VI - Bayesian asymptotics

In the frequentist perspective, most statistical procedures (estimators, tests, etc.) are evaluated through their *asymptotic properties.* This is because it is generally difficult to evaluate the frequentist properties of the considered procedure for a given sample size n.

In addition, as mentioned in Chapter 2, the frequentist risk does not allow to derive an optimal estimator from a decision theoretic point of view and thus asymptotic results are often the main justification for using a particular estimator.

Example: *Under mild conditions, the asymptotic properties of the maximum likelihood estimator (MLE)* $\hat{\theta}_n$ *are the following:*

1. *The MLE is convergent*

$$\hat{\theta}_n \to \theta_0 \text{ in probability}$$

   where θ_0 *is the 'true' parameter.*

2. *The MLE is asymptotically Normal,*

$$n^{1/2}(\hat{\theta}_n - \theta_0) \overset{\text{dist.}}{\Rightarrow} \mathcal{N}_d(0, I_1(\theta_0)^{-1})$$

   as the number n *of observations goes towards infinity.*

The first result indicates that $\hat{\theta}_n$ *gets closer and closer to* θ_0 *as the sample size increases. The second result gives a rough evaluation of the estimation error as* $\sqrt{n}I_1(\theta_0)^{-1/2}$*.*

Asymptotics for Bayesian methods

In the Bayesian framework asymptotic properties are less an issue:

- From a pure Bayesian perspective, all the inference is done conditionally to the observations $x_1, \ldots, x_n$ and therefore asymptotic results are not relevant to choose an estimator.
- Bayesian estimators are justified from a decision theoretic point of view (see Chapter 2).
- The uncertainty on the value of θ is already expressed through the spread of the posterior distribution. In particular, the posterior variance (for instance) is the Bayesian estimator of the quadratic loss incurred by the use of the posterior mean (see Problem Sheet 2).
- In connection with the previous point, it is always possible to derive credible regions that are exactly of level $1-\alpha$ in a Bayesian framework without resorting to asymptotic approximations (in contrast with frequentist confidence regions).

Nonetheless,

- A Bayes estimator which, for instance, would not be convergent would have little appeal.
- The asymptotic properties of Bayesian estimators explain why Bayesian methods are also appealing from a frequentist point of view.

The goal of this chapter is to present the most classical results on Bayesian asymptotics.

Set-up and notation

- We assume in this chapter that $X_1, \ldots, X_n$ are i.i.d. from the distribution on $\mathcal{X}_1$ having density $\tilde{f}(\cdot|\theta_0)$ for some $\theta_0 \in \Theta \subseteq \mathbb{R}^d$. In other words, we assume that the statistical model $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$ is well-specified.
- In this chapter we use the shorthand $X^{(n)} = (X_1, \ldots, X_n)$ so that the posterior distribution can be written as

$$\pi(\theta|X^{(n)}) = \frac{f(X^{(n)}|\theta)\pi(\theta)}{\int_\Theta f(X^{(n)}|\theta)\pi(\theta)\mathrm{d}\theta}, \quad f(X^{(n)}|\theta) = \prod_{i=1}^n \tilde{f}(X_i|\theta).$$

- As in Chapter 5, we denote by $l_n(\theta)$ the log-likelihood function; that is,

$$l_n(\theta) = \sum_{i=1}^n \log \tilde{f}(X_i|\theta), \quad \theta \in \Theta.$$

- As in Chapter 3, we denote by $\hat{\theta}_n$ the MLE of θ_0; that is

$$\hat{\theta}_n \in \operatorname*{argmax}_{\theta \in \Theta} l_n(\theta),$$

by $I_1(\theta)$ the Fisher information matrix for a single observation; that is

$$I_1(\theta) = \mathbb{E}_\theta\left[\frac{\partial \log \tilde{f}(X_1|\theta)}{\partial \theta}\frac{\partial \log \tilde{f}(X_1|\theta)}{\partial \theta^T}\right], \quad \theta \in \Theta,$$

and by $KL(\theta'|\theta)$ Kullback-Leibler (KL) divergence between $\tilde{f}(x_1|\theta')\mathrm{d}x_1$ and $\tilde{f}(x_1|\theta)\mathrm{d}x_1$; that is

$$KL(\theta'|\theta) = \mathbb{E}_{\theta'}\left[\log \frac{\tilde{f}(X_1|\theta')}{\tilde{f}(X_1|\theta)}\right], \quad (\theta, \theta') \in \Theta^2.$$

Consistency of posterior distributions

Definition 6.1 *We say that the sequence of posterior distributions $\pi(\theta|X^{(n)})$ is consistent if for every $\epsilon > 0$ we have*

$$\pi(\{\theta : \|\theta - \theta_0\| \geq \epsilon\}|X^{(n)}) \to 0, \quad \mathbb{P}_{\theta_0}\text{-almost surely.}$$

Informally speaking, the posterior distribution is consistent if, as the sample size n increases, it puts more and more mass around θ_0.

The following lemma provides an alternative but equivalent definition of consistent posterior distributions.

Lemma 6.1 *A sequence of posterior distributions $\pi(\theta|X^{(n)})$ is consistent if and only if*

$$\pi(\theta|X^{(n)})\mathrm{d}\theta \overset{\text{dist.}}{\Rightarrow} \delta_{\theta_0}, \quad \mathbb{P}_{\theta_0}\text{-almost surely.}$$

Proof: admitted.

In words, this lemma says that posterior consistency is equivalent to the convergence in distribution of the posterior distribution to a Dirac mass at θ_0.

Remark: As a corollary of Lemma 6.1, $\pi(\theta|X^{(n)})$ is consistent if and only if for any continuous and bounded function $g : \Theta \to \mathbb{R}$ we have

$$\lim_{n\to+\infty} \mathbb{E}_\pi[g(\theta)|X^{(n)}] = g(\theta_0), \quad \mathbb{P}_{\theta_0}\text{-almost surely.}$$

Schwartz's consistency theorem

The following well-known result provides a sufficient condition for posterior consistency.

