Variational Inference

* Bivariate Gaussian with known parameters

$$\ln q_1^*(z_1) \stackrel{c}{=} \mathbb{E}_{q_2^*(z_2)}\left[\ln p(z)\right]$$

where $\ln p(z) \stackrel{c/z_1}{=} -\frac{1}{2}(z-\mu)^T \Lambda (z-\mu)$

$$\stackrel{c/z_1}{=} -\frac{1}{2}\begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}^T \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}$$

$$\stackrel{c/z_1}{=} -\frac{1}{2}\begin{pmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{pmatrix}^T \begin{pmatrix} \Lambda_{11}(z_1-\mu_1) + \Lambda_{12}(z_2-\mu_2) \\ \Lambda_{21}(z_1-\mu_1) + \Lambda_{22}(z_2-\mu_2) \end{pmatrix}$$

$$\stackrel{c/z_1}{=} -\frac{1}{2}\Big[(z_1-\mu_1)\big(\Lambda_{11}(z_1-\mu_1) + \Lambda_{12}(z_2-\mu_2)\big) + (z_2-\mu_2)\big(\Lambda_{21}(z_1-\mu_1) + \Lambda_{22}(z_2-\mu_2)\big)\Big]$$

$$\stackrel{c/z_1}{=} -\frac{1}{2}\left[(z_1-\mu_1)^2\Lambda_{11} + 2(z_1-\mu_1)\Lambda_{12}(z_2-\mu_2)\right]$$

$$\ln q_1^*(z_1) \overset{c}{=} \mathbb{E}_{q_2(z_2)}\left[-\frac{1}{2}(z_1-\mu_1)^2\Lambda_{11} - (z_1-\mu_1)\Lambda_{12}(z_2-\mu_2)\right]$$

$$\overset{c}{=} \mathbb{E}_{q_2(z_2)}\left[-\frac{1}{2}(z_1^2+\mu_1^2-2z_1\mu_1)\Lambda_{11} - z_1\Lambda_{12}(z_2-\mu_2)\right]$$

$$\overset{c}{=} \mathbb{E}_{q_2(z_2)}\left[-\frac{1}{2}z_1^2\Lambda_{11} + z_1\left(\mu_1\Lambda_{11} - \Lambda_{12}(z_2-\mu_2)\right)\right]$$

$$\overset{c}{=} -\frac{1}{2}z_1^2\Lambda_{11} + z_1\left(\mu_1\Lambda_{11} - \Lambda_{12}\left(\mathbb{E}_{q_2(z_2)}[z_2]-\mu_2\right)\right)$$

This is a quadratic function of z_1, so $q_1^*(z_1)$ is a Gaussian distribution:

$$\ln q_1^*(z_1) = \ln \mathcal{N}(z_1; m_1, \gamma_1^{-1})$$

$$\overset{c}{=} -\frac{1}{2}(z_1-m_1)^2\gamma_1$$

$$\overset{c}{=} -\frac{1}{2}(z_1^2-2z_1m_1)\gamma_1$$

$$\overset{c}{=} -\frac{1}{2}z_1^2\gamma_1 + z_1m_1\gamma_1$$

By identification, we have

- $\boxed{\gamma_1 = \Lambda_{11}}$
- $m_1 \gamma_1 = \mu_1 \Lambda_{11} - \Lambda_{12}\left(\mathbb{E}_{q_2(z_2)}[z_2] - \mu_2\right)$

$$\Leftrightarrow \boxed{m_1 = \mu_1 - \Lambda_{11}^{-1}\Lambda_{12}\left(\mathbb{E}_{q_2(z_2)}[z_2] - \mu_2\right)}$$

By symmetry, we also have

$$q_2^*(z_2) = \mathcal{N}(z_2; m_2, \gamma_2^{-1})$$

with
$$\begin{cases} m_2 = \mu_2 - \Lambda_{22}^{-1}\Lambda_{21}\left(\mathbb{E}_{q_1(z_1)}[z_1] - \mu_1\right) \\ \gamma_2 = \Lambda_{22} \end{cases}$$

To fully specify $q_1^*(z_1)$ & $q_2^*(z_2)$ we need to compute:

