Exercise 1. ***Gaussian mean field variational inference***

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

$$y_1 \sim \mathcal{N}(y_1; 0, 1) \qquad y_2 \sim \mathcal{N}(y_2; 0, 1) \tag{1}$$
$$n \sim \mathcal{N}(n; 0, 1) \qquad x = y_1 + y_2 + n \tag{2}$$

where y_1, y_2, n are statistically independent.

(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 $\mathbf{C}$,

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

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. 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

$$q_i(h_i|\mathbf{v}) = \frac{1}{Z}\exp\left[\mathbb{E}_{q(\mathbf{h}_{\setminus i}|\mathbf{v})}\left[\log p(\mathbf{v}, \mathbf{h})\right]\right] \tag{4}$$

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 $q_i(h_i|\mathbf{v})$.

2. You may use that

$$p(y_1, y_2, x) = \mathcal{N}\left((y_1, y_2, x); \mathbf{0}, \mathbf{\Sigma}\right) \quad \mathbf{\Sigma} = \begin{pmatrix}1 & 0 & 1\\0 & 1 & 1\\1 & 1 & 3\end{pmatrix} \quad \mathbf{\Sigma}^{-1} = \begin{pmatrix}2 & 1 & -1\\1 & 2 & -1\\-1 & -1 & 1\end{pmatrix} \tag{5}$$

(c) In this example, the update rules result in two equations for two unknowns that can be solved in closed form. Derive the closed-form expression for the optimal mean-field approximation and compare it with the true conditional $p(y_1, y_2|x)$.

Exercise 2. ***Variational posterior approximation***

We have seen that maximising the evidence lower bound (ELBO) with respect to the variational distribution minimises the Kullback-Leibler divergence to the true posterior. We here investigate the nature of the approximation if the family of variational distributions does not include the true posterior.

(a) Assume that the true posterior for $\mathbf{x} = (x_1, x_2)$ is given by

$$p(\mathbf{x}) = \mathcal{N}(x_1; 0, \sigma_1^2)\mathcal{N}(x_2; 0, \sigma_2^2) \tag{6}$$

and that our variational distribution $q(\mathbf{x}; 0, \lambda^2)$ is

$$q(\mathbf{x}; \lambda^2) = \mathcal{N}(x_1; 0, \lambda^2)\mathcal{N}(x_2; 0, \lambda^2), \tag{7}$$

where $\lambda > 0$ is the variational parameter. Provide an equation for

$$J(\lambda) = \mathrm{KL}(q(\mathbf{x}; \lambda^2)||p(\mathbf{x})), \tag{8}$$

where you can omit additive terms that do not depend on λ.

(b) Determine the value of λ that minimises $J(\lambda) = \mathrm{KL}(q(\mathbf{x}; \lambda^2)||p(\mathbf{x}))$. Interpret the result and relate it to properties of the Kullback-Leibler divergence.