

New Tools for Consistency in Bayesian Nonparametrics

GABRIELLA SALINETTI

Università di Roma “La Sapienza”, Italy

saline@pow2.sta.uniroma1.it

SUMMARY

Posterior consistency and the parallel behaviour of consistency of maximum likelihood estimators is analyzed in nonparametric statistical problems. The framework is the hypo-Strong Law of Large Numbers, a form of “one-sided” Uniform Law of Large Numbers.

Keywords: CONSISTENCY; LAW OF LARGE NUMBERS; MAXIMUM LIKELIHOOD ESTIMATORS.

1. INTRODUCTION

Consistency in Bayesian nonparametric statistical problems continues to register an increasing attention, motivated by various and easy to share reasons, widely illustrated in the rich literature on the subject starting from Diaconis and Freedman (1986) to the recent Ghosh and Ramamoorthi (2002).

The difficulty in nonparametrics is in the dimension of the problem: the “parameter” itself is typically a probability measure, the prior is a probability measure on a space of probability measures and the resulting posterior, after seeing the data, is a random probability measure.

Posterior consistency at a given “parameter value” P_0 reduces then to almost sure (weak) convergence of the sequence of the posterior (random) probability measures to a probability measure concentrated in P_0 . On the other hand, once elicited the prior, the posterior depends on the data through the likelihood function and then its asymptotic behaviour depends on the asymptotic behaviour of the likelihood, or the log of it, a functional of the empirical process determined by the observations.

With this view it is easy to realize that the standard Strong Law of Large Numbers, certainly a key tool in analyzing the asymptotic behaviour of the likelihood, is too weak to determine consistency results; on the other hand Uniform Strong Laws of Large Numbers, often invoked in consistency of statistical functionals in nonparametric contexts, certainly produces consistency but they could reveal themselves too much demanding in applications.

Hypo-Strong Law of Large Numbers, somehow between standard and uniform laws, seems to delineate a natural framework to deal with consistency of statistical functionals including posterior consistency.

Hypo-Strong Law of Large Numbers is based on the notion of hypo-convergence of functions; this is basically convergence of the hypographs of the functions (the set below the graph) and its relevance derives from the fact that it delineates the “minimal” setting for convergence of the suprema (values and solutions). In this role, almost sure hypo-convergence finds its natural and fruitful application in consistency of those statistical functionals which are, or can

be expressed as, solutions of stochastic optimization problems; this is typically the case of Maximum Likelihood Estimators. The essential of hypo-convergence and hypo-Strong Law of Large Numbers is given in Section 3.

The parallel behaviour of consistency of Maximum Likelihood Estimators and posterior consistency, often observed in the literature on consistency analyses, has been the initial motivation to analyze the role and the implications of the hypo-Strong Law of Large Numbers in posterior consistency. The main implication is posterior consistency on “compacts”, a necessary prelude for posterior consistency; tightness of the posteriors is then a necessary and sufficient condition for posterior consistency; compactification or natural compact embedding can become then a way to verify posterior consistency. This is illustrated in Section 4.

Section 5 explores the parallelism between consistency of Maximum Likelihood Estimators and posterior consistency. Finally in Section 6, together with conclusive remarks, further developments are delineated.

2. PROBLEM SETTING

Let E be the finite dimensional Euclidean space where the observations take value and \mathcal{E} its Borel field. Let Θ be “parameter” space always assumed to be metric separable and complete, with its Borel field $\mathcal{B}(\Theta)$. For each $\theta \in \Theta$, P_θ is a probability measure on \mathcal{E} and let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. The prior beliefs are expressed by a prior probability measure Π on the Borel field $\mathcal{B}(\Theta)$. The observations $\{X_n, n \geq 1\}$ are random variables valued in E and modeled, given θ , as i.i.d. with probability law P_θ ; it will be always convenient to look at the observations as co-ordinate process on (Ω, \mathcal{A}) , $\Omega = E^\infty$, $\mathcal{A} = \mathcal{E}^\infty$.

The prior Π is updated by observing X_1, \dots, X_n in the posterior $\Pi(\cdot | X_1, \dots, X_n)$, conditional distribution on $\mathcal{B}(\Theta)$ given X_1, \dots, X_n . The sequence $\{\Pi(\cdot | X_1, \dots, X_n), n \geq 1\}$ is a sequence of random probability measures on $\mathcal{B}(\Theta)$, simply denoted $\Pi_n(B, \cdot)$ with $B \in \mathcal{B}(\Theta)$.

The sequence $\{\Pi_n, n \geq 1\}$ is said to be *consistent at θ_0* , if there exists $\Omega_0 \subset \Omega$ with $P_{\theta_0}^\infty(\Omega_0) = 1$ such that if $\omega \in \Omega_0$ then for every open neighborhood U of θ_0 $\Pi_n(U, \omega) \rightarrow 1$ or, equivalently, $\Pi_n(U^c, \omega) \rightarrow 0$, where U^c denotes the complement of U in Θ .

Equivalently, posterior consistency at θ_0 holds if for all $\omega \in \Omega_0$, the sequence of probability measures $\{\Pi_n(\cdot, \omega), n \geq 1\}$ weakly converges to Π_0 where Π_0 is the probability measure giving mass 1 to θ_0 .

Here the attention is restricted to the case where the probability measures P_θ are absolutely continuous with respect to the same σ -finite measure λ ; correspondent densities are denoted $f(\cdot; \theta)$ and it is understood that $\theta \neq \theta'$ implies $f(\cdot; \theta) \neq f(\cdot; \theta')$; in fact, we will identify Θ with the set $\{f(\cdot; \theta); \theta \in \Theta\}$ when no confusion can arise. The likelihood, for the observations $X_1(\omega), \dots, X_n(\omega)$ is $\prod_{i=1}^n f(X_i(\omega), \theta)$ and the posterior $\Pi_n(\cdot, \cdot)$ can now be expressed as

$$\Pi_n(B, \omega) = \frac{\int_B \prod_{i=1}^n f(X_i(\omega), \theta) \Pi(d\theta)}{\int_\Theta \prod_{i=1}^n f(X_i(\omega), \theta) \Pi(d\theta)}, \quad B \in \mathcal{B}(\Theta).$$

