Proceedings of A. Razmadze
Mathematical Institute
Vol. **160** (2012), 121–134

ON THE CRAMER-RAO INEQUALITY IN AN INFINITE DIMENSIONAL SPACE

E. NADARAYA, O. PURTUKHIA AND G. SOKHADZE

Abstract. An infinite-dimensional analogue of the Cramer-Rao inequality is given. The technique of smooth measures are used. Conditions of regularity are given and under these conditions a variant of maximal likelihood principle for the infinite-dimensional case is proposed. The consistency property of the maximum likelihood estimate is given.

რეზიუმე. მიღებულია კრამერ-რაოს უტოლობის უსასრულო განზომილებიანი ანალოგი. გამოყენებულია გლუვი ზომების ტექნიკა. ჩამოყალიბებულია რეგულარობის პირობები და ამ პირობებში შემოთავაზებულია მაქსიმალური დასაჯერობის პრინციპის უსასრულო განზომილებიანი ვარიანტი. მოცემულია მაქსიმალური დასაჯერობის შეფასების ძალდებულობის თვისება.

The Cramer-Rao (C-R) inequality and ensuing from it consequences play fundamental role in statistical analysis. Many important problems are being solved on the grounds of that analysis. But a range of such problems do not involve situations which are connected with random processes. Therefore the key point is to extend the methods of the C-R inequality to an infinite dimensional case. Contemporary state of the infinite dimensional analysis allows one to consider many basic problems of statistics from a more general point of view.

The approach based on the analysis of sensibility of a family of probabilistic measures is well-known (see, for e.g., E. Pitmen). The theory of smooth measures ([2]–[3]) gives us a good chance for generalization in this direction. The present paper realizes this chance by an example in which the C-R inequality is generalized to an infinite dimensional case. The outlines of such ideas have actually been given by E. Gobet in [4]. The idea consists in application of the theory of P. Malliavin calculus ([5]–[8]). In their work [9], J.M. Corcuera and A. Kohatsu-Higa have used the technique of stochastic calculus of variations (Malliavin calculus) and obtained the

2010 *Mathematics Subject Classification.* 62F12, 62M20, 62E10.
Key words and phrases. Estimate, logarithmic derivative of measures, Cramer-Rao inequality.

results for a finite dimensional (more precisely, for a one-dimensional) case. In the present work the use will be made of the methods of smooth measures allowing us to formulate a general look at the questions dealt with the C-R inequality and related with them problems. In [10], the theory of smooth measures has been used to estimate a logarithmic measure derivative.

1. The Logarithmic Measure Derivative

Let $\{\Omega, \Im, P\}$ be a complete probabilite space. $X = X(\omega;\theta)$ is a randon element with values in a linear space E and parameter $\theta \in \Theta$, where Θ a subset of a separable real Banach space Ξ has the norm $\|\cdot\|_\Xi$.

The basic problem of statistics is to estimate an unknown parameter θ. This estimation should be based on observations for realization of the given randon value $X_1, X_2, \dots, X_n, \dots$, a sequence (a sampling) of independent and identical to X distributed random values. We are required to construct such a statistics $T = T(X_1, X_2, \dots, X_n)$ which will be optimal (in a sense motivated in advance) to estimate θ.

Usually, this situation generates a sequence of statistical structures $\{\aleph, \Re, (P(\theta;\cdot), \theta \in \Theta)\}$, where $\aleph = E^n(n = 1, 2, \dots, \infty$ is a linear space generated by a sequence of random values $X_1, X_2, \dots, X_n$, $\Re$ σ is the algebra generated by observable sets, and $\{P(\theta;\cdot), \theta \in \Theta)\}$ is a system of probabilite measures (distributions) generated by the vector $Y = (X_1, X_2, \dots, X_n)$ by virtue of the relation $P(\theta; A) = P(Y^{-1}(A))$, $A \in \Re$. In the classical statistics, the main object of investigation is, namely, that statistical structure $\{\aleph, \Re, (P(\theta;\cdot), \theta \in \Theta)\}$.

On the other hand, there exist a vast variety of problems in which $X = X(\omega;\theta)$ is just the function, convenient to operate with by imposing certain analytic requirements of smoothness (the conditions of regularity) with respect to the parameter θ. In addition, there arises a good possibility to apply an apparatus of stochastic calculus of variations.

Thus we obtain double calculus: the first one relies on the study of properties of the statistical structure $\{\aleph, \Re, (P(\theta;\cdot), \theta \in \Theta)\}$, the smoothness of the measure family $P(\theta,\cdot)$, and the second, the direct stochastic methods whose object of investigation is $X(\omega,\theta)$.

In this connection, we will be interested in the family of distributions $\{P(\theta;A), \theta \in \Theta, A \in \Re\}$ from the point of view of their smoothness with respect to the both parameters θ and A. Here we cite some definitions, therminological agreements and properties.

Throughout the paper, it will be assumed that $\aleph$ is a separable, reflexive Banach space. For every fixed $\theta \in \Theta$, $P(\theta,\cdot)$ is of positive measure. If $h \in \aleph$ is some vector, then by $P_h(\theta; A)$ we denote a measure obtained by the shift $P_h(\theta; A) = P(\theta; A + h)$.