Theorem 6.1 (Schwartz's theorem) *Assume the following:*

(A1) For every $\eta > 0$, $\pi\big(\{\theta : KL(\theta_0|\theta) \leq \eta\}\big) > 0$*;*

(A2) For every $\epsilon > 0$ *there exists a sequence of tests* $(\phi_n)_{n\geq 1}$ *(i.e.* $\phi_n : \mathcal{X}_1^n \to \{0,1\}$*) such that, for some constants* $D_1, D_2 \in \mathbb{R}_{>0}$*,*

$$\mathbb{E}_{\theta_0}\big[\phi_n(X^{(n)})\big] \leq e^{-nD_1}, \qquad \sup_{\{\theta:\, \|\theta-\theta_0\|\geq\epsilon\}} \mathbb{E}_\theta\big[1-\phi_n(X^{(n)})\big] \leq e^{-nD_2}.$$

Then, for every $\epsilon > 0$ *we have*

$$\pi(\{\theta : \|\theta - \theta_0\| \geq \epsilon\}|X^{(n)}) \to 0, \quad \mathbb{P}_{\theta_0}\text{-almost surely.}$$

Proof: See Appendix 1.

Assumption (A1) ensures that the prior distribution π puts some mass on a neighbourhood of θ_0 (otherwise there would be no hope to get posterior consistency!).

Assumption (A2) is about the identifiability of θ_0. It assumes the existence of tests $(\phi_n)_{n\geq 1}$ that separate the singleton $\{\theta_0\}$ from the alternative $\{\theta : \|\theta - \theta_0\| \geq \epsilon\}$ in an uniform fashion.

Remark: Surprisingly, (A2) is equivalent to

(A2') For every $\epsilon > 0$ there exists a sequence of tests $(\phi_n)_{n\geq 1}$ such that

$$\lim_{n\to+\infty} \mathbb{E}_{\theta_0}\big[\phi_n(X^{(n)})\big] = \lim_{n\to+\infty} \sup_{\{\theta:\, \|\theta-\theta_0\|\geq\epsilon\}} \mathbb{E}_\theta\big[1-\phi_n(X^{(n)})\big] = 0.$$

Convergence rate of the posterior distributions

Definition 6.2 *We say that the sequence of posterior distributions* $\pi(\theta|X^{(n)})$ *converges to* θ_0 *at rate* $\epsilon_n \to 0$ *if for all sequences* $M_n \to +\infty$ *we have*

$$\pi\big(\{\theta : \|\theta - \theta_0\| \geq M_n \epsilon_n\}|X^{(n)}\big) \to 0, \quad \textit{in } \mathbb{P}_{\theta_0}\textit{-probability.}$$

The following result illustrates how the convergence rate of the posterior distribution is related to the convergence rate of point-estimates derived from it (such as the posterior mean or the posterior median; see Chapter 2).

Lemma 6.2 *Assume that the sequence of posterior distributions* $\pi(\theta|X^{(n)})$ *converges to* θ_0 *at rate* $\epsilon_n \to 0$ *and let* $\tilde{\theta}_n$ *be the centre of the smallest ball that contains posterior mass of at least* $1/2$*. Then, for every sequence* $M_n \to +\infty$ *we have*

$$\mathbb{P}_{\theta_0}\big(\|\tilde{\theta}_n - \theta_0\| \leq 2M_n\epsilon_n\big) \to 1.$$

Proof: Done in class.

In other words, Lemma 6.2 shows that the convergence rate of the estimator $\tilde{\theta}_n$ is at least ϵ_n.

A classical result for the convergence rate of posterior distributions

Theorem 6.2 *Assume that 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,*

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

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

Then, under a set of technical conditions[a] *on $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$ and on $\pi(\theta)$, we have*

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

for every sequence $M_n \to +\infty$.

Proof: See Problem Sheet 4 (for a particular model $\{\tilde{f}(\cdot|\theta), \theta \in \Theta\}$).

Remark: The assumption on the identifiability of θ_0 is stronger than in Schwartz's theorem (Theorem 6.1): Theorem 6.2 assumes the existence of tests ϕ_n that separate the singleton $\{\theta_0\}$ from the alternative $\{\|\theta - \theta_0\| > M_n/\sqrt{n}\}$, which is the complement of a shrinking ball around θ_0.

Remark: This result shows that the convergence rate of the estimator $\tilde{\theta}_n$ defined in Lemma 6.2 is at least $n^{-1/2}$.

[a] See e.g. Kleijn, B. J. K., and A. W. Van der Vaart. "The Bernstein-von-Mises theorem under misspecification." Electronic Journal of Statistics (2012): 354-381.

The Bernstein-von Mises theorem

Informally speaking, the Bernstein-von Mises theorem states that, as n increases, the posterior distribution behaves more and more like the $\mathcal{N}_d(\hat{\theta}_n, n^{-1}I_1(\theta_0)^{-1})$ distribution.

Theorem 6.3 (Bernstein von Mises theorem) *Under some technical conditions we have*

$$\sqrt{n}\big(\theta - \hat{\theta}_n\big)|X^{(n)} \overset{\text{dist.}}{\Rightarrow} \mathcal{N}_d(0, I_1(\theta_0)^{-1}), \quad \textit{in } \mathbb{P}_{\theta_0}\textit{-probability.}$$

Proof: See Appendix 2.

A first implication of this result is that a highest posterior density (HPD) region at level $(1-\alpha)$ (see Chapter 4) is asymptotically equivalent to a Wald $(1-\alpha)$-confidence interval based on the MLE; that is, a HPD at level $(1-\alpha)$ is a valid $(1-\alpha)$ confidence interval when the model is well-specified.

The following corollary gives the expansion of the log evidence we saw in Chapter 5.

Corollary 6.1 *Under some technical conditions,*

$$\log m(X^{(n)}) = l_n(\hat{\theta}_n) - \frac{d}{2}\log n + \mathcal{O}_{\mathbb{P}_{\theta_0}}(1).$$

Proof: See Appendix 3.

Posterior mean and maximum likelihood estimator

Theorem 6.4 *Under some technical conditions*[a] *we have*

$$\sqrt{n}\,(\mathbb{E}_\pi[\theta|X^{(n)}] - \hat{\theta}_n) \to 0, \quad \textit{in } \mathbb{P}_{\theta_0}\textit{-probability}$$

and therefore, if $\sqrt{n}\,(\hat{\theta}_n - \theta_0) \overset{\text{dist.}}{\Rightarrow} \mathcal{N}_d(0, I_1(\theta_0)^{-1})$, *we have*

$$\sqrt{n}\,(\mathbb{E}_\pi[\theta|X^{(n)}] - \theta_0) \overset{\text{dist.}}{\Rightarrow} \mathcal{N}_d(0, I_1(\theta_0)^{-1})$$

Proof: Admitted.