The expression clarifies that the posterior depends only on the empirical measure determined by the observations. In fact, with standard manipulations, $P_{\theta_0}^\infty$ -a.s., we have

$$\Pi_n(B, \omega) = \frac{\int_B \exp [nH_n(\omega, \theta)] \Pi(d\theta)}{\int_\Theta \exp [nH_n(\omega, \theta)] \Pi(d\theta)} \quad (1)$$

where

$$H_n(\omega, \theta) = \frac{1}{n} \sum_{i=1}^n \log \frac{f(X_i(\omega), \theta)}{f(X_i(\omega), \theta_0)} = \int_E \log \frac{f(x, \theta)}{f(x, \theta_0)} P_n(\omega, dx)$$

and $P_n(\omega, \cdot)$ is the empirical measure determined by the observations $X_1(\omega), \dots, X_n(\omega)$. Statements involving the probability measure $P_{\theta_0}^\infty$ correspond “formally” to a situation where the observations $\{X_n, n \geq 1\}$ are i.i.d. P_{θ_0} ; and from now on the underlying probability space is $(\Omega, \mathcal{A}, P_{\theta_0}^\infty)$. Observe that

$$H_0(\theta) = \int \log \frac{f(x, \theta)}{f(x, \theta_0)} P_{\theta_0}(dx) = - \int f(x, \theta_0) \log \frac{f(x, \theta_0)}{f(x, \theta)} \lambda(dx) = -K(\theta_0, \theta)$$

where $K(\theta_0, \theta)$ is the Kullback-Leibler divergence of P_θ from P_{θ_0} . Once rephrased $\Pi_n(U^c, \omega)$ in the form of (1), consistency is expressed as

$$\Pi_n(U^c, \omega) = \frac{\int_{U^c} \exp [nH_n(\omega, \theta)] \Pi(d\theta)}{\int_{\Theta} \exp [nH_n(\omega, \theta)] \Pi(d\theta)} \rightarrow 0 ; \quad (2)$$

it is evident that the asymptotic behaviour of Π_n depends on the asymptotic behaviour of H_n and its relation with H_0 ; in fact, as it will be motivated next, on the hypo-convergence $P_{\theta_0}^\infty$ -a.s. of $\{H_n\}$ to H_0 .

The basic assumptions maintained in the rest of the paper are now introduced.

For every $\delta > 0$ let $K_\delta(\theta_0)$ be the Kullback-Leibler neighborhood of θ_0

$$K_\delta(\theta_0) = \{\theta \in \Theta : K(\theta_0, \theta) < \delta\}.$$

Even if main applications refer to the Hellinger metric, or the equivalent L_1 -metric, the metric on Θ here will remain mostly not specified, the main goal being to identify situations of consistency or possible inconsistency issues depending on the metric adopted and the topological nature of the parameter space. It is also understood, until the metric is not specified, that we are inside “permissible” problems in the sense that when we write for example $\Pi(K_\delta(\theta_0))$ it is understood that $K_\delta(\theta_0) \in \mathcal{B}(\Theta)$.

As usually done in analyzing posterior consistency it will be assumed not only that $\Pi(U) > 0$ for every open neighborhood U of θ_0 , a necessary condition for posterior consistency, but also that θ_0 is in the Kullback-Leibler support of the prior Π , i.e. :

$$(A) \quad \Pi(K_\delta(\theta_0)) > 0 \quad \forall \delta > 0.$$

We also tacitly assume that in the metric adopted on Θ , for every open neighborhood U of θ_0

$$(B) \quad \inf_{\theta \in U^c} K(\theta_0, \theta) > 0,$$

a condition satisfied in the cases of interest.

As consequence of (A), through the Strong Law of Large Numbers we have (see for example Ghosh and Ramamoorthi (2002) or Barron et al. (1999) for the derivation)

$$\forall \delta > 0, \quad \liminf_{n \rightarrow \infty} e^{n\delta} \int \prod_{i=1}^n \frac{f(X_i(\omega), \theta)}{f(X_i(\omega), \theta_0)} \Pi(d\theta) = \infty \quad P_{\theta_0}^\infty\text{-a.s.} \quad (3)$$

Condition (A) and its consequence (3) controls the denominator in (2). We have for all $\omega \in \Omega_\delta$, $P_{\theta_0}^\infty(\Omega_\delta) = 1$, and n sufficiently large

$$\int \exp [nH_n(\omega, \theta)] \Pi(d\theta) > e^{-n\delta}. \quad (4)$$

In this case, as it will be seen later, it is easy to verify posterior consistency at θ_0 if $P_{\theta_0}^\infty$ -a.s., for any open neighborhood U of θ_0

$$\sup_{\theta \in U^c} H_n(\omega, \theta) \longrightarrow \sup_{\theta \in U^c} H_0(\theta) \quad (5)$$

or, more generally

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U^c} H_n(\omega, \theta) < 0. \quad (6)$$

The convergence expressed by conditions (5) or (6), or variations of these, as a.s. convergence of suprema is directly connected with the a.s. hypo-convergence of the sequence $\{H_n(\omega, \cdot)\}$ to $H_0(\cdot)$ and it introduces a preliminary motivation to analyze posterior consistency in presence of hypo-convergence.

3. HYPO-CONVERGENCE AND HYPO-STRONG LAW OF LARGE NUMBERS

Hypo-convergence for a sequence of functions is basically set convergence of their hypographs. For the purposes of this paper it will suffice to restrict the attention to functions defined on a metric separable and complete space (Θ, d) , even if some properties and results hold in more general settings.

For a function $H : \Theta \rightarrow \bar{\mathbb{R}}$, the hypograph is the set $\text{hypo } H = \{(\theta, \alpha) \in \Theta \times \mathbb{R} : H(\theta) \geq \alpha\}$, i.e. all the points below the graph of H .

A sequence of functions $\{H_n, n \geq 1\}$, *hypo-converges* to the function H at θ if for every sequence $\theta_m \rightarrow \theta$ and any subsequence $\{n_m\}$

$$\limsup_{m \rightarrow \infty} H_{n_m}(\theta_m) \leq H(\theta) \quad (7)$$

and there exists $\theta_n \rightarrow \theta$ such that

$$H(\theta) \leq \liminf_{n \rightarrow \infty} H_n(\theta_n). \quad (8)$$

If the two relations above hold for every $\theta \in \Theta$, the sequence $\{H_n, n \geq 1\}$ is said to *hypo-converge* to H , denoted $H_n \overset{h}{\rightarrow} H$, or $H = \text{hypo-}\lim_{n \rightarrow \infty} H_n$.

The sequence $\{H_n, n \geq 1\}$ is said to *hypo-converge* to H on a subset C of Θ if relations (7) and (8) hold when restricted to C , i.e. for every $\theta \in C$ and every sequence $\{\theta_n\}$ in C . When only condition (7) holds for every θ we simply write $\text{hypo-}\limsup_{n \rightarrow \infty} H_n \leq H$.

Basic and extended references to hypo-convergence, its motivations and its role in convergence of optimization problems are in Attouch (1984). Most of the literature on the subject concerns epi-convergence (convergence of epigraphs) and minimization problems, but any result on epi-convergence has its counterpart in the mirror setting of hypo-convergence. We refer to hypo-convergence, the choice being determined by the need to preserve the structure of the statistical problems where it will be used.

Hypo-convergence yields convergence of maximizers and optimal values in the sense stated below. The literature on the subject is quite rich, but here the basic aspects will suffice.

Proposition 3.1. *Suppose that $H_n \overset{h}{\rightarrow} H$. Then*

$$\sup_{\theta \in \Theta} H(\theta) \leq \liminf_{n \rightarrow \infty} \sup_{\theta \in \Theta} H_n(\theta). \quad (9)$$

Moreover, if there is a subsequence $\{\theta_{n_m}\}$ such that for every m , $\theta_{n_m} \in \text{argmax } H_{n_m}$ and $\theta_{n_m} \rightarrow \hat{\theta}$ then $\hat{\theta} \in \text{argmax } H$ and $\sup_{\theta \in \Theta} H_n(\theta) \rightarrow \sup_{\theta \in \Theta} H(\theta)$.

