Math. Scand. Vol. 78, No. 2, 1996, (293-319)
Preprint Ser. No. 39, 1992, Math. Inst. Aarhus

Consistency of Statistical Models in the Stationary Case

GORAN PESKIR

Recently established asymptotic likelihood theory [6] concerns the study of the asymptotic behavior of various maximum estimators with the maximum likelihood estimator as the leading example. The main objective of this study is to perform the estimation of the true parameter value for the unknown distribution. The estimation is based on the observations that form a sequence of independent and identically distributed random variables. The main purpose of this paper is to investigate the case where the given observations form a stationary ergodic sequence of random variables. The first step in this direction is devoted to the foundation of the problem itself. Although dropping independence one causes some difficulties it turns out that using the results and methods established in [6] and [18] we reach our primary ambition in this direction by characterizing the sets of all accumulation and limit points of maximum estimators under consideration. Then we pass to the problem of consistency. We show that slightly stronger conditions than those established in [6] imply consistency in the present case. Moreover by using the uniform law of large numbers that is recently established in the stationary case in [19], as well as the methods developed for this purpose, we deduce new conditions implying consistency. These conditions are of eventual total boundedness in the mean type. In this way the problem of consistency of the given statistical models is naturally connected with the infinitely dimensional (uniform) law of large numbers. In particular all of the derived results apply to the maximum likelihood estimator based on the stationary ergodic observations.

1. Stationary ergodic observations

Let $\Pi = (\,\pi_\theta \mid \theta \in \Theta_0\,)$ be a statistical model with a sample space $(S, \mathcal{A})$, reference measure μ, and parameter set Θ_0. In other words $(S, \mathcal{A}, \mu)$ is a measure space and π_θ is a probability measure on $(S, \mathcal{A})$ satisfying $\pi_\theta \ll \mu$ for all $\theta \in \Theta_0$. Then the *likelihood function* and the *log-likelihood function* for Π are defined as follows:

$$f(s,\theta) = \frac{d\pi_\theta}{d\mu}(s) \quad \text{and} \quad h(s,\theta) = \log f(s,\theta) \tag{1}$$

for all $(s,\theta) \in S \times \Theta_0$. Suppose a random phenomenon is considered that has the *unknown distribution* π belonging to Π. Then there exists $\theta_0 \in \Theta_0$ such that $\pi = \pi_{\theta_0}$ and we may

AMS 1980 *subject classifications.* Primary 28D05, 60B12, 60G07, 60G10, 62A10, 62F10, 62F12. Secondary 28A20, 41A30, 41A65.
Key words and phrases: Asymptotic likelihood theory, statistical model, stochastic process, analytic (metric space), the likelihood function, the log-likelihood function, the information function, approximating (empirical, asymptotic) maximums, consistent, stationary, ergodic, k-ergodic, completely ergodic, the unilateral shift, the uniform law of large numbers, eventually total bounded in the mean (from above).
(Second edition) © goran@imf.au.dk

define the *information function* as follows:

$$I(\theta) = \int_S f(s,\theta_0)\, h(s,\theta)\ \mu(ds) \tag{2}$$

for all $\theta \in \Theta_0$ for which the integral exists. Put $\beta = \sup_{\theta\in\Theta_0} I(\theta)$ and denote $M = \{ \theta \in \Theta_0 \mid I(\theta) = \beta \}$. If the following condition is satisfied:

$$\int_S f(s,\theta_0)\ |\log f(s,\theta_0)\ |\ \mu(ds) < \infty \tag{3}$$

then by the information inequality, see [7], we may conclude:

$$M = \{ \theta \in \Theta_0 \mid \pi_\theta = \pi \} \ \text{ and } \ I(\theta) < I(\theta_0) = \beta \ \text{ for } \ \pi_\theta \neq \pi \ . \tag{4}$$

Hence we see that under condition (3) *the problem of determining the unknown distribution* π *is equivalent to the problem of determining the set* M of all maximum points of the information function I on Θ_0 . It is easily verified that (3) is satisfied as soon as we have:

$$f(\,\cdot\,,\theta_0) \in L^p(\mu) \cap L^q(\mu) \tag{5}$$

for some $0 < p < 1 < q < \infty$. In order to approach the set M we may suppose that the observations $X_1, X_2, \ldots$ of the random phenomenon under consideration are available. In other words $\{ X_j \mid j \geq 1 \}$ is a sequence of identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in $(S, \mathcal{A})$ and the common distribution law π . If $X_1, X_2, \ldots$ are independent, then by the law of large numbers we have:

$$\frac{1}{n}\sum_{j=1}^{n} h(X_j,\theta) \ \rightarrow I(\theta) \quad P\text{-a.s.} \tag{6}$$

as $n \rightarrow \infty$ for all $\theta \in \Theta_0$ for which the integral in (2) exists. Thus it may occur that under possibly additional hypotheses certain maximum points $\hat{\theta}_n$ of the map on the left side in (6) on Θ_0 approach the maximum points of the map on the right side in (6) on Θ_0 , that is the set M . This principle is, more or less explicitly, well-known and goes back to Fisher's fundamental papers [3] and [4]. A large number of studies have followed. We do not wish to review the history of this development, but will point out classical works [1], [8], [9], [11], [13], [21] and [22], as well as the surveys [15] and [16] where more detailed information with additional references can be found. For some new developments see [6], [12], [14], [18] and [20]. Let us however emphasize here that whenever (6) is valid, the sense of this problem as well as the interpretation of its solution do not require any additional assumption on the sequence $\{ X_j \mid j \geq 1 \}$. Thus we may and do assume that *the likelihood function for general (possibly dependent) sequences* $\{ X_j \mid j \geq 1 \}$ *is the same as in the independent case*. In other words the function $h(s,\theta) = \log f(s,\theta)$ may be chosen as the criterion function. We think that this fact is by itself of theoretical (and practical) interest.

In this paper we consider and investigate a new case of the same problem where $X_1, X_2, \ldots$ *are no longer independent*. However we shall assume that the probabilistic structure of the observations under consideration does not depend of the moment when we begin with the observation. In other words we shall assume that the sequence $\{ X_j \mid j \geq 1 \}$ is *stationary*, that is:

$$P\{ (X_1, X_2, \dots) \in B \} = P\{ (X_n, X_{n+1}, \dots) \in B \} \tag{7}$$

for all $n \geq 1$ and all $B \in \mathcal{A}^{\mathbf{N}}$. In addition we shall assume that the sequence $\{ X_j \mid j \geq 1 \}$ is *ergodic*. In other words whenever for some $B \in \mathcal{A}^{\mathbf{N}}$ we have:

$$\{ (X_1, X_2, \dots) \in B \} = \{ (X_n, X_{n+1}, \dots) \in B \} \tag{8}$$

being valid for all $n \geq 1$, then we may conclude:

$$P\{ (X_1, X_2, \dots) \in B \} \in \{0, 1\} . \tag{9}$$

Under these hypotheses we may conclude by using Birkhoff's pointwise ergodic theorem that (6) is valid with degenerated $I(\theta)$ for all $\theta \in \Theta_0$ for which the integral in (2) exists and is finite. It turns out that this fact is important for the proofs presented below, and moreover the statistical nature lying behind justifies it well. To conclude this introduction it is worthwhile to recall that every sequence of independent and identically distributed random variables is stationary and ergodic. Thus all of the derived results apply to this case as well. The next section constitutes the main body of the paper, while the section following it contains the main results. In the last section we present some typical examples, which motivated the theory exposed below, as well as indicate its applications.

2. Foundation of the problem of consistency

In this section we shall introduce and investigate the setting for the problem of consistency of statistical models in the stationary case. For this we shall follow the course developed for the case of independent and identically distributed random variables in the framework of the asymptotic likelihood theory [6]. However it turns out that there exist some crucial differences that should be emphasized and raised problems solved. It is instructive to observe that these problems are of the same type as those that appeared in the general reversed submartingale case for establishing consistency, see [18] with references. In spite of that we shall see that the assumptions of stationarity and ergodicity in the present case can provide close relatives of the facts deduced in [6] that are sufficient for establishing consistency under slightly stronger hypotheses than those used in [6]. Some more details in this direction will be presented later. Let us turn to the setting itself.

We consider a stationary ergodic sequence of random variables $\{ X_j \mid j \geq 1 \}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$ and a common distribution law π . The measurable space $(S, \mathcal{A})$ is called the *sample space*, and the probability measure π is called the *true distribution*. In addition we suppose that an analytic metric space Θ_0 is given and fixed, and by $\mathcal{B}_0 = \mathcal{B}(\Theta_0)$ we denote the Borel σ-algebra on Θ_0 . The space Θ_0 is called the *parameter space*. It may be embedded into a compact metric space (Θ, d) that will be called the *compactified parameter space*. Any function f defined on Θ_0 with values in $\bar{\mathbf{R}}$ will by definition be extended on Θ putting $f(\theta) = -\infty$ for all $\theta \in \Theta \setminus \Theta_0$. Moreover we suppose that an $\mathcal{A} \times \mathcal{B}_0$-measurable map $h(s, \theta)$ from $S \times \Theta_0$ into $\bar{\mathbf{R}}$ is given and fixed. It is called the *criterion function*. According to the above rule we have $h(s, \theta) = -\infty$ for all $s \in S$ and all

$\theta \in \Theta \setminus \Theta_0$. Moreover we shall assume that $h(\,\cdot\,,\theta)$ belongs to $L(\pi)$ for all $\theta \in \Theta$, where $L(\pi)$ denotes the set of all functions from S into $\bar{\mathbf{R}}$ for which the π-integral exists in $\bar{\mathbf{R}}$. Thus the *information function* may be defined as follows:

$$I(\theta) = \int_S h(s,\theta)\ \pi(ds)$$

for all $\theta \in \Theta$. Note that $I(\theta) = -\infty$ for all $\theta \in \Theta \setminus \Theta_0$. Let us introduce the set of all maximum points of I on Θ_0 , that is:

$$M = \{\ \theta \in \Theta\ \mid\ I(\theta) = \beta\ \}$$

where we define $\beta = \sup_{\theta \in \Theta_0} I(\theta)$. The problem under our consideration may be stated as follows: *Given* $\{\ X_j\ \mid\ j \geq 1\ \}$ *with unknown* π *estimate* M *!* The sense of this problem as well as the interpretation of its solution rely upon the *information inequality* that appears in the framework of the above setting where the criterion function $h(s,\theta)$ equals the log-likelihood function $\log f(s,\theta)$ for a statistical model $\Pi = (\ \pi_\theta \mid \theta \in \Theta_0\ )$, see section 1. We shall refer the reader to [6] for more details and information in this direction. Let us however emphasize once again that no assumption on the sequence $\{\ X_j\ \mid\ j \geq 1\ \}$ itself is used for this purpose. Thus we may and do assume that the likelihood function for general (possibly dependent) stationary sequences $\{\ X_j \mid j \geq 1\ \}$ is the same as in the independent case. In order to perform the estimation in our problem we shall follow [6] and define the *empirical information function* as follows:

$$h_n(\omega,\theta) = \frac{1}{n} \sum_{j=1}^{n} h(X_j(\omega),\theta)$$

for all $(\omega,\theta) \in \Omega \times \Theta$ and all $n \geq 1$ with the convention $+\infty - \infty = -\infty$. By using Birkoff's pointwise ergodic theorem we obtain:

$$I(\theta) = \lim_{n\to\infty} h_n(\,\cdot\,,\theta) \quad P\text{-a.s.} \tag{1}$$

for all $\theta \in \Theta_0$ for which $h(\,\cdot\,,\theta)$ belongs to $L^1(\pi)$. This fact is essential. It indicates that maximum points $\hat{\theta}_n(\omega)$ of $h_n(\omega,\cdot)$ on Θ_0 for $\omega \in \Omega$ and $n \geq 1$ may under certain additional hypotheses approach the set M of maximum points of I on Θ_0 . Our further work strongly relies upon this belief. In order to exploit the preceding conclusions on (1) we shall proceed by introducing several auxiliary functions associated to the empirical information function:

$$h_n^*(\omega,B) = \sup_{\theta \in B} h_n(\omega,\theta) \tag{2}$$

$$H_0^*(\omega,B) = \liminf_{n\to\infty}\ h_n^*(\omega,B) \quad , \quad H^*(\omega,B) = \limsup_{n\to\infty}\ h_n^*(\omega,B) \tag{3}$$

$$\bar{H}_0(\omega,B) = \inf_{G \in \mathcal{G}(\Theta), G \supset B} H_0^*(\omega,G) \quad , \quad \bar{H}(\omega,B) = \inf_{G \in \mathcal{G}(\Theta), G \supset B} H^*(\omega,G) \tag{4}$$

$$\bar{H}_0(\omega,\theta) = \inf_{r>0} H_0^*(\omega,b(\theta,r)) \quad , \quad \bar{H}(\omega,\theta) = \inf_{r>0} H^*(\omega,b(\theta,r)) \tag{5}$$

$$\eta(\theta) = \inf_{n \geq 1} E^* \bar{h}_n(\theta) \quad , \quad \eta^*(B) = \inf_{n \geq 1} E^* h_n^*(B) \tag{6}$$

whenever $\omega \in \Omega$, $\theta \in \Theta$, $B \subset \Theta$ and $n \geq 1$. Here $\bar{h}_n(\omega, \cdot\,) = \lim_{r \downarrow 0} h_n^*(\omega, b(\,\cdot\,, r))$ denotes the upper semicontinuous envelope of $h_n(\omega, \cdot)$ on Θ for $\omega \in \Omega$ and $n \geq 1$, while E^* denotes the upper P-integral. Also $\mathcal{G}(\Theta)$ denotes the family of all open sets in Θ . According to [6] the functions $\bar{H}_0(\omega, B)$ and $\bar{H}(\omega, B)$ are called the *outer maximal functions*, the functions $\bar{H}_0(\omega, \theta)$ and $\bar{H}(\omega, \theta)$ are called the *upper information functions*, and the functions $\eta(\theta)$ and $\eta^*(B)$ are called the *mean value information functions*. In order to transfer the results of interest from section 2 in [18] to the present case we shall put $\mathcal{H} = (\{\, h_n(\omega, \theta), \mathcal{S}_X^n \mid n \geq 1 \,\} \mid \theta \in \Theta_0)$, where $\mathcal{S}_X^n$ denotes the permutation invariant σ-algebra of order n based on $X = (X_1, X_2, \, \dots\,)$, see [7]. Then each $h_n(\,\cdot\,, \theta)$ is $\mathcal{S}_X^n$-measurable for $n \geq 1$, and moreover by our hypotheses $\mathcal{H}$ is measurable in the sense of [18], that is $h_n(\omega, \theta)$ is $\mathcal{S}_X^n \times \mathcal{B}_0$-measurable for $n \geq 1$. Let $\mathcal{A}(\Theta)$ denote the family of all analytic sets in Θ . If B belongs to $\mathcal{A}(\Theta)$, then by the projection theorem, see [6], the map $\omega \mapsto h_n^*(\omega, B)$ is $(\mathcal{S}_X^n)^P$-measurable. However under the present hypotheses on $X_1, X_2, \, \dots$ we can not conclude that the σ-algebra $\mathcal{S}_X^\infty = \cap_{n=1}^\infty \mathcal{S}_X^n$ is degenerated, that is $P(F) \in \{0, 1\}$ for all $F \in \mathcal{S}_X^\infty$, as it was possible by using the Hewitt-Savage 0-1 law in the case where $X_1, X_2, \, \dots$ are independent and identically distributed (exchangeable). Actually we have:

$$\mathcal{S}_X^\tau \subset \mathrm{T}_X^\infty \subset \mathcal{S}_X^\infty$$

where T_X^∞ denotes the well-known tail σ-algebra based on $X = (X_1, X_2, \, \dots\,)$ and $\mathcal{S}_X^\tau$ denotes the σ-algebra of all shift invariant sets based on $X = (X_1, X_2, \, \dots\,)$, that is $F \in \mathcal{S}_X^\tau$ if and only if $F = (X_n, X_{n+1}, \, \dots\,)^{-1}(B)$ for some fixed $B \in \mathcal{A}^{\mathbf{N}}$ and all $n \geq 1$. Here τ denotes the unilateral shift in $S^{\mathbf{N}}$, that is $\tau(s_1, s_2, s_3, \, \dots\,) = (s_2, s_3, \, \dots\,)$ for all $(s_1, s_2, \, \dots\,) \in \mathcal{S}^{\mathbf{N}}$. Thus we could say that $\mathcal{H}$ is *degenerated* relative to a given family $\mathcal{C}$ of subsets of Θ , if we have:

(7) $$H^*(\,\cdot\,, B) = \text{const. and } H_0^*(\,\cdot\,, B) = \text{const. } P\text{-a.s.}$$

for all $B \in \mathcal{C}$. Then it is easily verified that *proposition 2.1, proposition 2.2, corollary 2.3 and proposition 2.4 in [18] are valid in the present case*, provided that in their hypotheses the words "of reversed submartingales" and "measurable" are removed and that the word "degenerated" is replaced by the words "degenerated relative to" a decent family of subsets of Θ in the sense explained above. For instance, in order to formulate statement (6) in proposition 2.1 $\mathcal{H}$ should be degenerated relative to $\cup_{\,\theta \in \bar{\Theta}_0} \{\, b(\theta, r) \mid r \leq r_\theta \}$ with some $r_\theta > 0$. We shall leave the strict formulation of these statements and all of the remaining details to the reader. It is instructive to notice that the submartingale property of $\mathcal{H}$ in [18] is not used for this purpose. Moreover exactly in the same way *proposition 2.5, corollary 2.6 and remark 2.7 in [18] may be carried over to the present case*. Most of the details will be omitted. The final results that will be of use in the next considerations may be stated as follows from (11) to (15) below. We consider the following sequences of maximum functions:

(8) A sequence of functions $\{\, \hat{\theta}_n \mid n \geq 1 \,\}$ from Ω into Θ is called a *sequence of empirical maximums*, if there exist a function $q : \Omega \to \mathbf{N}$ and a P-null set $N \in \mathcal{F}$ satisfying:

(i) $\hat{\theta}_n(\omega) \in \Theta_0 \;\; , \;\; \forall n \geq q(\omega) \; , \; \forall \omega \in \Omega \setminus N$

(ii) $h_n(\omega, \hat{\theta}_n(\omega)) = h_n^*(\omega, \Theta_0) \;\; , \;\; \forall n \geq q(\omega) \; , \; \forall \omega \in \Omega \setminus N$

(9) A sequence of functions $\{ \hat{\theta}_n \mid n \geq 1 \}$ from Ω into Θ is called a *sequence of asymptotic maximums*, if there exist a function $q : \Omega \to \mathbf{N}$ and a P-null set $N \in \mathcal{F}$ satisfying:

(i) $\hat{\theta}_n(\omega) \in \Theta_0 \quad , \quad \forall n \geq q(\omega) \ , \ \forall \omega \in \Omega \setminus N$

(ii) $\liminf_{n\to\infty} \ h_n(\omega, \hat{\theta}_n(\omega)) \ \geq \ H_0^*(\omega, \Theta_0) \quad , \quad \forall \omega \in \Omega \setminus N$

(10) A sequence of functions $\{ \hat{\theta}_n \mid n \geq 1 \}$ from Ω into Θ is called a *sequence of approximating maximums*, if there exist a function $q : \Omega \to \mathbf{N}$ and a P-null set $N \in \mathcal{F}$ satisfying:

(i) $\hat{\theta}_n(\omega) \in \Theta_0 \quad , \quad \forall n \geq q(\omega) \ , \ \forall \omega \in \Omega \setminus N$

(ii) $\liminf_{n\to\infty} \ h_n(\omega, \hat{\theta}_n(\omega)) \ \geq \ \beta \quad , \quad \forall \omega \in \Omega \setminus N$

where we recall that $\beta = \sup_{\,\theta \in \Theta_0} I(\theta)$.

It is easily verified that every sequence of empirical maximums is a sequence of asymptotic maximums and that every sequence of asymptotic maximums is a sequence of approximating maximums. Moreover sequences of approximating and asymptotic maximums always exist. For more details see [6] and [18]. In addition we shall introduce the following sets:

$$\hat{M} = \{ \ \theta \in \bar{\Theta}_0 \ \mid \ \bar{H}(\omega, \theta) \geq \beta \ \ P\text{-}a.s. \ \}$$

$$\hat{L} = \{ \ \theta \in \bar{\Theta}_0 \ \mid \ \bar{H}_0(\omega, \theta) \geq \beta \ \ P\text{-}a.s. \ \}$$

$$M^* = \{ \ \theta \in \bar{\Theta}_0 \ \mid \ \bar{H}(\omega, \theta) \geq H^*(\omega, \Theta) \ \ P\text{-}a.s. \ \}$$

$$M_0^* = \{ \ \theta \in \bar{\Theta}_0 \ \mid \ \bar{H}(\omega, \theta) \geq H_0^*(\omega, \Theta) \ \ P\text{-}a.s. \ \}$$

$$L_0^* = \{ \ \theta \in \bar{\Theta}_0 \ \mid \ \bar{H}_0(\omega, \theta) \geq H_0^*(\omega, \Theta) \ \ P\text{-}a.s. \ \} \ .$$

Then we have:

(11) $\hat{M}$ is exactly the set of all possible accumulation points of all possible sequences of approximating maximums

(12) $\hat{L}$ is exactly the set of all possible limit points of all possible sequences of approximating maximums

(13) M_0^* is exactly the set of all possible accumulation points of all possible sequences of asymptotic maximums

(14) L_0^* is exactly the set of all possible limit points of all possible sequences of asymptotic maximums

(15) If $\{ \hat{\theta}_n \mid n \geq 1 \}$ is a sequence of empirical maximums, then there exists a P-null set $N \in \mathcal{F}$ such that $\mathcal{C}\{\hat{\theta}_n(\omega)\} \subset M_0^*$, $\mathcal{C}\{\hat{\theta}_n(\omega)\} \cap M^* \neq \emptyset$ and $\mathcal{L}\{\hat{\theta}_n(\omega)\} \subset L_0^*$ for all $\omega \in \Omega \setminus N$.

