Online Variational Bayes Inference for High-Dimensional Correlated Data Review

Daeyoung Lim*
Department of Statistics
Korea University

May 28, 2016

1 Model specification

The mdoel that was proposed in this paper naturally assumes image data that consist of voxels where a voxel represents a single sample, or a data point, on a regularly spaced, three-dimensional grid. This data point can consist of a single piece of data such as an opacity or multiple pieces of data, such as a color in addition to opacity. The model is as follows:

$$\mathbf{Y}_{it} = \boldsymbol{\eta}_i \boldsymbol{\mu}_{it} + \mathbf{X}_{it}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_{it} \tag{1}$$

where

$$\begin{aligned}
\mathbf{Y}_{it} &: K \times 1 && (2)\\
\mathbf{X}_{it} &: K \times g && (3)\\
\boldsymbol{\mu}_{it} &: m \times 1 \, (m < K) && (4)\\
\boldsymbol{\eta}_{i} &: K \times m && (5)\\
\boldsymbol{\beta}_{i} &: g \times 1. && (6)
\end{aligned}$$

1.1 Priors

The priors are as follows:

$$\begin{aligned}
\boldsymbol{\mu}_{it} &\sim \mathcal{N}\left(\boldsymbol{\mu}_{i,t-1}, \theta_i^{-1}\mathbf{I}_m\right),\ \boldsymbol{\mu}_{i0} \sim \mathcal{N}\left(\boldsymbol{\mu}_0, v\mathbf{I}_m\right) && (7)\\
\boldsymbol{\epsilon}_{it} &\sim \mathcal{N}\left(\mathbf{0}, \sigma^{-2}\mathbf{I}_K\right) && (8)\\
\boldsymbol{\eta}_{ij} &\sim \mathcal{N}\left(\mathbf{0}, \tau^{-1}\left(\mathbf{I}_K - \rho\mathbf{C}\right)^{-1}\boldsymbol{\Omega}\right)^{*} && (9)\\
\left(\boldsymbol{\beta}_i, \theta_i\right) &\sim G,\ G \sim \mathrm{DP}\left(\alpha, \mathcal{N}\left(\boldsymbol{\beta}_{0r}, \boldsymbol{\Sigma}_{0r}\right)\mathrm{Ga}\left(a_\theta, b_\theta\right)\right) && (10)\\
\sigma^2 &\sim \mathrm{Ga}\left(a_\sigma, b_\sigma\right) && (11)\\
\tau &\sim \mathrm{Ga}\left(a_\tau, b_\tau\right) && (12)\\
\rho_\ell &= \frac{\ell}{M},\ \ell = 0, 1, \ldots, M-1, M-\epsilon && (13)\\
\phi_\ell &= \mathrm{P}\left(\rho = \rho_\ell\right) = \frac{1}{M+1} && (14)
\end{aligned}$$

*Prof. Taeryon Choi

in which elements in * need clarification as below.

$$\boldsymbol{\Omega} = \mathrm{Diag}\left(\frac{1}{d_{1+}}, \ldots, \frac{1}{d_{K+}}\right), \; d_{r+} = \sum_s d_{rs} \tag{15}$$

$$\mathbf{C} : K \times K, \; c_{rs} = \frac{d_{rs}}{d_{r+}}, \quad \mathbf{D} = (d_{rs}) : \text{proximity matrix} \tag{16}$$

$$d_{rs} = \begin{cases} 0 & \text{if } r = s \\ \|\varphi_r - \varphi_s\|^{-\phi}, \; \phi > 0 & \text{otherwise} \end{cases}. \tag{17}$$

2 Variational Bayes

Instead of directly dealing with the Dirichlet process itself, we will use the stick-breaking representation and reformulate it as follows. If we let $\Theta_i = \left(\boldsymbol{\beta}_i', \theta_i'\right)'$,

$$\Theta_i \sim G, \quad G = \sum_{r=1}^{\infty} \pi_r \delta_{\Theta_r^*}, \quad \pi_r = v_r \prod_{\ell<r} (1 - v_\ell), \quad \Theta_r^* = \left(\boldsymbol{\beta}_r^{*\prime}, \theta_r^{*\prime}\right)' \tag{18}$$

$$v_r \sim \mathrm{Be}(1, \alpha), \quad \alpha \sim \mathrm{Ga}(a_\alpha, b_\alpha), \boldsymbol{\beta}_r^* \sim \mathcal{N}(\boldsymbol{\beta}_{0r}, \boldsymbol{\Sigma}_{0r}), \; \theta_r^* \sim \mathrm{Ga}(a_\theta, b_\theta). \tag{19}$$

Then the parameters of interest are $\mathbf{W} = \left(V, \boldsymbol{\Theta}^*, \boldsymbol{z}, \boldsymbol{\eta}, \boldsymbol{\mu}, \rho, \tau, \sigma^2, \alpha\right)$. Then by mean-field assumption,

$$\begin{aligned} q(\mathbf{W}) = {} & q\left(\sigma^2\right) q(\alpha) q(\rho) q(\tau) \left(\prod_{r=1}^{R} q(\theta_r^*) q(\boldsymbol{\beta}_r^*) q(v_r)\right) \left(\prod_{i=1}^{n} q(z_i)\right) \\ & \times \left(\prod_{i=1}^{n} \prod_{i=1}^{T} q(\mu_{it})\right) \left(\prod_{i=1}^{n} q(\boldsymbol{\eta}_i)\right) \end{aligned}$$

where R is the truncation point of the Dirichlet process.

2.1 Updating variational distributions

The joint distribution of $\mathbf{Y}_{it}$ and $\mathbf{W}$ is

$$\begin{aligned} p(\mathbf{Y}_{it}, \mathbf{W}) = {} & \left(\prod_{i=1}^{n} \prod_{t=1}^{T} p\left(\mathbf{Y}_{it} | \boldsymbol{\mu}_{it}, \boldsymbol{\eta}_i, \sigma^2, A\right)\right) p\left(\sigma^2\right) p(\alpha) p(\rho) p(\tau) \left(\prod_{r=1}^{R} p(\theta_r^*) p(\boldsymbol{\beta}_r^*) p(v_r)\right) \left(\prod_{i=1}^{n} p(z_i)\right) \\ & \times \left(\prod_{i=1}^{n} \prod_{t=1}^{T} p(\boldsymbol{\mu}_{it})\right) \left(\prod_{i=1}^{n} p(\boldsymbol{\eta}_i)\right). \end{aligned}$$