This result is well-known; an elementary proof (in the setting of epi-convergence) is given in Dong and Wets (2000). As immediate consequence of (7) we have

Proposition 3.2. *If $\text{hypo-lim sup}_{n \rightarrow \infty} H_n \leq H$ then for every compact set C*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in C} H_n(\theta) \leq \sup_{\theta \in C} H(\theta)$$

Actually a form of characterization of convergence of suprema in terms of hypo-convergence can be given (see for example Atouch (1984, Theorem 2.11)) as follows

Proposition 3.3. *If $H_n \xrightarrow{h} H$ then the following are equivalent:*

- (i) $\sup_{\theta \in \Theta} H_n(\theta) \rightarrow \sup_{\theta \in \Theta} H(\theta)$ and $\text{argmax } H \neq \emptyset$
- (ii) $\exists \epsilon_n \downarrow 0$ and a relatively compact sequence $\{\theta_n\}$ such that for all n , $\theta_n \in \epsilon_n\text{-argmax } H_n = \{\theta : H_n(\theta) > \sup H_n - \epsilon_n\}$
- (iii) $\exists \epsilon_n \downarrow 0$ and a non empty relatively compact set C such that for all n

$$\min \left\{ \frac{1}{\epsilon_n}, \sup_{\theta \in \Theta} H_n(\theta) - \epsilon_n \right\} < \sup_{\theta \in C} H_n(\theta)$$

Remark - The sense of this result is that in presence of hypo-convergence, convergence of suprema can fail only when ϵ -optimal solutions “move out” of the space, living on complements of compact sets.

We turn now to random functions $H_n(\cdot, \cdot)$ defined on a probability space $(\Omega, \mathcal{A}, \mu)$ which map every $\omega \in \Omega$ into a function $H_n(\omega, \cdot) : \Theta \rightarrow \mathfrak{R}$.

For a sequence of such functions $\{H; H_n, n \geq 1\}$, μ -a.s. hypo-convergence of H_n to H means, as natural to expect, that for all $\omega \in \Omega \setminus N$, $\mu(N) = 0$,

$$H_n(\omega, \cdot) \xrightarrow{h} H(\omega, \cdot).$$

The theory of random semicontinuous functions, with upper (or lower) semicontinuous realizations, based on the topology generated by hypo- (respectively epi-) convergence, has received remarkable attention as the appropriate setting to deal with stochastic processes with semicontinuous realizations (Salinetti and Wets, 1986, 1990). In the specific setting of consistency of statistical estimators, as functionals of the empirical measures, it will suffice to refer to $P_{\theta_0}^\infty$ -a.s. convergence of random functions generated by a sequence of i.i.d. random variables to which we refer as *hypo-Strong Law of Large Numbers*.

Let $\{X_n, n \geq 1\}$ be a sequence of random variables valued in (E, \mathcal{E}) i.i.d. with probability law P_{θ_0} , regarded as co-ordinate process on $(\Omega, \mathcal{A}, \mu)$ with $\Omega = E^\infty$, $\mathcal{A} = \mathcal{E}^\infty$ and $\mu = P_{\theta_0}^\infty$. For a function $h : E \times \Theta \rightarrow \bar{\mathfrak{R}}$ let

$$H_n(\omega, \theta) = \frac{1}{n} \sum_{i=1}^n h(X_i(\omega), \theta) = \int_E h(x, \theta) P_n(\omega, dx)$$

where $P_n(\omega, \cdot)$ is the empirical probability measure determined by X_1, \dots, X_n and let

$$H_0(\theta) = \int_E h(x, \theta) P_{\theta_0}(dx).$$

The sequence of random functions $\{H_n\}$ satisfies the *hypo-Strong Law of Large Numbers* (hypo-SSLN) if for all $\omega \in \Omega \setminus N$, $P_{\theta_0}^\infty(N) = 0$

$$H_n(\omega, \cdot) \xrightarrow{h} H_0(\cdot).$$

We basically refer here to the version reproduced in Dong and Wets (2000), Theorem A.4. A close derivation aiming at consistency of Maximum Likelihood Estimators (MLE’s) is given in Hess (1996). The basic assumptions under which the hypo-SLLN holds require that the function $h : E \times \Theta \rightarrow \bar{\mathfrak{R}}$ is $\mathcal{E} \times \mathcal{B}(\Theta)$ -measurable and for each $x \in E$, $h(x, \cdot)$ is upper semicontinuous on Θ , i.e. h is a random upper semicontinuous function.

Theorem 3.4. Let (Θ, d) be a Polish space, $\{X_n, n \geq 1\}$ a sequence of i.i.d. random variables valued in (E, \mathcal{E}) with common probability law P_{θ_0} defined on \mathcal{E} with \mathcal{E} P_{θ_0} -complete and let $h : E \times \Theta \rightarrow \bar{\mathbb{R}}$ be a random upper semicontinuous function. Suppose that for every $\theta \in \Theta$ there exists a neighborhood U of θ and a measurable function $\alpha : E \rightarrow \bar{\mathbb{R}}$ with $\int \alpha^+ dP_{\theta_0} < \infty$ such that P_{θ_0} -a.s. $h(\cdot, \theta') \leq \alpha(\cdot)$ for all $\theta' \in U$. Then the hypo-SLLN holds, i.e. $P_{\theta_0}^\infty$ -a.s.

$$H_n(\omega, \cdot) \xrightarrow{h} H_0(\cdot).$$

Remark - For the problems that will be approached in the next section often it will be actually sufficient a one-sided version of the hypo-SLLN in the form

$$P_{\theta_0}^\infty\text{-a.s. hypo-}\limsup_{n \rightarrow \infty} H_n(\omega, \cdot) \leq H_0(\cdot).$$

In view of Propositions 3.2, 3.3, and Theorem 3.4 it is easy to realize the role of hypo-SLLN in consistency of statistical estimators when these are obtained, or can be expressed, as solution of optimization problems; this is quite frequent in the context of classical statistics and in fact a.s. hypo-convergence, and specifically hypo-SLLN, offers at the same time the theoretical setting (stochastic processes with semicontinuous realizations) for statistical functionals of empirical processes and a unified derivation of consistency properties. An approach of this type is pursued in Salinetti (1990) and Salinetti (2001) even if mainly restricted to the parametric case or regular “parameter” spaces.

Consistency of MLE’s based on a.s. hypo-convergence is obtained in Hess (1996) under quite general assumptions. The same problem is approached in Dong and West (2000) in a more general setting where “prior” information, mainly in the form of model assumptions, is formalized as “constraints” of the maximization problem.

At the same time, also posterior consistency, when based on convergence of stochastic suprema as sketched in (5) and (6) takes advantages of the potential offered by hypo-convergence, as it will be seen in the next section.

