Bayesian Modelling – Problem Sheet 4

Please hand in your solutions for Problems 3-6 by 6pm on Wednesday 20/03/2019

Problem 1

In this problem we prove a result which allows to control the denominator of the posterior distribution. This result plays an important role in Bayesian asymptotic theory.

Let $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$ be a family of p.d.f. on $\mathcal{X}_1$, $\delta > 0$ and $C > 0$ be some constants, and let $\pi(\theta)$ be some p.d.f. on Θ. Let

$$B_\delta = \left\{\theta \in \Theta : \, KL(\theta_0|\theta) \leq \delta^2, \mathbb{E}_{\theta_0}\Big[\Big(\log \frac{\tilde{f}(X_1|\theta)}{\tilde{f}(X_1|\theta_0)}\Big)^2\Big] \leq \delta^2\right\} \tag{1}$$

and, for $n \in \mathbb{N}$, let

$$D_{n,\delta,C} = \left\{x^{(n)} \in \mathcal{X}_1^n : \, \int_\Theta \prod_{i=1}^n \frac{\tilde{f}(x_i|\theta)}{\tilde{f}(x_i|\theta_0)} \pi(\theta)\mathrm{d}\theta \leq \pi(B_\delta) e^{-(1+C)n\delta^2}\right\}. \tag{2}$$

The goal of this problem is to show that

$$\mathbb{P}_{\theta_0}(D_{n,\delta,C}) \leq \frac{1}{C^2 n\delta^2}, \quad \forall n \geq 1. \tag{3}$$

1. Assume first that $\pi(B_\delta) = 1$.
   a) Show that, for all $n \geq 1$,

   $$\mathbb{P}_{\theta_0}(D_{n,\delta,C}) \\ \leq \mathbb{P}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^n \int_{B_\delta} \log \frac{\tilde{f}(X_i|\theta)}{\tilde{f}(X_i|\theta_0)} \pi(\theta)\mathrm{d}\theta - \mathbb{E}_{\theta_0}\left[\int_{B_\delta} \log \frac{\tilde{f}(X_1|\theta)}{\tilde{f}(X_1|\theta_0)} \pi(\theta)\mathrm{d}\theta\right]\right| \geq C\delta^2\right).$$

   Hint: Use Fubini's theorem and Jensen's inequality.

   b) Show that, for all $n \geq 1$,

   $$\mathbb{E}_{\theta_0}\left[\left(\int_{B_\delta} \log \frac{\tilde{f}(X_1|\theta)}{\tilde{f}(X_1|\theta_0)} \pi(\theta)\mathrm{d}\theta\right)^2\right] \leq \delta^2.$$

   Hint: Use Fubini's theorem and Jensen's inequality.

   c) Using the results in parts 1.a) and 1.b), show that (3) holds.

2. Using the results in part 1, show that (3) also holds when $\pi(B_\delta) < 1$.

Problem 2

Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables such that $X_1 \sim \tilde{f}(x_1|\theta_0)$ for some $\theta_0 \in \Theta \subset \mathbb{R}^d$ and where $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$ is a parametric model for a single observation. Let $\pi(\theta)$ be the prior distribution for 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 $X^{(n)} := (X_1, \dots, X_n)$.

In this question we assume that the following conditions hold:

(C_1) $\pi(\theta)$ is continuous and strictly positive on the set $(\theta_0 - \delta_\pi, \theta_0 + \delta_\pi)$ for some $\delta_\pi > 0$.

(C_2) For every sequence $M_n \to +\infty$ there exists a sequence of tests $(\phi_n)_{n\geq 1}$ such that $\mathbb{E}_{\theta_0}[\phi_n(X^{(n)})] \to 0$ and such that, for some constants $\epsilon > 0$ and $D > 0$, and for n large enough,

$$\mathbb{E}_\theta[1 - \phi_n(X^{(n)})] \leq e^{-D(\|\theta-\theta_0\|^2 \wedge \epsilon^2)}$$

for all θ such that $\|\theta - \theta_0\| \geq M_n/\sqrt{n}$.

