CONCENTRATION OF POSTERIOR DISTRIBUTIONS WITH MISSPECIFIED MODELS

Christophe ABRAHAM [a] and Benoît CADRE [b]

[a] Montpellier SupAgro, UMR Biométrie et Analyse des Systèmes,
2, Place P. Viala, 34060 Montpellier Cedex 1, France ;
[b] ENS Cachan - Bretagne, Campus de Ker Lann
Av. R. Schuman, 35170 Bruz

Abstract
We investigate the asymptotic properties of posterior distributions when the model is misspecified, i.e. it is comtemplated that the observations $x_1, ..., x_n$ might be drawn from a density in a family $\{h_\sigma, \sigma \in \Theta\}$ where $\Theta \subset I\!R^d$, while the actual distribution of the observations may not correspond to any of the densities h_σ. A concentration property around a fixed value of the parameter is obtained as well as concentration properties around the maximum likelihood estimate.

1. Introduction Let $x_1, x_2, \cdots$ be independent and identically distributed observations on some topological space $\mathcal{X}$, with common law Q on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where $\mathcal{B}(\Omega)$ denotes the borel σ-field of any topological space Ω. Throughout the paper, we assume that Q is absolutely continuous with respect to some probability ν on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and we denote by q its density. Let $\{h_\sigma, \sigma \in \Theta\}$ (the model) be a set of densities with respect to ν and π a prior distribution on the set $(\Theta, \mathcal{B}(\Theta))$.

Strasser (1976) studied the asymptotic of the posterior distribution when the model is correctly specified i.e. q is equal to h_θ for some $\theta \in \Theta$. In particular, it is shown that the posterior distribution of a univariate parameter is close to a normal distribution centered at the maximum likelihood estimate when the number of observations is large enough. If one does not assume that the probability model is correctly specified, it is natural to ask what happens to the properties of the posterior distribution. This question was apparently first considered in Berk (1966, 1970) where conditions under which a sequence of posterior distributions weakly converge to a degenerate distribution are given.

In this paper, we consider the multivariate case where $\Theta \subset I\!R^d$ with a misspecified model, i.e. the observations are drawn from a distribution with density q which is not assumed to correspond to any of the densities h_σ. The proofs are inspired by the proofs in Strasser (1976) and analogous asymptotic properties of the posterior distribution of a multivariate parameter are obtained under weaker assumptions. The technical results contained in this paper are given without proof in Abraham and Cadre (2002). They are, in some sense, the foundations of an article (Abraham and Cadre, 2004) in which we study the asymptotic of three measures of robustness in Bayesian Decision Theory. More precisely, let $\mathcal{D}$ be the decisions space, $l : \mathcal{D} \times \Theta \to I\!R$ be a loss function in a class $\mathcal{L}$ and denote by d_l^n a minimizer of the posterior expected loss associated with

l. By using the results of the present paper, we provide in Abraham and Cadre (2004) the asymptotic behavior of $\sup_{l\in\mathcal{L}} \|d_n^l - d_l^\theta\|$, where d_l^θ is a minimizer of $l(.,\theta)$ and θ is the true value of the parameter. This last expression can be viewed as a measure of global robustness with respect to the loss function. It can be noted that loss robustness includes prior robustness as a particular case.

The paper is organized as follows. In section 2, we set up the notations and the assumptions. Section 3 is dedicated to the concentration properties of the posterior distribution around a fixed value of the parameter and around the maximum likelihood estimate. In section 4, we study the convergence of posterior expectations.

2. Notations and hypotheses

Throughout the paper, $Q^{\otimes n}$ (resp. Q^∞) denotes the usual product distribution defined on $(\mathcal{X}^n, \mathcal{B}(\mathcal{X}^n))$ (resp. $(\mathcal{X}^\infty, \mathcal{B}(\mathcal{X}^\infty))$), where $\mathcal{X}^n = \prod_{k=1}^n \mathcal{X}$ and $\mathcal{X}^\infty = \prod_{k\geq 1} \mathcal{X}$. The space of parameters $\Theta \subset I\!R^d$ is assumed to be convex for the norm $\|.\|$ where $\|u\|$ denotes the maximum of the absolute values of the coordinates of a vector or a matrix u with real entries. If g is any Q-integrable borel function on $\mathcal{X}$, we write:

$$Q(g) = \int g(x)Q(dx).$$

For notational simplicity, any $\sup$, $\inf$ or integral taken over a subset T of $I\!R^d$ is understood to be a $\sup$, $\inf$ or integral over $T \cap \Theta$. Finally, we let for $\sigma \in \Theta$ and $x \in \mathcal{X}$:

$$f_\sigma(x) = -\log h_\sigma(x),$$

and

$$f_\sigma^{'}(x) = \Big(\frac{\partial}{\partial \sigma_i} f_\sigma(x)\Big)_{i=1,\cdots,d} \text{ and } f_\sigma^{''}(x) = \Big(\frac{\partial^2}{\partial\sigma_i \partial\sigma_j} f_\sigma(x)\Big)_{i,j=1,\cdots,d},$$

when it can be defined.

Denoting by $\bar{\Theta}$ the closure of Θ in a compact set containing Θ and by $\mathring{\Theta}$ the interior of Θ, we introduce the following assumptions on the model:

1 a) $\forall\sigma \in \Theta, \exists r > 0$ such that $\sup\{f_s, \ \|s-\sigma\| \leq r\}$ is Q-integrable;
b) $\exists\theta \in \mathring{\Theta}, \forall\sigma \in \bar{\Theta}$ with $\theta \neq \sigma : Q(f_\theta) < Q(f_\sigma)$;
c) $\forall x \in \mathcal{X}$, the application $\sigma \mapsto f_\sigma(x)$ defined on $\bar{\Theta}$ is continuous and twice continuously differentiable on $\mathring{\Theta}$;
d) $\forall\sigma \in \Theta, \exists h > 0$ such that $\sup\{\|f_s^{''}\|, \ \|s-\sigma\| \leq h\}$ is Q-integrable;
e) $\forall\sigma \in \Theta$, the matrix $A_\sigma = Q(f_\sigma^{''})$ is positive definite and the matrix I_θ defined by $A_\theta^{-1} Q(f_\theta^{'} f_\theta^{'T}) A_\theta^{-1}$ exists, and is invertible.

When the model is correctly specified, $q = h_\theta$ where θ is defined by 1b) and q is the density of Q with respect to ν . In such a case, the matrix I_θ defined in 1e) reduces to the inverse of the usual Fisher's information matrix under the classical assumption that $Q(f_\theta^{'} f_\theta^{'T}) = Q(f_\theta^{''})$.

In the following, let θ_n denote a maximum likelihood estimate. Under a misspecified model, it is known from white (1982) that θ_n is a natural estimator for the value of the parameter which minimizes the Kullback-Leibler Information Criterium $\sigma \to Q(f_\sigma) - Q(-\log(q))$. Assumption 1b) ensures that such a minimizer does exist and that it is equal to θ. Taking into account the previous remark, we can assume the following property for which sufficient conditions can be found in white (1982).

2 There exists a sequence $q_n \nearrow \infty$ when $n \nearrow \infty$ such that Q^∞-a.s., $q_n(\theta_n - \theta) \to 0$.

Finally, denote the prior distribution by π and assume the following assumptions.

3 On some neighborhood of θ, π is absolutely continuous with respect to the Lebesgue measure, the density p is continuous at θ and $p(\theta) > 0$;

4 There exists $t > 0$ such that Q^∞-a.s.:

$$\liminf_n n^t \pi\Big(\{\sigma \in \Theta \,:\, \|\sigma - \theta_n\| \leq \frac{1}{\sqrt{n}}\}\Big) > 0.$$

We let π_n be the posterior distribution i.e. for all $U \in \mathcal{B}(\Theta)$:

$$\pi_n(U) = \frac{\int_U \prod_{i=1}^n h_\sigma(x_i)\pi(d\sigma)}{\int_\Theta \prod_{i=1}^n h_\sigma(x_i)\pi(d\sigma)}.$$

The existence of π_n is studied in Berk (1970). Note that from 1a), for every $\sigma \in \Theta$, h_σ is positive Q-a.s. and the denominator of the posterior distribution is positive $Q^{\otimes n}$-a.s. as well. Thus, the posterior distribution does exist $Q^{\otimes n}$-a.s. since the denominator is finite $Q^{\otimes n}$-a.s. by the absolute continuity of Q with respect to ν.

