Bayesian Modelling – Problem Sheet 2 (Solutions)

Problem 1 (Jacobian formula in one dimension)

We show below the following slightly more general result:

> *Let $X \sim f_X$ be a random variable taking values in some open interval $U \subset \mathbb{R}$ and having a continuous probability density function $f_X : U \to \mathbb{R}_+$, let $g : V \to U$ be a bijective and continuously differentiable mapping, with $V \subset \mathbb{R}$ an open interval, and let $Y = g^{-1}(X)$. Then, $Y \sim f_Y$ where $f_Y : V \to \mathbb{R}_+$ is defined by*
> $$f_Y(y) = \Big|\frac{\mathrm{d}}{\mathrm{d}y} g(y)\Big| f_X(g(y)), \quad y \in V.$$

To show this result let F_X and F_Y be respectively the c.d.f. of X and Y, and note that, since g is bijective, g is either strictly increasing or strictly decreasing on the interval V.

Assume first that g is strictly increasing on V. Then,

$$F_Y(a) = \mathbb{P}(Y \le a) = \mathbb{P}(g^{-1}(X) \le a) = \mathbb{P}(X \le g(a)) = F_X(g(a)), \quad \forall a \in V,$$

and therefore,

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} g(y) f_X(g(y)) = \Big|\frac{\mathrm{d}}{\mathrm{d}y} g(y)\Big| f_X(g(y)), \quad \forall y \in V.$$

Assume now that g is strictly decreasing. Then, for all $a \in V$ we have

$$\begin{aligned} F_Y(a) = \mathbb{P}(Y \le a) = \mathbb{P}(g^{-1}(X) \le a) = 1 - \mathbb{P}(g^{-1}(X) > a) &= 1 - \mathbb{P}(X < g(a)) \\ &= 1 - \mathbb{P}(X \le g(a)) \\ &= 1 - F_X(g(a)) \end{aligned}$$

where the penultimate equality holds because X is a continuous random variable. Then,

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = -\frac{\mathrm{d}}{\mathrm{d}y} g(y) f_X(g(y)) = \Big|\frac{\mathrm{d}}{\mathrm{d}y} g(y)\Big| f_X(g(y)), \quad \forall y \in V.$$

Problem 2

Since $x_1, \dots, x_n$ are modelled as independent random variables the joint log-likelihood function of $x := (x_1, \dots, x_n)$ is given by

$$\log f(x|\theta) = \sum_{k=1}^{n} \log \tilde{f}(x_k|\theta), \quad \theta \in \Theta$$

and, consequently, using the linearity of the differential operator,

$$\frac{\partial \log f(x|\theta)}{\partial \theta} \frac{\partial \log f(x|\theta)}{\partial \theta^T} = \sum_{k=1}^{n} \frac{\partial \log \tilde{f}(x_k|\theta)}{\partial \theta} \frac{\partial \log \tilde{f}(x_k|\theta)}{\partial \theta^T}.$$

Then, using the linearity of the expectation operator, we deduce that

$$I_n(\theta) = \sum_{k=1}^{n} I_1(\theta) = n I_1(\theta)$$

so that the Jeffreys prior for this model is

$$\pi_J(\theta) \propto \sqrt{\det(I_n(\theta))} = \sqrt{\det(n\, I_1(\theta))} = n^{d/2}\sqrt{\det(I_1(\theta))} \propto \sqrt{\det(I_1(\theta))}$$

which is independent of n.