We say that the measure $P(\theta;\cdot)$ is differentiable along the vector h, if there exists a bounded linear functional on $\aleph$, denoted by $d_h P(\theta;\cdot)$, such that for every $A \in \Re$, the equality

$$P_h(\theta; A) - P(\theta; A) = d_h(\theta; A)h + \alpha(\theta, A; h),$$

holds, where $\alpha(\theta, A; h)$ is the function, such that $\alpha(\theta, A; th) = o(t)$, $t \in R$. This is the so-called measure differentiability due to Fomin (the detailed account of another, somewhat different from Fomin's differentiability, definitions, properties and interconnections between them can be found in monograph [2]).

In the case, where $\aleph$ is a separable, real Hilbert space with scalar product $(\cdot,\cdot)_\aleph$ and norm $||h||_\aleph$, $h \in \aleph$, we write $P_h(\theta; A) - P(\theta; A) = (d_h P(\theta; A), h)_\aleph + \alpha(\theta, A; h)$, and sometimes (when this does not put us to confusion) under the derivative $d_h P(\theta;\cdot)$ will be understood an element of Hilbert space. Surely, the function $d_h P(\theta;\cdot)$ is the σ-additive (of alternating signs) measure on $\Re$.

The measure derivative of higher order is defined by iteration in the course of determination of the derivative. Thus, for example, $d_k d_h P(\theta;\cdot) = d_k(d_h P(\theta;\cdot)h)k$, $k, h \in \aleph$. In particular, in the case of Hilbert space $\aleph$, we have $(d^{(2)}_{h,h} P(\theta;\cdot)h, h)_\aleph = (d_h(d_h P(\theta;\cdot), h)_\aleph, h)_\aleph$..

For the n times differentiable measure, the expansion

$$P(\theta; A + h) - P(\theta; A) = \sum_{k=1}^{n} \frac{1}{k!} d^{(k)}_{h,\ldots,h} P(\theta; A) + \alpha(\theta, A; h),$$

is valid, where $\varphi(t) = \alpha(\theta, A; th)$ is the $(n-1)$-multiply differentiable function, vanishing together with its derivatives in zero faster than the corresponding powers t.

The function $\psi_\theta(t) = P(\theta, A + th)$ is nonnegative and everywhere differentiable. If $P(\theta; A) = 0$, $A \in \Re$, then the point $t = 0$ will turn out to be the point of minimum for the function $\psi_\theta(t)$. Therefore $d_h(\theta; A) = 0$. Hence, by the Radon-Nikodym's theorem, there exists a measurable function $\beta_\theta(x; h)$ such that $\dfrac{d_h P(\theta; dx)}{P(\theta; dx)} = \beta_\theta(x, h)$. This function is called logarithmic measure derivative along the vector $h \in \aleph$.. The logarithmic derivative $\beta_\theta(x, h)$ is linear in the second argument. The vector h is called an admissible direction for measure $P(\theta;\cdot)$. A set of all admissible directions is called an admissible subspace.

Example 1. Let $H_+ \subset H \subset H_-$ be a triple of Hilbert spaces whose embedding operator $i : H_+ \to H$ is the Hilbert-Schmidt's operator. Such a triple is called a Hilbert-Schmidt structure, or an equipped Hilbert space with quasi-kernel embeddings. Let γ_θ be Gaussian measure in H_- with a unit correlation operator in H with θ mean, $\theta \in H_-$. An admissible space for γ_θ is H_+. In addition, if $h \in H_+$, then the logarithmic derivative of measure γ_θ along h is $(\theta - x, h)_H$.

In the theory of differentiable measures, of great importance is the fact that the formula of integration by parts is valid. Let $\aleph$ be the separable, real Hilbert space and $f(x)$ be a functional in that space. Assume that there exists its derivative directed with respect to the vector $h \in \aleph$, and $d_h f(x) = \lim\limits_{t\to 0} t^{-1}[f(x+th) - f(x)]$, and $d_h f(\cdot) \in L_1(P(\theta;\cdot))$ for the fixed $\theta \in \Theta$. Then, if measure $P(\theta;\cdot)$ is differentiable with respect to the direction h, then (see [2])

$$\int\limits_{\aleph} (d_h f(x), h)_{\aleph} P(\theta; dx) = -\int\limits_{\aleph} f(x) d_h P(\theta; dx) = \\ = -\int\limits_{\aleph} f(x)\beta_\theta(x;h)P(\theta;dx). \tag{1}$$

We can define logarithmic derivative along nonconstant directions (the so-called logarithmic gradient). Equality (1) may serve as a basis for such a definition, or we can act analogously to what we have done in determining the measure derivative along constant directions.

Let $z(x) : \aleph \to \aleph$ be the differentiable vector field possessing bounded derivative $\sup\limits_{x\in\aleph} ||z'(x)|| < \infty$. An integral flow, corresponding to $z(x)$, we denote by S_t, $t \in R$. This implies that

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

The family of measures $(P(\theta;\cdot)\theta \in \Theta)$ is associated with a class of measures $(P_t(\theta;\cdot)\theta \in \Theta, t \in R)$ due to the transformation $P_t(\theta; A) = P(\theta; S_t^{-1}(A))$, $A \in \Re$.