3. Concentration properties for the posterior distribution

3.1 Posterior concentration around the true parameter

Theorem 1 *Let $g \in L_1(\pi)$ be a positive fonction. Under assumptions 1a)-1c) and 3), for all $\delta > 0$ there exists $\eta > 0$ such that:*

$$Q^{\otimes n}\left(\int_{\|\sigma-\theta\|\geq\delta} g(\sigma)\pi_n(d\sigma) > e^{-\eta n}\right) \to 0.$$

We have divided the proof into a sequence of lemmas.

Lemma 1 *Under the assumptions of Theorem 1, for all $\delta > 0$, there exists $\varepsilon > 0$ such that:*

$$Q^{\otimes n}\left(\inf_{\|\sigma-\theta\|\geq\delta} \frac{1}{n}\sum_{i=1}^n f_\sigma(x_i) \leq Q(f_\theta) + \varepsilon\right) \to 0.$$

Proof Let $\mathcal{C} = \left\{\tau \in \bar{\Theta} \,:\, \|\tau - \theta\| \geq \delta\right\}$ and fix $\tau \in \mathcal{C}$. According to 1a) and 1b), there exists an open ball centered at τ with radius less than $r(\tau)$, denoted by $U(\tau)$, such that $Q(f_\theta) < Q(\inf_{\sigma\in U(\tau)} f_\sigma)$. Let:

$$\varepsilon = \frac{1}{2}\min_{j=1,\cdots,m}\left(Q(\inf_{\sigma\in U_j} f_\sigma) - Q(f_\theta)\right) > 0,$$

where $\{U_j,\, j=1,\cdots,m\}$ is a finite cover from $\{U(\tau),\, \tau\in\mathcal{C}\}$. Then, for all $n \geq 1$:

$$\begin{aligned} & Q^{\otimes n}\left(\inf_{\|\sigma-\theta\|\geq\delta} \frac{1}{n}\sum_{i=1}^{n} f_\sigma(x_i) \leq Q(f_\theta) + \varepsilon\right) \\ \leq \quad & Q^{\otimes n}\left(\bigcup_{j=1}^{m}\left\{\frac{1}{n}\sum_{i=1}^{n}\inf_{\sigma\in U_j} f_\sigma(x_i) \leq Q(\inf_{\sigma\in U_j} f_\sigma) - \varepsilon\right\}\right), \end{aligned}$$

and the rightmost term vanishes according to 1a) and the law of large numbers. □

Lemma 2 *Under the assumptions of Theorem 1, for all $\beta > 0$, there exists $\alpha > 0$ such that:*

$$Q^{\otimes n}\left(\sup_{\|\sigma-\theta\|\leq\alpha}\frac{1}{n}\sum_{i=1}^{n} f_\sigma(x_i) \geq Q(f_\theta) + \beta\right) \to 0.$$

Proof According to 1a), 1c) and Jennrich (1969, Theorem 2), one has for some $\alpha > 0$:

$$\sup_{\|\sigma-\theta\|\leq\alpha}\left|\frac{1}{n}\sum_{i=1}^{n} f_\sigma(x_i) - Q(f_\sigma)\right| \to 0, \quad Q^\infty\text{−a.s.}$$

The lemma is then obvious since, by 1a) and 1c), the application $\sigma \mapsto Q(f_\sigma)$ is continuous. □

One can now prove Theorem 1. Its proof is closed to that of Strasser (1976, Theorem 1).

Proof of Theorem 1 Let $\delta > 0$. We then choose $\varepsilon > 0$ as defined in Lemma 1. Fix $\eta < \varepsilon$. From $\beta = (\varepsilon - \eta)/2$, we also choose $\alpha > 0$ as defined in Lemma 2. Then, for all $n \geq 1$:

$$
\begin{aligned}
& Q^{\otimes n}\left(\int_{\|\sigma-\theta\|\geq\delta} g(\sigma)\pi_n(d\sigma) > \exp(-\eta n)\right) \\
= \; & Q^{\otimes n}\left(\frac{1}{n}\log\int_{\|\sigma-\theta\|\geq\delta} g(\sigma)\exp\left(-\sum_{i=1}^{n} f_\sigma(x_i)\right)\pi(d\sigma)\right. \\
& \left. -\frac{1}{n}\log\int_{\Theta}\exp\left(-\sum_{i=1}^{n} f_\sigma(x_i)\right)\pi(d\sigma) > -\eta\right) \\
\leq \; & Q^{\otimes n}\left(-\inf_{\|\sigma-\theta\|\geq\delta}\frac{1}{n}\sum_{i=1}^{n} f_\sigma(x_i) + \sup_{\|\sigma-\theta\|\leq\alpha}\frac{1}{n}\sum_{i=1}^{n} f_\sigma(x_i) > -\eta + \frac{a}{n}\right),
\end{aligned}
$$

