Записки научных
семинаров ПОМИ
Том 278, 2001 г.

A. Gupta, L. Klebanov

ILL-POSED PROBELMS IN STATISTICAL ESTIMATION THEORY

ABSTRACT. Connections between elements of the mathematical model used in the statistical estimation theory are investigated. In partisular, restrictions that must be imposed on the model in order this model to posess some natural properties are established.

Ill-posed problems are usually understood as certain results where small changes in the assumptions lead to arbitrary large changes in the conclusions. Such results are not very useful for practical applications, where the presumptions usually hold only approximately. Presumably, the ill-posedness of certain practical problems is due to the lack of their precise mathematical formulation. Consequently, we can deal with such problems by replacing a given ill-posed problem with another, well-posed problem, which in some sense is "close" to the original one.

The aim of the proposed work is to study the constraints that are imposed on the model elements in the theory of statistical estimation in the presence of certain conditions that seem to be desirable for the theory or its applications. The results derived here can be interpreted as recommendations that would allow applied statistician to justify the use of some statistical models. One of the main notations of the model in the theory of statistical parameter estimation is the niotation of a loss function. It is almost obviously that the use of unbounded loss functions leads to ill-posedness in the problem of statistical estimation theory. We shall show what class of loss functions is more natural from our point of view.

Let us move now to the notion of a model in the theory of statistical parameter estimation, and to a more exact formulation of the problem.

On a measurable space $(\mathfrak{X}, \mathfrak{F})$, define a family of probability measures $\{\mathbb{P}_\theta, \theta \in \Theta\}$, where Θ - is some abstract set. Let $\mathbf{x}$ be an observed value distributed according to one of the laws of $\mathbb{P}_\theta$, where the corresponding value of the parameter θ is unknown. The objective is to build an estimator $\gamma(\theta)$ from the observed value of $\mathbf{x}$, where $\gamma(\theta)$ is some given parametric function which maps parametric set Θ into r-measurable Euclidean set $\mathbb{R}^r$. Let $\mathfrak{R}^r$ denote the σ-algebra of Borel subsets of $\mathbb{R}^r$, and let π_r denote the set of probability measures on $(\mathbb{R}^r,\ \mathfrak{R}^r)$. We will call an estimator (or to be exact, a randomized estimator) of a parametric function γ, a map δ^* from $\mathfrak{X}$ into π_r, giving for each $\mathbf{x} \in \mathfrak{X}$ a corresponding probability measure $\delta^*_{\mathbf{x}}(\cdot)$ on $\mathfrak{R}^r$. In doing so we will always assume that for each $A \in \mathfrak{R}^r$ the quantity $\delta^*_{\mathbf{x}}(A)$ is a measurable function of an unknown $\mathbf{x}$. If for each $\mathbf{x} \in \mathfrak{X}$ the measure $\delta^*_{\mathbf{x}}$ is degenerate, then we will talk about a non-randomized estimator of a parametric function γ. Let us agree to denote such non-randomized estimators of γ with the symbol γ^*, supplementing it with various indices if necessary.

A randomized estimator can be interpreted in a following way: for an observation $\mathbf{x}$ a probability distribution $\delta^*_{\mathbf{x}}$ is chosen on $(\mathbb{R}^r,\ \mathfrak{R}^r)$. Then, we observe an auxiliary random variable with distribution $\delta^*_{\mathbf{x}}$. A value δ of this random variable will be accepted as a final estimator of $\gamma(\theta)$.

Among various estimators of a parametric function γ, we would like to chose the best one, in some meaning of the word. To formally measure the quality of an estimator, let us introduce a loss function $w = w(s,t)$. Here, the value $w(\delta, \gamma(\theta))$ denotes the losses incurred when δ is used as an estimator of $\gamma(\theta)$. In doing so let us naturally assume that,

$$w(s,t) \geq 0 \quad \text{and} \quad w(t,t) = 0 \quad \text{for} \quad s,\ t\ \in \mathbb{R}^r,$$

that is, the losses are always nonnegative, and if the value that is being estimated is obtained precisely, losses are absent. In the future, we shall always assume that these conditions are satisfied, unless specified otherwise.

Let δ^* be an estimator of γ. Then, the risk of δ^*, corresponding to a the family $\{\mathbb{P}_\theta,\ \theta \in \Theta\}$ and loss function w, is the mean loss

$$R_\theta(\delta^*, \gamma; P_\theta; w) = E_\theta \int\limits_{\mathbb{R}^r} w(\delta, \gamma(\theta)) d\delta^*_{\mathbf{x}}(\delta) =$$

$$= \int\limits_{\mathfrak{X}} d\mathbb{P}_\theta(\mathbf{x}) \int\limits_{\mathbb{R}^r} w(\delta, \gamma(\theta)) d\delta^*(\delta).$$

For brevity, we shall simply talk about "a risk of an estimator δ^*" and we may omit the dependency from certain elements. For example, if the family of distributions, loss function, and parametric function γ are clear from the contest, we shall simply write $R_\theta(\delta^*)$ (or $R_\theta\delta^*$) instead of $R_\theta(\delta^*, \gamma; \mathbb{P}_\theta, w)$. (Of course, it is always assumed that w satisfies necessary measure conditions, so that the corresponding integrals are defined. In the future these conditions may not be stated explicitly).

Note that although the usual loss functions are maps from $\mathbb{R}^r \times \mathbb{R}^r$ to $\mathbb{R}^1$, sometimes the so-called matrix loss functions are used. Such loss functions were studied in detail in works [10, 4, 12]. Here however we will mostly deal with scalar loss functions.

If δ estimates the value of $\gamma(\theta)$, then it is natural to assume that losses incurred in such estimation depend on the difference $\delta - \gamma(\theta)$, or on the norm $|\delta - \gamma(\theta)|$. For this reason in many applied problems the loss function w has the form

$$w(s,t) = v(s-t)$$

or

$$w(s,t) = v_1(|s-t|),$$

where v and v_1 are non-negative functions taking zero values when the argument itself is zero. Then it is often assumed that $v_1(z)$ increases for $z > 0$. It should be kept in mind that in many problems, it is convenient to consider estimators that belong to some specified class $\mathcal{K}^*$. (For example, the set of all non-randomized estimators can be considered as a candidate for $\mathcal{K}^*$.) In the sequel, we shall assume that estimators that belong to the class of interest have finite risk, unless specified otherwise.

Sometimes we may have to examine not just one parametric function γ, but all parametric functions from some set $\mathcal{K}$. In other words, for a given function $\gamma \in \mathcal{K}$ its estimator $\delta^* \in \mathcal{K}^*$, which depends on $\gamma \in \mathcal{K}$, should be proposed.

Let us call the family of distributions $\{\mathbb{P}_\theta\ \theta \in \Theta\}$ defined on $(\mathfrak{X},\ \mathfrak{F})$, the loss functions w, the set of parametric functions $\mathcal{K}$ and the class of estimators $\mathcal{K}^*$ – the model in a theory of statistical parameter estimation (or simply the "model").

An estimator $\tilde{\delta}^* \in \mathcal{K}^*$ of a parametric function γ is called admissible in the class $\mathcal{K}^*$ if the conditions

$$\delta^* \in \mathcal{K}^*, \quad R_\theta \delta^* \le R_\theta \tilde{\delta}^* \ \text{ for all } \ \theta \in \Theta$$

imply that

$$R_\theta \delta^* = R_\theta \tilde{\delta}^* \ \text{ for all } \ \theta \in \Theta.$$

Thus, if an estimator $\tilde{\delta}^*$ is admissible in the class $\mathcal{K}^*$, then in this class there is no other estimator whose risk would be uniformly grater than the risk of $\tilde{\delta}^*$, and there is at least one value of θ for which the risk is strictly less than the risk of $\tilde{\delta}^*$.