Clarify that $\mathcal{C}\{\hat{\theta}_n(\omega)\}$ and $\mathcal{L}\{\hat{\theta}_n(\omega)\}$ denote the sets of all accumulation and all limit points of the sequence $\{ \hat{\theta}_n(\omega) \mid n \geq 1 \}$ in Θ for $\omega \in \Omega$ respectively. Let us emphasize that statements (11)-(14) rely upon the existence of measurable approximating maximums theorem, see [17]. Moreover the proof of (11) in the present case contains a small detail that deserves to be mentioned. Namely in order to show that for any point $\theta \in \hat{M}$ there exists a sequence of approximating maximums $\{ \hat{\theta}_n \mid n \geq 1 \}$ satisfying $\theta \in \mathcal{C}\{\hat{\theta}_n(\omega)\}$ for all $\omega \in \Omega$ outside some P-null set $N \in \mathcal{F}$ we may strictly follow the proof of (1) in corollary 4.2 in [17] with $\beta(\omega)$ replaced by β . However in order to deduce the first inequality in this proof we may proceed as follows. First assume that $\beta < \infty$, then there exists a sequence $\{ \theta_m \mid m \geq 1 \}$ in Θ_0 such that $I(\theta_m) \geq \beta - 2^{-m}$ for all $m \geq 1$. Moreover for each $m \geq 1$ there exists a P-null set $N_m \in \mathcal{F}$ such that $h_n(\omega, \theta_m) \to I(\theta_m)$ for all $\omega \notin N_m$ as $n \to \infty$. Putting $N = \cup_{m=1}^{\infty} N_m$ we find:

$$\bar{H}_0(\omega, \bar{\Theta}_0) = \liminf_{n\to\infty} h_n^*(\omega, \Theta_0) \geq \liminf_{n\to\infty} h_n^*(\omega, \{ \theta_k \mid k \geq 1 \}) \geq I(\theta_m) \geq \beta - 2^{-m}$$

for all $\omega \notin N$ and all $m \geq 1$. Letting $m \to \infty$ we get:

$$\bar{H}_0(\omega, \bar{\Theta}_0) \geq \beta$$

for all $\omega \notin N$. The case $\beta = \infty$ may be handled in the same way. The rest of the proof of (1) in corollary 4.2 in [17] is the very same in the present case, and we shall omit the details. Concerning this remark it is instructive to notice that the constant β appears simultaneously in the definition of a sequence of approximating maximums as well as in the definition of $\hat{M}$, see (11).

These facts finish the establishment of the setting. Summarizing, we may conclude that (7) remains the central preliminary question that should be answered. Having this answer, all of the preceding results involving non-degenerated functions $H_0^*(\omega, B)$, $H^*(\omega, B)$, $\bar{H}_0(\omega, \theta)$, $\bar{H}(\omega, \theta)$, $\bar{H}_0(\omega, B)$ and $\bar{H}(\omega, B)$ may be formulated in the degenerated form under the additional hypothesis, that will be established in the next lemma. We shall leave the details in this direction to the reader. Moreover we shall see in the next section that the validity of (7) can be completely avoided in establishing consistency under slightly stronger hypotheses than those used in [6] and [18]. This approach differs from that one demonstrated in [6] and [18] where (7) is automatically satisfied for all analytic sets by the Hewitt-Savage 0-1 law. However in order to follow [6] in the present case, we shall turn out an answer to (7) in the next lemma. This process requires the following definitions.

Let $\{ \xi_j \mid j \geq 1 \}$ be a stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with values in a measurable space $(S, \mathcal{A})$, let $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$ denote the countable product of $(S, \mathcal{A})$, and let P_ξ denote the distribution law of $\xi = (\xi_1, \xi_2, \dots)$ as a map from Ω into $\mathcal{S}^{\mathbf{N}}$. Let τ denote the unilateral shift in $\mathcal{S}^{\mathbf{N}}$, that is $\tau(s_1, s_2, s_3, \dots) = (s_2, s_3, \dots)$ for all $(s_1, s_2, \dots) \in \mathcal{S}^{\mathbf{N}}$. Let $k \geq 1$ be given and fixed, then $\{ \xi_j \mid j \geq 1 \}$ is said to be k*-ergodic*, if τ^k is ergodic in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, P_\xi)$, that is $P_\xi(A) \in \{0,1\}$ whenever $A \in \mathcal{A}^{\mathbf{N}}$ and $\tau^{-k}(A) = A$. Moreover $\{ \xi_j \mid j \geq 1 \}$ is said to be *completely ergodic*, if it is k-ergodic for each $k \geq 1$. Note that $\{ \xi_j \mid j \geq 1 \}$ is ergodic, if and only if it is 1-ergodic. Moreover if $\{ \xi_j \mid j \geq 1 \}$ is $k \cdot l$-ergodic for some $k, l \geq 1$, then it is obviously k-ergodic. For more information in this direction we shall refer the reader to [19] (p.3-7). Now the lemma may be stated as follows.

Lemma 1.

Under the hypotheses of the above setting let us suppose that for given $B \in \mathcal{A}(\Theta)$ *there exists* $k \geq 1$ *such that:*

(1) $\sup_{\theta \in B} \mid h_k(\,\cdot\,,\theta) \mid < \infty$ *P-a.s.*

(2) $\{ X_j \mid j \geq 1 \}$ *is* k*-ergodic.*

Then there exist numbers $H^*(B)$ *and* $H_0^*(B)$ *in* $\bar{\mathbf{R}}$ *satisfying:*

(3) $\limsup_{n \to \infty} h_n^*(\,\cdot\,,B) = H^*(B)$ *P-a.s.*

(4) $\liminf_{n \to \infty} h_n^*(\,\cdot\,,B) = H_0^*(B)$ *P-a.s.*

Proof. Let us consider the map $g_n : S^{\mathbf{N}} \times \Theta \to \bar{\mathbf{R}}$ defined by:

$$g_n(s,\theta) = \frac{1}{n} \sum_{j=1}^{n} h(s_j,\theta)$$

for $s = (s_1, s_2, \dots) \in S^{\mathbf{N}}$, $\theta \in \Theta$ and $n \geq 1$. Let us denote:

$$g_n^*(s,B) = \sup_{\theta \in B} g_n(s,\theta)$$

for all $s \in S^{\mathbf{N}}$ with the given analytic set $B \in \mathcal{A}(\Theta)$. Let $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}})$ denote the countable product of $(S,\mathcal{A})$, and let P_X denote the distribution law of $X = (X_1, X_2, \dots)$ as a map from Ω into $S^{\mathbf{N}}$. Then by our hypotheses we see that each $g_n(s,\theta)$ is $\mathcal{A}^{\mathbf{N}} \times \mathcal{B}$-measurable and thus by the projection theorem, see [6], we may conclude that the map $s \mapsto g_n^*(s,B)$ is P_X-measurable for $n \geq 1$. Let τ denote the unilateral shift in $S^{\mathbf{N}}$, then by our hypotheses we have that τ^k is ergodic in $(S^{\mathbf{N}}, \mathcal{A}^{\mathbf{N}}, P_X)$. Therefore in order to establish that a P_X-measurable map f from $S^{\mathbf{N}}$ into $\bar{\mathbf{R}}$ is equal to a constant P_X-a.s., it is enough to show that this map is τ^k-invariant mod P_X , that is $f \circ \tau^k = f$ P_X-a.s., see [10] (p.5). Hence we may easily conclude that (3) will be established as soon as we have:

(5) $$\limsup_{n \to \infty} g_n^*(\tau^k(\,\cdot\,),B) = \limsup_{n \to \infty} g_n^*(\,\cdot\,,B) \quad P_X\text{-a.s.}$$

In order to deduce (5) let us note that we have:

$$g_n(\tau^k(s)\,,\theta) = \frac{1}{n} \sum_{j=1}^{n} h(s_{j+k},\theta) = \frac{n+k}{n} \cdot \frac{1}{n+k} \sum_{j=1}^{n+k} h(s_j,\theta) - \frac{1}{n} \sum_{j=1}^{k} h(s_j,\theta)$$

for all $s = (s_1, s_2\,, \dots) \in S^{\mathbf{N}}$, all $\theta \in \Theta$ and all $n \geq 1$. Taking supremum over all $\theta \in B$ and using (1) one can easily verify the validity of (5). This fact completes the proof of (3). Statement (4) may be proved in exactly the same way, and the proof is complete. □

Let us in addition remind that certain maximal inequalities were important to be established in the case of independent and identically distributed random variables in [6], as well as in the reversed submartingale case in [18]. We shall proceed and conclude this section by investigating these inequalities in the present case. It turns out that their close relatives may be established in the stationary case under our consideration. Moreover we shall see in the next section that these facts are sufficiently good for most of our purposes. Let us clarify that E^* denotes the P-upper integral. In particular if $X : \Omega \to \bar{\mathbf{R}}$ is P-measurable, then we have $E^*X = EX$ for $X \in L(P)$ and $E^*X = +\infty$ otherwise.

Lemma 2.

