CAVI Derivations

Clément Corbeau–Izorche

1 Model

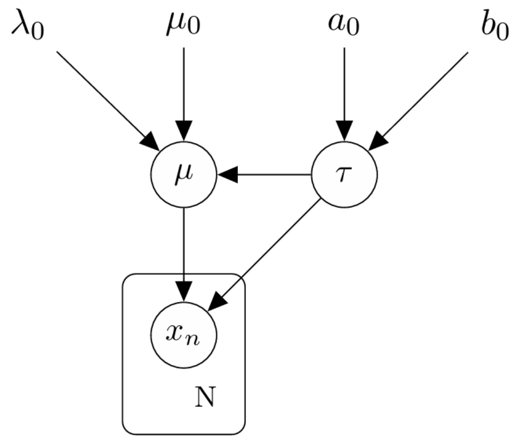

Figure 1.1: Directed graphical model (DGM)

Let us consider the model given by the equations (1), (2) and (3) , for which DGM is given in Figure 1.1.

$$p(\mathbf{x}|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{N/2} \exp\left(-\frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2\right) \tag{1}$$

$$p(\mu|\tau) = \mathcal{N}\left(\mu \mid \mu_0, (\lambda_0\tau)^{-1}\right) \tag{2}$$

$$p(\tau) = \mathrm{Gam}(\tau \mid a_0, b_0) \tag{3}$$

2 ML Estimates

We take the logarithm of the likelihood, given in Equation (1):

$$\ell(\mu, \tau) = \log p(\mathbf{x}|\mu, \tau) \stackrel{\pm}{=} \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_{n=1}^{N}(x_n - \mu)^2$$

Then, the parameters we are looking for are given by:

$$\mu_{\text{ML}} = \arg\max_{\mu} \ell(\mu), \quad \tau_{\text{ML}} = \arg\max_{\tau} \ell(\tau)$$

In order to get their expressions, we set the corresponding partial derivatives to zero:

- With respect to μ:

$$\begin{aligned}
\frac{\partial \ell}{\partial \mu} &= \tau\sum_{n=1}^{N} x_n - \tau N\mu \\
\frac{\partial \ell}{\partial \mu}(\mu_{ML}) = 0 &\implies \tau\sum_{n=1}^{N} x_n - \tau N\mu_{\text{ML}} = 0 \\
&\implies \mu_{\text{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n
\end{aligned}$$

- With respect to τ:

$$\begin{aligned}
\frac{\partial \ell}{\partial \tau} &= \frac{N}{2\tau} - \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2 \\
\frac{\partial \ell}{\partial \tau}(\tau_{ML}) = 0 &\implies \frac{N}{\tau_{\text{ML}}} = \sum_{n=1}^{N}(x_n - \mu_{\text{ML}})^2 \\
&\implies \tau_{\text{ML}} = \frac{N}{\sum_{n=1}^{N}(x_n - \mu_{\text{ML}})^2}
\end{aligned}$$

3 Exact Posterior

Using Bayes' theorem, the posterior distribution is given by:

$$p(\mu, \tau|\mathbf{x}) = \frac{p(\mathbf{x}|\mu, \tau)\, p(\mu, \tau)}{p(\mathbf{x})} \propto p(\mathbf{x}|\mu, \tau)\, p(\mu, \tau)$$

By taking the logarithm, we get:

$$\begin{aligned}
\log p(\mu, \tau|\mathbf{x}) &\stackrel{\pm}{=} \log p(\mathbf{x}|\mu, \tau) + \log p(\mu, \tau) \\
&= \log p(\mathbf{x}|\mu, \tau) + \log p(\mu|\tau) + \log p(\tau) \\
&= \frac{N}{2}\log\tau - \frac{\tau}{2}\sum_{n=1}^{N}(x_n-\mu)^2 + \frac{1}{2}\log\tau - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 + (a_0-1)\log\tau + b_0\log\tau \\
&\stackrel{\pm}{=} \left(\frac{N}{2} + a_0 - \frac{1}{2}\right)\log\tau - b_0\tau - \frac{\tau}{2}\left(N+\lambda_0\right)\mu^2 + 2\mu\tau\left(\sum_{n=1}^{N} x_n + \lambda_0\mu_0\right) \\
&\quad + \left(\sum_{n=1}^{N} x_n^2 + \lambda_0\mu_0^2\right) \\
&= \left(\frac{N}{2} + a_0 - \frac{1}{2}\right)\log\tau - b_0\tau - \frac{\tau}{2}\left(N+\lambda_0\right)\mu^2 + \frac{2\mu\tau}{N+\lambda_0}\left(\sum_{n=1}^{N} x_n + \lambda_0\mu_0\right) \\
&\quad + \frac{\sum_{n=1}^{N} x_n^2 + \lambda_0\mu_0^2}{N+\lambda_0} \\
&= \left(\frac{N}{2} + a_0 - \frac{1}{2}\right)\log\tau - \left(b_0 + \frac{1}{2}\left(\sum_{n=1}^{N} x_n^2 + \lambda_0\mu_0^2\right)\right)\tau \\
&\quad - \frac{\tau}{2}(N+\lambda_0)\left(\mu^2 - 2\mu\left(\frac{\sum_{n=1}^{N} x_n + \lambda_0\mu_0}{N+\lambda_0}\right)\right) \\
&= \left(\frac{N}{2} + a_0 - \frac{1}{2}\right)\log\tau - \left(b_0 + \frac{1}{2}\left(\sum_{n=1}^{N} x_n^2 + \lambda_0\mu_0^2\right)\right)\tau \\
&\quad - \frac{\tau}{2}(N+\lambda_0)\left(\mu^2 - 2\mu C + C^2 - C^2\right)
\end{aligned}$$

