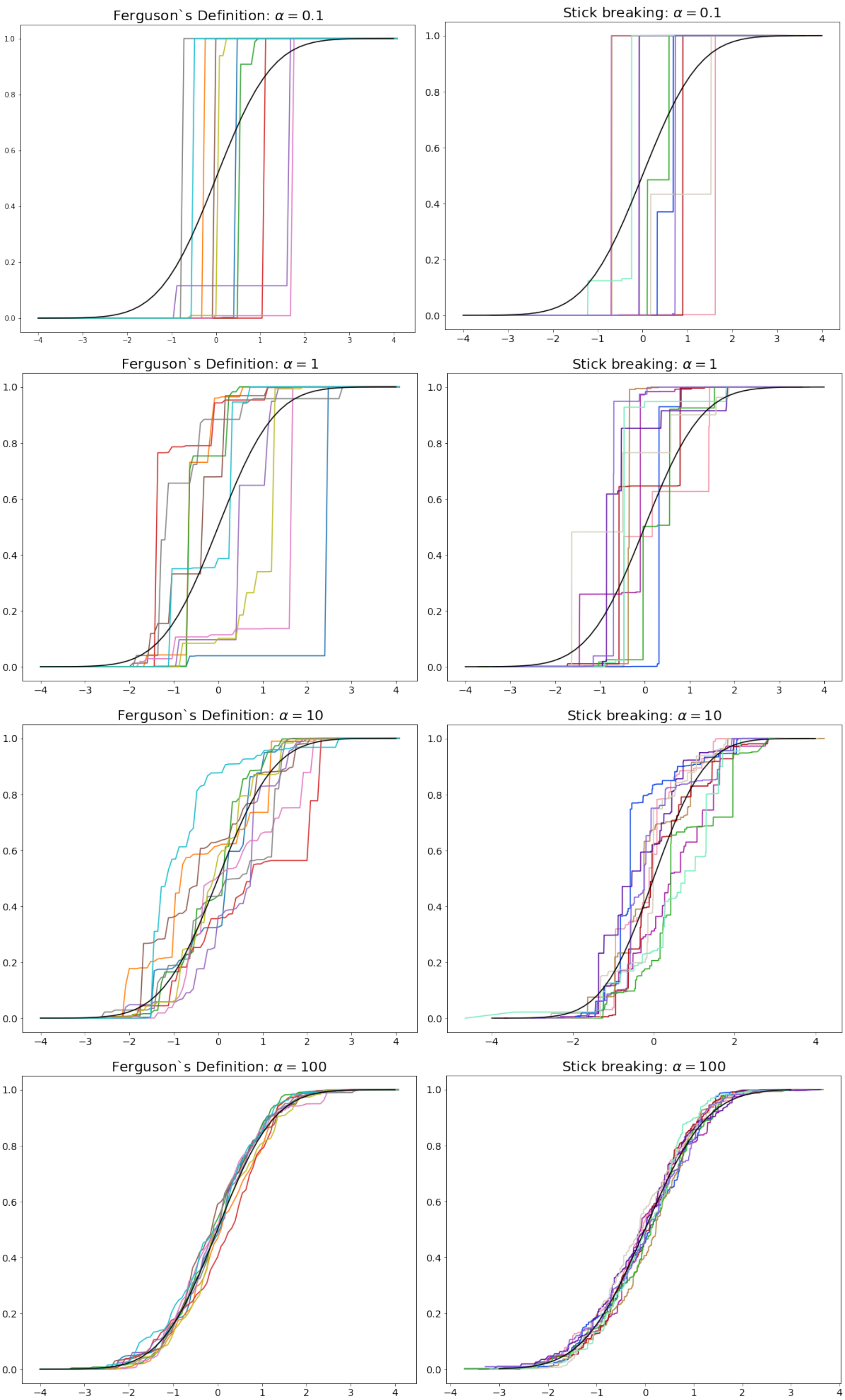

Figure 1: The left plots use Ferguson's original definition and the right plots use Sethuraman's constructive definition with various α. For reproducible code, see `https://github.com/msilva00/BNP_Homework/blob/master/HW1/HW1_Prob4.ipynb`

