Bayesian Modelling – Problem Sheet 1

Please hand in your solutions for Problems 3,4 and 5 on Wednesday 06/02/2019

Problem 1

Let $x_1, \dots, x_n$ be n observations that we model as independent $\mathcal{N}_d(\mu, \Sigma)$ random variables, with $\mu \in \mathbb{R}^d$ and Σ a d-dimensional covariance matrix. Both μ and Σ are unknown so that, abusing notation, $\theta = (\mu, \Sigma)$.

We consider the prior distribution $\pi(\theta)$ for which

$$\mu|\Sigma \sim \mathcal{N}_d\big(\mu_0, \kappa_0^{-1}\Sigma\big), \quad \Sigma \sim \mathcal{W}_d^{-1}(\Psi_0, \nu_0)$$

with $\kappa_0 > 0$, $\nu_0 > d-1$, $\mu_0 \in \mathbb{R}^d$ and Ψ_0 a $(d \times d)$ positive definite matrix. In the sequel $\mathcal{W}_d^{-1}(\Psi_0, \nu_0)$ denotes the inverse Whishart distribution whose density function is given by

$$f(\Sigma|\Psi_0, \nu_0) \propto |\Sigma|^{-\frac{\nu_0+d+1}{2}} \exp\Big(-\frac{1}{2}\mathrm{tr}(\Psi_0\Sigma^{-1})\Big).$$

Let $x = (x_1, ..., x_n)$ and $\bar{x}_n = \frac{1}{n}\sum_{k=1}^n x_k$. Show that the posterior distribution $\pi(\theta|x)$ is such that

$$\mu|(\Sigma, x) \sim \mathcal{N}_d\big(\mu_n, \kappa_n^{-1}\Sigma\big), \quad \Sigma|x \sim \mathcal{W}_d^{-1}(\Psi_n, \nu_n)$$

where

$$\mu_n = \frac{\kappa_0\mu_0 + n\bar{x}_n}{\kappa_0 + n}, \quad \kappa_n = \kappa_0 + n, \quad \nu_n = \nu_0 + n$$

and

$$\Psi_n = \Psi_0 + \sum_{k=1}^n (x_k - \bar{x}_n)(x_k - \bar{x}_n)^T + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{x}_n - \mu_0)(\bar{x}_n - \mu_0)^T.$$

Problem 2 (The Bayesian linear regression model)

We consider the following linear regression model:

$$y_i = \beta^T z_i + \epsilon_i, \quad i = 1, \dots, n$$

with $y_i \in \mathbb{R}$, $z_i \in \mathbb{R}^d$ and where the ϵ_i's are i.i.d. $\mathcal{N}_1(0, \sigma^2)$ random variables.

We assign to $\theta := (\beta, \sigma^2)$ the prior distribution $\pi(\theta)$ for which

$$\beta|\sigma^2 \sim \mathcal{N}_d(\beta_0, \sigma^2\Sigma_0), \quad \sigma^2 \sim \Gamma^{-1}(a_0, b_0)$$

where the fixed hyper-parameters are as follows: $\beta_0 \in \mathbb{R}^d$, $(a_0, b_0) \in \mathbb{R}^2_{>0}$ and Σ_0 is a d-dimensional covariance matrix. In the sequel $\Gamma^{-1}(a_0, b_0)$ denotes the inverse Gamma distribution whose density is given by

$$f(\sigma^2|a_0, b_0) \propto (\sigma^2)^{-a_0-1}e^{-b_0/\sigma^2}.$$

Let $y = (y_1, \dots, y_n)$, Z be the $n \times d$ matrix having z_i^T as i-th row and $x := (y_1, z_1, \dots, y_n, z_n)$ be the vector of observations.

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

$$\beta|(\sigma^2, x) \sim \mathcal{N}_d(\beta_n, \sigma^2\Sigma_n), \quad \sigma^2|x \sim \Gamma^{-1}(a_n, b_n)$$

   where

$$\mu_n = (Z^T Z + \Sigma_0^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + Z^T y), \quad \Sigma_n = (Z^T Z + \Sigma_0^{-1})^{-1}$$

   and

$$a_n = a_0 + \frac{n}{2}, \quad b_n = b_0 + \frac{1}{2}\big(y^T y - \mu_n^T \Sigma_n^{-1}\mu_n + \beta_0^T \Sigma_0^{-1}\beta_0\big).$$

2. Let

$$\hat{\beta} = (Z^T Z)^{-1} Z^T y$$

   be the OLS estimator of β and I_d be the d-dimensional identity matrix. Show that

$$\mathbb{E}_\pi[\beta|x] = M_0\mu_0 + (I_d - M_0)\hat{\beta}$$

   where $M_0 = (Z^T Z + \Sigma_0^{-1})^{-1}\Sigma_0^{-1}$; remark that M_0 and $I_d - M_0$ are positive definite matrices.

