VARIATIONAL BAYES LINEAR REGRESSION WITH SPIKE AND SLAB PRIORS

Alexander Pei
alexanderpei805@gmail.com

1 Introduction

Maybe someone will find this useful for something. References [1, 2, 3]

https://github.com/AverageSignalsEnjoyer/Variational-Bayes-Linear-Regression-Spike-and-Slab

1.1 Distributions

The given data $\mathbf{Y} \in \mathbb{R}^{S\times T}$ and $\mathbf{X} \in \mathbb{R}^{K\times T}$, where T is the number of time points. S and K are the number of variables in each of the data matrices. $\mathbf{W} \in \mathbb{R}^{S\times K}$ is the weighting matrix.

The observation model:

$$\mathbf{y}_{\cdot,t} \sim \mathcal{N}(\mathbf{y}_{\cdot,t} \mid \mathbf{W}\mathbf{x}_{\cdot,t}, \text{diag}(\boldsymbol{\tau})) \tag{1}$$

$\mathbf{y}_{\cdot,t} \in \mathbb{R}^{S\times 1}$ and $\mathbf{x}_{\cdot,t} \in \mathbb{R}^{K\times 1}$ are vectors representing the states at a particular time index t. $\boldsymbol{\tau} \in \mathbb{R}^{1\times S}$ is a vector of precision values for each of the observed states. $\text{diag}(\boldsymbol{\tau}) \in \mathbb{R}^{S\times S}$ is a diagonal matrix with the elements of $\boldsymbol{\tau}$ along the diagonal.

The spike and slab prior over weighting elements are as follows:

$$w_{sk} \mid z_{sk}, \alpha_k \sim \begin{cases} \mathcal{N}(w_{sk} \mid 0, \alpha_k^{-1}) & \text{if } z_{sk} = 1 \\ \mathcal{N}(w_{sk} \mid 0, \beta^{-1}) & \text{if } z_{sk} = 0 \end{cases} \tag{2}$$

w_{sk} is a single element in the matrix $\mathbf{W}$. $\mathbf{Z} \in \mathbb{R}^{S\times K}$ is a matrix of 1's and 0's indicating which elements in the weighting matrix are active. z_{sk} is a single element in the matrix $\mathbf{W}$. If $z_{sk} = 1$, then w_{sk} follows the "slab" distribution. α_k is a single element in the vector $\boldsymbol{\alpha} \in \mathbb{R}^{1\times K}$, is the precision of the weighting matrix distribution. If $z_{sk} = 0$, then w_{sk} follows the "spike" distribution. In this case, $\beta \gg \alpha_k$, meaning that the spike distribution has much higher precision. If β becomes arbitrarily large, the distribution approaches the Dirac delta distribution.

z_{sk} follows a Bernoulli distribution:

$$z_{sk} \sim \text{Ber}(z_{sk} \mid \pi_k) \tag{3}$$

where π_k is an element in the vector $\boldsymbol{\pi} \in \mathbb{N}^{1\times K}$ that specifies the prior probability of an element in the weighting matrix being active (1) or not (0).

Each of the precision values α_k and τ_s follow a gamma distribution:

$$\alpha_k \sim \text{Gamma}(\alpha_k \mid c_{\alpha k}, d_{\alpha k}) \tag{4}$$

$$\tau_s \sim \text{Gamma}(\tau_s \mid c_{\tau s}, d_{\tau s}) \tag{5}$$

where the parameters in the Gamma distriubtion are hyperparameters for each of the precision values.

1.2 Probabilties

The full joint distribution has the following form:

$$p(\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\tau}) = p(\mathbf{Y} \mid \mathbf{W}, \boldsymbol{\tau}) p(\mathbf{W} \mid \mathbf{Z}, \boldsymbol{\alpha}) p(\mathbf{Z}) p(\boldsymbol{\alpha}) p(\boldsymbol{\tau}) \tag{6}$$

