C. R. Acad. Sci. Paris, Ser. I 335 (2002) 495–498

Statistique/*Statistics*
(Probabilités/*Probability Theory*)

Asymptotic properties of posterior distributions derived from misspecified models

Christophe Abraham [a], **Benoît Cadre** [b]

[a] ENSAM-INRA, UMR biométrie et analyse des systèmes, 2, place P. Viala, 34060 Montpellier cedex 1, France

[b] Département de probabilités et statistiques, Université Montpellier II, CC 051, place E. Bataillon, 34095 Montpellier cedex 5, France

Received 1 May 2002; accepted after revision 26 July 2002

Note presented by Paul Deheuvels.

Abstract We investigate the asymptotic properties of posterior distributions when the model is misspecified, i.e. it is contemplated that the observations $x_1, \ldots, x_n$ might be drawn from a density in a family $\{h_\sigma,\ \sigma \in \Theta\}$ where $\Theta \subset \mathbb{R}^d$, 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. ***To cite this article: C. Abraham, B. Cadre, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 495–498.***

Propriétés asymptotiques des lois a posteriori sous un modèle incorrect

Résumé Nous étudions les propriétés asymptotiques des lois a posteriori lorsque la distribution des observations est mal spécifiée, c'est-à-dire lorsque la loi a posteriori est construite à partir d'une famille de densités $\{h_\sigma,\ \sigma \in \Theta\}$ où $\Theta \subset \mathbb{R}^d$, alors que la vraie loi des observations peut ne correspondre à aucune densité h_σ. Nous obtenons des propriétés de concentration de la loi a posteriori autour d'une valeur fixe du paramètre ainsi que des propriétés de concentration autour de l'estimateur du maximum de vraisemblance. ***Pour citer cet article : C. Abraham, B. Cadre, C. R. Acad. Sci. Paris, Ser. I 335 (2002) 495–498.***

1. Introduction

Let $x_1, x_2, \ldots$ 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 [5] 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

E-mail addresses: abraham@ensam.inra.fr (C. Abraham); cadre@stat.math.univ-montp2.fr (B. Cadre).

S1631-073X(02)02520-7/FLA

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 [3] and [4] 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 $\Theta \subset \mathbb{R}^d$ 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 [5] 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, in some sense, the foundations of the article [2] in which we study the asymptotic of three measures of robustness in Bayesian Decision Theory. More precisely, let $\mathcal{D}$ be the decisions space, $l : \mathcal{D} \times \Theta \to \mathbb{R}$ 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. We provide in [2], for instance, the asymptotic behavior of $\sup_{l\in\mathcal{L}} \|d_l^n - d_l^\theta\|$, where d_l^θ is a minimizer of $l(\cdot,\theta)$ and θ is the true value of the parameter.

The paper is organized as follows. In Section 2, we set up the notations and the assumptions. In the third section, we have compiled three theorems about the asymptotic properties of the posterior distribution. The results of this paper are announced without proofs. For the proofs we refer the reader to [1].

2. Notations and hypotheses

Throughout the paper, $Q^{\otimes n}$ (resp. $Q^{\otimes \mathbb{N}}$) denotes the usual product distribution defined on $(\mathcal{X}^n, \mathcal{B}(\mathcal{X}^n))$ $(\mathcal{X}^\infty, \mathcal{B}(\mathcal{X}^\infty))$ respectively, where $\mathcal{X}^n = \prod_{k=1}^n \mathcal{X}$ and $\mathcal{X}^\infty = \prod_{k\geqslant 1} \mathcal{X}$. The space of parameters $\Theta \subset \mathbb{R}^d$ is assumed to be convex for the norm $\|\cdot\|$ 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:

$$Q(g) = \int g(x)Q(\mathrm{d}x).$$

For notational simplicity, any sup, inf or integral taken over a subset T of $\mathbb{R}^d$ 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

$$f'_\sigma(x) = \left(\frac{\partial}{\partial\sigma_i} f_\sigma(x)\right)_{i=1,\dots,d} \quad \text{and} \quad f''_\sigma(x) = \left(\frac{\partial^2}{\partial\sigma_i\partial\sigma_j} f_\sigma(x)\right)_{i,j=1,\dots,d},$$

when it can be defined.