(C_3) There exist constants $\delta > 0$ and $c_\star > 0$ such that, for all θ such that $\|\theta - \theta_\star\| \leq \delta$,

$$KL(\theta_0|\theta) \leq c_\star^2\|\theta - \theta_0\|^2, \quad \mathbb{E}_{\theta_0}\Big[\Big(\log\frac{\tilde{f}(X_1|\theta)}{\tilde{f}(X_1|\theta_0)}\Big)^2\Big] \leq c_\star^2\|\theta - \theta_0\|^2.$$

Let $(M_n)_{n\geq 1}$ be an arbitrary sequence in $\mathbb{R}_{>0}$ such that $\lim_{n\to+\infty} M_n = +\infty$ and $\lim_{n\to+\infty} M_n n^{-1/2} = 0$. The goal of this problem is to show that, under (C_1)-(C_3),

$$\pi\big(\{\theta : \|\theta - \theta_0\| \geq M_n\, n^{-1/2}\}\,\big|X^{(n)}\big) \to 0, \quad \text{in } \mathbb{P}_{\theta_0}\text{-probability}, \tag{4}$$

i.e. that $\pi(\theta|X^{(n)})$ converges to θ_0 at rate $n^{-1/2}$.

1. We first show that

$$\limsup_{n\to+\infty} \mathbb{E}_{\theta_0}\big[\pi\big(\{\theta :\, \epsilon_n \leq \|\theta - \theta_0\| < \epsilon\}|X^{(n)}\big)\big] = 0 \tag{5}$$

where $\epsilon > 0$ is as in (C_2) and where $\epsilon_n = M_n n^{-1/2}$.

Let $M = \sqrt{D/4}$, with D as in (C_2), and $\delta_n = M\epsilon_n$.

a) Using the result of Problem 1 show that

$$\limsup_{n\to+\infty} \mathbb{E}_{\theta_0}\big[\pi\big(\{\theta :\, \epsilon_n \leq \|\theta - \theta_0\| < \epsilon\}|X^{(n)}\big)\big]$$
$$\leq \limsup_{n\to+\infty} \mathbb{E}_{\theta_0}\big[\mathbb{1}_{D^c_{n,\delta_n,1}}(X^{(n)})(1 - \phi_n(X^{(n)}))\pi\big(\{\theta :\, \epsilon_n \leq \|\theta - \theta_0\| < \epsilon\}|X^{(n)}\big)\big]$$

where for every $\delta > 0$ the set $D_{n,\delta,1}$ is defined in (2) (with $C = 1$).

b) Let J be the smallest integer such that $(J+1)\epsilon_n > \epsilon$ and let

$$\Theta_{n,j} = \Big\{\theta:\, j\epsilon_n \le |\theta-\theta_0| \le \min\big((j+1)\epsilon_n, \epsilon\big)\Big\}, \quad j=1,\dots,J.$$

Show that, for n large enough,

$$\mathbb{E}_{\theta_0}\big[\mathbb{1}_{D^c_{n,\delta_n,1}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi\big(\Theta_{n,j}|X^{(n)}\big)\big] \le e^{2M^2M_n^2 - Dj^2M_n^2}\frac{\pi(\Theta_{n,j})}{\pi(B_{\delta_n})}, \quad \forall j\in\{1,\dots,J\}$$

where for every $\delta>0$ the set B_δ is defined in (1).

c) Show that, under (C_1) and (C_3), there exists a constant $\underline{L}>0$ such that, for n large enough, $\pi(B_{\delta_n}) \ge \underline{L}\,\delta_n^d$.

d) Show that, under (C_1), there exists a constant $\bar{L}>0$ such that, for n large enough, $\pi(\Theta_{n,j}) \le \bar{L}\epsilon_n^d$ for all $j\in\{1,\dots,J\}$.

Hint: Remark that, in (C2), $\epsilon>0$ can be taken arbitrarily small.

e) Using the results in parts 1.b)-1.d), and recalling that $M=\sqrt{D/4}$, show that there exists a constant $\bar{C}>0$ such that, for n large enough,

$$\mathbb{E}_{\theta_0}\big[\mathbb{1}_{D^c_{n,\delta_n,1}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi\big(\{\theta:\, \epsilon_n \le \|\theta-\theta_0\| < \epsilon\}|X^{(n)}\big)\big] \le \bar{C}\,e^{-\frac{1}{2}DM_n^2}$$

and deduce that (5) holds.

