ST451 Bayesian Machine Learning
Week 5

Exercises

1. Consider the model where we have independent observations $y = (y_1, \dots, y_n)$ from the $\mathrm{N}(\mu, \sigma^2)$ with the $\mathrm{N}(\mu_0, \omega^2\sigma^2)$ and $\mathrm{IGamma}(\alpha_0, \beta_0)$ being the priors for μ and σ^2 respectively. The aim is to perform variational inference for this model under the mean field approximation framework. Derive the explicit algorithm for this problem and present its steps.

   Answer: Note that the for $\theta = (\mu, \sigma^2)$ is specified as $\pi(\theta) = \pi(\mu, \sigma^2) = \pi(\mu|\sigma^2)\pi(\sigma^2)$. The joint density $\pi(y, \theta) = f(y|\theta)\pi(\theta)$ for the parameters θ and data y can be written as

$$
\begin{aligned}
\pi(y,\theta) &= f(y|\mu,\sigma^2)\pi(\mu|\sigma^2)\pi(\sigma^2) \\
&\propto (\sigma^2)^{-n/2}\exp\left(\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}\right)(\sigma^2)^{-1/2}\exp\left(-\frac{(\mu-\mu_0)^2}{2\omega^2\sigma^2}\right)(\sigma^2)^{-\alpha_0-1}\exp\left(-\frac{\beta_0}{\sigma^2}\right)
\end{aligned}
$$

   Taking logs we get

$$
\log\pi(y,\theta) = -\frac{n+1}{2}\log\sigma^2 - \frac{\frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2}{\sigma^2} - \frac{(\mu-\mu_0)^2}{2\omega^2\sigma^2} - (\alpha_0+1)\log\sigma^2 - \frac{\beta_0}{\sigma^2} + c
$$

   We will now derive the VB components $q(\mu)$ and $q(\sigma^2)$. For $q(\mu)$ we can focus on the terms involving μ.

$$
\begin{aligned}
\log q(\mu|y,\phi) &= \mathbb{E}_{q(\sigma^2)}\left[-\frac{\frac{1}{2}\sum_{i=1}^n (y_i-\mu)^2}{\sigma^2} - \frac{\frac{1}{2}(\omega^2)^{-1}(\mu-\mu_0)^2}{\sigma^2}\right] + c \\
&= -\frac{1}{2}\left[\sum_{i=1}^n (y_i-\mu)^2 + (\omega^2)^{-1}(\mu-\mu_0)^2\right]\mathbb{E}_{q(\sigma^2)}(1/\sigma^2) + c.
\end{aligned}
$$

   Denoting by $\sigma_\mu^2 = 1/\mathbb{E}_{q(\sigma^2)}(1/\sigma^2)$ and exponentiating gives

$$
q(\mu|y,\phi) \propto \exp\left(-\frac{\mu^2 - 2\mu\frac{(\omega^2)^{-1}\mu_0 + \sum_i y_i}{n+(\omega^2)^{-1}}}{2\frac{\sigma_\mu^2}{n+(\omega^2)^{-1}}}\right),
$$

   which is recognised as the $\mathrm{N}(\mu_\phi, \sigma_\phi^2)$, where

$$
\begin{aligned}
\mu_\phi &= \frac{(\omega^2)^{-1}\mu_0 + \sum_{i=1}^n y_i}{n+(\omega^2)^{-1}} \\
\sigma_\phi^2 &= \frac{\sigma_\mu^2}{n+(\omega^2)^{-1}} = \frac{1/\mathbb{E}_{q(\sigma^2)}(1/\sigma^2)}{n+(\omega^2)^{-1}}
\end{aligned}
$$

   For $q(\sigma^2|y,\phi)$ we take as before

$$
\begin{aligned}
\log q(\sigma^2|y,\phi) &= \mathbb{E}_{q(\mu)}\left[-\left(\frac{n+1}{2}+\alpha_0+1\right)\log\sigma^2 - \frac{\beta_0 + \frac{1}{2}\left((\omega^2)^{-1}(\mu-\mu_0)^2 + \sum_{i=1}^n (y_i-\mu)^2\right)}{\sigma^2}\right] + c \\
&= -\left(\frac{n+1}{2}+\alpha_0+1\right)\log\sigma^2 - \frac{\beta_0 + \frac{1}{2}\mathbb{E}_{q(\mu)}\left[(\omega^2)^{-1}(\mu-\mu_0)^2 + \sum_{i=1}^n (y_i-\mu)^2\right]}{\sigma^2} + c
\end{aligned}
$$