Under the hypotheses of the above setting let us suppose that $B \in \mathcal{A}(\Theta)$ *is given and fixed. Then for any given and fixed* $d \geq 1$ *we have:*

$$h_n^*(\omega, B) \leq \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n} \sum_{j=1}^{\sigma_n} \sup_{\theta \in B} \Big\{ \frac{1}{d} \sum_{i=1}^{d} h(X_{i+(j-1)\cdot d}(\omega)\, ,\theta) \Big\} + \frac{1}{n} \sum_{\sigma_n \cdot d < j \leq n} \sup_{\theta \in B} h(X_j(\omega)\, ,\theta) \tag{1}$$

$$E^* h_n^*(B) \leq \frac{\sigma_n \cdot d}{n}\, E^* h_d^*(B) + \frac{n - \sigma_n \cdot d}{n}\, E^* h_1^*(B) \tag{2}$$

being valid for all $\omega \in \Omega$ *and all* $n \geq 1$ *, where* $\sigma_n = [n/d]$ *denotes the integer part of* n/d *. In particular we may conclude:*

$$\limsup_{n \to \infty} E h_n^*(B) = \inf_{n \geq 1} E h_n^*(B) < \infty \tag{3}$$

whenever $E h_1^*(B) < \infty$ *, as well as:*

$$E^* h_{n \cdot d}^*(B) \leq E^* h_d^*(B) \tag{4}$$

for all $n \geq 1$ *. Moreover we have:*

$$h_{n+k}^*(\omega, B) \leq \frac{n}{n+k} \sup_{\theta \in B} \Big\{ \frac{1}{n} \sum_{j=1}^{n} h(X_{j+k}(\omega)\, ,\theta) \Big\} + \frac{k}{n+k} \sup_{\theta \in B} \Big\{ \frac{1}{k} \sum_{j=1}^{k} h(X_j(\omega)\, ,\theta) \Big\} \tag{5}$$

$$E^* h_{n+k}^*(B) \leq \frac{n}{n+k}\, E^* h_n^*(B) + \frac{k}{n+k}\, E^* h_k^*(B) \tag{6}$$

being valid for all $\omega \in \Omega$ *and all* $n, k \geq 1$ *.*

Proof. Since B is analytic, then by the projection theorem, see [6], the map $\omega \mapsto$

$\sup_{\theta \in B} h_n(\omega,\theta)$ is P-measurable. Therefore $E^* h_n^*(B) = E h_n^*(B)$ if $h_n^*(\,\cdot\,,B) \in L(P)$ and $E^* h_n^*(B) = +\infty$ otherwise. In order to establish (1) one may notice that we have:

$$(7) \qquad h_n(\omega,\theta) = \frac{1}{n}\sum_{j=1}^{n} h(X_j(\omega)\,,\theta) = \frac{1}{n}\sum_{j=1}^{\sigma_n \cdot d} h(X_j(\omega)\,,\theta) + \frac{1}{n}\sum_{\sigma_n \cdot d < j \leq n} h(X_j(\omega)\,,\theta)$$

$$= \frac{1}{n}\sum_{j=1}^{\sigma_n}\sum_{i=1}^{d} h(X_{i+(j-1)\cdot d}(\omega)\,,\theta) + \frac{1}{n}\sum_{\sigma_n \cdot d < j \leq n} h(X_j(\omega)\,,\theta)$$

$$= \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n}\sum_{j=1}^{\sigma_n} \left\{ \frac{1}{d}\sum_{i=1}^{d} h(X_{i+(j-1)\cdot d}(\omega)\,,\theta) \right\} + \frac{1}{n}\sum_{\sigma_n \cdot d < j \leq n} h(X_j(\omega)\,,\theta)$$

being valid for all $(\omega,\theta) \in \Omega \times \Theta$ and all $n \geq 1$. Hence (1) follows straightforward by taking supremum over all $\theta \in B$. Statements (2) and (4) follow by (1) and stationarity. Moreover statement (5) is obvious and statement (6) follows straightforward by (5) and stationarity in the same manner. Finally if $E h_1^*(B) < \infty$, then letting $n \to \infty$ in (2) we may conclude:

$$\limsup_{n \to \infty} E h_n^*(B) \leq E h_d^*(B)$$

since $(\sigma_n \cdot d)/n \to 1$ as $n \to \infty$. Taking infimum over all $d \geq 1$ we get (3). These facts complete the proof. □

3. Consistency of statistical models in the stationary case

In this section we shall introduce and investigate the concept of consistency of statistical models in the stationary case. Throughout the whole section we work within the setting that is introduced in the preceding section. We begin by introducing the definition of consistency. The criterion function $h(s,\theta)$ is said to be *π-consistent* on a given subset Γ of Θ , if for every sequence of approximating maximums $\{\ \hat{\theta}_n \mid n \geq 1\ \}$ there exists a P-null set $N \in \mathcal{F}$ such that $\mathcal{C}\{\hat{\theta}_n(\omega)\} \cap \Gamma \subset M$, for all $\omega \in \Omega \setminus N$. The criterion function $h(s,\theta)$ is said to be *π-consistent*, if it is π-consistent on Θ . Thus $h(s,\theta)$ is π-consistent on Γ if and only if *every accumulation point of any sequence of approximating maximums that belongs to Γ is a maximum point of the information function I on Θ_0* . By (2.11) we see that the following statements are equivalent:

(1) $h(s,\theta)$ is π-consistent on Γ

(2) $h(s,\theta)$ is π-consistent on $\Gamma \cap (\hat{M} \setminus M)$

(3) $\Gamma \cap \hat{M} \subset M$

(4) $\bar{H}(\omega,\theta) < \beta$ for all $\omega \in \Omega$ that belong to some $F_\theta \in \mathcal{F}$ satisfying $P(F_\theta) > 0$ whenever $\theta \in \Gamma \setminus M$.

Moreover if $\{\ \hat{\theta}_n \mid n \geq 1\ \}$ is a Γ-tight sequence of approximating maximums, that is $\mathcal{C}\{\hat{\theta}_n(\omega)\} \subset \Gamma$ for all $\omega \in \Omega$ outside some P-null set $N \in \mathcal{F}$, and $h(s,\theta)$ is π-consistent on Γ , then we have:

(5) $\mathcal{C}\{\hat{\theta}_n(\omega)\} \subset M$

(6) $\lim_{n\to\infty} d(\hat{\theta}_n(\omega), M) = 0$

for all $\omega \in \Omega$ outside some P-null set $N \in \mathcal{F}$. Thus in this case any sequence of approximating (empirical or asymptotic) maximums converges to the set of all maximum points of the information function I on Θ_0 . It is instructive to observe that we always have $M \subset \hat{L} \subset \hat{M}$. Therefore if $\hat{M} \cap \Gamma \subset M$, then $M \cap \Gamma = \hat{L} \cap \Gamma$. In other words if h is π-consistent on Γ , then every point $\theta \in M \cap \Gamma$ may be reached as the limit point of a sequence of approximating maximums. Our next aim is to obtain conditions for π-consistency of the criterion function. Let us for this introduce the *set of all* L^1*-dominated points* of $h(s,\theta)$ as follows:

$$\Theta_d = \Big\{ \ \theta \in \Theta \ \Big| \ \exists r > 0 \ \text{ such that } \int_S \sup_{\xi \in b(\theta,r)} h(s,\xi)\ \pi(ds) < \infty \ \Big\} \ .$$

Note that Θ_d is an open subset of Θ . Furthermore let us introduce the *set of all upper semicontinuity points of* $h(s,\theta)$ as follows:

$$\Theta_u = \Big\{ \ \theta \in \Theta \ \Big| \ \exists k \geq 1 \ \text{ such that } \ h_n(\omega,\,\cdot\,) \ \text{ is } \ P\text{-a.s.} \\ \text{upper semicontinuous at } \theta \ \text{ for all } \ n \geq k \ \Big\} \ .$$

Note if the map $h(s,\,\cdot\,)$ is π-a.s. upper semicontinuous at a given point $\theta \in \Theta$, then θ belongs to Θ_u . Also note that the present definitions slightly differ of those given in [6]. Finally, let us introduce the *set of all non-trivial points of* $h(s,\theta)$ as follows:

$$\Theta_f = \Big\{ \ \theta \in \Theta \ \Big| \ \sup_{\xi \in b(\theta,r)} I(\xi) > -\infty \ \text{ for all } \ r > 0 \ \Big\} \ .$$

Note that $\Theta_f \subset \bar{\Theta}_0$. The next theorem offers conditions for π-consistency of the criterion function.

Theorem 1. (**Consistency in the stationary case**)
Under the hypotheses of the setting of section 2 suppose that Γ *is a subset of* Θ *. Then we have:*

(1) *If* $\beta = -\infty$ *, then* $h(s,\theta)$ *is* π*-consistent on* Γ *if and only if* $\Gamma \subset \Theta_0 \cup (\Theta \setminus \bar{\Theta}_0)$

(2) *If* $\beta > -\infty$ *, then* $h(s,\theta)$ *is* π*-consistent on* Γ *if and only if* $\Gamma \subset M \cup (\Theta \setminus \hat{M}) \cup (\Theta_u \cap \Theta_d \cap \Theta_f)$

(3) *If* $h(s,\theta)$ *is* π*-consistent on* Γ *and* $\Gamma \cap M = \{\theta_0\}$ *, then* $\hat{\theta}_n \to \theta_0$ *P-a.s. as* $n \to \infty$ *for every* Γ*-tight sequence of approximating maximums.*

Proof. In the case of (1) we have $M = \Theta_0$ and $\hat{M} = \bar{\Theta}_0$. Thus (1) follows by (3) above. Statement (3) is a straightforward consequence of the definition of consistency. In order to complete the proof it remains to establish (2). For this first suppose that $h(s,\theta)$ is π-consistent on Γ .

Then by (3) above we see that $\Gamma \subset M \cup (\Theta \setminus \hat{M})$ and the first part of (2) is complete. Conversely suppose that $\Gamma \subset M \cup (\Theta \setminus \hat{M}) \cup (\Theta_u \cap \Theta_d \cap \Theta_f)$. Then $\Gamma \cap \hat{M} \subset M \cup (\Theta_u \cap \Theta_d \cap \Theta_f)$ and therefore $(\Gamma \cap \hat{M}) \setminus M \subset \Theta_u \cap \Theta_d \cap \Theta_f$. Hence by (3) above we see that the proof of (2) will be completed as soon as we have:

$$((\Gamma \cap \hat{M}) \setminus M) \cap (\Theta_u \cap \Theta_d \cap \Theta_f) = \emptyset \ . \tag{4}$$

We shall establish this fact by showing that:

$$\bar{H}(\,\cdot\,,\theta) = I(\theta) \quad P\text{-a.s.} \tag{5}$$

for all $\theta \in \Theta_u \cap \Theta_d \cap \Theta_f$. So let $\theta \in \Theta_u \cap \Theta_d \cap \Theta_f$ be a given point, then obviously $\bar{H}(\,\cdot\,,\theta) \geq I(\theta)$ P-a.s. In order to prove the converse inequality we may proceed as follows. Since θ belongs to Θ_d , then there exists $r > 0$ such that:

$$\int_S \sup_{\xi \in b(\theta,r)} h(s,\xi)\ \pi(ds) < \infty \ . \tag{6}$$

Hence by (4) in lemma 2.2 we easily find by using the monotone convergence theorem and the fact that θ belongs to Θ_u that we have:

$$\eta(\theta) = \inf_{n \geq 1} E\bar{h}_n(\theta) = \inf_{n \geq k} E\bar{h}_n(\theta) = I(\theta) \tag{7}$$

being valid for all $k \geq 1$. Let us in addition choose $k_\theta \geq 1$ large enough to satisfy $2^{-k_\theta} \leq r$, and let $k \geq k_\theta$ be given and fixed. Put $B_k = b(\theta, 2^{-k})$ and $h^*(s, B_k) = \sup_{\xi \in B_k} h(s,\xi)$ for all $s \in S$. Then by (6) and the fact that $\theta \in \Theta_f$ we may conclude:

$$-\infty < \int_S h^*(s\,,B_k)\ \pi(ds)\ < +\infty \ . \tag{8}$$

Let $d \geq 1$ be given and fixed, and let us define:

$$Y_j^*(\omega) = \sup_{\xi \in B_k} \Big(\frac{1}{d} \sum_{i=1}^{d} h(X_{i+(j-1)\cdot d}(\omega)\,,\xi) \Big)$$

for all $\omega \in \Omega$ and all $j \geq 1$. Then by the projection theorem, see [6], each Y_j^* is P-measurable for $j \geq 1$. Moreover by (2.3) and (2.4) in [19] we may easily conclude that the sequence $\{ Y_j^* \mid j \geq 1 \}$ is stationary and ergodic. By (1) in lemma 2.2 we have:

$$h_n^*(\omega, B_k) \leq \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n} \sum_{j=1}^{\sigma_n} Y_j^*(\omega) + \frac{1}{n} \sum_{\sigma_n \cdot d < j \leq n} h^*(X_j(\omega)\,, B_k) \tag{9}$$

for all $\omega \in \Omega$ and all $n \geq 1$, where $\sigma_n = [n/d]$ denotes the integer part of n/d . By (8) we easily find that $Y_1^* \in L^1(P)$, and therefore by Birkhoff's pointwise ergodic theorem we may conclude:

(10) $$\frac{1}{n}\sum_{j=1}^{n} Y_j^* \to EY_1^* = Eh_d^*(B_k) \quad P\text{-a.s.}$$

as $n \to \infty$. By the same argument and the fact that $(\sigma_n \cdot d)/n \to 1$ as $n \to \infty$ we have:

(11) $$\frac{1}{n}\sum_{\sigma_n \cdot d < j \le n} h^*(X_j, B_k) = $$

$$= \frac{1}{n}\sum_{j=1}^{n} h^*(X_j, B_k) - \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n \cdot d} \sum_{j=1}^{\sigma_n \cdot d} h^*(X_j, B_k) \to 0 \quad P\text{-a.s.}$$

as $n \to \infty$. Now by (9), (10) and (11) we may deduce:

(12) $$\limsup_{n\to\infty} h_n^*(\,\cdot\,, B_k) \le Eh_d^*(B_k) \quad P\text{-a.s.}$$

for all $d \ge 1$. Taking infimum over all $d \ge 1$ we get:

$$\limsup_{n\to\infty} h_n^*(\,\cdot\,, B_k) \le \inf_{d\ge 1} Eh_d^*(B_k) \quad P\text{-a.s.}$$

Letting $k \to \infty$ and using the monotone convergence theorem we may conclude:

(13) $$\bar{H}(\,\cdot\,,\theta) \le \inf_{k\ge 1}\inf_{d\ge 1} Eh_d^*(B_k) = \inf_{d\ge 1}\inf_{k\ge 1} Eh_d^*(B_k)$$

$$= \inf_{d\ge 1} E\bar{h}_d(\theta) = \eta(\theta) = I(\theta) \quad P\text{-a.s.}$$

Thus (5) is proved and the proof is complete. □

Remark 2.

In exactly the same way as for (12) in the preceding proof we may conclude:

(1) $$H^*(\,\cdot\,, B) \le \inf_{n\ge 1} Eh_n^*(B) \quad P\text{-a.s.}$$

whenever $B \in \mathcal{A}(\Theta)$ with $\sup_{\xi\in B} h(\,\cdot\,,\xi) \in L^1(\pi)$. Moreover if $\sup_{\xi\in B} |\, h(\,\cdot\,,\xi)\,| \in L^1(\pi)$, then by (3) in lemma 2.1, Fatou's lemma, and (3) in lemma 2.2 we easily get:

(2) $$\sup_{\theta\in B} I(\theta) \le H^*(B) = \inf_{n\ge 1} Eh_n^*(B) = \limsup_{n\to\infty} Eh_n^*(B) \ .$$

These facts and the method presented in the proof of theorem 1 may be used in an attempt to obtain under additional hypotheses the results in the present case that correspond to those established in proposition 3.1 and proposition 3.2 in [18] with the origin in [6]. However they are irrelevant for our further purposes and we shall resist of doing it here. We shall turn to the question when the first inequality in (2) becomes an equality.

The main applicability of theorem 1 may be in essence presented as follows. Suppose that under the hypotheses of the setting of section 2 we have a subset Γ of Θ satisfying:

(7) $\sup_{\xi \in b(\theta,r)} I(\xi) > -\infty$ for all $r > 0$ and all $\theta \in \Gamma$

(8) $\int_S \sup_{\xi \in b(\theta,r_\theta)} h(s,\xi)\ \pi(ds) < \infty$ for all $\theta \in \Gamma$ with some $r_\theta > 0$

(9) $h(s,\cdot\,)$ is upper semicontinuous at θ for all $s \in S$ outside some π-null set $N_\theta \in \mathcal{A}$, being valid for all $\theta \in \Gamma$.

Then every accumulation point of any sequence of approximating (empirical or asymptotic) maximums that belongs to Γ is a maximum point of I on Θ_0. This fact completes our main purpose. Conditions (7)–(9) are in most cases easily verified.

A close look at the proof of theorem 1 enables one to verify that under (7)-(9) we have $I(\theta) = \eta(\theta) = \bar{H}(\theta)$ for all $\theta \in \Gamma$, see (13) in this proof. Hence we easily find that $I(\theta) = \bar{I}(\theta)$ for all $\theta \in \Gamma$, where $\bar{I}$ denotes the upper semicontinuous envelope of I. In other words I is upper semicontinuous on Γ. Thus in order to find some new conditions implying consistency it is not very restrictive to assume that I is upper semicontinuous on the candidate subset of Θ. Having this fact in mind we shall now present another way toward consistency that relies upon the uniform law of large numbers recently established in the stationary case in [19]. This approach is already applied in the reversed submartingale case, see [18]. In the present case we may proceed as follows. By (2.11) we know that the set of all accumulation points of all possible sequences of approximating maximums equals $\hat{M} = \{\ \theta \in \bar{\Theta}_0 \mid \bar{H}(\,\cdot\,,\theta) \geq \beta \ \ P\text{-a.s.}\ \}$. Moreover we have:

$$\bar{H}(\omega,\theta) = \inf_{r>0} \limsup_{n\to\infty} h_n^*(\omega, b(\theta,r))$$

for all $\omega \in \Omega$ and all $\theta \in \Theta$. Hence we see that conditions implying:

$$\limsup_{n\to\infty} h_n^*(\,\cdot\,, b(\theta,r_m)) = \sup_{\xi \in b(\theta,r_m)} I(\xi) \quad P\text{-a.s.}$$

for some sequence $\{\, r_m \mid m \geq 1 \}$ satisfying $r_m \downarrow 0$ as $m \to \infty$ will have for a consequence:

$$\bar{H}(\theta) = \bar{I}(\theta)$$

where $\theta \in \Theta$ is a given point. Since the set:

$$\tilde{M} = \{\ \theta \in \bar{\Theta}_0 \mid \bar{I}(\theta) \geq \beta\ \}$$

is closed and contains M, then we have $\bar{M} \subset \tilde{M}$. Conversely if $\theta \in \tilde{M}$, then there exists a sequence $\{\ \theta_n \mid n \geq 1\ \}$ in Θ satisfying:

$$d(\theta_n,\theta) \leq 2^{-n} \quad \text{and} \quad I(\theta_n) \geq \beta - 2^{-n}$$

for all $n \geq 1$. Thus if $\theta_n \to \theta$ with $I(\theta_n) \to \beta$ implies $I(\theta) = \beta$ for every $\theta \in \tilde{M}$, then we have $\tilde{M} = M = \bar{M}$. This will be for instance true if I has the closed graph $gr(I) = \{ (\theta, I(\theta)) \mid \theta \in \Theta_0 \}$, or if I is upper semicontinuous on $\tilde{M}$. It is instructive to notice that I is always upper semicontinuous on M , as well as that for every $\theta \in \tilde{M}$ we actually have $\bar{I}(\theta) = \beta$. The next result relies upon this idea and the uniform law of large numbers in the stationary case [19]. It requires the following definitions. Let T be a subset of Θ , then $\Gamma(T)$ denotes the family of all finite covers of T . Recall that a *finite cover* of T is any family of non-empty subsets $A_1, \dots, A_n$ of T satisfying $\cup_{j=1}^n A_j = T$. We shall in what follows say that a finite cover $\gamma = \{ A_1, \dots, A_n \}$ of T is *analytic*, if every A_j is analytic in Θ for $j = 1, \dots, n$. Note that in this case T must be also analytic in Θ . The criterion function $h(s, \theta)$ will be called *eventually total bounded in π-mean* on T , if the following condition is satisfied:

(10) For each $\varepsilon > 0$ there exists an analytic cover $\gamma_\varepsilon \in \Gamma(T)$ such that:

$$\inf_{n \geq 1} \int_\Omega \sup_{\theta', \theta'' \in A} \mid h_n(\omega, \theta') - h_n(\omega, \theta'') \mid P(d\omega) < \varepsilon$$

for all $A \in \gamma_\varepsilon$.

Note that the present definition of eventual total boundedness in the mean slightly differs from the one used in [19] since we require that the elements of γ_ε in (10) are analytic sets. This is done in order to avoid some technical difficulties that may appear if the integrand in (10) is not measurable, see theorem 3.1, proposition 3.2 and remark 3.3 in [19]. Note that this can not happen under our present hypothesis on γ_ε in (10) since by the projection theorem, see [6], we may easily conclude that the integrand in (10) is P-measurable. Another possibility to avoid those difficulties is to assume that the map $X = (X_1, X_2, \dots)$ is P_X-perfect as a map from Ω into $S^{\mathbf{N}}$, see [19]. This is for instance true for the canonical representation of X , see section 1 in [19]. We shall leave the formulation and verification of the next results under this hypothesis instead of the requirement that γ_ε in (10) is analytic to the reader. Also we shall refer the reader to theorem 3.7, theorem 3.9 and theorem 3.10 in [19] for equivalent formulations of (10) and more information in this direction. However let us emphasize here that condition (10) is in all decent cases equivalent to the uniform law of large numbers being valid on T , see corollary 3.5 and corollary 3.8 in [19]. We turn to the next result itself.

Theorem 3.

Under the hypotheses of the setting of section 2 suppose that Γ is a subset of Θ such that for each $\theta \in \Gamma \cap \hat{M}$ there exists $r_\theta > 0$ satisfying:

(1) $$\int_S \sup_{\xi \in b(\theta, r_\theta)} \mid h(s, \xi) \mid \; \pi(ds) < \infty$$

(2) *h is eventually total bounded in π-mean on $b(\theta, r_\theta)$.*

Then we have:

(3) $$\sup_{\xi \in b(\theta, r_\theta)} \mid h_n(\,\cdot\,, \xi) - I(\xi) \mid \to 0 \;\; P\text{-a.s. and in } L^1(P) \text{ as } n \to \infty \text{, for all } \theta \in \Gamma \cap \hat{M}$$

(4) $\bar{H}(\theta) = \bar{I}(\theta)$ *for all* $\theta \in \Gamma \cap \hat{M}$

(5) $\Gamma \cap \hat{M} = \{ \theta \in \Gamma \mid \bar{I}(\theta) = \beta \}$.