- $\mathbb{E}_{q_2(z_2)}[z_2] = m_2$
- $\mathbb{E}_{q_1(z_1)}[z_1] = m_1$

Univariate Gaussian with unknowns parameters

→ Data: $\underline{x} = \{x_1, x_2, \dots, x_N\}$

→ Likelihood:

$$p(\underline{x} \mid \mu, \tau) = \prod_{i=1}^{N} p(x_i \mid \mu, \tau)$$

$$= \prod_{i=1}^{N} \mathcal{N}(x_i ; \mu, \tau^{-1})$$

$$= \left(\frac{\tau}{2\pi}\right)^{N/2} \exp\left(-\frac{\tau}{2} \sum_{i=1}^{N} (x_i - \mu)^2\right)$$

→ Priors:

$$p(\mu \mid \tau) = \mathcal{N}(\mu ; \mu_0, (\lambda_0 \tau)^{-1}) = \left(\frac{\lambda_0 \tau}{2\pi}\right)^{1/2} \exp\left(-\frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2\right)$$

$$p(\tau) = \mathcal{G}(\tau, a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)} \tau^{(a_0 - 1)} \exp(-\tau b_0)$$

→ Complete-data likelihood:

$$p(\underline{x}, \mu, \tau) = p(\underline{x} \mid \mu, \tau)\, p(\mu \mid \tau)\, p(\tau)$$

→ True posterior

$$p(\mu, \tau \mid \underline{x}) = p(\mu \mid \underline{x}, \tau)\, p(\tau \mid \underline{x})$$

- $$\ln p(\mu \mid \underline{x}, \tau) \stackrel{c}{=} \ln p(\underline{x}, \mu, \tau)$$

$$\stackrel{c}{=} \ln p(\underline{x} \mid \mu, \tau) + \ln p(\mu \mid \tau)$$

$$\stackrel{c}{=} -\frac{\tau}{2} \sum_{i=1}^{N} (x_i - \mu)^2 - \frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2$$

$$\stackrel{c}{=} -\frac{\tau}{2} \sum_{i=1}^{N} (\mu^2 - 2x_i \mu) - \frac{\lambda_0 \tau}{2} (\mu^2 - 2\mu \mu_0)$$

$$\stackrel{c}{=} -\frac{1}{2} \mu^2 (N\tau + \lambda_0 \tau) + \mu \left(\tau \sum_{i=1}^{N} x_i + \mu_0 \lambda_0 \tau \right)$$

We recognize $\ln p(\mu \mid \underline{x}, \sigma^2) = \ln \mathcal{N}(\mu ; \mu_*, \lambda_*^{-1})$

$$\stackrel{c}{=} -\frac{\lambda_*}{2} (\mu - \mu_*)^2$$

$$\stackrel{c}{=} -\frac{\lambda_*}{2} \mu^2 + \lambda_* \mu_* \mu$$

By identification,

$$\boxed{\lambda_* = N\tau + \lambda_0 \tau}$$

and $\lambda_* \mu_* = \tau \sum_{i=1}^{N} x_i + \mu_0 \lambda_0 \tau$

Let us define $\bar{x} \triangleq \frac{1}{N} \sum_{i=1}^{N} x_i$, we have.

$$\mu_* = \frac{1}{\lambda^*}\left(N\tau\bar{x} + \mu_0 \lambda_0 \tau\right)$$

$$\boxed{\mu_* = \frac{N\tau}{N\tau + \lambda_0 \tau}\bar{x} + \frac{\lambda_0 \tau}{N\tau + \lambda_0 \tau}\mu_0}$$

- $p(\tau | \underline{x}) = \int p(\tau, \mu | \underline{x})\, d\mu$ (def. by marginalization)

$$\ln p(\tau, \mu | \underline{x}) \overset{c}{=} \ln p(\underline{x} | \mu, \tau) + \ln p(\mu | \tau) + \ln p(\tau)$$

$$\overset{c}{=} \frac{N}{2}\ln(\tau) - \frac{\tau}{2}\sum_{i=1}^{N}(x_i - \mu)^2 + \frac{1}{2}\ln(\tau) - \frac{\lambda_0 \tau}{2}(\mu - \mu_0)^2$$

$$+ (a_0 - 1)\ln(\tau) - \tau b_0$$

$$\ln p(\tau,\mu|\underline{x}) \overset{c}{=} \ln(\tau)\left(\frac{N}{2}+\frac{1}{2}+a_0-1\right)$$

$$-\tau\left(b_0+\frac{\lambda_0}{2}(\mu-\mu_0)^2+\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right)$$

$$\Leftrightarrow p(\tau,\mu|\underline{x}) \propto \tau^{a_0+\frac{N}{2}-1} \times \exp\left(-\tau\left(b_0+\frac{\lambda_0}{2}(\mu-\mu_0)^2+\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right)\right)$$