The probability of the data given the weighting matrix and covariance matrix factorizes over a product of time points:

$$\begin{aligned} p(\mathbf{Y} \mid \mathbf{W}, \boldsymbol{\tau}) &= \prod_{t=1}^{T} \mathcal{N}(\mathbf{y}_{\cdot,t} \mid \mathbf{W}\mathbf{x}_{\cdot,t}, \text{diag}(\boldsymbol{\tau})) \\ &= \prod_{t=1}^{T} \prod_{s=1}^{S} \mathcal{N}(y_{st} \mid \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t} \rangle, \tau_s) \end{aligned} \tag{7}$$

The second equality uses the fact that the covariance matrix is diagonal, thus the multivariate normal distribution factors into a product of univariate normal distributions. y_{st} is a particular element in the matrix $\mathbf{Y}$. $\mathbf{w}_{s,\cdot} \in \mathbb{R}^{1 \times K}$ is a vector in the matrix $\mathbf{W}$.

The probability of the weighting matrix given the binary matrix and precision:

$$\begin{aligned} p(\mathbf{W} \mid \mathbf{Z}, \boldsymbol{\alpha}) p(\mathbf{Z}) &= p(\mathbf{W}, \mathbf{Z} \mid \boldsymbol{\alpha}) \\ &= \prod_{s=1}^{S} \prod_{k=1}^{K} p(w_{sk}, z_{sk} \mid \alpha_k) \end{aligned} \tag{8}$$

where

$$p(w_{sk}, z_{sk} \mid \alpha_k) = \left[\mathcal{N}(w_{sk} \mid 0, \alpha_k^{-1}) \pi_k\right]^{z_{sk}} \left[\mathcal{N}(w_{sk} \mid 0, \beta^{-1})(1 - \pi_k)\right]^{1 - z_{sk}} \tag{9}$$

Note in the above equation, only one of the normal distributions survives since $z_{sk} \in \{0, 1\}$, and the probability of this happening for either case follows the Bernoulli distribution with parameter π_k.

The remaining two distribution factorize as a product of Gamma distributions for each individual component:

$$\begin{aligned} p(\boldsymbol{\alpha}) &= \prod_{k=1}^{K} \text{Gamma}(\alpha_k \mid c_{\alpha k}, d_{\alpha k}) \\ p(\boldsymbol{\tau}) &= \prod_{s=1}^{S} \text{Gamma}(\tau_s \mid c_{\tau s}, d_{\tau s}) \end{aligned} \tag{10}$$

1.3 Variational inference

We approximate the true posterior distribution with an approximate distribution of the parameters with the following form:

$$\begin{aligned} q(\mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\tau}) &= q(\mathbf{W}, \mathbf{Z}) q(\boldsymbol{\alpha}) q(\boldsymbol{\tau}) \\ &= \left[\prod_{s=1}^{S} \prod_{k=1}^{K} q(w_{sk}, z_{sk})\right] \left[\prod_{k=1}^{K} q(\alpha_k)\right] \left[\prod_{s=1}^{S} q(\tau_s)\right] \end{aligned} \tag{11}$$

Solution for the variational factor in general:

$$\log q^*(\theta_j) = \mathbb{E}_{q(\mathbf{\Theta}/\{\theta_j\})}\left[\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\tau})\right] + \text{const.} \tag{12}$$

where $\theta_j \in \mathbf{\Theta} = \{\mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\tau}\}$ is a particular parameter in the set of all variational parameters $\mathbf{\Theta}$. The notation $q(\mathbf{\Theta}/\{\theta_j\})$ is the distribution over all variational parameters, except θ_j.

