A Simple Application of Variational Bayes

Babak Shahbaba

Department of Statistics, University of California at Irvine, CA, USA

Consider the following model:

$$
\begin{aligned}
x_i &\sim N(\mu, \sigma^2), \qquad i = 1, \ldots, n \\
\mu &\sim N(\mu_0, \sigma_0^2) \\
\sigma^2 &\sim \text{Inv-Gamma}(\alpha, \beta)
\end{aligned}
$$

$$P(\mu, \sigma^2|x) = \frac{P(\mu, \sigma^2, x)}{P(x)}$$

where

$$P(\mu, \sigma^2, x) = \sigma^{-1} \exp[-\frac{\Sigma(x_i - \mu)^2}{2\sigma^2}] \exp[-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}](\sigma^2)^{-\alpha-1} \exp[-\frac{\beta}{\sigma^2}]$$

As an illustrative example, we set $\mu = -1$ and $\sigma^2 = 1$. We generate 5 random samples from $N(-1, 2)$. Given these samples as observed data, the posterior distribution, $P(\mu, \sigma^2|x)$, based on Gibbs sampling is shown in Figure 1. Our goal is to approximate this distribution with $Q(\mu, \sigma^2|\theta)$. We assume

$$Q(\mu, \sigma^2|\theta) = Q(\mu|\theta)Q(\sigma^2|\theta)$$

More specifically, we assume

$$
\begin{aligned}
Q(\mu|m, v) &= N(m, v^2) \\
Q(\sigma^2|a, b) &= \text{Inv-Gamma}(a, b) \\
Q(\mu|m, v)Q(\sigma^2|a, b) &\propto v^{-1} \exp[-\frac{(\mu - m)^2}{2v^2}].\frac{b^a}{\Gamma(a)}(\sigma^2)^{-a-1} \exp(-\frac{b}{\sigma^2})
\end{aligned}
$$

We minimize $-\mathcal{L}(Q)$ with respect to $\theta = (m, v, a, b)$,

$$\mathcal{L}(Q) = E_Q[\log P(\mu, \sigma^2, x)] - E_Q[\log Q(\mu, \sigma^2|\theta)]$$

subject to $v, a, b > 0$. We have

$$
\begin{aligned}
\log P(\mu, \sigma^2, x) &= -n \log \sigma - \frac{\Sigma(x_i - \mu)^2}{2\sigma^2} - \frac{(\mu - \mu_0)^2}{2\sigma_0^2} - (\alpha + 1) \log \sigma^2 - \frac{\beta}{\sigma^2} \\
\log Q(\mu, \sigma^2|\theta) &= -\log v - \frac{(\mu - m)^2}{2v^2} + a \log b - \log \Gamma(a) - (a + 1) \log \sigma^2 - \frac{b}{\sigma^2}
\end{aligned}
$$

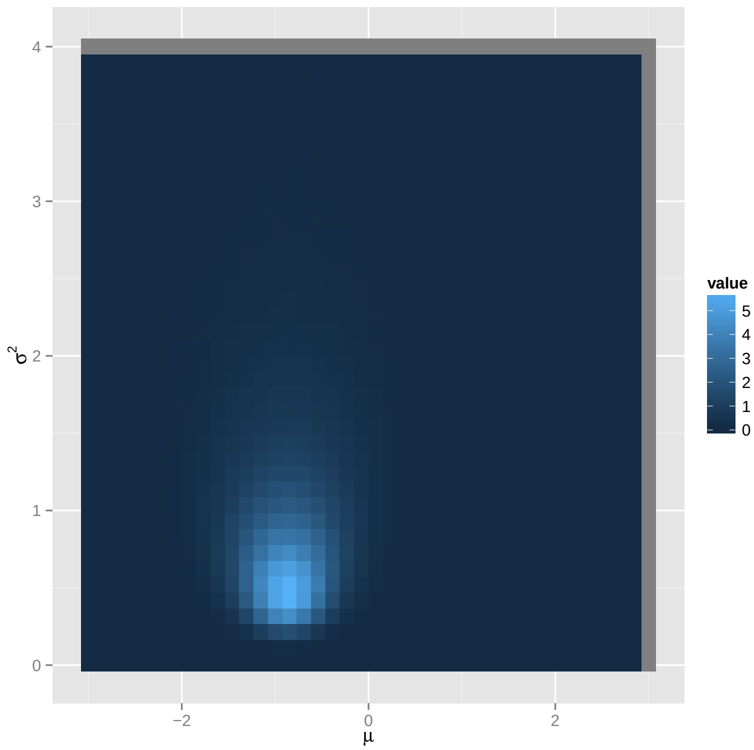

Fig 1: Posterior distribution of (μ, σ^2) using the Gibbs sampler.

