Georgian Math. J. ? (????), 1–13
DOI 10.1515/GMJ.????.???

On the limit properties of maximal likelihood estimators in a Hilbert space

Petre Babilua, Elizbar Nadaraya and Grigol Sokhadze

Abstract. For infinite-dimensional Hilbert space is formulated maximal likelihood principle. In this paper is proved theorems about consistency and asymptotical normality.

Keywords. Maximum likelihood estimation, infinite dimensional spaces, consistency, asymptotical normality.

2010 Mathematics Subject Classification. 28A15, 62F12, 62B15, 62C15, 62F10.

1 Introduction

The method of maximal likelihood estimators (MLEs) is one of the fundamental methods for parametric estimation. It was proposed by R. Fisher and was used in the course of many decades both in applied statistics and in theoretical studies. The properties of MLEs were studied in a series of works. In the recent papers [1], [2], the maximal likelihood principle is formulated for infinite-dimensional spaces. Many statistical problems, in particular estimation problems, are of infinite-dimensional nature and the extension of this method to such spaces is important. Due to Malliavin's calculus and, speaking more generally, the smooth measure calculus it became possible to generalize the MLE principle to the infinite-dimensional case (see [3]).

In the present paper, using the definitions and methods of [1], [2] the limit properties of MLEs are investigated for Hilbert (generally speaking, infinite-dimensional) spaces.

Let $\mathbb{E}$ be a linear space, $\mathbb{B}$ be a separable real Banach space with norm $\|\mathbf{b}\|_{\mathbb{B}}$, $\mathbf{b} \in \mathbb{B}$, Θ be a subspace of the space $\mathbb{B}$, $\Theta \subset \mathbb{B}$. Θ plays the role of a parametric set. $\{\Omega, \mathfrak{F}, P\}$ is a complete probability space. We are going to consider a random element $X = X(\omega, \theta)$ on $\Omega \times \Theta$ with values in $\mathbb{E}$. If on $\mathbb{E}$ we define a σ-algebra $\mathfrak{B}$ of its subsets, such that for each $\theta \in \Theta$ and measurable set $E \subset \mathbb{E}$ we have $X^{-1}(E) \in \mathfrak{F}$, then the variable X on $\mathbb{E}$ generates the class of measures $\{p_\theta,\ \theta \in \Theta\}$. These measures are defined by the relation $p_\theta(E) = P(X^{-1}(E))$. The set $\{p_\theta,\ \theta \in \Theta\}$ is called the distribution of a random element $X = X(\omega, \theta)$.

We perform observations of a random element X. As a result we obtain the

sampling which is a sequence of independent random variables $X_1, X_2, \ldots, X_n$. The distribution of each of these variables coincides with the distribution of X. The parameter θ is assumed to be an unknown variable and, based on observations, it can be evaluated by means of some measurable function, i.e. by the statistics $\widehat{\theta}_n = T = T(X_1, X_2, \ldots, X_n)$. The meaning of optimality is specified separately. In any case, the estimator $\widehat{\theta}_n$ of the unknown parameter θ is considered to be "good" (consistent) if $\widehat{\theta}_n$ infinitely tends to θ as $n \to \infty$ in the specified sense. We obtain, generally speaking, a sequence of statistical structures $\{\mathbb{E}^n, \mathfrak{B}^n, \{P_\theta,\ \theta \in \Theta\}\}$. At the same time, $\mathfrak{B}^n$ has the meaning of a set of observed sets, while $\{P_\theta,\ \theta \in \Theta\} = \{p_\theta,\ \theta \in \Theta\}^n$. In other words, if $Y = (X_1, \ldots, X_n)$, then $P_\theta(A) = P(Y^{-1}(A))$, $A \in \mathfrak{B}^n$. The structure $\{\mathbb{E}^n, \mathfrak{B}^n, \{P_\theta,\ \theta \in \Theta\}\}$ is called the statistical structure of an iterated sampling. In statistics, this structure $\{\mathbb{E}^n, \mathfrak{B}^n, \{P_\theta, \theta \in \Theta\}\}$ is the object of investigation.

For some problems it is more convenient to use the function $X = X(\omega, \theta)$ the more so that the achievements (in particular Malliavin's stochastic variational calculus and the properties of smooth measures) of stochastic analysis have recently made it possible to study some analytical properties of a random function $X(\omega, \theta)$.