By inspection we can identify $q(\sigma^2)$ to be the IGamma$(\alpha_\phi, \beta_\phi)$, where

$$\begin{aligned}
\alpha_\phi &= \alpha_0 + \tfrac{n+1}{2}, \\
\beta_\phi &= \beta_0 + \tfrac{1}{2}\left(S_y^2 - 2\mathbb{E}(\mu)S_y + n\mathbb{E}(\mu^2)\right) + \frac{1}{2\omega^2}\left(\mu_0^2 - 2\mu_0\mathbb{E}(\mu) + \mathbb{E}(\mu^2)\right), \\
\mathbb{E}(\mu) &= \mathbb{E}_{q(\mu)}(\mu), \quad \mathbb{E}(\mu^2) = \mathbb{E}_{q(\mu)}(\mu^2), \\
S_y &= \sum_i y_i, \quad S_y^2 = \sum_i y_i^2.
\end{aligned}$$

So overall we set $q(\mu|\mu_\phi, \sigma_\phi^2)$ =N$(\mu_\phi, \sigma_\phi^2)$ and $q(\tau|\alpha_\phi, \beta_\phi)$ =IGamma$(\alpha_\phi, \beta_\phi)$.

Then we look for the q parameters $\phi = (\mu_\phi, \sigma_\phi^2, \alpha_\phi, \beta_\phi)$ that maximise ELBO by setting

$$\begin{aligned}
\mu_\phi &= \frac{(\omega^2)^{-1}\mu_0 + \sum_{i=1}^n y_i}{n + (\omega^2)^{-1}}, \\
\alpha_\phi &= \alpha_0 + \tfrac{n+1}{2},
\end{aligned}$$

and initialising $\sigma_\phi^2 = (\sigma_\phi^2)^0$ and $\beta_\phi = \beta_\phi^0$. To complete the updates we note that if g is the Gamma$(\alpha_\phi, \beta_\phi)$, we can write $1/\mathbb{E}_{q(\sigma^2)}(1/\sigma^2) = \mathbb{E}_g(\sigma^2) = \beta_\phi/\alpha_\phi$. Similarly $\mathbb{E}(\mu) = \mu_\phi$ and $\mathbb{E}(\mu^2) = \sigma_\phi^2 + \mu_\phi^2$. Hence, the update fro from iteration n to $n+1$ will take the following form

$$\begin{aligned}
(\sigma_\phi^2)^{n+1} &= \frac{(\beta_\phi)^n/\alpha_\phi}{n + (\omega^2)^{-1}}, \\
\mathbb{E}(\mu^2) &= (\sigma_\phi^2)^{n+1} + \mu_\phi^2, \\
(\beta_\phi)^{n+1} &= \beta_0 + \tfrac{1}{2}\left(S_y^2 - 2\mu_\phi S_y + n\mathbb{E}(\mu^2)\right) + \frac{1}{2\omega^2}\left(\mu_0^2 - 2\mu_0\mu_\phi + \mathbb{E}(\mu^2)\right)
\end{aligned}$$

2. Following the previous exercise, simulate data from this model, using parameters of your preference, and apply the algorithm for variational inference for the parameters you used. Provide evidence of agreement between the estimated and true values of the parameters used.

   Answer: See file ending in ...exercise2.ipynb

3. Consider the model for the VIX index mentioned in the lecture slides

$$Y_t = Y_{t-1} + \kappa(\mu - Y_{t-1})\delta + \epsilon_t,$$

   where Y_t is VIX at time t, and ϵ_t are independent error terms following the student-t distribution with scale σ and 3 degrees of freedom. Download the VIX series (file vix_201518.csv) and fit this model using automatic Variational Bayes via RStan.

   Hint: In RStan the code $x \sim$ student_t(df,m,s) indicates that x follows the student-t distribution with df degrees of freedom, location mu and scale s.

   Answer: See file ending in ...exercise3.Rmd