An estimator $\tilde{\delta}^* \in \mathcal{K}^*$ is called minimax in class $\mathcal{K}^*$, if

$$\sup_{\theta\in\Theta} R_\theta \tilde{\delta}^* = \inf_{\delta^*\in\mathcal{K}^*} \sup_{\theta\in\Theta} R_\theta \delta^*.$$

An estimator $\tilde{\delta}^* \in \mathcal{K}^*$ is called an optimal estimator of a parametric function $\gamma \in \mathcal{K}$ in class $\mathcal{K}^*$ if for any estimator $\delta^* \in \mathcal{K}^*$, the following inequality holds

$$R_\theta \tilde{\delta}^* \leq R_\theta \delta^*, \quad \theta \in \Theta.$$

Clearly, if an estimator $\tilde{\delta}^*$ is optimal in class $\mathcal{K}^*$, then it also is admissible and minimax in $\mathcal{K}^*$.

Let K be some set of estimators, and $\mathcal{K}_1^* \subset \mathcal{K}^*$ be a subset of $\mathcal{K}^*$. Let us say that $\mathcal{K}_1^*$ is a complete subset of $\mathcal{K}^*$ if for any estimator $\delta^* \in \mathcal{K}^*$ of a parametric function γ there exists an estimator $\tilde{\delta}^* \in \mathcal{K}^*$ of the same parametric function such that the following identity holds:

$$R_\theta \tilde{\delta}^* \leq R_\theta \delta^* \quad \text{for all} \quad \theta \in \Theta.$$

If $\mathcal{K}^*$ is a set of all estimators with finite risks, then we will call its complete subclass a complete class.

As it was already mentioned, our aim is to investigate the relationship among the elements of the model in the theory of statistical estimation of parameters and to study how certain conveniences that seem to be natural impose limitations on the model. Under such conveniences we can understand, for example, the requirement of:

1. completeness of a class of non-randomized estimators;
2. completeness of a class of estimators that are symmetric functions of observations obtained via repeated sampling scheme;
3. the use of "difference" loss functions

   $$w(s,t) = v(s-t)$$

   or

   $$w(s,t) = v_1(|s-t|),$$

   with preservation of a possibility to improve any estimator by a function of a sufficient statistic;
4. many other conditions.

On the other hand, even when studying the main problem of the theory of statistical estimation one has to remember that in practice a model (or some of its elements) can be known only approximately. That is why it is desirable to choose well-posed formulation of problems.

The requirements 1–3 mentioned above are connected mostly to the problem of loss function selection.

1. Reducing functions

First, we introduce the notion of a family of reducing functions which strengthens the notion of a convex function. Let us have a family of real functions of a real argument $\{W_t(\cdot), t \in T\}$, where T is a certain abstract set.

Definition 1.1 (See [9]). *We say that $\{W_t(\cdot), t \in T\}$ is a family of reducing functions if for each cumulative distribution function $G(s)$ on R^1 there exists a point $s^* = s^*(G)$ such that*

$$\int_{R^1} W_t(s)dG(s) \geqslant W_t(s^*), \quad \forall t \in T, \tag{1.1}$$

if the integral on the left hand side exists and is finite. We assume that s^* is a measurable function of G.

(Measurability of s^* as a function of G is understood here in the following sense. Let $G_u(s)$ be a family of c.d.f.'s which is a measurable function of u for each fixed $s \in R^1$. Then we assume that $s^*(G_u)$ is also a measurable function of u.)

If in Definition 1.1 we would require, instead of the existence of s^*, the following equality

$$S^* = \int_{R^1} s dG(s),$$

then (1.1) would reduce to the Jensen's inequality and would be equivalent to convexity of each function $W_t(s)$ from the considered family.

On the other hand, if a family consists of one continuous function $W_{t_0}(s)$ reaching its minimal and maximal values, then by the mean value theorem

$$\int_{R^1} W_{t_0}(s)dG(s) = W_{t_0}(s^*)$$

and there is no convexity conditions.

Next, we formulate three more definitions related to the above one.

Definition 1.2 (See [9]). *We say that $\{W_t(\cdot), t \in T\}$ is a family of strongly reducing functions if it consists of reducing functions and inequality (1.1) is strict for each non-degenerated c.d.f. $G(s)$.*

Definition 1.3 (See [9]). *We say that $\{W_t(\cdot), t \in T\}$ is a family of weakly reducing functions if for each real numbers s_1, s_2 and each rational $p \in (0,1)$ there exists a point $s^* = s^*(s_1, s_2; p)$ such that*

$$pW_t(s_1) + (1-p)W_t(s_2) \geqslant W_t(s^*), \forall t \in T, \tag{1.2}$$

where s^* is a measurable function of s_1 and s_2.

Definition 1.4. *We say that $\{W_t(\cdot), t \in T\}$ is a continuous family of weakly reducing functions if*

1. *The functions $W_t(s)$ are continuous,*
2. *The family is weakly reducing,*
3. *The point s^* satisfying (1.2) can be chosen in such a way that it becomes a continuous function of s_1, s_2, and p.*

Clearly, if $\{W_t(\cdot), t \in T\}$ is reducing, then it is weakly reducing. The opposite implication is not in general true. Consider the following example.

Example 1.1. Let the set T consist only of three elements $\{1\}$, $\{2\}$, $\{3\}$, i.e., $T = \{1, 2, 3\}$. Let us define functions

$$W_1(s) = \begin{cases} 1 & \text{if } s \text{ is irrational,} \\ 0 & \text{if } s \text{ is rational,} \end{cases}$$
$$W_2(s) = \begin{cases} 0 & \text{if } s \text{ is irrational,} \\ s & \text{if } s \text{ is rational,} \end{cases}$$
$$W_3(s) = \begin{cases} 0 & \text{if } s \text{ is irrational,} \\ -s & \text{if } s \text{ is rational.} \end{cases}$$

The family $\{W_t(\cdot), t \in T\}$ is weakly reducing but not reducing.

Proof. We need to define s^* satisfying (1.2). If s_1 and s_2 are irrational numbers, and $p \in (0,1)$ is an arbitrary rational number, then we can take for $s^*(s_1, s_2; p)$ an arbitrary irrational number or zero.

If s_1 and s_2 are rational, then it is easy to see that (1.2) holds if and only if s^* is rational and $s^* = ps_1$ (equivalently $s^* = (1-p)s_2$).

If the constructed family was reducing, then equation (1.2) should be satisfied for rational s_1 and s_2 and irrational p. But for rational s_1 and s_2 and $t = 1$ we should have $W_1(s^*) \leqslant 0$, i.e., s^* should be rational. For $t = 2$ we find that

$$ps_1 + (1-p)s_2 \geqslant s^*,$$

and for $t = 3$:

$$-(ps_1 + (1-p)s_2) \geqslant -s^*,$$

i.e., s^* has to have the form

$$s^* = ps_1 + (1-p)s_2,$$

which, in view of the irrationality of p, leads to a contradiction with the rationality of s^*. $\square$

Let us formulate here one simple result which establishes an equivalent condition to the weakly reducing family of functions.

