Variational Approximation for Beta Mixture

Daeyoung Lim*
Department of Statistics
Korea University

January 31, 2016

1 Model Specification

For identification, p_i has been switched to $y_i \in (0,1)$.

$$
\begin{aligned}
f\left(y_i|s_j, m_j\right) &= \frac{\Gamma\left(s_j\right)}{\Gamma\left(s_j m_j\right)\Gamma\left(s_j\left(1-m_j\right)\right)} y_i^{s_j m_j - 1}\left(1-y_i\right)^{s_j\left(1-m_j\right)-1} \\
p\left(y|m, s, \lambda, Z\right) &= \prod_{i=1}^{N}\prod_{j=1}^{M}\left\{f\left(y_i|s_j, m_j\right)\right\}^{1\left[Z_i=j\right]} \\
p\left(m, s, \lambda, Z\right) &\propto \prod_{j=1}^{M}\left[\lambda_j^{\sum_{i=1}^{N} 1\left[Z_i=j\right]} \cdot \lambda_j^{\underline{a}-1} \cdot m_j^{\underline{n}_{m_1}-1}\left(1-m_j\right)^{\underline{n}_{m_0}-1} \cdot s_j^{\underline{a}_s - 1}\exp\left\{-s_j/\underline{b}_s\right\}\right] \\
\log p\left(y, \theta\right) &\propto \log p\left(y|m, s, \lambda, Z\right) + \log p\left(m, s, \lambda, Z\right)
\end{aligned}
$$

where $\theta = (m, s, \lambda, Z)$.

2 Coordinate Ascent

2.1 $q\left(m_i\right)$

$$
\begin{aligned}
\log q\left(m\right) &\propto \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{M} 1\left[Z_i = j\right]\log f\left(y_i|s_j, m_j\right) + \sum_{j=1}^{M}\left[\left(\underline{n}_{m_1}-1\right)\log m_j + \left(\underline{n}_{m_0}-1\right)\log\left(1-m_j\right)\right]\right] \\
&\propto \sum_{j=1}^{M}\left[\sum_{i=1}^{N} q\left(Z_i=j\right)\mathbb{E}\left[s_j\right]\log\left(\frac{y_i}{1-y_i}\right)m_j + \left(\underline{n}_{m_1}-1\right)\log m_j + \left(\underline{n}_{m_0}-1\right)\log\left(1-m_j\right)\right]
\end{aligned}
$$

Ignoring the summation with respect to j,

$$
\begin{aligned}
q\left(m_j\right) &\propto \exp\left\{\sum_{i=1}^{N} q\left(Z_i=j\right)\mathbb{E}\left[s_j\right]\log\left(\frac{y_i}{1-y_i}\right)m_j\right\} \cdot m_j^{\underline{n}_{m_1}-1}\left(1-m_j\right)^{\underline{n}_{m_0}-1} \\
&\propto \exp\left\{-C_m^q m_j\right\} m_j^{\underline{n}_{m_1}-1}\left(1-m_j\right)^{\underline{n}_{m_0}-1}, \quad C_m^q > 0
\end{aligned}
$$

*Prof. Taeryon Choi

Let

$$\mathcal{I}(a,\alpha,\beta)=\int_0^1 e^{-ax}x^{\alpha-1}(1-x)^{\beta-1}\,dx.$$

Therefore,

$$\begin{aligned}
q(m_j) &= \frac{\exp\left\{-C_m^q m_j\right\} m_j^{\underline{n}_{m_1}-1}(1-m_j)^{\underline{n}_{m_0}-1}}{\mathcal{I}\left(C_m^q,\underline{n}_{m_1},\underline{n}_{m_0}\right)} \\
C_m^q &\leftarrow -\sum_{i=1}^{N}\phi_{ij}\mathbb{E}[s_j]\log\left(\frac{y_i}{1-y_i}\right) \\
q(Z_i=j) &= \phi_{ij}.
\end{aligned}$$

2.2 $q(s_j)$

$$\begin{aligned}
\log q(s) &\propto \sum_{i=1}^{N}\sum_{j=1}^{M} 1[Z_i=j]\log f(y_i|s_j,m_j)+\sum_{j=1}^{M}(\underline{a}_s-1)\log s_j - s_j/\underline{b}_s \\
q(s_j) &\propto \exp\left\{\left(\sum_{i=1}^{N}\phi_{ij}\left(\mathbb{E}[m_j]\log y_i+(1-\mathbb{E}[m_j])\log(1-y_i)\right)-\frac{1}{\underline{b}_s}\right)s_j\right\}s_j^{\underline{a}_s-1} \\
&= \mathrm{Gamma}\left(\underline{a}_s,\beta_{s_j}\right) \\
\beta_{s_j} &\leftarrow \left(-\sum_{i=1}^{N}\phi_{ij}\left(\mathbb{E}[m_j]\log y_i+(1-\mathbb{E}[m_j])\log(1-y_i)\right)+\frac{1}{\underline{b}_s}\right)^{-1}
\end{aligned}$$