where $a = \log\pi(\{\sigma\in\Theta\,:\,\|\sigma-\theta\|\leq\alpha\}) - \log\int g(\sigma)\pi(d\sigma)$. Note that, according to 3), $a > -\infty$. Consequently,

$$
\begin{aligned}
& Q^{\otimes n}\left(\int_{\|\sigma-\theta\|\geq\delta} g(\sigma)\pi_n(d\sigma) > \exp(-\eta n)\right) \\
\leq \; & Q^{\otimes n}\left(-Q(f_\theta) - \varepsilon + \sup_{\|\sigma-\theta\|\leq\alpha}\frac{1}{n}\sum_{i=1}^{n} f_\sigma(x_i) > -\eta + \frac{a}{n}\right) \\
& + Q^{\otimes n}\left(\inf_{\|\sigma-\theta\|\geq\delta}\frac{1}{n}\sum_{i=1}^{n} f_\sigma(x_i) \leq Q(f_\theta) + \varepsilon\right),
\end{aligned}
$$

and the rightmost terms vanish according to Lemmas 1 and 2. □

3.2 Posterior concentration around the maximum likelihood estimate

For any $n \geq 1$ and $x \in \mathcal{X}^n$, we shall use throughout the following notations:

$$
\begin{aligned}
T(\sigma) &= \sqrt{n}\, I_\theta^{-1/2}(\sigma - \theta_n),\ \sigma\in\Theta; \\
A_n(\sigma) &= \frac{1}{n}\sum_{i=1}^{n} f_\sigma^{''}(x_i),\ \sigma\in\Theta; \\
W_n^k &= \left\{\sigma\in\Theta\,:\,\|T(\sigma)\| \leq \sqrt{k\log n}\right\},\ k>0; \\
V_n &= \left\{\sigma\in\Theta\,:\,\|\sigma-\theta_n\| \leq \frac{1}{\sqrt{n}}\right\}.
\end{aligned}
$$

Theorem 2 *Assume that 1)-4) hold. Then, for all $r > 0$ and $c > 0$, there exists $k > 0$ such that:*

$$
Q^{\otimes n}\left(\pi_n(\Theta\setminus W_n^k) > cn^{-r}\right) \to 0.
$$

The proof will be divided into three steps.

Lemma 3 *Assume that 1)-3) hold.*
i) For all $n \geq 1$, $\sigma \in \Theta$ and $x \in \mathcal{X}^n$, let $\hat{\theta}_n(\sigma) \in \Theta$ such that $\|\hat{\theta}_n(\sigma) - \theta_n\| \leq \|\theta_n - \sigma\|$. Then, for all $\varepsilon > 0$, there exists a sequence of events $(E_N(\varepsilon))_{N \geq 1}$ such that $Q^\infty(E_N(\varepsilon)^c) \to 0$ as $N \to \infty$. Furthermore, for all $\varepsilon, k > 0$, one has for all $N \geq 1$ large enough:

$$E_N(\varepsilon) \subset \bigcap_{n \geq N} \left\{ \sup_{\sigma \in V_n} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| \leq \varepsilon,\ \sup_{\sigma \in W_n^k} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| \leq \varepsilon \right\}.$$

ii) For all $k > 0$, we have Q^∞-a.s.:

$$\sup_{\sigma \in W_n^k} \|A_n(\sigma) - A_\theta\| \to 0.$$

Proof i) First note that according to 1c) and 1d), the application $\sigma \mapsto A_\sigma$ defined on Θ is continuous. Fix $\varepsilon > 0$. We then have for some $r > 0$:

$$\sup_{\|\sigma - \theta\| \leq r} \|A_\sigma - A_\theta\| \leq \frac{\varepsilon}{2}.$$

