Theory Probab. Appl. Vol. 38, No. 2, 1993, (360-367)
Preprint Ser. No. 38, 1992, *Math. Inst. Aarhus*

Measure Compact Sets of Functions and Consistency of Statistical Models

GORAN PESKIR

Sufficient conditions for consistency of statistical models are deduced by using Fréchet-Šmulian's necessary and sufficient conditions for conditionally compactness relative to the topology of convergence in measure being imposed on the family of associated densities by so-called a control measure . The method relies upon the facts established in [4] where sufficient conditions for consistency are deduced by using necessary and sufficient conditions for conditionally compactness of the family of associated densities relative to the topology of pointwise convergence. In the case where a statistical model under consideration admits a finite control measure the present conditions for conditionally compactness become weaker and easily verified. However the present approach relies upon the fact that any control measure that yields consistency must be nice enough in such a way that the upper oscillation of densities on infinitesimally small balls behaves smoothly relative to the distribution of the random phenomenon under consideration.

1. Measure compact sets of functions

Let (T,d) be a separable metric space, let S be a non-empty set, and let $\mathcal{M} \subset T^S$ be a family of functions from S into T . Let τ_S denote the topology of pointwise convergence on S , and let cl_S denote the τ_S-closure in T^S . Then $\mathcal{M}$ is said to be *pointwise compact metrizable* in T^S , if $cl_S(\mathcal{M})$ is compact and metrizable in the τ_S-topology. It can be shown that $\mathcal{M}$ is pointwise compact metrizable in T^S if and only if the following two conditions are satisfied, see [4]:

(1) The set $T_s = \{ \ f(s) \ | \ f \in \mathcal{M} \ \}$ is conditionally compact in T for all $s \in S$

(2) The map $\delta(u,v) = \sup_{f \in \mathcal{M}} d\,(f(u)\,,f(v))$ defines a *separable* pseudometric on S .

Moreover if in this case $D = \{ \ d_j \ | \ j \geq 1 \ \}$ denotes a countable δ-dense set in S , then by:

$$\rho(f,g) \,=\, \sum_{n=1}^{\infty} \frac{1}{2^n} \arctan\Big(d\,(f(d_n)\,,g(d_n))\Big) \tag{3}$$

a separable metric on $cl_S(\mathcal{M})$ is defined such that the ρ-topology, the τ_S-topology and the τ_D-topology on $cl_S(\mathcal{M})$ coincide. Furthermore we have:

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, reference measure, parameter set, analytic metric space, sample space, the likelihood function, the log-likelihood function, the unknown (true) distribution, the (upper, empirical) information function, consistent, maximum likelihood estimator, pointwise compact metrizable, separable, conditionally compact, Fréchet-Šmulian compactness criteria, control measure, upper semicontinuous, the Hewitt-Savage 0-1 law, the projection theorem, the permutation invariant σ-algebras. goran@imf.au.dk

(4) $$\mathcal{B}(S,\delta) = \sigma(\mathcal{M}) = \sigma(cl_S(\mathcal{M}))\ .$$

For more details see [4]. This result applies to the problem of consistency of statistical models as follows, see [4] and [5]. Let $\Pi = (\ \pi_\theta \mid \theta \in \Theta_0\ )$ be a statistical model with a sample space $(S,\mathcal{A})$, reference measure μ , likelihood function $f(s,\theta)$ and log-likelihood function $h(s,\theta) = \log f(s,\theta)$ for $(s,\theta) \in S \times \Theta_0$ with Θ_0 being an analytic metric space. In other words $(S,\mathcal{A},\mu)$ is a measure space and each π_θ is a probability measure on $(S,\mathcal{A})$ such that $\pi_\theta \ll \mu$ and $f(s,\theta) = (d\pi_\theta/d\mu)(s)$ for $(s,\theta) \in S \times \Theta_0$. The map $(s,\theta) \mapsto f(s,\theta)$ is supposed to be measurable on $S \times \Theta_0$. Moreover we suppose that $\pi = \pi_{\theta_0}$ for some $\theta_0 \in \Theta_0$ is the unknown (true) distribution of a random phenomenon, and that $\{\ X_j \mid j \geq 1\ \}$ is a sequence of independent and identically distributed random variables defined on a probability space $(\Omega,\mathcal{F},P)$ with values in $(S,\mathcal{A})$ and common distribution law π . The information function is defined as follows:

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

for all $\theta \in \Theta_0$ where the integral is supposed to exist. We put $M_h(\pi)$ to denote the set of all maximum points of I on Θ_0 . The empirical information function is defined as follows:

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

for all $(\omega,\theta) \in \Omega \times \Theta_0$ and all $n \geq 1$. We put $h_n^*(\omega,B)$ to denote $\sup_{\theta\in B} h_n(\omega,\theta)$ whenever $B \subset \Theta_0$, $\omega \in \Omega$ and $n \geq 1$. Assuming the validity of the following two conditions:

(5) If $\pi_\theta = \pi$ for some $\theta \in \Theta_0$, then $\theta = \theta_0$

(6) $$\int_S f(s,\theta_0)\ |\log f(s,\theta_0)|\ \mu(ds)\ < \infty$$

and using (1.2.4) in [5] we may conclude that $M_h(\pi) = \{\theta_0\}$. If in addition the following three conditions are satisfied:

(7) $$\inf_{n\geq 1} E^* h_n^*(\Theta_0) < \infty$$

(8) The set $\mathcal{M} = \{\ f(\,\cdot\,,\theta) \mid \theta \in \Theta_0\ \}$ is pointwise compact metrizable in $[0,\infty]^S$

(9) If $\{\ \theta_j \mid j \geq 1\ \}$ is a sequence in Θ_0 such that $f(s,\theta_n) \to f(s,\theta_0)$ for all $s \in S$, then $\theta_n \to \theta_0$ as $n \to \infty$

then h is π-consistent, see [4]. In other words every sequence of approximating maximums converges P-a.s. to θ_0 . In particular any sequence of maximum likelihood estimators converges P-a.s. to θ_0 . For more details see [4] and [5]. Note that we also have convergence in $L^1(\mu)$ in (9) since $\int f(s,\theta_n)\ \mu(ds) = \int f(s,\theta_0)\ \mu(ds) = 1$ with $f(\,\cdot\,,\theta_n)$ being non-negative for all $n \geq 1$. By (1) and (2) above we see that condition (8) is equivalent to the following two conditions:

(10) $$\sup_{\theta\in\Theta_0} f(s,\theta)\ < \infty \quad \text{for all} \quad s \in S$$

(11) The map $\rho(s',s'') = \sup_{\theta \in \Theta_0} \mid f(s',\theta) - f(s'',\theta) \mid$ defines a separable pseudo-metric on S .

The main idea of the present paper may be stated as follows. Having in mind Fréchet-Šmulian's conditions for conditionally compactness we shall replace the topology of pointwise convergence in (8) with the topology of convergence in measure. We proceed by giving details in this direction. Let ν be a measure on $(S,\mathcal{A})$, then $M(\nu)$ denotes the set of all ν-measurable functions from S into $\mathbf{R}$. For $f \in M(\nu)$ we define:

(12) $$\| f \|_0 = \inf \{ \alpha > 0 \mid \nu\{ \mid f \mid > \alpha \} \le \alpha \}$$

and we put $L^0(\nu)$ to denote the set of all $f \in M(\nu)$ satisfying $\| f \|_0 < \infty$. It is easily verified that we have:

(13) $$L^0(\nu) = \{ f \in M(\nu) \mid \lim_{\beta \to \infty} \nu\{ \mid f \mid > \beta \} = 0 \} .$$