Thus, we can use the double calculus: one of them is based on the study of such a property of the statistical structure $\{\mathbb{E}^n, \mathfrak{B}^n, \{P_\theta,\ \theta \in \Theta\}\}$ as the smoothness of the family of measures P_θ, and the other employs direct stochastic methods, for which the object of the study is $X(\omega, \theta)$. In particular, we are interested in the family of distributions $\{P_\theta(A),\ \theta \in \Theta, A \in \mathfrak{B}^n\}$ from the viewpoint of smoothness with respect to both parameters θ and A. Below we give some definitions, notations and properties.

2 Differentiable measures

In the sequel it will always be assumed that $\mathbb{E}$ is a real, separable, reflexive Banach space. $P_\theta(\,\cdot\,)$ is a positive measure for each fixed $\theta \in \Theta$. If $h \in \mathbb{E}^n$ is some vector, then $P_{\theta,h}(A)$ denotes the measure obtained by means of bias:

$$P_{\theta,h}(A) = P_\theta(A + h).$$

We say that the measure $P_\theta(\,\cdot\,)$ is differentiable along a vector $h \in \mathbb{E}^n$, if there exists a bounded linear functional $d_h P_\theta$, such that the equality

$$P_{\theta,h}(A) - P_\theta(A) = d_h P_\theta(A)h + O(\|h\|^2)$$

is fulfilled for each $A \in \mathfrak{B}^n$. (For the properties of differentiable measures see [4].)

When $\mathbb{E}$, and thereby $\mathbb{E}^n$, too, is a separable real Hilbert space with scalar product $\langle x, y\rangle_{\mathbb{E}^n}$, $x, y \in \mathbb{E}^n$, and norm $\|x\|_{\mathbb{E}^n}$, $x \in \mathbb{E}^n$, we write

$$P_{\theta,h}(A) - P_\theta(A) = \big\langle d_h P_\theta(A), h\big\rangle_{\mathbb{E}^n} + O\big(\|h\|^2_{\mathbb{E}^n}\big)$$

and under the derivative we mean an element of the Hilbert space $\mathbb{E}^n$. It is understood that the function $d_h P_\theta(\,\cdot\,)h$ is a σ-additive (alternating) measure on $\mathfrak{B}^n$.

A higher order derivative of the measure is defined by iteration of the definition of a derivative. Thus, for instance, $d_k d_h P_\theta = d_k(d_h P_\theta h)k$, $k, h \in \mathbb{E}^n$. In particular for the Hilbert space we have

$$\big\langle d^{(2)}_{h,h} P_\theta h, h\big\rangle_{\mathbb{E}^n} = \big\langle d_h\langle d_h P_\theta, h\rangle_{\mathbb{E}^n}, h\big\rangle_{\mathbb{E}^n}.$$

If $P_\theta(\,\cdot\,)$ is a differentiable measure, then the function $\varphi_\theta(t) = P_\theta(A + th)$ is non-negative and everywhere differentiable with respect to t. If the set A is such that $P_\theta(A) = 0$, then the point $t = 0$ is the point of minimum for the function $\varphi_\theta(t)$. Therefore

$$\frac{d}{dt}\,\varphi_\theta(t)\Big|_{t=0} = 0,$$

i.e. $d_h P_\theta(A) = 0$. Thus $d_h P_\theta \ll P_\theta$. According to the Radon–Nikodym theorem, there exists a measurable function $\beta_\theta(x; h)$, such that

$$\frac{d_h P_\theta(dx)}{P_\theta(dx)} = \beta_\theta(x; h).$$

This function is called the logarithmic derivative of the measure P_θ along a vector $h \in \mathbb{E}^n$. The logarithmic derivative $\beta_\theta(x; h)$ is linear with respect to the second argument. A vector h is called an admissible direction for the measure P_θ. The set of all admissible directions is called an admissible subspace.

For the sake of simplicity, our argumentation in the sequel will always involve the space $\mathbb{E}$ because all the definitions and properties are automatically applicable to the space $\mathbb{E}^n$ as a direct product of spaces.