The conditions under which hypo-SLLN holds do not appear particularly restrictive in applications, at least for Maximum Likelihood Estimators cases; the uppersemicontinuity of the integrand is often a consequence of the metric adopted on Θ ; examples are worked out in Dong and West (2000) and Hess (1996) and some easy applications are in the next section.

4. POSTERIOR CONSISTENCY AND HYPO-CONVERGENCE

In accordance to Section 2, posterior consistency requires that for every $\omega \in \Omega_0$, $P_{\theta_0}^\infty(\Omega_0) = 1$, and every open neighborhood U of θ_0 ,

$$\Pi_n(U^c, \omega) = \frac{\int_{U^c} \exp [nH_n(\omega, \theta)] \Pi(d\theta)}{\int_{\Theta} \exp [nH_n(\omega, \theta)] \Pi(d\theta)} \rightarrow 0.$$

A preliminary consideration easily leads to

Proposition 4.1. If θ_0 belongs to the Kullback-Leibler support of Π and $P_{\theta_0}^\infty$ -a.s. for every open neighborhood U of θ_0

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U^c} H_n(\omega, \theta) \leq \sup_{\theta \in U^c} H_0(\theta) \quad (10)$$

then posterior consistency at θ_0 holds.

Proof. By (B) there exists $\epsilon > 0$ such that $\sup_{\theta \in U^c} H_0(\theta) < -\epsilon$. Let $\Omega_1 = \Omega \setminus N_1$, $P_{\theta_0}^\infty(N_1) = 0$, the set where (10) and the consequence (4) of (A) for $\delta < \epsilon$ hold. For $\omega \in \Omega_1$ and n sufficiently large we have

$$\Pi_n(U^c, \omega) \leq e^{-n(\epsilon-\delta)} \Pi(U^c)$$

and the result follows. \square

The result, based on the convergence of suprema (10), induces to study posterior consistency in presence of a.s. hypo-convergence of the sequence $\{H_n(\omega, \cdot)\}$ to $H_0(\cdot)$ appealing to the recognized fact that hypo-convergence is the “minimal” setting for convergence of suprema (values and solutions).

However, the connection between hypo-convergence and posterior consistency is deeper and related to the tightness of the sequence of posteriors.

Since posterior consistency at θ_0 is also $P_{\theta_0}^\infty$ -a.s. weak convergence of $\{\Pi_n(\cdot, \omega)\}$ to $\Pi_0(\cdot)$ and (Θ, d) is a Polish space, a necessary condition for posterior consistency at θ_0 is $P_{\theta_0}^\infty$ -a.s. tightness of the sequence $\{\Pi_n(\cdot, \omega), n \geq 1\}$, i.e. for every $\omega \in \Omega_0$, $P_{\theta_0}^\infty(\Omega_0) = 1$, and $\alpha > 0$ it has to exist a compact subset C_α of Θ such that $\Pi_n(C_\alpha^c, \omega) < \alpha$ for all n .

If $P_{\theta_0}^\infty$ -a.s., tightness is guaranteed (for example we know that Π has compact support), then for all $\omega \in \Omega_0$ and every neighborhood U of θ_0 we have

$$\Pi_n(U^c, \omega) = \Pi_n(U^c \cap C_\alpha, \omega) + \Pi_n(U^c \cap C_\alpha^c, \omega) < \Pi_n(U^c \cap C_\alpha, \omega) + \alpha.$$

Thus posterior consistency at θ_0 follows if we can prove that for every α and every neighborhood U of θ_0 , $\Pi_n(U^c \cap C_\alpha, \omega) \rightarrow 0$.

Actually the sequence $\{\Pi_n, n \geq 1\}$ is posterior consistent at θ_0 if and only if there exists Ω_0 , $P_{\theta_0}^\infty(\Omega_0) = 1$, such that for every $\omega \in \Omega_0$

- (i) $\{\Pi_n(\cdot, \omega), n \geq 1\}$ is tight
- (ii) for any open neighborhood U of θ_0 and any compact set C , $\Pi_n(U^c \cap C, \omega) \rightarrow 0$.

We now show that condition (ii) holds if $P_{\theta_0}^\infty$ -a.s., the sequence $\{H_n(\omega, \cdot)\}$ hypo-converges to $H_0(\cdot)$. In fact just as consequence of Proposition 3.2 we have

Proposition 4.2. *Suppose that θ_0 belongs to the Kullback-Leibler support of Π and $P_{\theta_0}^\infty$ -a.s. the following conditions hold*

- (i) $\{\Pi_n(\cdot, \omega), n \geq 1\}$ is tight
- (ii) $H_n(\omega, \cdot) \xrightarrow{h} H_0(\cdot)$

then $\{\Pi_n\}$ is posterior consistent at θ_0 .

Proof. Let Ω_0 , $P_{\theta_0}^\infty(\Omega_0) = 1$, be such that on it (i) and (ii) hold. By tightness, for every $\alpha > 0$ there exists C_α compact such that $\Pi_n(C_\alpha^c, \omega) < \alpha$. Let U be an open neighborhood of θ_0 . Let $\epsilon > 0$ such that $\sup_{U^c \cap C_\alpha} H_0(\theta) < -\epsilon$. Let $0 < \delta < \epsilon$ and Ω_δ , $P_{\theta_0}^\infty(\Omega_\delta) = 1$, such that, for all $\omega \in \Omega_\delta$ and n sufficiently large, (4) holds.

Hypo-convergence, by Proposition 3.2, yields

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U^c \cap C_\alpha} H_n(\omega, \theta) \leq \sup_{\theta \in U^c \cap C_\alpha} H_0(\theta) < -\epsilon. \quad (11)$$

For n sufficiently large the last relation with (4) yields

$$\Pi_n(U^c \cap C_\alpha, \omega) < e^{-n(\epsilon-\delta)} \Pi(U^c \cap C_\alpha);$$

it follows

$$\Pi_n(U^c \cap C_\alpha, \omega) \rightarrow 0 \quad \text{and} \quad \limsup_{n \rightarrow \infty} \Pi_n(U^c, \omega) \leq \alpha.$$

The argument repeated for every α gives the conclusion. \square

Remark - The role played by hypo-convergence a.s. in the proof is in obtaining (11). It is useful to observe that only the “hypo-limsup” part of it enters and actually it will suffice simply the “hypo-limsup” part on compact sets.

Relation (11) is evidently a key point, tightness being in any case a necessary condition. It is relevant to observe that under general conditions (see for example Barron et al. (1999)) $P_{\theta_0}^\infty$ -a.s. we have

$$H_n(\omega, \cdot) \longrightarrow H_0(\cdot) \quad \Pi\text{-a.s.},$$

however “pointwise” convergence is not sufficient for convergence of suprema.