2. We now show that

$$\limsup_{n\to+\infty}\mathbb{E}_{\theta_0}\big[\pi\big(\{\theta:\, \|\theta-\theta_0\|\ge\epsilon\}|X^{(n)}\big)\big] = 0 \tag{6}$$

where we recall that $\epsilon>0$ is as in (C_2).

Let $\tilde{M}=\sqrt{d/2}$ and $\gamma_n = \tilde{M}\sqrt{\log(n)/n}$.

a) Show that, for n large enough,

$$\mathbb{E}_{\theta_0}\big[\mathbb{1}_{D^c_{n,\gamma_n,1}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi\big(\{\theta:\, \|\theta-\theta_0\|\ge\epsilon\}|X^{(n)}\big)\big] \le \frac{e^{-\log(n)2\tilde{M}^2+\frac{d}{2}\log(n)}}{\underline{L}\tilde{M}^d}(\log(n))^{-d/2}$$

with $\underline{L}>0$ is as in part 1.c). Recall that for every $\delta>0$ the set $D_{n,\delta,1}$ is defined in (2) (with $C=1$).

Hint: Use similar computations as in part 1.b).

b) Using the result in part 2.a) and similar computations as in part 1.a), show (6).

3. Using (5) and (6), show (4).

Problem 3

- (**Markov's inequality**) Let X be a real valued random variable, $p \geq 1$ be such that $\mathbb{E}[|X|^p] < +\infty$ and $\epsilon > 0$. Show that

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{\mathbb{E}[|X|^p]}{\epsilon^p}.$$

- (**Law of large numbers**) Let $(Z_k)_{k\geq 1}$ be a sequence of i.i.d. real-valued random variables such that $\mathbb{E}[Z_1^4] < +\infty$.

  1. Assume first that $\mathbb{E}[Z_1] = 0$.

     a) Show that $\mathbb{E}\big[\big(\sum_{k=1}^n Z_k\big)^4\big] = n\mathbb{E}[Z_1^4] + 3n(n-1)\mathbb{E}[Z_1^2]^2$.

     b) Using the result of part 1.a), show that

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n Z_k = 0, \quad \text{almost surely.}$$

  2. Using the result of part 1.b), show that

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n Z_k = \mathbb{E}[Z_1], \quad \text{almost surely.} \tag{7}$$

  Remark: Recall that for (7) to hold it is enough that $\mathbb{E}[|Z_1|] < +\infty$. The assumption $\mathbb{E}[Z_1^4] < +\infty$ is only made to (greatly) simplify the proof.

Problem 4

Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables such that $X_1 \sim \tilde{f}(x_1|\theta_0)$ for some $\theta_0 \in \Theta := \mathbb{R}^d$ and where $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$ is a parametric model for a single observation. Let $\pi(\theta)$ be the prior distribution for 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 $X^{(n)} := (X_1, \dots, X_n)$.

The goal of this problem is to prove a weaker version of Schwartz's theorem, establishing that

$$\pi\big(\{\theta : \|\theta - \theta_0\| \geq \epsilon\}\big|X^{(n)}\big) \to 0 \quad \text{in } \mathbb{P}_{\theta_0} - \text{probability}, \quad \forall \epsilon > 0. \tag{8}$$

To this aim, we assume that Condition (A2) given in Theorem 6.1 of the lecture notes and Condition (C3) of Problem 2 hold.

1. Using the result of Problem 1 show that, for every $\delta > 0$ and $\epsilon > 0$,

$$\begin{aligned}
&\limsup_{n\to+\infty} \mathbb{E}_{\theta_0}\big[\pi\big(\{\theta:\ \|\theta-\theta_0\| \geq \epsilon\}|X^{(n)}\big)\big] \\
&\leq \limsup_{n\to+\infty} \mathbb{E}_{\theta_0}\big[\mathbb{1}_{D^c_{n,\delta,1}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi\big(\{\theta:\ \|\theta-\theta_0\| \geq \epsilon\}|X^{(n)}\big)\big]
\end{aligned}$$

   where the set $D_{n,\delta,1}$ is defined in (2) (with $C = 1$) and where ϕ_n is as in Condition (A2) of Theorem 6.1.

