Online Appendix: Gibbs Sampler and VB

September 1, 2023

1 The Model

$$y = Z\theta + \epsilon \tag{1}$$

where y is $T \times 1$ vector of responses, Z is the $T \times k$ matrix of standardized regressors, ϵ is the $T \times 1$ vector of $i.i.d$ Normal errors with mean 0 and unknown variance σ^2.

2 The Priors

Following Bhattacharya, Pati, Pillai, and Dunson (2015) (BPPD), we elicit hierarchical DL prior as following:

$$\theta_j|\phi,\tau \sim DE(\phi_j\tau), \ \ \phi_j \sim Dir(a,...,a) \tag{2}$$

Conditional prior for θ_j is $DE(\phi_j\tau)$ implies a zero mean Double Exponential or Lapalace distribution with the density $f(\theta_j) = (2\phi_j\tau)^{-1}exp(-\frac{|\theta_j|}{\phi_j\tau})$ for $\theta_j \in \mathbb{R}$.

Next, we set a Gamma priors for τ and σ^{-2}:

$$\tau \sim G(ka, 1/2), \ \ \sigma^{-2} \sim G(\nu, S). \tag{3}$$

The above hierarchical DL prior for θ_j can be expressed as:

$$\theta_j \sim N(0, \psi_j\phi_j^2\tau^2), \ \ \psi_j \sim Exp(1/2) \tag{4}$$

Hence, the prior of θ is $N(\mathbf{0}, V)$ where $V = diag(\psi_1\phi_1^2\tau^2, ..., \psi_k\phi_k^2\tau^2)$.

3 The Posteriors

Multiply the likelihood and the priors, we have

$$\begin{aligned}
& L(y|\theta,\sigma^2)p(\theta|\phi,\tau,\sigma^2)p(\tau)p(\phi)p(\sigma^2) \\
\propto & \frac{1}{(\sqrt{\sigma^2})^T} exp[-\frac{1}{2\sigma^2}(y-Z\theta)'(y-Z\theta)] \times \prod_{j=1}^{k} \frac{1}{2\phi_j\tau} exp[-\sum_{j=1}^{k}\frac{|\theta_j|}{\phi_j\tau}] \times \\
& \tau^{ka-1}exp(-\frac{\tau}{2}) \times \prod_{j=1}^{k}\phi_j^{a-1} \times (\sigma^{-2})^{\nu-1}exp(-S/\sigma^2)
\end{aligned} \tag{5}$$

or

$$\begin{aligned}
& L(y|\theta,\sigma^2)p(\theta|\phi,\tau,\psi,\sigma^2)p(\tau)p(\phi)p(\psi)p(\sigma^2) \\
\propto & \frac{1}{(\sqrt{\sigma^2})^T} exp[-\frac{1}{2\sigma^2}(y-Z\theta)'(y-Z\theta)] \times \prod_{j=1}^{k} \frac{1}{\sqrt{\psi_j\phi_j^2\tau^2}} exp[-\sum_{j=1}^{k}\frac{\theta_j^2}{2\psi_j\phi_j^2\tau^2}] \times \\
& \tau^{ka-1}exp(-\frac{\tau}{2}) \times \prod_{j=1}^{k}\phi_j^{a-1} \times \prod_{j=1}^{k} exp(-\psi_j/2) \times (\sigma^{-2})^{\nu-1}exp(-S/\sigma^2)
\end{aligned} \tag{6}$$

In (6), the items involving θ are $exp[-\frac{1}{2\sigma^2}(y-Z\theta)'(y-Z\theta)] \times \prod_{j=1}^{k}\frac{1}{\sqrt{\psi_j\phi_j^2\tau^2}} exp[-\sum_{j=1}^{k}\frac{\theta_j^2}{2\psi_j\phi_j^2\tau^2}]$. Thus, the conditional posterior of θ is $N(\bar{b}, \overline{V})$, with $\overline{V} = [\sigma^{-2}Z'Z + V^{-1}]^{-1}$ and $\bar{b} = \sigma^{-2}\overline{V}Z'y$.