The argument used in the last Proposition actually allows to conclude

Proposition 4.3. *If θ_0 belongs to the Kullback-Leibler support of Π and $P_{\theta_0}^\infty$ -a.s. $H_n(\omega, \cdot) \xrightarrow{h} H_0(\cdot)$, then for every compact subset C of Θ not containing θ_0*

$$\Pi_n(C, \omega) \rightarrow 0. \quad (12)$$

Moreover $\{\Pi_n\}$ is posterior consistent at θ_0 if and only if $P_{\theta_0}^\infty$ -a.s., $\{\Pi_n(\cdot, \omega)\}$ is tight.

Proof. If C does not contain θ_0 there exists an open neighborhood U of θ_0 with $U \cap C = \emptyset$ so that $C \subset U^c$ and then, by (B), $\sup_{\theta \in C} H_0(\theta) < 0$. The argument used in the last Proposition then shows (12). Tightness, as sufficient condition, is Proposition 4.2; the necessity is obvious. \square

Let us consider now the tightness condition. If θ_0 belongs to the Kullback-Leibler support of Π , a sufficient condition for $P_{\theta_0}^\infty$ -a.s. tightness is that for every $\omega \in \Omega_0 = \Omega \setminus N_0$, $P_{\theta_0}^\infty(N_0) = 0$, there exists a compact subset C of Θ such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in C^c} H_n(\omega, \theta) < 0. \quad (13)$$

The argument to show it is the usual one: $P_{\theta_0}^\infty$ -a.s., appealing to (A) and (13), for $\epsilon > 0$ and $\delta_k \downarrow 0$ we get

$$\Pi_n(C^c, \omega) = \frac{\int_{C^c} \exp [nH_n(\omega, \theta)] \Pi(d\theta)}{\int \exp [nH_n(\omega, \theta)] \Pi(d\theta)} < e^{-n(\epsilon - \delta_k)} \Pi(C^c)$$

and this is enough to state the tightness for $\Pi_n(\cdot, \omega)$.

The argument above with Proposition 4.2 gives

Proposition 4.4. *If θ_0 belongs to the Kullback-Leibler support of Π and $P_{\theta_0}^\infty$ -a.s. the following conditions hold*

(i) *there exists a compact C such that*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in C^c} H_n(\omega, \theta) < 0$$

(ii) $H_n(\omega, \cdot) \xrightarrow{h} H_0(\cdot)$

then $\{\Pi_n\}$ is posterior consistent at θ_0

The development followed here aims at pointing out the role of hypo-convergence in posterior consistency: hypo-SLLN ensures the right posterior convergence on compact sets not

containing θ_0 ; possible inconsistency can only derive from lack of tightness and this is recognized in the tail behaviour of $H_n(\omega, \cdot)$ where tails are complements of compact sets.

Of course if Θ is compact then the hypo-SLLN implies posterior consistency. This is also the case when Θ is locally compact and convex with convex metric, because in this case the concavity of $H_n(\omega, \cdot)$ and the strict concavity of $H_0(\cdot)$ recovers on the tightness.

The specific nature of these results, consistency on compacts and tail problems, where tails are complements of compacts, suggests to analyze posterior consistency in suitable compact embedding of the parameter space, where conditions for hypo-SLLN are easily satisfied. Two cases are briefly examined: the first one, relative to discrete observations is well-known Freedman (1963). The revisitation aims at illustrating the procedure and the way it can be extended in the more general situation of the second example.

Example 1. [Discrete observations] The observations take values in a countable set $I = \{x_1, x_2, \dots\}$ and the parameter space consists of all the probability distributions on I . As in Freedman (1963), let $S = \{f : I \rightarrow [0, 1]\}$ with the product topology so S is compact and metrizable; $L = \{p \in S : \sum p(x_r) \leq 1\}$, with the relative topology, is compact too. The parameter space $\Theta = \{\theta \in L : \sum \theta(x_r) = 1\}$ is a G_δ set of L . Let Π be the prior on $\mathcal{B}(\Theta)$, the Borel field of Θ with the relative topology of L ; suppose that θ_0 belongs to the Kullback-Leibler support of Π and $-\int \log \theta_0(x) P_{\theta_0}(dx) < \infty$. In accordance with Proposition 4.1 posterior consistency at every θ_0 follows if $P_{\theta_0}^\infty$ -a.s., for every open neighborhood U of θ_0

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U^c} H_n(\omega, \theta) < 0. \tag{14}$$

We obtain it passing to the compact embedding L . Let us consider on the compact space L the functions $\hat{H}_n(\omega, p) = \int \log \frac{p(x)}{\theta_0(x)} P_n(\omega, dx)$ and $\hat{H}_0(p) = \int \log \frac{p(x)}{\theta_0(x)} P_{\theta_0}(dx)$, extensions of $H_n(\omega, \theta)$ and $H_0(\theta)$ defined on Θ . It is not difficult to prove that \hat{H}_0 is uppersemicontinuous on L and for every open neighborhood U_L of θ_0 in L , $\sup_{U_L^c} \hat{H}_0 < 0$. It is also easy to verify that hypo-SLLN holds for the sequence $\{\hat{H}_n\}$; this follows after observing that the conditions of Theorem 3.4 are satisfied by the function $h(x, \theta) = \log \frac{p(x)}{\theta_0(x)}$; observe that $p(x) \leq 1$, so the dominant function α is $-\log \theta_0(x)$. Since U_L^c is compact, it follows by Proposition 3.2, $P_{\theta_0}^\infty$ -a.s.

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U_L^c} \hat{H}_n(\omega, p) \leq \sup_{\theta \in U_L^c} \hat{H}_0(p) < 0.$$

On the other hand for any open neighborhood U of θ_0 in Θ , with U^c complement of U in Θ , we have $U^c \subset U_L^c$ where U_L is an open neighborhood of θ_0 in L ; so $\sup_{\theta \in U^c} H_n(\omega, \theta) \leq \sup_{\theta \in U_L^c} \hat{H}_n(\omega, p)$ and (14) follows.

The argument above shows, as in Freedman (1963, Theorem 3.2),

The posterior is consistent at every θ_0 in the Kullback-Leibler support of Π such that $-\int \log \theta_0(x) P_{\theta_0}(dx) < \infty$.

Example 2. [Unimodal densities] Without entering in technical details, we consider the case where the parameter space consists of unimodal and uppersemicontinuous (u.sc) densities on the real line, dominated by an u.sc. function ψ . The mode is the same, conventionally put at 0 and the modal value is $+\infty$. Let $L_0 = \{f : \mathbb{R} \rightarrow [0, \infty], f \text{ unimodal}, f(0) = \infty, \int f d\lambda \leq 1\}$, where λ is the Lebesgue measure, $\Theta_0 = \{\theta \in L_0 : \int f d\lambda = 1\}$ so the parameter space is $\Theta = \{\theta \in \Theta_0, \theta \leq \psi\}$. The space L_0 can be equipped with a metric based on the Levy distance for monotone functions; with this metric (L_0, d) is metric and compact as shown in Reiss (1973),

Θ_0 is a G_δ set in L_0 and convergence $d(f_n, f) \rightarrow 0$ in L_0 means pointwise convergence on the continuity set of f .

