Truly Nonparametric Online Variational Inference for Hierarchical Dirichlet Processes: Supplementary Material

Michael Bryant and Erik B. Sudderth
Department of Computer Science, Brown University, Providence, RI
mbryantj@gmail.com, sudderth@cs.brown.edu

A. Variational objective expansion

$$\begin{aligned}
\mathcal{L}(q) \quad = \quad & \mathbb{E}_q[\log p(\mathbf{w}|\mathbf{z},\boldsymbol{\phi})] + \mathbb{E}_q[\log p(\mathbf{z}|\boldsymbol{\pi})] + \mathbb{E}_q[\log p(\boldsymbol{\pi}|\alpha\beta)] + \mathbb{E}_q[\log p(\boldsymbol{\phi}|\eta)] \\
& + \mathbb{E}_q[\log p(\beta|\gamma)] - \mathbb{E}_q[\log q(\mathbf{z}|\boldsymbol{\varphi})] - \mathbb{E}_q[\log q(\boldsymbol{\pi}|\boldsymbol{\theta})] - \mathbb{E}_q[\log q(\boldsymbol{\phi}|\boldsymbol{\lambda})] \\
= \quad & \sum_{j=1}^{D}\sum_{w=1}^{W}\sum_{k=1}^{\infty} n_{w(j)}\mathbb{E}_q[\mathbb{I}(z_{jw}=k)]\mathbb{E}_q[\log\phi_{kw}] \\
& + \sum_{j=1}^{D}\sum_{w=1}^{W}\sum_{k=1}^{\infty} n_{w(j)}\mathbb{E}_q[\mathbb{I}(z_{jw}=k)]\mathbb{E}_q[\log\pi_{jk}] \\
& + \sum_{j=1}^{D}\left\{\log\Gamma(\alpha) - \sum_k \log\Gamma\left(\alpha\beta_k\right) + \sum_k(\alpha\beta_k-1)\mathbb{E}_q[\log\pi_{jk}]\right\} \\
& + \sum_{k=1}^{\infty}\left\{\log\Gamma(W\eta) - W\log\Gamma(\eta) + \sum_w(\eta-1)\mathbb{E}_q[\log\phi_{kw}]\right\} \\
& + \left\{K\log\gamma + (\gamma-1)\log T_{K+1} - \sum_{k=1}^{K}\log T_k\right\} \\
& - \sum_{j=1}^{D}\sum_{w=1}^{W}\sum_{k=1}^{\infty} n_{w(j)}\mathbb{E}_q[\mathbb{I}(z_{jw}=k)]\log\mathbb{E}_q[\mathbb{I}(z_{jw}=k)] \\
& - \sum_{j=1}^{D}\left\{\log\Gamma\left(\sum_k\theta_{jk}\right) - \sum_k\log\Gamma(\theta_{jk}) + \sum_k(\theta_{jk}-1)\mathbb{E}_q[\log\pi_{jk}]\right\} \\
& - \sum_{k=1}^{\infty}\left\{\log\Gamma(\sum_w\lambda_{kw}) - \sum_w\log\Gamma(\lambda_{kw}) + \sum_w(\lambda_{kw}-1)\mathbb{E}_q[\log\phi_{kw}]\right\}
\end{aligned}$$

The form of the stick-breaking prior (line 5 in equation 2) comes from performing a change of variables from stick-breaking proportions to stick-breaking weights, and truncating the model at K components. The T_k are tail sums, and are defined

$$T_k \quad = \quad 1 - \sum_{l=1}^{k-1}\beta_l.$$

See the appendix of [15] for a complete derivation.