Theorem 1.1. *A family $\{W_t(\cdot), t \in T\}$ is weakly reducing if and only if for each integer $n \geqslant 2$ and each choice of real numbers $s_1, \dots, s_n$ there exists a point $s_n^* = s_n^*(s_1, \dots, s_n)$, being a measurable function of $s_1, \dots, s_n$ and such that*

$$\frac{1}{n}\sum_{j=1}^{n} W_t(s_j) \geqslant W_t(s_n^*), \quad \forall t \in T. \tag{1.3}$$

Proof. We assume that the family $\{W_t(\cdot), t \in T\}$ is weakly reducing and we prove that equality (1.3) holds. Indeed, for $p = 1/2$ from (1.2) we obtain (1.3) for $n = 2$. If (1.3) holds for some $n \geqslant 2$, then

$$\begin{aligned}\frac{1}{n+1}\sum_{j=1}^{n+1} W_t(s_j) &\geqslant \frac{n}{n+1}W_t(s_n^*) + \frac{1}{n+1}W_t(s_{n+1}) \\ &\geqslant W(s^*(s_n^*, s_{n+1}; n/(n+1))).\end{aligned}$$

Thus (1.3) follows by the mathematical induction.

If the family $\{W_t(\cdot), t \in T\}$ satisfies (1.3) for all $n \geqslant 2$ and $p \in (0,1)$ is a rational number of the form k/n, then by choosing in (1.3) the first k numbers among s_j's equal to s' and the remaining $n-k$ equal to s'', we obtain

$$\frac{k}{n}W_t(s') + \Big(1 - \frac{k}{n}\Big)W_t(s'') \geqslant W_t(s_n^*(s', \dots, s', s'', \dots, s'')),$$

which coincides with (1.2). $\square$

Remark 1.1. Clearly, s_n^* in (1.3) can be chosen to be a symmetric function of arguments $s_1, \dots, s_n$.

We demonstrate that a continuous weakly reducing family satisfying some additional conditions becomes reducing.

Theorem 1.2. *Let a family $\{W_t(\cdot), t \in T\}$ be continuous weakly reducing. We assume that for each $t \in T$ there exist (finite of infinite) limits $\lim_{s\to\infty} W_t(s)$ and $\lim_{s\to-\infty} W_t(s)$, and*

$$\lim_{s\to\infty} W_t(s) = \lim_{s\to-\infty} W_t(s) > W_t(x) \tag{1.4}$$

for all $x \in R^1$. Then the family $\{W_t(\cdot), t \in T\}$ is reducing.

Proof. Using the mathematical induction it is easy to show the following property. If the family $\{W_t(\cdot), t \in T\}$ is weakly reducing, then for all real numbers $p_1, \dots, p_n$ satisfying $p_j \geqslant 0$, $\sum_{j=1}^n p_j = 1$, there exists a point $s_n^* = s_n^*(s_1, \dots, s_n; p_1, \dots, p_n)$ such that

$$\sum_{j=1}^n p_j W_t(s_j) \geqslant W_t(s_n^*), \quad \forall t \in T.$$

Now, let $G(s)$ be an arbitrary c.d.f. We can select $s_j^{(n)}$, $p_j^{(n)}$, $(j = 1, \dots, n)$ in such a way that for all $t \in T$ the following condition is satisfied

$$\lim_{n\to\infty} \sum_{j=1}^n P_j^{(n)} W_t(s_j^{(n)}) = \int_{R^1} W_t(s) dG(s).$$

Let us consider the sequence of resulting points s_n^*, $n = 1, 2, \dots$. We demonstrate that it is bounded. Indeed, if there exists a subsequence $s_{n_k}^*$ for which $\lim_{k\to\infty} |s_{n_k}^*| = \infty$, then by (1.4) we would have

$$\begin{aligned}\lim_{s\to\pm\infty} W_t(s) &= \lim_{k\to\infty} W_t(s_{n_k}^*) \leqslant \lim_{k\to\infty} \sum_{j=1}^{n_k} p_j^{(n_k)} W_t(s_j^{(n_k)}) \\ &= \int_{R^1} W_t(s) dG(s) < \lim_{s\to\pm\infty} W_t(s).\end{aligned}$$

This is a contradiction and thus $\{s_n^*\}_{n=1}^\infty$ is bounded.

Let us consider a subsequence $\{s_{n_l}^*\}_{l=1}^\infty$ of the sequence $\{s_n^*\}_{n=1}^\infty$ which is convergent to some point s^*. By the continuity of $W_t(s)$ we have

$$\begin{aligned}W_t(s^*) &= \lim l \to \infty W_t(s_{n_l}^*) \\ &\leqslant \lim_{l\to\infty} \sum_{j=1}^{n_l} p_j^{(n_l)} W_t(s_j^{(n_l)}) = \int_{R^1} W_t(s) dG(s). \quad \square\end{aligned}$$

For classes obtained by shifting some functions, the reducing property becomes equivalent of convexity. The following theorem specifies this statement.

Theorem 1.3. *Let $v(s)$ be a positive locally integrable function on R^1 such that for each $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that*

$$\inf\{v(s) : |s| \geqslant \epsilon\} > v(0) + \delta. \tag{1.5}$$

In order for the family $\{v(s-t), t \in R^1\}$ to be weakly reducing it is necessary and sufficient that the function v is convex.

Proof. If the function v is convex, then

$$\begin{aligned} pv(s_1 - t) + (1-p)v(s_2 - t) &\geqslant v(p(s_1 - t) + (1-p)(s_2 - t)) \\ &= v(ps_1 + (1-p)s_2 - t) \end{aligned}$$

and condition (1.2) is satisfied with the function $s^* = ps_1 + (1-p)s_2$, i.e., the family $\{v(s-t), t \in R^1\}$ is weakly reducing.

Now, let us assume that the family $\{v(s-t), t \in R^1\}$ is weakly reducing.

First we show the convexity of the function v under the additional assumption that it is continuously differentiable.

We demonstrate that if a positive twice continuously differentiable function v satisfies the condition (1.5) and it is not convex, then there exists a point ξ for which $v'(\xi) > 0$ and $v''(\xi) < 0$. Indeed, let $U = \{s \geqslant 0 : v'(s) > 0\}$. By (1.5), $U \neq \emptyset$. We need to show that for some $\xi \in U$, $v''(\xi) < 0$. Assume the opposite, i.e., $v''(s) \geqslant 0$ for all $s \in U$. Since v' is continuous, for each point $s_0 \in U$ there exists an interval (α, β) such that $s_0(\alpha, \beta) \subseteq U$. Let $[a, b]$ be a maximal interval satisfying the following

1) $(\alpha, \beta) \subseteq [a, b]$;

2) $v''(s) \geqslant 0$ for all $s \in [a, b]$.

Since $v'(s_0) \geqslant 0$ and $v''(s) \geqslant 0$ for $s \in [a, b]$, $v'(s) > 0$ for $s \in [s_0, b]$, i.e., $[s_0, b] \subseteq U$. In particular, if $b < \infty$, then $v'(b) > 0$. Therefore $[b-\eta, b+\eta] \subseteq U$ for some $\eta > 0$. However, by the assumptions, $v''(s) \geqslant 0$ for $s \in U$, and thus also for $s \in [b, b+\eta]$, which contradicts that $[a, b]$ was maximal. This contradiction does not hold only if $b = \infty$. In this way the interval $[s_0, \infty) \subseteq U$ for each $s_0 \in U$. Set $\gamma = \inf U$, so $U = (\gamma, \infty)$. From (1.5) it follows that the function v should have points of increase in an arbitrary small neighborhood of zero, i.e. $\gamma = 0$. Thus the positive function v should be convex which is in contradiction with the assumption. We have thus shown existence of ξ:

$$v'(\xi) > 0, \quad v''(\xi) < 0 \ (\xi > 0).$$

Assume now, that the positive twice differentiable function v satisfying (1.5) is not convex and the family $\{v(s-t), t \in R^1\}$ is weakly reducing. In this case we can find numbers $\xi > 0$ and ϵ_0, $0 < \epsilon_0 < \xi$ such that $v'(s) > 0$, $v''(s) < 0$ for $[\xi - \epsilon_0, \xi + \epsilon_0]$. Let $\delta_0 > 0$ satisfy the condition

$$\inf\{v(s) : |s| \geqslant \epsilon_0/2\} > v(0) + \delta_0.$$

Consider the points s_2, $0 < s_2 < \min(\epsilon_0, \delta_0)$, and $s_1 = -s_2$. Since the family $\{v(s-t), t \in R^1\}$ is weakly reducing, we should have

$$\frac{1}{2}(v(s_1 - t) + v(s_2 - t) \geqslant v(s^* - t), \quad t \in R^1 \tag{1.6}$$

for some s^*, which depends of s_1 and s_2.

