Proceedings of the 7th International Conference on Applied Informatics
Eger, Hungary, January 28–31, 2007. Vol. 1. pp. 223–228.

Note on strong consistency of maximum likelihood estimators for dependent observations*

Erika Fülöp, Gyula Pap

Department of Applied Mathematics and Probability Theory
Faculty of Informatics, University of Debrecen, Hungary

1. Introduction

Consider a *statistical experiment* $\big(X, \mathcal{X}, \{\mathsf{P}_\theta : \theta \in \Theta\}\big)$, where $(X, \mathcal{X})$ is a measurable space, $\Theta \subset \mathbb{R}^p$ is a non-empty set and $\{\mathsf{P}_\theta : \theta \in \Theta\}$ is a family of probability measures on it. The set Θ is called the *parameter set*, an element $x \in X$ is called an *observation*, a measurable function $T : X \to \Theta$ is called a *statistic*.

Let $(\Omega, \mathcal{A}, \mathsf{P})$ be a probability space, and for each $\theta \in \Theta$, let $\xi^{(\theta)} : \Omega \to \mathbb{R}^d$ be a random vector (sample) such that its distribution P_θ on $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)\big)$ is absolutely continuous with a density function (likelihood function) $x \mapsto L(x;\theta)$ from $\mathbb{R}^d$ into $[0,\infty)$. Then $\Lambda(x;\theta) := \log L(x;\theta) \in [-\infty,\infty)$, $x \in \mathbb{R}^d$, is the loglikelihood function, where we put $\log 0 := -\infty$.

Consider the statistical experiment $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathsf{P}_\theta : \theta \in \Theta\}\big)$. The main task of the parameter estimation in this statistical experiment is to find a statistic $T : \mathbb{R}^d \to \Theta$ to estimate the true (but unknown) value $\theta_0 \in \Theta$ of the parameter based on a sample $\xi^{(\theta_0)}$ such that T is good in the sense that the random vector $T(\xi^{(\theta_0)}) := T \circ \xi^{(\theta_0)}$ is close to θ_0.

Definition 1.1. Based on an observation $x \in \mathbb{R}^d$, a **maximum likelihood estimator** for the parameter in the statistical experiment $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathsf{P}_\theta : \theta \in \Theta\}\big)$ is a value $\widehat{\theta}(x) \in \Theta$ with

$$L(x;\widehat{\theta}(x)) = \sup_{\theta\in\Theta} L(x;\theta). \tag{1.1}$$

We say that there exists a **measurable maximum likelihood estimator** for the parameter in the statistical experiment $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathsf{P}_\theta : \theta \in \Theta\}\big)$ if there exists a measurable function (statistic) $\widehat{\theta} : \mathbb{R}^d \to \Theta$ such that (1.1) holds for all $x \in \mathbb{R}^d$.

*This research has been supported by the Hungarian Scientific Research Fund under Grant No. OTKA–T048544/2005.

A maximum likelihood estimator for a statistical experiment does not necessarily exist, and even if it exists, it is not necessarily unique. If $\widehat{\theta} : \mathbb{R}^d \to \Theta$ is a measurable function then $\widehat{\theta} \circ \xi^{(\theta)} : \Omega \to \Theta$ is a random vector (a measurable function) for all $\theta \in \Theta$. The following lemma gives a sufficient condition for the existence of a measurable maximum likelihood estimator. The proof can be found, e.g. in Jennrich [3, Lemma 2].

Lemma 1.2. *If for each $x \in \mathbb{R}^d$, the function $\theta \mapsto L(x;\theta)$ is continuous on Θ, then $x \mapsto \sup_{\theta\in\Theta} L(x;\theta)$ is a measurable function on $\mathbb{R}^d$. If, in addition, Θ is compact, then there exists a measurable maximum likelihood estimator for the parameter in the statistical experiment $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \{\mathsf{P}_\theta : \theta \in \Theta\}\big)$.*

Now let d_n, $n \in \mathbb{N}$, be positive integers, and for each $n \in \mathbb{N}$ and $\theta \in \Theta$, let $\xi_n^{(\theta)} : \Omega \to \mathbb{R}^{d_n}$ be a random vector such that its distribution $\mathsf{P}_{n,\theta}$ on $\big(\mathbb{R}^{d_n}, \mathcal{B}(\mathbb{R}^{d_n})\big)$ is absolutely continuous with a density function $x \mapsto L_n(x;\theta)$ from $\mathbb{R}^{d_n}$ into $[0,\infty)$. Consider the sequence $\big(\mathbb{R}^{d_n}, \mathcal{B}(\mathbb{R}^{d_n}), \{\mathsf{P}_{n,\theta} : \theta \in \Theta\}\big)_{n\in\mathbb{N}}$ of statistical experiments.