Let Π be a prior on the Borel field of Θ_0 with $\Pi(\Theta) = 1$ and let θ_0 be in the Kullback-Leibler support of Π and such that $\int \log \frac{\psi(x)}{\theta_0(x)} P_{\theta_0}(dx) < \infty$. It is easy to realize that the topological structure of the problem is analogous to the previous case. Also in this case on the compact space $L = \{f \in L_0, f \leq \psi\}$ the extensions $\hat{H}_n(\omega, f) = \int \log \frac{f(x)}{\theta_0(x)} P_n(\omega, dx)$ and $\hat{H}_0(f) = \int \log \frac{f(x)}{\theta_0(x)} P_{\theta_0}(dx)$, satisfy the conditions of Theorem 3.4 for the hypo-SLLN; \hat{H}_0 is u.s.c. on L and for every open neighborhood U_L of θ_0 in L , $\sup_{U_L^c} \hat{H}_0(f) < 0$. The argument above allows to conclude similarly to the previous case that *the posterior is consistent at any θ_0 in the Kullback-Leibler support of Π such that $\int \log \frac{\psi(x)}{\theta_0(x)} P_{\theta_0}(dx) < \infty$.*

It is relevant to observe that since in this class of unimodal functions, convergence in the metric d is equivalent to convergence in L_1 , as shown in Reiss (1973), we actually get posterior consistency in the L_1 -metric. Extensions to more general situations are possible, for example when the mode is not fixed but seats in a compact interval.

Remark In both examples the parameter space has a global dominant function, constantly 1 in the first case and ψ in the second case. This is evidently more restrictive than what requested by Theorem 3.4; in this direction “sieves”-type constraints could lead to workable extensions.

5. POSTERIOR CONSISTENCY AND CONSISTENCY OF MLE’S

In this section the parallel behaviour of posterior consistency and consistency of Maximum Likelihood Estimators is analyzed and differences are highlighted. The parallelism between the two consistencies is delineated in Ghosh and Ramamoorthi (2002) and reference to consistency of Maximum Likelihood Estimators often appears in the literature when dealing with posterior consistency. An explicit connection between posterior consistency and consistency of Maximum Likelihood Estimators is in fact already contained in Strasser (1981). With the notations of Section 2 the Maximum Likelihood Estimator for θ_0 , when existing, is $\hat{\theta}_n(\omega) = \arg \max_{\theta \in \Theta} H_n(\omega, \theta)$. Existence and measurability questions are not the main concern here; we will refer to Approximate Maximum Likelihood Estimators (AMLE) where the sets of α -approximate MLE’s, $\alpha > 0$, are given by

$$A_{n,\alpha}(\omega) = \{\theta \in \Theta : H_n(\omega, \theta) > \sup H_n(\omega, \theta) - \alpha\};$$

AMLE’s consistency for θ_0 means here that there exists Ω_0 , $P_{\theta_0}^\infty(\Omega_0) = 1$, such that for all $\omega \in \Omega_0$ and any sequence $\alpha_n \downarrow 0$, for every open neighborhood U of θ_0

$$A_{n,\alpha_n}(\omega) \subset U$$

for n sufficiently large; in fact a restrictive notion of consistency of AMLE’s which however serves the purpose of comparison.

As easy completion of Proposition 4.2 we immediately have:

Proposition 5.1. *If θ_0 belongs to the Kullback-Leibler support of Π and $P_{\theta_0}^\infty$ -a.s., for every open neighborhood U of θ_0 we have*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U^c} H_n(\omega, \theta) < 0 \tag{15}$$

then AMLE’s consistency and posterior consistency at θ_0 both hold.

Proof. Posterior consistency at θ_0 is Proposition 4.1. For $\omega \in \Omega_0$, with $P_{\theta_0}^\infty(\Omega_0) = 1$, $\{\alpha_n\}$ a sequence of positive numbers decreasing to 0, U any neighborhood of θ_0 , since $H_n(\omega, \theta_0) = 0$, relation (15) implies that $A_{n, \alpha_n}(\omega) \subset U$ for n sufficiently large and AMLE's consistency follows. \square

Beyond any definition of AMLE's it seems relevant to observe that most of the literature on consistency of AMLE's is based on a system of conditions where it is easy to recognize, in addition to the uppersemicontinuity of the maps $\theta \mapsto \log f(x, \theta)$, for all x , assumptions as "local dominance" of the type considered in Theorem 3.4 and "tail" assumptions which include global dominance, semi-dominance and global uniformity as in Perlman (1972) or, more recently, one-sided bracketing condition as in Dudley (1998). Uppersemicontinuity and local dominance imply hypo-SLLN; then, based on tail conditions, consistency of AMLE's for θ_0 is obtained showing that, $P_{\theta_0}^\infty$ -a.s., (15) holds.

In all these cases then, which include Huber (1967) and Wald (1949), posterior consistency and AMLE's consistency both hold.

In general however, consistency of AMLE's does not imply (15) unless additional regularity assumptions are introduced such as local compactness (Perlman, 1972); on the other hand consistency of AMLE's can be obtained without passing through (15), as for example in (Hess, 1996).

Of course the same parallel behaviour is observed under the conditions of Proposition 4.4 as it can be easily proved. We have:

Proposition 5.2. *If θ_0 belongs to the Kullback-Leibler support of Π and $P_{\theta_0}^\infty$ -a.s. the following conditions hold*

(i) *there exists a compact C such that*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in C^c} H_n(\omega, \theta) < 0$$

(ii) $H_n(\omega, \cdot) \xrightarrow{h} H_0(\cdot)$

then posterior consistency and AMLE's consistency both hold.

The result suggests a final comment on the relation between the two consistencies. In view of all the above in dealing with AMLE's consistency, almost sure hypo-convergence is necessary as minimal requirement for convergence of stochastic suprema; conditions on tails based on the existence of convenient compact sets are sufficient. The situation is in a sense mirrored for posterior consistency: tightness is a necessary condition, almost sure hypo-convergence is sufficient, with the additional remark that, if tightness is guaranteed, almost sure hypo-convergence on compacts is sufficient.

6. SOME CONCLUSIONS AND POSSIBLE DEVELOPMENTS

Hypo-SLLN is a functional limit wich prelude to convergence of stochastic suprema; then a preliminary requirement for consistency of AMLE's and a potential useful tool to verify posterior consistency through convergence of stochastic suprema. It does not imply, alone, posterior consistency; it implies consistency in presence of tightness and somehow it produces "approximate" consistency as consistency on compacts; it locates sources of inconsistency in the tails identified with complements of compact sets, thus alerting on the behaviour of the prior on these sets. Moreover, it is a tool to obtain posterior consistency in those circumstances where there are "natural" compactifications and the crucial properties of the functionals involved are

preserved in the extension to the compact embedding. The real point is the existence of suitable compactifications.

The case of unimodal u.s.c. densities, regarded from a different perspective, indicates a way to proceed in this direction, relying on the topology generated by hypo-convergence. It is easy to show that convergence in the metric of unimodal u.s.c. functions is equivalent to hypo-convergence of them and actually many of the properties, compactness included, could be easily obtained as consequence of this fact.