By taking in (1.6) $t = 0$, we see that $v(0) \leqslant v(s^*) < v(0) + \delta_0$. From this, it follows that $|s^*| \leqslant \epsilon_0/2$ and $s_1 - \epsilon_0/2 < s^* < s_2 + \epsilon_0/2$. We now take in (1.6) $t = -\xi$, so that the points $s_1 - t$, $s_2 - t$, $s^* - t$ belong to the interval $[\xi - \epsilon_0, \xi + \epsilon_0]$. In this interval, the function v is concave and increasing. Thus, in order for (1.6) to hold with $t = -\xi$, it is necessary that

$$s^* - t < \frac{1}{2}((s_1 - t) + (s_2 - t)),$$

so that

$$s^* < 0. \tag{1.7}$$

Let us take $t = \xi$ in (1.6). Then, the points $s_1 - t$, $s_2 - t$, and $s^* - t$ belong to the interval $[-\xi - \epsilon_0, -\xi + \epsilon_0]$. Because v is positive, in view of the definition of the interval $[\xi - \epsilon_0, \xi + \epsilon_0]$, the function v concave and decreasing on the interval $[-\xi - \epsilon_0, -\xi + \epsilon_0]$. Consequently, if (1.6) holds for $t = \xi$, it is necessary that

$$s^* - t > \frac{1}{2}((s_1 - t) + (s_2 - t)),$$

and so $s^* > 0$. Thus, we obtain a contradiction with (1.7), derived from the assumption that v is not convex.

Thus far we have shown that the weak reducing property of the family $\{v(s-t), t \in R^1\}$ implies the convexity of the function v under the additional (when compared with the conditions of the theorem) assumption about twice continuous differentiability of the function v. To prove the theorem in full generality it is enough to smooth the function v. Namely, let $\omega_\rho(x) \geqslant 0$ be an infinitely differentiable function satisfying the conditions: $\omega_\rho(x) = 0$ if $|x| \geqslant \rho$, $\int_{-\infty}^{\infty} \omega_\rho(x)dx = 1$, $(\rho > 0)$. Since the family $\{v(s-t), t \in R^1\}$ is weakly reducing, for each s_1, s_2 and rational $p \in (0,1)$ the following holds

$$pv(s_1-t)+(1-p)v(s_2-t) \geqslant v(s^*-t), \forall t \in R^1, \tag{1.8}$$

where s^* depends only on s_1, s_2, and p. We multiply both the sides of (1.8) by $\omega_\rho(t-\tau)$ and integrate with respect to t over the interval $(-\infty, \infty)$, to obtain

$$pv_\rho(s_1-\tau)+(1-p)v_\rho(s_2-\tau) \geqslant v_\rho(s^*-\tau), \forall \tau \in R^1, \tag{1.9}$$

where

$$v_\rho(z) = \int\limits_{-\infty}^{\infty} v_(\theta)\omega_\rho(z-\theta)d\theta.$$

Obviously, $v_\rho(z)$ is a infinitely differentiable function and $v_\rho(z) \to v(z)$ when $\rho \to 0$ for almost all z. Condition (1.9) means that the family $\{v\rho(s-t), t \in R^1\}$ is weakly reducing and it is clear that (for sufficiently small $\rho > 0$) this family satisfies condition (1.5). By the previous argument, the function $v_\rho(z)$ is convex and consequently so iz $v(z)$ (as a limit of convex functions). $\square$

Now we consider strongly reducing families generated by shifts of a function.

Theorem 1.4. *Let $v(s)$ be a positive, locally integrable function on R^1, satisfying condition* (1.5). *The family $\{v(s-t), t \in R^1\}$ strongly reducing if and only if the function v is strongly convex.*

Proof. If the function $v(s)$ strongly convex, then the family $\{v(s-t), t \in R^1\}$ is strongly reducing by Jensen's inequality.

Assume now that the family $\{v(s-t), t \in R^1\}$ is strongly reducing. Since strong reducing implies weak reducing, by Theorem 1.3 the function v is convex. We prove that v can not have linear components. Assume the opposite and let $v(x) = ax+b$ for $x \in [\alpha, \beta]$, $(0 < \alpha < \beta)$. From the definition of the strongly reducing property it follows that for each continuous c.d.f. G on R^1 there exists a point $s^* = s^*(G)$ such that

$$\int\limits_{R^1} v(s-t)dG(s) > v(s^*-t), \forall t \in R^1.$$

Let a c.d.f. G be concentrated with equal probability at two points s_1 and s_2, where $s_2 = -s_1 = s$, $0 < s < (\beta-\alpha)/2$. The last equality for this choice of G takes the form

$$v(s-t)+v(-s-t)>2v(s^*-t), \forall t \in R^1. \tag{1.10}$$

For $t=0$, by (1.10) and the positivity of v, we have

$$v(s^*)<v(s),$$

so that

$$-s<s^*<s$$

and $v(z)$ is monotone for $z \geqslant 0$.

By taking $t=-(\alpha+\beta)/2$ in (1.10), the linearity of v on (α,β) implies

$$a(s+(\alpha+\beta)/2)+a(-s+(\alpha+\beta)/2)+2b>2a((\alpha+\beta)/2+s^*)+2b,$$

i.e., $s^*<0$. But for $t=(\alpha+\beta)/2$, it follows from (1.10), by an analogous argument that $s^*>0$. The obtained contradiction proves strong convexity of the function v. □

Now we present several examples of reducing and non-reducing families.

Example 1.2. An arbitrary family $\{v_t(s), t \in T\}$ of convex and real functions on R^1 is reducing.

The proof of this fact is obvious.

Example 1.3. Let $\{v_t(s), t \in T\}$ be an arbitrary family of functions defined on the same interval (a,b) on convex on this interval. We assume that a function h is a one-to-one mapping of R^1 on (a,b). Then the family $\{w_t(s), t \in T\}$, where $w_t(s)=v_t(h(s))$, is reducing.

Proof. Let $G(s)$ be an arbitrary c.d.f. for which there exists the integral

$$\int\limits_{R^1} w_t(s)dG(s).$$

Set

$$s^*=h^{-1}\Big(\int\limits_{R^1} h(s)dG(s)\Big).$$

By Jensen's inequality, we have for $v_t(s)$:

$$\int\limits_{R^1} w_t(s)dG(s) = \int\limits_{R^1} v_t(h(s))dG(s) \geqslant$$

$$\geqslant v_t\Big(\int\limits_{R^1} h(s)dG(s)\Big) = w_t(s^*), \quad \forall t \in T.$$

The existence of the integral $\int_{R^1} h(s)dG(s)$ follows. $\square$

Clearly, one can obtain examples of reducing families characterizing functions which are not convex in any rational sense.

Example 1.4. Let $\{w_t(s), t \in T\}$ be a family of functions which achieve the minimum at the same point s^* (not dependent on t). Then such a family of functions is reducing.

The proof is trivial.

From Theorem 1.4 we have the next result.

Example 1.5. Let $w(z) = |z|$. Then the family $\{w(s-t), t \in R^1\}$ is reducing but not strongly reducing.