Let us denote for all $n \geq 1$:

$$H_n(\varepsilon) = \left\{ \|\theta_n - \theta\| > \frac{1}{q_n} \text{ or } \sup_{\|\sigma - \theta\| \leq r} \|A_n(\sigma) - A_\sigma\| > \frac{\varepsilon}{2} \right\},$$

where q_n is defined by 2). According to 1c), 1d), 2) and Jennrich (1969, Theorem 2), one has:

$$Q^\infty\left(\bigcap_{n \geq N} H_n(\varepsilon)^c \right) \to 1, \text{ as } N \to \infty.$$

Let for all $N \geq 1$:

$$E_N(\varepsilon) = \bigcap_{n \geq N} H_n(\varepsilon)^c.$$

The first property of i) is then proved. Moreover, let $N \geq 1$ be such that for all $n \geq N$, $n^{-1/2} + q_n^{-1} < r$. One has for all $n \geq N$:

$$E_N(\varepsilon) \subset \left\{ \sup_{\sigma \in V_n} \|\hat{\theta}_n(\sigma) - \theta\| \leq r \right\},$$

by the very definition of $\hat{\theta}_n(\sigma)$. Consequently, for all $n \geq N$:

$$E_N(\varepsilon) \subset \left\{ \sup_{\sigma \in V_n} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| \leq \varepsilon \right\}.$$

Let now $k > 0$. In a similar fashion, we can choose N large enough so that for all $n \geq N$:

$$E_N(\varepsilon) \subset \left\{ \sup_{\sigma \in W_n^k} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| \leq \varepsilon \right\},$$

hence i).

ii) Note that for all $n \geq 1$ and $k > 0$:

$$W_n^k \subset \left\{ \sigma \in \Theta \,:\, \|\sigma - \theta_n\| \leq c_0 \sqrt{\frac{\log n}{n}} \right\},$$

where $c_0 = k^{1/2} \left(\inf_{\|u\|=1} \|I_\theta^{-1/2} u\| \right)^{-1}$. Assertion ii) is then straightforward. $\square$

We introduce now the notations, for all $\delta > 0$, $k > 0$ and $n \geq 1$:

$$\begin{aligned} S_n^k(\delta) &= \left\{ \sigma \in \Theta \,:\, \frac{1}{s}\sqrt{\frac{k \log n}{n}} \leq \|\sigma - \theta_n\| \leq \delta \right\}; \\ B_\delta &= \{ \sigma \in \Theta \,:\, \|\sigma - \theta\| \leq \delta \}, \end{aligned}$$

where $s = \sup_{\|u\|=1} \|I_\theta^{-1/2} u\|$.

Lemma 4 *Let $\hat{\theta}_n(\sigma)$ be defined as in Lemma 3, for all $n \geq 1$, $\sigma \in \Theta$ and $x \in \mathcal{X}^n$, and assume that 1c), 1d) and 2) hold. Then, for all $c > 0$ there exists $\delta > 0$ such that for all $k > 0$:*

$$Q^{\otimes n} \left(\sup_{\sigma \in S_n^k(\delta)} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| > c \right) \to 0.$$

Proof Let $k, c > 0$ and $n \geq 1$. By continuity, one has for some $\delta > 0$:

$$\sup_{\|\sigma - \theta\| \leq 3\delta} \|A_\sigma - A_\theta\| \leq \frac{c}{2}.$$

Then,

$$\begin{aligned} & Q^{\otimes n} \left(\sup_{\sigma \in S_n^k(\delta)} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| > c \right) \\ \leq\ & Q^{\otimes n} \left(\sup_{\sigma \in S_n^k(\delta)} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| > c,\ S_n^k(\delta) \cup \{\theta_n\} \subset B_{2\delta} \right) \\ & + Q^{\otimes n} \left(S_n^k(\delta) \cup \{\theta_n\} \text{ not included into } B_{2\delta} \right) \\ \leq\ & Q^{\otimes n} \left(\sup_{\sigma \in B_{3\delta}} \|A_n(\sigma) - A_\sigma\| \geq \frac{c}{2} \right) + Q^{\otimes n}(\|\theta_n - \theta\| > \delta) \\ & + Q^{\otimes n} \left(S_n^k(\delta) \cup \{\theta_n\} \text{ not included into } B_{2\delta} \right). \end{aligned}$$

The first term on the right-hand side of the inequality vanishes according to Jennrich (1969, Theorem 2). The second term also vanishes according to 2), hence the lemma. $\square$