1.3.1 $q^*(\alpha_k)$

$$\begin{aligned}
\log q^*(\alpha_k) &= \mathbb{E}_{q(\mathbf{\Theta}/\{\alpha_k\})}\left[\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\tau})\right] + \text{const.} \\
&= \mathbb{E}_{q(\mathbf{\Theta}/\{\alpha_k\})}\left[\log p(\mathbf{W}, \mathbf{Z} \mid \boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha})\right] + \text{const.} \\
&= \mathbb{E}_{q(\mathbf{\Theta}/\{\alpha_k\})}\left[\log \prod_{s=1}^{S}\prod_{k=1}^{K} p(w_{sk}, z_{sk} \mid \alpha_k) + \log \prod_{k=1}^{K} p(\alpha_k)\right] + \text{const.} \\
&= \mathbb{E}_{q(\mathbf{\Theta}/\{\alpha_k\})}\left[\sum_{s=1}^{S} \log p(w_{sk}, z_{sk} \mid \alpha_k) + \log p(\alpha_k)\right] + \text{const.} \\
&= \sum_{s=1}^{S} \mathbb{E}_{q(\mathbf{\Theta}/\{\alpha_k\})}\left[\log p(w_{sk}, z_{sk} \mid \alpha_k)\right] + \log p(\alpha_k) + \text{const.} \\
&= \sum_{s=1}^{S} \mathbb{E}_{q(w_{sk}, z_{sk})}\left[\log p(w_{sk}, z_{sk} \mid \alpha_k)\right] + \log p(\alpha_k) + \text{const.}
\end{aligned} \tag{13}$$

$$\begin{aligned}
\log p(w_{sk}, z_{sk} \mid \alpha_k) &= z_{sk} \log\left[\mathcal{N}(w_{sk} \mid 0, \alpha_k^{-1})\pi_k\right] + (1 - z_{sk}) \log\left[\mathcal{N}(w_{sk} \mid 0, \beta^{-1})(1 - \pi_k)\right] \\
&= z_{sk} \log\left[\mathcal{N}(w_{sk} \mid 0, \alpha_k^{-1})\pi_k\right] + \text{const.} \\
&= z_{sk}\left[\frac{1}{2}\log\alpha_k - \frac{1}{2}\log 2\pi - \frac{1}{2}\alpha_k w_{sk}^2 + \log \pi_k\right] + \text{const.} \\
&= z_{sk}\left[\frac{1}{2}\log\alpha_k - \frac{1}{2}\alpha_k w_{sk}^2\right] + \text{const.}
\end{aligned} \tag{14}$$

$$\begin{aligned}
\log p(\alpha_k) &= \log \text{Gamma}(\alpha_k \mid c_{\alpha k}, d_{\alpha k}) \\
&= -\log \Gamma(c_{\alpha k}) + (c_{\alpha k} - 1)\log \alpha_k - d_{\alpha k}\alpha_k + c_{\alpha k} \log d_{\alpha k} \\
&= (c_{\alpha k} - 1)\log \alpha_k - d_{\alpha k}\alpha_k + \text{const.}
\end{aligned} \tag{15}$$

Combining the equations we have

$$\log q^*(\alpha_k) = \left(c_{\alpha k} + \frac{1}{2}\sum_{s=1}^{S} \mathbb{E}_{q(w_{sk}, z_{sk})}\left[z_{sk}\right] - 1\right)\log\alpha_k - \left(d_{\alpha k} + \frac{1}{2}\sum_{s=1}^{S} \mathbb{E}_{q(w_{sk}, z_{sk})}\left[z_{sk} w_{sk}^2\right]\right)\alpha_k + \text{const.} \tag{16}$$

which has the form of a Gamma distribution.

1.3.2 $q^*(\tau_s)$

$$\begin{aligned}
\log q^*(\tau_s) &= \mathbb{E}_{q(\mathbf{\Theta}/\{\tau_s\})}\left[\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\tau})\right] + \text{const.} \\
&= \mathbb{E}_{q(\mathbf{\Theta}/\{\tau_s\})}\left[\log p(\mathbf{Y} \mid \mathbf{W}, \boldsymbol{\tau}) + \log p(\boldsymbol{\tau})\right] + \text{const.} \\
&= \mathbb{E}_{q(\mathbf{\Theta}/\{\tau_s\})}\left[\log \prod_{t=1}^{T}\prod_{s=1}^{S} \mathcal{N}(y_{st} \mid \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle, \tau_s) + \log \prod_{s=1}^{S} p(\tau_s)\right] + \text{const.} \\
&= \mathbb{E}_{q(\mathbf{\Theta}/\{\tau_s\})}\left[\sum_{t=1}^{T}\sum_{s=1}^{S} \log \mathcal{N}(y_{st} \mid \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle, \tau_s) + \log p(\tau_s)\right] + \text{const.} \\
&= \sum_{t=1}^{T} \mathbb{E}_{q(\mathbf{\Theta}/\{\tau_s\})}\left[\log \mathcal{N}(y_{st} \mid \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle, \tau_s)\right] + \log p(\tau_s) + \text{const.}
\end{aligned} \tag{17}$$