Definition 1.3. For each $n \in \mathbb{N}$, let $T_n : \mathbb{R}^{d_n} \to \Theta$ be a measurable function.

The sequence $(T_n)_{n\in\mathbb{N}}$ is called a **strongly consistent estimator** of the true value $\theta_0 \in \Theta$ of the parameter if

$$T_n(\xi_n^{(\theta_0)}) \to \theta_0 \qquad \text{a.s. (almost surely)}$$

as $n \to \infty$, i.e.

$$\mathsf{P}\big\{\lim_{n\to\infty} T_n(\xi_n^{(\theta_0)}) = \theta_0\big\} = 1.$$

Assume that the parameter set $\Theta \subset \mathbb{R}^p$ is compact, and for each $n \in \mathbb{N}$ and $x \in \mathbb{R}^{d_n}$, the likelihood function $\theta \mapsto L_n(x;\theta)$ is continuous on Θ. Then, by Lemma 1.2, for each $n \in \mathbb{N}$, there exists a measurable maximum likelihood estimator $\widehat{\theta}_n : \mathbb{R}^{d_n} \to \Theta$ for the parameter in the statistical experiment $\big(\mathbb{R}^{d_n}, \mathcal{B}(\mathbb{R}^{d_n}), \{\mathsf{P}_{n,\theta} : \theta \in \Theta\}\big)$.

Heijmans and Magnus [2] mention without proof a sufficient condition for strong consistency of maximum likelihood estimators for the parameter in the sequence of statistical experiments $\big(\mathbb{R}^{d_n}, \mathcal{B}(\mathbb{R}^{d_n}), \{\mathsf{P}_\theta : \theta \in \Theta\}\big)_{n\in\mathbb{N}}$, which is contained in the following statement as the implication (ii) $\Rightarrow$ (iii). A neighborhood N of a point $\theta \in \Theta$ is an open subset of Θ which contains θ.

Theorem 1.4. *Let $\theta_0 \in \Theta$. Consider the following statements:*

(i) *there exists a sequence $(k_n)_{n\in\mathbb{N}}$ of real numbers with $\liminf_{n\to\infty} k_n > 0$, and for every $\theta \in \Theta \setminus \{\theta_0\}$, there exist a neighborhood $N(\theta,\theta_0)$ of θ and a quantity $I(\theta,\theta_0)$ such that $\inf_{\phi\,\in\, N(\theta,\theta_0)} I(\phi,\theta_0) > 0$ and*

$$\lim_{n\to\infty} \sup_{\phi\,\in\, N(\theta,\theta_0)} \left| \frac{1}{k_n}\big(\Lambda_n(\xi_n^{(\theta_0)};\phi) - \Lambda_n(\xi_n^{(\theta_0)};\theta_0)\big) + I(\phi,\theta_0) \right| = 0 \quad a.s. \quad (1.2)$$

(ii) *for every* $\theta \in \Theta \setminus \{\theta_0\}$, *there exist a neighborhood* $N(\theta, \theta_0)$ *of* θ *such that*

$$\limsup_{n\to\infty} \sup_{\phi \in N(\theta,\theta_0)} \big(\Lambda_n(\xi_n^{(\theta_0)}; \phi) - \Lambda_n(\xi_n^{(\theta_0)}; \theta_0)\big) < 0 \qquad a.s. \tag{1.3}$$

(iii) *every sequence* $(\widehat{\theta}_n)_{n\in\mathbb{N}}$ *of measurable maximum likelihood estimators is a strongly consistent estimtor of the true value* $\theta_0 \in \Theta$*;*

(iv) *for every neighborhood* N *of* θ_0 *we have*

$$\limsup_{n\to\infty} \left[\sup_{\phi \in \Theta\setminus N} \Lambda_n(\xi_n^{(\theta_0)}; \phi) - \sup_{\phi \in N} \Lambda_n(\xi_n^{(\theta_0)}; \phi)\right] \leqslant 0 \qquad a.s. \tag{1.4}$$

Then (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv).

Note that by Lemma 1.2, $\sup_{\phi \in N(\theta,\theta_0)} L_n(\xi_n^{(\theta_0)}; \phi)$ is a random variable with values in $[0, +\infty]$, hence $\sup_{\phi \in N(\theta,\theta_0)} \Lambda_n(\xi_n^{(\theta_0)}; \phi)$ is also a random variable with values in $[-\infty, +\infty]$. Moreover,

$$\sup_{\phi \in N(\theta,\theta_0)} \big(\Lambda_n(\xi_n^{(\theta_0)}; \phi) - \Lambda_n(\xi_n^{(\theta_0)}; \theta_0)\big) = \left(\sup_{\phi \in N(\theta,\theta_0)} \Lambda_n(\xi_n^{(\theta_0)}; \phi)\right) - \Lambda_n(\xi_n^{(\theta_0)}; \theta_0)$$

is defined almost surely, since $\mathsf{P}\{\Lambda_n(\xi^{(\theta_0)}; \theta_0) = -\infty\} = 0$. Indeed, this follows from $\mathsf{P}\{L_n(\xi^{(\theta_0)}; \theta_0) = 0\} = \int_{\{x\in\mathbb{R}^d \,:\, L_n(x;\theta_0)=0\}} L_n(x; \theta_0)\,\mathrm{d}x = 0$.

