VARIATIONAL INFERENCE ON RVD2 MODEL

Y. HE

1. Model Structure

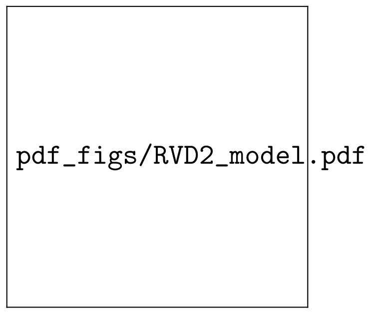

Figure 1. RVD2 Graphical Model

The joint distribution over the latent and observed variables for data at location j in replicate i given the parameters is

$$p\left(r_{ji}, \theta_{ji}, \mu_j | n_{ji}; \mu_0, M_0, M_j\right) = p\left(r_{ji} | \theta_{ji}, n_{ji}\right) p\left(\theta_{ji} | \mu_j; M_j\right) p\left(\mu_j; \mu_0, M_0\right), \tag{1}$$

where

$$p\left(\mu_j; \mu_0, M_0\right) = \frac{\Gamma(M_0)}{\Gamma(\mu_0 M_0)\Gamma(M_0(1-\mu_0))} \mu_j^{M_0\mu_0 - 1}(1-\mu_j)^{M_0(1-\mu_0)-1},$$
$$p\left(\theta_{ji} | \mu_j; M_j\right) = \frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))} \theta_{ji}^{M_j\mu_j - 1}(1-\theta_{ji})^{M_j(1-\mu_j)-1},$$
$$p\left(r_{ji} | \theta_{ji}, n_{ji}\right) = \frac{\Gamma(n_{ji}+1)}{\Gamma(r_{ji}+1)\Gamma(n_{ji}-r_{ji}+1)} \theta_{ji}^{r_{ji}}(1-\theta_{ji})^{n_{ji}-r_{ji}}.$$

Integrating over the latent variables θ_{ji} and μ_j yields the marginal distribution of the data,

$$p\left(r_{ji} | n_{ji}; \mu_0, M_0, M_j\right) = \int_{\mu_j} \int_{\theta_{ji}} p\left(r_{ji} | \theta_{ji}, n_{ji}\right) p\left(\theta_{ji} | \mu_j; M_j\right) p\left(\mu_j; \mu_0, M_0\right) d\theta_{ji} d\mu_j. \tag{2}$$

Finally, the log-likelihood of the data set is

$$\log p\left(r|n;\mu_0,M_0,M\right)=\sum_{j=1}^{J}\sum_{i=1}^{N}\log\int_{\mu_j}\int_{\theta_{ji}}p\left(r_{ji}|\theta_{ji},n_{ji}\right)p\left(\theta_{ji}|\mu_j;M_j\right)p\left(\mu_j;\mu_0,M_0\right)d\theta_{ji}d\mu_j. \tag{3}$$

2. Variational Inference

2.1. **Factorization.** We propose the following factorized variational distribution to approximate the true posterior over latent variables:

$$q(\mu,\theta)=q(\mu)q(\theta)=\prod_{j=1}^{J}q(\mu_j)\prod_{i=1}^{N}q(\theta_{ji}). \tag{4}$$