In the theory of differentiable measures, the validity of the formula of integration by parts is an important fact. Let $\mathbb{E}$ be a separable, real Hilbert space and $f(x)$ be the functional on $\mathbb{E}$. Let us assume that there exists its derivative along a vector $h \in \mathbb{E}$:

$$d_h f(x) = \lim_{t\to 0} \frac{f(x + th) - f(x)}{t} = \langle f'(x), h\rangle_{\mathbb{E}},$$

and $d_h f \in \mathcal{L}_1(P_\theta)$ for $\theta \in \Theta$. Then if the measure P_θ is differentiable along h, we have

$$\int_{\mathbb{E}} \langle f'(x), h\rangle_{\mathbb{E}} P_\theta(dx) = -\int_{\mathbb{E}} f(x)\beta_\theta(x; h) P_\theta(dx).$$

The logarithmic derivative of the measure can be defined along non-constant vectors (this is the so-called logarithmic gradient). Let $z(x) : \mathbb{E} \to \mathbb{E}$ be the differentiable vector field with bounded derivative $\sup\limits_{x\in\mathbb{E}} \|z'(x)\| < \infty$. Denote by S_t, $t \in R$ the integral flow corresponding to $z(x)$. This means that

$$\frac{dS_t}{dt} = z(S_t), \quad S_0 = I.$$

By the transformation $P_\theta^t(A) = P_\theta(S_t^{-1}(A))$, $A \in \mathfrak{B}$, to the family of measures $\{P_\theta,\ \theta \in \Theta\}$ there corresponds the class of measures $\{P_\theta^t,\ \theta \in \Theta,\ t \in R\}$.

We say that the measure P_θ is differentiable along the vector field $z(x)$ if there exists a (by all means alternating) measure (alternating) $D_z P_\theta$, such that for any bounded and differentiable function $\varphi : \mathbb{E} \to E$, $\varphi \in C^1(\mathbb{E}; R)$ we have

$$\int\limits_{\mathbb{E}} \varphi(x) D_z P_\theta(dx) = -\lim_{t\to 0} \int\limits_{\mathbb{E}} \varphi(x) \frac{P_\theta^t - P_\theta}{t}(dx) = -\int\limits_{\mathbb{E}} \varphi'(x) z(x) P_\theta(dx).$$

If at the same time $D_z P_\theta \ll P_\theta$, then the corresponding Radon–Nikodym density is called the logarithmic derivative P_θ along the vector field $z(x)$:

$$\beta_\theta(x; z) = \frac{D_z P_\theta(dx)}{P_\theta(dx)}.$$

Let $\mathbb{H} \subset \mathbb{E}$ be the Hilbert space embedded in $\mathbb{E}$, and let the embedding operator be the Hilbert-Schmidt operator. Then we can consider the Hilbert-Schmidt structure $\mathbb{E}^* \subset \mathbb{H} \subset \mathbb{E}$. Let us define an important class of measures $\mathfrak{M}$ on $\mathbb{E}$, for which there exists a measurable, locally bounded function $\ell : \mathbb{E} \to \mathbb{E}$, such that for each constant vector $h \in \mathbb{E}^*$ there exists a logarithmic derivative of the measure P_θ along that has the form

$$\beta_\theta(x; h) = \ell(\theta; x)h = \big\langle \ell(\theta; x), h \big\rangle_{\mathbb{H}}.$$

In that case, we also say that the measure possesses the logarithmic gradient $\ell(\theta; x)$.

As is known (see [5]), if $P_\theta \in \mathfrak{M}$ and the vector field $z : \mathbb{E} \to \mathbb{E}^*$ is bounded together with its derivative, then the measure P_θ possesses the logarithmic gradient and the representation

$$\beta_\theta(x; z(x)) = \big\langle \ell(\theta; x), z(x) \big\rangle_{\mathbb{H}} + tr z'(x)$$

is fulfilled. This continuity functional can be extended for smooth vector fields $z(x) : \mathbb{E} \to \mathbb{H}$ as a measurable linear functional. In stochastic analysis, for Gaussian spaces this functional is an extended Skorokhod stochastic integral or, which is the same, the conjugate of the Malliavin derivative.

Let us present some well-known properties of the logarithmic derivative.

Proposition 1. *Let the following conditions be fulfilled:*

(i) *Measures P_θ are differentiable along a vector $h \in \mathbb{E}$;*

(ii) *Functions f and g are differentiable along $h \in \mathbb{E}$;*

(iii) *$f, g \in \mathcal{L}_1(d_h P)$ and $f'(x)h$, $g'(x)h \in \mathcal{L}_1(P)$, $f(x)g(x)\beta_\theta(x;h) \in \mathcal{L}_1(P)$.*

Then

$$\int_{\mathbb{E}} (f'(x)h)g(x)P_\theta(dx) = -\int_{\mathbb{E}} f(x)(g'(x)h)P_\theta(dx) - \int_{\mathbb{E}} f(x)g(x)\beta_\theta(x;h)P_\theta(dx).$$

Proposition 2. *Let the measures P_θ be differentiable along a vector $h \in \mathbb{E}$, the function $\varphi(t) = \beta_\theta(x+th;h)$ be everywhere differentiable and*

$$\varphi'(0) = \beta'_\theta(x;h)h \in \mathcal{L}_2(P_\theta).$$

Then

(i) *The measure P_θ is twice differentiable along h;*

(ii) $d^2_{h,h}P_\theta = \{\beta'_\theta(x;h)h + \beta^2_\theta(x;h)\}P_\theta$;

(iii) $\int\limits_{\mathbb{E}} \beta^2_\theta(x;h)P_\theta(dx) = -\int\limits_{\mathbb{E}} \beta'_\theta(x;h)P_\theta(dx)$.

