Exercise 1. *Variational posterior approximation*

We have seen that maximising the evidence lower bound (ELBO) with respect to the variational distribution q minimises the Kullback-Leibler divergence to the true posterior p. We here assume that q and p are probability density functions so that the Kullback-Leibler divergence between them is defined as

$$\mathrm{KL}(q||p) = \int q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} \mathrm{d}\mathbf{x} = \mathbb{E}_q \left[\log \frac{q(\mathbf{x})}{p(\mathbf{x})} \right]. \tag{1}$$

(a) You can here assume that $\mathbf{x}$ is one-dimensional so that p and q are univariate densities. Consider the case where p is a bimodal density but the variational densities q are unimodal. Sketch a figure that shows p and a variational distribution q that has been learned by minimising $\mathrm{KL}(q||p)$. Explain qualitatively why the sketched q minimises $\mathrm{KL}(q||p)$.

(b) Assume that the true posterior $p(\mathbf{x}) = p(x_1, x_2)$ factorises into two Gaussians of mean zero and variances σ_1^2 and σ_2^2,

$$p(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{x_1^2}{2\sigma_1^2}\right] \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{x_2^2}{2\sigma_2^2}\right]. \tag{2}$$

Assume further that the variational density $q(x_1, x_2; \lambda^2)$ is parameterised as

$$q(x_1, x_2; \lambda^2) = \frac{1}{2\pi\lambda^2} \exp\left[-\frac{x_1^2 + x_2^2}{2\lambda^2}\right] \tag{3}$$

where λ^2 is the variational parameter that is learned by minimising $\mathrm{KL}(q||p)$. If σ_2^2 is much larger than σ_1^2, do you expect λ^2 to be closer to σ_2^2 or to σ_1^2? Provide an explanation.

Exercise 2. *Gaussian mean field variational inference*

Assume random variables y_1, y_2, x are generated according to the following process

$$y_1 \sim \mathcal{N}(y_1; 0, 1) \qquad y_2 \sim \mathcal{N}(y_2; 0, 1) \tag{4}$$

$$n \sim \mathcal{N}(n; 0, 1) \qquad x = y_1 + y_2 + n \tag{5}$$

where y_1, y_2, n are statistically independent.

The three questions below can be solved independently of each other.

(a) y_1, y_2, x are jointly Gaussian. Determine their mean and their covariance matrix, and hence their joint distribution.

(b) The conditional $p(y_1, y_2|x)$ is Gaussian with mean $\mathbf{m}$ and covariance $\mathbf{C}$,

$$\mathbf{m} = \frac{x}{3}\begin{pmatrix}1\\1\end{pmatrix} \qquad \mathbf{C} = \frac{1}{3}\begin{pmatrix}2 & -1\\-1 & 2\end{pmatrix} \tag{6}$$

Since x is the sum of three random variables that have the same distribution, it makes intuitive sense that, given x, the conditional mean of y_1, y_2 is $1/3$ of the observed value of x.

©Gutmann, University of Edinburgh, 2026 CC BY 4.0

Moreover, y_1 and y_2 are negatively correlated since an increase in y_1 must be compensated with a decrease in y_2.

We now approximate the posterior $p(y_1, y_2|x)$ with mean field variational inference. Denoting the variational distribution by $q(\mathbf{y}|x) = q(y_1|x)q(y_2|x)$, derive the update rules for the marginals $q(y_i|x)$.

Hints:

1. For a model $p(\mathbf{v}, \mathbf{h})$ on observed variables $\mathbf{v}$ and unobserved variables $\mathbf{h}$, in mean-field variational inference, each q_i is iteratively updated as

$$q_i(h_i|\mathbf{v}) = \frac{1}{Z} \exp\left[\mathbb{E}_{q(\mathbf{h}_{\setminus i}|\mathbf{v})}\left[\log p(\mathbf{v}, \mathbf{h})\right]\right] \tag{7}$$

   where $q(\mathbf{h}_{\setminus i}|\mathbf{v}) = \prod_{j \neq i} q_j(h_j|\mathbf{v})$ is the product of all marginals without marginal $q_i(h_i|\mathbf{v})$. It is typically easier to work with the update rule in the log-domain.

2. Omitting constants, the log pdf of $p(y_1, y_2, x)$ is

$$\log p(y_1, y_2, x) = -\frac{1}{2}\left(2y_1^2 + 2y_2^2 + x^2 + 2y_1y_2 - 2y_1x - 2y_2x\right) + \text{const} \tag{8}$$

3. A Gaussian density $\mathcal{N}(y; m, \sigma^2)$ with mean m and variance σ^2 takes, in the log-domain, the following form:

$$\log \mathcal{N}(y; m, \sigma^2) = -\frac{1}{2}\frac{(y-m)^2}{\sigma^2} + \text{const} \tag{9}$$

$$= -\frac{1}{2}\left(\frac{y^2}{\sigma^2} - 2y\frac{m}{\sigma^2}\right) + \text{const} \tag{10}$$

(c) The update rules from the previous questions shows that the marginal variational distributions $q_1(y_1|x)$ and $q_2(y_2|x)$ are both Gaussian with variance equal to $1/2$ and their means are iteratively updated as

$$m_1 = \frac{1}{2}(x - m_2) \qquad\qquad m_2 = \frac{1}{2}(x - m_1) \tag{11}$$

These are two equations for two unknowns that we can solve in closed form. Derive the closed-form expression for the optimal mean-field approximation and compare it with the true conditional $p(y_1, y_2|x)$.

©Gutmann, University of Edinburgh, 2026 CC BY 4.0