Remarks:

1. The first result shows that the difference between the MLE $\hat{\theta}_n$ and $\mathbb{E}_\pi[\theta|X^{(n)}]$ decreases quickly with n (i.e. faster than $n^{-1/2}$).
2. The second result shows that the estimator $\mathbb{E}_\pi[\theta|X^{(n)}]$ has the same asymptotic distribution as the MLE.

Altogether, these two results show that the estimator $\mathbb{E}_\pi[\theta|X^{(n)}]$ is asymptotically equivalent to the MLE, justifying the use of Bayesian methods from a frequentist perspective.

[a] See e.g. Chapter 1 of Ghosh, J. K. and Ramamoorthi, R.V. *Bayesian Nonparametrics.* Springer-Verlag New York (2013).

Bayesian asymptotics under misspecified models

In this chapter we assumed that the model was well-specified; that is, that there exists a $\theta_0 \in \Theta$ such that $X_1 \sim \tilde{f}(x_1|\theta_0)\mathrm{d}x_1$.

In practice, this assumption is never verified meaning that a model is always misspecified and it is therefore important to understand the asymptotic behaviour of Bayesian quantities in this context.

Let $\theta_0 \in \mathrm{argmax}_{\theta\in\Theta}\, \mathbb{E}_0[\log \tilde{f}(X_1|\theta)]$ and

$$V_{\theta_0} = -\frac{\partial^2 \mathbb{E}_0\big[\log \tilde{f}(X_1|\theta_0)\big]}{\partial\theta\,\partial\theta^T}, \quad \dot{l}_{\theta_0}(X_1) = \frac{\partial \log \tilde{f}(X_1|\theta_0)}{\partial\theta},$$

where the notation $\mathbb{E}_0$ is used to denote expectation under the true distribution of X_1.

Remark: If the model is well specified then θ_0 is as before the true parameter and $V_{\theta_0} = I_1(\theta_0)$ (under some additional conditions).

Then, under some technical conditions,

- the MLE is such that $\hat{\theta}_n \to \theta_0$ in $\mathbb{P}_0$-probability while

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{\text{dist.}}{\Rightarrow} \mathcal{N}_d\big(0, V_{\theta_0}^{-1}\mathbb{E}_0[\dot{l}_{\theta_0}(X_1)\dot{l}_{\theta_0}(X_1)^T]V_{\theta_0}^{-1}\big)$$

- the Bernstein von Misses theorem (Theorem 6.3) becomes

$$\sqrt{n}\big(\theta - \hat{\theta}_n)|X^{(n)} \overset{\text{dist.}}{\Rightarrow} \mathcal{N}_d(0, V_{\theta_0}^{-1}), \quad \text{in } \mathbb{P}_{\theta_0}\text{-probability}$$

  while the results in Theorem 6.1 and in Theorem 6.2 still hold[a].

Consequently,

1. The two 'procedures' converge to θ_0.
2. Since in general $V_{\theta_0}^{-1} \neq V_{\theta_0}^{-1}\mathbb{E}_0[\dot{l}_{\theta_0}(X_1)\dot{l}_{\theta_0}(X_1)^T]V_{\theta_0}^{-1}$, a HPD at level $(1-\alpha)$ is not a valid $(1-\alpha)$ confidence interval when the model misspecified.

[a]See Kleijn, B. J. K., and A. W. Van der Vaart. "The Bernstein-von-Mises theorem under misspecification." Electronic Journal of Statistics (2012): 354-381.

Appendix 1: Proof of Theorem 6.1

We start with two preliminary results.

Lemma 6.3 *Let* $(Y_n)_{n\geq 1}$ *be a sequence of random variables such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_n| \geq \epsilon) < +\infty, \quad \forall \epsilon > 0.$$

Then, $\lim_{n\to+\infty} Y_n = 0$, $\mathbb{P}$*-almost surely.*

Proof: This is a direct consequence of the Borel-Cantelli lemma.

Lemma 6.4 *Let* $(g_n)_{n\geq 1}$ *be a sequence of non-negative (measurable) functions* $g_n : \Theta \to [0, +\infty)$*. Then,*

$$\liminf_{n\to+\infty} \int_{\Theta} g_n(\theta)\pi(\theta)\mathrm{d}\theta \geq \int_{\Theta} \liminf_{n\to+\infty} g_n(\theta)\pi(\theta)\mathrm{d}\theta.$$

Proof: This is a direct consequence of Fatou's lemma.

To prove Theorem 6.1, remark first that, for every $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}_{\theta_0}(|\phi_n(X^{(n)})| \geq \epsilon) \leq \sum_{n=1}^{\infty} \frac{\mathbb{E}_{\theta_0}[\phi_n(X^{(n)})]}{\epsilon} \leq \sum_{n=1}^{\infty} \frac{e^{-nD_1}}{\epsilon} < +\infty$$

where the first inequality uses Markov's inequality (see Problem Sheet 4, Problem 2) while the second one holds under (A1).

Then, by Lemma 6.3,

$$\lim_{n\to+\infty} \phi_n(X^{(n)}) = 0, \quad \mathbb{P}_{\theta_0}\text{-almost surely.} \tag{1}$$

Appendix 1: Proof of Theorem 6.1 (continued)

Next, let $\epsilon > 0$ and $V = \{\theta : \ \|\theta - \theta_0\| \geq \epsilon\}$. We show below that

$$\limsup_{n\to+\infty} \pi(V|X^{(n)}) = 0 \quad \mathbb{P}_{\theta_0}\text{-almost surely.}$$

To this end note first that ($\mathbb{P}_{\theta_0}$-a.s.),

$$\begin{aligned}
&\limsup_{n\to+\infty} \pi(V|X^{(n)}) \\
&\leq \limsup_{n\to+\infty} \phi_n(X^{(n)})\pi(V|X^{(n)}) + \limsup_{n\to+\infty} \pi(V|X^{(n)})\big(1-\phi_n(X^{(n)})\big) \\
&\leq \limsup_{n\to+\infty} \phi_n(X^{(n)}) + \limsup_{n\to+\infty} \pi(V|X^{(n)})\big(1-\phi_n(X^{(n)})\big) \\
&= \limsup_{n\to+\infty} \pi(V|X^{(n)})\big(1-\phi_n(X^{(n)})\big)
\end{aligned} \tag{2}$$

where the equality follows from (1).