Note that (1.3) and (1.4) may also be written in the form

$$\limsup_{n\to\infty} \frac{1}{L_n(\xi_n^{(\theta_0)}; \theta_0)} \sup_{\phi \in N(\theta,\theta_0)} L_n(\xi_n^{(\theta_0)}; \phi) < 1 \qquad \text{a.s.} \tag{1.5}$$

and

$$\limsup_{n\to\infty} \frac{\sup_{\phi \in \Theta\setminus N} L_n(\xi_n^{(\theta_0)}; \phi)}{\sup_{\phi \in N} L_n(\xi_n^{(\theta_0)}; \phi)} \leqslant 1 \qquad \text{a.s.} \tag{1.6}$$

respectively.

The assumption (1.2) means that an almost sure expansion is valid uniformly in ϕ for $\Lambda_n(\xi_n^{(\theta_0)}; \phi) - \Lambda_n(\xi_n^{(\theta_0)}; \theta_0)$, namely,

$$\Lambda_n(\xi_n^{(\theta_0)}; \phi) - \Lambda_n(\xi_n^{(\theta_0)}; \theta_0) = -I(\phi, \theta_0)k_n + o(k_n) \qquad \text{a.s.}$$

as $n \to \infty$ uniformly in $\phi \in N(\theta, \theta_0)$. An obvious candidate for $I(\phi, \theta_0)k_n$ is the main term in the expansion of the Kullback-Leibler information, i.e.

$$\mathsf{E}\big(\Lambda_n(\xi_n^{(\theta_0)}; \phi) - \Lambda_n(\xi_n^{(\theta_0)}; \theta_0)\big) = -I(\phi, \theta_0)k_n + o(k_n).$$

as $n \to \infty$ (See e.g. Heijmans and Magnus [2].) In the standard i.i.d. case one may choose $k_n = n$, but in more general situations this will be not the case.

The main advantage of Theorem 1.4 is that the condition (i) can be checked in several situations, hence this is a useful tool for proving strong consistency of maximum likelihood estimators for dependent observations.

One can also find similar conditions for weak consistency, see e.g. Heijmans and Magnus [2], Fazekas [1].

2. Proof of Theorem 1.4

We start with showing (i) $\Rightarrow$ (ii). First observe that

$$\begin{aligned} &\frac{1}{k_n} \sup_{\phi \in N(\theta,\theta_0)} \big(\Lambda_n(\xi_n^{(\theta_0)};\phi) - \Lambda_n(\xi_n^{(\theta_0)};\theta_0)\big) \\ &\quad\leqslant \sup_{\phi \in N(\theta,\theta_0)} \left| \frac{1}{k_n}\big(\Lambda_n(\xi_n^{(\theta_0)};\phi) - \Lambda_n(\xi_n^{(\theta_0)};\theta_0)\big) + I(\phi,\theta_0) \right| - \inf_{\phi \in N(\theta,\theta_0)} I(\phi,\theta_0), \end{aligned}$$

hence (i) implies

$$\limsup_{n\to\infty} \frac{1}{k_n} \sup_{\phi \in N(\theta,\theta_0)} \big(\Lambda_n(\xi_n^{(\theta_0)};\phi) - \Lambda_n(\xi_n^{(\theta_0)};\theta_0)\big) \leqslant - \inf_{\phi \in N(\theta,\theta_0)} I(\phi,\theta_0) \qquad \text{a.s.}$$

Thus $\liminf_{n\to\infty} k_n > 0$ and $\inf_{\phi \in N(\theta,\theta_0)} I(\phi,\theta_0) > 0$ clearly yields (ii).

