Statistical Machine Learning
Sharif University of Technology
Hamid R. Rabiee

April 12, 2020
CE 40-956

TH Quiz 4 (Indian Buffet Process)

Due April 19, 2020 (11:59 pm)

1. Assume $\pi_1 \sim Poisson(\alpha_1)$ and $\pi_2 \sim Poisson(\alpha_2)$. Show that:

$$(\pi_1 + \pi_2) \sim Poisson(\alpha_1 + \alpha_2).$$

2. An Indian Buffet Process with parameter $\alpha = 4$ is running.

   (a) $Z_{n,k}$ indicates the presence of k'th feature (dish) in n'th sample (customer). Find probability distribution for number of non-zero elements in the 1'st, 2'nd, and 20'th rows of Z.

   (b) Implement IBP with $\alpha = 4$, run it for 1000 times (terminate after 20 customers entered) and then calculate $\sum_i Z_{1,i}$, $\sum_i Z_{2,i}$ and $\sum_i Z_{20,i}$. Plot the histogram of observed draws for these random variables in a chart. Do the results approve your calculations?

3. An improved version of IBP is introduced as two parameter IBP. The generative process of two parameter IBP is as follows:

   - First customer orders π_1 dishes; where $\pi_1 \sim Poisson(\alpha)$.
   - For all (n $>$ 1), n'th customer do two things:
     - Taste each existing dish with probability $\frac{m_k}{\beta+n-1}$; where β is the second parameter and m_k is number of previous customers who have tasted k'th dish.
     - Order π_n new dishes; where $\pi_n \sim Poisson(\frac{\alpha\beta}{\beta+n-1})$

   (a) Find probability distribution for the number of non zero elements in k'th row of Z matrix.

   (b) Find probability distribution for total number of dishes after n customers visit this buffet.

   (c) Implement two parameter IBP. Your implementation must include visualization of Z matrix for the first 50 customers. Run your implementation for all combinations of $\alpha \in \{5, 10, 20\}$ and $\beta \in \{1, 2, 4\}$. Visualize the results (Z matrices).

   (d) Try to interpret effects of each parameter on the results.