In fact the topology generated by hypo-convergence is compact and metrizable (Dolecki et al. 1983); thus when the prior information selects a.s. u.s.c. densities, not necessarily unimodal, a compact embedding is available. Of course, in the Bayesian setting this way to proceed has to be accompanied by a device to elicit a prior on the space of u.s.c. functions. In this direction the theory developed for random u.s.c. functions states that a probability measure on the Borel field generated by the hypo-topology determines and is uniquely determined by a Choquet capacity or hitting function (Salinetti and Wets, 1986); this function, with properties which are the natural extension of the properties of the distribution function of a random variable, can play the role of a general device to assign nonparametric prior on the space of u.s.c. functions.

REFERENCES

- Attouch, H. (1984). *Variational Convergence for Functions and Operators*. Boston, MA: Pitman.
- Barron, A., Schervish, M. J. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27**, 536–561.
- Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–67. (with discussion).
- Dolecki, S., Salinetti, G. and Wets, R. J.-B. (1983). Convergence of functions: equi-semicontinuity. *Ann. Math. Statist.* **276**, 409–430.
- Dong, M. X. and Wets, R. J.-B. (2000). Estimating density functions: a constrained maximum likelihood approach. *J. Nonparametr. Statist.* **12**, 549–595.
- Dudley, R. M. (1998). *Consistency of M-estimators and one-sided bracketing. High dimensional probability (Oberwolfach, 1996)*. Basel: Birkhäuser, 33–58.
- Freedman, D. A. (1963). On the asymptotic behavior of Bayes estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386–1403.
- Ghosh, J. K. and Ramamoorthi, R. V. (2002). Bayesian nonparametrics Unpublished manuscript.
- Hess, C. (1996). Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. *Ann. Statist.* **24**, 1298–1315.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp.* **1** (L. M. LeCam and J Neyman, eds.). Berkeley: Univ. California Press, 221–233.
- Perlman, M. D. (1972). On the strong consistency of approximate maximum likelihood estimators. *Proc. Fifth Berkeley Symp.* **1** (L. M. LeCam and J Neyman, eds.). Berkeley: Univ. California Press,
- Reiss, R.-D. (1973). On the measurability and consistency of maximum likelihood estimates for unimodal densities. *Ann. Statist.* **1**, 888–901.
- Salinetti, G. (1990). On structural relationships between probability and statistics: Empirical processes and statistical functionals (Italian). In *Proceedings of the Italian Statistical Society*.
- Salinetti, G. (2001). Consistency of statistical estimators: the epigraphical view. In *Stochastic optimization: algorithms and applications (Gainesville, FL, 2000)*. Dordrecht: Kluwer, 365–383.
- Salinetti, G. and Wets, R. J.-B. (1986). On the convergence in distribution of 15 measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima. *Math. Oper. Res.* **11**, 385–419.
- Salinetti, G. and Wets, R. J.-B. (1990). Random semicontinuous functions. In *Lectures in applied mathematics and informatics*. Manchester: University Press, 330–353.
- Strasser, H. (1981). Consistency of maximum likelihood and Bayes estimates. *Ann. Statist.* **9**, 1107–1113.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595–601.

DISCUSSION

JAYANTA K. GHOSH (*Purdue University, USA*)

This is a very clear, enjoyable paper. I begin by quickly summing up,

I. $\Pi(\theta \in U^c | X_1, X_2, \dots, X_n)$

$$= \frac{\int_{U^c} e^{nHn(\theta, X_1, \dots, X_n)} \Pi(d\theta)}{\int_{\Theta} e^{nHn(\theta, X_1, \dots, X_n)} \Pi(d\theta)}$$

where nHn is the log likelihood ratio as in the paper.

Under the relatively simple condition that θ_0 is in Kullback – Leibler support of Π , Schwartz (1965, ZW) showed the denominator

$$= 0(e^{-n\delta}) \text{ a.s. } P_{\theta_0}, \quad (1)$$

for every positive δ , however small.

So to show $\Pi(\theta \in U^c | X_1, X_2, \dots, X_n)$

$$\rightarrow 1 \text{ a.s. } P_{\theta_0},$$

one has to show the numerator

$$= O(e^{-n\delta_0}) \text{ a.s. } P_{\theta_0} \text{ for some } \delta_0 > 0 \quad (2)$$

Since the numerator is bounded by

$$\exp\left\{n \sup_{\theta \in U^c} H_n(\theta, X_1, X_2, \dots, X_n)\right\}$$

one way of showing posterior consistency is to study conditions under which the limit (or lim sup) of

$$\sup_{\theta \in U^c} H_n(\theta, X_1, \dots, X_n) \leq -\delta_0 \text{ a.s. } P_{\theta_0} \quad (3)$$

GS points out that a minimal and natural setting for studying (3) is “hypo-convergence” of H_n . This is well-established by her. She also exhibits several interesting implications of “hypo-convergence”. From these follow her interesting observation that one needs hypo convergence and tightness for (2) and hence for posterior consistency.

II. Comments, suggestions

- A. The essence of this idea is well-known but traditionally one uses “uniform convergence” instead of hypo convergence which is much weaker.
- B. The snag is that GS has not told us how to replace the many Uniform Strong Laws of Large Number (USLLN) by Hypo-SLLN’s with conditions that are easy to verify. Will bracketing entropy and uniform strong laws based on them help?
- C. Also, to handle high or infinite dimensional Θ , one may have to develop Hypo-convergence theorems in the context of sieves.

Finally, using (3) to prove (2) may not always be efficient for infinite dimensional (non-parametric) examples. Integrals are better behaved than maxima — a fact that is easy to accept for Bayesians. Bayesians prefer to integrate the likelihood instead of maximizing.

For example (as shown by Schwartz) if U is a weak neighbourhood, then (2) holds *without any assumptions*, whereas we do not know any such general fact about (3).

Also if $\sup_{\theta \in U^c} H_n(\theta, X_1, X_2, \dots, X_n)$ converges to infinity a.s. P_{θ_0} , the methods of hypo convergence will not work.

There is a famous counter example of Bahadur with countable Θ where the mle is not consistent. It is treated in detail in Lehmann's "Point Estimation". Ghosh and Ramamoorthi (forthcoming book on Bayesian Nonparametrics) establish $\sup_{\theta \in U^c} H_n(\theta, X_1, X_2, \dots, X_n)$ converges to infinity a.s. P_{θ_0} . But by Doob's theorem the posterior is consistent at any θ_0 in the support of the prior.

SUBHASHIS GHOSAL (*North Carolina State University, U. S. A.*) and
AAD VAN DER VAART (*Free University Amsterdam, The Netherlands*)

This interesting paper attempts to relax the need for uniform convergence of log likelihood ratios in consistency theorems. Salinetti's proof (e.g., Propositions 4.4 and 4.1) for posterior consistency involves two steps. First it is noted that a standard consistency proof goes through if

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U^c} H_n(\omega, \theta) < 0.$$

She then shows that when the space is compact, hypo-convergence yields the condition above. In non-compact spaces, U^c must be intersected with a compact set C and the posterior probability of the remaining part must be bounded by other methods.

