1. Consider a univariate Gaussian distribution $\mathcal{N}(x|\mu, \tau^{-1})$ having conjugate Gaussian-gamma prior given by (1), and a data set $\mathbf{x} = \{x_1, \ldots, x_N\}$ of i.i.d. observations. Show that the posterior distribution is also a Gaussian-gamma distribution of the same functional form as the prior, and write down expressions for the parameters of this posterior distribution. (Exercise 2.44 [Bishop, 2006])

$$p(\mu, \lambda) = \mathcal{N}(\mu|\mu_0, (\beta\lambda)^{-1}) \, \mathrm{Gam}(\lambda|a, b) \tag{1}$$

2. Analyze example 10.1.3 from [Bishop, 2006]. a) Generate data from the following model:

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

   where the prior for μ and τ are given by conjugate priors as

$$p(\mu|\tau) = \mathcal{N}\left(\mu|\mu_0, (\lambda_0\tau)^{-1}\right)$$
$$p(\tau) = \mathrm{Gam}\left(\tau|a_0, b_0\right)$$

   b) Adjust the variational model using the data.

3. Suppose that $p(x)$ is some fixed distribution and that we wish to approximate it using a Gaussian distribution $q(x) = \mathcal{N}(x|\mu, \Sigma)$. By writing down the form of the KL divergence $\mathrm{KL}(p||q)$ for a Gaussian $q(x)$ and then differentiating, show that minimization of $\mathrm{KL}(p||q)$ with respect to μ and Σ leads to the result that μ is given by the expectation of x under $p(x)$ and that Σ is given by the covariance. (Exercise 10.4, [Bishop, 2006])

References

[Bishop, 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.