2.2. **Derivation of $q(\mu)$ and $q(\theta)$.**

2.2.1. *Derivation of $q(\theta)$.*

$$\begin{aligned}
\log q_\theta^*(\theta) &= E_\mu\left[\log p\left(r,\mu,\theta|n;\phi\right)\right]\\
&= E_\mu\left[\log p\left(r|\theta,n\right)\right]+E_\mu\left[\log p\left(\theta|\mu;M\right)\right]+E_\mu\left[\log p\left(\mu;\mu_0,M_0\right)\right]\\
&= \sum_{j=1}^{J}\sum_{i=1}^{N}E_\mu\left[\log p\left(r_{ji}|\theta_{ji},n_{ji}\right)\right]+\sum_{j=1}^{J}\sum_{i=1}^{N}E_\mu\left[\log p\left(\theta_{ji}|\mu_j;M_j\right)\right]+con.\\
&= \sum_{j=1}^{J}\sum_{i=1}^{N}E_\mu\left[\log\left(\frac{\Gamma(n_{ji}+1)}{\Gamma(r_{ji}+1)\Gamma(n_{ji}-r_{ji}+1)}\theta_{ji}^{r_{ji}}(1-\theta_{ji})^{n_{ji}-r_{ji}}\right)\right]\\
&\quad+\sum_{j=1}^{J}\sum_{i=1}^{N}E_\mu\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_jM_j)\Gamma(M_j(1-\mu_j))}\theta_{ji}^{M_j\mu_j-1}(1-\theta_{ji})^{M_j(1-\mu_j)-1}\right)\right]+con.\\
&= \sum_{j=1}^{J}\sum_{i=1}^{N}E_\mu\left[\log\left(\theta_{ji}^{r_{ji}+M_j\mu_j-1}(1-\theta_{ji})^{n_{ji}-r_{ji}+M_j(1-\mu_j)-1}\right)\right]+con.\\
&= \sum_{j=1}^{J}\sum_{i=1}^{N}\left[\log\left(\theta_{ji}^{r_{ji}+M_jE_\mu[\mu_j]-1}(1-\theta_{ji})^{n_{ji}-r_{ji}+M_j(1-E_\mu[\mu_j])-1}\right)\right]+con.
\end{aligned} \tag{5}$$

Exponentiating both sides, we can see that $q_\theta^*(\theta)$ is a product of beta distributions.

2.2.2. *Derivation of $q(\mu)$.*

$$
\begin{aligned}
\log q_\mu^*(\mu) &= E_\theta \left[\log p\left(r, \mu, \theta | n; \phi\right) \right] \\
&= E_\theta \left[\log p\left(r | \theta, n\right) \right] + E_\theta \left[\log p\left(\theta | \mu; M\right) \right] + E_\theta \left[\log p\left(\mu; \mu_0, M_0\right) \right] \\
&= \sum_{j=1}^{J} \sum_{i=1}^{N} E_\theta \left[\log p\left(\theta_{ji} | \mu_j; M_j\right) \right] + \sum_{j=1}^{J} E_\theta \left[\log p\left(\mu_j; \mu_0, M_0\right) \right] + con. \\
&= \sum_{j=1}^{J} \sum_{i=1}^{N} E_\theta \left[\log \left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))} \theta_{ji}^{M_j \mu_j - 1} (1-\theta_{ji})^{M_j(1-\mu_j)-1} \right) \right] \\
&\quad + \sum_{j=1}^{J} E_\theta \left[\log \left(\frac{\Gamma(M_0)}{\Gamma(\mu_0 M_0)\Gamma(M_0(1-\mu_0))} \mu_j^{M_0 \mu_0 - 1} (1-\mu_j)^{M_0(1-\mu_0)-1} \right) \right] + con. \\
&= \sum_{j=1}^{J} \log \left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))} \right) \\
&\quad + \sum_{j=1}^{J} \sum_{i=1}^{N} \left\{ \left(M_j \mu_j - 1\right) E_\theta \left[\log \theta_{ji} \right] + \left(M_j(1-\mu_j) - 1\right) E_\theta \left[\log\left(1-\theta_{ji}\right) \right] \right\} \\
&\quad + \sum_{j=1}^{J} \left\{ \left(M_0 \mu_0 - 1\right) \log \mu_j + \left(M_0(1-\mu_0) - 1\right) \log(1-\mu_j) \right\} + con.
\end{aligned} \tag{6}
$$

It can be seen that the distribution function of $q_\mu^*(\mu)$ is not any known distribution. We propose to approximate $q_\mu^*(\mu)$ using Beta distribution.