Moreover it is well-known that $(L^0(\nu), \| \cdot \|_0)$ forms an F-space, see [1] (p.51,146). Convergence in $L^0(\nu)$ is equivalent to convergence in ν-measure. The space $L^0(\nu)$ need not admit any non-zero continuous linear functional, see [1] (p.329). Moreover if $\nu(S) < \infty$, then a necessary and sufficient condition that there exists a non-zero continuous linear functional on $L^0(\nu)$ is that there exists an atom for ν , see [11]. It is useful to observe that the space $L^0(\nu)$ as defined here equals to the space $TM(S,\mathcal{A},\nu,\mathbf{R})$ of totally ν-measurable functions from S into $\mathbf{R}$ as defined in [1] (p.329). The F-norm of $f \in TM(S,\mathcal{A},\nu,\mathbf{R})$ is given by:

$$\mid f \mid = \inf_{\alpha > 0} (\alpha + \nu\{ \mid f \mid > \alpha \})$$

and it is easily verified that we have:

$$\| f \|_0 \le \mid f \mid \le 2 \| f \|_0 .$$

Moreover if $\nu(S) < \infty$, then any left continuous increasing concave bounded function $\varphi : [0,\infty[\to [0,\infty[$ such that $\varphi(x) > 0$ for $x > 0$ and $\varphi(0) = \lim_{x \downarrow 0} \varphi(x) = 0$ may be used to define an F-norm of $f \in L^0(\nu)$ as follows:

$$\| f \|_\varphi = \int_S \varphi (\mid f \mid) \, d\nu$$

in such a way that the $\| \cdot \|_\varphi$-topology and the $\| \cdot \|_0$-topology on $L^0(\nu)$ coincide, see [6]. Compactness in the space $L^0(\nu)$ may be characterized nicely, see [1] (p.330), [2] and [19]. For this we shall recall that any family $\gamma = \{ \Delta_1, \dots, \Delta_n \}$ of non-empty subsets of S satisfying $S = \cup_{j=1}^n \Delta_j$ is called *a finite cover* of S . If we moreover have $\gamma \subset \mathcal{A}$, then γ is said to be *measurable*. The family of all finite covers of S will be denoted by $\Gamma(S)$. Let us also clarify that a subset of a metric space is called *conditionally compact* if its closure is compact.

Theorem 1. (Fréchet-Šmulian)

Let $(S,\mathcal{A},\nu)$ be a measure space, and let Ξ_0 be a subset of $L^0(\nu)$. Then Ξ_0 is conditionally *compact in* $L^0(\nu)$ *if and only if for every* $\varepsilon > 0$ *there exist* $M_\varepsilon > 0$ *, a measurable cover* $\gamma_\varepsilon = \{ \Delta_1, \dots, \Delta_{n_\varepsilon} \} \in \Gamma(S)$ *, and a set* $D_{f,\varepsilon} \in \mathcal{A}$ *depending on* $f \in \Xi_0$ *such that:*

(i) $\nu(D_{f,\varepsilon}) \le \varepsilon$

(ii) $|\, f(s) \,| \le M_\varepsilon$ *for all* $s \in S \setminus D_{f,\varepsilon}$

(iii) $\sup_{s',s'' \in \Delta_j \setminus D_{f,\varepsilon}} |\, f(s') - f(s'') \,| \le \varepsilon$ *for all* $j = 1, \dots, n_\varepsilon$

being valid for all $f \in \Xi_0$ *, see [2] and [19].*

It is instructive to notice a certain degree of freedom that the exceptional set $D_{f,\varepsilon}$ in this characterization yields, compare with theorem 6 (p.260) and theorem 5 (p.266) in [1]. In the case where S is a decent subset of the real line somewhat different characterizations of conditionally compactness in $L^0(\nu)$ may be found in [7] and [9]. In addition we shall put $\Xi_0 = \{ f(\,\cdot\,,\theta) \mid \theta \in \Theta_0 \}$, then we have $\Xi_0 \subset L^1(\mu) \subset L^0(\mu)$. However it turns out that any measure ν on $(S, \mathcal{A})$ such that $\Xi_0 \subset L^0(\nu)$ may be of interest. Such a measure ν will be called *a control measure* for the statistical model Π . In addition ν will denote a control measure for Π . The closure of Ξ_0 in $L^0(\nu)$ will be denoted by Ξ . Thus ξ belongs to Ξ if and only if there exists a sequence $\{ \xi_n \mid n \ge 1 \}$ in Ξ_0 satisfying $\xi_n \to \xi$ in ν-measure. The space $L^0(\nu)$ will be considered relative to the F-norm $\|\cdot\|_0$ defined by (12) above. However other norms inducing the topology on $L^0(\nu)$ might also be used. We shall denote $b_\Xi(\xi, r) = \{ \xi' \in \Xi \mid \| \xi' - \xi \|_0 < r \}$ whenever $\xi \in \Xi$ and $r > 0$. Moreover we shall say that the statistical model Π is *π-upper semicontinuous relative to the control measure* ν , if the following condition is satisfied:

(14) $$\sup_{\xi' \in b_\Xi(\xi, r)} \xi' \;\to\; \xi \quad \text{in } \pi\text{-measure}$$

as $r \downarrow 0$ for all $\xi \in \Xi$. Note that $b_\Xi(\xi, r)$ equals to the set of all ξ' in Ξ satisfying $\nu\{ \,|\, \xi' - \xi \,| > r \,\} \le r$ for $\xi \in \Xi$ and $r > 0$. Hence we see that condition (14) may be viewed as an oscillation condition on the family Ξ_0 . We shall see later that condition (14) together with Fréchet-Šmulian's compactness conditions in theorem 1 above take a central place in establishing consistency. Thus the control measure ν has to be chosen having these conditions as criteria. Roughly speaking we may say that a statistical model will be consistent as soon as it admits a nice control measure. We shall in addition introduce a map $\kappa : \Theta_0 \to \Xi_0$ in a natural way by putting:

$$\kappa(\theta) = f(\,\cdot\,, \theta)$$

for all $\theta \in \Theta_0$. It turns out to be useful to assume that $(\,id \times \kappa\,)(\Theta_0) = \{ \,(\,\theta\,, f(\,\cdot\,,\theta)\,) \mid \theta \in \Theta_0 \,\}$ is an analytic set in $\Theta_0 \times \Xi_0$. This fact will be for instance true if κ is a continuous map from Θ_0 into Ξ_0 , see (0.13) in [5]. Moreover from this fact we may conclude that $\kappa(A)$ is analytic in Ξ_0 whenever A is analytic in Θ_0 . Indeed $\kappa(A)$ is the projection of $(\,id \times \kappa\,)(\Theta_0) \cap (A \times \Xi)$ on Ξ and thus the statement follows from the fact that Ξ is analytic, see (0.15) in [5]. We shall define the criterion function $g : S \times \Xi \to \bar{\mathbf{R}}$ in a natural way:

$$g(s, \xi) = \log \xi(s)$$

for all $(s,\xi) \in S \times \Xi$. Note that we have:

(15) $$h(s,\theta) = g(s,\kappa(\theta))$$

for all $s \in S$ whenever $\theta \in \Theta_0$. We shall in addition recall that the upper information function associated to h is defined by:

$$\bar{H}(\theta) = \lim_{r \downarrow 0} H^*(\, b_\Theta(\theta,r)\,)$$

for all $\theta \in \Theta$ with (Θ,d) being a suitable chosen compact metric space containing Θ_0 and $b_\Theta(\theta,r) = \{\, \theta' \in \Theta \mid d(\theta',\theta) < r \,\}$ whenever $\theta \in \Theta$ and $r > 0$. The map $h(s,\,\cdot\,)$ is extended on Θ by putting $h(s,\theta) = -\infty$ for $s \in S$ and $\theta \in \Theta \setminus \Theta_0$. Also recall that $H^*(B)$ stands for $\limsup_{n\to\infty} h_n^*(\,\cdot\,,B)$ which is by the Hewitt-Savage 0-1 law and the projection theorem degenerated whenever B is analytic in Θ . The maps g_n , g_n^* , G^* and $\bar{G}$ that correspond to the criterion function g are supposed to be defined in an analogous way. To distinguish the information functions associated to h and g we shall use the notation $I_h(\theta)$ and $I_g(\xi)$ for $\theta \in \Theta$ and $\xi \in \Xi$ respectively. Let us also clarify that we put $\mathcal{C}\{\theta_n\}$ to denote the set of all accumulation points of a sequence $\{\, \theta_n \mid n \geq 1 \,\}$ in Θ . A similar notation will be used for sequences in Ξ . Recall that $\mathcal{S}_X^n$ denote the permutation invariant σ-algebra of order n based on $X = (X_1, X_2, \,\dots\,)$ for $n \geq 1$. Given $\theta \in \Theta$ and $r > 0$ we shall use $b_{\Theta_0}(\theta,r)$ to denote $b_\Theta(\theta,r) \cap \Theta_0$. Also let us recall that $\hat{M}_h(\pi)$ is defined to be the set of all $\theta \in \bar{\Theta}_0$ satisfying $\bar{H}(\theta) \geq \beta_h(\pi)$ with $\beta_h(\pi) = \sup_{\theta\in\Theta_0} I_h(\theta)$. This fact finishes the establishment of the setting under the next considerations. For more informations in this direction we shall refer the reader to [3], [4], [5], [6], [8], [10], [12], [13], [14], [17] and [18].