The expectation in the above equation is under the distribution $\prod_{k=1}^{K} q(w_{sk}, z_{sk})$

$$\begin{aligned}\log \mathcal{N}(y_{st} \mid \mathbf{w}_{s,\cdot} \cdot \mathbf{x}_{\cdot,t}, \tau_s) &= \frac{1}{2}\log\tau_s - \frac{1}{2}\log 2\pi - \frac{1}{2}\tau_s\left(y_{st} - \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle\right)^2 \\ &= \frac{1}{2}\log\tau_s - \frac{1}{2}\tau_s\left(y_{st} - \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle\right)^2 + \text{const.}\end{aligned} \tag{18}$$

$$\begin{aligned}\left(y_{st} - \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle\right)^2 &= \left(y_{st} - \sum_{k=1}^{K} w_{sk}x_{kt}\right)^2 \\ &= y_{st}^2 - y_{st}\sum_{k=1}^{K} 2w_{sk}x_{kt} + \sum_{k=1}^{K}\sum_{\substack{k'=1\\k'\neq k}}^{K} 2w_{sk}w_{sk'}x_{kt}x_{k't} + \sum_{k=1}^{K} w_{sk}^2 x_{kt}^2\end{aligned} \tag{19}$$

$$\begin{aligned}&\mathbb{E}_{q(\boldsymbol{\Theta}/\{\tau_s\})}\left[\log\mathcal{N}(y_{st} \mid \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle, \tau_s)\right] \\ &= \frac{1}{2}\log\tau_s - \frac{1}{2}\tau_s\mathbb{E}_{q(\boldsymbol{\Theta}/\{\tau_s\})}\left[y_{st}^2 - y_{st}\sum_{k=1}^{K} 2w_{sk}x_{kt} + \sum_{k=1}^{K}\sum_{\substack{k'=1\\k'\neq k}}^{K} 2w_{sk}w_{sk'}x_{kt}x_{k't} + \sum_{k=1}^{K} w_{sk}^2 x_{kt}^2\right] \\ &= \frac{1}{2}\log\tau_s - \frac{1}{2}\tau_s\Bigg[y_{st}^2 - y_{st}\sum_{k=1}^{K} 2\mathbb{E}_{q(w_{sk},z_{sk})}\left[w_{sk}\right]x_{kt} \\ &\qquad + \sum_{k=1}^{K}\sum_{\substack{k'=1\\k'\neq k}}^{K} 2\mathbb{E}_{q(w_{sk},z_{sk})}\left[w_{sk}\right]\mathbb{E}_{q(w_{sk'},z_{sk})}\left[w_{sk'}\right]x_{kt}x_{k't} + \sum_{k=1}^{K}\mathbb{E}_{q(w_{sk},z_{sk})}\left[w_{sk}^2\right]x_{kt}^2\Bigg]\end{aligned} \tag{20}$$

$$\begin{aligned}\log q^*(\tau_s) &= \left(c_{\tau s} + \frac{T}{2} - 1\right)\log\tau_s \\ &\quad - \Bigg(d_{\tau s} + \frac{1}{2}\sum_{t=1}^{T}\Bigg[y_{st}^2 - y_{st}\sum_{k=1}^{K} 2\mathbb{E}_{q(w_{sk},z_{sk})}\left[w_{sk}\right]x_{kt} + \\ &\quad + \sum_{k=1}^{K}\sum_{\substack{k'=1\\k'\neq k}}^{K} 2\mathbb{E}_{q(w_{sk},z_{sk})}\left[w_{sk}\right]\mathbb{E}_{q(w_{sk'},z_{sk})}\left[w_{sk'}\right]x_{kt}x_{k't} + \sum_{k=1}^{K}\mathbb{E}_{q(w_{sk},z_{sk})}\left[w_{sk}^2\right]x_{kt}^2\Bigg]\Bigg)\tau_s\end{aligned} \tag{21}$$