In (5), the items involving τ are $\frac{1}{\tau^k}exp[-\sum_{j=1}^{k}\frac{|\theta_j|}{\phi_j\tau}] \times \tau^{ka-1}exp(-\frac{\tau}{2})$, which can be written as $\tau^{ka-k-1}exp[-\frac{1}{2}(\tau + \frac{1}{\tau}(\sum_{j=1}^{k}\frac{2|\theta_j|}{\phi_j}))]$. Thus the conditional posterior of τ is $giG(ka - k, 1, \sum_{j=1}^{k}\frac{2|\theta_j|}{\phi_j})$.[1]

To find the conditional posteriors of ϕ_j, we integrate τ out following BPPD. Collecting the terms involving ϕ_j, we have $\prod_{j=1}^{k}(\phi_j^{a-1}\frac{1}{\phi_j})\int_{\tau=0}^{\infty}\tau^{ka-k-1}exp(-\frac{\tau}{2})exp[-\sum_{j=1}^{k}\frac{|\theta_j|}{\phi_j\tau}]d\tau$. Thus, we can derive the conditional posterior of ϕ_j as following: First, we have $\xi_j \sim giG(a-1, 1, 2|\theta_j|)$. Next let et $\Xi = \sum_{j=1}^{k}\xi_j$. The conditional posterior of ϕ_j can then be found to be ξ_j/Ξ.

In (6), the terms involving ψ_js are $\prod_{j=1}^{k}\psi_j^{-\frac{1}{2}}exp[-\sum_{j=1}^{k}\frac{\theta_j^2}{2\psi_j\phi_j^2\tau^2}] \times \prod_{j=1}^{k}exp(-\psi_j/2)$. Thus

[1] $y \sim giG(p, a, b)$ if $f(y) \propto y^{p-1}exp[-\frac{1}{2}(ay + b/y)]$.

the conditional posterior of $1/\psi_j$ is Inverse Gaussian with mean $\sqrt{\frac{\phi_j^2\tau^2}{\theta_j^2}}$ and scale parameter 1.

In (6), the terms involving σ^{-2} are

$$\frac{1}{(\sqrt{\sigma^2})^T} exp[-\frac{1}{2\sigma^2}(y - Z\theta)'(y - Z\theta)] \times (\sigma^{-2})^{\nu-1} exp(-S/\sigma^2)$$

Thus the conditional posterior of σ^{-2} is $G(\frac{T}{2} + \nu, \frac{1}{2}(y - Z\theta)'(y - Z\theta) + S)$.

4 Variational Bayes

The optimal VB q densities are as follows:

4.1 $q(\theta)$

$$q(\theta) \sim N(\overline{\theta}, \overline{V}), \tag{7}$$

where

$$\overline{V} = (\frac{\frac{T}{2} + \nu}{\overline{S}} Z'Z + V^{-1})^{-1}$$

$$\overline{\theta} = (\frac{\frac{T}{2} + \nu}{\overline{S}})\overline{V} Z'y$$

$$V^{-1} = diag(\overline{\psi_1^{-1}}\ \overline{\phi_1^{-2}}\overline{\tau^{-2}}, ..., (\overline{\psi_k^{-1}}\ \overline{\phi_k^{-2}}\overline{\tau^{-2}})$$

$$E[\log(|V^{-1}|)] = \sum_{j=1}^{k} E[\log(\psi_j) + 2\log(\phi_j)] + 2kE[\log(\tau)],$$

where

$$E[\log(\psi_j) + 2\log(\phi_j)] = \int_0^\infty q(\psi_j)\log(\psi_j)d\psi_j + 2\int_0^\infty q(\phi_j)\log(\phi_j)d\phi_j$$

and

$$E[\log(\tau)] = \int_0^\infty q(\tau)\log(\tau)d\tau$$

4.2 $q(\sigma^{-2})$

$$q(\sigma^{-2}) \sim G(\frac{T}{2} + \nu, \overline{S}), \tag{8}$$

where

$$\overline{S} = \frac{1}{2}[||y - Z\overline{\theta}||^2 + tr(Z'Z\overline{V})] + S$$

Hence

$$\overline{\sigma}^{-2} = \frac{\frac{T}{2} + \nu}{\overline{S}},$$

and

$$E[\log(\sigma^{-2})] = \psi(\nu + \frac{T}{2}) - log(\overline{S})$$

where $\psi(\bullet)$ is the Digamma function.