We say that the measure $P(\theta;\cdot)$ is differentiable along the vector field $z(x)$, if there exists the measure (necessarily of alternating signs) $D_z P(\theta; A)$ such that for any bounded and differentiable function $\varphi : \aleph \to R$, $\varphi \in C^1(\aleph; R)$ we have

$$\int\limits_{\aleph} \varphi(x) D_z P(\theta; dx) = -\lim_{t\to 0} \int\limits_{\aleph} \varphi(x) \frac{P_t(\theta) - P(\theta)}{t}(dx).$$

Hence after the transformation, we obtain

$$\int\limits_{\aleph} \varphi(x) D_z P(\theta; dx) = -\int\limits_{\aleph} \varphi'(x) z(x) P(\theta; dx).$$

If, in addition, $D_z P(\theta;\cdot) << P(\theta;\cdot)$, then the Radon-Nikodym's density is called the logarithmic derivative $P(\theta;\cdot)$ along the vector field $z(x)$:

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

Let H be the embedding in the $\aleph$ Hilbert space whose embedding operator is the Hilbert-Schmitd's operator. Then we can consider the Hilbert-Schmidt's structure $\aleph^* \subset H \subset \aleph$. We distinguish an important class of measures $\mathcal{L}$ for which there exists a measurable, locally bounded function $\lambda : \aleph \to \aleph$ such that for every constant direction $h \in \aleph^*$ there exists the logarithmic derivative along h of the form $\beta_\theta(x;h) = \lambda(\theta;x)h = (\lambda(\theta,x),h)_H$. In this case we say that the measure possesses the logarithmic gradient $\lambda(\theta;x)$.

If $P(\theta) \in \mathcal{L}$, and the vector field $z : \aleph \to \aleph^*$ is bounded together with its derivative, then for the measure $P(\theta)$ there exists the logarithmic gradient (see [3]), and

$$\beta_\theta(x;z(x)) = \langle \lambda(\theta;x), z(x)\rangle + trz'(x).$$

This functional with respect to the continuity can be extended to smooth vector fields $z(x) : \aleph \to H$.

Example 2. In the conditions of Example 1, we consider the vector field $z(x) : H_- \to H_-$, possessing the bounded derivative: $\sup_{x\in H_-} ||z'(x)|| < \infty$. If $z : H_- \to H$, then the logarithmic gradient exists. But if it is known additionally that $z : H_- \to H_+$, then $\beta_\theta(x;z) = (\theta - x, z(x))_H + trz'(x)$.

Here we cite some properties of the logarithmic derivative; their proof can be found in [2].

Proposition 1. *Let the following conditions be fulfilled:*
(i) *The measures $P = P(\theta,\cdot)$ are differentiable along the vector $h \in \aleph$;*
(ii) *The functions f and g are differentiable along $h \in \aleph$;*
(iii) *$f, g \in L_1(d_h P)$ and $f'(x)h, g'(x)h \in L_1(P)$;*
(iv) *$(f'(x)h)g(x), f(x)(g'(x)h), f(x)g(x)\beta_\theta(x;h) \in L_1(P)$.*
Then

$$\int_\aleph (f'(x)h)g(x)P(\theta;dx) = -\int_\aleph f(x)(g'(x)h)P(\theta;dx) = -\int_\aleph f(x)g(x)\beta_\theta(x;h)P(\theta;dx). \quad (2)$$

Proposition 2. *Let the measures $P = P(\theta,\cdot)$ be differentiable along the vector $h \in \aleph$ and the function $\varphi(t) = \beta_\theta(x+th;H)$ be everywhere differentiable with $\beta'_\theta(x;h)h \in L_2(P)$. Then:*
(i) *$P(\theta,\cdot)$ is twice differentiable along h;*

(ii) $d^2_{h,h}P(\theta,\cdot) = \lfloor \beta'_\theta(x;h)h + (\beta_\theta(x;h))^2 \rfloor P(\theta;\cdot)$;
(iii) $\int\limits_{\aleph} (\beta_\theta(x;h))^2 P(\theta;dx) = -\int\limits_{\aleph} \beta'_\theta(x;hP(\theta;dx)$.

We will need measure smoothness with respect to the parameter, as well. Let, as above, we have the statistical structure $\{\aleph, \Re(P(\theta;\cdot), \theta \in \Theta)\}$, where $\aleph$ is the separable, real Banach space, and let Θ be a smooth manyfold imbedded into another separable, real Banach space Ξ. For any fixed $A \in \Re$ and for the vector $\vartheta \in \Xi$, let us consider the derivative of the function $\tau(\theta) = P(\theta;A)$ at the point θ along ϑ. We denote this derivative as follows: $d_\theta P(\theta;A)\vartheta$. For the fixed θ and ϑ, this derivative is the measure of alternating signs. It is easy to see that $d_\theta P(\theta,\cdot)\vartheta << P(\theta,\cdot)$, and by the Radon-Nikodym's theorem, there exists the measutable function $l_\theta(x;\vartheta) = \dfrac{d_\theta P(\theta;dx)\vartheta}{P(\theta;dx)}$. $l_\theta(x;\vartheta)$ which is called logarithmic derivative of measure with respect to the parameter $P(\theta;\cdot)$.

When Ξ is the separable Hilbert space, by $\mathbf{K}$ we denote a space of measures for which the logarithmic derivative with respect to the parameter is representable in the form of a scalar product $l_\theta(x;\vartheta) = (\mathbf{k}(x,\theta),\vartheta)_\Xi$. In addition, $\mathbf{k}(x,\theta)$ will be called a vector logarithmic gradient with respect to the parameter. For Examples 1 and 2, $\lambda(x,\theta) = \theta - x$ and $\mathbf{k}(x,\theta) = x - \theta$.

