Reflection - VI for DP Mixtures - Blei, Jordan
Nicholas Hoernle **April 9, 2018**

1. A countably infinite number of parameters is relative. If we only observe N data points then although there can be infinitely many components, we certainly can see no more than N clusters in the data. Thus any number of parameters greater than N is effectively unbounded. In practice, the number of truncated parameters can be $<< N$ as long as it is $>> K$ (K the number of clusters in the data). So in short, we can truncate and still be nonparametric Bayesians. This idea is formalized by the *degree L weak limit approximation* to the Dirichlet process where: $GEM_L(\alpha) \triangleq Dir(\alpha/L, \ldots, \alpha/L)$ (i.e. the truncated Dirichlet distribution approximates the GEM distribution). I was interested to note the authors treatment of the difference between truncating the number of components and truncating the number of *variational* components. I did not follow their reasoning for why the later is a more reasonable assumption to make.

2. Definitely! The generative model assumes we can continue to generate data indefinitely, thus the number of clusters will grow as we generate more data. Inference is in the setting where some data has already been generated and we are trying to infer clusters from these observations.

3. The authors place a prior on α as it is not clear what concentration parameter to use. If they had more domain knowledge (about how clustered the data are), then the prior might not be necessary. The choice of the $Gamma(s_1, s_2)$ prior seems natural as it is conjugate to the $Beta(1, \alpha)$ distribution. Moreover, s_1 specifies an expected concentration parameter value and s_2 (the inverse scale parameter) can be set small for a weakly informative prior.