$$
\mu_j \sim \text{Beta}(\delta_j, \gamma_j) \tag{7}
$$

2.3. **ELBO compute.** Using Jensen's inequality, the log-likelihood of the data is lower-bounded:

$$
\begin{aligned}
\log p\left(r | \phi\right) &= \log \int_\mu \int_\theta p\left(r, \mu, \theta\right) d\theta d\mu \\
&= \log \int_\mu \int_\theta p\left(r, \mu, \theta\right) \frac{q\left(\mu, \theta\right)}{q\left(\mu, \theta\right)} d\theta d\mu \\
&\geq \int_\mu \int_\theta q\left(\mu, \theta\right) \log \frac{p\left(r, \mu, \theta\right)}{q\left(\mu, \theta\right)} d\theta d\mu \\
&= E_q \left[\log p\left(r, \mu, \theta\right) \right] - E_q \left[\log q\left(\mu, \theta\right) \right] \\
&\triangleq \mathcal{L}(q, \phi),
\end{aligned} \tag{8}
$$

where $\phi = (\mu_0, M_0, M)$

Actually,

$$
\begin{aligned}
\log p\left(r|\phi\right) &= \log \int_\mu \int_\theta p\left(r,\mu,\theta\right) d\theta d\mu \\
&= \log \int_\mu \int_\theta p\left(r,\mu,\theta\right) \frac{q\left(\mu,\theta\right)}{q\left(\mu,\theta\right)} d\theta d\mu \\
&= \int_\mu \int_\theta q\left(\mu,\theta\right) \log \frac{p\left(r,\mu,\theta\right)}{q\left(\mu,\theta\right)} d\theta d\mu - \int_\mu \int_\theta q\left(\mu,\theta\right) \log \frac{p\left(\mu,\theta|r\right)}{q\left(\mu,\theta\right)} d\theta d\mu \\
&= \mathcal{L}(q,\phi) + KL\left(q(\mu,\theta)||p(\mu,\theta|r)\right),
\end{aligned}
\tag{9}
$$

Maximizing the ELBO is equivalent to minimizing the KL-divergence between the variational distribution and the true posterior.

Writing out ELBO $\mathcal{L}(q,\phi)$, we have

$$
\begin{aligned}
\mathcal{L}(q,\phi) &= E_q\left[\log p\left(r,\mu,\theta|n;\phi\right)\right] - E_q\left[\log q\left(\mu,\theta\right)\right] \\
&= E_q\left[\log p\left(r|\theta,n\right)\right] + E_q\left[\log p\left(\theta|\mu;M\right)\right] + E_q\left[\log p\left(\mu;\mu_0,M_0\right)\right] - E_q\left[\log q\left(\mu\right)\right] - E_q\left[\log q\left(\theta\right)\right]
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
E_q\left[\log p\left(r|\theta,n\right)\right] &= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log p\left(r_{ji}|\theta_{ji},n_{ji}\right)\right] \\
&= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log\left(\frac{\Gamma(n_{ji}+1)}{\Gamma(r_{ji}+1)\Gamma(n_{ji}-r_{ji}+1)}\theta_{ji}^{r_{ji}}(1-\theta_{ji})^{n_{ji}-r_{ji}}\right)\right] \\
&= \sum_{j=1}^{J}\sum_{i=1}^{N} \log\left(\frac{\Gamma(n_{ji}+1)}{\Gamma(r_{ji}+1)\Gamma(n_{ji}-r_{ji}+1)}\right) \\
&\quad + \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[r_{ji}\log\theta_{ji} + (n_{ji}-r_{ji})\log(1-\theta_{ji})\right] \\
&= \sum_{j=1}^{J}\sum_{i=1}^{N} \log\left(\frac{\Gamma(n_{ji}+1)}{\Gamma(r_{ji}+1)\Gamma(n_{ji}-r_{ji}+1)}\right) \\
&\quad + \sum_{j=1}^{J}\sum_{i=1}^{N} \left\{r_{ji}E_q\left[\log\theta_{ji}\right] + (n_{ji}-r_{ji})E_q\left[\log(1-\theta_{ji})\right]\right\}
\end{aligned}
\tag{11}
$$