We now need to integrate w.r.t. μ:

$$\int p(\tau,\mu|\underline{x})\,d\mu \propto \tau^{a_0+\frac{N}{2}-1} e^{-\tau b_0} \underbrace{\int e^{-\tau\left[\frac{\lambda_0}{2}(\mu-\mu_0)^2+\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\right]} d\mu}_{(*)}$$

$$(*) = \int \exp\left\{-\tau\left[\frac{\lambda_0}{2}(\mu^2+\mu_0^2-2\mu\mu_0)+\frac{1}{2}\sum_{i=1}^{N}(x_i^2+\mu^2-2x_i\mu)\right]\right\} d\mu$$

$$= \exp\left\{-\tau\left(\frac{\lambda_0}{2}\mu_0^2+\frac{1}{2}\sum_{i=1}^{N}x_i^2\right)\right\} \underbrace{\int \exp\left\{-\tau\left[\frac{\lambda_0}{2}(\mu^2-2\mu\mu_0)+\frac{N}{2}\mu^2-\mu\sum_i x_i\right]\right\} d\mu}_{(**)}$$

$$(**) = \int \exp\left\{ -\tau \left[\mu^2 \frac{\lambda_0 + N}{2} - \mu\left(\lambda_0\mu_0 + \sum_i x_i \right) \right] \right\} d\mu$$

$$= \int_{-\infty}^{+\infty} \exp\left\{ -\mu^2 \frac{\tau(\lambda_0+N)}{2} + \mu\tau\left(\lambda_0\mu_0 + \sum_i x_i\right) \right\} d\mu$$

$$\boxed{a = \frac{\tau(\lambda_0+N)}{2} \qquad b = -\frac{\tau}{2}\left(\lambda_0\mu_0 + \sum_i x_i\right)}$$

We need to compute

$$I = \int_{-\infty}^{+\infty} \exp\left(-ax^2 - 2bx\right) dx$$

$$= \int_{-\infty}^{+\infty} \exp\left(-ax^2 - 2bx - \frac{b^2}{a} + \frac{b^2}{a}\right) dx$$

$$= \exp\left(\frac{b^2}{a}\right) \int_{-\infty}^{+\infty} \exp\left(-\left(\sqrt{a}\,x + \frac{b}{\sqrt{a}}\right)^2\right) dx$$

$\underline{\text{cov}}$: $t = \sqrt{a}\,x + \frac{b}{\sqrt{a}} \Rightarrow dt = \sqrt{a}\,dx$

$$I = \exp\left(\frac{b^2}{a}\right) \times \frac{1}{\sqrt{a}} \underbrace{\int_{-\infty}^{+\infty} \exp\left(-t^2\right) dt}_{= \sqrt{\pi} \text{ (Gaussian integral)}} = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{a}\right)$$

So,

$$(**) = \sqrt{\frac{2\pi}{\tau(\lambda_0+N)}}\ \exp\left(\frac{\tau^{\cancel{2}}}{4}\left(\lambda_0\mu_0+\sum_i x_i\right)^2 \cdot \frac{2}{\cancel{\tau}(\lambda_0+N)}\right)$$

$$= \sqrt{\frac{2\pi}{\tau(\lambda_0+N)}}\ \exp\left(\tau\,\frac{\left(\lambda_0\mu_0+\sum_i x_i\right)^2}{2(\lambda_0+N)}\right)$$

and

$$(*) = \sqrt{\frac{2\pi}{\tau(\lambda_0+N)}}\ \exp\left\{-\tau\left(\frac{\lambda_0}{2}\mu_0^2 - \frac{\left(\lambda_0\mu_0+\sum_i x_i\right)^2}{2(\lambda_0+N)} + \frac{1}{2}\sum_{i=1}^{N} x_i^2\right)\right\}$$

and

$$\underbrace{\int p(\tau,\mu\,|\,\underline{x})\,d\mu}_{\parallel \atop p(\tau|\underline{x})} \propto \tau^{a_0+\frac{N}{2}-1}\ \exp\left\{-\tau\left(b_0 + \frac{\lambda_0}{2}\mu_0^2 - \frac{\left(\lambda_0\mu_0+\sum_i x_i\right)^2}{2(\lambda_0+N)} + \frac{1}{2}\sum_i x_i^2\right)\right\}$$