3 Smoothness with respect to the parameter

We need to investigate the smoothness of the family of measures with respect to the parameter. Suppose that as above we have the statistical structure $\{\mathbb{E}, \mathfrak{B}, P_\theta,\ \theta \in \Theta\}$, where $\mathbb{E}$ is a separable real Banach space, and Θ is a compactum into the other separable real Banach space $\mathbb{B}$. For any fixed $A \in \mathfrak{B}$ and vector $\vartheta \in \mathbb{B}$ we consider the function derivative $\psi(\theta) = P_\theta(A)$ at a point θ along ϑ. This derivative is denoted by $d_\theta P_\theta(A)\vartheta$. For fixed θ and ϑ, it is an alternating measure. It is easy to verify that $d_\theta P_\theta \vartheta \ll P_\theta$ and by the Radon–Nikodym theorem there exists a measurable function

$$\rho_\theta(x;\vartheta) = \frac{d_\theta P_\theta(dx)\vartheta}{P_\theta(dx)}.$$

$\rho_\theta(x;\vartheta)$ is called the logarithmic derivative of the measure P_θ with respect to the parameter.

When $\mathbb{B}$ is a separable Hilbert space, we denote $\mathcal{M}$ by the space of measures, for which the logarithmic derivative with respect to the parameter can be represented as a scalar product $\rho_\theta(x;\vartheta) = \langle r(x;\theta), \vartheta\rangle_{\mathbb{B}}$. We call $r(x;\theta)$ a vector logarithmic gradient with respect to the parameter.

Remark. We should make a special remark that if Θ consists of one point $\Theta = \{\theta_0\}$, then we obtain the family of measures which is constant with respect to the parameter. In that case, the derivative with respect to the parameter is equal to 0: $r(x;\theta_0) = 0$. We will need this obvious fact in the sequel.

For the family of measures $\{P_\theta,\ \theta \in \Theta\}$, which possesses the logarithmic derivative with respect to the parameter along ϑ, there exists a measure μ that dominates this family. If Θ is a real interval, then, as is known ([6]), all measures P_θ are mutually equivalent and

$$\frac{P_{\theta_2}(dx)}{P_{\theta_1}(dx)} = \exp \int_{\theta_1}^{\theta_2} \rho_\theta(x;1)\, d\theta, \quad \theta_1, \theta_2 \in \Theta.$$

4 Regularity conditions

Let us list the so-called regularity conditions, the fulfillment of which is needed for our further discussion.

Condition 1. For a random element $X = X_\theta = X_\theta(\omega) : \Theta \times \Omega \to \mathbb{E}$ there exists a derivative $\frac{d}{d\theta} X_\theta = X'$ along $\vartheta \in \mathbb{B}_0 \subset \mathbb{B}$, where $\mathbb{B}_0$ is a subspace of $\mathbb{B}$. This derivative is the linear mapping $\mathbb{B} \to \mathbb{E}$ for each $\theta \in \Theta$. Also, for any $\vartheta \in \mathbb{B}_0$, $\theta \in \Theta$ we have $\|X'\vartheta\|_{\mathbb{E}} \in \mathcal{L}_2(\Omega, P)$.

Condition 2. The function $f(x) = E\{X'\vartheta|\ X = x\}$ is strongly continuous for all $\vartheta \in \mathbb{B}_0$, $\theta \in \Theta$.

Condition 3. The family of measures $\{P_\theta,\ \theta \in \Theta\}$ possesses a logarithmic derivative with respect to the parameter along constant vectors from a dense in $\mathbb{B}$ subspace $\mathbb{B}_0 \subset \mathbb{B}$ and $\rho_\theta(x;\vartheta) \in \mathcal{L}_2(\mathbb{E}, P_\theta)$, $\vartheta \in \mathbb{B}_0$, $\theta \in \Theta$.

