# Dirichlet Process

**Xiaobao Wu**
Tsinghua University
xiaobao.wu1996@gmail.com

In general, Dirichlet Process is not as straightforward as Gaussian Process. The definition and construction are extremely opaque even after perusing. In this note, we conclude some main points of Dirichlet Process to explain it in a simple and direct way.

## 1 Motivation

Dirichlet Process is a method of Nonparametric Bayesian. One simple motivation of Dirichlet Process is to determine the number of clusters in mixture models like GMM. If the parameters of each data point are drawn from a continuous distribution, the probability is zero that two data points are from a same distribution and the number of clusters is the size of data points. Therefore, we need a discrete function to handle this issue.

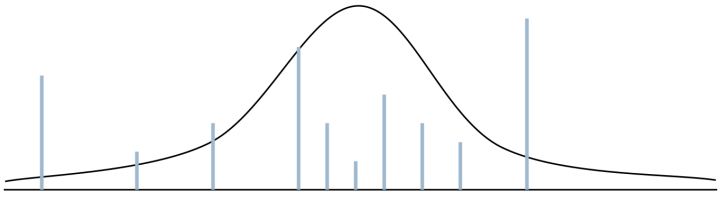

Figure 1: Illustration of $G$ and $H$ in Dirichlet Process.

## 2 Definition of Dirichlet Process

$$G \sim \mathrm{DP}(\alpha, H) \tag{1}$$

$$\Leftrightarrow (G(a_1), G(a_2), \ldots, G(a_K)) \sim \mathrm{Dir}(\alpha H(a_1), \alpha H(a_2), \ldots, \alpha H(a_K)) \tag{2}$$

$H$ is called the base function. As shown in Figure 1, $G$ can be considered as the discrete version of $H$ where $\alpha$ determines the degree. If $\alpha = 0$, only one probability measure equal to 1 exists. If $\alpha = \infty$, $G$ is $H$. $G$ is also called the random probability measure. $a_1, \ldots, a_K$ are the partitions of $H$. $G(a_i)$ means the measure sum of $G$ in partition $a_i$ and $H(a_i)$ means the measure sum of $H$ in partition $a_i$. The expectation and variance are

$$E[G(a_i)] = H(a_i) \tag{3}$$

$$Var[G(a_i)] = \frac{H(a_i)(1 - H(a_i))}{\alpha + 1} \tag{4}$$

## 3 Stick Breaking Construction

We consider the variable $\pi_i$ where

$$\theta_i \sim H \tag{5}$$
$$\beta_i \sim Beta(1, \alpha) \tag{6}$$
$$\pi_1 = \beta_1 \tag{7}$$
$$\pi_i = \beta_i \prod_{k=1}^{i-1}(1 - \beta_k) \tag{8}$$

$\pi_i$ is the probability measure of point $\theta_i$. Here we construct the distribution $G$ as

$$G(\cdot) = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i}(\cdot) \tag{9}$$

$\delta$ is the Dirac delta function, also called point mass function where $\delta_{\theta_i}(\theta) = 1$ if $\theta = \theta_i$ and 0 otherwise.

## 4 Posterior Dirichlet Process

$$G \sim DP(\alpha, H) \tag{10}$$
$$\theta_1, \ldots, \theta_N \overset{i.i.d.}{\sim} G \tag{11}$$

We need to calculate the posterior as

$$p(G|\theta_1, \ldots, \theta_N) \propto p(\theta_1, \ldots, \theta_N|G)p(G)$$

where $G = p(\theta_1, \ldots, \theta_N|G)$. However, it is inconvinent to directly use the Dirichlet Process as $p(G)$. In another way, we can turn to the conjugate property of Multinomial and Dirichlet distribution. For any partition $(a_1, \ldots, a_K)$, $n_i$ is the number of $\theta_j, j = 1, \ldots, N$ in $a_i$. We have

$$p(G(a_1), \ldots, G(a_K)|n_1, \ldots, n_K) \tag{12}$$
$$\propto Mult(n_1, \ldots, n_K|G(a_1), \ldots, G(a_K)Dir(\alpha H(a_1), \ldots, \alpha H(a_K)) \tag{13}$$
$$= Dir(\alpha H(a_1) + n_1, \ldots, \alpha H(a_2) + n_2) \tag{14}$$
$$= DP(\alpha + N, \frac{\alpha H + \sum_{i=1}^{N} \delta_{\theta_i}}{\alpha + N}) \tag{15}$$

Thus, the posterior of a Dirichlet Process is also a Dirichlet Process.

## 5 Chinese Resturant Process

$n_{l,\neg i}$ means the number of data points in class $l$ except $i$.

$$p(z_i = m|\mathbf{z}_{\neg i}) \tag{16}$$
$$= \frac{p(z_i = m, \mathbf{z}_{\neg i})}{p(\mathbf{z}_{\neg i})} \tag{17}$$
$$= \frac{\int p(z_i = m, \mathbf{z}_{\neg i}|p_1, \ldots, p_K)Dir(p_1, \ldots, p_K|\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})d(p_1, \ldots, p_K)}{\int p(\mathbf{z}_{\neg i}|p_1, \ldots, p_K)Dir(p_1, \ldots, p_K|\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})d(p_1, \ldots, p_K)} \tag{18}$$
$$= \frac{\Gamma(n_{m,\neg i} + \frac{\alpha}{K} + 1)\prod_{l=1, l\neq m}^{K}\Gamma(n_{l,\neg i} + \frac{\alpha}{K})}{\Gamma(\alpha + N)} \times \frac{\Gamma(\alpha + N - 1)}{\prod_{l=1}^{K}\Gamma(n_{l,\neg i} + \frac{\alpha}{K})} \tag{19}$$
$$= \frac{n_{m,\neg i} + \frac{\alpha}{K}}{N + \alpha - 1} \tag{20}$$
$$= \frac{n_{m,\neg i}}{N + \alpha - 1} \quad \text{(If } K \text{ is infinity)} \tag{21}$$

In CRP, if a new customer comes, he will choose an old table with the probaility $\frac{\sum_{l=1}^{K} n_{m,\neg i}}{N+\alpha-1} = \frac{N-1}{N+\alpha-1}$ and a new table with $\frac{\alpha}{N+\alpha-1}$.