4.3 $q(\tau)$

$$q(\tau) \sim giG[ka - k, 1, \sum_{j=1}^{k} 2(\overline{\theta}_j^2 + \overline{V}_{jj})^{1/2} \frac{1}{\overline{\overline{\phi_j}}}], \tag{9}$$

Let $\chi = \sum_{j=1}^{k} 2(\overline{\theta}_j^2 + \overline{V}_{jj})^{1/2} \frac{1}{\overline{\overline{\phi_j}}}$, we have

$$\overline{\tau} = \frac{\sqrt{\chi} K_{ka-k+1}(\sqrt{\chi})}{K_{ka-k}(\sqrt{\chi})}$$

and

$$\overline{\tau^2} = \overline{\tau}^2 + \chi[\frac{K_{ka-k+2}(\sqrt{\chi})}{K_{ka-k}(\sqrt{\chi})} - (\frac{K_{ka-k+1}(\sqrt{\chi})}{K_{ka-k}(\sqrt{\chi})})^2]$$

where $K_*[\bullet]$ is the modified Bessel functions of the second kind.

4.4 $q(\psi_j)$

$$q(\frac{1}{\psi_j}) \sim iG(\sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\theta}_j^2 + \overline{V}^{jj}}}, 1), \tag{10}$$

Let $\rho = \sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\theta}_j^2 + \overline{V}^{jj}}}$,

$$\overline{\frac{1}{\psi_j}} = \rho$$

and

$$\overline{\psi}_j = 1 + 1/\rho$$

Note that to calculate $ELBO$, we need to use the following optimal q density of ψ_j.[2]

$$q(\psi_j) = (\frac{\psi_j}{2\pi})^{1/2} \exp\{\frac{-(\psi_j - \rho)^2}{2\rho^2\psi_j}\} \tag{11}$$

4.5 $q(\phi_j)$

$$q(\xi_j) \sim giG(a-1, 1, 2\sqrt{\overline{\theta}_j^2 + (\overline{V}_{jj})^2}) \tag{12}$$

Let $\varpi = 2\sqrt{\overline{\theta}_j^2 + (\overline{V}_{jj})^2}$, we have

$$\overline{\xi}_j = \frac{\sqrt{\varpi} K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})},$$

and

$$var(\xi_j) = \varpi\{\frac{K_{a+1}(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})} - [\frac{K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})}]^2\}.$$

where $var(\bullet)$ denotes the variance.

Scaling ξ, we have

$$\overline{\phi}_j = \frac{\overline{\xi}_j}{\sum^k \overline{\xi}_j},$$

and

$$\overline{\phi_j^2} = \overline{\phi}_j^2 + \frac{var(\xi_j)}{(\sum^k \overline{\xi}_j)^2}$$

Thus, the optimal q density of ϕ_j takes the following form:

$$q(\phi_j) \sim giG[a-1, \sum^k \overline{\xi}_j, (2\sqrt{\overline{\theta}_j^2 + (\overline{V}_{jj})^2})/(\sum^k \overline{\xi}_j)] \tag{13}$$

[2] If x is distributed as $f(x)$, then $y = 1/x$ is distributed as $\frac{1}{y^2} f(\frac{1}{y})$.

4.6 *ELBO*

The evidence lower bound (*ELBO*) is as follows:[3]

$$\begin{aligned} ELBO &= E\{\log p(\mathbf{y}, \theta, \sigma^2, \phi, \tau, \psi)\} - E\{\log q(\theta, \sigma^2, \phi, \tau, \psi)\} \\ &= E\{\log p(\mathbf{y}|\theta, \sigma^2, \phi, \tau, \psi)\} + E\{\log p(\theta)\} + E\{\log p(\sigma^2)\} \\ &+ E\{\log p(\phi)) + E(\log p(\psi)\} + E\{\log p(\tau)\} \\ &- E\{\log(q(\theta))\} - E\{\log q(\sigma^2)\} - E\{\log q(\tau)\} - E\{\log q(\phi)\} - E\{\log q(\psi)\} + ... \end{aligned} \tag{14}$$

where

$$E\{\log p(\mathbf{y}|\theta, \sigma^2, \phi, \tau, \psi)\} = -\frac{T}{2}\log(2\pi) - \frac{T}{2}[-\psi(\nu + \frac{T}{2}) + log(\overline{S})] - \frac{\overline{S} - S}{(\frac{\overline{S}}{\nu + \frac{T}{2}})}, \tag{15}$$

