MLHEP 2021 - Bayesian Methods Basics - problem set

1. [Basic Bayesian reasoning] During medical checkup, one of the tests indicates a serious disease. The test has high accuracy 99% (probability of true positive is 99%, probability of true negative is 99%). However, the disease is quite rare, and only one person of 10000 is affected. Calculate the probability that the examined person has the disease.

2. [Data modeling: Bayesian vs frequentist] Consider a toy one-dimensional linear regression problem. Let $\mathcal{D} = \{x, y\}$ be a training dataset of K points, where $x \in \mathbb{R}^K$ is a vector of features and $y \in \mathbb{R}^K$ is a vector of target variables. In order to train a linear regression model on the training data, we assume that the likelihood has the form of normal distribution:
$$p(y|x,w) = \prod_{i=1}^{K} \mathcal{N}(y_i|wx_i, 1) = \mathcal{N}(y|wx, I),$$
where $w \in \mathbb{R}$ is a trainable weight.

   We would like to estimate w in a frequentist and Bayesian way. In order to perform an analytical Bayesian inference, we choose a conjugate prior. The conjugate prior distribution for the normal likelihood is a Normal distribution. We choose the following prior form:
$$p(w) = \mathcal{N}(w|0, \alpha^{-1}), \quad \alpha \in \mathbb{R}^+$$
   (a) Compute the maximum likelihood estimate for w.
   (b) Check that the Normal distribution is indeed the conjugate distribution for the Normal likelihood.
   (c) Compute the posterior distribution $p(w|x, y)$.
   (d) Compute a maximum a posteriori estimate for w and compare it with the maximum likelihood estimate.
   * Compute the posterior predictive distribution $p(y_{new}|x_{new}, x, y) = \int p(y_{new}|x_{new}, w)p(w|x, y)\mathrm{d}w$.

3. [Variational Inference] Consider a toy one-dimensional binary classification problem. Let $\mathcal{D} = \{x, y\}$ be a training dataset of K points, where $x \in \mathbb{R}^K$ is a vector of features and $y \in \{-1, 1\}^K$ is a vector of target variables. In order to train a linear classification model on the training data, we assume that the likelihood has the following form:
$$p(y|x,w) = \prod_{i=1}^{K} \sigma(y_i x_i w), \quad \sigma(z) = \frac{1}{1+\exp(-z)},$$
where $w \in \mathbb{R}$ is a trainable weight.

   We would like to estimate w in Bayesian way, therefore we choose a prior distribution:
$$p(w) = \mathcal{N}(w|0, \alpha^{-1}), \quad \alpha \in \mathbb{R}^+$$
   (a) Check that our likelihood and prior distributions are not conjugate.
   (b) Suggest the parametric family of distributions for the approximate posterior $q(w) \approx p(w|x, y)$.
   (c) Write down the optimization problem for the training with parametric variational inference.

In tasks 2 and 3, we use the following parametrization for normal distribution:
$$\mathcal{N}(x|\mu, \lambda^{-1}) = \sqrt{\frac{\lambda}{(2\pi)}} \exp\left(-\frac{1}{2}\lambda(x-\mu)^2\right)$$