2. Consistency theorem

In this section we work within the setting that is introduced in the preceding section. It concerns a statistical model $\Pi = (\, \pi_\theta \mid \theta \in \Theta_0 \,)$ and a control measure ν . In order to present the main result of the paper we shall first state two lemmas.

Lemma 1.

Under the hypotheses of the setting of section 1 let $\theta \in \bar{\Theta}_0$ be a given point. Then there exists a sequence $\hat{\theta}_n : \Omega \to \Theta_0$ for $n \geq 1$ and a P-null set $N \in \mathcal{F}$ satisfying:

(1) $\hat{\theta}_n$ *is $\mathcal{S}_X^n$-measurable for all* $n \geq 1$

(2) $d\,(\hat{\theta}_n(\omega)\,,\theta) \rightrightarrows 0$ *uniformly for* $\omega \in \Omega$ *as* $n \to \infty$

(3) $\bar{H}(\theta) \leq \limsup_{n\to\infty} h_n(\omega,\hat{\theta}_n(\omega))$ *for all* $\omega \in \Omega \setminus N$.

Moreover if C is an analytic subset of Ξ_0 and $\bar{G}_C$ denotes the upper information function associated to the restriction of g to $S \times C$, then there exists a P-null set $N_C \in \mathcal{F}$ such that for any sequence $\hat{\theta}_n : \Omega \to \Theta_0$ with $n \geq 1$ we have:

(4) $$\limsup_{n\to\infty} h_n(\omega,\hat\theta_n(\omega)) \le \sup_{\xi\in\mathcal{C}\{\kappa(\hat\theta_n(\omega))\}} \bar G_C(\xi)$$

whenever $\omega \in \Omega \setminus N_C$ *and* $\kappa(\hat\theta_n(\omega)) \in C$ *for all* $n \ge k$ *with some* $k \ge 1$.

Proof. Statements (1), (2) and (3) follow straight forward by theorem 4.1 in [13] using remark 4.3 in [13]. Statement (4) follows easily by applying (3) in corollary 2.3 with (10) in proposition 2.1 in [14] to C as the parameter set and using (1.15). These facts complete the proof. □

Lemma 2.

Under the hypotheses of the setting of section 1 let us define:

(1) $$\bar F(\theta,\xi) = \lim_{r\downarrow 0} G^*\big(b_\Xi(\xi,r)\cap\kappa(b_{\Theta_0}(\theta,r))\big)$$

for $(\theta,\xi)\in\Theta\times\Xi$. *Moreover let us put* $\Delta(\theta)$ *to denote the set of all* $\xi\in\Xi$ *for which there exists a sequence* $\{\theta_n \mid n\ge 1\}$ *in* Θ_0 *such that* $\theta_n\to\theta$ *and* $f(\cdot,\theta_n)\to\xi$ *in* ν*-measure as* $n\to\infty$. *Then for every* $\theta\in\bar\Theta_0$ *there exists* $\xi\in\Delta(\theta)$ *such that* $\bar H(\theta)=\bar F(\theta,\xi)$.