Example 1.6. For some $a > 0$, let us set

$$w(z) = \begin{cases} 0: & \text{for } \ z \in (-a, a), \\ 1: & \text{for } \ z \notin (-a, a). \end{cases}$$

Then, the family $\{w(s-t), t \in R^1\}$ is not weakly reducing.

This result is not a consequence of Theorem 1.3 since the condition (1.5) does not hold.

Proof. Let a c.d.f. $G(s)$ be concentrated at the points $-a/2$, $a/2$, assigning them the same probabilities. If the considered family of functions is weakly reducing then we should have

$$(w(-a/2-t) + w(a/2-t))/2 \geqslant w(s^* - t) \tag{1.11}$$

for some values of s^*, not depending on t, and for all $t \in R^1$.

Let us first take $t = -a$ in (1.11). Thus we have $1/2 \geqslant w(s^* + a)$, so that $s^* + a < a$ and $s^* < 0$. Now, put $t = a$ in (1.11) to obtain $1/2 \geqslant w(s^* - a)$, i.e., $s^* - a > -a$, $s^* > 0$. The obtained contradiction concludes the proof. $\square$

2. The classification of a class of estimators according to their completeness types

In this section, as in many others, we are interested in the scheme of repeated selection, i.e., the case when

$$\mathcal{X} = \mathcal{X}_n = X^n, \ \mathcal{A} = \mathcal{A}_n = A^n, \ P_\theta = P_\theta^{(n)} = Q_\theta \times \cdots \times Q_\theta,$$

where $\{Q_\theta, \theta \in \Theta\}$ is a family of probabilistic measures on the measurable space (X, A). For $x \in \mathcal{X}$ we write $x = (x_1, \dots, x_n)$, where $x_j \in X$. In this way, the vector x can be treated as a choice of n independent and identically distributed observations $x_1, \dots, x_n$.

Below we define four completeness properties for classes of estimators and we investigate its mutual relations.

A). Assume that on the basis of $n-1$ observations $x_1, \dots, x_{n-1}$, we have an estimator δ_{n-1}^* of a certain parameter function $\gamma : \Theta \to R^1$. We want to know whether on the basis of n observations $x_1, \dots, x_n$ one can obtain an estimator δ_n^* which would be essentially better than δ_{n-1}^* for each measure $P_\theta^{(n)}$, i.e., for which the following inequality would hold

$$R_\theta(\delta_n^*, \gamma; P_\theta^{(n)}; w) < R_\theta(\delta_{n-1}^*, \gamma; P_\theta^{n-1}; w), \ \forall \theta \in \Theta. \tag{2.1}$$

In other words, we are interested in the question if the new observation x_n contains information about the parameter function γ which could be exploited for reducing the risk of estimation of this parameter function.

Clearly, if for some $\theta_0 \in \Theta$ the estimator δ_{n-1}^* is constant with $P_{\theta_0}^{(n-1)}$-probability one and $\delta_{n-1}^* = \gamma(\theta_0)$, then

$$R_{\theta_0}(\delta_{n-1}^*, \gamma; P_{\theta_0}; w) = 0$$

and thus the risk of such an estimator can not be essentially reduced. (We say that δ_{n-1}^* is constant with $P_{\theta_0}^{(n-1)}$-probability one if δ_{n-1}^* is a degenerated measure concentrated at some point $\gamma_{n-1}^*(x)$ and $P_{\theta_0}^{(n-1)}\{\gamma_{n-1}^*(x) = \text{const}\} = 1$.) Thus, we are interested only in improving the estimators δ_{n-1}^* which are not constant. Such an approach does not restrict the generality, as the constant estimators in the given sense are useless (they do not follow the parameter).

If in the inequality (2.1) we fix all elements in the model except the loss function, then the obtained property becomes a property of the loss function which characterizes a possibility of reducing the risk by using the information included in the n-th observation. This is reflected in the following definition.

Definition 2.1 (See [7, 8]). *We say that a loss function w satisfies the condition of complete information (CI-condition), if for each family of distributions $\{Q_\theta, \theta \in \Theta\}$, each integer $n \geqslant 2$, each parameter function γ, and an arbitrary random estimator δ^*_{n-1} ($\delta^*_{n-1} \neq const\ P_\theta^{(n-1)}$–almost surely for each $\theta \in \Theta$) based on $n-1$st observations, there exists an estimator δ^*_n based on n observations for which inequality* (2.1) *holds.*

The CI condition means that in order to construct "good" estimators one has to use all available observations. On the other hand, using the loss function which does not satisfy the condition of complete information is not reasonable. In the extreme case, it requires additional assumptions and can lead to ignoring some of equally important observations.

B). It is interesting to consider non-randomized estimators. Indeed, randomized estimators use for their construction some additional random mechanism. This mechanism, as it appears at the first sight, should often lead to reduction of estimation accuracy, i.e., should appear to be useless. Obviously, this statement is, in general, incorrect but it is interesting to explain situations in which it can happen. We again want to modify the property of completeness of the class of non-randomized estimators so it will be related only to a loss function.

Definition 2.2 (See [8]). *We say that a loss function w satisfies the condition of lack of randomization (LR-condition), if for each family of distributions $\{P_\theta, \theta \in \Theta\}$, each parameter function γ and an arbitrary randomized estimator δ^* there exists an estimator γ^* which belongs to the class $K^*_{n.r.}$ of all non-randomized estimators and such that*

$$R_\theta(\gamma^*, \gamma; P_\theta; w) \leqslant R_\theta(\delta^*, \gamma; P_\theta; w),\ \forall \theta \in \Theta. \tag{2.2}$$

In other words, for each randomized estimator there exists a non-randomized estimator for which the risk is not larger than for this randomized estimator, i.e., when estimating (for an arbitrary family of distributions) one can restrict to non-randomized estimators.

C). Let us have a scheme of repeated choice of the sample size n. We assume that we admit only non-randomized estimators based on this choice. Since all observations $x_1, \dots, x_n$ are equivalent (i.e., independent and identically distributed), thus it seems reasonable to consider for construction of estimator only symmetric functions of the observations $x_1, \dots, x_n$. To address this we introduce the following definition.

Definition 2.3 (See [6, 5]). *We say that a loss function w satisfy the condition of symmetrization (S condition), if for each family of proba-*

bilistic measures $\{Q_\theta, \theta \in \Theta\}$, *an integer* $n \geqslant 2$, *parameter function* γ *and its non-randomized estimator* $\gamma^* : \mathcal{X}_n \to R^1$, *there exists a statistic* $\hat{\gamma}^* \in K^*_{n.r.}$, *symmetric with respect to the coordinates of* $x = (x_1, \ldots, x_n)$, *such that*

$$R_\theta(\hat{\gamma}^*, \gamma; P_\theta^{(n)}; w) \leqslant R_\theta(\gamma^*, \gamma; P_\theta^{(n)}; w), \ \forall \theta \in \Theta. \tag{2.3}$$

The S-condition can be interpreted in the following way. We have n measurements giving independent measurements x_j, $(j = 1, \ldots, n)$ of the same physical quantity $\gamma(\theta)$. In this situation an arbitrary "good" assessment of the value of $\gamma(\theta)$ is independent of the order in which the observations were received from measurement devices (or observers) and then statistically analyzed.

D). The last requirement is a certain strengthening of the S condition.