$$E\{\log p(\theta)\} = -\frac{k}{2}\log(2\pi) - \frac{1}{2}E[\log|\mathbf{V}|] - \frac{1}{2}[\overline{\theta}^{'}\mathbf{V}^{-1}\overline{\theta} + tr(\mathbf{V}^{-1}\overline{\mathbf{V}})], \tag{16}$$

$$E\{\log p(\sigma^{-2})\} \quad = \quad \nu \log S - \log\Gamma(\nu) - (\nu - 1)[-\psi(\nu + \frac{T}{2}) + log(\overline{S})] - (\frac{S}{\frac{\overline{S}}{\nu + \frac{T}{2}}}), \tag{17}$$

$$E\{\log q(\theta)\} = -(\frac{k}{2} + \frac{k}{2}\log(2\pi) + \frac{1}{2}\log|\overline{\mathbf{V}}_i|), \tag{18}$$

$$E\{\log q(\sigma^{-2})\} = -[\nu + \frac{T}{2} - \log(\overline{S}) + \log(\Gamma(\nu + \frac{T}{2})) - (\nu + \frac{T}{2} - 1)\psi(\nu + \frac{T}{2})]. \tag{19}$$

[3]It is seen that the *ELBO* described below contains many constant terms. These terms need to be dropped from *ELBO*s in VB iterations to speed up the process.

Notice that

$$
\begin{aligned}
&E\{\log p(\mathbf{y}|\theta,\sigma^2,\phi,\tau,\psi)\} + E\{\log p(\theta)\} + E\{\log p(\sigma^2)\} - E\{\log(q(\theta))\} - E\{\log q(\sigma^2)\}\\
&= \frac{k}{2} - \frac{T}{2}\log(2\pi) + \frac{1}{2}\log(|\overline{V}|) - \frac{1}{2}E[\log(|V|)] - \frac{1}{2}[\overline{\theta}' V^{-1}\overline{\theta} + tr(V^{-1}\overline{V})] + \nu\log(\underline{s}) - \log\Gamma(\nu)\\
&- (\nu + \frac{T}{2})\log(\overline{s}) + \log(\nu + \frac{T}{2})\\
&= \frac{1}{2}\log(|\overline{V}|) - \frac{1}{2}E[\log(|V|)] - \frac{1}{2}[\overline{\theta}' V^{-1}\overline{\theta} + tr(V^{-1}\overline{V})] - (\nu + \frac{T}{2})\log(\overline{s}) + Const.
\end{aligned}
\tag{20}
$$

Above terms are the baseline $ELBO$ discussed in GKP2018. The additional terms are as follows:

$$
\begin{aligned}
E\{\log p(\tau)\} &= -\log\Gamma(ka) - ka[\int_0^\infty (q(\tau)\log\tau)d\tau] - 0.5\overline{\tau}\\
&= -ka[\int_0^\infty (q(\tau)\log\tau)d\tau] - 0.5\overline{\tau} + Const.
\end{aligned}
\tag{21}
$$

$$
E\{\log p(\psi)\} = \sum_{j=1}^{k}(\log\frac{1}{2} - 0.5\overline{\psi}_j) = -\frac{1}{2}\sum_{j=1}^{k}\overline{\psi}_j + Const. \tag{22}
$$

$$
\begin{aligned}
&E\{\log p(\phi) = \log\Gamma(ka) - k\log\Gamma(a) - (a-1)\sum_{j=1}^{k}[\int_0^\infty (q(\phi_j)\log\phi_j)d\phi_j]\\
&= -(a-1)\sum_{j=1}^{k}[\int_0^\infty (q(\phi_j)\log\phi_j)d\phi_j]
\end{aligned}
\tag{23}
$$

$$
E\{\log q(\tau)\} = \int_0^\infty q(\tau)\log q(\tau)d\tau. \tag{24}
$$

$$
E\{\log q(\psi)\} = \sum_{j=1}^{k}\int_0^\infty q(\psi_j)\log q(\psi_j)d\psi_j \tag{25}
$$

$$
E\{\log q(\phi)\} = \sum_{j=1}^{k}\int_0^\infty q(\phi_j)\log q(\phi_j)d\phi_j \tag{26}
$$

Terms in $ElBO$ that do not have clear analytical forms can be numerically estimated.

References

[1] Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. (2015), Dirichlet-laplace priors for optimal shrinkage, *Journal of the American Statistical Association*, 110, 1479-1490.