02477 Practice exam set ($\approx 50\%$ of a full exam set)

- The exam will be a 4 hours with all aids allowed except internet
- This test exam set amounts to approximately 50% of a full exam set
- The solution to the actual exam must be handed in digitally as a PDF.
- Report all results with 2 digits after the decimal point.
- Explain how you arrived at your results, document intermediate results etc.

Part 1

Consider the following regression model

$$y(x) = f(x) + e = w_0 + w_1 x^2 + w_2 \sin x + w_3 x + e, \tag{1}$$

such that $y_n = f(x_n) + e_n$, where $x_n, y_n \in \mathbb{R}$ are input and targets, respectively. The additive noise $e_n \in \mathbb{R}$ is assumed to i.i.d from a zero-mean Gaussian distribution, i.e. $e_n \sim \mathcal{N}(0, \beta^{-1})$ for $\beta > 0$.

Let $\mathbf{x} = [2.29, -1.8, -0.06, 3.72, 2.6, -5.93, -0.15]$ and $\mathbf{y} = [3.17, -4.53, -0.78, 3.15, 4.76, -1.96, -1.32]$ denote the vector of inputs and targets, respectively, for a dataset with $N = 5$ observations.

Let $\mathbf{w} = [w_0, w_1, w_2, w_3] \in \mathbb{R}^4$ denote the parameter vector.

Question 1.1: Compute and report a maximum likelihood estimate for $\mathbf{w}$ and β.

Question 1.2: Compute the posterior predictive distribution $p(y^*|\boldsymbol{y}, x^* = 1)$, where $y^* = y(x^*)$ using a plug-in approximation based on the maximum likelihood estimators for $\boldsymbol{w}$ and β. Report the mean, standard deviation and a 95% credibility interval for y^*

Next, we impose i.i.d Gaussian priors on all regression coefficients $w_j \sim \mathcal{N}(0, \alpha^{-1})$ for $j = 0, 1, 2, 3$ and assume $\alpha = 1$ and $\beta = \frac{1}{2}$.

Question 1.3: Compute and report the posterior mean and marginal posterior standard deviation for each regression coefficient in $\mathbf{w}$.

Question 1.4: Compute the analytical posterior predictive density $p(y^*|\boldsymbol{y}, x^*)$ for $x^* = 1$.

Question 1.5: State the analytical expression for the marginal likelihood $p(\boldsymbol{y}|\alpha, \beta)$ and compute the value of $\log p(\boldsymbol{y}|\alpha = 1, \beta = \frac{1}{2})$.

Consider now the following hyperprior distribution for α and β:

$$p(\alpha, \beta) = \text{Gamma}(\alpha|1, 1)\text{Gamma}(\beta|1, 1) \tag{2}$$

Question 1.6: Use the Metropolis-Hastings algorithm to generate posterior samples from the distribution $p(\alpha, \beta|\boldsymbol{y})$. Run 2 chains for 2000 iterations each. Initialize the first chain using $\alpha = 1$ and $\beta = 1$ and the second chain using $\alpha = 10$ and $\beta = 10$. Choose an appropriate proposal variance and justify your choice. Plot the trace of both parameters.

Question 1.7: Use the samples to compute a Monte Carlo estimate for the posterior mean of α and β and report the MCSE for both estimates.

Part 2

Suppose the outcome of $N = 31$ independent Bernoulli trials generated $y = 7$ successes. Let $\theta \in [0, 1]$ denote the probability of success. Assume a Binomial likelihood, i.e. $p(y|\theta) = \text{Bin}(y|N, \theta)$ with the following prior distribution for θ:

$$p(\theta) = \frac{3}{7}\text{Beta}(\theta|2, 10) + \frac{4}{7}\text{Beta}(\theta|10, 2) \tag{3}$$

Question 2.1: Compute the prior probability of the event $\theta > \frac{1}{2}$.

Question 2.2: Compute the analytical marginal likelihood $p(y)$ and evaluate $p(y = 7)$.

Part 3

Consider the generalized linear model with a Poisson likelihood

$$y_n|\boldsymbol{w}, x_n \sim \text{Poisson}(\lambda_n) \tag{4}$$
$$\lambda_n = e^{w_0 + w_1 x_n} \tag{5}$$
$$\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \alpha^{-1}\boldsymbol{I}), \tag{6}$$

where $\boldsymbol{w} = [w_0, w_1]$ for the following dataset $\mathcal{D} = \{x_n, y_n\}$, for $N = 5$, where $\boldsymbol{x} = [1, 2, 4, 8, 10]$ and $\boldsymbol{y} = [5, 4, 1, 0, 0]$. Assume $\alpha = \frac{1}{4}$.

Question 3.1: Plot the contours of the prior distribution, the log likelihood and the posterior for the ranges $w_0 \in [-3.5, 3.5]$ and $w_1 \in [-3.5, 3.5]$.

Question 3.2: Write the logarithm of the joint distribution $p(\boldsymbol{y}, \boldsymbol{w})$ and absorb all terms that are constant wrt. $\boldsymbol{w}$ into a constant $K \in \mathbb{R}$.

Next, assume $\boldsymbol{w}_{MAP} = \left[2.1575, -0.5201\right]^T$ is a MAP estimator for $\boldsymbol{w}$.

Question 3.3: Compute the Hessian of $\log p(\boldsymbol{y}, \boldsymbol{w})$ with respect to $\boldsymbol{w}$ and evaluate it at the mode of $p(\boldsymbol{w}|\boldsymbol{y})$.

If you did not answer the previous question, assume the Hessian at the mode is

$$\mathbf{H} = \begin{bmatrix} -9 & -17 \\ -17 & -48 \end{bmatrix} \tag{7}$$

Question 3.4: Construct a Laplace approximation of $p(\boldsymbol{w}|\boldsymbol{y})$.

Question 3.5: Compute the mean and variance of the posterior predictive probability $p(y^*|\boldsymbol{y}, x^* = 0)$, where $y^* = y(x^*)$ via the Laplace approximation and Monte Carlo sampling. Use $S = 1000$ Monte Carlo samples.

Question 3.6: What would happen to the posterior predictive distribution $p(y^*|\boldsymbol{y}, x^* = 0)$ if $\alpha \to \infty$? Explain your reasoning.