Note that for Inv-Gamma distribution, we have

$$E_Q(\frac{1}{\sigma^2}) = a/b$$
$$E_Q(\log \sigma^2) = \log b - \psi(a)$$

where ψ is the digamma function, whose derivative is ψ_1, the trigamma function.

We start by minimizing $-\mathcal{L}(Q)$ with respect to m and v keeping a and b fixed at their current values. We have

$$\begin{aligned}
E_Q[\log P(\mu, \sigma^2, x)] &= E[-\frac{\Sigma(x_i - \mu)^2}{2\sigma^2} - \frac{(\mu - \mu_0)^2}{2\sigma_0^2}] \\
&= -\frac{1}{2}E(\frac{1}{\sigma^2})(nv^2 + \Sigma(x_i - m)^2) - \frac{1}{2\sigma_0^2}(v^2 + (\mu_0 - m)^2) \\
&= -\frac{a}{2b}(nv^2 + \Sigma(x_i - m)^2) - \frac{1}{2\sigma_0^2}(v^2 + (\mu_0 - m)^2) \\
E_Q[\log Q(\mu, \sigma^2|\theta)] &= -\log v - \frac{1}{v^2}v^2 \\
&= -\log v - \frac{1}{2}
\end{aligned}$$

Therefore,

$$\mathcal{L}(Q) = -\frac{a}{2b}(nv^2 + \Sigma(x_i - m)^2) - \frac{1}{2\sigma_0^2}(v^2 + (\mu_0 - m)^2) + \log v + \frac{1}{2}$$

Next we find the partial derivatives of $\mathcal{L}(Q)$ with respect to m and v,

$$\begin{aligned}
\partial\mathcal{L}(Q)/\partial m &= \frac{-a}{2b}(2nm - 2\Sigma x_i) - \frac{1}{2\sigma_0^2}(2m - 2\mu_0) \\
&= \frac{-a}{b}(nm - \Sigma x_i) - \frac{1}{\sigma_0^2}(m - \mu_0) \\
\partial\mathcal{L}(Q)/\partial v &= -\frac{a}{b}nv - \frac{v}{\sigma_0^2} + \frac{1}{v} \\
&= (-\frac{an}{b} - \frac{1}{\sigma_0^2})v + \frac{1}{v}
\end{aligned}$$

Next, we minimize $-\mathcal{L}(Q)$ with respect to a and b keeping m and v fixed at their current values. We have

$$\begin{aligned}
E_Q[\log P(\mu, \sigma^2, x)] &= -\frac{n}{2}E(\log\sigma^2) - \frac{nv^2 + \Sigma(x_i - m)^2}{2}E(\frac{1}{\sigma^2}) - (\alpha + 1)E(\log\sigma^2) - \beta E(\frac{1}{\sigma^2}) \\
&= -\frac{n}{2}[\log b - \psi(a)] - \frac{nv^2 + \Sigma(x_i - m)^2}{2}\frac{a}{b} - (\alpha + 1)[\log b - \psi(a)] - \beta\frac{a}{b} \\
&= -(\alpha + n/2 + 1)[\log b - \psi(a)] - \frac{nv^2 + \Sigma(x_i - m)^2 + 2\beta}{2}\frac{a}{b} \\
E_Q[\log Q(\mu, \sigma^2|\theta)] &= -\frac{1}{2} + a\log b - \log\Gamma(a) - (a + 1)[\log b - \psi(a)] - a
\end{aligned}$$

Therefore,

$$\mathcal{L}(Q) = (a - \alpha - n/2)[\log b - \psi(a)] - \frac{nv^2 + \Sigma(x_i - m)^2 + 2\beta}{2}\frac{a}{b} + \frac{1}{2} - a\log b + \log\Gamma(a) + a$$

We now find the partial derivatives with respect to a and b,

$$\begin{aligned}
\partial\mathcal{L}(Q)/\partial a &= \log b - \psi(a) - (a-\alpha-n/2)\psi_1(a) - \frac{nv^2+\Sigma(x_i-m)^2+2\beta}{2b} - \log b + \psi(a) \\
&= -(a-\alpha-n/2)\psi_1(a) - \frac{nv^2+\Sigma(x_i-m)^2+2\beta}{2b} \\
\partial\mathcal{L}(Q)/\partial b &= \frac{a-\alpha-n/2}{b} + \frac{nv^2+\Sigma(x_i-m)^2+2\beta}{2}\frac{a}{b^2} - \frac{a}{b} \\
&= \frac{-\alpha-n/2}{b} + \frac{a[nv^2+\Sigma(x_i-m)^2+2\beta]}{2b^2}
\end{aligned}$$