Next, write

$$\pi(V|X^{(n)})(1-\phi_n) = \frac{(1-\phi_n)\int_V \pi(\theta)\prod_{k=1}^n \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)}\mathrm{d}\theta}{\int_\Theta \pi(\theta)\prod_{k=1}^n \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)}\mathrm{d}\theta} \tag{3}$$

and we now study the denominator of the term appearing on the r.h.s. of the equality sign.

Let $\eta > 0$, $K_\eta = \{\theta : \ KL(\theta_0|\theta) \leq \eta\}$ and $\theta \in K_\eta$. (Remark that under (A1) $\pi(K_\delta) > 0$ and thus $K_\delta \neq \emptyset$). Then, by the law of large numbers,

$$\lim_{n\to+\infty} \left| \frac{1}{n}\sum_{k=1}^n \log\frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} - \mathbb{E}_{\theta_0}\left[\log\frac{\tilde{f}(X_1|\theta)}{\tilde{f}(X_1|\theta_0)}\right]\right| = 0, \quad \mathbb{P}_{\theta_0} - a.s.$$

so that, for any $\delta > 0$, there exists ($\mathbb{P}_{\theta_0}$-a.s) an $n_\delta \geq 1$ such that

$$\frac{1}{n}\sum_{k=1}^n \log\frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \geq -KL(\theta_0|\theta) - \delta, \quad \forall n \geq n_\delta.$$

Appendix 1: Proof of Theorem 6.1 (continued)

Since $\theta \in K_\eta$, we have $KL(\theta_0|\theta) \leq \eta$ and thus ($\mathbb{P}_{\theta_0}$-a.s)

$$\frac{1}{n}\sum_{k=1}^{n} \log \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \geq -(\delta+\eta), \quad \forall n \geq n_\delta.$$

Applying this result with $\delta = \eta$ implies that, for any $\theta \in K_\eta$, there exists ($\mathbb{P}_{\theta_0}$-a.s) a $n_\delta \geq 1$ such that

$$\frac{1}{n}\sum_{k=1}^{n} \log \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \geq -2\eta \Leftrightarrow \prod_{k=1}^{n} \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \geq e^{-n2\eta}, \quad \forall n \geq n_\delta.$$

Therefore ($\mathbb{P}_{\theta_0}$-a.s),

$$\begin{aligned}
\liminf_{n\to+\infty} e^{2n\eta} \int_\Theta \pi(\theta) \prod_{k=1}^{n} \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \mathrm{d}\theta &\geq \liminf_{n\to+\infty} e^{2n\eta} \int_{K_\eta} \pi(\theta) \prod_{k=1}^{n} \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \mathrm{d}\theta \\
&\geq \int_{K_\eta} \pi(\theta) \liminf_{n\to+\infty} e^{2n\eta} \prod_{k=1}^{n} \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \mathrm{d}\theta \\
&= \pi(K_\eta)
\end{aligned}$$

where the second inequality uses Lemma 6.4.

Then, using (3) we have ($\mathbb{P}_{\theta_0}$-a.s),

$$\begin{aligned}
&\limsup_{n\to+\infty} \pi(V|X^{(n)})(1-\phi_n(X^{(n)}) \\
&\leq \frac{\limsup_{n\to+\infty} e^{2n\eta}\big(1-\phi_n(X^{(n)})\big) \int_V \pi(\theta) \prod_{k=1}^{n} \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \mathrm{d}\theta}{\liminf_{n\to+\infty} e^{2n\eta} \int_\Theta \pi(\theta) \prod_{k=1}^{n} \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \mathrm{d}\theta} \qquad (4) \\
&\leq \frac{\limsup_{n\to+\infty} e^{2n\eta}\big(1-\phi_n(X^{(n)})\big) \int_V \pi(\theta) \prod_{k=1}^{n} \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)} \mathrm{d}\theta}{\pi(K_\eta)}.
\end{aligned}$$

where $\pi(K_\eta) > 0$ under (A1).

Appendix 1: Proof of Theorem 6.1 (end)

To proceed further let

$$Z_n = \left(1-\phi_n(X^{(n)})\right)\int_V \pi(\theta)\prod_{k=1}^n \frac{\tilde{f}(X_k|\theta)}{\tilde{f}(X_k|\theta_0)}\mathrm{d}\theta$$

and remark that

$$\begin{aligned}
\mathbb{E}_{\theta_0}[Z_n] &= \int_{\mathcal{X}_1^n}\left(\left(1-\phi_n(x^{(n)})\right)\int_V \pi(\theta)\prod_{k=1}^n \frac{\tilde{f}(x_k|\theta)}{\tilde{f}(x_k|\theta_0)}\mathrm{d}\theta\right)\prod_{i=1}^n \tilde{f}(x_i|\theta_0)\mathrm{d}x_i \\
&= \int_{\mathcal{X}_1^n}\left(\left(1-\phi_n(x^{(n)})\right)\int_V \pi(\theta)\prod_{k=1}^n \tilde{f}(x_k|\theta)\mathrm{d}\theta\right)\prod_{i=1}^n \mathrm{d}x_i \\
&= \int_V \left(\int_{\mathcal{X}_1^n}\left(1-\phi_n(x^{(n)})\right)\prod_{i=1}^n \tilde{f}(x_i|\theta)\mathrm{d}x_i\right)\pi(\theta)\mathrm{d}\theta \\
&= \int_V \mathbb{E}_\theta[1-\phi_n(X^{(n)})]\pi(\theta)\mathrm{d}\theta \\
&\leq \sup_{\theta\in V}\mathbb{E}_\theta[1-\phi_n(X^{(n)})] \\
&\leq e^{-nD_2}
\end{aligned}$$

where the third equality holds by Tonelli's theorem and the last inequality holds under (A2).

Then, taking $\eta > 0$ sufficiently small so that $\beta := D_2 - 2\eta > 0$, we have (using Markov's inequality for the first inequality)

$$\sum_{n=1}^\infty \mathbb{P}_{\theta_0}(|e^{2n\eta}Z_n| \geq \epsilon) \leq \sum_{n=1}^\infty \frac{e^{2n\eta}\mathbb{E}_{\theta_0}[Z_n]}{\epsilon} \leq \sum_{n=1}^\infty \frac{e^{-n\beta}}{\epsilon} < +\infty, \quad \forall \epsilon > 0.$$

Consequently, by Lemma 6.3,

$$\lim_{n\to+\infty} e^{2n\eta}Z_n = 0, \quad \mathbb{P}_{\theta_0}\text{-almost surely}$$

which, together with (2) and (4), completes the proof.