Condition 4. The family of measures $\{P_\theta,\ \theta \in \Theta\}$ possesses the logarithmic derivative $\beta_\theta(x;h)$ along constant vectors from a dense in $\mathbb{E}$ subspace $\mathbb{E}_0 \subset \mathbb{E}$ and $\beta_\theta(x;h) \in \mathcal{L}_2(\mathbb{E}, P_\theta)$, $h \in \mathbb{E}_0$, $\theta \in \Theta$.

Condition 5. The statistics $T = T(x) : \mathbb{E} \to R$ is such that the equality

$$d_\vartheta \int\limits_{\mathbb{E}} T(x) P_\theta(dx) = \int\limits_{\mathbb{E}} T(x) d_\vartheta P_\theta(dx)$$

is valid.

It is important to note that there exists an analytic connection between the introduced notion of a distribution derivative with respect to the spatial argument and the notion of a derivative with respect to the parameter. Hence the following proposition is true.

Proposition 3 ([1], [2]). *Under regularity conditions* 1–4*, for the logarithmic derivatives* $\beta_\theta(x; h)$ *and* $\rho_\theta(x; \vartheta)$ *we have the equality*

$$\rho_\theta(x; \vartheta) = -\beta_\theta(x; K_{\theta,\vartheta}(x)),$$

where

$$K_{\theta,\vartheta}(x) = E\Big\{ \Big(\frac{d}{d\theta} X \Big) \vartheta \Big|\ X = x \Big\}.$$

5 Maximal likelihood principle

Let $\{\mathbb{E}, \mathfrak{B}, P_\theta, \theta \in \Theta\}$ be the statistical structure corresponding to a random element $X = X_\theta$. Here $\mathbb{E}$ is a separable, real, reflexive Banach space, $\mathfrak{B}$ is the σ-algebra of Borel sets. $\Theta \subset \mathbb{B}$ is a compact subset of the separable real Banach space $\mathbb{B}$. We remind that regularity conditions 1–5 are assumed to be fulfilled.

Let $g(\theta) = E_\theta T(X)$, where $T : \mathbb{E} \to R$ is a measurable mapping (statistics). The derivative of the function $g(\theta)$ along the vector ϑ is denoted $g'_\vartheta(\theta)$. The following proposition is true.

Proposition 4 (Cramer–Rao inequality) [1], [2]). *If regularity conditions* 1–5 *are fulfilled, then*

$$\operatorname{Var} T(X) \geq \frac{(g'_\vartheta(\theta))^2}{E_\theta \rho^2_\theta(X; \vartheta)} = \frac{(g'_\vartheta(\theta))^2}{E_\theta \beta^2_\theta(X; E(X'_\theta \vartheta|\ X))}\,.$$

Definition. The value $E_\theta \rho^2_\theta(X; \vartheta)$ is called the Fisher information along ϑ and denoted by $\mathcal{I}(\theta)\vartheta$. Therefore

$$\mathcal{I}(\theta)\vartheta = E_\theta \rho^2_\theta(X; \vartheta) = E_\theta \beta^2_\theta\big(X; E(X'_\theta \vartheta|\ X)\big).$$

Let us consider the structure of the iterated sampling

$$\big\{\mathbb{E}^n, \mathfrak{B}^n, \{P_\theta, \theta \in \Theta\}\big\} = \big\{\mathbb{E}, \mathfrak{B}, \{p_\theta, \theta \in \Theta\}\big\}^n.$$

Theorem 1 ([1], [2]). *If there exists the logarithmic derivative $\rho_\theta(x;\vartheta)$ of the family p_θ with respect to the parameter in the statistical structure $\{\mathbb{E}, \mathfrak{B}, \{p_\theta, \theta \in \Theta\}\}$, then for the iterated sampling structure $\{\mathbb{E}^n, \mathfrak{B}^n, \{P_\theta, \theta \in \Theta\}\}$ there also exists the logarithmic derivative $L_\theta((x_1,\dots,x_n);(\vartheta,\dots,\vartheta))$ of the family P_θ with respect to the parameter, along $(\vartheta,\dots,\vartheta)$ and*

$$L_\theta\big((x_1,\dots,x_n);(\vartheta,\dots,\vartheta)\big) = \sum_{k=1}^n \rho_\theta(x_k,\vartheta) = -\sum_{k=1}^n \beta_\theta\Big(x_k; E\big\{X_k'\vartheta \big|\ X_k = x_k\big\}\Big).$$

Using this theorem, we can formulate the maximal likelihood principle in the considered case.