2.3 $q(\lambda)$

$$\begin{aligned}
\log q(\lambda) &\propto \sum_{j=1}^{M}\left[\left(\sum_{i=1}^{N}\phi_{ij}+\underline{a}-1\right)\log\lambda_j\right] \\
q(\lambda_j) &\propto \lambda_j^{\sum_{i=1}^{N}\phi_{ij}+\underline{a}-1}
\end{aligned}$$

Therefore, $\lambda\sim\mathrm{Dir}\left(\underline{a}_q\right)$ where the j^{th} element of $\underline{a}_q$ is given as $\sum_{i=1}^{N}\phi_{ij}+\underline{a}$.

2.4 $q(Z)$

$$
\begin{aligned}
\log q(Z) &\propto \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{M} 1[Z_i = j]\log f(y_i|s_j, m_j) + \sum_{j=1}^{M}\left[\left(\sum_{i=1}^{N} 1[Z_i = j]\right)\log \lambda_j\right]\right] \\
&\propto \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{M} 1[Z_i = j]\left((s_j m_j - 1)\log y_i + (s_j - s_j m_j - 1)\log(1 - y_i)\right) + \sum_{j=1}^{M}\sum_{i=1}^{N} 1[Z_i = j]\log \lambda_j\right] \\
&\propto \sum_{i=1}^{N}\sum_{j=1}^{M} 1[Z_i = j]\left[\log\left\{y_i^{\langle s_j\rangle\langle m_j\rangle - 1}(1 - y_i)^{\langle s_j\rangle - \langle s_j\rangle\langle m_j\rangle - 1}\langle\lambda_j\rangle\right\}\right] \\
&\propto \sum_{i=1}^{N}\sum_{j=1}^{M} \log\left[\left(\frac{y_i}{1 - y_i}\right)^{\langle s_j\rangle\langle m_j\rangle}\cdot\frac{(1 - y_i)^{\langle s_j\rangle}\langle\lambda_j\rangle}{y_i(1 - y_i)}\right]^{1[Z_i = j]} \\
q(Z) &\propto \prod_{i=1}^{N}\prod_{j=1}^{M}\left[\left(\frac{y_i}{1 - y_i}\right)^{\langle s_j\rangle\langle m_j\rangle}\cdot\frac{(1 - y_i)^{\langle s_j\rangle}\langle\lambda_j\rangle}{y_i(1 - y_i)}\right]^{1[Z_i = j]}
\end{aligned}
$$

Since we chose to denote $q(Z_i = j)$ with ϕ_{ij},

$$
\phi_{ij} \leftarrow \left(\frac{y_i}{1 - y_i}\right)^{\langle s_j\rangle\langle m_j\rangle}\cdot\frac{(1 - y_i)^{\langle s_j\rangle}\langle\lambda_j\rangle}{y_i(1 - y_i)}
$$

3 Lower Bound

$$
\begin{aligned}
\mathcal{L} &= \mathbb{E}[\log p(y, \theta)] - \mathbb{E}[\log q(\theta)] \\
&= \mathbb{E}[\log p(y|m, s, \lambda, Z)] + \mathbb{E}[\log p(m, s, \lambda, Z)] - \mathbb{E}[\log q(m, s, \lambda, Z)]
\end{aligned}
$$

3.1 $\mathbb{E}[\log p(y, \theta)]$

$$
\begin{aligned}
\mathbb{E}[\log p(y, \theta)] = &\sum_{i=1}^{N}\sum_{j=1}^{M} \phi_{ij}\langle\log\Gamma(s_j)\rangle - \langle\log\Gamma(s_j m_j)\rangle - \langle\log\Gamma(s_j - s_j m_j - 1)\rangle + (\langle s_j\rangle\langle m_j\rangle - 1)\log y_i \\
&+ (\langle s_j\rangle(1 - \langle m_j\rangle) - 1)\log(1 - y_i)
\end{aligned}
$$

- The expectation of $\log\Gamma(m_j)$ is not analytically tractable but it certainly is numerically through Monte-Carlo methods and the strong law of large numbers. The sampling of m_j is carried out through acceptance-rejection sampling method if we select the instrument density with care bearing $(0, 1)$ as its support.

$$
\mathbb{E}[g(X)] \approx \frac{1}{n}\sum_{i=1}^{n} g(X_i)
$$

- The same logic applies for the expectations of $\log\Gamma(s_j m_j)$ and $\log\Gamma(s_j - s_j m_j - 1)$.
- The expectation of m_j is as follows:

$$
\langle m_j\rangle = \frac{\mathcal{I}\left(C_m^q, \underline{n}_{m_1} + 1, \underline{n}_{m_0}\right)}{\mathcal{I}\left(C_m^q, \underline{n}_{m_1}, \underline{n}_{m_0}\right)}
$$