Below we try to argue why the notion of hypo-convergence is only moderately useful in the present context.

First, it does not yield global results, but must be complemented by showing tightness of the posterior sequence, which can be almost as hard as the original problem. Instead of a fixed compact, one could consider a sieve consisting of an increasing sequence of suitable compact sets. Then the posterior probability of the tail can be bounded by showing that its prior probability is exponentially small. However, hypo-convergence does not seem to yield the desired conclusion for the central part when it increases with n .

Secondly, because of averaging by the prior, convergence of the supremum of the log likelihood ratios is more than necessary for the consistency of the posterior. The inadequacy of the approach based on bounds for the supremum is particularly pronounced in a countable parameter space where posterior consistency holds under minimal conditions by Doob's (1948) theorem.

Thirdly, the concept of hypo-convergence appears conceptually more complicated than the condition (*) it is supposed to derive. The sufficient condition for hypo-convergence as in Theorem 3.4 is a device already available in Wald's consistency proof, and can be used directly to derive (*).

Finally, we note that (*) implies the classical testing condition of Schwartz (1965) for posterior consistency and hence it is not possible to derive any new consistency result using hypo-convergence. To see this, consider the test that accepts $H_0 : \theta = \theta_0$ when $\sup_{\theta \in U^c} H_n(\omega, \theta) < -\varepsilon$ for some suitable $\varepsilon > 0$. Then (*) implies that the type I error probability is bounded away from 1. The type II error probability at any $\theta \in U^c$ is bounded by

$$P_\theta \left(\sup_{\theta' \in U^c} H_n(\omega, \theta') < -\varepsilon \right) \leq P_\theta (H_n(\omega, \theta) < -\varepsilon) \leq P_\theta \left(\prod_{i=1}^n \frac{f^{1/2}(X_i, \theta_0)}{f^{1/2}(X_i, \theta)} \geq e^{n\varepsilon/2} \right) \leq e^{-n\varepsilon/2}$$

by Markov's inequality. The desired test can now be constructed using the i.i.d. structure.

The hypothesis of hypo-convergence does help to construct the desired tests. Because in a compact parameter space, the desired tests exist by the finiteness of the metric entropy (see e.g., Ghosal et al., 1999), this is again only of moderate usefulness.

REPLY TO THE DISCUSSION

First, I would like to thank the discussants for their useful suggestions and comments. The suggestions are all accepted and worth to be further developed; the comments, especially the critical ones, are all constructive and help to clarify potential and limits of hypo-convergence.

In answering the comments the role of a.s. hypo-convergence in convergence of stochastic suprema will be kept separated from its role in presence of tightness or in suitable compact embeddings. Preliminary to all that it has to be remarked that the approach to posterior consistency based on hypo-convergence does not replace neither competes with uniformly consistent tests; these, when existing, provide a complete answer. Also, the reference to a compact parameter space as particular instance, can be of limited interest but this is not the mainpoint as it will be tried to argue.

The fact that if a.s., for every open neighborhood U we have

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in U^c} H_n(\omega, \theta) < 0 \quad (1)$$

then posterior consistency, as well as MLE's consistency hold, discloses the nature of the often observed parallel behaviour of the two consistencies; since most of the literature on MLE's provide conditions under which (1) holds, posterior consistency, as already observed, can be obtained in all those circumstances where the prior information, including modeling assumptions, and the topology chosen, design a space Θ where these conditions are satisfied. In that, a.s. hypo-convergence is a preliminary "necessary" requirement, not sufficient in general, for the convergence of stochastic suprema. This will be the case if Θ is compact but that can be of limited interest in most applications. Weak compactness is easier to achieve either because is provided by the metric adopted on Θ or requiring for example that the solutions of the optimization problem are bounded. In this case, a.s. Mosco-hypo-convergence, a somewhat stronger notion of hypo-convergence, provides convergence of suprema (Dong and Wets, 2000). This path parallels with the possibility offered by upper bracketing entropy with the consequent "upper" Glivenko-Cantelli classes which remains a key tool.

In the same direction to obtain posterior consistency through convergence of suprema, the suggestion to develop the hypo-convergence in the context of sieves is appealing; the a.s. convergence of $\sup_{\theta \in S_n} H_n(\omega, \theta)$ to $H_0(\theta_0)$ along the sieve $\{S_n\}$ relies on the hypo-convergence of $\hat{H}_n(\omega, \theta)$ to H_0 , with $\hat{H}_n(\omega, \theta) = H_n(\omega, \theta)$ if $\theta \in S_n$ and $-\infty$ otherwise; it will provide also convergence of $\sup_{S_n \cap U^c} H_n(\omega, \theta)$ to $\sup_{U^c} H_0(\theta)$, hence posterior consistency if the prior $\Pi(S_n)$ is exponentially small. Some aspects concerning hypo-convergence in the context of sieves have been already touched in Dong and Wets (2000) in the form of penalized maximum likelihood, a sort of "dual" of the method of sieves.

Some additional comments are necessary on a.s. hypo-convergence, tightness and suitable compact embedding.

Hypo-convergence, when holding, highlights that possible sources of inconsistency are on the tails of $H_n(\omega, \cdot)$ identified with complements of compact subsets of Θ . Small prior probability on the tails can compensate the possible explosive behaviour of these and accounts for those situations, as in Bahadur's example, where one has posterior consistency but not consistency of MLE's. On the other hand, a.s. tightness of the posteriors is a necessary condition; when holding, a.s. hypo-convergence on compacts is sufficient for posterior consistency. This delineates an approach based on conditions which jointly provide a.s. tightness and hypo-convergence. A.s. hypo-convergence on compacts is not difficult to obtain; in general tightness could reveal hard to verify but, since it is unavoidable, I am still convinced that it remains an interesting direction to work on; just as minor hint a possibility is offered by a revisitation, based on the prior Π , of the tail conditions of the type considered in Huber (1967).

The fact that inconsistency seats on complements of compact sets and the nature of inconsistency as highlighted in Freedman (1963) suggest an approach to the consistency problem based on a “suitable” compact embedding where, in the topology adopted, a.s. hypo-convergence is easily obtained. The way to proceed has been illustrated for two classes of problems, one well-known in the context of posterior consistency and the other already considered in consistency of MLE’s. Proceeding with this approach towards more general situations and considering the fact that consistency depends on the topology adopted, raise the questions of what is the “suitable” compact embedding, if there exists any, which is the relation with the metric in which the original consistency problem has been formulated and finally, in the Bayesian context, how to specify the prior. These questions are open, but in restricting the attention to upper semicontinuous densities, as indicated in the conclusive remarks, the topology of hypo-convergence of the densities emerges as an appealing candidate.

ADDITIONAL REFERENCES IN THE DISCUSSION

- Doob, J. L. (1948). Application of the theory of martingales. *Coll. Int. du CNRS, Paris*, 22–28.
- Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27**, 143–158.
- Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10–26.