In the sequel, it will always be assumed that $\mathbb{E}$ and $\mathbb{B}$ are Hilbert spaces. Let $X_1,\dots,X_n$ be the sampling from a random element $X = X_\theta$. θ is the unknown parameter which we have to estimate by means of the sampling. Assume further that there exists the logarithmic derivative $\rho_\theta(x;\vartheta)$ with respect to the parameter, along any vectors $\vartheta \in \mathbb{B}_0$, of the distribution P_θ corresponding to X_θ. The derivative has the form $\rho_\theta(x;\vartheta) = \langle r(x;\theta),\vartheta\rangle_{\mathbb{B}}$. Here $\mathbb{B}_0$ is a dense subset of $\mathbb{B}$.

Assuming the existence and uniqueness conditions to be fulfilled, we call the root of the equation

$$\sum_{k=1}^n \rho_\theta(x_k;\vartheta) = 0 \ \ \forall\,\vartheta \in \mathbb{B}_0 \tag{1}$$

the maximal likelihood estimator $\widehat{\theta}_n$ along the direction $\vartheta \in \mathbb{B}_0$ with respect to θ, if the operator $\frac{d}{d\theta}\,\rho_\theta(x;\vartheta)$ is negatively defined.

By Proposition 3, equation (1) can be replaced by

$$\sum_{k=1}^n \left\{ \big\langle \ell(x_k;\theta), K_{\theta,\vartheta}(x_k)\big\rangle_{\mathbb{H}} + tr\,\frac{d}{dx}\,K_{\theta,\vartheta}(x_k)\vartheta \right\} = 0 \ \ \forall\,\vartheta \in \mathbb{B}_0. \tag{2}$$

Note that in equations (1) and (2), x_k, $k = 1,2,\dots,n$, are the observed values of X_k in the experiment.

Example. Suppose that the sampling $X_1, X_2, \dots, X_n$ of a Gaussian value with unknown mean θ and unit correlation operator in $\mathbb{H}$ is considered in the equipped Hilbert space $\mathbb{E}^* \subset \mathbb{H} \subset \mathbb{E}$. We obtain $\beta_\theta(x;h) = \langle \theta - x, h\rangle_{\mathbb{H}}$, $h \in \mathbb{E}^*$. It is

obvious that $X_\theta = N + \theta$, where N is a canonical Gaussian value with zero mean. $X'(\theta) = I$, $X'(\theta)h = h$ and therefore

$$K_{\theta,\vartheta}(x_k) = E\big\{X'_k(\theta)h|\ X_k(\theta) = x_k\big\} = h.$$

Thus (2) takes the form

$$\sum_{k=1}^{n} \langle \theta - x_k, h\rangle_{\mathbb{H}} = 0.$$

Hence

$$\langle \widehat{\theta}, h\rangle_{\mathbb{H}} = \frac{1}{n}\sum_{k=1}^{n} \langle x_k, h\rangle_{\mathbb{H}}$$

and therefore

$$\widehat{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{x}.$$

Also,

$$\sum_{k=1}^{n} \frac{d_h}{d\theta}\,\langle x - \theta, h\rangle_{\mathbb{H}} = -\|h\|^2_{\mathbb{H}} \le 0.$$

6 Consistency of the maximal likelihood estimator

Let the statistical structure $\{\mathbb{E}, \mathfrak{B}, P_\theta, \theta \in \Theta\}$ be such that P_θ possesses the logarithmic derivative $\rho_\theta(x;\vartheta)$ with respect to the parameter along a constant vector $\vartheta \in \mathbb{B}_0$, where $\mathbb{B}_0$ is a subspace of $\mathbb{B}$.

Let us introduce the Kullback–Leibler type distance function for a pair of measures:

$$D(\theta_1, \theta_2) = E_{\theta_1}\Big\{\rho_{\theta_1}(X; \theta_2 - \theta_1) - \rho_{\theta_2}(X; \theta_2 - \theta_1)\Big\}. \tag{1}$$

For example, in the equipped Hilbert space $\mathbb{E}^* \subset \mathbb{H} \subset \mathbb{E}$, for the Gaussian measures μ_1 and μ_2 with means θ_1 and θ_2, respectively, and unit correlation operators the distance is

$$D(\mu_1, \mu_2) = -\big\langle \theta_1 - \theta_2, \theta_1 - \theta_2\big\rangle_{\mathbb{H}} = -\|\theta_1 - \theta_2\|^2_{\mathbb{H}}.$$