Appendix 2[a]: A proof of Theorem 6.3

We assume in this Appendix that $\Theta = \mathbb{R}$ (and thus that $d = 1$). This assumption is made to simplify the presentation but what follows can be easily generalized to any $d \geq 1$.

Below we shall consider the following assumptions.

(B1) $\tilde{f}(x|\theta) > 0$ for all $(x, \theta) \in \mathcal{X}_1 \times \Theta$.

(B2) $l(\theta, x) := \log \tilde{f}(x|\theta)$ is thrice differentiable with respect to θ in a neighbourhood $(\delta_0 - \theta_0, \theta_0 + \delta_0)$ of θ_0. If $\dot{l}$, $\ddot{l}$ and $\dddot{l}$ stand for the first, second and third derivatives, then $\mathbb{E}_{\theta_0}[\dot{l}(X_1, \theta_0)]$ and $\mathbb{E}_{\theta_0}[\ddot{l}(X_1, \theta_0)]$ are both finite and

$$\sup_{\theta \in (\delta_0 - \theta_0, \theta_0 + \delta_0)} |\dddot{l}(X_1, \theta_0)| \leq M(x), \quad \text{and } \mathbb{E}_{\theta_0}[M(X_1)] < +\infty.$$

(B3) Interchange of the order of expectation with respect to θ_0 and differentiation at θ_0 are justified, so that

$$\mathbb{E}_{\theta_0}[\dot{l}(X_1, \theta_0)] = 0, \quad \mathbb{E}_{\theta_0}[\ddot{l}(X_1, \theta_0)] = -\mathbb{E}_{\theta_0}[\dot{l}(X_1, \theta_0)]^2.$$

(B4) $I_1(\theta_0) = \mathbb{E}_{\theta_0}[\dot{l}(X_1, \theta_0)]^2 > 0$.

(B5) $\hat{\theta}_n \to \theta_0$ in $\mathbb{P}_{\theta_0}$-probability.

(B6) For any $\delta > 0$ there exists an $\epsilon > 0$ such that

$$\lim_{n \to +\infty} \mathbb{P}_{\theta_0}\Big(\sup_{\{\theta : |\theta - \theta_0| > \delta\}} \frac{1}{n}\big(l_n(\theta) - l_n(\theta_0)\big) \leq -\epsilon \Big) = 1.$$

(B7) The prior distribution has density $\pi(\theta)$ (w.r.t. the Lebesgue measure) which is continuous and positive at θ_0.

[a]This Appendix is based on Chapter 1 of Ghosh, J. K. and Ramamoorthi, R.V. *Bayesian Nonparametrics*. Springer-Verlag New York (2013).

Appendix 2: A proof of Theorem 6.3 (continued)

Let $H = \sqrt{n}(\theta - \hat{\theta}_n)$. Then, if θ has density $\pi(\theta|X^{(n)})$ then, by the change of variable formula, H has density $\pi^*(h|X^{(n)})$ defined by

$$\pi^*(h|X^{(n)}) = \frac{\pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big)\prod_{k=1}^n \tilde{f}\big(X_k|\hat{\theta}_n + \frac{h}{\sqrt{n}}\big)}{\int_{\mathbb{R}} \pi\big(\hat{\theta}_n + \frac{h'}{\sqrt{n}}\big)\prod_{k=1}^n \tilde{f}\big(X_i|\hat{\theta}_n + \frac{h'}{\sqrt{n}}\big)\mathrm{d}h'}, \quad h \in \mathbb{R}.$$

Then, we have the following result.

Theorem 6.5 *Assume $\Theta = \mathbb{R}$ and that (B1)-(B6) hold. Then,*

$$\int_{\mathbb{R}^d} \left|\pi^*(h|X^{(n)}) - \sqrt{\frac{I_1(\theta_0)}{2\pi}} e^{-\frac{h^2 I_1(\theta_0)}{2}}\right| \mathrm{d}h \to 0, \quad \textit{in } \mathbb{P}_{\theta_0}\textit{-probability.}$$

Proof: The proof follows from Lemme 6.5-6.9 stated and proved below.

Remark: The convergence in Theorem 6.5 is for the total variation metric which is stronger than the weak convergence.

Appendix 2: A proof of Theorem 6.3 (continued)

Lemma 6.5 *Assume (B4). Then, a sufficient condition for the result of Theorem 6.5 to hold is that*

$$\int_{\mathbb{R}} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\theta_0) e^{-\frac{h^2 I_1(\theta_0)}{2}} \right| \mathrm{d}h \to 0 \qquad (5)$$

in $\mathbb{P}_{\theta_0}$-probability.

Proof: Let $C_n = \int_{\mathbb{R}} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} \mathrm{d}h$. Then, noting that $\pi^*(h|X^{(n)})$ can be rewritten as

$$\pi^*(h|X^{(n)}) = \frac{\pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)}}{C_n}$$

we have

$$\begin{aligned} &\int_{\mathbb{R}^d} \left| \pi^*(h|X^{(n)}) - \sqrt{\frac{I_1(\theta_0)}{2\pi}} e^{-\frac{h^2 I_1(\theta_0)}{2}} \right| \mathrm{d}h \\ &= C_n^{-1} \int_{\mathbb{R}} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - C_n \sqrt{\frac{I_1(\theta_0)}{2\pi}} e^{-\frac{h^2 I_1(\theta_0)}{2}} \right| \mathrm{d}h. \end{aligned}$$

By (5), $C_n \to \pi(\theta_0)\sqrt{2\pi/I_1(\theta_0)}$ in $\mathbb{P}_{\theta_0}$-probability and thus to prove the lemma it is enough to show that, in $\mathbb{P}_{\theta_0}$-probability,

$$I^{(n)} := \int_{\mathbb{R}} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - C_n \sqrt{\frac{I_1(\theta_0)}{2\pi}} e^{-\frac{h^2 I_1(\theta_0)}{2}} \right| \mathrm{d}h \to 0.$$

Appendix 2: A proof of Theorem 6.3 (continued)

To this end let

$$I_1^{(n)} = \int_{\mathbb{R}} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\theta_0) e^{-\frac{h^2 I_1(\theta_0)}{2}} \right| \mathrm{d}h$$

$$I_2^{(n)} = \int_{\mathbb{R}} \left| \pi(\theta_0) e^{-\frac{h^2 I_1(\theta_0)}{2}} - C_n \sqrt{\frac{I_1(\theta_0)}{2\pi}} e^{-\frac{h^2 I_1(\theta_0)}{2}} \right| \mathrm{d}h$$

so that $I \leq I_1 + I_2$ where, by (5), $I_1^{(n)} \to 0$ in $\mathbb{P}_{\theta_0}$-probability.