For the family of measures $(P(\theta;\cdot), \theta \in \Theta)$ possessing the logarithmic derivative with respect to the parameter along ϑ, there exists the measure ν dominating this family. It is known ([11]) that all measures $P(\theta;\cdot)$ are mutually equivalent, and $\dfrac{P(\theta_2;dx)}{P(\theta_1;dx)} = \exp \int\limits_{\theta_1}^{\theta_2} l_\theta(x;\vartheta)d\theta$.

2. The Regularity Conditions

A statistical structure is a notion, derivative from the probabilistic space and from a random value. Therefore in some conditions of regularity the above two notions of logarithmic derivative should be connected. Here we point out these conditions (the conditions of regularity).

Condition 1. $X(\theta) = X(\theta;\omega) : \Theta \times \Omega \to \aleph$, and there exists the derivative $X'(\theta)$ with respect to θ along $\vartheta \in \Xi_0$, where $\Xi_0 \subset \Xi$ is the subspace of Ξ. This derivative is a linear mapping $\Xi \to \aleph$ for every $\theta \in \Theta$. For any $\vartheta \in \Xi_0$ and $\theta \in \Theta$, we have $||\Xi'(\theta)\vartheta||_\aleph \in L_2(\Omega, P)$.

Condition 2. $E\{X'(\theta)\vartheta | X(\theta) = x\}$ is strongly continuous as the function x for all $\vartheta \in \Xi_0$, $\theta \in \Theta$.

Condition 3. The family of measures $(P(\theta;\cdot), \theta \in \Theta$ possess the logarithmic derivative with respect to the parameter along constant directions from the dense in Ξ subspace $\Xi_0 \subset \Xi$, and $l_\theta(x;\vartheta) \in L_2(\aleph, P(\theta))$, $\vartheta \in \Xi_0$, $\theta \in \Theta$.

Condition 4. The family of measures $(P(\theta;\cdot), \theta \in \Theta)$ possess the logarithmic derivative along constant ditections from the dense in $\aleph$ subspace $\aleph_0 \subset \aleph$ and $\beta_\theta(x;h) \in L_2(\aleph, P(\theta))$, $h \in \aleph_0$, $\theta \in \Theta$.

Lemma 1. *Under the conditions of regularity* 1)–4), *for the logarithmic derivatives* $\beta_\theta(x;h)$ *and* $l_\theta(x;\vartheta)$, *the equality*

$$l_\theta(x;\vartheta) = -\beta_\theta(x;\boldsymbol{K}_{\theta,\vartheta}(x)) \text{ where } \boldsymbol{K}_{\theta,\vartheta}(x) = E\left\{\frac{d}{d\theta}X(\theta)\vartheta|X(\theta)=x\right\} \quad (3)$$

holds.

Proof. By the definition, $P(\theta;A) = P(X^{-1}(\theta;A))$. Let $f(x)$ be the bounded, continuous differentiable along $h \in \aleph$ real-valued function. By the change of variable formula,

$$\int_\aleph f(x)P(\theta;dx) = Ef(X(\theta)).$$

We differentiate both parts with respect to θ along ϑ. Thus we obtain

$$\int_\aleph f(x)d_\theta P(\theta;dx)\vartheta = E\frac{d}{dx}f(X(\theta))\frac{d}{d\theta}X(\theta)\vartheta,$$

or

$$\int_\aleph f(x)l_\theta(x;\vartheta)P(\theta;dx) = \int_\aleph f'(x)E\{X'(\theta)\vartheta|X(\theta)=x\}P(\theta;dx).$$

Denote $K_{\theta,\vartheta}(x) = E\left\{\frac{d}{d\theta}X(\theta)\vartheta|X(\theta)=x\right\}$, and write

$$\int_\aleph f'(x)K_{\theta,\vartheta}(x)P(\theta;dx) = -\int_\aleph f(x)\beta_\theta(x,K_{\theta,\vartheta}(x))P(\theta;dx).$$

Since $f(x)$ is arbitrary, we obtain (3). □

Example 3. Let $\Xi = R^2$, $\Theta = R \times (0,\infty)$, $\Xi_0 = R^2$, $\aleph = R$. For a random $X(\theta_1,\theta_2)$, the distribution $P(\theta;A)$, $A \in \mathbf{B} = \Re$, where $\mathbf{B}$, the Borel σ-algebra in R, is prescribed by density

$$P(\theta_1,\theta_2;A) = \frac{1}{\sqrt{2\pi}\theta_2}\int_A \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2^2}\right\}dx.$$

Then for any $\vartheta = (\vartheta_1,\vartheta_2)^T$, we have

$$l_\theta(x:\vartheta) = \frac{(x-\theta_1)\theta_2\vartheta_1 - \theta_2^2\vartheta_2 + (x-\theta_1)^2\vartheta_2}{\theta_2^3}.$$

On the other hand, if by N we denote a standard normal distribution in R, then we can write $X(\theta_1,\theta_2)=\theta_2 N+\theta_1$. Hence

$$X'(\theta)\vartheta=\vartheta_1+\frac{X(\theta)-\theta_1}{\theta_2}\vartheta_2.$$

Respectively, $E\{X'(\theta)\vartheta|X(\theta)=x\}=\vartheta_1+\dfrac{x-\theta_1}{\theta_2}\vartheta_2=z(x)$. In this situation, $\lambda(\theta;x)=\dfrac{\theta_1-x}{\theta_2^2}$ and $\mathbf{k}(x,\theta)=\left(\dfrac{x-\theta_1}{\theta_2^2},\dfrac{(x-\theta_1)^2-\theta_2^2}{\theta_2^3}\right)^T$. Clearly, we have

$$\beta_\theta(x;z(x))=\frac{\theta_1-x}{\theta_2^2}\left(\vartheta_1+\frac{x-\theta_1}{\theta_2}\vartheta_2\right)+\frac{\vartheta_2}{\theta_2}=-l_\theta(x;\vartheta).$$