Proof of Theorem 2 For all $k, \delta > 0$ and $n \geq 1$, we have:

$$\begin{aligned} \Theta \setminus W_n^k &\subset \left\{ \sigma \in \Theta \,:\, \|\sigma - \theta_n\| \geq \frac{1}{s}\sqrt{\frac{k \log n}{n}} \right\} \\ &\subset S_n^k(\delta) \bigcup \{\sigma \in \Theta \,:\, \|\sigma - \theta_n\| > \delta\}. \end{aligned}$$

But, according to 2) and Theorem 1, one has for all $r, c, \delta > 0$:

$$Q^{\otimes n}\left(\pi_n(\{\sigma \in \Theta \,:\, \|\sigma - \theta_n\| > \delta\}) > cn^{-r}\right) \to 0,$$

and one only needs to prove that for all $r, c > 0$, there exist $k, \delta > 0$ such that:

$$Q^{\otimes n}\left(\pi_n(S_n^k(\delta)) > cn^{-r}\right) \to 0.$$

Let us first remark that according to 1b) and 2), $Q^{\otimes n}(\theta_n \notin \mathring{\Theta}) \to 0$. Hence one can assume without loss of generality that $\theta_n \in \mathring{\Theta}$ for all $n \geq 1$. Since Θ is convex, by Taylor's formula, for all $n \geq 1$, $x \in \mathcal{X}$, $\sigma \in \Theta$, there exists $\hat{\theta}_n(\sigma) \in \Theta$ such that:

$$f_\sigma(x) = f_{\theta_n}(x) + \frac{1}{2}(\sigma - \theta_n)^T f''_{\hat{\theta}_n(\sigma)}(x)(\sigma - \theta_n), \tag{1}$$

with

$$\|\hat{\theta}_n(\sigma) - \theta_n\| \leq \|\theta_n - \sigma\|. \tag{2}$$

Let $\delta, k > 0$ and $n \geq 1$. We have:

$$\pi_n(S_n^k(\delta)) = \frac{R_n(\delta, k)}{D_n},$$

where:

$$\begin{aligned} R_n(\delta, k) &= \int_{S_n^k(\delta)} \exp(-\frac{n}{2}(\sigma - \theta_n)^T A_n(\hat{\theta}_n(\sigma))(\sigma - \theta_n))\pi(d\sigma); \\ D_n &= \int_\Theta \exp(-\frac{n}{2}(\sigma - \theta_n)^T A_n(\hat{\theta}_n(\sigma))(\sigma - \theta_n))\pi(d\sigma). \end{aligned}$$

We also denote:

$$b_\theta = \inf_{\|u\|=1} |u^T A_\theta u| \text{ and } \lambda = \sup_{\|u\|=1} (\sum_{i=1}^d |u_i|)^2.$$

By 1e), one has $b_\theta > 0$. On the event

$$\left\{ \sup_{\sigma \in S_n^k(\delta)} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| \leq \frac{b_\theta}{2\lambda} \right\},$$

we have for all $\tau \in \Theta$ and $\sigma \in S_n^k(\delta)$:

$$\begin{aligned}
\tau^T A_n(\hat{\theta}_n(\sigma))\tau &\geq \tau^T A_\theta \tau - |\tau^T(A_n(\hat{\theta}_n(\sigma)) - A_\theta)\tau| \\
&\geq (b_\theta - \sup_{\sigma \in S_n^k(\delta)} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\|\lambda)\|\tau\|^2 \\
&\geq \frac{b_\theta}{2}\|\tau\|^2,
\end{aligned}$$

and consequently, one gets on the same event:

$$\begin{aligned}
R_n(\delta, k) &\leq \int_{S_n^k(\delta)} \exp(-\frac{nb_\theta}{4}\|\sigma - \theta_n\|^2)\pi(d\sigma) \\
&\leq n^{-(b_\theta k)/(4s^2)}\,\pi(B_\delta^n), \qquad (3)
\end{aligned}$$

where $B_\delta^n = \{\sigma \in \Theta : \|\sigma - \theta_n\| \leq \delta\}$. Moreover, according to Lemma 3, there exists $N \geq 1$ such that for all $n \geq N$, one has on the event $E_N(a_\theta/\lambda)$ (where $a_\theta = \sup_{\|u\|=1} |u^T A_\theta u|$), for all $\tau \in \Theta$ and $\sigma \in V_n$:

$$\begin{aligned}
\tau^T A_n(\hat{\theta}_n(\sigma))\tau &\leq \tau^T A_\theta \tau + |\tau^T(A_n(\hat{\theta}_n(\sigma)) - A_\theta)\tau| \\
&\leq (a_\theta + \sup_{\sigma \in V_n} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\|\lambda)\|\tau\|^2 \\
&\leq 2a_\theta\|\tau\|^2,
\end{aligned}$$

and hence, on the same event:

$$\begin{aligned}
D_n &\geq \int_{V_n} \exp(-\frac{n}{2}(\sigma - \theta_n)^T A_n(\hat{\theta}_n(\sigma))(\sigma - \theta_n))\pi(d\sigma) \\
&\geq \exp(-a_\theta)\pi(V_n). \qquad (4)
\end{aligned}$$

According to (3) and (4), for all $n \geq N$, we have on the event

$$E_N\left(\frac{a_\theta}{\lambda}\right) \bigcap \left\{ \sup_{\sigma \in S_n^k(\delta)} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| \leq \frac{b_\theta}{2\lambda} \right\},$$

the inequality:

$$\pi_n(S_n^k(\delta)) \leq \exp(a_\theta)\pi(B_\delta^n)\frac{n^{-(b_\theta k)/(4s^2)}}{\pi(V_n)}.$$

Hence, for all $n \geq N, r, c, \delta, k > 0$:

$$\begin{aligned}
Q^{\otimes n}\left(\pi_n(S_n^k(\delta)) > cn^{-r}\right) &\leq Q^{\otimes n}\left(\exp(a_\theta)\pi(B_\delta^n)\frac{n^{-(b_\theta k)/(4s^2)}}{\pi(V_n)} > cn^{-r}\right) \\
&\quad + Q^{\infty}\left(E_N(\frac{a_\theta}{\lambda})^c\right) \\
&\quad + Q^{\otimes n}\left(\sup_{\sigma \in S_n^k(\delta)} \|A_n(\hat{\theta}_n(\sigma)) - A_\theta\| > \frac{b_\theta}{2\lambda}\right).
\end{aligned}$$

For some choice of δ, the latter term vanishes according to Lemma 4. Moreover, the second term of the right hand side also vanishes by Lemma 3. Finally, the first term of the right hand side tends to 0 according to assumption 4), if k is such that:

$$r < \frac{b_\theta k}{4s^2} - t,$$

and the proof is complete. $\square$

4. Convergence of posterior expectations

In the sequel, F_n denotes the law $\pi \circ T^{-1}$ and K_n is the ball in Θ with center 0 and radius $\sqrt{\log n}$.

Theorem 3 *Assume that 1)-3) hold. Let $g : \Theta \to I\!R$ be a Borel function such that for some $\kappa > 0$:*

$$\int_\Theta |g(\sigma)| \ \exp(\kappa \|I_\theta^{1/2}\sigma\|^2) F_\theta(d\sigma) < \infty,$$

where F_θ is a centered normal distribution with variance matrix $I_\theta^{-1/2} A_\theta^{-1} I_\theta^{-1/2}$. Then,

$$\int_{K_n} g(\sigma) F_n(d\sigma) \to \int_\Theta g(\sigma) F_\theta(d\sigma), \quad \textit{as } n \to \infty,$$

in $Q^{\otimes n}$-probability.

Proof We can assume, with no loss of generality, that $\int_\Theta g(\sigma)F_\theta(d\sigma) = 0$.
For $n \geq 1$, $x \in \mathcal{X}$ and $\sigma \in \Theta$, let us introduce again $\hat\theta_n(\sigma)$ defined in the proof of Theorem 2, and satisfying (1) and (2). We then have according to (1), if $W_n := W_n^1$:

$$\int_{K_n} g(\tau) F_n(d\tau) = \int_{W_n} g(T(\sigma))\pi_n(d\sigma) = \frac{P_n}{D_n},$$

where D_n is the random variable of the proof of Theorem 2 and

$$P_n = \int_{W_n} g(T(\sigma)) \exp(-\frac{n}{2}(\sigma - \theta_n)^T A_n(\hat\theta_n(\sigma))(\sigma - \theta_n))\pi(d\sigma).$$