where $C = \frac{\sum_{n=1}^{N} x_n + \lambda_0\mu_0}{N+\lambda_0}$

Then,

$$\log p(\mu,\tau|\mathbf{x}) = \left(\frac{N}{2}+a_0-\frac{1}{2}\right)\log\tau - \left(b_0+\frac{1}{2}\left(\sum_{n=1}^{N}x_n^2+\lambda_0\mu_0^2\right)\right)\tau$$
$$+\frac{\tau}{2}(N+\lambda_0)C^2-\frac{\tau}{2}(N+\lambda_0)(\mu+C)^2$$
$$= \left(\frac{N}{2}+a_0-\frac{1}{2}\right)\log\tau - \left(b_0+\frac{1}{2}\left(\sum_{n=1}^{N}x_n^2+\lambda_0\mu_0^2-(N+\lambda_0)C^2\right)\right)\tau$$
$$-(N+\lambda_0)(\mu+C)^2\frac{\tau}{2}$$

Finally, we identify the expression of a $Normal-Gamma$ distribution: $p(\mu,\tau|\mathbf{x}) = \mathcal{NG}(a_0^*, b_0^*, \mu_0^*, \lambda_0^*)$ with the following parameters:

$$\begin{cases} a_0^* = a_0 + \dfrac{N}{2}, \\ b_0^* = b_0 + \dfrac{1}{2}\left(\displaystyle\sum_{n=1}^{N}x_n^2+\lambda_0\mu_0^2-\frac{\left(\sum_{n=1}^{N}x_n+\lambda_0\mu_0\right)^2}{N+\lambda_0}\right), \\ \mu_0^* = \dfrac{\sum_{n=1}^{N}x_n+\lambda_0\mu_0}{N+\lambda_0}, \\ \lambda_0^* = \lambda_0 + N \end{cases}$$

4 Variational Posterior

Joint distribution

The joint distribution is given by:

$$p(\mathbf{x}, \mu, z) = p(\mathbf{x}|\mu, z)p(\mu|\tau)p(\tau)$$

By taking the logarithm, it follows:

$$\begin{aligned}
\log p(\mathbf{x}, \mu, z) &= \log p(\mathbf{x}|\mu, z) + \log p(\mu|\tau) + \log p(\tau), \\
&= \frac{N}{2} \log \frac{\tau}{2\pi} - \frac{\tau}{2} \sum_{n=1}^{N} (x_n - \mu)^2 + \frac{1}{2} \log \frac{\lambda_0 \tau}{2\pi} - \frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2 \\
&\quad + a_0 \log b_0 - \log \Gamma(a_0) + (a_0 - 1) \log \tau - b_0 \tau
\end{aligned}$$

CAVI updates

We consider a factorized variational approximation to the posterior distribution given by:

$$q(\mu, \tau) = q(\mu)q(\tau),$$

CAVI equation with respect to μ:

$$
\begin{aligned}
\log q^*(\mu) &\stackrel{+}{=} E_\tau[\log p(\mathbf{x}, \mu, \tau)] \\
&\stackrel{+}{=} E_\tau \left[-\frac{\tau}{2} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2 \right] \\
&\stackrel{+}{=} -\frac{1}{2} E_\tau[\tau] \left(\sum_{n=1}^{N} (x_n^2 + \mu^2 - 2x_n\mu) \right) - \frac{\lambda_0}{2} E_\tau[\tau](\mu^2 + \mu_0^2 - 2\mu\mu_0) \\
&\stackrel{+}{=} -\frac{1}{2} E_\tau[\tau] \left(N\mu^2 - 2\mu \sum_{n=1}^{N} x_n \right) - \frac{\lambda_0}{2} E_\tau[\tau](\mu^2 - 2\mu\mu_0) \\
&= \left(\sum_{n=1}^{N} x_n + \lambda_0\mu_0 \right) E_\tau[\tau]\mu - \frac{(N + \lambda_0)}{2} E_\tau[\tau]\mu^2
\end{aligned}
$$