Example 4. Let $X(\theta)$ be a normally distributed random element with values in the separable, real Hilbert space H with 0 mean and with kernel correlation operator $\theta=B$. Let $P(\theta)$ be the corresponding Gaussian measure in H. In this case, $\aleph=H$, $\Re$, is the Borel σ-algebra in it, $\Xi=\mathbf{L}_1(H,H)$ is the Banach space of kernel operators in H with the norm $||K||_1=trK$, $\Theta\subset\mathbf{L}_1(H,H)$ is the space of linear operators C such that $B^{-1/2}CB^{-1/2}$ is the kernel operator in H. We have $X(B)=B^{1/2}N$, where N is the canonical Gaussian value in H. As a direction, we take the operator $C\in \mathrm{L}_{-1}(H,H)$. Calculations provide us with $X'(B)C=1/2\left(B^{-1/2}CB^{-1/2}(X(B))\right)$ and $E\{X'(B)C|X(B)=1/2B^{-1/2}CB^{-1/2}x=z(x)$. Respectively,

$$l_B(x;C)=-\beta_B(x;z(x))=\frac{1}{2}\left(B^{-1/2}CB^{-1/2}x,B^{-1}x\right)_H-\frac{1}{2}tr_H B^{-1/2}CB^{-1/2}.$$

3. The Cramer-Rao Inequality

Let $\{\aleph,\Re,(P(\theta,\cdot),\theta\in\Theta)\}$ be the statistical structure corresponding to a random element $X(\omega)=X(\theta,\omega)$. Here, $\aleph$ is the separable, real, reflexive Banach space, $\Re$ is the σ-algebra of Borel sets, and $\Theta\subset\Xi$ is an open subset of the separable, real Banach space Ξ. The conditions of regularity 1)–4) will be assumed to be fulfilled.

Let $g(\theta)=E_\theta(T(X))$, where $T:\aleph\to R$ is a measurable mapping (statistics). For the statistics we take one more condition of regularity.

Condition 5. For the statistics $T=T(x):\aleph\to R$, the equality

$$d_\vartheta\int_\aleph T(x)P(\theta;dx)=\int_\aleph T(x)d_\vartheta P(\theta;dx)$$

is valid.

Theorem 1 (The Cramer-Rao Inequality)**.** *Let the conditions of regularity* 1.–5. *be fulfilled. Then*

$$VarT(X) \geq \frac{(g'_\vartheta(\theta))^2}{E_\theta l_\theta^2(X;\vartheta)}. \tag{4}$$

Proof. Having differentiated $g(\theta)$ along $\vartheta \in \Xi$, we get

$$d_\vartheta E_\theta T(X) = d_\vartheta \int\limits_{\aleph} T(x)P(\theta;dx) = \int\limits_{\aleph} T(x) d_\vartheta P(\theta;dx)\vartheta =$$

$$= \int\limits_{\aleph} T(x) l_\theta(x,\vartheta) P(\eta;dx) = E_\theta T(X) l_\theta(X;\vartheta).$$

Thus

$$d_\vartheta E_\theta T(X) = E_\theta T(X) l_\theta(X;\vartheta). \tag{5}$$

In (5), we put $T(X) = 1$. We obtain $E_\theta l_\theta(X;\vartheta) = 0$. Therefore

$$d_\vartheta E_\theta(T(X)) = E_\theta((T(X) - g(\theta)) l_\theta(X;\vartheta)),$$

and hence

$$\left(d_\vartheta E_\theta(T(X))\right)^2 \leq E_\theta \left(T(X) - g(\theta)\right)^2 \cdot E_\theta l_\theta^2(X;\vartheta)$$

which yields

$$VarT(X) \geq \frac{\left(g'_\vartheta(\theta)\right)^2}{E_\theta l_\theta^2(X;\vartheta)}. \qquad \square$$

Corollary 1. *Taking into account our Lemma, inequality* (4) *takes the form*

$$VarT(X) \geq \frac{\left(g'_\vartheta(\theta)\right)^2}{E_\theta \beta_\theta^2(X; E(X'(\theta)\vartheta|X))}.$$

Example 5. Let there be observed an $X_k(\theta_1,\theta_2) = \dfrac{1}{\theta_1} e_k + \theta_2$ random value, where every e_k $(k = 1,2,\ldots,n)$ is an exponentially distributed random value with distribution density

$$p(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

Let $X_1, X_2, \ldots, X_n$ be observations; $X(\theta) = (X_1(\theta_1,\theta_2), \ldots, X_n(\theta_1,\theta_2))$. Then

$$X'(\theta) = \begin{pmatrix} -\dfrac{1}{\theta_1^2} e_1 & 1 \\ \cdots\cdots & \cdots\cdots \\ -\dfrac{1}{\theta_1^2} e_n & 1 \end{pmatrix} = \begin{pmatrix} -\dfrac{X_1(\theta) - \theta_2}{\theta_1} & 1 \\ \cdots\cdots & \cdots\cdots \\ -\dfrac{X_n(\theta) - \theta_2}{\theta_1} & 1 \end{pmatrix}, \ \theta = (\theta_1, \theta_2).$$