(b) In addition to prior c.d.f. realizations, obtain, for each value of α, the corresponding prior distribution for the mean functional

$$\mu(G) = \int t dG(t)$$

and for the variance functional

$$\sigma^2(G) = \int t^2 dG(t) - \left\{ \int t dG(t) \right\}^2$$

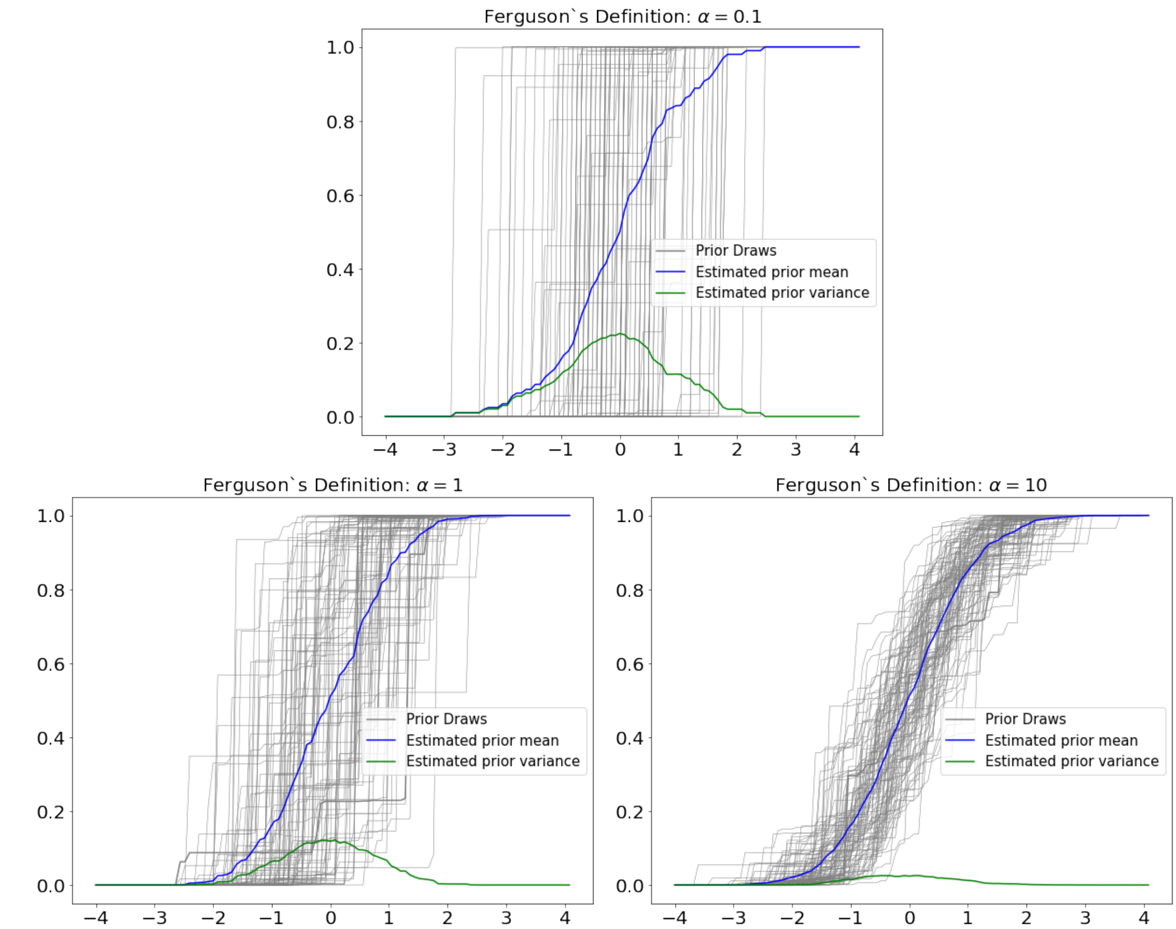

Figure 2: Using Ferguson's definition of the DP, I can compute the functional mean and the functional variance for each simulation, fixing the value of α. For reproducible code, see `https://github.com/msilva00/BNP_Homework/blob/master/HW1/HW1_Prob4.ipynb`

(c) Consider also a 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, $G|\alpha \sim DP(\alpha, N(0,1)$, with a prior assigned to the precision parameter α from its prior. You can work with a gamma prior for α and 2-3 different choices for the gamma prior parameters.