Problem 3

1. The log-likelihood function is given by

$$\log f(x|\theta) = -\frac{d}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{d}(x_i - \theta_i)^2, \quad \theta \in \Theta$$

so that, for every $i, j = 1, \dots, d$,

$$\frac{\partial \log f(x|\theta)}{\partial \theta_i} = (x_i - \theta_i), \quad \frac{\partial^2 \log f(x|\theta)}{\partial \theta_i \partial \theta_j} = -\delta_{ij}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. It follows that $I(\theta) = I_d$ and thus

$$\pi_J(\theta) \propto [\det(I(\theta))]^{1/2} = 1, \quad \theta \in \Theta.$$

2. It is easily checked that $\theta|x \sim \mathcal{N}_d(x, I_d)$ so that $\mathbb{E}_\pi[\theta|x] = x$.

3. We have, using the result derived in part 2,

$$\mathbb{E}_\pi[\eta|x] = \mathbb{E}_\pi\left[\|\theta\|^2|x\right] = \sum_{i=1}^{d}\mathbb{E}_\pi[\theta_i^2|x] = \sum_{i=1}^{d}(x_i^2 + 1) = \|x\|^2 + d.$$

4. **IMPORTANT:** Henceforth we use the notation $R(\theta, \delta_c)$ for the frequentist risk of δ_c that is used throughout the course instead of the notation $R(\|\theta\|^2, \delta_c)$ introduced in the statement of the question.

Let $c \in \mathbb{R}$ and $\theta \in \Theta$. Then,

$$\begin{aligned}
R(\theta, \delta_c) &= \int_{\mathbb{R}^d} L(\theta, \delta_c(x)) f(x|\theta) \mathrm{d}x \\
&= \int_{\mathbb{R}^d} (\|\theta\| - \delta_c(x))^2 f(x|\theta) \mathrm{d}x \\
&= \int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2 - c)^2 f(x|\theta) \mathrm{d}x \\
&= \int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2)^2 f(x|\theta) \mathrm{d}x + c^2 - 2c \int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2) f(x|\theta) \mathrm{d}x.
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial R(\theta, \delta_c)}{\partial c} &= 2c - 2 \int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2) f(x|\theta) \mathrm{d}x \\
\frac{\partial^2 R(\theta, \delta_c)}{\partial c^2} &= 2 > 0
\end{aligned}$$

so that the unique minimizer c_θ^* of the mapping $c \mapsto R(\theta, \delta_c)$ is given by

$$c_\theta^* = \int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2) f(x|\theta) \mathrm{d}x = \|\theta\|^2 - \left(\|\theta\|^2 + d\right) = -d$$

and is thus independent of θ.

Lastly, because for every $\theta \in \Theta$ the minimizer of $c \mapsto R(\theta, \delta_c)$ is unique and given by $c_\theta^* = -d$, for every $c \neq -d$ we have

$$R(\theta, \delta_{-d}) < R(\theta, \delta_c), \quad \forall \theta \in \Theta$$

and, in particular,

$$R(\theta, \delta_{-d}) < R(\theta, \delta_d), \quad \forall \theta \in \Theta \tag{1}$$

showing that the estimator $\mathbb{E}_\pi[\eta|\cdot]$ is uniformly dominated by the estimator δ_{-d}.

5. Let $c \in \mathbb{R}$ and $\theta \in \Theta$. Then, using the computations done in part 4,

$$\begin{aligned}
R(\theta, \delta_c) &= \int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2 - c)^2 f(x|\theta) \mathrm{d}x \\
&= \int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2)^2 f(x|\theta) \mathrm{d}x + c^2 + 2cd
\end{aligned}$$

where

$$\begin{aligned}
\int_{\mathbb{R}^d} (\|\theta\|^2 - \|x\|^2)^2 f(x|\theta) \mathrm{d}x &= \|\theta\|^4 + \int_{\mathbb{R}^d} \|x\|^4 f(x|\theta) \mathrm{d}x - 2\|\theta\|^2 (\|\theta\|^2 + d) \\
&= \int_{\mathbb{R}^d} \|x\|^4 f(x|\theta) \mathrm{d}x - \|\theta\|^4 - 2\|\theta\|^2 d.
\end{aligned}$$