3.2 $\mathbb{E}\left[\log p\left(m, s, \lambda, Z\right)\right]$

For the prior distributions, normalizing constants of these distributions are of no use since they do not change in every iteration which also indicates they do not contribute to the lower bound at all. We will thus only consider the following:

$$
\begin{aligned}
\mathbb{E}\left[\log p\left(\theta\right)\right] \propto \sum_{j=1}^{M}\Bigg[&\sum_{i=1}^{N}\phi_{ij}\left\langle\log\lambda_j\right\rangle + \left(\underline{a}-1\right)\left\langle\log\lambda_j\right\rangle + \left(\underline{n}_{m_1}-1\right)\left\langle\log m_j\right\rangle + \left(\underline{n}_{m_0}-1\right)\left\langle\log\left(1-m_j\right)\right\rangle \\
&+\left(\underline{a}_s-1\right)\left\langle\log s_j\right\rangle - \left\langle s_j\right\rangle/\underline{b}_s\Bigg]
\end{aligned}
$$

- The variational distribution of λ is dirichlet distribution with the j^{th} element of its parameter vector $\underline{a}_q$ is $\sum_{i=1}^{N}\phi_{ij}+\underline{a}$. We know that the marginal distribution of λ_j follows beta distribution. We first present a general formulation of the theorem. Let $X=\left(X_1,\ldots,X_K\right)\sim\text{Dir}\left(\alpha\right)$ where $\alpha_0=\sum_{i=1}^{K}\alpha_i$. Then,

$$
\begin{aligned}
\mathbb{E}\left[X_i\right] &= \frac{\alpha_i}{\alpha_0} \\
\text{Var}\left[X_i\right] &= \frac{\alpha_i\left(\alpha_0-\alpha_i\right)}{\alpha_0^2\left(\alpha_0+1\right)} \\
\text{Cov}\left[X_i,X_j\right] &= \frac{-\alpha_i\alpha_j}{\alpha_0^2\left(\alpha_0+1\right)}, \quad \text{if } i\neq j \\
X_i &\sim \text{Beta}\left(\alpha_i,\alpha_0-\alpha_i\right)
\end{aligned}
$$

- The above item suggests that

$$
\lambda_j\sim\text{Beta}\left(\sum_{i=1}^{N}\phi_{ij}+\underline{a},\sum_{j=1}^{M}\sum_{i=1}^{N}\phi_{ij}-\sum_{i=1}^{N}\phi_{ij}+\left(M-1\right)\underline{a}\right).
$$

- Let $Y\sim\text{Beta}\left(\alpha,\beta\right)$.

$$
\mathbb{E}\left[\log Y\right]=\varphi\left(\alpha\right)-\varphi\left(\alpha+\beta\right)
$$

  where φ is the digamma function.

- Thus

$$
\text{E}\left[\log\lambda_j\right]=\varphi\left(\sum_{i=1}^{N}\phi_{ij}+\underline{a}\right)-\varphi\left(\sum_{j=1}^{M}\sum_{i=1}^{N}\phi_{ij}+M\underline{a}\right).
$$

3.3 $\mathbb{E}\left[\log q\left(m, s, \lambda, Z\right)\right]$

3.3.1 $\mathbb{E}\left[\log q\left(m\right)\right]$

$$
q\left(m_j\right)=\frac{\exp\left\{-C_m^q m_j\right\}m_j^{\underline{n}_{m_j}-1}\left(1-m_j^{\underline{n}_{m_0}-1}\right)}{\mathcal{I}\left(C_m^q,\underline{n}_{m_1},\underline{n}_{m_0}\right)}
$$

$$
\begin{aligned}
\log q\left(m_j\right) &= -C_m^q m_j+\left(\underline{n}_{m_1}-1\right)\log m_j+\left(\underline{n}_{m_0}-1\right)\log\left(1-m_j\right)-\log\mathcal{I}\left(C_m^q,\underline{n}_{m_1},\underline{n}_{m_0}\right) \\
\mathbb{E}\left[\log q\left(m_j\right)\right] &= -C_m^q\left\langle m_j\right\rangle+\left(\underline{n}_{m_1}-1\right)\left\langle\log m_j\right\rangle+\left(\underline{n}_{m_0}-1\right)\left\langle\log\left(1-m_j\right)\right\rangle-\log\mathcal{I}\left(C_m^q,\underline{n}_{m_1},\underline{n}_{m_0}\right) \\
\left\langle m_j\right\rangle &= \frac{\mathcal{I}\left(C_m^q,\underline{n}_{m_1}+1,\underline{n}_{m_0}\right)}{\mathcal{I}\left(C_m^q,\underline{n}_{m_1},\underline{n}_{m_0}\right)}
\end{aligned}
$$