Let $\{P_\theta, \theta \in \Theta\}$ be a family of probabilistic measures on a measurable space $(\mathcal{X}, \mathcal{A})$ for which there exists a sufficient statistics T for the parameter $\theta \in \Theta$. (For the reasons of introduction of the concept of sufficient statistic see, for example, [11, 2]. Below when saying about sufficiency in the RB condition, we assume that for considered families there exists a regular conditional probability given a value of the sufficient statistics T.) If a loss function $w(s,t)$ is convex in s for each fixed t, then by the Rao–Blackwell theorem (see, e.g., [13,1]), non-randomized estimators which are functions of the sufficient statistic T constitute a complete subclass of the class of all non-randomized estimators $K^*_{n.r.}$. The Rao–Blackwell theorem does not say anything about necessity of convexity of $w(s,t)$ in s. Thus, it is of interest to describe the loss functions for which an analog of the Rao–Blackwell theorem holds, i.e., loss functions satisfying the following definition.

Definition 2.4. *We say that a loss function* w *satisfies the Rao–Blackwell condition (RB-condition), if for an arbitrary measurable space* $(\mathcal{X}, \mathcal{A})$ *on which there is defined a family of distributions* $\{P_\theta, \theta \in \Theta\}$ *having a sufficient statistics* T, *and for an arbitrary non-randomized estimator* γ^* *of parameter function* γ, *there exists a non-randomized estimator* $\tilde{\gamma}^*$ *of the same parameter function which depends only on the sufficient statistic* T *and such that*

$$R_\theta(\tilde{\gamma}^*, \gamma; P_\theta; w) \leqslant R_\theta(\gamma^*, \gamma; P_\theta; w) \ \forall \theta \in \Theta. \tag{2.4}$$

Clearly, if a loss function w satisfies the RB condition, then it also satisfies the S-condition.

Note that the requirement that one of the four formulated conditions is satisfied, in a sense, reduces the role of "randomness" in estimation. The CI-condition means that the addition of a new observation can be used for reduction of the role of randomness by reducing the estimation risk. The LR-condition discards additional random mechanism in the construction of an estimator. The S-condition rejects randomness related to the order of obtaining data. The RB-condition allows for rejection of the non-informative component in the selection, since a sufficient statistics "contains all information" about a considered family.

There exist theorems stating that in the scheme of repeated selection symmetric randomized estimators constitutes a complex class. From our point of view this statement does not have much essence since randomness related to the order in which data are obtained is equivalent to the random mechanism used for computation of the value of an estimator.

Thus, through Definitions 2.1–2.4, we have introduced four properties of the loss function, induced by the conditions of completeness of certain classes of estimators. We now turn to the description of the loss functions satisfying these conditions.

Theorem 2.1 (See [9]). *Let a loss function w satisfy one of the following conditions*

a) $\sup_{s,t} w(s,t) < \infty$;

b) *for each $s_1 \geqslant s_2 \geqslant t \geqslant s_3 \geqslant s_4$ the following conditions are satisfied*

$$w(s_1,t) \geqslant w(s_2,t), \ w(s_4,t) \geqslant w(s_3,t), \ \sup_s w(s,t) = \infty.$$

Then the function $w(s,t)$ satisfies the CI-condition if and only if the family $\{w(s,t), t \in R^1\}$ is strongly reducing.

Proof. Let us assume that $w(s,t)$ satisfies the CI-condition and we will show that the family $\{w(s,t), t \in R^1\}$ is strongly reducing.

We need to demonstrate that for an arbitrary continuous c.d.f. $G(\delta)$ on R^1 there exist degenerated c.d.f. $H(\delta)$ concentrated at a point γ^* which depends on G and such that for all $t \in R^1$

$$\int_{R^1} w(\delta,t)dG(\delta) > \int_{R^1} w(\delta,t)dH(\delta) = w(\gamma^*,t). \tag{2.5}$$

Indeed, let $X = \{0,1\}$ and let $\mathcal{A}$ be a set of all subsets of X. Set $\mathcal{X} = X^2$, $P_\theta = Q_\theta \times Q_\theta$, $x = (x_1,x_2) \in \mathcal{X}$. Let $G_1(\delta)$ be an arbitrary non-degenerated c.d.f. on R^1, and let $\gamma(\theta)$ be an arbitrary (temporarily) parameter function. We assume that an estimator δ_1^* is such that for $x = 1$ it induces on R^1 a distribution which coincide with $G_1(\delta)$ and in other cases it is arbitrary but not constant almost everywhere with respect to measures Q_θ, $\theta \in [0,1]$. By the condition of the complete information there exists an estimator $\delta_2^* : \mathcal{X} \to \Pi_1$ for which

$$R_\theta(\delta_2^*, \gamma; P_\theta; w) < R_\theta(\delta_1^*, \gamma; Q_\theta; w), \quad \theta \in [0,1]. \tag{2.6}$$

Set $\theta = 0$. The measure Q_0 is concentrating at the point $\{1\}$, while the measure P_0 at the point $\{1,1\}$. If δ_2^* with $x = (1,1)$ is inducing the distribution $G_2(\delta)$ on R^1, then inequality (2.6) produces

$$\int\limits_{R^1} w(\delta,t)dG_1(\delta) > \int\limits_{R^1} w(\delta,t)dG_2(\delta), \tag{2.7}$$

where $\gamma(0) = t$. Note that the value of t is arbitrary, since the parameter function γ can be arbitrary.

If in (2.7) the c.d.f. $G_2(\delta)$ is degenerated, then we obtain (2.5). Assume the opposite, that is that $G_2(\delta)$ is a non-degenerate c.d.f. Since G_1 is an arbitrary non-degenerate c.d.f., then we can repeat our arguments, taking as the initial function $G_2(\delta)$. Then, we obtain that there exists a non-degenerated c.d.f. $G_3(\delta)$, for which

$$\int\limits_{R^1} w(\delta,t)dG_2(\delta) > \int\limits_{R^1} w(\delta,t)dG_3(\delta).$$

Repeating all these arguments we obtain a sequence of non-degenerated c.d.f. G_1, G_2, $\ldots, G_m, \ldots$, for which the following inequalities hold

$$\begin{aligned} \int\limits_{R^1} w(\delta,t)dG_1(\delta) > \int\limits_{R^1} w(\delta,t)dG_2(\delta) > \cdots > \\ > \cdots > \int\limits_{R^1} w(\delta,t)dG_m(\delta) > \cdots . \end{aligned} \tag{2.8}$$

Let $\{G_\xi, \xi \in \Xi\}$ (where Ξ is some linearly ordered set) be the family of all non-degenerated c.d.f.'s satisfying the condition

$$\int\limits_{R^1} w(\delta,t)dG_1(\delta) > \int\limits_{R^1} w(\delta,t)dG_\xi(\delta) > \int\limits_{R^1} w(\delta,t)dG_{\xi'}(\delta)$$

for $\xi' > \xi$ and all t. Fix an arbitrary $t = t_0$ and set

$$d = \inf_{\xi \in \Xi} \int\limits_{R^1} w(\delta, t_0) dG_\xi(\delta).$$

There exists a sequence $\xi_1, \xi_2, \ldots, \xi_m, \ldots$ such that

$$\lim_{m \to \infty} \int\limits_{R^1} w(\delta, t_0) dG_{\xi_m}(\delta) = d.$$

Without loss of generality, we can assume that $G_m = G_{\xi_m}$ and that the sequence G_m is convergent (otherwise we select a convergent subsequence). We set

$$H(\delta) = \lim_{m \to \infty} G_m(\delta).$$

Clearly, $0 \leqslant H(\delta) \leqslant 1$ and $H(\delta)$ is a monotone function. If $H(\delta)$ was non-degenerate c.d.f., then instead of G_1 we would take again $H(\delta)$ and construct a c.d.f. $H_1(\delta)$ satisfying

$$\int\limits_{R^1} w(\delta, t) dH(\delta) > \int\limits_{R^1} w(\delta, t) dH_1(\delta)$$

for all t. This would mean that

$$\int\limits_{R^1} w(\delta, t) dG_\xi(\delta) > \int\limits_{R^1} w(\delta, t) dH_1(\delta), \forall \xi \in \Xi,$$

$$\inf_{\xi \in \Xi} \int\limits_{R^1} w(\delta, t) dG_\xi(\delta) > \int\limits_{R^1} w(\delta, t) dH_1(\delta).$$

The assumption that H_1 is non-degenerate has led us to the contradiction with the definition of the family $\{G_\xi, \xi \in \Xi\}$ and the number d. This means that H_1 has to be a degenerate c.d.f. and, in such a case, while $H(\delta)$ is a non-degenerated c.d.f., the statement (2.5) (with H replaced by H_1) is proven.