In addition, using the fact that $\|x\|^2$ is distributed according to a noncentral χ^2 distribution with d degrees of freedom and non-centrality parameter $\|\theta\|^2$, we have

$$\int_{\mathbb{R}^d} \|x\|^4 f(x|\theta)\mathrm{d}x = 2(d + 2\|\theta\|^2) + (d + \|\theta\|^2)^2$$

so that,

$$R(\theta, \delta_c) = 4\|\theta\|^2 + c^2 + 2d(c+1) + d^2.$$

Hence,

$$\sup_{\theta\in\Theta} R(\theta, \delta_c) = +\infty, \quad \forall c \in \mathbb{R}.$$

Problem 4

1. Let $\bar{x}_n = n^{-1}\sum_{k=1}^n x_k$. Then, the joint likelihood function is given by

   $$f(x|\theta) = \prod_{k=1}^n \tilde{f}(x_k|\theta) = \frac{\theta^{n\bar{x}_n}}{x_1! \cdots x_n!} e^{-n\theta}, \quad \theta \in \Theta$$

   and thus

   $$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \theta^{\alpha_0 + n\bar{x}_n - 1} e^{-\theta(\beta_0+n)}.$$

   We therefore deduce that

   $$\theta|x \sim \text{Gamma}(\alpha_n, \beta_n)$$

   with $\alpha_n = \alpha_0 + n\bar{x}_n$ and $\beta_n = \beta_0 + n$.

   The prior is conjugate because the prior distribution and the posterior distribution belong to the same parametric family of distributions.

2. Using the result of Problem Sheet 1, Problem 3, part 2, we have

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

3. Let $g : \Theta \to [0, +\infty)$ be defined by

   $$g(\theta) = L(\theta, \mathbb{E}_\pi[\theta|x]), \quad \theta \in \Theta.$$

   Then, we are asked to compute the estimator $\delta^\pi(x)$ of $g(\theta)$ obtained by minimizing the posterior expected loss associated to a quadratic loss function.

   For this problem the set of possible decisions is $\mathcal{D} = g(\Theta)$ and the loss function $L : \Theta \times \mathcal{D} \to [0, +\infty)$ we are considering is defined by

   $$L(\theta, d) = \big(g(\theta) - d\big)^2, \quad (\theta, d) \in \Theta \times \mathcal{D}.$$

   Then, as seen during the lectures,

   $$\delta^\pi(x) = \mathbb{E}_\pi\big[g(\theta)|x\big] = \mathbb{E}_\pi\big[L(\theta, \mathbb{E}_\pi[\theta|x])|x\big] = \mathbb{E}_\pi\big[(\theta - \mathbb{E}_\pi[\theta|x])^2|x\big] = \text{Var}_\pi(\theta|x)$$

   with $\text{Var}_\pi(\theta|x)$ given in part 2.

4. a) Using the result of part 1, it follows that

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto f(x|\theta) \propto \theta^{\alpha_n'-1}e^{-\beta_n'\theta}$$

with $\alpha_n' = 1 + n\bar{x}_n$ and $\beta_n' = n$. Therefore, under the Laplace's prior,

$$\theta|x \sim \mathrm{Gamma}(\alpha_n', \beta_n'),\ \mathbb{E}_\pi[\theta|x] = \frac{\alpha_n'}{\beta_n'} = \frac{1}{n} + \bar{x}_n,\ \mathrm{Var}_\pi(\theta|x) = \frac{\alpha_n'}{(\beta')_n^2} \qquad (2)$$

so that the improper Laplace prior density yields a well-defined posterior distribution. This latter can be obtained as the limit, as $\beta_0 \to 0$, of the posterior distribution derived in part 1 under the conjugate prior density (with $\alpha_0 = 1$). Note that when $\beta_0 \to 0$ the prior variance converges to $+\infty$.

