Bayesian Modelling – Problem Sheet 7

Problem 1

Let $X_1, \ldots, X_n$ be n independently and identically distributed random variables such that $X_1 \sim \text{Laplace}(\theta_0)$ for some $\theta_0 \in \Theta := (0, +\infty)$, an unknown parameter. To estimate θ_0 we have an observation $x := (x_1, \ldots, x_n)$ of $(X_1, \ldots, X_n)$. The likelihood of $x_1 \in \mathbb{R}$ given θ is

$$\tilde{f}(x_1|\theta) = \frac{1}{2\theta} e^{-\frac{|x_1|}{\theta}}.$$

Below we denote by $f(x|\theta)$ the full likelihood; that is $f(x|\theta) = \prod_{k=1}^n \tilde{f}(x_k|\theta)$, while the prior distribution for θ is assumed to be the Inv-Gamma(α_0, α_0) distribution, with $\alpha_0 \in (0, +\infty)$ a fixed hyper-parameter.

Reminder: For parameters $\alpha, \beta > 0$, the density of the Inv-Gamma(α, β) distribution is given by

$$f_{\alpha,\beta}(z) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} e^{-\frac{\beta}{z}}, \quad z \in (0, +\infty).$$

1. Show that the density of the posterior distribution of θ given x is given by

$$\pi(\theta|x) = \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \theta^{-\alpha_n - 1} e^{-\frac{\beta_n}{\theta}}, \quad \theta \in \Theta$$

   with $\alpha_n = \alpha_0 + n$ and $\beta_n = \alpha_0 + \sum_{k=1}^n |x_k|$. Is the prior conjugate? Explain your answer.

2. Show that, for $n \geq 2$, the posterior mean and the posterior variance are given by

$$\mathbb{E}_\pi[\theta|x] = \frac{\beta_n}{\alpha_n - 1}, \quad \text{Var}_\pi(\theta|x) = \frac{\beta_n^2}{(\alpha_n - 1)^2(\alpha_n - 2)}.$$

   Hint: Recall that, for any $t > 0$, the relationship $\Gamma(t+1) = t\,\Gamma(t)$ holds.

3. Show that, under a quadratic loss function, $\mathbb{E}_\pi[\theta|\cdot]$ is not a Bayes estimator when $\alpha_0 \in (0, 1)$ and $n = 1$.

4. Let $X^{(n)} = (X_1, \ldots X_n)$. Show that $\lim_{n \to +\infty} \mathbb{E}_\pi[\theta|X^{(n)}] = \theta_0$ (almost surely) while $\lim_{n \to +\infty} \text{Var}_\pi(\theta|X^{(n)}) = 0$ (almost surely).

5. We now consider the loss function $L : \Theta \times \Theta \to \mathbb{R}_+$ defined by

$$L(\theta, d) = \left(\frac{d}{\theta} - 1\right)^2, \quad (\theta, d) \in \Theta^2.$$

   Show that $\delta^\pi(x)$, the minimizer of the corresponding posterior expected loss, is given by

$$\delta^\pi(x) = \frac{\beta_n}{\alpha_n + 1}$$

with β_n and α_n as in part 1. Show also that $\lim_{n\to+\infty} \delta^\pi(X^{(n)}) = \theta_0$ (almost surely).

6. We finally want to test the point null hypothesis $H_0 : \theta = 1$ against the alternative $H_1 : \theta \neq 1$. To this aim, we consider the prior density $\tilde{\pi}(\theta)$ defined by

$$\tilde{\pi}(\theta) = 0.5\,\mathbb{1}_{\{1\}}(\theta) + 0.5\,\mathbb{1}_{\{\theta\neq 1\}}(\theta)\pi(\theta), \quad \theta \in \Theta.$$

   a) Show that $B_{01}^\pi(x)$, the Bayes factor for this test, is given by

$$B_{01}^\pi(x) = \frac{\beta_n^{\alpha_n}\Gamma(\alpha_0)}{\alpha_0^{\alpha_0}\Gamma(\alpha_n)e^{\beta_n-\alpha_0}}$$

   with β_n and α_n as in part 1.

   b) Show that, as $\alpha_0 \to 0$, $B_{01}^\pi(x) \to +\infty$. What do you conclude about the use of a vague prior distribution in the context of hypothesis testing?

