polya-gamma

Bert van der Veen

2025-04-29

Let the marginal log-likelihood for a logistic GLLVM be:

$$
\begin{aligned}
\mathcal{L}(\Theta) &= \sum_{i=1}^{n} \log \int \prod_{j=1}^{m} p(y_{ij}, \mathbf{z}_i) d\mathbf{z}_i \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \log \mathbb{E}_{\mathbf{z}_i} [\exp\{\eta_{ij}\}^{y_{ij}} / \{1 + \exp(\eta_{ij})\} h(\mathbf{z}_i)],
\end{aligned} \tag{1}
$$

with $\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top \boldsymbol{\gamma}_j$. The above expectation of the first term does not possess a closed-form solution. Akin to Gibbs sampling of logistic regression, we apply a polya-gamma data augmentation and additional introduce the auxiliary variable $\omega \sim PG(b, c) = \frac{\exp(-c^2\omega/2) f(\omega; b, 0)}{\mathbb{E}\{\exp(-\frac{1}{2}c^2\omega)\}}$, a polya-gamma random variable, so that the joint log-likelihood instead becomes:

$$
\mathcal{L}(\Theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \log p(y_{ij} | \omega_{ij}, \mathbf{z}_i) + \log p(\omega; 1, 0) - \frac{1}{2} \mathbf{z}_i^\top \mathbf{z}_i, \tag{2}
$$

with corresponding marginal log-likelihood:

$$
\mathcal{L}(\Theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \log \mathbb{E}_{p(\mathbf{z}_i)} [\mathbb{E}_{p(\omega_{ij}; 1, 0)} \{p(y_{ij}, \omega_{ij}, \mathbf{z}_i)\}]. \tag{3}
$$

Applying Jensen's inequality, we have the typical VA ELBO:

$$
\mathcal{L}_{VA}(\Theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{E}_{\mathbf{z}_i, \omega_{ij}} \{\log p(y_{ij} | \omega_{ij}, \mathbf{z}_i)\} + \mathbb{E}_{\omega_{ij}} \{\log p(\omega; 1, 0)\} - \mathbb{E}_{\mathbf{z}_i} (\frac{1}{2} \mathbf{z}_i^\top \mathbf{z}_i) - \mathbb{E}_{\omega_{ij}} \{\log q(\omega)\} - \mathbb{E}_{\mathbf{z}_i} \{\log q(\mathbf{z}_i)\}, \tag{4}
$$

where $\mathbb{E}_{\mathbf{z}_i}\{\log p(y_{ij}|\omega_{ij}, \mathbf{z}_i)\} = -\log 2 + (y_{ij} - \frac{1}{2}) \mathbb{E}_{\mathbf{z}_i}(\eta_{ij}) - \frac{1}{2} \mathbb{E}_{\omega_{ij}}(\omega) \mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)$ under the assumption $q(\omega_{ij}, \mathbf{z}_i) = q(\omega_{ij}) q(\mathbf{z}_i)$.

it is straightforward to show that if we formulate $q(\mathbf{z}_i | \omega_{ij}) q(\omega_{ij})$, the first distribution is conditionally normal and thus possesses a closed form solution w.r.t. integration of ω_{ij}. Thus, due to the polya-gamma augmentation, so we can assume (as usual) $q(\mathbf{z}_i) = \mathcal{N}(\mathbf{a}_i, \mathbf{A}_i)$ where $\mathbf{a}_i$ and $\mathbf{A}_i$ are free variational parameters to estimate. For $q(\omega_{ij})$:

$$
\begin{aligned}
\log q(\omega_{ij}) &\propto \mathbb{E}_{\mathbf{z}_i} \{\mathcal{L}(\Theta)\} \\
&\propto -\frac{1}{2} \omega \mathbb{E}_{\mathbf{z}_i} (\eta_{ij}^2) + \log p(\omega; 1, 0),
\end{aligned} \tag{5}
$$

which we recognize as the (unnormalized) exponentially tilted polya-gamma distribution with $b = 1$ and $c = \sqrt{\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)}$, i.e., $q(\omega_{ij}) = PG\{1, \sqrt{\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)}\}$.The terms due to the KL-divergence for ω_{ij} are given by:

$$\begin{aligned}
\mathbb{E}_{\omega_{ij}}\{\log p(\omega_{ij}; 1, 0)\} - \mathbb{E}_{\omega_{ij}}\{\log q(\omega_{ij})\} &= \mathbb{E}_{\omega_{ij}}\{\log p(\omega_{ij}; 1, 0)\} - \mathbb{E}_{\omega_{ij}}[\log \frac{\exp\{-\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)\omega_{ij}/2\}p\{\omega_{ij}; 1, 0\}}{\cosh^{-1}\{\sqrt{\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)}/2\}}] \\
&= \frac{1}{2}\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)\mathbb{E}_{\omega_{ij}}(\omega_{ij}) - \log[\cosh\{\sqrt{\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)}/2\}],
\end{aligned} \tag{6}$$

where $\mathbb{E}_{\omega_{ij}}\{\log(p(\omega_{ij}; 1, 0)\}$ from the prior cancels out with the same term from the entropy of the variational distribution written as an exponentially tilted polya-gamma variable with $b = 1$ and $c = 0$. Consequently, the ELBO is now:

$$\begin{aligned}
\mathcal{L}_{VA}(\Theta) = -nm\log(2) + \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \frac{1}{2})\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}) - \frac{1}{2}\mathbb{E}_{\omega_{ij}}(\omega_{ij})\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2) + \frac{1}{2}\mathbb{E}_{\omega_{ij}}(\omega_{ij})\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2) - \log[\cosh\{\sqrt{\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2)}/2\}] \\
+ \sum_{i=1}^{n}\{\log\det(\mathbf{A}_i) - \operatorname{tr}(\mathbf{A}_i) - \mathbf{a}_i^\top\mathbf{a}_i\},
\end{aligned} \tag{7}$$

so that we see that the terms involving ω_{ij} cancel out, and we only need to calculate the normalising constant of the variational distribution w.r.t. $\omega \sim PG(1, 0)$, which has the solution $\cosh^{-1}(c/2)$.

Finally, with $\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top\boldsymbol{\gamma}_j$, so that $\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}) = \tilde{\eta_{ij}}$, we have $\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2) = \tilde{\eta}_{ij}^2 + \operatorname{tr}(\boldsymbol{A}_i\boldsymbol{\gamma}_j\boldsymbol{\gamma}_j^\top)$.

Plugging in the different results, with the well known results for $\mathbb{E}_{\mathbf{z}_i}(\frac{1}{2}\mathbf{z}_i^\top\mathbf{z}_i)$ and $\mathbb{E}_{\mathbf{z}_i}\{\log q(\mathbf{z}_i)\}$, equation (4) becomes:

$$\begin{aligned}
\mathcal{L}_{VA}(\Theta) =& -nm\log(2) + \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \frac{1}{2})\tilde{\eta_{ij}} - \log\{\cosh(\sqrt{\tilde{\eta_{ij}}^2 + \boldsymbol{\gamma}_j^\top\boldsymbol{A}_i\boldsymbol{\gamma}_j}/2)\} + \frac{1}{2}\sum_{i=1}^{n}\{\log\det(\mathbf{A}_i) - \operatorname{tr}(\mathbf{A}_i) - \mathbf{a}_i^\top\mathbf{a}_i\} \\
=& \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \frac{1}{2})\tilde{\eta_{ij}} - \frac{1}{2}\sqrt{\tilde{\eta_{ij}}^2 + \boldsymbol{\gamma}_j^\top\boldsymbol{A}_i\boldsymbol{\gamma}_j} - \log\{1 + \exp(-\sqrt{\tilde{\eta_{ij}}^2 + \boldsymbol{\gamma}_j^\top\boldsymbol{A}_i\boldsymbol{\gamma}_j})\} + \frac{1}{2}\sum_{i=1}^{n}\{\log\det(\mathbf{A}_i) - \operatorname{tr}(\mathbf{A}_i) - \mathbf{a}
\end{aligned} \tag{8}$$

Finally, we note that for the quadratic model with $\eta_{ij} = \beta_{0j} + \mathbf{z}_i^\top\boldsymbol{\gamma}_j - \mathbf{z}_i^\top\mathbf{D}_j\mathbf{z}_i$ the solution to $\mathbb{E}_{\mathbf{z}_i}(\eta_{ij}^2) = \operatorname{var}(\eta_{ij}) + \tilde{\eta}_{ij}^2$ is the same as for the probit model.