Lemma. *Let a family $\{P_\theta, \theta \in \Theta\} \in \mathcal{M}$ be uniquely defined by the parameter, i.e. $[P_{\theta_1} = P_{\theta_2}] \Longleftrightarrow [\theta_1 = \theta_2]$ and the logarithmic derivative $r(x;\theta)$ have the continuous negative-definite derivative with respect to θ. In that case, if $D(\theta_1, \theta_2) \ge 0$, then $P_{\theta_1} = P_{\theta_2}$ and vice versa.*

Proof. We at once obtain that $D(\theta_1,\theta_2)\le 0$. Indeed,

$$\begin{aligned}
D(\theta_1,\theta_2) &= E_{\theta_1}\Big\{\rho_{\theta_1}(X;\theta_2-\theta_1)-\rho_{\theta_2}(X;\theta_2-\theta_1)\Big\}\\
&= E_{\theta_1}\Big\{\big\langle r(x;\theta_1),\theta_2-\theta_1\big\rangle_{\mathbb{B}}-\big\langle r(x;\theta_2),\theta_2-\theta_1\big\rangle_{\mathbb{B}}\Big\}\\
&= E_{\theta_1}\Big\langle r(x;\theta_1)-r(x;\theta_2),\theta_2-\theta_1\Big\rangle_{\mathbb{B}}\\
&= E_{\theta_1}\Big\langle r'_\theta\big(X;\theta_1+\tau(\theta_2-\theta_1)\big)(\theta_2-\theta_1),\theta_2-\theta_1\Big\rangle_{\mathbb{B}}\\
&\le 0,\quad 0\le\tau\le 1.
\end{aligned}$$

Hence it follows that if $D(\theta_1,\theta_2)\ge 0$, then $\theta_1=\theta_2$, which implies $P_{\theta_1}=P_{\theta_2}$ and vice versa. □

Theorem 2. *If Θ is a compact set and $\{P_\theta,\theta\in\Theta\}\in\mathcal{M}$ is uniquely defined by the parameter, then the maximal likelihood estimator is consistent.*

Proof. Let $\widehat{\theta}_n$ be the solution of equation (1) (or (2)) and θ_0 be the true value of the parameter θ. According to the strong law of large numbers, for any θ we have

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}\big\langle r(X_k;\theta),\theta_0-\theta\big\rangle_{\mathbb{B}}-\frac{1}{n}\sum_{k=1}^{n}\big\langle r(X_k;\theta_0),\theta_0-\theta\big\rangle_{\mathbb{B}}\\
&\quad=\frac{1}{n}\sum_{k=1}^{n}\Big\langle r(X_k;\theta)-r(X_k;\theta_0),\theta_0-\theta\Big\rangle_{\mathbb{B}}\\
&\qquad\xrightarrow{a.s.}E_{\theta_0}\Big\{\big\langle r(X;\theta)-r(X;\theta_0),\theta_0-\theta\big\rangle_{\mathbb{B}}\Big\}\\
&\qquad=-D(\theta_0,\theta)\ge 0. \qquad (2)
\end{aligned}$$

Now we take into account that

$$\sum_{k=1}^{n}\big\langle r(x_k;\widehat{\theta}_n),\vartheta\big\rangle_{\mathbb{B}}=0,\quad \vartheta\in\mathbb{B}_0\ \text{ (by equation (1)),}$$

$$\sum_{k=1}^{n}\big\langle r(x_k;\theta_0),\vartheta\big\rangle_{\mathbb{B}}=0,\quad \vartheta\in\mathbb{B}_0\ \text{ (by the Remark in Section 3).}$$

Then equality (2) is fulfilled if $\theta=\lim\limits_{n\to\infty}\widehat{\theta}_n$. □

7 Asymptotic normality of the maximal likelihood estimator

Theorem 3. *If regularity conditions* 1–5 *are fulfilled, then for the maximal likelihood estimator* $\widehat{\theta}_n$ *we have the asymptotic convergence to the normal law*

$$\sqrt{n}\,(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}(\theta)).$$

Proof. Denote

$$-\frac{1}{n}\sum_{k=1}^{n} \rho_\theta(X_k;\vartheta) = L_n(\theta),$$

then

$$-\frac{1}{n}\sum_{k=1}^{n} \frac{d}{d\theta}\,\rho_{\,\theta}(X_k;\vartheta) = L_n'(\theta).$$

We have

$$L_n(\widehat{\theta}_n) - L_n(\theta_0) = L_n'(\widetilde{\theta})(\widehat{\theta}_n - \theta_0),$$

where

$$\widetilde{\theta} = \widehat{\theta}_n + \tau(\theta_0 - \widehat{\theta}_n), \quad 0 \le \tau \le 1.$$