Proof. Let $\theta\in\bar\Theta_0$ be a given point, then there exist a sequence $\hat\theta_n:\Omega\to\Theta_0$ and a P-null set $N\in\mathcal{F}$ satisfying (1)-(3) in lemma 1. Put C_i to denote $\kappa(b_{\Theta_0}(\theta,2^{-i}))$ for all $i\ge 1$, and let $j\ge 1$ be given and fixed. By (2) in lemma 1 there exists $k\ge 1$ such that $\hat\theta_n(\omega)\in b_{\Theta_0}(\theta,2^{-j})$ for all $n\ge k$ and all $\omega\in\Omega$. Let $\omega\in\Omega\setminus N$ and put $\hat\xi_n(\omega)$ to denote $\kappa(\hat\theta_n(\omega))=f(\cdot,\hat\theta_n(\omega))$, then we have $\hat\xi_n(\omega)\in C_j$ for all $n\ge k$. Let g_j denote the restriction of g to $S\times C_j$, then by (3) and (4) in lemma 1 applied to C_j we can find a P-null set $N_j\in\mathcal{F}$ such that:

(2) $$\bar H(\theta) \le \limsup_{n\to\infty} h_n(\omega,\hat\theta_n(\omega)) \le \sup_{\xi\in\mathcal{C}\{\hat\xi_n(\omega)\}} \bar G_j(\xi)$$

whenever ω does not belong to N_j . Since $\bar G_j$ is upper semicontinuous, then it attains its maximum on the compact set $\mathcal{C}\{\hat\xi_n(\omega)\}$ for any $\omega\in\Omega\setminus N$. Thus by (2) for any $\omega\in\Omega\setminus(N\cup N_j)$ there exists $\xi_j\in\mathcal{C}\{\hat\xi_n(\omega)\}\subset\Delta(\theta)$ such that:

(3) $$\bar H(\theta) \le \bar G_j(\xi_j) \le G^*(b_\Xi(\xi_j,2^{-j})\cap C_j)$$

being valid for all $j\ge 1$. By compactness of $\Delta(\theta)$ there exists $\xi\in\mathcal{C}\{\xi_j\}$. Therefore for any given $r>0$ there exist $\xi_{j_1},\xi_{j_2},\dots$ such that $2^{-j_1}<r$ and $b_\Xi(\xi_{j_m},2^{-j_m})\subset b_\Xi(\xi,r)$ for all $m\ge 1$. Hence by (3) we may conclude:

$$\bar H(\theta) \le G^*(b_\Xi(\xi_{j_m},2^{-j_m})\cap C_{j_m}) \le G^*(b_\Xi(\xi,r)\cap\kappa(b_{\Theta_0}(\theta,r))) .$$

being valid for all $m\ge 1$. Letting $r\downarrow 0$ we see that $\bar H(\theta)\le\bar F(\theta,\xi)$, while the converse inequality follows straight forward by definition (1). These facts complete the proof. □

Theorem 3.

Under the hypotheses of the setting of section 1 let us suppose that the following six conditions *are satisfied:*

(1) $$\int_S f(s,\theta_0)\ |\log f(s,\theta_0)\ |\ \mu(ds)\ <\infty$$

(2) *If* $f(\,\cdot\,,\theta_n) \rightarrow f(\,\cdot\,,\theta_0)$ *ν-a.s. for some sequence* $\{\ \theta_n\ \mid\ n\geq 1\ \}$ *in* Θ_0 *, then* $\theta_n \rightarrow \theta_0$ *as* $n\rightarrow\infty$

(3) *The set* $\Xi_0 = \{\, f(\,\cdot\,,\theta)\mid \theta\in\Theta_0\}$ *is conditionally compact in* $L^0(\nu)$ *. In other words for every* $\varepsilon>0$ *there exist* $M_\varepsilon>0$ *, a measurable cover* $\gamma_\varepsilon = \{\ \Delta_1,\dots,\Delta_{n_\varepsilon}\ \}\in\Gamma(S)$ *, and a set* $D_{\theta,\varepsilon}\in\mathcal{A}$ *depending on* $\theta\in\Theta_0$ *such that:*

