# GEM Process

Caden Hewlett

2024-04-12

(separate file for easy merge)

## Implementation: GEM Distribution

In order to properly implement a DPMM, we must establish some finite approximation of $\text{DP}(\alpha\mathbb{G}_0)$. The approach implemented in the software applied in this work is known as a stick-breaking or a Griffiths, Engen, and McCloskey (GEM) process The purpose of the GEM process in terms of a Dirichlet Process is to generate weights $\{\pi_k\}$, which will be assigned to pulls from the base measure to approximate a sampled measure in the neighborhood of $\mathbb{G}_0$.

The general idea behind GEM weighing is to take a "stick" with unit length and break it at a location decided by a $\beta_1 \sim \text{beta}(1, \alpha)$ random pull, which I will denote $\pi_1$. Then, we break the remaining stick length in two by a second $\beta_2 \sim \text{beta}(1, \alpha)$ random pull. Hence, $\pi_2 = (1 - \beta_1)\beta_2$, which can be understood as "the remaining stick length after the first break, broken at the second random break location." Then, we can discretize this concept for $k = 1, 2, \ldots, K$.

$$\pi \sim \text{GEM}(\alpha) : \text{Let } \beta_k \overset{\text{iid}}{\sim} \text{beta}(1, \alpha), \text{ then } \pi_k = \beta_k \prod_{i=1}^{k-1}(1 - \beta_i)$$

Where $\pi_k$ can be considered the $k$-th value returned from the GEM distribution. The resulting realization approximates a random $K$-dimensional probability measure. For computational purposes, we treat $\text{GEM}(\alpha)$ as a discrete distribution for a reasonably large choice of $K$ since the residual distance of the sum from one quickly converges to zero as $K$ increases (Xing 2014). Critically, the sole parameter $\alpha$, known as the "concentration" controls the sampled measure variability in the finite-approximated Dirichlet Process.
A higher $\alpha$ indicates less confidence in the base measure, yielding greater dispersion and a higher number of clusters of measures about $\mathbb{G}_0$. Conversely, a lower $\alpha$ indicates more confidence in $\mathbb{G}_0$ resulting in fewer, larger clusters. (Sethuraman 1994). The finite approximation of this property can be seen in Figure 1.

Sethuraman, Jayaram. 1994. "A Constructive Definition of Dirichlet Priors." *Statistica Sinica*.

Xing, Eric P. 2014. "Hierarchical Dirichlet Processes." Carnegie Mellon University; Online. https://www.cs.cmu.edu/~epxing/Class/10708-14/scribe_notes/scribe_note_lecture20.pdf.