1.3.3 $q^*(w_{sk}, z_{sk})$

$$\begin{aligned}\log q^*(w_{sk}, z_{sk}) &= \mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\log p(\mathbf{Y}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\alpha}, \boldsymbol{\tau})\right] + \text{const.} \\ \log q^*(w_{sk}, z_{sk}) &= \mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\log p(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\tau}) + \log p(\mathbf{W}, \mathbf{Z} \mid \boldsymbol{\alpha})\right] + \text{const.}\end{aligned} \tag{22}$$

$$
\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\log p(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\tau})\right] \\
&= \mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\log \prod_{t=1}^{T}\prod_{s=1}^{S}\mathcal{N}(y_{st} \mid \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle, \tau_s)\right] \\
&= \mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\sum_{t=1}^{T}\sum_{s=1}^{S}\left(\frac{1}{2}\log\tau_s - \frac{1}{2}\log 2\pi - \frac{1}{2}\tau_s\left(y_{st} - \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle\right)^2\right)\right] \\
&= \mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\sum_{t=1}^{T}\sum_{s=1}^{S} -\frac{1}{2}\tau_s\left(y_{st} - \langle \mathbf{w}_{s,\cdot}, \mathbf{x}_{\cdot,t}\rangle\right)^2\right] \\
&= -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\sum_{s=1}^{S}\tau_s\left(y_{st}^2 - y_{st}\sum_{k=1}^{K}2w_{sk}x_{kt} + \sum_{k=1}^{K}\sum_{\substack{k'=1\\k'\neq k}}^{K}2w_{sk}w_{sk'}x_{kt}x_{k't} + \sum_{k=1}^{K}w_{sk}^2x_{kt}^2\right)\right] \\
&= -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\sum_{s=1}^{S}\tau_s\left(-y_{st}\sum_{k=1}^{K}2w_{sk}x_{kt} + \sum_{k=1}^{K}\sum_{\substack{k'=1\\k'\neq k}}^{K}2w_{sk}w_{sk'}x_{kt}x_{k't} + \sum_{k=1}^{K}w_{sk}^2x_{kt}^2\right)\right] \\
&= -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\tau_s\left(-2y_{st}w_{sk}x_{kt} + \sum_{\substack{k'=1\\k'\neq k}}^{K}2w_{sk}w_{sk'}x_{kt}x_{k't} + w_{sk}^2x_{kt}^2\right)\right] \\
&= -\frac{1}{2}\sum_{t=1}^{T}\mathbb{E}_{q(\tau_s)}\left[\tau_s\right]\mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[-2y_{st}w_{sk}x_{kt} + \sum_{\substack{k'=1\\k'\neq k}}^{K}2w_{sk}w_{sk'}x_{kt}x_{k't} + w_{sk}^2x_{kt}^2\right]
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\log p(\mathbf{W}, \mathbf{Z} \mid \boldsymbol{\alpha})\right] \\
&= \mathbb{E}_{q(\boldsymbol{\Theta}/\{w_{sk},z_{sk}\})}\left[\log \prod_{s=1}^{S}\prod_{k=1}^{K} p(w_{sk}, z_{sk} \mid \alpha_k)\right] \\
&= \mathbb{E}_{q(\alpha_k)}\left[z_{sk}\log\left[\mathcal{N}(w_{sk} \mid 0, \alpha_k^{-1})\pi_k\right] + (1-z_{sk})\log\left[\mathcal{N}(w_{sk} \mid 0, \beta^{-1})(1-\pi_k)\right]\right] \\
&= z_{sk}\left(\frac{1}{2}\mathbb{E}_{q(\alpha_k)}\left[\log\alpha_k\right] - \frac{1}{2}\log 2\pi - \frac{1}{2}\mathbb{E}_{q(\alpha_k)}\left[\alpha_k\right]w_{sk}^2 + \log\pi_k\right) \\
&\qquad + (1-z_{sk})\left(\frac{1}{2}\log\beta - \frac{1}{2}\log 2\pi - \frac{1}{2}\beta w_{sk}^2 + \log(1-\pi_k)\right)
\end{aligned}
\tag{24}
$$