If, in addition, the function I *satisfies any of the following two equivalent conditions:*

(6) I *is upper semicontinuous on* $\Gamma \cap \hat{M}$

(7) $cl(gr(I)) \cap ((\Gamma \cap \hat{M}) \times \{\beta\}) \subset gr(I)$, *or equivalently if* $\theta_n \to \theta$ *and* $I(\theta_n) \to \beta$ *with* $\theta \in \Gamma \cap \hat{M}$, *then* $I(\theta) = \beta$

then $h(s,\theta)$ *is* π*-consistent on* Γ .

Proof. Let $\theta \in \Gamma \cap \hat{M}$ be given and fixed and let $r_\theta > 0$ be chosen in such a way that (1) and (2) are satisfied. Then (3) follows straightforward by theorem 3.1, proposition 3.2 and remark 3.3 in [19]. In particular we get:

(8) $$\sup_{\xi \in b(\theta,r)} h_n(\,\cdot\,,\xi) \;\to\; \sup_{\xi \in b(\theta,r)} I(\xi) \quad P\text{-a.s.}$$

as $n \to \infty$ for all $0 < r \leq r_\theta$. In other words we have:

(9) $$H^*(b(\theta,r)) \;=\; \sup_{\xi \in b(\theta,r)} I(\xi)$$

for all $0 < r \leq r_\theta$. Thus taking infimum over all $0 < r \leq r_\theta$ we get (4). Statement (5) is an easy consequence of (4). Moreover since θ belongs to $\hat{M}$, then there exists a sequence $\{ \theta_n \mid n \geq 1 \}$ in Θ such that $d(\theta_n,\theta) \leq 2^{-n}$ and $I(\theta_n) \geq \beta - 2^{-n}$ for all $n \geq 1$. Hence we see that $\theta_n \to \theta$ and $I(\theta_n) \to \beta$. Thus if (7) is satisfied, then we get $I(\theta) = \beta$. In this way we may conclude that $\Gamma \cap \hat{M} \subset M$ and thus $h(s,\theta)$ is π-consistent on Γ by (3) above. Moreover it is easily verified that in the presence of (4) statement (6) is equivalent to statement (7). These facts complete the proof. □

The preceding result is in a way a straightforward consequence of the uniform law of large numbers in the stationary case. However let us notice that in this way we have obtained even more than it is needed. Namely we have established (8) in the proof of theorem 3 by using the validity of (3) in the same theorem. Our next aim is to show that applying the methods used in the proof of the uniform law of large numbers [5] and [19], a more direct approach to (8) could be done. For that reason we shall relax the condition of eventual total boundedness in the mean as follows. The criterion function $h(s,\theta)$ will be called *eventually total bounded in* π*-mean from above* on a given subset T of Θ , if the following condition is satisfied:

(11) For each $\varepsilon > 0$ there exist an analytic cover $\gamma_\varepsilon = \{ A_1, \ldots, A_{m_\varepsilon} \} \in \Gamma(T)$ and points $\theta_1 \in A_1, \ldots, \theta_{m_\varepsilon} \in A_{m_\varepsilon}$ such that:

(i) $$\inf_{n \geq 1} \int_\Omega \sup_{\theta \in A_j} \big(h_n(\omega,\theta) - h_n(\omega,\theta_j) \big)^+ \; P(d\omega) < \varepsilon$$

(ii) $\sup_{\xi \in A_j} \, (\, I(\theta_j) - I(\xi) \,)^+ < \varepsilon$

for all $j = 1, \ldots, m_\varepsilon$.

It is easily verified that (10) implies (11). Moreover it turns out that condition (1) in theorem 3 may be slightly weakened to the form that already appeared in theorem 1.

Theorem 4.

Under the hypotheses of the setting of section 2 suppose that Γ *is a subset of* Θ *such that for each* $\theta \in \Gamma \cap \hat{M}$ *there exists* $r_\theta > 0$ *satisfying:*

(1) $$\int_S \sup_{\xi \in b(\theta, r_\theta)} h(s,\xi) \, \pi(ds) < \infty$$

(2) h *is eventually total bounded in* π*-mean from above on* $b(\theta, r_\theta)$.

Then we have:

(3) $\sup_{\xi \in b(\theta, r_\theta)} \, (\, h_n(\,\cdot\,,\xi) - I(\xi) \,)^+ \to 0$ *P-a.s. and in* $L^1(P)$ *as* $n \to \infty$ *, for all* $\theta \in \Gamma \cap \hat{M}$

(4) $\bar{H}(\theta) = \bar{I}(\theta)$ *for all* $\theta \in \Gamma \cap \hat{M}$

(5) $\Gamma \cap \hat{M} = \{ \, \theta \in \Gamma \mid \bar{I}(\theta) = \beta \, \}$.

If, in addition, the function I *satisfies any of the following two equivalent conditions:*

(6) I *is upper semicontinuous on* $\Gamma \cap \hat{M}$

(7) $cl(gr(I)) \cap ((\Gamma \cap \hat{M}) \times \{\beta\}) \subset gr(I)$ *, or equivalently if* $\theta_n \to \theta$ *and* $I(\theta_n) \to \beta$ *with* $\theta \in \Gamma \cap \hat{M}$ *, then* $I(\theta) = \beta$

then $h(s,\theta)$ *is* π*-consistent on* Γ .

Proof. Let $\theta \in \Gamma \cap \hat{M}$ be given and fixed and let $r_\theta > 0$ be chosen in such a way that (1) and (2) are satisfied.

(3): Let us denote $B = b(\theta, r_\theta)$ and let $\varepsilon > 0$ be given. Then by (2) there exist an analytic cover $\gamma_\varepsilon = \{ \, A_1, \ldots, A_{m_\varepsilon} \} \in \Gamma(B)$ and points $\theta_1 \in A_1, \ldots, \theta_{m_\varepsilon} \in A_{m_\varepsilon}$ such that:

(8) $$\inf_{n \geq 1} \int_\Omega \sup_{\theta \in A_j} \, (\, h_n(\omega,\theta) - h_n(\omega,\theta_j) \,)^+ \; P(d\omega) < \varepsilon$$

(9) $$\sup_{\xi \in A_j} \, (\, I(\theta_j) - I(\xi) \,)^+ < \varepsilon$$

for all $j = 1, \ldots, m_\varepsilon$. Moreover we have:

$$\sup_{\xi \in B} \, (h_n(\omega,\xi) - I(\xi))^+ = \max_{1 \leq j \leq m_\varepsilon} \, \sup_{\xi \in A_j} \, (h_n(\omega,\xi) - I(\xi))^+$$

$$\leq \max_{1 \leq j \leq m_\varepsilon} \sup_{\xi \in A_j} (h_n(\omega,\xi) - h_n(\omega,\theta_j))^+ + \max_{1 \leq j \leq m_\varepsilon} (h_n(\omega,\theta_j) - I(\theta_j))^+$$
$$+ \max_{1 \leq j \leq m_\varepsilon} \sup_{\xi \in A_j} (I(\theta_j) - I(\xi))^+$$

for all $\omega \in \Omega$. By Birkhoff's pointwise ergodic theorem and (9) we easily get:

(10)
$$\limsup_{n\to\infty} \sup_{\xi \in B} (h_n(\omega,\xi) - I(\xi))^+ \leq$$
$$\leq \max_{1 \leq j \leq m_\varepsilon} \limsup_{n\to\infty} \sup_{\xi \in A_j} (h_n(\omega,\xi) - h_n(\omega,\theta_j))^+ + \varepsilon \quad P\text{-a.s.}$$

Replacing $h(s,\xi)$ by $g(s,\xi) = h(s,\xi) - h(s,\theta_j)$ for $\xi \in A_j$ with $j = 1,\dots,m_\varepsilon$ we get a new criterion function satisfying (1) with h replaced by g . Thus by (8) and (10) we may easily conclude that the proof of (3) will be completed as soon as we show that:

(11)
$$\limsup_{n\to\infty} \sup_{\xi \in A} (h_n(\,\cdot\,,\xi))^+ \leq \inf_{n \geq 1} \int_\Omega \sup_{\xi \in A} (h_n(\omega,\xi))^+ \, P(d\omega) \quad P\text{-a.s.}$$

for any analytic set A in Θ satisfying:

(12)
$$\int_S \sup_{\xi \in A} h(s,\xi) \; \pi(ds) < \infty \; .$$

Let $d \geq 1$ be given and fixed. Then by (7) in the proof of lemma 2.2 we may easily conclude:

(13)
$$\sup_{\xi \in A} (h_n(\omega,\xi))^+ \leq \frac{\sigma_n \cdot d}{n} \cdot \frac{1}{\sigma_n} \sum_{j=1}^{\sigma_n} Y_j^*(\omega) \; + \; \frac{1}{n} \sum_{\sigma_n \cdot d < j \leq n} \sup_{\xi \in A} (h(X_j(\omega)\,,\xi))^+$$

being valid for all $n \geq d$ and all $\omega \in \Omega$ with $\sigma_n = [\, n/d \,]$, and where:

$$Y_j^*(\omega) \; = \; \sup_{\xi \in A} \; \Big(\frac{1}{d} \sum_{i=1}^{d} h(X_{i+(j-1)\cdot d}(\omega)\,,\xi) \; \Big)^+$$

for $\omega \in \Omega$, $\xi \in A$ and $j \geq 1$. In the same way as for (10) in the proof of theorem 1 we find:

(14)
$$\frac{1}{n} \sum_{j=1}^{n} Y_j^* \to EY_1^* = \int_\Omega \sup_{\xi \in A} (h_d(\omega,\xi))^+ P(d\omega) \quad P\text{-a.s.}$$

as well as for (11) in the proof of theorem 1 by using (12) that we have:

(15)
$$\frac{1}{n} \sum_{\sigma_n \cdot d < j \leq n} \sup_{\xi \in A} (h(X_j,\xi))^+ \to 0 \quad P\text{-a.s.}$$

as $n \to \infty$. Thus letting $n \to \infty$ in (13) and using (14) and (15) we get:

$$\limsup_{n\to\infty} \sup_{\xi \in A} (h_n(\,\cdot\,,\xi))^+ \leq \int_\Omega \sup_{\xi \in A} (h_d(\omega,\xi))^+ \, P(d\omega) \quad P\text{-a.s.}$$

being valid for all $d \geq 1$. Hence (11) follows straightforward by taking infimum over all $d \geq 1$.