Finally, we identify $q^*(\mu) = \mathcal{N}(\mu^*, \lambda^{*-1})$, with the following parameters:

$$
\begin{cases}
\mu^* = \dfrac{\sum_{n=1}^{N} x_n + \lambda_0\mu_0}{N + \lambda_0}, \\
\lambda^* = (N + \lambda_0) E_\tau[\tau]
\end{cases}
$$

CAVI equation with respect to τ:

$$
\begin{aligned}
\log q^*(\tau) &\stackrel{+}{=} E_\mu[\log p(\mathbf{x}, \mu, \tau)] \\
&\pm E_\mu \left[\frac{N}{2} \log \tau - \frac{\tau}{2} \sum_{n=1}^{N} (x_n - \mu)^2 + \frac{1}{2} \log \tau - \frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2 + (a_0 - 1) \log \tau - b_0 \tau \right] \\
&= \left(\frac{N}{2} + \frac{1}{2} + a_0 - 1 \right) \log \tau - \left(b_0 + \frac{1}{2} E_\mu \left[\sum_{n=1}^{N} (x_n - \mu)^2 + \lambda_0 (\mu - \mu_0)^2 \right] \right) \tau
\end{aligned}
$$

Focusing on the last term, it follows:

$$
\begin{aligned}
E_\mu\left[\sum_{n=1}^{N}(x_n-\mu)^2+\lambda_0(\mu-\mu_0)^2\right] &= \sum_{n=1}^{N} E_\mu[x_n^2+\mu^2-2x_n\mu]+\lambda_0 E_\mu[\mu^2+\mu_0^2-2\mu\mu_0] \\
&= \sum_{n=1}^{N} x_n^2 + N E_\mu[\mu^2] - 2E_\mu[\mu]\sum_{n=1}^{N} x_n \\
&\quad + \lambda_0 E_\mu[\mu^2] + \lambda_0\mu_0^2 - 2\lambda_0\mu_0 E_\mu[\mu] \\
&= \sum_{n=1}^{N} x_n^2 + (N+\lambda_0)E_\mu[\mu^2] - 2\left(\lambda_0\mu_0+\sum_{n=1}^{N} x_n\right)E_\mu[\mu] \\
&\quad + \lambda_0\mu_0^2
\end{aligned}
$$

Putting all together, we get:

$$
\begin{aligned}
q^*(\tau) &\stackrel{+}{=} \left(\frac{N+1}{2}+a_0-1\right)\log\tau \\
&\quad - \left(b_0+\frac{1}{2}\left(\sum_{n=1}^{N} x_n^2 + (N+\lambda_0)E_\mu[\mu^2] - 2\left(\lambda_0\mu_0+\sum_{n=1}^{N} x_n\right)E_\mu[\mu]+\lambda_0\mu_0^2\right)\right)\tau
\end{aligned}
$$

Finally, we identify $q^*(\tau)=\text{Gamma}(a^*,b^*)$, with the following parameters:

$$
\begin{cases}
a^* = a_0 + \dfrac{N+1}{2}, \\
b^* = b_0 + \dfrac{1}{2}\left[\displaystyle\sum_{n=1}^{N} x_n^2 + (N+\lambda_0)E_\mu[\mu^2] - 2\left(\lambda_0\mu_0+\sum_{n=1}^{N} x_n\right)E_\mu[\mu]+\lambda_0\mu_0^2\right]
\end{cases}
$$

Furthermore, we know:

$$
E_\mu[\mu]=\mu^* \quad \text{and} \quad E_\mu[\mu^2]=\text{Var}(\mu)+E_\mu[\mu]^2=\lambda^{*-1}+\mu^{*2}.
$$

ELBO

The ELBO (Evidence Lower BOund) is given by:

$$
\begin{aligned}
\mathcal{L}(q) &= E_{q(\mu),q(\tau)}\left[\log\frac{p(x,\mu,\tau)}{q(\mu,\tau)}\right] \\
&= E_q\left[\log p(x|\mu,\tau)\right] + E_q\left[\log p(\mu|\tau)\right] + E_q\left[\log p(\tau)\right] - E_q\left[\log q(\mu)\right] - E_q[\log q(\tau)]
\end{aligned}
$$