Since $L_n(\widehat{\theta}_n) = 0$, we obtain

$$\sqrt{n}\,(\widehat{\theta}_n - \theta_0) = -\sqrt{n}\,(L_n'(\widetilde{\theta}))^{-1} L_n(\theta_0). \tag{1}$$

By the strong law of large numbers we have

$$L_n'(\theta) \xrightarrow{a.s.} E_{\theta_0}\,\frac{d}{d\theta}\,\rho_{\,\theta}(X;\vartheta)$$

but

$$E_{\theta_0}\,\frac{d}{d\theta}\,\rho_{\,\theta}(X;\vartheta)\vartheta = E_{\theta_0}\rho_\theta^2(X;\vartheta) = \mathcal{I}(\theta)\vartheta.$$

We also take into account that, by the consistency theorem, $\widetilde{\theta} \to \theta_0$ as $n \to \infty$. Finally, we study the expression $L_n(\theta_0)$. It consists of independent, equally distributed random values, for which $E_{\theta_0}\rho_\theta(X_k;\vartheta) = 0$ and $\operatorname{Var}\rho_\theta(X_k;\vartheta) = \mathcal{I}(\theta)\vartheta$. According to the central limit theorem,

$$-\sqrt{n}\,L_n(\theta_0) \xrightarrow{d} N(0;\mathcal{I}(\theta)).$$

Then from (1) we obtain

$$\sqrt{n}\,(\widehat{\theta}_n - \theta_0) = -\sqrt{n}\,(L_n'(\widetilde{\theta}))^{-1} L_n(\theta_0) \xrightarrow{d} N(0;\mathcal{I}^{-1}(\theta)). \qquad \square$$

Acknowledgement

The ideas of the proof of Theorems 2 and 3 are borrowed from the unpublished works of splendid mathematician Guy Lebanon (G. Lebanon, Consistency of the Maximum Likelihood Estimator, 2008; G. Lebanon, Asymptotic Efficiency of the Maximum Likelihood Estimator, 2009).

Bibliography

[1] E. Nadaraya, P. Babilua, M. Patsatsia, and G. Sokhadze, On the Cramer–Rao inequality in an infinite dimensional space, *Bull. Georgian Natl. Acad. Sci. (N.S.)* **6**:1 (2012), 5–13.

[2] E. Nadaraya, O. Purtukhia, and G. Sokhadze, On the Cramer–Rao inequality in an infinite dimensional space, *Proc. A. Razmadze Math. Inst.* **160** (2012), 121–134.

[3] J. M. Corcuera and A. Kohatsu-Higa, Statistical inference and Malliavin calculus, Dalang, Robert C. (ed.) et al., *Seminar on stochastic analysis, random fields and applications VI. Centro Stefano Franscini, Ascona, Italy*, May 19–23, 2008. *Basel: Birkhauser* (ISBN 978-3-0348-0020-4/pbk; 978-3-0348-0021-1/ebook). *Progress in Probability* 63, 59–82 (2011).

[4] Bogachev, Vladimir I. Differentiable measures and the Malliavin calculus, *Mathematical Surveys and Monographs*, 164. *American Mathematical Society, Providence, RI*, 2010.

[5] Yu. L. Daletskiĭ and Ya. I. Belopol'skaya, Stochastic equations and differential geometry, *"Vyshcha Shkola", Kiev*, 1989 (in Russian).

[6] Yu. L. Daletskii and G. A. Sokhadze, Absolute continuity of smooth measures, *Funktsional. Anal. i Prilozhen.* **22**:2 (1988), 77–78 (in Russian); translation in *Funct. Anal. Appl.* **22**:2 (1988), 149–150.

Received ??????????????????.

Author information

Petre Babilua, Faculty of Exact and Natural Sciences, Department of Mathematics,
I. Javakhishvili Tbilisi State University, 2 University St. Tbilisi 0186, Georgia.
E-mail: petre.babilua@tsu.ge

Elizbar Nadaraya, Faculty of Exact and Natural Sciences, Department of Mathematics,
I. Javakhishvili Tbilisi State University, 2 University St. Tbilisi 0186, Georgia.
E-mail: elizabar.nadaraya@tsu.ge

Grigol Sokhadze, I. Vekua Institute of Applied Mathematics of I. Javakhishvili Tbilisi State University, 2 University St. Tbilisi 0186, Georgia.
E-mail: grigol.sokhadze@tsu.ge