Collecting the quadratic terms for w_{sk}^2

$$
-\frac{1}{2}\left(\sum_{t=1}^{T}\mathbb{E}_{q(\tau_s)}\left[\tau_s\right]x_{kt}^2 + z_{sk}\mathbb{E}_{q(\alpha_k)}\left[\alpha_k\right] + (1-z_{sk})\beta\right)w_{sk}^2 \coloneqq -\frac{1}{2}\gamma_{sk}w_{sk}^2
\tag{25}
$$

Collecting the linear terms for w_{sk}

$$
\left(\sum_{t=1}^{T}\mathbb{E}_{q(\tau_s)}\left[\tau_s\right]y_{st}x_{kt} - \sum_{t=1}^{T}\mathbb{E}_{q(\tau_s)}\left[\tau_s\right]\sum_{\substack{k'=1\\k'\neq k}}^{K}\mathbb{E}_{q(w_{sk'},z_{sk'})}\left[w_{sk'}\right]x_{kt}x_{k't}\right)w_{sk} \coloneqq \rho_{sk}w_{sk}
\tag{26}
$$

Thus $q^*(w_{sk}, z_{sk})$ has a form

$$q^*(w_{sk}, z_{sk}) = \frac{1}{\mathcal{Z}} \exp\left(-\frac{1}{2}\gamma_{sk} w_{sk}^2 + \rho_{sk} w_{sk}\right) \pi_k^{z_{sk}} (1-\pi_k)^{1-z_{sk}} \exp\left(\frac{1}{2} z_{sk} \mathbb{E}_{q(\alpha_k)}\left[\log \alpha_k\right]\right) \beta^{1-z_{sk}} \tag{27}$$

$$\begin{aligned} q^*(z_{sk}=1) &= \int_{-\infty}^{\infty} q^*(w_{sk}, z_{sk}=1) dw_{sk} \\ &= \frac{1}{\mathcal{Z}} \pi_k \exp\left(\frac{1}{2}\mathbb{E}_{q(\alpha_k)}\left[\log \alpha_k\right]\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\gamma_{sk} w_{sk}^2 + \rho_{sk} w_{sk}\right) dw_{sk} \\ &= \frac{1}{\mathcal{Z}} \pi_k \exp\left(\frac{1}{2}\mathbb{E}_{q(\alpha_k)}\left[\log \alpha_k\right]\right) \sqrt{2\pi} \exp\left(\rho_{sk}^2 / 2\gamma_{sk}\right) / \sqrt{\gamma_{sk}} \end{aligned} \tag{28}$$

$$\begin{aligned} q^*(z_{sk}=0) &= \int_{-\infty}^{\infty} q^*(w_{sk}, z_{sk}=0) dw_{sk} \\ &= \frac{1}{\mathcal{Z}} (1-\pi_k) \exp\left(\frac{1}{2}\log \beta\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\gamma_{sk} w_{sk}^2 + \rho_{sk} w_{sk}\right) dw_{sk} \\ &= \frac{1}{\mathcal{Z}} (1-\pi_k) \exp\left(\frac{1}{2}\log \beta\right) \sqrt{2\pi} \exp\left(\rho_{sk}^2 / 2\gamma_{sk}\right) / \sqrt{\gamma_{sk}} \end{aligned} \tag{29}$$

2 Appendix

$$\mathcal{N}(y \mid w, \tau^{-1}) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}(y-w)^2\right) \tag{30}$$

$$\text{Gamma}(\alpha \mid c, d) = \frac{1}{\Gamma(c)} \alpha^{c-1} \exp(-d\alpha) d^c \tag{31}$$

$$\text{Ber}(z \mid k) = k^z (1-k)^{1-z} \tag{32}$$

References

[1] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[2] Yong See Foo and Heejung Shim. A comparison of bayesian inference techniques for sparse factor analysis, 2021.

[3] Seth M Hirsh, David A Barajas-Solano, and J Nathan Kutz. Sparsifying priors for bayesian uncertainty quantification in model discovery. *Royal Society Open Science*, 9(2):211823, 2022.