Dirichlet Process I

Lecturer: Prof. Michael Jordan *Scribe: Daniel Schonberg* dschonbe@eecs.berkeley.edu

22.1 Dirichlet Distribution

The Dirichlet distribution was introduced in the last lecture. Posterior inference under a Dirichlet prior will be summarized here.

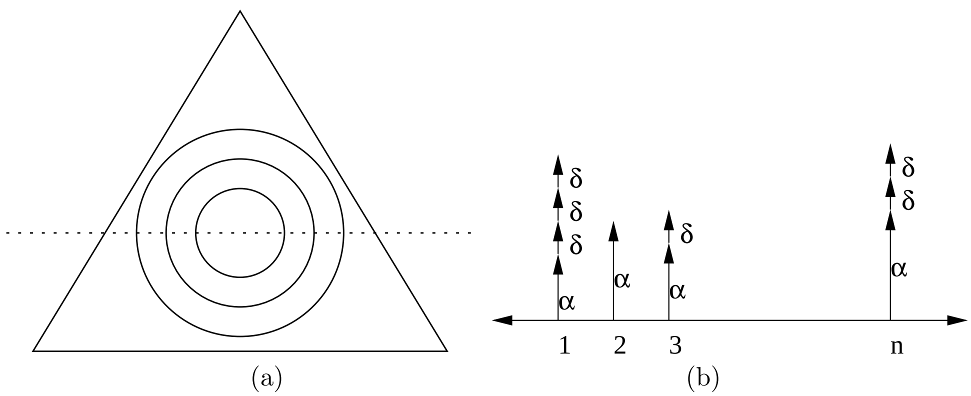

Figure 1: (a) A simplex over which the distribution can vary. (b) An initial estimate modified by sample observations.

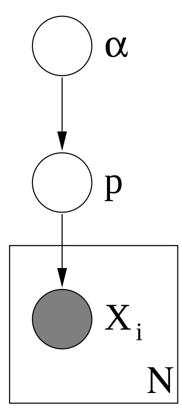

Figure 2: A graphical model depicting a Dirichlet prior and a multinomial likelihood.

We need a distribution for the parameters, which reside over a simplex, as in Figure 1(a). A graphical model is shown in Figure 2. The estimate is based upon the prior distribution and the effects of the observed data as in Figure 1(b)

$$p(p_1, \ldots, p_k | x_1, \ldots, x_n) \sim Dir(\alpha_1 + \sum_{i=1}^{n} \delta_1(x_i), \ldots, \alpha_k + \sum_{i=1}^{n} \delta_k(x_i) \tag{1}$$

$$p(x_{n+1} = j | x_1, \ldots, x_n, \alpha_1, \ldots, \alpha_n) = \frac{\alpha_j}{\alpha_+ + n} + \frac{1}{\alpha_+ + n} \sum_{i=1}^{n} \delta_j(x_i) \tag{2}$$

$$= \frac{\alpha_j}{\alpha_+} \frac{\alpha_+}{\alpha_+ + n} + \frac{n}{\alpha_+ + n} \frac{1}{n} \sum_{i=1}^{n} \delta_j(x_i) \tag{3}$$

where $\alpha_+ = \sum_{i=1}^{n} \alpha_i$.

We can compute the expected value of the p_i's as $E[p_i] = \alpha_i / \alpha_+$.

This model is related to the Polya Urn Model. In that model, we take one ball out of a bin, we then put that ball and another one of the same color back in the bin.

22.2 Dirichlet Process - View 1

In this view of the Dirichlet Process, we begin with any probability measure G_0 on the reals. The values for each of the partitions β is given according to this measure as suggested in Figure 3.

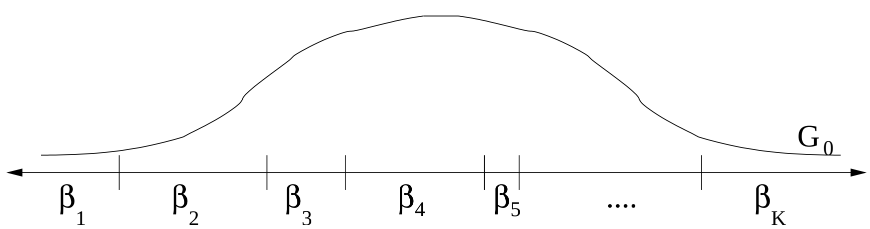

Figure 3: Here, G_0 can be any probability measure. The various β values are determined by the measure in the corresponding region.