$$
\begin{aligned}
E_q\left[\log p\left(\theta|\mu;M\right)\right] &= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log p\left(\theta_{ji}|\mu_j;M_j\right)\right] \\
&= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\theta_{ji}^{M_j\mu_j-1}(1-\theta_{ji})^{M_j(1-\mu_j)-1}\right)\right] \\
&= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right] + \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log\left(\theta_{ji}^{M_j\mu_j-1}(1-\theta_{ji})^{M_j(1-\mu_j)-1}\right)\right] \\
&= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right] \\
&\quad + \sum_{j=1}^{J}\sum_{i=1}^{N}\left\{E_q\left[(M_j\mu_j-1)\log\theta_{ji}\right] + E_q\left[(M_j(1-\mu_j)-1)\log\left(1-\theta_{ji}\right)\right]\right\} \\
&= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right] \\
&\quad + \sum_{j=1}^{J}\sum_{i=1}^{N}\left\{M_j E_q\left[\mu_j\right] E_q\left[\log\theta_{ji}\right] - E_q\left[\log\theta_{ji}\right] + \left(M_j - 1 - M_j E_q\left[\mu_j\right]\right) E_q\left[\log\left(1-\theta_{ji}\right)\right]\right\} \\
&= N * \sum_{j=1}^{J} E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right] \\
&\quad + \sum_{j=1}^{J}\sum_{i=1}^{N}\left\{M_j E_q\left[\mu_j\right] E_q\left[\log\theta_{ji}\right] - E_q\left[\log\theta_{ji}\right] + \left(M_j - 1 - M_j E_q\left[\mu_j\right]\right) E_q\left[\log\left(1-\theta_{ji}\right)\right]\right\}
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
E_q\left[\log p\left(\mu;\mu_0,M_0\right)\right] &= \sum_{j=1}^{J} E_q\left[\log p\left(\mu_j;\mu_0,M_0\right)\right] \\
&= \sum_{j=1}^{J} E_q\left[\log\left(\frac{\Gamma(M_0)}{\Gamma(\mu_0 M_0)\Gamma(M_0(1-\mu_0))}\mu_j^{M_0\mu_0-1}(1-\mu_j)^{M_0(1-\mu_0)-1}\right)\right] \\
&= \sum_{j=1}^{J} \log\frac{\Gamma(M_0)}{\Gamma(\mu_0 M_0)\Gamma(M_0(1-\mu_0))} \\
&\quad + \sum_{j=1}^{J}\left\{(M_0\mu_0-1)E_q\left[\log\mu_j\right]+(M_0(1-\mu_0)-1)E_q\left[\log(1-\mu_j)\right]\right\} \\
&= J * \log\frac{\Gamma(M_0)}{\Gamma(\mu_0 M_0)\Gamma(M_0(1-\mu_0))} \\
&\quad + \sum_{j=1}^{J}\left\{(M_0\mu_0-1)E_q\left[\log\mu_j\right]+(M_0(1-\mu_0)-1)E_q\left[\log(1-\mu_j)\right]\right\}
\end{aligned} \tag{13}
$$

Therefore, in order to compute ELBO, we need to compute the following expectations with respect to variational distribution: $E_q\left[\log\theta_{ji}\right]$, $E_q\left[\log\left(1-\theta_{ji}\right)\right]$, $E_q\left[\log\mu_j\right]$, $E_q\left[\log(1-\mu_j)\right]$, $E_q\left[\mu_j\right]$ and $E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right]$.

From

$$
\begin{aligned}
\theta_{ji} &\sim \text{Beta}(\alpha_{ji},\beta_{ji}) \\
\mu_j &\sim \text{Beta}(\delta_j,\gamma_j)
\end{aligned}
$$

We know

$$
\begin{aligned}
E_q\left[\log\theta_{ji}\right] &= \psi(\alpha_{ji})-\psi(\alpha_{ji}+\beta_{ji}) \\
E_q\left[\log\left(1-\theta_{ji}\right)\right] &= \psi(\beta_{ji})-\psi(\alpha_{ji}+\beta_{ji}) \\
E_q\left[\mu_j\right] &= \frac{\delta_j}{\delta_j+\gamma_j} \\
E_q\left[\log\mu_j\right] &= \psi(\delta_j)-\psi(\delta_j+\gamma_j) \\
E_q\left[\log(1-\mu_j)\right] &= \psi(\gamma_j)-\psi(\delta_j+\gamma_j)
\end{aligned} \tag{14}
$$

There is no analytical representation for $E_q\left[\log\Gamma(\mu_j M_j)\right]$ and $E_q\left[\log\Gamma(M_j(1-\mu_j))\right]$. Therefore, we propose to use trapezoidal numerical integration to approximate these two expectations.