Let $\kappa_1 = 2\kappa/\lambda$, where λ has been defined in the proof of Theorem 2. Following the arguments of the proof of (4), one can obtain the existence of $N \geq 1$ such that for all $n \geq N$, we have on the event $E_N(\kappa_1)$:

$$D_n \geq \exp(-\frac{1}{2}(a_\theta + \kappa_1\lambda))\pi(V_n).$$

For the sake of simplicity, one can assume by 3) that π admits a density on V_n, and hence on the same event, we have:

$$D_n \geq \exp(-\frac{1}{2}(a_\theta + \kappa_1\lambda)) \inf_{\sigma \in V_n} p(\sigma) 2^d (dn)^{-d/2}. \tag{5}$$

Moreover, note that for all $n \geq 1$:

$$W_n \subset \left\{ \sigma \in \Theta \,:\, \|\sigma - \theta_n\| \leq c_0 \sqrt{\frac{\log n}{n}} \right\},$$

where $c_0 = \left(\inf_{\|u\|=1} \|I_\theta^{-1/2} u\| \right)^{-1}$. Consequently, by 2) and 3), one can assume (for simplicity) that π has a density on W_n. Then, for all $n \geq 1$:

$$P_n = n^{-d/2} |\det I_\theta^{1/2}| \int_\Theta g(\tau) \exp(-\frac{1}{2}(I_\theta^{1/2}\tau)^T A_\theta (I_\theta^{1/2}\tau)) X_n(\tau) d\tau,$$

where for all $\tau \in \Theta$:

$$X_n(\tau) = I_{K_n}(\tau) p(T^{-1}(\tau)) \exp(-\frac{1}{2}(I_\theta^{1/2}\tau)^T (A_n(\hat{\theta}_n(T^{-1}(\tau))) - A_\theta)(I_\theta^{1/2}\tau)).$$

But, for all $n \geq N$, we have on the event $E_N(\kappa_1)$:

$$X_n(\tau) \leq \sup_{\tau \in K_n} p(T^{-1}(\tau)) \exp(\kappa \|I_\theta^{1/2}\tau\|^2), \quad \forall \tau \in \Theta$$

since $T^{-1}(\tau) \in W_n$ as soon as $\tau \in K_n$. Moreover, according to 2), 3), and Lemma 3, one also has Q^∞-a.s.:

$$\lim_n X_n(\tau) = p(\theta), \ \forall \tau \in \Theta.$$

We deduce from Lebesgue's Theorem that on $E_N(\kappa_1)$, we have Q^∞-a.s.:

$$\lim_n n^{d/2} P_n = 0. \tag{6}$$

Let $c > 0$. We have for all $n \geq N$:

$$Q^{\otimes n}\left(|\int_{K_n} g(\sigma) F_n(d\sigma)| > c \right) \quad \leq \quad Q^\infty \left(\frac{|P_n|}{D_n} > c, \, E_N(\kappa_1) \right) + Q^\infty \left(E_N(\kappa_1)^c \right).$$

The latter term vanishes by Lemma 3. Finally, according to (5), the first term of the right hand side is smaller than:

$$Q^\infty \left(|P_n| > c \exp(-\frac{1}{2}(a_\theta + \kappa_1 \lambda)) \inf_{\sigma \in V_n} p(\sigma) 2^d n^{-d/2}, \, E_N(\kappa_1) \right)$$

which vanishes according to (6) and assumption 3). $\square$

References

Abraham, C., Cadre, B. (2002). Asymptotic properties of posterior distributions derived from misspecified models. *C.R. Math. Acad. Sci. Paris* 335, pp. 495-498.

Abraham, C., Cadre, B. (2004). Asymptotic Global Robustness in Bayesian Decision Theory. *Ann. Statist.* 32, pp. 1341-1366.

Berk, R.H. (1966). Limiting Behavior of Posterior Distributions When the Model Is Incorrect. *Ann. Math. Statist.* 37, pp. 51-58.
Berk, R.H. (1970). Consistency a Posteriori. *Ann. Math. Statist.* 41, pp. 894-906.
Strasser, H. (1976). Asymptotic properties of posterior distributions. *Z. Wahrsch. Verw. Gebiete* 35, pp. 269-282.
Jennrich, R.I. (1969). Asymptotic properties of non-linear least square estimators. *Ann. Math. Statist.* 40, pp. 633-643.
White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50, pp. 1-25.