Dirichlet Process

Diego Garrido

Jupyter Notebook

1 Introduction

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

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

$$C_k = V_k \prod_{j=1}^{k-1} (1 - V_k) \tag{2}$$

$$\Phi_1, \Phi_2, \ldots \sim_{iid} G_0 \tag{3}$$

is called a Dirichlet Process (DP) with base measure G and concentration paremeter $\alpha > 0$, and we denote its law by $DP(\alpha, G_0)$.

Aplication

2 Implementation

To sample from a dirichlet process a base measure G_0 is required, in this case we choose a gaussian base measure with $\mu = 0$ and $\sigma \in \{1, 10\}$ for $\alpha \in \{1, 10\}$. To ensure termination in a reasonable number of steps we add a tolerance in the stick breaking process, i.e., $\sum C_k \approx 1$.

In Figure 1 we observe random measures sampled from a Dirichlet Process with normal base measure. For concentration $\alpha = 1$ the atoms exhibit high variance. For larges values of the concentration like $\alpha = 10$, we have more atoms and the sizes of the atoms become more uniform.

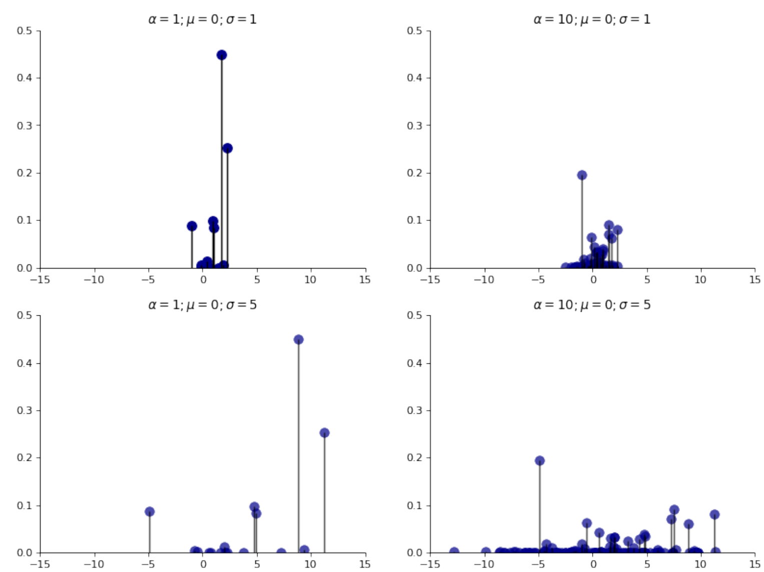

Figure 1: Random measure sampled from a Dirichlet Process with normal base measure. Height is proportional to mixture components.