Moreover, according to the entropy of beta distribution random variable,

$$\begin{aligned}E_q\left[\log q\left(\mu\right)\right] &= \sum_{j=1}^{J} E_q\left[\log q(\mu_j)\right]\\ &= -\sum_{j=1}^{J}\left\{\log(B(\delta_j,\gamma_j)) - (\delta_j-1)\psi(\delta_j) - (\gamma_j-1)\psi(\gamma_j) + (\delta_j+\gamma_j-2)\psi(\delta_j+\gamma_j)\right\}\end{aligned} \tag{15}$$

$$\begin{aligned}E_q\left[\log q\left(\theta\right)\right] &= \sum_{j=1}^{J}\sum_{i=1}^{N} E_q\left[\log q(\theta_{ji})\right]\\ &= -\sum_{j=1}^{J}\sum_{i=1}^{N}\\ &\quad \left\{\log(B(\alpha_{ji},\beta_{ji})) - (\alpha_{ji}-1)\psi(\alpha_{ji}) - (\beta_{ji}-1)\psi(\beta_{ji}) + (\alpha_{ji}+\beta_{ji}-2)\psi(\alpha_{ji}+\beta_{ji})\right\}\end{aligned} \tag{16}$$

2.4. **Optimizing Model Parameters $\phi = \{\mu_0, M_0, M\}$.**

2.4.1. *Optimizing μ_0.* The ELBO with respect to μ_0 is

$$\mathcal{L}_{[\mu_0]} = -J*\log\Gamma(\mu_0 M_0) - J*\log\Gamma(M_0(1-\mu_0)) + M_0\mu_0\sum_{j=1}^{J}\left\{E_q\left[\log\mu_j\right] - E_q\left[\log(1-\mu_j)\right]\right\}. \tag{17}$$

Take the derivative with respect to μ_0 and set it equal to zero,

$$\mathcal{L}'_{[\mu_0]} = -J*M_0\psi(\mu_0 M_0) + J*M_0\psi(M_0(1-\mu_0)) + M_0\sum_{j=1}^{J}\left\{E_q\left[\log\mu_j\right] - E_q\left[\log(1-\mu_j)\right]\right\} = 0, \tag{18}$$

the update for μ_0 can be numerically computed.

2.4.2. *Optimizing M_0.* The ELBO with respect to M_0 is

$$\mathcal{L}_{[M_0]} = J * \log \frac{\Gamma(M_0)}{\Gamma(\mu_0 M_0)\Gamma(M_0(1-\mu_0))} + M_0 \sum_{j=1}^{J} \{\mu_0 E_q [\log \mu_j] + (1-\mu_0) E_q [\log(1-\mu_j)]\} \quad (19)$$

Take the derivative with respect to M_0 and set it equal to zero,

$$\begin{aligned}\mathcal{L}'_{[M_0]} &= \log \frac{\Gamma(M_0)}{\Gamma(\mu_0 M_0)\Gamma(M_0(1-\mu_0))} + M_0 \sum_{j=1}^{J} \{\mu_0 E_q [\log \mu_j] + (1-\mu_0) E_q [\log(1-\mu_j)]\} \\ &= \psi(M_0) - \mu_0 \psi(\mu_0 M_0) - (1-\mu_0)\psi(M_0(1-\mu_0)) \\ &+ \sum_{j=1}^{J} \{\mu_0 E_q [\log \mu_j] + (1-\mu_0) E_q [\log(1-\mu_j)]\} \\ &= 0\end{aligned} \quad (20)$$

the update for M_0 can be numerically computed.