Thus, we can assume that either H is a degenerated c.d.f. and then (2.5) follows, or H is not a c.d.f.

We shall that H is a c.d.f. In view of the above comments, it will be sufficient for a complete proof of (2.5).

Assume first that the condition b) of the theorem holds and let A_1, A_2 be two numbers satisfying the inequalities $A_1 > t > A_2$. Then,

$$\int\limits_{R^1} w(\delta,t)dG_m(\delta) \geqslant \int\limits_{\delta\geqslant A_1} w(A_1,t)dG_m(\delta) = w(A_1,t)(1-G_m(A_1)),$$

$$\int\limits_{R^1} w(\delta,t)dG_m(\delta) \geqslant \int\limits_{\delta\leqslant A_2} w(A_2,t)dG_m(\delta) = w(A_2,t)G_m(A_2).$$

By (2.8), we have

$$1-G_m(A_1) \leqslant \int\limits_{R^1} w(\delta,t)dG_1(\delta)/w(A_1,t),$$

$$G_m(A_2) \leqslant \int\limits_{R^1} w(\delta,t)dG_1(\delta)/w(A_2,t).$$

The last inequalities, together with the condition b) of the theorem, imply the uniform smallness of the tails of the distribution G_m, $m = 1,2,\ldots$. Therefore, the limiting function H should be a c.d.f. Thus, under the condition b), the function H is a c.d.f.

We assume now that the condition a) holds. Set

$$A = \sup_{s,t} w(s,t)$$

and consider

$$\rho(s,t) = A - w(s,t).$$

The quantity ρ have the meaning of a winnings, and instead of minimizing the mean loss we can consider maximizing the mean winnings. Obviously, $\rho(s,t) \geqslant 0$ and instead of (2.8) we have the equivalent inequality

$$\int\limits_{R^1} \rho(\delta,t)dG_1(\delta) < \int\limits_{R^1} \rho(\delta,t)dG_2(\delta) < \cdots < \int\limits_{R^1} \rho(\delta,t)dG_m(\delta) < \cdots.$$

Thus, for the limiting function $H(\delta)$, we have

$$\int\limits_{R^1} \rho(\delta,t)dG_m\delta < \int\limits_{R^1} \rho(\delta,t)dH(\delta).$$

From this, it follows that without loss of generality the function H can be assumed to be a c.d.f. Indeed, since $H(\delta) = \lim_{m\to\infty} G_m(\delta)$, the function H is monotone and $0 \leqslant H(-\infty) \leqslant H(\delta) \leqslant H(\infty) \leqslant 1$. Passing, if needed, to the function $H(\delta) - H(-\infty)$ (such a replacement does not change the values of integrals) we can assume that $H(-\infty) = 0$. If $H(\infty) = 1$, then H is a c.d.f. If $H(\infty) < 1$, then we define

$$H_1(\delta) = H(\delta) + F(\delta),$$

where $F(\delta)$ is an arbitrary monotone function satisfying the conditions

$$\lim_{\delta\to-\infty} F(\delta) = 0, \lim_{\delta\to\infty} F(\delta) = 1 - H(\infty).$$

Then, H_1 is a c.d.f. and

$$\int\limits_{R^1} \rho(\delta,t)dG_m(\delta) < \int\limits_{R^1} \rho(\delta,t)dH(\delta) < \int\limits_{R^1} \rho(\delta,t)dH_1(\delta).$$

Thus, the assumption that H_1 is non-degenerate contradicts the definitions of the family $\{G_\xi, \xi \in \Xi\}$ and the number d.

Thus, we can assume that the above argument completes the construction of the degenerate c.d.f. H and (2.5) is established. Consequently, the family $\{w(s,t), t \in R^1\}$ has to be strongly reducing.

Now, we assume that the family $\{w(s,t), t \in R^1\}$ is strong reducing, and we shall prove that the loss function $w(s,t)$ satisfies the condition of the full information.

Let $\{Q_\theta, \theta \in \Theta\}$ be a family of probability measures over a measurable space $(X, \mathcal{A})$. Consider a sample $(x_1, \ldots, x_n) = x$ $(n \geqslant 2)$ from this domain and assume that $\delta^*_{n-1} : (x_1, \ldots, x_{n-1}) \to \delta^*_{n-1,(x_1,\ldots,x_{n-1})}(\cdot)$ is an arbitrary randomized estimator of some parameter function γ which is based on $n-1$ observations. We shall construct a probability measure on R^1 by setting for $x = (x_1, \ldots, x_n)$

$$\delta^*_{n,x}(\cdot) = (\delta^*_{n-1,(x_1,\ldots,x_{n-1})}(\cdot) + \delta^*_{n-1,(x_2,\ldots,x_n)}(\cdot))/2,$$

and letting $\tilde{\delta}^*_{n,x}(s)$ be a c.d.f. corresponding to this measure.

Clearly, if the estimator δ^*_{n-1} is not constant almost surely with respect to each measure $P_\theta^{(n-1)}$, then $\tilde{\delta}^*_{n,x}(s)$ is a non-degenerate c.d.f. By the strong reducing property of the family $\{w(s,t), t \in R^1\}$, there exists a point $\gamma^*_n(x)$ such that

$$\int\limits_{R^1} w(\delta, \gamma(\theta))d_n\tilde{\delta}^*(\delta) > w(\gamma^*_n(x), \gamma(\theta)), \forall\theta \in \Theta.$$

However,

$$
\begin{aligned}
R_\theta(\delta^*_{n-1},\gamma;P_\theta^{(n-1)};w) &= \int\limits_{\mathcal{X}_{n-1}} dP_\theta^{(n-1)}(x)\int\limits_{R^1} w(\delta,\gamma(\theta))d\delta^*_{n-1,(x_1,\ldots,x_n)}(\delta)\\
&= \int\limits_{\mathcal{X}} dP_\theta^{(n)}(x)\int\limits_{R^1} w(\delta,\gamma(\theta))d\tilde{\delta}^*_n(\delta)\\
&> \int\limits_{\mathcal{X}} w(\gamma^*_n(x),\gamma(\theta))dP_\theta^{(n)}(x)\\
&= R_\theta(\gamma^*_n,\gamma;P_\theta^{(n)};w),
\end{aligned}
$$

which concludes the proof of the theorem. $\square$

The following simple theorem yields a description of the loss functions which satisfy the condition of redundancy of randomization.

Theorem 2.2 (See [9]). *A loss function $w(s,t)$ satisfies the condition of redundancy of randomization if and only if the family $\{w(s,t), t \in R^1\}$ is reducing.*

Proof. Assume first that the family $\{w(s,t), t \in R^1\}$ is reducing. Let $\{\mathcal{X},\mathcal{A},P_\theta\}$, $\theta \in \Theta$, be an arbitrary measurable space with a family of probability measures on it. Let γ be an arbitrary parameter function, and let δ^* be its randomized estimator. Since the family $\{w(s,t), t \in R^1\}$ is reducing, for an arbitrary c.d.f. G on R^1 there exists $s^*(G)$ such that

$$\int\limits_{R^1} w(\delta,t)dG(\delta) \geqslant w(s^*(G),t),\ \ \forall t \in R^1.$$

Set $\gamma^*(x) = s^*(\delta^*_x)$ to obtain

$$\int\limits_{R^1} w(\delta,\gamma(\theta))d\delta^*_x(\delta) \geqslant w(\gamma^*(x),\gamma(\theta)).$$

By taking the expected value of both the sides of the above expression, we obtain

$$R_\theta\delta^* = E_\theta \int\limits_{R^1} w(\delta,\gamma(\theta))d\delta^*_x(\delta) \geqslant E_\theta w(\gamma^*(x),\gamma(\theta)) = R_\theta\gamma^*,$$

so that w satisfies the LR-condition.