If we choose the direction $\vartheta=\begin{pmatrix}\vartheta_1\\ \vartheta_2\end{pmatrix}$, then

$$X'(\theta)\vartheta=\begin{pmatrix} -\dfrac{X_1(\theta)-\theta_2}{\theta_1}\vartheta_1+\vartheta_2\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ -\dfrac{X_n(\theta)-\theta_2}{\theta_1}\vartheta_1+\vartheta_2\end{pmatrix}.$$

Therefore

$$z(x)=E\{X'(\theta)\vartheta|X(\vartheta)=x\}=\begin{pmatrix} -\dfrac{x_1-\theta_2}{\theta_1}\vartheta_1+\vartheta_2\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ -\dfrac{x_n-\theta_2}{\theta_1}\vartheta_1+\vartheta_2\end{pmatrix}.$$

For the exponential distribution X, the logarithmic derivative is $\lambda(x)=-\dfrac{1}{\theta_1}I\ (x\geq\theta_2)$. If $\Lambda=\begin{pmatrix}\lambda(x)\\ \cdots\cdots\\ \lambda(x)\end{pmatrix}$, then

$$\beta_\theta(x;h)=(\Lambda,h)_{R^n}=-\frac{I(x\geq\theta_2)}{\theta_1}\sum_{k=1}^{n}h_k.$$

Finally,

$$\beta_\theta(x;z(x))=-\frac{I(\min x_k\geq\theta_2)}{\theta_1}\sum_{k=1}^{n}\left[t-\frac{x_k-\theta_2}{\theta_1}\vartheta_1+\vartheta_2\right]+$$

$$+tr\begin{pmatrix} -\dfrac{\vartheta_1}{\theta_1} & 0 \cdots\cdots 0\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ 0 & 0\ \cdots\cdots\ -\dfrac{\vartheta_1}{\theta_1}\end{pmatrix}=\frac{I(\min x_k\geq\theta_2)}{\theta_1^2}\sum_{k=1}^{n}x_k-$$

$$-\frac{nI(\min x_k\geq\theta_2)\theta_2\vartheta_1}{\theta_1^2}-\frac{nI(\min x_k\geq\theta_2)\vartheta_2}{\theta_1}-\frac{n\vartheta_1}{\theta_1}.$$

4. The Method of Maximal Likelihood

Let $\{\aleph,\Re,(P(\theta),\theta\in\Theta)\}$ be the statistical structure corresponding to a random element $X=X(\theta)=X(\theta,\omega)$, $\omega\in\Omega$, where $\aleph$ is the separable, real Banach space, $\Re$ is the σ-algebra of the Borel subsets, and Θ is an open subset of another separable, real Banach space Ξ. We assume that the family of measures $(P(\theta,\cdot),\theta\in\Theta)$ possess the logarithmic derivative with respect to the parameter $l_\theta(x,\vartheta)$ along $\vartheta\in\Xi$. Then by Theorem 1, there exists the logarithmic derivative with respect to the measure, and $l_\theta(x,\vartheta)=-\beta_\theta(x,E\{X'(\theta)\vartheta|X(\theta)=x\}$.

Consider a structure of repeated sampling $\{\aleph, \Re, (P(\theta), \theta \in \Theta)\}^n = \{\aleph^n, \Re^n, (P^n(\theta), \theta \in \Theta)\}$.

Theorem 2. *If there exists the logarithmic derivative $l_\theta(x, \vartheta)$ with respect to the parameter in the statisical structure $\{\aleph, \Re, (P(\theta), \theta \in \Theta)\}$, then there likewise exists the logarithmic derivative $L_\theta((x_1, \dots, x_n), \vartheta^n)$ with respect to the parameter, along $\vartheta^n \overset{def}{=} (\vartheta, \dots, \vartheta)$, for the structure of repeated sampling $\{\aleph, \Re, (P(\theta), \theta \in \Theta)\}^n$, and we have*

$$L_\theta((x_1, \dots, x_n), \vartheta^n) = \sum_{k=1}^n l_\theta(x_k, \vartheta) = \\ = -\sum_{k=1}^n \beta_\theta\big(x_n, E\{X_k'(\theta)\vartheta | X_k(\theta) = x_k\}\big). \tag{6}$$

Proof. Since by the condition, there exists $d_\theta^\vartheta P(\theta)$, it is not difficult to calculate that there likewise exists $d_\theta^{\vartheta,\dots,\vartheta} P^n(\theta) = \sum_{k=1}^n d_\theta^\vartheta(x_k, \vartheta) \prod_{\substack{j=1 \\ j\neq k}}^n P(\theta)$. This last one is absolutely continuous with respect to $P^n(\theta)$. By the Radon-Nikodym's theorem, we find that the statements of the theorem, as well as formula (6), are valid. □

It follows from the theorem that in the case under consideration we can formulate the principle of maximal likelihood.

Let $X_1, X_2, \dots, X_n$ be sampling from the random value $X(\theta)$, where θ is an unknown parameter to be estimated by means of sampling. Assume that there exists the logarithmic derivative $l_\theta(x, \vartheta)$ with respect to the parameter, along any vector $\vartheta \in \Xi_0$, of distribution $P(\theta)$, which corresponds to $X(\theta)$ and has the form $l_\theta(x, \vartheta) = \langle \lambda(x, \theta), \vartheta \rangle$. Here, Ξ_0 is the dense subset of Ξ.