(4): Since we obviously have:

$$\sup_{\xi\in b(\theta,r)} (h_n(\omega,\xi) - I(\xi))^+ \geq \sup_{\xi\in b(\theta,r)} h_n(\omega,\xi) - \sup_{\xi\in b(\theta,r)} I(\xi)$$

for all $\omega \in \Omega$, all $n \geq 1$ and all $0 < r \leq r_\theta$, from (3) we may conclude:

$$\limsup_{n\to\infty} \sup_{\xi\in b(\theta,r)} h_n(\omega,\xi) \leq \sup_{\xi\in b(\theta,r)} I(\xi) \quad P\text{-a.s.}$$

for all $0 < r \leq r_\theta$. Letting $r \downarrow 0$ we get:

(16) $$\bar{H}(\,\cdot\,,\theta) \leq \bar{I}(\theta) \quad P\text{-a.s.}$$

Moreover since θ belongs to $\hat{M}$, then $\bar{H}(\,\cdot\,,\theta) \geq \beta$ P-a.s. However by definition of $\bar{I}$ we may easily conclude that $\bar{I}(\theta) \leq \beta$ for all $\theta \in \Theta$. Thus (4) follows straightforward by (16). The rest of the proof is exactly the same as the corresponding last part of the proof of theorem 3. These facts complete the proof. □

4. Examples of application and concluding remarks

There are many examples of statistical models covered by the preceding results. We do not wish to review them all here, but will point out examples in [6] (p.34,62,70) and [7] (chapters 12 and 13 with exercises) which can be easily modified to treat the stationary (ergodic) case under our consideration. Other important examples may be found in the references. In all of these examples we consider the setting introduced in section 1 (the criterion function $h(s,\theta)$ is the log-likelihood function) and apply results of theorem 1, theorem 3, and theorem 4 in section 3. In particular we remind on conditions (7)-(9) in section 3 which are in most of the cases easily verified. We may observe that condition (10) in section 3 (which is at the basis of theorems 3 and 4) involves Blum-DeHardt's condition (metric entropy with bracketing) as a particular case (by removing the infimum and putting $n = 1$), see [2] (p.39-44). Blum-DeHardt's condition is, so far, the best known sufficient condition for the uniform law of large numbers. Condition (11) in section 3 might be viewed as its (asymptotic) refinement towards consistency. The task of verification of the underlying conditions in the examples stated above is easy, and we shall leave the remaining details to the reader. Of course there are examples which are not covered by these conditions, but they require individual treatments and will be not considered here. Our main purpose was to unify as many examples as possible, but under the common and simple conditions. We clarify that the main novelty in the applications just described is the fact that we only assume stationarity (and ergodicity) of the underlying sequence of observations. In this way we generalize and extend the previous results that rely upon independence. We are unaware of a similar result and think that this extension is by itself of theoretical and practical interest.

In order to illustrate the facts just described, we present two typical (statistical) examples which in essence motivated the theory exposed above and indicate its application. It should be noted in the first example that the criterion function $h(s,\theta)$ is not necessarily the log-likelihood function for the underlying statistical model. This fact indicates a broader scope of application of our approach and results, and this will be additionally explained in the last part of this section (following the second example below).

Example 1. (Consistency of medians in the stationary case)

Let $h(s,\theta) = -|t(s)-\theta|$ for $(s,\theta) \in \mathbf{R}^2$, where $t \in L^1(\pi)$. Thus $S = \Theta_0 = \mathbf{R}$, and we shall set $\Theta = \bar{\mathbf{R}}$ (with the usual topology). Let $\{X_j \,|\, j \geq 1\}$ be a stationary ergodic sequence of real valued random variables which are defined on $(\Omega, \mathcal{F}, P)$ and have the common distribution law π . Then $I(\theta) = -E|t(X_1)-\theta|$, and thus $M = \{\, \theta \in \mathbf{R} \mid \theta$ is a *median* of $t(X_1) \,\}$:

$$\theta \in M \iff P\{\, t(X_1) < \theta \,\} \leq \tfrac{1}{2} \leq P\{\, t(X_1) \leq \theta \,\} \; .$$

Since $h_n(\omega,\theta) = -\frac{1}{n}\sum_{j=1}^n |t(X_j)-\theta|$, we see that $\hat{\theta}_n(\omega)$ maximizes $h_n(\omega,\cdot)$ on Θ_0 , if and only if $\hat{\theta}_n(\omega)$ is a *sample median* of $t(X_1(\omega)),\dots,t(X_n(\omega))$:

$$\tfrac{1}{n}\Big(card\big\{\, 1 \leq j \leq n \mid t(X_j(\omega)) < \hat{\theta}_n(\omega) \,\big\}\Big) \leq \tfrac{1}{2} \leq \tfrac{1}{n}\Big(card\big\{\, 1 \leq j \leq n \mid t(X_j(\omega)) \leq \hat{\theta}_n(\omega) \,\big\}\Big)$$

for $\omega \in \Omega$. In order to apply our results above, we shall verify the three conditions (7)-(9) following remark 2 in section 3 with $\Gamma = \Theta = \bar{\mathbf{R}}$. First, since $-\infty < I(\theta)$ for all $\theta \in \mathbf{R}$, the condition (7) is evident. Second, since $h(s,\theta) \leq 0$ for all $(s,\theta) \in S \times \Theta$, the condition (8) is obviously satisfied. Finally, since $\theta \mapsto h(s,\theta)$ is continuous on Θ_0 , and moreover $\lim_{n\to\pm\infty} h(s,\theta_n) = h(s,\pm\infty) = -\infty$ whenever $\theta_n \to \pm\infty$ with $s \in S$, we see that the condition (9) is fulfilled as well. Thus by theorem 1 in section 3 we may conclude that h is π-consistent (on Θ). In other words, every accumulation point of any sequence of approximating maximums $\{\, \hat{\theta}_n \mid n \geq 1 \,\}$ (for instance, of any one which satisfies:

$$h_n(\omega, \hat{\theta}_n(\omega)) \geq \sup_{\theta \in \Theta_0} h_n(\omega,\theta) - \varepsilon_n(\omega)$$

for some $\varepsilon_n(\omega) \to 0$ with $\omega \in \Omega$) belongs to the set M . As a particular case we obtain that *every accumulation point of any sample median* of the sequence $\{\, (t(X_1),\dots,t(X_n)) \mid n \geq 1 \,\}$ *belongs to the set of all medians* of $t(X_1)$. In particular, if $t(X_1)$ has a unique median $m\big(t(X_1)\big)$, then:

$$\hat{\theta}_n(\omega) \to m\big(t(X_1)\big) \quad \text{for } P\text{-a.s. } \omega \in \Omega$$

whenever $\{\, \hat{\theta}_n \mid n \geq 1 \,\}$ is a sequence of sample medians of $\{\, (t(X_1),\dots,t(X_n)) \mid n \geq 1 \,\}$. These facts are known to be valid if the sequence $X_1, X_2, \dots$ is assumed to be independent and identically distributed, see [6] p.34-35, but they are not accessible by those methods and results without the assumption of independence. Thus, the results above generalize this and extend to the stationary case, in a rather straightforward way, by applying our main results in section 3 above.

Example 2. (Consistency of the generalized inverse Gaussian distribution in the case of stationary observations)

The *generalized inverse Gaussian distribution* is the distribution on $S =]0,\infty[$ having density:

$$f(s,\theta) = \frac{s^{\lambda-1}}{c(\theta)} \exp\left(-\frac{1}{2}\left(\frac{\chi}{s} + \psi s\right)\right) \tag{1}$$

for $s \in]0,\infty[$ and $\theta = (\lambda,\chi,\psi) \in \Theta_0$, where $\Theta_0 = \Theta_1 \cup \Theta_2 \cup \Theta_3$ with:

$$\Theta_1 = \{\ (\lambda,\chi,\psi)\ |\ \lambda \in \mathbf{R}\ ,\ \chi > 0\ ,\ \psi > 0\ \}$$
$$\Theta_2 = \{\ (\lambda,\chi,\psi)\ |\ \lambda > 0\ ,\ \chi = 0\ ,\ \psi > 0\ \}$$
$$\Theta_3 = \{\ (\lambda,\chi,\psi)\ |\ \lambda < 0\ ,\ \chi > 0\ ,\ \psi = 0\ \}$$

and the map $\theta \mapsto c(\theta)$ is defined by:

$$c(\lambda,\chi,\psi) = \begin{cases} 2\left(\frac{\chi}{\psi}\right)^{\lambda/2} K_\lambda\big(\sqrt{\chi\psi}\big) & ,\ (\lambda,\chi,\psi) \in \Theta_1 \\ 2^\lambda\,\chi^{-\lambda}\,\Gamma(\lambda) & ,\ (\lambda,\chi,\psi) \in \Theta_2 \\ 2^{-\lambda}\,\chi^\lambda\,\Gamma(-\lambda) & ,\ (\lambda,\chi,\psi) \in \Theta_3 \end{cases} \tag{2}$$

where K_λ is the modified Bessel function (of the third kind) with index λ :

$$K_\lambda(x) = \int_0^\infty t^{-\lambda-1} \exp\Big(-\frac{1}{2}\,x\Big(t+\frac{1}{t}\Big)\Big)\ dt \tag{3}$$

for $\lambda \in \mathbf{R}$ and $x > 0$. Special cases of (1) are the *gamma distribution* ($\lambda > 0$, $\chi = 0$), the *distribution of a reciprocal gamma variate* ($\lambda < 0$, $\psi = 0$), the *inverse Gaussian distribution* ($\lambda = -\frac{1}{2}$), and the *distribution of a reciprocal inverse Gausssian variate* ($\lambda = \frac{1}{2}$). Other important cases are $\lambda = 0$ (the *hyperbola distribution*) and $\lambda = 1$. To the best of our knowledge the consistency of the generalized inverse Gaussian distribution in the case of independent observations has been firstly recorded in [6] (see p.111-112). This has been obtained as a consequence of a general result on the consistency of the exponential statistical models. The result and method appear to be rather involved. Here we shall generalize this and extend to the case of stationary observations, thus going beyond the scope of the result and method in [6]. Moreover, the proof of consistency indicated below seems to be more direct and transparent, even in the case of independent observations, thus serving nicely as an application of the theory presented above.

Suppose we are given a stationary ergodic sequence $\{\ X_j\ |\ j \geq 1\ \}$ of random variables defined on $(\Omega,\mathcal{F},P)$ with values in $]0,\infty[$ and common distribution law π which has the density of the form (1). Then the log-likelihood function is given by:

$$h(s,\theta) = (\lambda-1)\log s - \frac{\chi}{2s} - \frac{\psi s}{2} - \log c(\theta)$$

for $s \in S$ and $\theta = (\lambda,\chi,\psi) \in \Theta_0$. In order to prove that $h(s,\theta)$ is π-consistent on Θ_0 , we shall verify the three conditions (7)-(9) following remark 2 in section 3. First, note that evidently $\theta \mapsto h(s,\theta)$ is continuous on Θ_1 for $s \in S$. Moreover, it is rather straightforward by using definition (2) and expression (3) to check that $\theta \mapsto h(s,\theta)$ is continuous on $\Theta_2 \cup \Theta_3$. Thus (9) above is fulfilled with $\Gamma = \Theta_0$. Second, a similar elementary verification shows that (8) above is fulfilled with $\Gamma = \Theta_0$. Finally, condition (7) is evident (since by the essence of this condition one has a freedom of taking the supremum over the whole ball around the points in $\Gamma = \Theta_0$ to be examined). Thus, by the result of theorem 1 in section 3, we may conclude that $h(s,\theta)$ is π-consistent on Θ_0 . In other words, every accumulation point of any sequence of approximating maximums $\{\ \hat{\theta}_n\ |\ n \geq 1\ \}$ (for instance, of any one which satisfies:

$$\frac{1}{n}\sum_{j=1}^{n}\log f(X_j(\omega),\hat{\theta}_n(\omega)) \geq \sup_{\theta\in\Theta_0}\left(\frac{1}{n}\sum_{j=1}^{n}\log f(X_j(\omega),\theta)\right) - \varepsilon_n(\omega)$$

for some $\varepsilon_n(\omega) \to 0$ with $\omega \in \Omega$) converges P-a.s. to the true parameter value (the point $\theta_0 \in \Theta_0$ for which $\pi \sim f(\,\cdot\,,\theta_0)$). In this context (3) and (4) from section 1 are to be recalled.