In addition, $I_2^{(n)}$ can be rewritten as

$$I_2^{(n)} = \left| \pi(\theta_0) - C_n \sqrt{\frac{I_1(\theta_0)}{2\pi}} \right| \int_{\mathbb{R}} e^{-\frac{h^2 I_1(\theta_0)}{2}} \mathrm{d}h$$

and thus, as $C_n \to \pi(\theta_0)\sqrt{2\pi/I_1(\theta_0)}$ in $\mathbb{P}_{\theta_0}$-probability while, under (B4)

$$\int_{\mathbb{R}} e^{-\frac{h^2 I_1(\theta_0)}{2}} \mathrm{d}h < +\infty,$$

we have $I_2^{(n)} \to 0$ in $\mathbb{P}_{\theta_0}$-probability and the proof of Lemma 6.5 is complete.

Appendix 2: A proof of Theorem 6.3 (continued)

Lemma 6.6 *Assume (B1)-(B6) and let* $I_n = -\frac{1}{n}\sum_{i=1}^n \ddot{l}(X_i, \hat{\theta}_n)$. *Then, a sufficient condition for* (5) *to hold is that*

$$\int_{\mathbb{R}} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\hat{\theta}_n) e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \to 0$$

in $\mathbb{P}_{\theta_0}$*-probability.*

Proof: Under (B2) and (B5), with $\mathbb{P}_{\theta_0}$-probability tending to one we have (using the mean value theorem)

$$I_n = -\frac{1}{n}\sum_{i=1}^n \ddot{l}(X_i, \theta_0) - (\hat{\theta}_n - \theta_0)\frac{1}{n}\sum_{i=1}^n \dddot{l}(X_i, \theta_n')$$

for some θ_n' between $\hat{\theta}_n$ and θ_0. Hence, with $\mathbb{P}_{\theta_0}$-probability tending to one,

$$\left| I_n - \Big(-\frac{1}{n}\sum_{i=1}^n \ddot{l}(X_i, \theta_0)\Big)\right| \le |\hat{\theta}_n - \theta_0| \frac{1}{n}\sum_{i=1}^n M(X_i)$$

where , under (B2),

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n M(X_i) = \mathbb{E}_{\theta_0}[M(X_1)] < +\infty$$

while, under (B3),

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^n \ddot{l}(X_i, \theta_0) = -I_1(\theta_0).$$

Consequently,

$$I_n \to I_1(\theta_0), \qquad \text{in } \mathbb{P}_{\theta_0}\text{-probability.} \tag{6}$$

Appendix 2: A proof of Theorem 6.3 (continued)

Next, remark that

$$
\begin{aligned}
&\int_{\mathbb{R}} \left| \pi(\theta_0) e^{-\frac{h^2 I_1(\theta_0)}{2}} - \pi(\hat{\theta}_n) e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \\
&\leq \pi(\theta_0) \int_{\mathbb{R}} \left| e^{-\frac{h^2 I_1(\theta_0)}{2}} - e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h + \left| \pi(\theta_0) - \pi(\hat{\theta}_n) \right| \int_{\mathbb{R}} e^{-\frac{h^2 I_n}{2}} \mathrm{d}h \qquad (7) \\
&= \pi(\theta_0) \int_{\mathbb{R}} \left| e^{-\frac{h^2 I_1(\theta_0)}{2}} - e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h + \left| \pi(\theta_0) - \pi(\hat{\theta}_n) \right| \sqrt{\frac{2\pi}{I_n}}
\end{aligned}
$$

where, using (6) and under (B4), (B5) and (B7), the second term appearing on the r.h.s. of the last inequality sign converges to zero in $\mathbb{P}_{\theta_0}$-probability.

In addition, using the fact that $|e^x - 1| \leq |x| e^{|x|}$ for all $x \in \mathbb{R}$,

$$
\begin{aligned}
\int_{\mathbb{R}} \left| e^{-\frac{h^2 I_1(\theta_0)}{2}} - e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h &= \int_{\mathbb{R}} e^{-\frac{h^2 I_1(\theta_0)}{2}} \left| 1 - e^{-\frac{h^2 (I_n - I_1(\theta_0))}{2}} \right| \mathrm{d}h \\
&\leq \frac{\left| I_n - I_1(\theta_0) \right|}{2} \int_{\mathbb{R}} h^2 e^{-\frac{h^2 (I_1(\theta_0) - |I_n - I_1(\theta_0)|)}{2}} \mathrm{d}h.
\end{aligned}
$$

Using (6) and under (B4), with $\mathbb{P}_{\theta_0}$-probability tending to one $I_1(\theta_0) - |I_n - I_1(\theta_0)| > \epsilon$ for some $\epsilon > 0$ so that, with $\mathbb{P}_{\theta_0}$-probability tending to one,

$$
\int_{\mathbb{R}} h^2 e^{-\frac{h^2 (I_1(\theta_0) - |I_n - I_1(\theta_0)|)}{2}} \mathrm{d}h \leq C
$$

for some $C < +\infty$. Together with (6), this shows that

$$
\int_{\mathbb{R}} \left| e^{-\frac{h^2 I_1(\theta_0)}{2}} - e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \to 0, \qquad \text{in } \mathbb{P}_{\theta_0}\text{-probability.}
$$

Then, the result follows from (7) and the triangle inequality.

Appendix 2: A proof of Theorem 6.3 (continued)

Lemma 6.7 *Assume (B1)-(B6) and let* $\delta > 0$ *and* $A_1 = \{h \in \mathbb{R} : |h| > \delta/\sqrt{n}\}$. *Then,*

$$\int_{A_1} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\hat{\theta}_n) e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \to 0$$

in $\mathbb{P}_{\theta_0}$*-probability and where* I_n *is as in Lemma 6.6.*

Proof: We have

$$\int_{A_1} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\hat{\theta}_n) e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h$$
$$\leq \int_{A_1} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} \mathrm{d}h + \pi(\hat{\theta}_n) \int_{A_1} e^{-\frac{h^2 I_n}{2}} \mathrm{d}h$$

where the first integral converges to zero in $\mathbb{P}_{\theta_0}$-probability under (B6) while under (B4) the second one converges to zero in $\mathbb{P}_{\theta_0}$-probability using (6) and usual tail estimates for the normal distribution.

