Supplementary Material: Effective Monte Carlo Variational Inference for Binary-Variable Probabilistic Programs

Geng Ji
Computer Science Department, UC Irvine

Erik B. Sudderth
Computer Science Department, UC Irvine

A Data-augmented Variational Inference for Latent Feature Relational Model

The mean-field variational distribution for the latent features is $q(z) = \prod_{i=1}^{N} \prod_{d=1}^{D} q(z_{id})$, in which N is the number of entities and D is the feature dimension. Each dimension z_{id} is a Bernoulli distribution $q(z_{id}) \sim \text{Bernoulli}(q_{id})$, with the activation probability $q_{id} \triangleq q(z_{id} = 1)$ as the free parameter.

The regular ELBO (Eq. (4) in the main paper) of the latent feature relational model discussed in Sec. 2 is

$$\begin{aligned}
&\mathcal{L}(q(z)) \qquad \text{(A.1)}\\
&= \sum_{i=1}^{N} \mathbb{E}_{q(z)} \Bigg[\sum_{d=1}^{D} z_{id} \log \rho + (1 - z_{id}) \log(1 - \rho) \\
&\quad + \sum_{j>i}^{N} x_{ij} \log \Phi\Big(w_0 + \sum_{d=1}^{D} w_d z_{id} z_{jd} \Big) \\
&\quad + \sum_{j>i}^{N} (1 - x_{ij}) \log \Big(1 - \Phi\Big(w_0 + \sum_{d=1}^{D} w_d z_{id} z_{jd} \Big) \Big) \\
&\quad - \sum_{d=1}^{D} z_{id} \log q_{id} + (1 - z_{id}) \log(1 - q_{id}) \Bigg]
\end{aligned}$$

By using the thresholded Gaussian data augmentation trick in Albert and Chib (1993), we introduce an auxiliary variable y_{ij} for each pair of entities. Then the second and third row of Eq. (A.1) would be equivalent to

$$\begin{aligned}
&\sum_{i=1}^{N} \sum_{j>i}^{N} \log \int 1\{y_{ij} \geq 0\}^{x_{ij}} 1\{y_{ij} < 0\}^{1 - x_{ij}} \\
&\qquad \mathcal{N}(y_{ij} | w_0 + \sum_{d=1}^{D} w_d z_{id} z_{jd}, 1)\, \mathrm{d}y_{ij} \\
&\geq \sum_{i=1}^{N} \sum_{j>i}^{N} \mathbb{E}_{q(y_{ij})} \left[\log \frac{p(x_{ij}|y_{ij}) p(y_{ij}|z)}{q(y_{ij})} \right] \qquad \text{(A.2)}
\end{aligned}$$

in which $p(x_{ij}|y_{ij}) \triangleq 1\{y_{ij} \geq 0\}^{x_{ij}} 1\{y_{ij} < 0\}^{1 - x_{ij}}$, and $p(y_{ij}|z) \triangleq \mathcal{N}(y_{ij} | w_0 + \sum_{d=1}^{D} w_d z_{id} z_{jd}, 1)$. The greater-than-or-equal-to sign comes from applying Jensen't inequality to the log function.

Bring Eq. (A.2) back to Eq. (A.1), we get a lower bound of the original ELBO. It's also mathematically equivalent to the ELBO of a data-augmented model in which a latent variable y_{ij} is added to each pair of entities. Following Eq. (5), the optimal coordinate-ascent variational factor of y_{ij} can be derived into a truncated normal distribution:

$$\begin{aligned}
&q(y_{ij}) \qquad \text{(A.3)}\\
&= \begin{cases} \mathcal{TN}_+\big(y_{ij} \mid w_0 + \sum_d w_d q_{id} q_{jd}, 1\big), \text{ if } x_{ij} = 1; \\ \mathcal{TN}_-\big(y_{ij} \mid w_0 + \sum_d w_d q_{id} q_{jd}, 1\big), \text{ if } x_{ij} = 0. \end{cases}
\end{aligned}$$

From Eq. (A.2) we could see the data-augmented ELBO is a quadratic function of z, so the coordinate update for the latent feature $q(z)$ can be computed efficiently in quadratic time:

$$\begin{aligned}
&q_{id} \qquad \text{(A.4)}\\
&= \Phi\Big(\log \rho - \log(1 - \rho) + \sum_{j \neq i} q_{jd} w_d \big(\\
&\qquad \mathbb{E}_{q(x_{ij})}[x_{ij}] - w_0 - \frac{1}{2} w_d - \sum_{e \neq d} w_e q_{ie} q_{je} \big) \Big),
\end{aligned}$$

in which the mean of a truncated normal $\mathbb{E}_q[x_{ij}]$ could be computed from the unit Gaussian CDF.