Denoting by $\overline{\Theta}$ 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 $\sup\{f_s,\ \|s-\sigma\| \leqslant r\}$ is Q-integrable;
   (b) $\exists\theta \in \mathring{\Theta}$, $\forall\sigma \in \overline{\Theta}$ 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 $\overline{\Theta}$ is continuous and twice continuously differentiable on $\mathring{\Theta}$;
   (d) $\forall\sigma \in \Theta$, $\exists h > 0$ such that $\sup\{\|f''_s\|,\ \|s-\sigma\| \leqslant h\}$ 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

$$A_\theta^{-1} Q\left(f'_\theta f'^{\mathrm{T}}_\theta\right) A_\theta^{-1}$$

exists, and is invertible.

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

In the following, let θ_n denote a maximum likelihood estimate. Under a misspecified model, it is known from [6] 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 (1)(b) 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 [6].

(2) There exists a sequence $q_n \nearrow \infty$ when $n \nearrow \infty$ such that $Q^{\otimes \mathbb{N}}$-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 $Q^{\otimes \mathbb{N}}$-a.s.:

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

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

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

The existence of π_n is studied in [4]. The absolute continuity of Q with respect to ν and assumption (1)(a) entails the existence of π_n $Q^{\otimes n}$-a.s.

3. Concentration properties for the posterior distribution

Theorem 3.1 provides a concentration property of the posterior distribution around $\theta \in \Theta$ while Theorem 3.2 deals with concentration in a neighborhood of a maximum likelihood estimate.

THEOREM 3.1. – *Let $g \in L_1(\pi)$ be a positive fonction. Under assumptions* (1)(a)–(1)(c) *and* (3), *for all $\delta > 0$ there exists $\eta > 0$ such that*:

$$Q^{\otimes n}\left(\int_{\|\sigma-\theta\|\geqslant\delta} g(\sigma)\pi_n(\mathrm{d}\sigma) > \mathrm{e}^{-\eta n}\right) \to 0, \quad \textit{as } n \to \infty.$$

For any $n \geqslant 1$ and $(x_1, \ldots, x_n) \in \mathcal{X}^n$, we shall use throughout the following notations:

$$T_n(\sigma) = \sqrt{n} I_\theta^{-1/2}(\sigma - \theta_n), \quad \sigma \in \Theta;$$
$$W_n^k = \left\{\sigma \in \Theta : \|T_n(\sigma)\| \leqslant \sqrt{k \log n}\right\}, \quad k > 0.$$

THEOREM 3.2. – *Assume that* (1)–(4) *hold. Then, for all $r > 0$ and $c > 0$, there exists $k > 0$ such that*:

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

In the sequel, F_n denotes the law $\pi_n \circ T_n^{-1}$ and K_n denotes the ball in Θ with center 0 and radius $\sqrt{\log n}$. F_n can be viewed as a measure of the gap between the parameter σ with posterior distribution π_n and the maximum likelihood estimate θ_n. Roughly, Theorem 3.3 says that F_n converges to a normal distribution.

THEOREM 3.3. – *Assume that* (1)–(3) *hold. Let* $g : \Theta \to \mathbb{R}$ *be a borel function such that for some* $\kappa > 0$:

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

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

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

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

References

[1] C. Abraham, B. Cadre, Asymptotic properties of posterior distributions derived from misspecified models, Rapport 02-05 INRA-ENSAM-UMII.

[2] C. Abraham, B. Cadre, Asymptotic global robustness in Bayesian decision theory, submitted.

[3] R.H. Berk, Limiting behavior of posterior distributions when the model is incorrect, Ann. Math. Statist. 37 (1966) 51–58.

[4] R.H. Berk, Consistency a posteriori, Ann. Math. Statist. 41 (1970) 894–906.

[5] H. Strasser, Asymptotic properties of posterior distributions, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 35 (1976) 269–282.

[6] H. White, Maximum likelihood estimation of misspecified models, Econometrica 50 (1982) 1–25.