Appendix 2: A proof of Theorem 6.3 (continued)

Lemma 6.8 *Assume (B1)-(B6) and let* $c > 0$ *and* $A_2 = \{h \in \mathbb{R} : |h| < c\log(\sqrt{n})\}$. *Then,*

$$\int_{A_2} \left| \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\hat{\theta}_n) e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \to 0$$

in $\mathbb{P}_{\theta_0}$*-probability and where* I_n *is as in Lemma 6.6.*

Proof: Because $\hat{\theta}_n \to \theta_0$ in $\mathbb{P}_{\theta_0}$-probability under (B5), with $\mathbb{P}_{\theta_0}$-probability tending to one we have, under (B2) and using a second order Taylor expansion around $\hat{\theta}_n$,

$$\begin{aligned} l_n(\hat{\theta}_n + h/\sqrt{n}) &= l_n(\hat{\theta}_n) + \frac{1}{2}\Big(\frac{h}{\sqrt{n}}\Big)^2 \sum_{i=1}^n \ddot{l}(X_i, \hat{\theta}_n) \\ &\quad + \frac{1}{6}\Big(\frac{h}{\sqrt{n}}\Big)^3 \sum_{i=1}^n \dddot{l}(X_i, \theta_n') \\ &= -\frac{h^2 I_n}{2} + R_n(h) \end{aligned} \tag{8}$$

for some θ_n' between θ_0 and $\hat{\theta}_n$ and with

$$R_n(h) = \frac{1}{6}\Big(\frac{h}{\sqrt{n}}\Big)^3 \sum_{i=1}^n \dddot{l}(X_i, \theta_n'). \tag{9}$$

To proceed further remark that, for n large enough,

$$\begin{aligned} \sup_{h\in A_2} R_n(h) &= \sup_{h\in A_2} \frac{1}{6}\Big(\frac{h}{\sqrt{n}}\Big)^3 \sum_{i=1}^n \dddot{l}(X_i, \theta_n') \\ &\le \frac{c^3}{6}\frac{(\log(\sqrt{n}))^3}{n} \Big(\frac{1}{\sqrt{n}} \sum_{i=1}^n M(X_i)\Big) \\ &\to 0 \end{aligned} \tag{10}$$

in $\mathbb{P}_{\theta_0}$-probability since, under (B2), $\frac{1}{\sqrt{n}}\sum_{i=1}^n M(X_i) \overset{\text{dist.}}{\Rightarrow} \mathcal{N}_1\big(\mathbb{E}_{\theta_0}[M(X_1)], \mathbb{E}_{\theta_0}[M(X_1)^2] - \mathbb{E}_{\theta_0}[M(X_1)]^2\big)$.

Appendix: A proof of Theorem 6.3 (continued)

Then, using (8),

$$\begin{aligned}
&\int_{A_2} \left| \pi\big(\hat\theta_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat\theta_n + h/\sqrt{n}) - l_n(\hat\theta_n)} - \pi(\hat\theta_n) e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \\
&\quad = \int_{A_2} \left| \pi\big(\hat\theta_n + \frac{h}{\sqrt{n}}\big) e^{-\frac{h^2 I_n}{2} + R_n(h)} - \pi(\hat\theta_n) e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \\
&\quad \le \int_{A_2} \pi\big(\hat\theta_n + \frac{h}{\sqrt{n}}\big) e^{-\frac{h^2 I_n}{2}} \left| e^{R_n(h)} - 1 \right| \mathrm{d}h \\
&\quad + \sup_{h \in A_2} \left| \pi\big(\hat\theta_n + \frac{h}{\sqrt{n}}\big) - \pi(\hat\theta_n) \right| \int_{A_2} e^{-\frac{h^2 I_n}{2}} \mathrm{d}h.
\end{aligned}$$

Under (B5) and (B7),

$$\sup_{h \in A_2} \left| \pi\big(\hat\theta_n + \frac{h}{\sqrt{n}}\big) - \pi(\hat\theta_n) \right| \to 0, \quad \text{in } \mathbb{P}_{\theta_0}\text{-probability}$$

while, using (6) and under (B4), with $\mathbb{P}_{\theta_0}$-probability tending to one $\int_{A_2} e^{-\frac{h^2 I_n}{2}} \mathrm{d}h < C$ for some $C < +\infty$. Hence,

$$\int_{A_2} \pi\big(\hat\theta_n + \frac{h}{\sqrt{n}}\big) \left| e^{-\frac{h^2 I_n}{2} + R_n(h)} - e^{-\frac{h^2 I_n}{2}} \right| \mathrm{d}h \to 0$$

in $\mathbb{P}_{\theta_0}$-probability.

In addition, using the fact that $|e^x - 1| \le |x| e^{|x|}$ for all $x \in \mathbb{R}$,

$$\begin{aligned}
&\int_{A_2} \pi\Big(\hat\theta_n + \frac{h}{\sqrt{n}}\Big) e^{-\frac{h^2 I_n}{2}} \left| e^{R_n(h)} - 1 \right| \mathrm{d}h \\
&\le \sup_{h \in A_2} \Big(|R_n(h)| \pi\big(\hat\theta_n + \frac{h}{\sqrt{n}}\big) \Big) \int_{A_2} e^{-\frac{h^2 I_n}{2} + |R_n(h)|} \mathrm{d}h \\
&\to 0
\end{aligned}$$

in $\mathbb{P}_{\theta_0}$-probability using (6) and (10) and under (B5)-(B7). The proof of Lemma 6.7 is complete.