2. Show that, for every $\delta > 0$, $\epsilon > 0$ and $n \geq 1$, we have

$$\mathbb{E}_{\theta_0}\big[\mathbb{1}_{D^c_{n,\delta,1}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi\big(\{\theta:\, \|\theta-\theta_0\| \geq \epsilon\}|X^{(n)}\big)\big] \leq \frac{e^{2n\delta^2}}{\pi(B_\delta)}e^{-nD_2}$$

with $D_2 > 0$ as in Condition (A2) of Theorem 6.1 and where the set B_δ is defined in (1).

3. Remark that under Condition (C3) of Problem 2, for very $\gamma > 0$ sufficiently small we have

$$\{\theta:\, \|\theta-\theta_0\| < \gamma/c_\star\} \subset B_\gamma$$

with $c_\star$ as in (C3). Use this result to provide a simple sufficient condition on the prior distribution $\pi(\theta)$ which ensures that, for every $\delta > 0$, there exists a constant $c_\delta > 0$ such that $\pi(B_\delta) \geq c_\delta$.

4. Assuming that $\pi(\theta)$ satisfies the condition of part 3, and using the results in parts 1-2, show that

$$\limsup_{n\to+\infty} \mathbb{E}_{\theta_0}\big[\pi\big(\{\theta:\, \|\theta-\theta_0\| \geq \epsilon\}|X^{(n)}\big)\big] = 0, \quad \forall\epsilon > 0.$$

5. Using the result in part 4, show (8).

Problem 5

Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. $\mathcal{N}_1(\theta_0, 1)$ random variables, with $\theta_0 \in \Theta := \mathbb{R}$, and for every $\theta \in \Theta$ let $\tilde{f}(\cdot|\theta)$ be the p.d.f. of the $\mathcal{N}_1(\theta, 1)$ distribution, with $\theta \in \Theta$.

In this problem we show that $(X_k)_{k\geq 1}$ and $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$ verify Condition (A2) given in Theorem 6.1 of the lecture notes as well as Conditions (C2) and (C3) of Problem 2, and thus that the posterior distribution $\pi(\theta|X^{(n)})$ is both consistent (in the sense of Definition 6.1) and converges to θ_0 at rate $n^{-1/2}$ (provided that $\pi(\theta)$ has positive mass around θ_0).

1. We first show that Condition (A2) of Theorem 6.1 holds.
   a) Show that,
   $$\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}\mathrm{d}t \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}, \quad \forall x > 0.$$
   b) Let $\epsilon > 0$, $n \in \mathbb{N}$ and $\psi_n(X^{(n)}) = \mathbb{1}(|\bar{X}_n - \theta_0| \geq \epsilon/2)$ where $\bar{X}_n = n^{-1}\sum_{k=1}^n X_k$. Using the result in part 1.a) show that
   $$\mathbb{E}_{\theta_0}[\psi_n(X^{(n)})] \leq \frac{4\,e^{-n\epsilon^2/8}}{\sqrt{n}\epsilon\sqrt{2\pi}}, \quad \mathbb{E}_\theta[1-\psi_n(X^{(n)})] \leq \frac{4\,e^{-n|\theta-\theta_0|^2/8}}{\sqrt{n}|\theta-\theta_0|\sqrt{2\pi}}$$
   for all θ such that $|\theta - \theta_0| \geq \epsilon$.
   Hint: For this question it may be useful to recall that $|x-y| \geq |x-z| - |z-y|$ for any real numbers x, y, z

c) Using the results in part 1.b), show that Condition (A2) of Theorem 6.1 holds.

2. Using the results in part 1.b), show that Condition (C2) of Problem 2 holds.

3. We now show that Condition (C3) of Problem 2 holds.

   a) Show that, for any $(\theta, \tilde{\theta}) \in \mathbb{R}^2$, we have

   $$KL(\tilde{\theta}|\theta) = \frac{(\theta - \tilde{\theta})^2}{2}.$$

   b) Show that

   $$\mathbb{E}_{\tilde{\theta}}\left[\left(\log \frac{\tilde{f}(X_1|\tilde{\theta})}{\tilde{f}(X_1|\theta)}\right)^2\right] = \frac{(\theta - \tilde{\theta})^2}{4}\Big((\theta + \tilde{\theta})^2 + 4(1 + \tilde{\theta}^2) - 4\tilde{\theta}(\theta + \tilde{\theta})\Big).$$

   c) Using the results in part 3.a)-3.b), show that Condition (C3) of Problem 2 holds.