Next, we extend the Dirichlet process by adding a gamma prior for α.

$$\alpha \sim Gamma(3, 3)$$

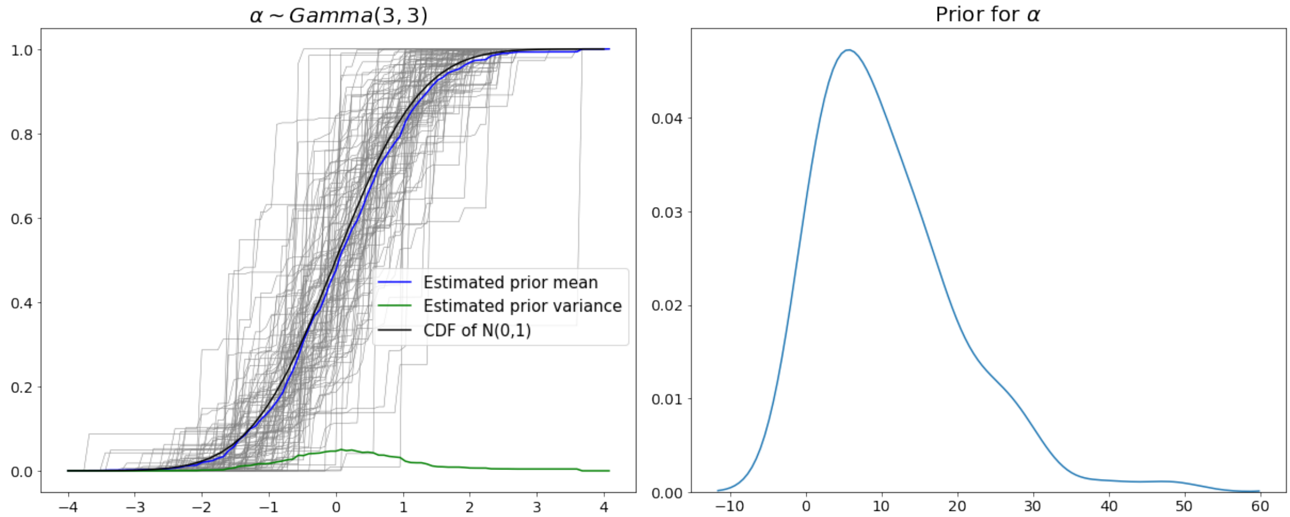

Figure 3: Here I simulate from the MDP using Ferguson's original definition given draws for α from a Gamma prior. The result considering $\alpha \sim Gamma(3,3)$ is similar to the result in part a, when $\alpha = 1$ because the mean of the α prior distribution is close to 1.

Last, consider the following Gamma prior

$$\alpha \sim Gamma(1,10)$$

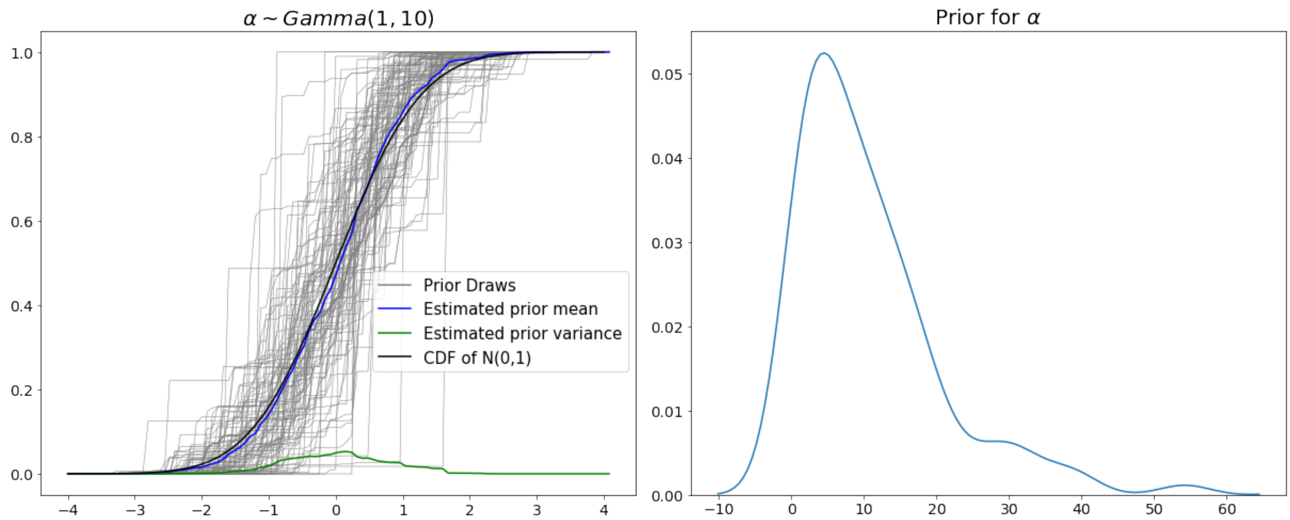

Figure 4: Here I simulate from the MDP using Ferguson's original definition given draws for α from a Gamma prior. The result considering $\alpha \sim Gamma(1,10)$ has low variance and also small mean, putting heavy emphasis on highly discrete DP draws.