Problem 2

Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables taking values in $\mathcal{X}_1 \subset \mathbb{R}^k$ and such that $X_1 \sim \tilde{f}(x_1|\theta_0)$ for some $\theta_0 \in \Theta \subset \mathbb{R}^d$, where $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$ is a parametric model for a single observation. Let $\pi(\theta)$ be a strictly positive proper prior density on Θ and, for $n \geq 1$, let $\pi(\theta|X^{(n)}) \propto \pi(\theta)\prod_{k=1}^n \tilde{f}(X_k|\theta)$ be the posterior distribution based on the observation $X^{(n)} := (X_1, \dots, X_n)$. We assume below that $\pi(\theta|X^{(n)})$ is consistent.

1. Assuming that Θ is a bounded set, show that
   a) $\lim_{n\to+\infty} \mathbb{E}_\pi[\theta|X^{(n)}] = \theta_0$, $\mathbb{P}_{\theta_0}$-almost surely.
   b) $\lim_{n\to+\infty} \rho^\pi(X^{(n)}) = 0$, $\mathbb{P}_{\theta_0}$-almost surely, where $\rho^\pi(X^{(n)})$ is the posterior expected loss of $\mathbb{E}_\pi[\theta|X^{(n)}]$ under the quadratic loss function.

2. Consider the test $H_0 : \theta \in \Theta_0$ against the alternative $H_1 : \theta \in \Theta_1 := \Theta \setminus \Theta_0$, where $\Theta_0 \subsetneq \Theta$ is a non-empty open set, and let $B_{01}^\pi(X^{(n)})$ be the Bayes factor for this test. Show that, for every $C > 0$,

$$\lim_{n\to+\infty} \mathbb{P}_{\theta_0}\big(B_{01}^\pi(X^{(n)}) \geq C\big) = \begin{cases} 1, & \theta_0 \in \Theta_0 \\ 0, & \theta_0 \in \Theta_1. \end{cases}$$

Problem 3

Let $\mathcal{Y} = \{1, 2\}$ and $(Y_t)_{t\geq 0}$ be a sequence of random variables taking values in the state space $\mathcal{Y}$ and such that $\mathbb{P}(Y_0 = 1) = \lambda_1$.

1. Prove that if $\lim_{t\to+\infty} \mathbb{E}[|Y_t - Y|^2] = 0$ (for some random variable Y) then $Y_t \to Y$ in probability.

2. Assume that the Y_t's are i.i.d. and let P be the corresponding transition matrix.

   a) Write down P. For what value(s) of λ_1 is P an irreducible and aperiodic transition matrix?

   b) What is the stationary distribution of P? Is it unique?

3. Let Q_1 and Q_2 be two transition matrices on $\mathcal{Y}$ having $\mu \in \mathcal{P}(\mathcal{Y})$ as invariant distribution.

   a) Show that, for every $\alpha \in [0, 1]$, the matrix $R = \alpha Q_1 + (1-\alpha)Q_2$ is a transition matrix having μ as invariant distribution.

   b) Show that $R = Q_1 Q_2$ is a transition matrix having μ as invariant distribution.

4. Let $y = (y_0, \ldots, y_t)$ be $t + 1$ observations in $\mathcal{Y}$ that we model as a Markov (λ, P)-process and let θ be the vector that contains all the parameters of the model.

   a) What is Θ, the parameter space?

   b) Write down $f(y|\theta)$, the likelihood function of the model.

   c) Assuming that with prior probability $1/2$ the observations are i.i.d. and using the result in part 2.a), propose a prior distribution for θ.

   d) Write down the DAG for (Y, θ).

   e) Write down a Metropolis-Hastings algorithm that can be used in practice to approximate $\pi(\theta|y)$, the posterior distribution corresponding to the likelihood function defined in part 4.b) and the prior distribution specified in part 4.c). Explain why and in which sense the algorithm you propose is valid to approximate $\pi(\theta|y)$.