b) Let $\theta > 0$ and $\mathcal{X} = \big(\{0\} \cup \mathbb{N}\big)^n$. Then using (2) (and with $\mathrm{d}x$ the counting measure),

$$\begin{aligned}
R(\theta, \delta^\pi) &= \int_{\mathcal{X}} L(\theta, \delta^\pi(x)) f(x|\theta)\mathrm{d}x \\
&= \int_{\mathcal{X}} (\theta - \mathbb{E}_\pi[\theta|x])^2 f(x|\theta)\mathrm{d}x \\
&= \int_{\mathcal{X}} (\theta - n^{-1} - \bar{x}_n)^2 f(x|\theta)\mathrm{d}x \\
&= e^{-n\theta}\sum_{k=0}^{\infty}(\theta - n^{-1} - k/n)^2 \frac{(n\theta)^k}{k!}
\end{aligned}$$

where the last equality uses the hint.

Hence, using the fact that $\mathbb{E}[Y] = \mathrm{Var}(Y) = \lambda$ if $Y \sim \mathrm{Poisson}(\lambda)$, $\lambda > 0$, we have

$$\begin{aligned}
R(\theta, \delta^\pi) &= (\theta - n^{-1})^2 + n^{-2}\Big(e^{-n\theta}\sum_{k=0}^{\infty} k^2 \frac{(n\theta)^k}{k!}\Big) - 2n^{-1}(\theta - n^{-1})\Big(e^{-n\theta}\sum_{k=0}^{\infty} k \frac{(n\theta)^k}{k!}\Big) \\
&= \theta^2 + n^{-2} - 2\theta n^{-1} + n^{-2}(n^2\theta^2 + n\theta) - 2n^{-1}\theta(n\theta) + 2n^{-2}(n\theta) \\
&= n^{-2} + \theta n^{-1}
\end{aligned}$$

and therefore,

$$r(\pi) = \int_0^\infty R(\theta, \delta^\pi)\pi(\theta)\mathrm{d}\theta \propto \int_0^\infty \big(n^{-2} + \theta n^{-1}\big)\mathrm{d}\theta = +\infty.$$

5. a) Since

$$\log \tilde{f}(x_1|\theta) = \log(x_1!) + x_1 \log(\theta) - \theta$$

we get

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

Consequently, the Fisher information matrix for one observation is given by

$$I_1(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \log \tilde{f}(X_1|\theta)}{\partial\theta^2}\right] = \frac{\mathbb{E}_\theta[X_1]}{\theta^2} = \frac{1}{\theta}$$

and thus (see Problem 2 for the equality)

$$\pi_J(\theta) \propto \sqrt{I_n(\theta)} = \sqrt{n\, I_1(\theta)} \propto \frac{1}{\theta^{1/2}}.$$

Finally, because

$$\int_0^\infty \theta^{-1/2}\mathrm{d}\theta = 2\theta^{1/2}\Big|_0^\infty = +\infty$$

we conclude that the Jeffreys prior is improper.

b) Using similar computations as in part 1 we easily check that, under the Jeffreys prior,

$$\theta|x \sim \mathrm{Gamma}(\tilde{\alpha}_n, \tilde{\beta}_n)$$

with $\tilde{\alpha}_n = n\bar{x}_n + 0.5 > 0$ and $\tilde{\beta}_n = n > 0$. Consequently, the Jeffreys prior yields a valid posterior distribution in this model.

Problem 5

1. Let $g : (0, +\infty) \to \Theta$ be defined by

$$g(y) = \frac{y}{1+y}, \quad y \in (0, +\infty).$$

Then, the prior density $\pi^*(\eta)$ of $\eta = g^{-1}(\theta)$ implied by $\pi(\theta)$ can be computed using the change-of-variable rule (see Problem 1)

$$\pi^*(\eta) = \pi(g(\eta))\left|\frac{\mathrm{d}g}{\mathrm{d}\eta}(\eta)\right| = \frac{1}{(1+\eta)^2}, \quad \eta \in (0, +\infty).$$

Since

$$\int_0^\infty \pi^*(\eta)\mathrm{d}\eta = \frac{1}{1+\eta}\Big|_\infty^0 = 1$$

the prior density $\pi^*(\eta)$ is a proper density function on $(0, +\infty)$. Note that this result is not surprising because $\pi(\theta)$ is a well-defined probability distribution on Θ and thus $\eta = g^{-1}(\theta)$ is a well-defined random variable.

