Statistical Machine Learning
Sharif University of Technology
Hamid R. Rabiee

March 19, 2020
CE 40-956

TH Quiz 3 (Dirichlet Process - Chinese Restaurant Process)

Due April 7, 2020 (11:59 pm)

1. Let $G \sim DP(\alpha, G_0)$ be a dirichlet process.
   (a) If we increase/decrease α, predict how this change will affect the similarity between distributions G and G_0.
   (b) According to your answer to the last question, How will you choose α if distribution of your data consists of many/few clusters? Justify your answer.

2. Imagine a Chinese Restaurant Process running with the concentration parameter equal to α. If there are 10 customers in the restaurant and cluster (table) assignments are:

$$c_1, c_1, c_2, c_3, c_1, c_3, c_3, c_1, c_4, c_4$$

   (a) Compute the probability of this occurrence.
   (b) Change the order of cluster assignments such that the final clusters be equal to the last part. Then compute the probability of this occurrence again. Is CRP exchangeable?

3. Assume $G \sim DP(\alpha, G_0)$ is a Dirichlet process where G_0 is a probability measure defined over a set (Θ).
   (a) Find the mean of $G(A)$ for every $A \subset \Theta$.
   (b) Find the variance of $G(A)$ for every $A \subset \Theta$.

4. Assume $(\pi_1, \pi_2, \ldots, \pi_k) \sim Dirichlet(\alpha_1, \alpha_2, \ldots, \alpha_k)$. Show that:

$$(\pi_1 + \pi_2, \pi_3, \ldots, \pi_k) \sim Dirichlet(\alpha_1 + \alpha_2, \alpha_3, \ldots, \alpha_k).$$