Stick-Breaking Process and Dirichlet Process

Diego Garrido

Jupyter Notebook

1 Dirichlet Distribution

A multivariate generalization of the beta distribution is the Dirichlet distribution, which has support over the probability simples, defined by

$$\triangle_K = \{C : C_k \geq 0, \sum_{k=1}^{K} C_k = 1\} \tag{1}$$

The pdf is defined as follows:

$$Dir(C|\alpha) \triangleq \frac{\prod_{k=1}^{K} C_k^{\alpha_k - 1} \mathbb{I}(C \in \triangle_K)}{B(\alpha)} \tag{2}$$

In Figure 1 we have samples from a symmetric Dirichlet Distribution, lower α results in very sparse distribution, with many zeros and high variance, while higher α makes the size of the atoms more similar with higher density in the center of the simplex, note that with $\alpha = 1$ the distribution is uniform over the simplex (i.e. no region with higher density). In addition, a higher dimension results in less variance and decrease in the size of the atom because there are more dimensions over which the mass must be distributed.

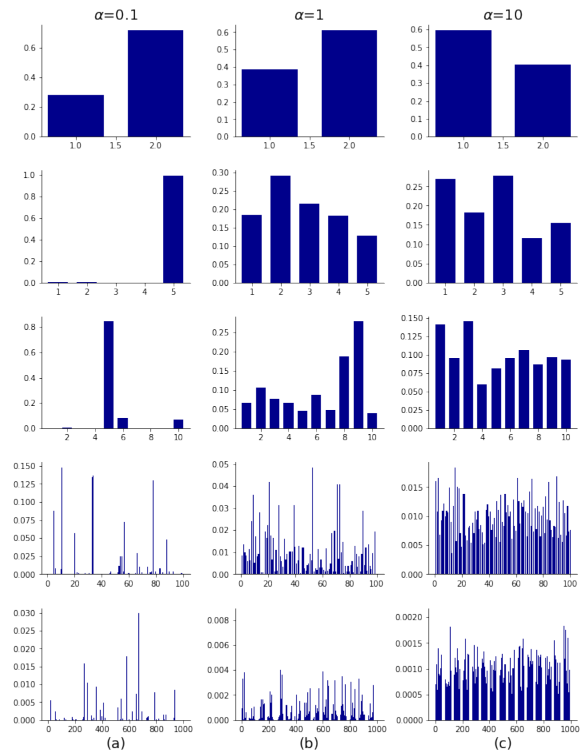

Figure 1: Samples from a symmetric Dirichlet Distribution, that is $Dir(\frac{\alpha}{K} 1_K)$, with $\alpha \in \{0.1, 1, 1\}$ and $K \in \{2, 5, 10, 100, 1000\}$.

2 Dirichlet Process

We use the stick-breaking point construction of a Dirichlet Process. If $\alpha > 0$ and if G is a probability measure on Ω_ϕ the random discrete probability measure $\Theta := \sum C_k \delta_{\Phi_k}$ generated by

$$V_1, V_2, \ldots \sim_{iid} Beta(1, \alpha) \tag{3}$$

$$C_k = V_k \prod_{j=1}^{k-1} (1 - V_k) \tag{4}$$

$$\Phi_1, \Phi_2, \ldots \sim_{iid} G_0 \tag{5}$$

is called a Dirichlet Process (DP) with base measure G and concentration paremeter $\alpha > 0$, and we denote its law by $DP(\alpha, G_0)$. In practice, when we sample from this distribution we need to add a tolerance (1e-8) so that the process ends in a reasonable number of steps.

In Figure 2 we have some samples from a stick-breaking process (left) and dirichlet process (right) with base measure $\mathcal{N}(0, 1)$ with concentration parameter $\alpha \in \{0.1, 0.2, 0.6, 6, 60\}$. The higher the alpha, the less variance and the greater number of atoms, on the contrary, small values of alpha show a high variance and a lower number of atoms, additionally they show a greater variance in the number of atoms according to the random seed.

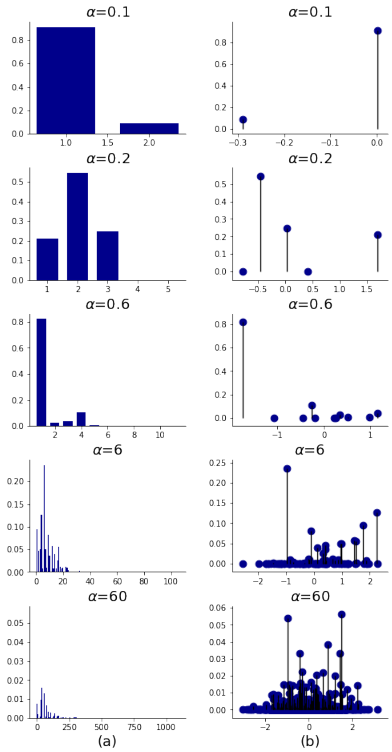

Figure 2: Random measures sample from a Dirichlet process with normal base measure $\mathcal{N}(0,1)$ with concentration parameter $\alpha \in \{0.1, 0.2, 0.6, 6, 60\}$. *(a)* Samples from stick-breaking process. *(b)* Samples from Dirichlet process with those mixture weights.

We want to sample from $\Theta \sim Dir(\alpha, G)$ by N times:

$$\Theta := \sum C_k \delta_{\Phi_k} \tag{6}$$

from the stick-breaking process we know $C_{1:K}$ and $\Phi_{1:K}$, where K is finite with probability 1, then we can sample from this discrete distribution a sequence $\phi_1, \phi_2, \dots$ by:

$$L_i \sim Mult(C) \tag{7}$$
$$\phi_i = \Phi_{L_i} \tag{8}$$

In Figure 3 we have some samples from random measures sampled from a Dirichlet Process. To higher N the possibility of repeating a previous ϕ is greater and this occurs with greater probability for smaller α, this is because the number of possibilities/atoms is less. In addition, to higher N the empirical distribution converge to the true discrete distribution.

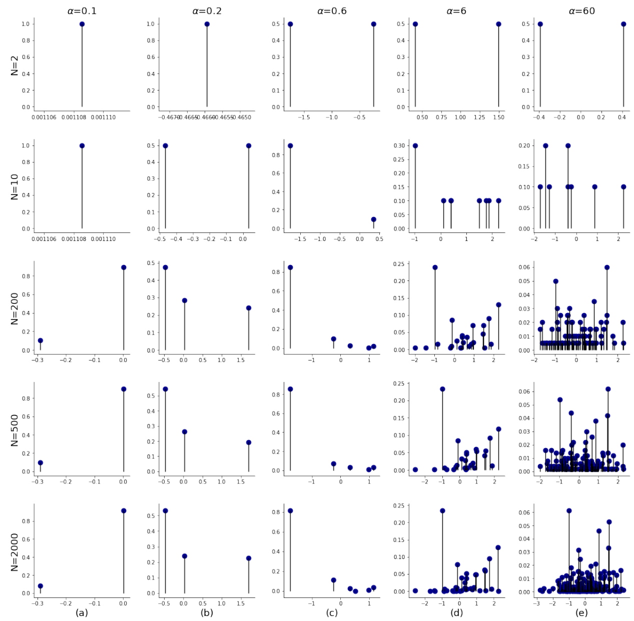

Figure 3: Samples from random measures samples from a Dirichlet Process with ormal base measure $\mathcal{N}(0,1)$ with concentration parameter $\alpha \in \{0.1, 0.2, 0.6, 6, 60\}$. The atom sizes is normalized based on the number of samples.