We recognize $p(\tau|x) = G(\tau, \alpha, \beta)$

$$\propto \tau^{(\alpha-1)}\ \exp(-\tau\beta)$$

ere

$$\boxed{\begin{aligned} &\cdot\ \alpha = a_0 + \frac{N}{2} \\ &\cdot\ \beta = b_0 + \frac{1}{2}\sum_{i=1}^{N} x_i^2 + \frac{\lambda_0}{2}\mu_0^2 - \frac{(\lambda_0\mu_0 + N\bar{x})^2}{2(\lambda_0+N)} \end{aligned}}$$

We can further rewrite β as follows :

$$\beta = b_0 + \frac{1}{2}\sum_{i=1}^{N}(x_i - \bar{x} + \bar{x})^2 + \frac{\lambda_0}{2}\mu_0^2 - \frac{\lambda_0\mu_0^2 + 2\lambda_0\mu_0 N\bar{x} + N^2\bar{x}^2}{2(\lambda_0+N)}$$

$$= b_0 + \frac{1}{2}\sum_{i=1}^{N}\left((x_i-\bar{x})^2 + \bar{x}^2 + 2\bar{x}(x_i-\bar{x})\right) + \frac{\lambda_0}{2}\mu_0^2 - \frac{\lambda_0^2\mu_0^2 + 2\lambda_0\mu_0 N\bar{x} + N^2\bar{x}^2}{2(\lambda_0+N)}$$

$$= b_0 + \frac{1}{2}\sum_{i=1}^{N}(x_i-\bar{x})^2 + \frac{N}{2}\bar{x}^2 + \cancel{N\bar{x}^2} - \cancel{N\bar{x}^2} + \frac{\lambda_0}{2}\mu_0^2 - \frac{\lambda_0^2\mu_0^2 + 2\lambda_0\mu_0 N\bar{x} + N^2\bar{x}^2}{2(\lambda_0+N)}$$

$$= b_0 + \frac{1}{2}\sum_i (x_i-\bar{x})^2 + \mu_0^2\left(\frac{\lambda_0}{2} - \frac{\lambda_0^2}{2(\lambda_0+N)}\right) + \bar{x}^2\left(\frac{N}{2} - \frac{N^2}{2(\lambda_0+N)}\right) - \frac{\lambda_0 N}{(\lambda_0+N)}\mu_0\bar{x}$$

$$= b_0 + \frac{1}{2}\sum_i (x_i-\bar{x})^2 + \frac{2\lambda_0(\lambda_0+N) - 2\lambda_0^2}{4(\lambda_0+N)}\mu_0^2 + \frac{2N(\lambda_0+N) - 2N^2}{4(\lambda_0+N)}\bar{x}^2 - \frac{\lambda_0 N}{(\lambda_0+N)}\mu_0\bar{x}$$

$$= b_0 + \frac{1}{2}\sum_i (x_i-\bar{x})^2 + \frac{\lambda_0 N\mu_0^2 + \lambda_0 N\bar{x}^2 - 2\lambda_0 N\mu_0\bar{x}}{2(\lambda_0+N)}$$

$$= b_0 + \frac{1}{2}\sum_i (x_i-\bar{x})^2 + \frac{\lambda_0 N}{2(\lambda_0+N)}(\bar{x}-\mu_0)^2$$

In summary :

$$p(\tau \mid x) = \mathcal{G}(\tau ; \alpha, \beta)$$

with $\quad \alpha = a_0 + \frac{N}{2}$

$$\beta = b_0 + \frac{1}{2} \sum_i (x_i - \bar{x})^2 + \frac{\lambda_0 N}{2(\lambda_0 + N)} (\bar{x} - \mu_0)^2$$

▷ Variational inference

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

- $$\ln q^*_\mu(\mu) \overset{c}{=} \mathbb{E}_{q_\tau(\tau)}\left[\ln p(\underline{x}, \mu, \tau) \right]$$

$$\overset{c}{=} \mathbb{E}_{q_\tau(\tau)}\left[\ln p(\underline{x} \mid \mu, \tau) + \ln p(\mu \mid \tau) \right]$$

$$\overset{c}{=} \mathbb{E}_{q_\tau(\tau)}\left[-\frac{\tau}{2} \sum_i (x_i - \mu)^2 - \frac{\lambda_0 \tau}{2} (\mu - \mu_0)^2 \right]$$

$$\overset{c}{=} -\frac{1}{2} \mathbb{E}_{q_\tau(\tau)}[\tau] \left(N\mu^2 - 2\mu \sum_i x_i + \lambda_0 \mu^2 - 2\lambda_0 \mu_0 \mu \right)$$

$$= -\frac{1}{2} \mathbb{E}_{q_\tau(\tau)}[\tau] \left(\mu^2 (N + \lambda_0) - 2\mu (N\bar{x} + \lambda_0 \mu_0) \right)$$