A process is a Dirichlet process if the following equation holds for all partitions:

$$(p(\beta_1), \ldots, p(\beta_k)) \sim Dir(\alpha G_0(\beta_1), \alpha G_0(\beta_2), \ldots, \alpha G_0(\beta_k)) \tag{4}$$

We denote a sample from the Dirichlet process as $G \sim DP(\alpha, G_0)$, where α is a concentration parameter and G_0 is the base probability measure (see Figure 4).

Some properties are as follows.

- $E[G] = G_0$. This is analogous to the fact that $E[p_i] = \alpha_i / \alpha_+$.
- $p(G | \theta_1, \ldots, \theta_n) = DP(\alpha + n, \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_{\theta_i})$. This effect of the θ's on the distribution of G is suggested in Figure 5.

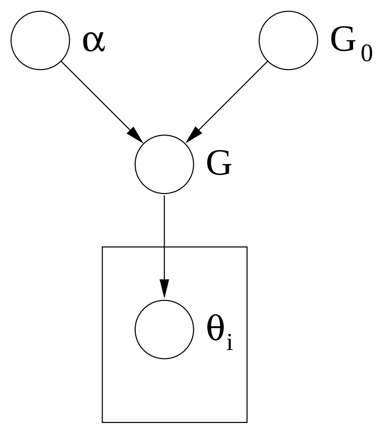

Figure 4: The graphical model for a Dirichlet Process generating the θ parameters.

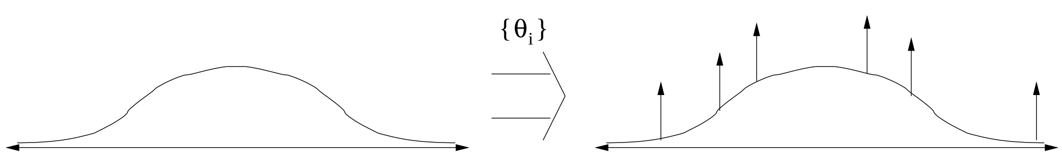

Figure 5: An example of how the observed values will change the distribution estimate.

- $E[G|\theta_1, \ldots, \theta_n] = \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_i$

We now examine the marginal probabilities for a new θ.

$$p(\theta_{n+1} = \theta_i \text{ for } 1 \leq i \leq n | \theta_1, \ldots, \theta_n, \alpha, G_0) = \frac{1}{\alpha + n} \sum_{j=1}^{n} \delta_{\theta_j}(\theta_{n+1}) \tag{5}$$

$$p(\theta_{n+1} \neq \theta_i \text{ for } 1 \leq i \leq n | \theta_1, \ldots \theta_n, \alpha, G_0) = \frac{\alpha}{\alpha + n} \tag{6}$$

These conditionals can be interpreted in terms of the so-called *Chinese restaurant process*. In this process, a restaurant possesses a countably infinite collection of empty tables. As customers arrive, they can either sit at a new table or join someone else's table. An example is shown in Figure 6.

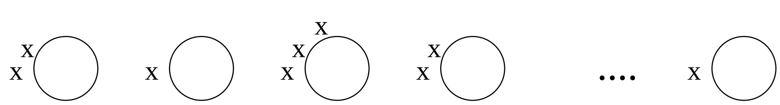

Figure 6: A diagram of the Chinese restaurant process. In this process, people arrive and are seated at tables. They either sit at a new table or join an already occupied table.

Suppose for example that $\alpha = 1$ and $n = 1$. The first customer occupies the first table. The next customer will either start a new table with probability 1/2 or join the first table with probability 1/2. After N

customers have arrived, the number of occupied tables is random and grows as $O(\log N)$. There is an isomorphism between the tables and the cycle structure of random permutations.