(i) $\nu(D_{\theta,\varepsilon})\leq\varepsilon$

(ii) $|\,f(s,\theta)\,|\leq M_\varepsilon$ *for all* $s\in S\setminus D_{\theta,\varepsilon}$

(iii) $\displaystyle\sup_{s',s''\in\Delta_j\setminus D_{\theta,\varepsilon}} |\ f(s',\theta)-f(s'',\theta)\ |\leq\varepsilon$ *for all* $j=1,\dots,n_\varepsilon$

being valid for all $\theta\in\Theta_0$

(4) $$\int_S \xi\ d\mu\leq 1 \ \textit{ for all }\ \xi\in\Xi\setminus\Xi_0$$

(5) *The statistical model* $\Pi = (\ \pi_\theta\ \mid\ \theta\in\Theta_0\ )$ *is upper semicontinuous relative to the control measure* ν *. In other words the following condition is satisfied:*

$$\sup_{\xi'\in b_\Xi(\xi,r)} \xi' \ \rightarrow\ \xi \quad \textit{in } \pi\textit{-measure}$$

as $r\downarrow 0$ *whenever* $\xi\in\Xi$

(6) $$\inf_{r>0}\int_S \sup_{\xi'\in b_\Xi(\xi,r)} \log\xi'(s)\ \pi(ds)\ <\infty \quad \textit{whenever }\ \xi\in\Xi$$

where $\pi=\pi_{\theta_0}$ *for fixed* $\theta_0\in\Theta_0$ *. Then* $\hat{M}_h(\pi)=M_h(\pi)=\{\,\theta_0\,\}$ *and* h *is* π*-consistent in the sense of [5]. In particular any sequence of maximum likelihood estimators converges* P*-a.s. to* θ_0 *.*

Proof. If $\pi_\theta=\pi$ for some $\theta\in\Theta_0$, then by (2) we may easily establish that θ must be equal to θ_0 . Thus by (1) and (1.2.4) in [5] we have $M_h(\pi)=\{\,\theta_0\,\}$. Since the map $x\mapsto\log x$ is continuous, then by (4) we may deduce:

$$\sup_{\xi'\in b_\Xi(\xi,r)} \log\,\xi' \ \rightarrow\ \log\,\xi \quad \text{in } \pi\text{-measure}$$

as $r\downarrow 0$ whenever $\xi\in\Xi$. By (5) and the monotone convergence theorem hence we may conclude:

$$\int_S \sup_{\xi'\in b_\Xi(\xi,r)} \log\,\xi'(s)\ \pi(ds)\ \rightarrow\ \int_S \log\,\xi(s)\ \pi(ds)$$

as $r \downarrow 0$ whenever $\xi \in \Xi$. Thus if we put $\xi_0 = \kappa(\theta_0) = f(\,\cdot\,,\theta_0)$, then by the reversed submartingale convergence theorem we may easily deduce:

$$(7) \qquad \bar{G}(\xi) = I_g(\xi)$$

for all $\xi \in \Xi$. Moreover for any $\xi \in \Xi$ and all $s \in S$ we have:

$$\xi_0(s) \cdot \log \xi(s) \le \xi(s) - \xi_0(s) + \xi_0(s) \cdot \log \xi_0(s)$$

with the strict inequality whenever $\xi_0(s) \neq \xi(s)$. Taking the μ-integral over S hence we get:

$$I_g(\xi) \le \int_S \xi \, d\mu - 1 + I_g(\xi_0)$$

for all $\xi \in \Xi$ with the strict inequality whenever $\mu\{\xi \neq \xi_0\} > 0$. Thus by (4) we may conclude:

$$(8) \qquad I_g(\xi) < I_g(\xi_0) = I_h(\theta_0) = \beta_h(\pi)$$

for all $\xi \in \Xi$ with $\mu\{\xi \neq \xi_0\} > 0$. Let us now take an arbitrary point $\theta \in \bar{\Theta}_0 \setminus \{\theta_0\}$, and let ξ be a point from $\Delta(\theta)$. Then there exists a sequence $\{\theta_n \mid n \ge 1\}$ in Θ_0 such that $\theta_n \to \theta$ and $f(\,\cdot\,,\theta_n) \to \xi$ in ν-measure as $n \to \infty$. Thus there exist $n_1 < n_2 < \dots$ such that $f(\,\cdot\,,\theta_{n_k}) \to \xi$ ν-a.s. for $k \to \infty$. Hence by (2) we see that $\mu\{\xi \neq \xi_0\} > 0$ and thus by (7) and (8) we may conclude:

$$(9) \qquad \bar{G}(\xi) < \beta_h(\pi)$$

for all $\xi \in \Delta(\theta)$ whenever $\theta \in \bar{\Theta}_0 \setminus \{\theta_0\}$. However by lemma 2 we see that for any $\theta \in \bar{\Theta}_0$ there exists $\xi \in \Delta(\theta)$ such that $\bar{H}(\theta) = \bar{F}(\theta,\xi)$. Moreover by definition (1) in lemma 2 and (9) we have $\bar{F}(\theta,\xi) \le \bar{G}(\xi) < \beta_h(\pi)$. Hence we see that $\bar{H}(\theta) < \beta_h(\pi)$ whenever $\theta \in \bar{\Theta}_0 \setminus \{\theta_0\}$. This fact completes the proof. □

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

REFERENCES

[1] DUNFORD, N. *and* SCHWARTZ, J. T. (1958). *Linear Operators, Part I: General Theory*. Interscience Publ. Inc., New York.

[2] FRÉCHET, M. (1927). Sur les ensembles de fonctions mesurables. *Fund. Math.* 9 (25-32).

[3] HOFFMANN-JØRGENSEN, J. (1991). The general monotone convergence theorem and measurable selections. *Math. Inst. Aarhus, Preprint Ser.* No. 13, (37 pp).

[4] HOFFMANN-JØRGENSEN, J. (1991). Pointwise compact metrizable sets of functions and consistency of statistical models. *Math. Inst. Aarhus, Preprint Ser.* No. 15, (6 pp).

[5] HOFFMANN-JØRGENSEN, J. (1992). Asymptotic likelihood theory. *Proc. Funct. Anal.* III

(Dubrovnik 1989), *Various Publ. Ser.* No. 40 (5-192).

[6] HOFFMANN-JØRGENSEN, J. (1994). *Probability With a View Toward Statistics*. Chapman and Hall.

[7] IZUMI, S. (1939). On the compactness of a class of functions. *Proc. Imp. Acad. Tokyo* 15 (111-113).

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

[9] MEDVEDEV, Yu. T. (1953). Two criteria of compactness of families of functions. *Doklady Akad. Nauk SSSR* 90 (337-340) (Russian).

[10] MØLLER, A. M. (1990). Consistent sampling from a finite number of types. *Math. Inst. Aarhus, Preprint Ser.* No. 36, (31 pp).

[11] NIKODÝM, O. M. (1931). Contribution à la théorie des fonctionnelles linéaires en connexion avec la théorie de la mesure des ensembles abstraits. *Mathematica, Cluj*, 5 (130-141).

[12] PESKIR, G. (1991). On separability of families of reversed submartingales. *Math. Inst. Aarhus, Preprint Ser.* No. 27, (26 pp). *Proc. Probab. Banach Spaces* IX (Sandbjerg 1993), *Progr. Probab.* Vol. 35, Birhauser, Boston, 1994 (36-53).

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

[14] 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).

[15] PESKIR, G. (1992). Uniform convergence of reversed martingales. *Math. Inst. Aarhus, Preprint Ser.* No. 21, (27 pp). *J. Theoret. Probab.* 8, 1995 (387-415).

[16] 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).

[17] PESKIR, G. (1992). Consistency of statistical models in the stationary case. *Math. Inst. Aarhus, Preprint Ser.* No. 39, (21 pp). *Math. Scand.* 78, 1996 (293-319).

[18] PITMAN, E.J.G. (1979). *Some Basic Theory for Statistical Inference*. Chapman and Hall.

[19] ŠMULIAN, V. L. (1944). On compact sets in the space of measurable functions. *Mat. Sbornik* 57 (343-356) (Russian).

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