2. The required probability is

$$\pi^*(\{\eta \le a\}) = \int_0^a \pi^*(\eta)\mathrm{d}\eta = \frac{1}{1+\eta}\Big|_a^0 = \frac{a}{1+a}, \quad a \ge 0.$$

When $a = 3$, the probability is $3/4$.

3. The prior density $\pi^*(\eta)$ is not non-informative in Laplace's sense, as it positively discriminates the values from $(0, 3]$ against the values $(3, +\infty)$. Note that this result is not surprising in light of part 1: since η is a well-defined random variable on $(0, +\infty)$ its density cannot be non-informative in Laplace's sense.

Problem 6

1. We have

$$\log \tilde{f}(x_1|\theta) = \log\theta + (x_1 - 1)\log(1-\theta), \quad \theta \in \Theta$$

and therefore

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

Consequently (using the fact that $\mathbb{E}_\theta[X_1] = \theta^{-1}$)

$$I_1(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \log \tilde{f}(X_1|\theta)}{\partial\theta^2}\right] = \frac{1}{\theta^2} + \frac{\mathbb{E}_\theta[X_1] - 1}{(1-\theta)^2} = \frac{1}{\theta^2} + \frac{\theta^{-1} - 1}{(1-\theta)^2} = \frac{1}{\theta^2(1-\theta)}$$

so that the Jeffreys prior is defined by (see Problem 2 for the equality)

$$\pi_J(\theta) \propto \sqrt{I_n(\theta)} = \sqrt{nI_1(\theta)} \propto \theta^{-1}(1-\theta)^{-1/2}, \quad \theta \in \Theta.$$

To show that the Jeffreys prior is improper, let $t = (1-\theta)^{1/2}$ so that

$$\begin{aligned}\int_0^1 \frac{\mathrm{d}\theta}{\theta(1-\theta)^{1/2}} = \int_0^1 \frac{2}{1-t^2}\mathrm{d}t = \int_0^1 \left(\frac{1}{1+t} + \frac{1}{1-t}\right)\mathrm{d}t \\ &= \big(\log(1+t) - \log(1-t)\big)\Big|_0^1 \\ &= +\infty\end{aligned}$$

and therefore the Jeffreys prior is improper.

Remark: To show that the Jeffreys prior is improper it would have been enough to notice that $\pi_J(\theta)$ corresponds to the degenerated Beta distribution $\text{Beta}(1/2, 0)$ (recall that the $\text{Beta}(\alpha, \beta)$ distribution is defined for $\alpha, \beta > 0$).

2. Since $x_1, \ldots, x_n$ are modelled as independent random variables the likelihood function of $x := (x_1, \ldots, x_n)$ is given by

$$f(x|\theta) = \prod_{k=1}^{n} \tilde{f}(x_k|\theta) = \theta^n(1-\theta)^{n(\bar{x}_n - 1)}, \quad \theta \in \Theta$$

where $\bar{x}_n = \frac{1}{n}\sum_{k=1}^n x_k$.
Consequently,

$$\pi(\theta|x) \propto f(x|\theta)\pi_J(\theta) \propto \theta^{\alpha_n - 1}(1-\theta)^{\beta_n - 1}$$

where $\alpha_n = n$ and $\beta_n = n(\bar{x}_n - 1) + 1/2$. Since $\alpha_n > 0$, $\beta_n > 0$, we deduce that

$$\theta|x \sim \text{Beta}(\alpha_n, \beta_n)$$

and thus $\pi_J(\theta)$ yields a proper posterior density function.