It turns out that the specific ordering of the customers is irrelevant; i.e., the Chinese restaurant process yields an exchangeable distribution on partitions of the integers.

Recall that an infinite set of random variables is said to be *infinitely exchangeable* if for every finite subset $\{x_1, x_2, \ldots, x_n\}$ we have

$$p(x_1, x_2, \ldots, x_n) = p(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$$

for any permutation π.

Theorem 1 (De Finetti) *A set of random variables is infinitely exchangeable if and only if*

$$p(x_1, \ldots, x_n) = \int \prod_{i=1}^{n} p(x_i|\theta) dP(\theta) \tag{7}$$

for some θ and some measure $P(\theta)$.

In particular, one possible choice for $P(\theta)$ is the Dirichlet process (where the preferred notion is "G" instead of "θ").

22.3 Dirichlet Process - View 2

The second representation we consider is the stick-breaking view. This representation gives a more concrete meaning to "$G \sim DP(\alpha, G_0)$."

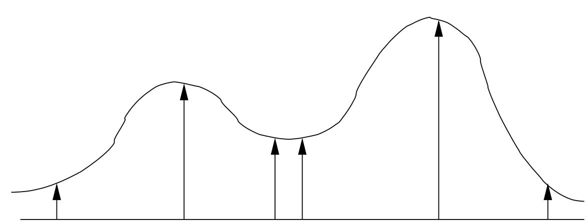

Figure 7: A sample measure and the resulting discrete Dirichlet Process.

It turns out that there is an explicit representation for samples G from the Dirichlet process. This representation is as follows:

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k} \tag{8}$$

In this equation, both the π_k and the θ_k are random. The π_k are distributed as follows:

$$\pi_k = \pi'_k \prod_{l=1}^{k-1} (1 - \pi_l), \tag{9}$$

where $\pi'_k \sim Beta(1, \alpha)$. The θ_k are distributed as $\theta_k \sim G_0$.

A diagram of the stick-breaking process is shown in Figure 8.

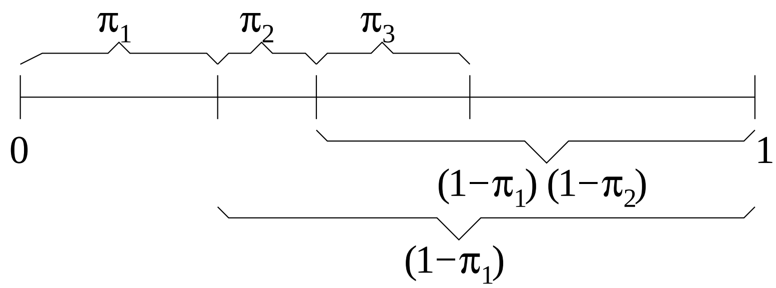

Figure 8: A diagram of the stick-breaking interpretation of the Dirichlet Process.

From Figure 8, we can obtain that $\pi_1 = \pi'_1$, $\pi_2 = \pi'_2(1 - \pi_1)$, and so on.

In the next lecture we will consider using Dirichlet processes in the context of mixture models (see Figure 9). The basic interpretation is that G is a generator of atoms that serve as parameters of mixture components in a mixture model.

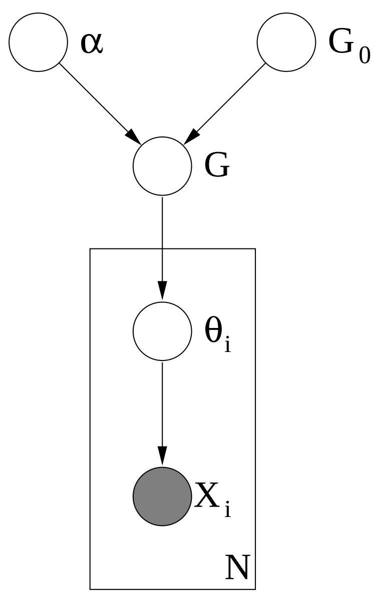

Figure 9: The graphical model for the Dirichlet Process mixture model..