Now we prove (ii) $\Rightarrow$ (iii). For a subset $H \subset \Theta$, consider the random variable

$$S_n(H) := \frac{1}{L_n(\xi_n^{(\theta_0)};\theta_0)} \sup_{\phi\in H} L_n(\xi_n^{(\theta_0)};\phi),$$

defined almost surely. By (ii), for every $\theta \in \Theta \setminus \{\theta_0\}$, there exist a neighborhood $N(\theta,\theta_0)$ of θ such that

$$\limsup_{n\to\infty} S_n(N(\theta,\theta_0)) < 1 \qquad \text{almost surely.} \tag{2.1}$$

We have to show that for each neighborhood N of θ_0, there exists a random variable $n_0(N)$ with values in $\mathbb{N}$, such that $\widehat{\theta}_n\big(\xi_n^{(\theta_0)}\big) \in N$ for all $n \geqslant n_0(N)$ almost surely. The set $\Theta \setminus N$ is compact (since it is closed and Θ is compact), and

$$\bigcup_{\theta \in \Theta\setminus N} N(\theta,\theta_0) \supset \Theta \setminus N$$

is an open covering of $\Theta \setminus N$, hence there exists a finite subcovering of $\Theta \setminus N$, i.e. we can find finite number of points $\theta_1, \dots, \theta_r \in \Theta \setminus N$, such that

$$\bigcup_{k=1}^{r} N(\theta_k,\theta_0) \supset \Theta \setminus N.$$

We obtain

$$\sup_{\phi\,\in\,\Theta\setminus N} L_n(\xi_n^{(\theta_0)};\phi) \leqslant \sup_{\phi\,\in\,\bigcup_{k=1}^r N(\theta_k,\theta_0)} L_n(\xi_n^{(\theta_0)};\phi) = \max_{1\leqslant k\leqslant r} \sup_{\phi\,\in\,N(\theta_k,\theta_0)} L_n(\xi_n^{(\theta_0)};\phi).$$

Dividing by $L_n(\xi_n^{(\theta_0)};\theta_0)$ both sides of this inequality yields

$$\frac{1}{L_n(\xi_n^{(\theta_0)};\theta_0)} \sup_{\phi\,\in\,\Theta\setminus N} L_n(\xi_n^{(\theta_0)};\phi) \leqslant \max_{1\leqslant k\leqslant r} S_n(N(\theta_k,\theta_0)).$$

Hence by (2.1),

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{L_n(\xi_n^{(\theta_0)};\theta_0)} \sup_{\phi\,\in\,\Theta\setminus N} L_n(\xi_n^{(\theta_0)};\phi) &\leqslant \limsup_{n\to\infty} \max_{1\leqslant k\leqslant r} S_n(N(\theta_k,\theta_0))\\ &= \max_{1\leqslant k\leqslant r} \limsup_{n\to\infty} S_n(N(\theta_k,\theta_0)) < 1\end{aligned}$$

almost surely. Consequently, there exists a random variable $n_0(N)$ with values in $\mathbb{N}$, such that

$$\sup_{\phi\,\in\,\Theta\setminus N} L_n(\xi_n^{(\theta_0)};\phi) < L_n(\xi_n^{(\theta_0)};\theta_0) \qquad \text{for all } n \geqslant n_0(N) \text{ almost surely.}$$

For each $n\in\mathbb{N}$ and $x\in\mathbb{R}^d$, the inequality $\sup_{\phi\,\in\,\Theta\setminus N} L_n(x;\phi) < L_n(x;\theta_0)$ implies $\widehat{\theta}_n(x)\in N$ since $\widehat{\theta}_n$ is a maximum likelihood estimator. Consequently,

$$\widehat{\theta}_n(\xi_n^{(\theta_0)}) \in N \qquad \text{for all } n \geqslant n_0(N) \text{ almost surely,}$$

thus we obtain (iii).

Finally we show (iii) $\Rightarrow$ (iv). By (iii), for an arbitrary neighborhood N of θ_0, there exists a random variable $n_0(N)$ with values in $\mathbb{N}$, such that $\widehat{\theta}_n\big(\xi_n^{(\theta_0)}\big)\in N$ for all $n\geqslant n_0(N)$ almost surely, hence

$$\sup_{\phi\,\in\,\Theta\setminus N} L_n(\xi_n^{(\theta_0)};\phi) < \sup_{\phi\,\in\,N} L_n(\xi_n^{(\theta_0)};\phi) \qquad \text{for all } n \geqslant n_0(N) \text{ almost surely.}$$

Consequently,

$$\sup_{\phi\,\in\,\Theta\setminus N} \Lambda_n(\xi_n^{(\theta_0)};\phi) - \sup_{\phi\,\in\,N'} \Lambda_n(\xi_n^{(\theta_0)};\phi) < 0 \qquad \text{for all } n \geqslant n_0(N) \text{ almost surely,}$$

hence we conclude (iv).

References

[1] Fazekas, I., Asymptotic properties of maximum likelihood estimators of parameters of a spatio-temporal econometric model, *Theory of Stoch. Process.,* 2(18), No. 1-2, (1996), 124–136.

[2] HEIJMANS, R. D. H., MAGNUS, J. R., Consistent maximum likelihood estimation with dependent observations, *J. Econometrics,* 32, (1986), 253–285.

[3] JENNRICH, R. I., Asymptotic properties of non-linear least squares estimators, *Ann. Math. Statist.*, 40(2), (1969), 633–643.