We can simply use the coordinate descent algorithm (Luo and Tseng, 1992) to find Q. For simplicity, we assume $a=1$ and minimize $-\mathcal{L}(Q)$ in terms of $m, v,$, and b as follows:

$$\begin{aligned}
\partial\mathcal{L}(Q)/\partial m &= \frac{-a}{b}(nm-\Sigma x_i) - \frac{1}{\sigma_0^2}(m-\mu_0) = 0 \\
m &= \frac{n\bar{x} + \frac{b}{a}\frac{\mu_0}{\sigma_0^2}}{(n+\frac{b}{a}\frac{1}{\sigma_0^2})} \\
\partial\mathcal{L}(Q)/\partial v &= (-\frac{an}{b} - \frac{1}{\sigma_0^2})v + \frac{1}{v} = 0 \\
v &= \sqrt{\frac{\frac{b}{a}}{n+\frac{b}{a}\frac{1}{\sigma_0^2}}} \\
\partial\mathcal{L}(Q)/\partial b &= \frac{-\alpha-n/2}{b} + \frac{a[nv^2+\Sigma(x_i-m)^2+2\beta]}{2b^2} = 0 \\
b &= \frac{a[nv^2+\Sigma(x_i-m)^2+2\beta]}{2\alpha+n}
\end{aligned}$$

Therefore, we use Algorithm 1 for updating the parameters.

Algorithm 1 Coordinate descent algorithm

Initialize m, v and b
for $\ell = 1$ to L **do**

$m^{(\ell+1)} = \frac{n\bar{x}+\frac{b^{(\ell)}}{a}\frac{\mu_0}{\sigma_0^2}}{(n+\frac{b^{(\ell)}}{a}\frac{1}{\sigma_0^2})}$

$v^{(\ell+1)} = \sqrt{\frac{\frac{b^{(\ell)}}{a}}{n+\frac{b^{(\ell)}}{a}\frac{1}{\sigma_0^2}}}$

$b^{(\ell+1)} = \frac{a[n[v^{(\ell+1)}]^2+\Sigma(x_i-m^{(\ell+1)})^2+2\beta]}{2\alpha+1}$

end for

The approximate posterior distribution is shown in Figure 2.

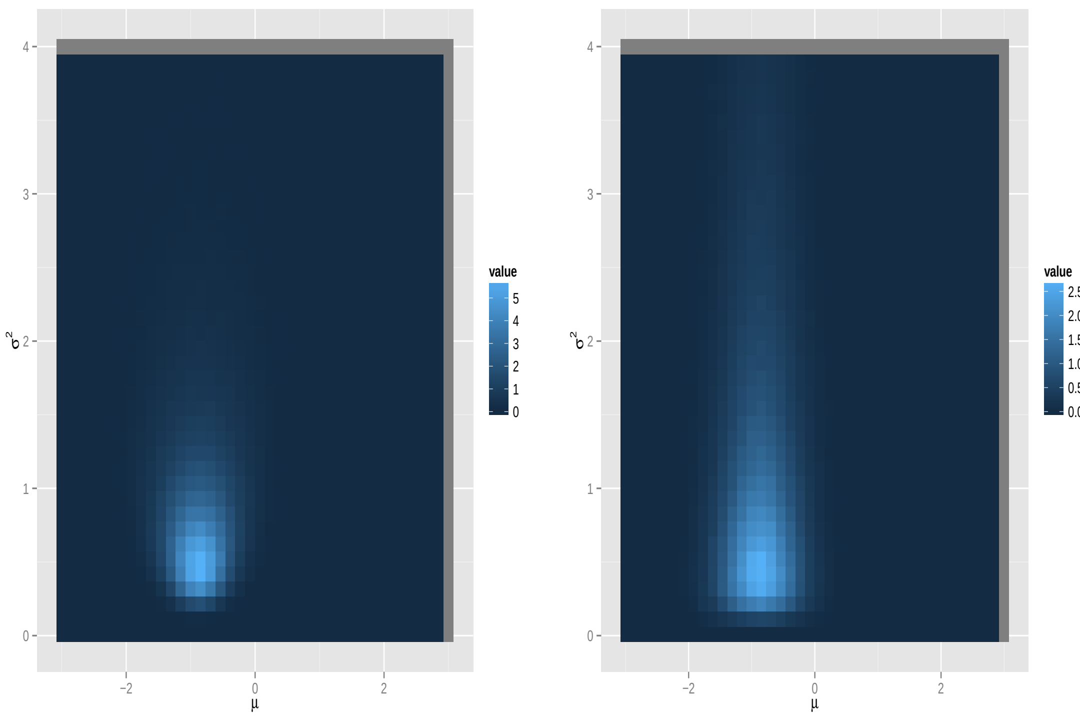

Fig 2: Approximate posterior distribution using variational Bayes. Left panel: True posterior distribution; Right panel: Variational Bayes approximation.