Supplementary Material: Memoized Online Variational Inference for Dirichlet Process Mixture Models

Michael C. Hughes and Erik B. Sudderth
Department of Computer Science, Brown University, Providence, RI 02912
mhughes@cs.brown.edu, sudderth@cs.brown.edu

Abstract

This document contains supplementary mathematics and algorithm descriptions to help readers understand our new learning algorithm. First, in Sec. 1 we offer detailed model description and update equations for a DP-GMM with zero-mean, full-covariance Gaussian likelihood. Second, in Sec. 2 we provide step-by-step discussion of our birth move algorithm, providing a level-of-detail at which the interested reader could implement our approach.

1 DP mixtures with zero-mean Gaussian observations

To review, consider the generic DP mixture model defined in the main text.

$$G \sim \mathrm{DP}(\alpha_0 H), \quad G \triangleq \sum_{k=1}^{\infty} w_k \delta_{\phi_k}, \quad v_k \sim \mathrm{Beta}(1, \alpha_0), \quad w_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell). \tag{1}$$

This process produces mixture weights w_k from a *stick-breaking* process and data-generating parameters ϕ_k from base measure H. Each data item n chooses an assignment $z_n \sim \mathrm{Cat}(w)$, and then draws observations $x_n \sim F(\phi_{z_n})$. We assume both H and F belong to exponential families

$$p(\phi_k \mid \lambda_0) = \exp\left\{\lambda_0^T t_0(\phi_k) - a_0(\lambda_0)\right\}, \qquad p(x_n \mid \phi_k) = \exp\left\{\phi_k^T t(x_n) - a(\phi_k)\right\}. \tag{2}$$

We now make this process concrete, providing the complete model and variational approximation for the particular case where observed data consists of a length-D column vector x_n and the observation model $F(x|\phi_k)$ is zero-mean Gaussian.

1.1 Zero-mean Gaussian Observation Model $F(x|\phi_k)$

For the Gaussian case, we parameterize the Gaussian likelihood $F(x|\phi_k)$ for component k by a D-length mean vector μ_k and a $D \times D$ symmetric, positive definite precision matrix Λ_k. Let $\phi_k = (\mu_k, \Lambda_k)$. For the zero-mean likelihood, we assume $\mu_k = 0$ for all k. This leaves only precision matrix Λ_k as a parameter of interest. The likelihood of x_n when assigned to component k is

$$p(x_n|z_n = k) = \mathrm{Normal}(x_n|0, \Lambda_k^{-1}) \tag{3}$$

$$\log p(x_n|z_n = k) = -\frac{D}{2}\log[2\pi] + \frac{1}{2}\log|\Lambda_k| - \frac{1}{2}x_n^T \Lambda_k x_n \tag{4}$$

$$= -\frac{D}{2}\log[2\pi] + \frac{1}{2}\log|\Lambda_k| - \frac{1}{2}\mathrm{tr}(\Lambda_k x_n x_n^T) \tag{5}$$

where $|P|$ represents the *determinant* of a square matrix P.