Appendix 2: A proof of Theorem 6.3 (continued)

Lemma 6.9 *Assume (B1)-(B6) and let* $c, \delta > 0$ *and* $A_3 = \{h \in \mathbb{R} : c\log(\sqrt{n}) \leq |h| \leq \delta\sqrt{n}\}$. *Then, for* δ *small enough and* c *large enough,*

$$\int_{A_3} \left|\pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big)e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\hat{\theta}_n)e^{-\frac{h^2 I_n}{2}}\right| \mathrm{d}h \to 0$$

in $\mathbb{P}_{\theta_0}$*-probability and where* I_n *is as in Lemma 6.6.*

Proof: We have,

$$\begin{aligned}
&\int_{A_3} \left|\pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big)e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} - \pi(\hat{\theta}_n)e^{-\frac{h^2 I_n}{2}}\right| \mathrm{d}h \\
&\leq \int_{A_3} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big)e^{-\frac{h^2 I_n}{2} + R_n(h)}\mathrm{d}h + \pi(\hat{\theta}_n)\int_{A_3} e^{-\frac{h^2 I_n}{2}}\mathrm{d}h
\end{aligned} \tag{11}$$

with $R_n(h)$ defined in (9).

For the second integral we have (for n large enough)

$$\begin{aligned}
\int_{A_3} e^{-\frac{h^2 I_n}{2}}\mathrm{d}h &\leq 2e^{-\frac{I_n c^2 \log(\sqrt{n})^2}{2}}\big(\delta\sqrt{n} - c\log(\sqrt{n})\big) \\
&\leq 2e^{-\frac{I_n c \log(\sqrt{n})}{2}}\big(\delta\sqrt{n} - c\log(\sqrt{n})\big) \\
&= n^{-\frac{c I_n}{4}}\big(\delta\sqrt{n} - c\log(\sqrt{n})\big)
\end{aligned}$$

where, by (6) and under (B4), $I_n \to I_1(\theta_0) > 0$ in $\mathbb{P}_{\theta_0}$-probability. Then, by taking c sufficiently large, $\int_{A_3} e^{-\frac{h^2 I_n}{2}}\mathrm{d}h \to 0$ in $\mathbb{P}_{\theta_0}$-probability so that, under (B5) and (B6),

$$\pi(\hat{\theta}_n)\int_{A_3} e^{-\frac{h^2 I_n}{2}}\mathrm{d}h \to 0 \tag{12}$$

in $\mathbb{P}_{\theta_0}$-probability.

Appendix 2: A proof of Theorem 6.3 (end)

Next, let $\gamma > 0$ and remark that for any $h \in A_3$ we have with $\mathbb{P}_{\theta_0}$-probability tending to one and under (B2) and (B5),

$$\begin{aligned} |R_n(h)| &\leq \frac{1}{6}\Big(\frac{h}{\sqrt{n}}\Big)^3 \sum_{i=1}^{n} \big|\dddot{l}(X_i, \theta'_n)\big| \\ &\leq \delta \frac{h^2}{6}\frac{1}{n}\sum_{i=1}^{n} \big|\dddot{l}(X_i, \theta'_n)\big| \\ &\leq \delta \frac{h^2}{6}\frac{1}{n}\sum_{i=1}^{n} M(X_i) \\ &\leq \delta \frac{h^2}{6}\Big(\mathbb{E}_{\theta_0}[M(X_1)] + \gamma\Big). \end{aligned}$$

Therefore, as $I_n \to I_1(\theta_0) > 0$ in $\mathbb{P}_{\theta_0}$-probability, for $\delta > 0$ sufficiently small we have

$$|R_n(h)| < \frac{h^2 I_n}{4}, \quad \text{with } \mathbb{P}_{\theta_0}\text{-probability tending to one}$$

implying that

$$-\frac{h^2 I_n}{2} + R_n(h) < -\frac{h^2}{4} I_n, \quad \text{with } \mathbb{P}_{\theta_0}\text{-probability tending to one.}$$

Therefore, with $\mathbb{P}_{\theta_0}$-probability tending to one,

$$\begin{aligned} \int_{A_3} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{-\frac{h^2 I_n}{2} + R_n(h)} \mathrm{d}h &\leq \sup_{h \in A_3} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) \int_{A_3} e^{-\frac{h^2 I_n}{4}} \mathrm{d}h \\ &\to 0 \end{aligned}$$

as $n \to +\infty$. Together with (11)-(12) this complete the proof of Lemma 6.8.

Appendix 3: Proof of Corollary 6.1

The following result is a direct consequence of Theorem 6.5.

Corollary 6.2 *Assume $\Theta = \mathbb{R}$ and that (B1)-(B6) holds. Then,*

$$\log m(X^{(n)}) = l_n(\hat{\theta}_n) - \frac{1}{2}\log n + \mathcal{O}_{\mathbb{P}_{\theta_0}}(1).$$

Proof: By Lemme 6.6-6.9,

$$\int_{\mathbb{R}} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} \mathrm{d}h \to \int_{\mathbb{R}} \pi(\theta_0) e^{-\frac{h^2 I_1(\theta_0)}{2}} \mathrm{d}h \tag{13}$$

in $\mathbb{P}_{\theta_0}$-probability where

$$\int_{\mathbb{R}} \pi(\theta_0) e^{-\frac{h^2 I_1(\theta_0)}{2}} \mathrm{d}h = \pi(\theta_0)\sqrt{\frac{2\pi}{I_1(\theta_0)}}$$

and where (using the change of variable formula)

$$\int_{\mathbb{R}} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} \mathrm{d}h = \sqrt{n} e^{-l_n(\hat{\theta}_n)} m(X^{(n)}). \tag{14}$$

Therefore, since the mapping $x \mapsto \log(x)$ is continuous, (13) implies that

$$\begin{aligned}
&\log\Big(\int_{\mathbb{R}} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} \mathrm{d}h\Big) \\
&\qquad\qquad = -l_n(\hat{\theta}_n) + \frac{1}{2}\log(n) + \log m(X^{(n)}) \\
&\qquad\qquad \to \log \pi(\theta_0) + \frac{1}{2}\log(2\pi) - \frac{1}{2}\log I_1(\theta_0)
\end{aligned}$$

as $n \to +\infty$ and in $\mathbb{P}_{\theta_0}$-probability. The proof is complete.

Remark: In the general case $d \geq 1$, (14) becomes

$$\int_{\mathbb{R}} \pi\big(\hat{\theta}_n + \frac{h}{\sqrt{n}}\big) e^{l_n(\hat{\theta}_n + h/\sqrt{n}) - l_n(\hat{\theta}_n)} \mathrm{d}h = n^{d/2} e^{-l_n(\hat{\theta}_n)} m(X^{(n)})$$

and we recover the expansion given in Corollary 6.1.