As is known, all measures $P(\theta)$ are equivalent to each other. Let $\theta_0 \in \Xi_0$ be a fixed point. Consider the likelihood function

$$\frac{dP(\theta)}{dP(\theta_0)}(x) = \rho(x, \theta).$$

It can be easily seen that for $P \in \mathrm{L}$ the above equality results in

$$\frac{\rho_\theta'(x, \theta)\vartheta}{\rho(x, \theta)} = l_\theta(x, \vartheta).$$

For the sampling $X_1, X_2, \dots, X_n$, the likelihood function is

$$L(X_1, X_2, \dots, X_n, \theta; \vartheta) = \prod_{k=1}^n \rho(X_k, \theta).$$

According to the likelihood principle, the estimate of maximal likelihood will be called the value $\theta = \widehat{\theta}$ which supplies the likelihood function L with maximum (provided that such value of the parameter θ exists). Since

$$\ln L(X_1, X_2, \dots, X_n, \theta; \vartheta) = \sum_{k=1}^{n} \ln \rho(X_k, \theta),$$

the condition for maximum allows us to formulate this definition in terms of the logarithmic derivative with respect to the parameter:

The estimate of maximal likelihood $\widehat{\theta}$ with respect to the direction ϑ is called the root (if exists) of the equation

$$\sum_{k=1}^{n} l_\theta(x_k, \vartheta) = 0, \quad \forall \vartheta \in \Xi \tag{7}$$

with respect to θ, under the condition that the expression $\dfrac{d}{d\theta} l(x, \theta)$ is defined negatively.

By Lemma 1, equation (7) can be replaced by

$$\sum_{k=1}^{n} \{\mathbf{k}(x_k, \theta), K(x_k, \theta)\vartheta)_{\aleph} + trK'_x(x_k, \theta)\vartheta\} = 0, \quad \forall \vartheta \in \Xi_0. \tag{8}$$

x_k in formulas (7) and (8) are the values of Xi, experimentally.

Example 6. Let in the equipped Hilbert space $H_+ \subset H \subset H_-$ be considered the sampling $X_1, X_2, \dots, X_n$ from the canonical Gaussian value with an unknown mean θ, for which $\beta_\theta(x, h) = (\theta - x, h)_H$, $h \in H_+$. Clearly, $X(\theta) = N + \theta$, where N is the canonical Gaussian value with zero mean. $X'(\theta) = I$, $X'(\theta)h = h$, and hence $E\{X'_k(\theta)h | X_k(\theta) = x\} = h$. Thus, (8) takes the form

$$\sum_{k=1}^{n} (\theta - x_k, h)_H = 0,$$

whence

$$(\widehat{\theta}, h)_H = \frac{1}{n} \sum_{k=1}^{n} (X_k, h)_H \text{ and } \widehat{\theta} = \frac{1}{n} \sum_{k=1}^{n} X_k = \overline{X}\,[1].$$

In addition, $\dfrac{1}{n} \displaystyle\sum_{k=1}^{n} \dfrac{d^h}{d\theta} (x - \theta, h)_H = -n||h||_H^2 \le 0.$

---

[1] This equality is obtained under the condition $h \in H_+$ and $\theta - x_k \in H_-$. If j is the embedding operator $j : H_+ \to H$ and $j^* : H_- \to H$, we can rewrite

$$(j^* \widehat{\theta}, jh)_H = \frac{1}{n} \sum_{k=1}^{n} (j^* X_k, jh)_H, \quad \text{and again} \quad \widehat{\theta} = \frac{1}{n} \sum_{k=1}^{n} X_k = \overline{X}.$$

If $h \in H$, then, as is known, the expression $\beta_\theta(x, h) = (\theta - x, h)_H$ extends with respect to the continuity as a measurable linear functional. The obtained in such a way so-called Skorokhod's integral solves our problem.

As an application, we consider a random process $x(t) = \varphi(t) + w(t)$, where $w(t)$ is a standard Wiener's process, $\varphi \in C[,\infty) = \Xi$ is an unknown component of the observable process. Surely, $Ex(t) = \varphi(t)$. In this case, $H_+ = C'[0,\infty)$, $H_- = L_2[0,\infty)$.

If $x_1(t), x_2(t), \ldots, x_n(t)$ are observed, then $\widehat{\varphi}(t) = \dfrac{1}{n}\sum_{k=1}^{n} x_k(t)$.

If $H = R^n$ is finite-dimensional, we obtain the estimate of maximal likelihood along any vector $h = (h_1, \ldots, h_n)$:

$$\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}) = \left(\frac{1}{n}\sum_{k=1}^{n} X_k^1, \ldots, \frac{1}{n}\sum_{k=1}^{n} X_k^m\right).$$

5. Consistensity of the Maximal Likelihood Estimate

Let the statistical structure $\{\aleph, \Re, (P(\theta), \theta \in \Theta)\}$ possess the logarithmic derivative with respect to the parameter along the constant vector $\vartheta \in \Xi_0 \ - \ l_\theta(x, \vartheta)$. We introduce a Kulbak-Leybler type function of distance for pairs of measures:

$$D(\theta_1, \theta_2) = E_{\theta_1}\{l_{\theta_1}(x, \theta_2 - \theta_1) - l_{\theta_2}(x, \theta_2 - \theta_1)\}. \tag{9}$$

Example 7. In the equipped Hilbert space $H_+ \subset H \subset H_-$, for canonical Gaussian measures μ_1 and μ_2 with, respectively, θ_1 and θ_2 means, the distance is $D(\mu_1, \mu_2) = (\theta_1 - \theta_2, \theta_2 - \theta_1)_H = -||\theta_2 - \theta_1||_H^2$.