<u>Expected log joint probability:</u>

$$
\begin{aligned}
1)\ E_q\left[\log p(\mathbf{x}|\mu,\tau)\right] &= \sum^N E_q\left[\log\mathcal{N}(\mu,\tau^{-1})\right] \\
&= \sum^N E_q\left[\log\left(\left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\tau}{2}(x_n-\mu)^2}\right)\right] \\
&= \sum^N E_q\left[\frac{1}{2}\log\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2}(x_n-\mu)^2\right] \\
&= \sum^N \frac{1}{2}\left(E_{q(\tau)}\left[\log\tau\right] - \log(2\pi)\right) - \sum^N \frac{E_{q(\tau)}[\tau]}{2} E_\mu[(x_n-\mu)^2] \\
&= \frac{N}{2}E_{q(\tau)}[\log\tau] - \frac{N}{2}\log(2\pi) - \frac{E_{q(\tau)}[\tau]}{2}\sum^N E_{q(\mu)}[(x_n-\mu)^2] \\
&= \frac{N}{2}\left(E_{q(\tau)}[\log\tau] - \log(2\pi)\right) - \frac{E_{q(\tau)}[\tau]}{2}\sum^N\left(x_n^2 + E_{q(\mu)}[\mu^2] - 2x_n E_{q(\mu)}[\mu]\right) \\
&= \frac{N}{2}\left(E_{q(\tau)}[\log\tau] - \log(2\pi)\right) \\
&\quad - \frac{E_{q(\tau)}[\tau]}{2}\left(NE_{q(\mu)}[\mu^2] + \sum^N x_n^2 - 2E_{q(\mu)}[\mu]\sum^N x_n\right)
\end{aligned}
$$

$$
\begin{aligned}
2)\ E_q\left[\log p(\mu|\tau)\right] &= E_q\left[\log\mathcal{N}\left(\mu_0,(\lambda_0\tau)^{-1}\right)\right] \\
&= E_q\left[\log\left(\left(\frac{\lambda_0\tau}{2\pi}\right)^{1/2} e^{-\frac{\lambda_0\tau}{2}(\mu_0-\mu)^2}\right)\right] \\
&= E_q\left[\frac{1}{2}\log\left(\frac{\lambda_0\tau}{2\pi}\right) - \frac{\lambda_0\tau}{2}(\mu_0-\mu)^2\right] \\
&= \frac{1}{2}\log\left(\frac{\lambda_0}{2\pi}\right) + \frac{1}{2}E_{q(\tau)}\left[\log\tau\right] - \frac{\lambda_0 E_{q(\tau)}[\tau]}{2}\left(E_{q(\mu)}[\mu^2] + \mu_0^2 - 2\mu_0 E_{q(\mu)}[\mu]\right)
\end{aligned}
$$

3) $E_q[\log p(\tau)] = E_q[a_0 \log b_0 - \log G(a_0) + (a_0 - 1)\log \tau - b_0 \tau]$

$$= a_0 \log b_0 - \log G(a_0) + (a_0 - 1)E_{q(\tau)}[\log \tau] - b_0 E_{q(\tau)}[\tau]$$

Variational entropy:

Respectively for a Normal distribution and for a Gamma distribution, the entropies are given by:

4) $-E_{q(\mu)}[\log q(\mu)] = H(\mathcal{N}(\mu^*, \lambda^{*-1})) = \frac{1}{2}\log\left(\frac{2\pi e}{\lambda^*}\right)$

5) $-E_{q(\tau)}[\log q(\tau)] = H(\mathcal{G}(a^*, b^*)) = a^* - \log b^* + \log \Gamma(a^*) + (1 - a^*)\psi(a^*)$

Full ELBO:

Finally, the full ELBO is given by:

$$\begin{aligned}
\mathcal{L}(q) =& \frac{N}{2}\left(E_{q(\tau)}[\log \tau] - \log(2\pi)\right) - \frac{E_{q(\tau)}[\tau]}{2}\left(N E_{q(\mu)}[\mu^2] + \sum^N x_n^2 - 2E_{q(\mu)}[\mu]\sum^N x_n\right) \\
&+ \frac{1}{2}\log\left(\frac{\lambda_0}{2\pi}\right) + \frac{1}{2}E_{q(\tau)}\left[\log \tau\right] - \frac{\lambda_0 E_{q(\tau)}[\tau]}{2}\left(E_{q(\mu)}[\mu^2] + \mu_0^2 - 2\mu_0 E_{q(\mu)}[\mu]\right) \\
&+ a_0 \log b_0 - \log G(a_0) + (a_0 - 1)E_{q(\tau)}[\log \tau] - b_0 E_{q(\tau)}[\tau] - \frac{1}{2}\log\left(\frac{2\pi e}{\lambda^*}\right) \\
&- a^* + \log b^* - \log G(a^*) - (1 - a^*)\psi(a^*)
\end{aligned}$$