Problem 6

Let $X_1, \dots, X_n$ be n random variables that we model as independent $\mathcal{N}_1(\theta, 1)$ random variables, with $\theta \in \Theta := \mathbb{R}$. We assign to θ the $\mathcal{N}_1(\mu_0, \sigma_0^2)$ distribution as prior distribution, so that the posterior distribution $\pi(\theta|X^{(n)})$ is given by

$$\pi(\theta|X^{(n)}) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\Big(-\frac{(\theta - \mu_n)^2}{2\sigma_n^2}\Big), \quad \theta \in \Theta$$

where

$$\mu_n = \frac{1}{1 + n\sigma_0^2}\mu_0 + \frac{\sigma_0^2}{1 + n\sigma_0^2}\sum_{k=1}^{n} X_k, \qquad \sigma_n^2 = \frac{\sigma_0^2}{1 + n\sigma_0^2}.$$

The maximum likelihood estimator of θ is given by $\hat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$.

1. Give the Highest Posterior Density (HPD) region at level $(1 - \alpha)$, with $\alpha \in (0, 1)$.

2. We first assume that the model is well-specified, i.e. that $X_1, \dots, X_n$ are independent and identically distributed random variables such that $X_1 \sim \mathcal{N}_1(\theta_0, 1)$ for some $\theta_0 \in \Theta$.

   a) Let $S_n = \sqrt{n}(\theta - \hat{\theta}_n)$ and $\pi^*(s|X^{(n)})$ be the probability density function of the posterior distribution of S_n given $X^{(n)}$.

   Show that, for every $s \in \mathbb{R}$,

   $$\lim_{n\to+\infty} \pi^*(s|X^{(n)}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \quad \mathbb{P}_{\theta_0} - \text{almost surely.} \tag{9}$$

   Remark that, by Scheffé's lemma, (9) implies that, as $n \to +\infty$, we have $\sqrt{n}(\theta - \hat{\theta}_n)|X^{(n)} \overset{\text{dist.}}{\Longrightarrow} \mathcal{N}_1(0, 1)$, $\mathbb{P}_{\theta_0}$-almost surely.

b) Give the confidence interval at level $1-\alpha$ centred at $\hat{\theta}_n$.

c) Use these results to compare, as $n \to +\infty$, the credible interval at level $1-\alpha$ computed in part 1 and the confidence interval at level $1-\alpha$ computed in part 2.b).

3. We now assume that the model is misspecified. More precisely, we assume that $X_1, \dots, X_n$ are independent and identically distributed random variables such that $X_1 \sim t_6(\theta_0, 1)$ for a $\theta_0 \in \Theta$. Here, $t_6(\theta_0, 1)$ denotes the Student-t distribution with location parameter θ_0, scale parameter 1 and 6 degrees of freedom. Note that $\mathbb{E}_0[X_1] = \theta_0$ while $\mathrm{Var}_0(X_1) = 1.5$ where, as in the lecture notes, the subscript 0 refers to the true distribution of the observations

   a) Using the result in part 2.a) deduce that, as $n \to +\infty$, we have $\sqrt{n}(\theta - \hat{\theta}_n)|X^{(n)} \xrightarrow{\text{dist.}} \mathcal{N}_1(0,1)$, $\mathbb{P}_0$-almost surely.

   b) Show that, as $n \to +\infty$, we have $\sqrt{n}\big(\hat{\theta}_n - \theta_0\big) \xrightarrow{\text{dist.}} \mathcal{N}_1(0, 1.5)$ and give the confidence interval at level $1-\alpha$ centred at $\hat{\theta}_n$.

   c) Use these results to compare, as $n \to +\infty$, the credible interval at level $1-\alpha$ computed in part 1. and the confidence interval at level $1-\alpha$ computed in part 3.c).

4. What do you conclude about the frequentist properties of credible intervals?