3. Assume that $\Sigma_0 = c_0\tilde{\Sigma}_0$ where $c_0 > 0$ and $\tilde{\Sigma}_0$ is a d-dimensional covariance matrix. Show that, as $c_0 \to +\infty$, $\mathbb{E}_\pi[\beta|x] \to \hat{\beta}$ and interpret this result.

Problem 3

Let $x_1, \dots, x_n$ be n observations that we model as independent Gamma(λ, θ) random variables; that is, the likelihood of observation x_1 is given by

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

where $\lambda \in (0, +\infty)$ is a known parameter while $\theta \in \Theta := (0, +\infty)$ is an unknown parameter. We consider a Gamma(α_0, β_0) prior distribution on Θ, where $\alpha_0, \beta_0 \in (0, +\infty)$ are fixed hyper-parameters.

1. Derive the posterior distribution of θ given $x^{(n)} := (x_1, \dots, x_n)$.

2. Compute $\mathbb{E}_\pi[\theta^k|x^{(n)}]$ for any integer $k \geq 1$. Deduce the expression of the posterior mean and of the posterior variance of θ.

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

3. Analyse the behaviour of the posterior mean and variance of θ as $n \to +\infty$ assuming that the model is well-specified; that is, assuming that the observations are i.i.d. and such that $X_1 \sim$ Gamma(λ, θ_0) for some $\theta_0 \in \Theta$.

4. Using decision theory and following the Bayesian approach, propose an estimate of $e^{-\theta}$ based on the observation $x^{(n)}$. More precisely you need to (i) specify a set of possible decisions $\mathcal{D}$, (ii) specify a loss function $L : \Theta \times \mathcal{D} \to [0, +\infty)$ and (ii) compute $\gamma^{\pi}(x^{(n)})$, the estimate of $e^{-\theta}$ obtained by minimizing the resulting posterior expected loss.

5. The linear exponential (LINEX) loss function $L : \Theta \times \Theta \to [0, +\infty)$ is defined by

$$L(\theta, d) = e^{\kappa(d-\theta)} - \kappa(d - \theta) - 1, \quad (\theta, d) \in \Theta \times \Theta$$

with $\kappa > 0$ a fixed parameter. Compute $\delta^{\pi}(x^{(n)})$, the estimate of θ obtained by minimizing the resulting posterior expected loss.

Problem 4

Let $f(\cdot|\theta)$ be the probability density function of the $\mathcal{N}_1(\theta, 1)$ distribution, with $\theta \in \Theta := \mathbb{R}$. Assume that $\theta \sim \mathcal{N}_1(0, 1)$ and consider the loss function defined by

$$L(\theta, d) = e^{\frac{3\theta^2}{4}} (\theta - d)^2, \quad (\theta, d) \in \mathbb{R}^2.$$

1. Show that the estimator $\delta^{\pi} : \mathbb{R} \to \Theta$ is unique and such that $\delta^{\pi}(x) = 2x$ for all $x \in \mathcal{X} := \mathbb{R}$.

2. Show that $r(\pi) = +\infty$.

3. Show that the maximum likelihood estimator $\delta_0(x) = x$ uniformly dominates δ^{π}; i.e. show that

$$R(\theta, \delta_0) < R(\theta, \delta^{\pi}), \quad \forall \theta \in \Theta.$$

Problem 5

Let $\Theta \subseteq \mathbb{R}^d$ be a convex set, $\pi(\theta)$ be a prior distribution on Θ, $L : \Theta \times \Theta \to [0, +\infty)$ be the loss function defined by

$$L(\theta, d) = \tilde{L}(\theta - d), \quad \forall (\theta, d) \in \Theta^2$$

with $\tilde{L} : \mathbb{R}^d \to [0, +\infty)$ such that

$$\rho(\pi, d|x) < +\infty, \quad \forall d \in \Theta, \quad \forall x \in \mathcal{X}.$$

1. Assume that $\tilde{L}$ is convex on $\mathbb{R}$ and show that, for any $x \in \mathcal{X}$, the mapping $d \mapsto \rho(\pi, d|x)$ is convex.

2. Assume that $\tilde{L}$ is strictly convex on $\mathbb{R}$ and show that, for any $x \in \mathcal{X}$, the mapping $d \mapsto \rho(\pi, d|x)$ is strictly convex. Deduce that the estimator δ^{π} is unique.