We recognize $\ln q^*_\mu(\mu) = \ln \mathcal{N}(\mu; \mu_N, \lambda_N^{-1}) \overset{c}{=} -\frac{1}{2} \lambda_N \mu^2 + \lambda_N \mu_N \mu$

and identify

$$\boxed{\lambda_N = \mathbb{E}_{q_\tau(\tau)}[\tau] (\lambda_0 + N)}$$

$$\mu_N = \frac{1}{\lambda_N} \mathbb{E}_{q_\tau(\tau)}[\tau] \left(\lambda_0 \mu_0 + N \bar{x} \right)$$

$$\boxed{\mu_N = \frac{\lambda_0 \mu_0 + N \bar{x}}{\lambda_0 + N}}$$

Note that $\begin{cases} \lim\limits_{N \to +\infty} \mu_N = \bar{x} = \frac{1}{N} \sum_i x_i \\ \lim\limits_{N \to +\infty} \lambda_N^{-1} = 0 \end{cases}$

which gives a dirac centered on the ML estimate of the mean.

$$\bullet \ \ln q^*_\tau(\tau) \overset{c}{=} \mathbb{E}_{q_\mu(\mu)}\left[\ln p(\underline{x}, \mu, \tau)\right]$$

$$\overset{c}{=} \mathbb{E}\left[\ln p(\underline{x} \mid \mu, \tau) + \ln p(\mu \mid \tau) + \ln p(\tau)\right]$$

$$\overset{c}{=} \mathbb{E}\left[\frac{N}{2}\ln(\tau) - \frac{\tau}{2}\sum_i (x_i - \mu)^2 + \frac{1}{2}\ln(\tau) - \frac{\lambda_0 \tau}{2}(\mu - \mu_0)^2 \right.$$

$$\left. + (a_0 - 1)\ln(\tau) - b_0 \tau \right]$$

$$\overset{c}{=} \ln(\tau)\left[\frac{N}{2} + \frac{1}{2} + a_0 - 1\right] - \tau\left[b_0 + \frac{1}{2}\mathbb{E}\left[\sum_i (x_i - \mu)^2\right] + \frac{\lambda_0}{2}\mathbb{E}\left[(\mu - \mu_0)^2\right]\right]$$

We recognize $\ln q^*_\tau(\tau) = \ln \mathcal{G}(\tau, a_N, b_N) \overset{c}{=} (a_N - 1)\ln(\tau) - b_N \tau$

and identify

$$\boxed{\begin{aligned} a_N &= a_0 + \frac{N+1}{2} \\ b_N &= b_0 + \frac{1}{2}\mathbb{E}_{q_\mu(\mu)}\left[\sum_i (x_i - \mu)^2 + \lambda_0 (\mu - \mu_0)^2\right] \end{aligned}}$$

The required expectat° are given by :

$$\begin{cases} \mathbb{E}_{q_\tau(\tau)}[\tau] = \dfrac{a_N}{b_N} \\ \mathbb{E}_{q_\mu(\mu)}[\mu] = \mu_N \\ \mathbb{E}_{q_\mu(\mu)}[\mu^2] = \mathbb{E}_{q_\mu(\mu)}[\mu]^2 + \mathbb{E}_{q_\mu(\mu)}\left[\left(\mu - \mathbb{E}_{q_\mu(\mu)}[\mu]\right)^2\right] \\ \qquad = \mu_N^2 + \lambda_N^{-1} \end{cases}$$

We can rewrite b_N as :

$$b_N = b_0 + \frac{1}{2}\mathbb{E}\left[\sum_i x_i^2 + N\mu^2 - 2\mu\sum_i x_i + \lambda_0\mu^2 + \lambda_0\mu_0^2 - 2\lambda_0\mu\mu_0\right]$$

$$b_N = b_0 + \frac{1}{2}\left(\sum_i x_i^2 + \mathbb{E}[\mu^2](\lambda_0 + N) - 2\mathbb{E}[\mu](\lambda_0\mu_0 + N\bar{x}) + \lambda_0\mu_0^2\right)$$