Thus, apart from the fact that we have extended the result of [6] to the stationary case, where the method in [6] (relying upon independence) is not directly applicable, we have obtained a more transparent proof of this result as well. (In this context, and in general as well, it is instructive to observe that in essence the only extra-condition, which distinguishes the stationary case from the independent case, is the condition (7) above.)

In the remainder we explain the role of the preceding results in the area of stochastic processes and applications. Let us for this consider a sequence of stochastic processes $\{\,(Z_n(t))_{t\in T} \mid n \geq 1\}$ defined on the probability space $(\Omega,\mathcal{F},P)$ and having the common time set T . Let $\hat{t}_n(\omega)$ be a maximum point of $Z_n(\omega,\,\cdot\,)$ on T , that is:

$$Z_n(\omega,\hat{t}_n(\omega)) = \sup_{t\in T} Z_n(\omega,t) \tag{4}$$

for $\omega \in \Omega$ and $n \geq 1$. Then the preceding results amounts to the study of the asymptotic behavior of the maximum points $\hat{t}_n(\omega)$ of $Z_n(\omega,\,\cdot\,)$ for $n \to \infty$. We think that this problem appears worthy of consideration. Under the hypotheses in this paper we have:

$$Z_n(\,\cdot\,,t) \longrightarrow L(t) \quad P\text{-a.s.} \tag{5}$$

as $n \to \infty$. We consider the set $M \subset T$ of all maximum points of the degenerated limiting process L on T , and ask when does $\hat{t}_n(\omega)$ approach M for $n \to \infty$ and $\omega \in \Omega$. It may happen that the supremum in (4) is not attained, and thus we relax condition (4) by requiring:

$$Z_n(\omega,\hat{t}_n(\omega)) \geq \Big(\sup_{t\in T} Z_n(\omega,t) - \varepsilon_n(\omega)\Big) \wedge n \tag{6}$$

for $\omega \in \Omega$ and $n \geq 1$ with $\varepsilon_n(\omega) \to 0$ as $n \to \infty$. From (5) and (6) we could get:

$$\liminf_{n\to\infty} Z_n(\omega,\hat{t}_n(\omega)) \geq \sup_{t\in T} L(t) \quad P\text{-a.s.} \tag{7}$$

In this way a *sequence of approximating maximums* $\{\,\hat{t}_n\}_{n\geq 1}$ is obtained (recall (8)-(10) in section 2 and the sentence following it). In this paper we consider the process:

$$n\,Z_n(\omega,t) = \sum_{j=1}^{n}\gamma(X_j(\omega),t) \tag{8}$$

where $\{\,X_j \mid j \geq 1\}$ is a stationary ergodic sequence of random variables, and $\gamma(x,t)$ is a real valued function. We recall that the case in [18] has been studied where $\{\,Z_n(t)\,\}_{n\geq 1}$ is assumed to be a reversed submartingale for each fixed $t \in T$. In the present case theorem 1, theorem 3, and theorem 4 in section 3 offer solutions for the problem just described. In order to indicate possible applications of these results, we give three examples which follow the same pattern and can easily be modified to treat the new cases. The problem which motivates such considerations is a *problem of maximization (with a random noise)*, but of sums with a large number of summands (corresponding to the large n below), so that the exact computations fail. The results above indicate how to overcome such a difficulty and find an approximative solution (which eventually reaches the exact one). (The "good" rate of convergence, in the form of an asymptotic normality, is conjectured to hold in these and similar cases as well, but this will be not considered here.) We are in general unaware of similar results. Throughout $\{ X_j \mid j \geq 1 \}$ denotes a stationary ergodic sequence of random variables.

Example 3.

Let $X_1 \sim N(0,1)$ be from the standard Gaussian distribution with density function $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$ for $x \in \mathbf{R}$. Let C denote the unique number from $]0, \pi/2[$ that satisfies $\tan(C) = C^{-1}$, and let T be a compact set in $\mathbf{R}$ containing C . If $\hat{t}_n(\omega)$ maximizes the process:

$$n Z_n(\omega, t) = (\sin t) \sum_{j=1}^{n} \cos(t X_j(\omega))$$

over $t \in T$ (in the sense of (6) or (7) above), then $\hat{t}_n \to C$ P-a.s. as $n \to \infty$. This fact readily follows from theorem 1 in section 3 by putting $h(x,t) = (\sin t) \cos(tx)$ and using that $\int_{-\infty}^{\infty} \exp(-x^2/2) \cos(tx)\ dx = \sqrt{2\pi} \exp(-t^2/2)$ for $t \in \mathbf{R}$. It should be noted that the given C is a unique maximum point of $(\sin t) \exp(-t^2/2)$ for $t \in \mathbf{R}$.

Example 4.

Let $X_1 \sim Exp(1)$ be from the exponential distribution with density function $f(x) = \exp(-x)$ for $x \in \mathbf{R}_+$. Let T be a compact set in $\mathbf{R}_+$ containing 1 . If $\hat{t}_n(\omega)$ maximizes the process:

$$n Z_n(\omega, t) = \sum_{j=1}^{n} \sin(t X_j(\omega))$$

over $t \in \mathbf{T}$ (in the sense of (6) or (7) above), then $\hat{t}_n \to 1$ P-a.s. as $n \to \infty$. This fact readily follows from theorem 1 in section 3 by putting $h(x,t) = \sin(tx)$ and using that $\int_0^{\infty} \exp(-x) \sin(tx)\ dx = t/(1+t^2)$ for $t \in \mathbf{R}_+$. It should be noted that 1 is a unique maximum point of $t/(1+t^2)$ for $t \in \mathbf{R}_+$.

Example 5.

Let $X_1 \sim C(1,0)$ be from the Cauchy distribution with density function $f(x) = 1/\pi(1+x^2)$ for $x \in \mathbf{R}$. Let T be a compact set in $\mathbf{R}_+$. If $\hat{t}_n(\omega)$ maximizes the process:

$$n Z_n(\omega, t) = \sum_{j=1}^{n} X_j(\omega) \sin(t X_j(\omega))$$

over $t \in T$ (in the sense of (6) or (7) above), then $\hat{t}_n \to \min(T)$ P-a.s. as $n \to \infty$. This fact readily follows from theorem 1 in section 3 by putting $h(x,t) = x \sin(tx)$ and using that $\int_0^{\infty} x \sin(tx)/(1+x^2)\ dx = (\pi/2) \exp(-t)$ for $t \in \mathbf{R}_+$. It should be noted that $\min(T)$ is a unique maximum point of $\exp(-t)$ for $t \in T$.

Acknowledgment. The author would like to thank his supervisor, Professor J. Hoffmann-Jørgensen, for instructive discussions and valuable comments.

REFERENCES

[1] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

[2] DUDLEY, R. M. (1984). A course on empirical processes. *Ecole d'Eté de Probabilitiés de Saint-Flour, XII-1982. Lecture Notes in Math.* No. 1097, Springer-Verlag (1-142).

[3] FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London*, Ser. A, Vol. 22 (309-368).

[4] FISHER, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Phil. Soc.* 22 (700-725).

[5] HOFFMANN-JØRGENSEN, J. (1984). Necessary and sufficient conditions for the uniform law of large numbers. *Proc. Conf. Probab. Banach Spaces V (Medford 1984), Lecture Notes in Math.* No. 1153, Springer-Verlag (127-137).

[6] HOFFMANN-JØRGENSEN, J. (1992). Asymptotic likelihood theory. *Proc. Conf. Functional Anal. III (Dubrovnik 1989), Various Publ. Ser.* No. 40, 1994 (5-192).

[7] HOFFMANN-JØRGENSEN, J. (1994). *Probability with a View toward Statistics*, Volumes I and II. Chapman & Hall.

[8] HUZURBAZAR, V. S. (1948). The likelihood equation, consistency and the maxima of the likelihood function. *Annals of Eugenics* 14 (185-200).

[9] KIEFER, J. *and* WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27 (887-906).

[10] KRENGEL, U. (1985). *Ergodic Theorems*. Walter de Gruyter & Co., Berlin.

[11] KULLBACK, S. *and* LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* 22 (79-86).

[12] LARSEN, F. S. (1990). Asymptotic likelihood theory – zero point estimation. *Math. Inst. Aarhus, Preprint Ser.* No. 38, (56 pp).

[13] LE CAM, L. *and* SCHWARTZ, L. (1960). A necessary and sufficient condition for the existence of consistent estimates. *Ann. Math. Statist.* 31 (140-150).

[14] MØLLER, A. M. (1990). Consistent sampling from a finite number of types. *Institute of Mathematics, University of Aarhus, Preprint Series* No. 36 (31 pp).

[15] NORDEN, R. H. (1972). A survey of maximum likelihood estimation. *Internat. Statist. Rev.* 40 (329-354).

[16] NORDEN, R. H. (1972). A survey of maximum likelihood estimation (Part 2). *Internat. Statist. Rev.* 41 (39-58).

[17] PESKIR, G. (1991). The existence of measurable approximating maximums. *Math. Inst. Aarhus, Preprint Ser.* No. 29, (13 pp). *Math. Scand.* Vol. 77, No. 1, 1995 (71-84).

[18] PESKIR, G. (1992). Consistency of statistical models described by families of reversed

submartingales. *Math. Inst. Aarhus, Preprint Ser.* No. 9, (26 pp). *Probab. Math. Statist.* 18, 1998 (289-318).

[19] PESKIR, G. *and* WEBER, M. (1992). Necessary and sufficient conditions for the uniform law of large numbers in the stationary case. *Math. Inst. Aarhus, Preprint Ser.* No. 27, (26 pp). *Proc. Funct. Anal.* IV (Dubrovnik 1993), *Various Publ. Ser.* No. 43, 1994 (165-190).

[20] PESKIR, G. (1992). Measure compact sets of functions and consistency of statistical models. *Math. Inst. Aarhus, Preprint Ser.* No. 38, (10 pp). *Theory Probab. Appl.* 38, 1993 (360-367).

[21] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* 20 (595-601).

[22] WOLFOWITZ, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* 20 (601-602).

Goran Peskir
Department of Mathematical Sciences
University of Aarhus, Denmark
Ny Munkegade, DK-8000 Aarhus
home.imf.au.dk/goran
goran@imf.au.dk