2.4.3. *Optimizing M.*

$$\begin{aligned}\mathcal{L}_{[M]} = &\sum_{j=1}^{J} E_q \left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right] \\ &+ M_j \sum_{j=1}^{J} \sum_{i=1}^{N} \{E_q [\mu_j] E_q [\log \theta_{ji}] + (1 - E_q [\mu_j]) E_q [\log(1-\theta_{ji})]\}\end{aligned} \quad (21)$$

uppose

$$f(\mu) = \log\left(\frac{\Gamma(M)}{\Gamma(\mu M)\Gamma(M(1-\mu))}\right)$$

then

$$\begin{aligned}f'(\mu) &= -M\psi(\mu M) + M\psi(M(1-\mu)) \\ f''(\mu) &= -M^2\psi'(\mu M) - M^2\psi'(M(1-\mu)) < 0\end{aligned} \quad (22)$$

where $\psi(\mu)$ is the Digamma function, and $\psi'(\mu) = \frac{\partial \psi(\mu)}{\partial \mu}$ is the Trigamma function. As trigamma function $\psi'(\mu)$ is positive, $f''(\mu)$ is negative. Thus, $f(\mu)$ is a concave function. We can approximate $f(\mu)$ using first-order Taylor expansion around point μ°, which is

$$\begin{aligned} f(\mu) &\leq f(\mu^\circ) + f'(\mu^\circ) \cdot (\mu - \mu^\circ) \\ &= \log\left(\frac{\Gamma(M)}{\Gamma(\mu^\circ M)\Gamma(M(1-\mu^\circ))}\right) + (-M\psi(\mu^\circ M) + M\psi(M(1-\mu^\circ))) \cdot (\mu - \mu^\circ). \end{aligned}$$

A upper bound approximation for $E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right]$ around point μ_j° can be represented as

$$\begin{aligned} E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right] &\leq \log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j^\circ M_j)\Gamma(M_j(1-\mu_j^\circ))}\right) \\ &\quad + \left(-M_j\psi(\mu_j^\circ M_j) + M_j\psi(M_j(1-\mu_j^\circ))\right) \cdot (E_q(\mu_j) - \mu_j^\circ). \end{aligned}$$

The equality holds if and only if $\mu_j^\circ = E_q(\mu_j)$. Therefore, at this particular point,

$$E_q\left[\log\left(\frac{\Gamma(M_j)}{\Gamma(\mu_j M_j)\Gamma(M_j(1-\mu_j))}\right)\right] = \log\left(\frac{\Gamma(M_j)}{\Gamma(E_q(\mu_j)M_j)\Gamma(M_j(1-E_q(\mu_j)))}\right).$$

Then

$$\begin{aligned} \mathcal{L}_{[M]} &= \sum_{j=1}^{J} \log\left(\frac{\Gamma(M_j)}{\Gamma(E_q(\mu_j)M_j)\Gamma(M_j(1-E_q(\mu_j)))}\right) \\ &\quad + M_j \sum_{j=1}^{J}\sum_{i=1}^{N} \left\{E_q[\mu_j]\, E_q[\log\theta_{ji}] + (1-E_q[\mu_j])\, E_q[\log(1-\theta_{ji})]\right\} \end{aligned} \tag{23}$$

The partial derivative is

$$\begin{aligned} \frac{\partial \mathcal{L}_{[M]}}{\partial\, M_j} &= \psi(M_j) - E_q(\mu_j)\psi(E_q(\mu_j)M_j) - (1-E_q(\mu_j))\psi((1-E_q(\mu_j))M_j) \\ &\quad + \sum_{j=1}^{J}\sum_{i=1}^{N} \left\{E_q[\mu_j]\, E_q[\log\theta_{ji}] + (1-E_q[\mu_j])\, E_q[\log(1-\theta_{ji})]\right\} \end{aligned} \tag{24}$$

the update for M_j can be numerically computed.

Algorithm 1 RVD2 Variational Inference

Initialize $q(\theta, \mu)$ and $\hat{\phi}$
repeat
  repeat
    for j = 1 to J **do**
      for i = 1 to N **do**
        Optimize $\mathcal{L}(q, \hat{\phi})$ over $q(\theta_{ji}; \delta_{ji}) = \text{Beta}(\delta_{ji})$
      end for
    end for
    for j = 1 to J **do**
      Optimize $\mathcal{L}(q, \hat{\phi})$ over $q(\mu_j; \gamma_j) = \text{Beta}(\gamma_j)$
    end for
  until change in $\mathcal{L}(q, \hat{\phi})$ is small
  Set $\hat{\phi} \leftarrow \arg\max_{\phi} \mathcal{L}(q, \phi)$
until change in $\mathcal{L}(q, \hat{\phi})$ is small