Lemma 2. *Let the family be defined uniquely by the parameter, i.e., if $P(\theta_1) = P(\theta_2)$, then $\theta_1 = \theta_2$, and vice versa. And if $D(\theta_1, \theta_2) \geq 0$, then $P(\theta_1) = P(\theta_2)$, and vice versa.*

Proof. We note immediately that $D(\theta_1, \theta_2) \leq 0$. Indeed,

$$\begin{aligned} D(\theta_1, \theta_2) &= E_{\theta_1}\{l_{\theta_1}(x, \theta_2 - \theta_1) - l_{\theta_2}(x, \theta_2 - \theta_1)\} = \\ &= E_{\theta_1}(\lambda(X, \theta_1) - \lambda(X, \theta_2), \theta_2 - \theta_1)_\Xi = \\ &= E_{\theta_1}(\lambda'_\theta(X, \theta_1 + \tau(\theta_2 - \theta_1))(\theta_2 - \theta_1), \theta_2 - \theta_1)_\Xi \leq 0, \ (0 \leq \tau \leq 1). \end{aligned}$$

Thus we can see that if $D(\theta_1, \theta_2) \geq 0$, then $\theta_1 = \theta_2$, which implies $P(\theta_1) = P(\theta_2)$, and vice versa. □

Theorem 3. *If Ξ_0 is a convex precompact set, then the estimate of maximal likelihood is consistent.*

Proof. Let $\widehat{\theta}$ be a solution of equation (8) or (9). Consider the difference

$$\varphi_n(t) = \frac{1}{n}\ln L(x_1, \ldots, x_n, \theta; t(\widehat{\theta} - \theta)) - \frac{1}{n}\ln L(x_1, \ldots, x_n, \widehat{\theta}; t(\widehat{\theta} - \theta)).$$

Since $\widehat{\theta}$ for any t is a point of maximum, $\varphi_n(t)$ decreases with respect to t on $[0,1]$. On the other hand, by the strong law of large numbers,

$$\varphi_n(t) = \frac{1}{n} \ln \frac{L(x_1, \dots, x_n, \theta; t(\widehat{\theta} - \theta))_{a.s}}{L(x_1, \dots, x_n, \widehat{\theta}; t(\widehat{\theta} - \theta))} \to \varphi(t),$$

and $\varphi_n(t)$ is likewise the decreasing function. Therefore, $\varphi'(t) \le 0$. But $\varphi'(t) = P \lim\limits_n \varphi_n'(t)$ and $\varphi_n' \overset{a.s}{\to} E_{\widehat{\theta}}\{l_\theta(X, \widehat{\theta} - \theta) - l_{\widehat{\theta}}(X, \widehat{\theta} - \theta)\} = D(\widehat{\theta}, \theta) \le 0$.

By Lemma 2, we have $\theta = \widehat{\theta}$. □

References

1. E. Pitmen, Principles of the theory of statistical conclusions. (Russian) *Moscow, Mir*, 1986.
2. V. V. Bogachev, Differentiable measures and Malliavin calculus. (Russian) *Moscow-Izhevsk: Sci. Invest. Center Regular and Chaotic Dynamics*, 2008.
3. Yu. L. Daletskii and Ja. I. Belopol'skaya, Stochastic equations and differential geometry. (Russian) *Vishcha shkola, Kiev*, 1989.
4. E. Gobet, Local asymptotic mixed normality property for elliptic diffusion: a Malliavin Calculus approach. *Bernoulli* **7** (2001), No. 6, 899–912.
5. P. Malliavin, Stochastic calculus of variations and hypoelliptic operators. *Proceedings of the International Symposium on Stochastic Differential Equations* (*Res. Inst. Math. Sci., Kyoto Univ., Kyoto*, (1976), 195–263, *Wiley, New York-Chichester-Brisbane*, 1978.
6. P. Malliavin, Calculus of variations, stochastic integrals and the de Rham complex on the Wiener space. *C. R. Acad. Sci. Paris Ser. I Math.* **299** (1984), No. 8, 347–350.
7. P. Malliavin, Stochastic analysis. *Springer-Verlag, Berlin*, 1997.
8. D. Nualart, The Malliavin calculus and related topics. Second edition. *Probability and its Applications* (*New York*), *Springer-Verlag, Berlin*, 2006.
9. J. M. Corcuera and A. Kohatsu-Higa, Stochastic inference and Malliavin calculus. *Math. Preprint Series, IMUB Institute de Matematica* **410** (2008).
10. E. Nadaraya and G. Sokhadze, On the statistical estimate of the logarithmic derivative of a measure in a Hilbert space. *Georgian Math. J.* **17** (2010), 741–747.
11. Yu. L. Daletskii and G. Sokhadze, Absolute continuity of smooth measures. (Russian) *Functional. Anal. i Prilozhen.* **22** (1988), No. 2, 77–78; *translation in Funct. Anal. Appl.* **22** (1988), No. 2, 149–150.

(Received 11.04.2012)

Authors' address:

I. Javakhishvili Tbilisi State University
2, University Str., Tbilisi 0186
Georgia
E-mail: elizbar.nadaraya@tsu.ge;
o.purtukhia@gmail.com;
giasokhi1@i.ua