Now, we assume that the function w satisfies the condition of redundancy of randomization. We shall show that the family $\{w(s,t), t \in R^1\}$ is reducing.

Let $\mathcal{X} = \{0,1\}$, and let $\mathcal{A}$ be the set of all subsets of $\mathcal{X}$. Let $P_\theta(\{0\}) = \theta$ and $P_\theta(\{1\}) = 1-\theta$, where $\theta \in [0,1]$. Let $\gamma(\theta)$ be an arbitrary parameter function, and let $G(\delta)$ be an arbitrary c.d.f. on R^1. Consider a randomized estimator δ^* which for $x=1$ is inducing on R^1 a distribution coinciding with $G(\delta)$, otherwise being arbitrary. By the LR-condition, there exists a non-randomized estimator γ^* for which

$$R_\theta \gamma^* \leqslant R_\theta \delta^*,\ \theta \in [0,1].$$

For $\theta = 0$ the last inequality gives

$$\int\limits_{R^1} w(\delta, \gamma(0)) dG(\delta) \geqslant w(\gamma^*(1), \gamma(0)).$$

Because $\gamma(0)$ and $G(\delta)$ are arbitrary, the above inequality implies that the family $\{w(s,t), t \in R^1\}$ is reducing. $\square$

It is not difficult to obtain a theorem describing the loss functions satisfying the symmetrization condition.

Theorem 2.3 (See [9]). *A loss function w satisfies the condition of symmetrization if and only if the family $\{w(s,t), t \in R^1\}$ is weakly reducing.*

Proof. Assume first that the family $\{w(s,t), t \in R^1\}$ is weakly reducing. Let $(\mathcal{X}, A)$ be an arbitrary measurable space with a given family $\{Q_\theta, \theta \in \Theta\}$ of probabilistic measures defined on it, and let $n \geqslant 2$ be an arbitrary integer. Set $\mathcal{X} = X^n$, $\mathcal{A} = A^n$, $P_\theta = Q_\theta \times \cdots \times Q_\theta$. Let $\gamma^*(x_1, \ldots, x_n)$ be an estimator of a parameter function $\gamma(\theta)$, based on the vector of observations $x = (x_1, \ldots, x_n)$. Since $\{w(s,t), t \in R^1\}$ is weakly reducing, by Theorem 1.1 there exists a statistic $\tilde{\gamma}^*$ such that

$$\frac{1}{n!} \sum_\sigma w(\gamma^*(\sigma(x_1, \ldots, x_n)), t) \geqslant w(\tilde{\gamma}^*(x_1, \ldots, x_n), t), t \in R^1. \qquad (2.9)$$

Here, the summation is over all $n!$ permutations σ of the values $x_1, \ldots, x_n$. By Remark 1.1, the quantity γ^* can be treated as a symmetric function of $\gamma^*(\sigma(x_1, \ldots, x_n))$ and thus also of the coordinates $x_1, \ldots, x_n$. Substituting in (2.9) the value of the parametric function $\gamma(\theta)$ and computing the expected values of both sides, we obtain

$$R_\theta \gamma^* \geqslant R_\theta \tilde{\gamma}^*, \theta \in \Theta.$$

The last inequality proves the S-condition for the loss function w.

We assume now that a loss function w satisfies the S-condition, and we shall show that the family $\{w(s,t), t \in R^1\}$ is weakly reducing.

Let $X = \{0,1\}$, let A be the set of all subsets of X. Let $Q_\theta(\{0\}) = \theta$ and $Q_\theta(\{1\}) = 1-\theta$, where $\theta \in (0,1)$, and let $n \geqslant 2$ be an arbitrary integer. Set $\mathcal{X} = X^n$, $\mathcal{A} = A^n$, $P_\theta = Q_\theta \times dots \times Q_\theta$. Let $t \in R^1$ be arbitrary. Consider an arbitrary parameter function γ, satisfying the condition $\lim_{\theta\to 1}\gamma(\theta) = t$ and constant in a certain interval $[\delta, 1)$ $(0 < \delta < 1)$. Assume that $\gamma^*(x_1,\dots,x_n)$ is an estimator of γ and $\gamma^*(0,\dots,0) = t$, with other values of γ^* arbitrary. By the S-condition, there exists a symmetric in $x_1,\dots,x_n$ estimator of $\tilde\gamma^*$ for which

$$R_\theta\tilde\gamma^* \leqslant R_\theta\gamma^*, \theta \in (0,1). \tag{2.10}$$

Let us take the limit when $\theta \to 1$ in (2.10). By the condition $\lim_{\theta\to 1}\gamma(\theta) = t = \gamma^*(0,\dots,0)$, the limit of the right hand side is equal to zero. Consequently, the limit of the left hand side should be equal to zero as well. This is possible only when $t = \tilde\gamma^*(0,\dots,0)$.

Dividing both sides of (2.10) by $(1-\theta)$ and then passing again to the limit $\theta \to 1$, we find

$$\begin{aligned} w(\gamma^*(1,0,\dots,0),t) + w(\gamma^*(0,1,0,\dots,0),t) + \dots + \\ + w(\gamma^*(0,0,\cdots,1),t) \geqslant nw(\tilde\gamma^*(1,0,\dots,0),t). \end{aligned}$$

Since the values $\gamma^*(1,0,\dots,0)$, $\gamma^*(0,1,\dots,0)$,$\dots$, $\gamma^*(0,0,\dots,1)$ can be chosen arbitrarily and $t \in R^1$ is also arbitrary, the last inequality and Theorem 1.1 demonstrate that the family $\{w(s,t), t \in R^1\}$ is weakly reducing. $\square$

We turn to studying the Rao–Blackwell condition.

Theorem 2.4. *In order for a loss function w to satisfy the RB-condition it is necessary and sufficient that the family $\{w(s,t), t \in R^1\}$ be reducing.*

Proof. Let $\{P_\theta, \theta \in \Theta\}$ be an arbitrary family of probability measures on a space $(\mathcal{X}, \mathcal{A})$, admitting a sufficient statistics T. The condition that the family $\{w(s,t), t \in R^1\}$ is reducing means that

$$E_\theta\{w(\gamma^*(x),t)|T\} \geqslant w(\tilde\gamma^*(T),t),\ \forall t \in R^1,$$

i.e.,

$$E_\theta\{E_\theta\{w(\gamma^*(x),t)|T\}\} \geqslant E_\theta w(\tilde\gamma^*(T),t).$$

This means that the RB-condition is valid.

Now we assume that w satisfies the RB-condition. Consider two random variables. The variable x_1 takes only two values – zero or one, and the variable x_2 can take arbitrary real values. The resulting joint distributions have the form

$$P_\theta\{x_1 = 0, x_2 < s\} = \theta G(s),$$
$$P_\theta\{x_1 = 1, x_2 < s\} = (1-\theta)G(s), \theta \in [0,1],$$

where $G(s)$ is an arbitrary (fixed) c.d.f. Obviously, $T = x_1$ is a sufficient statistics for the parameter θ. Let $\gamma^*(x_1, x_2)$ be an estimator of an arbitrary parameter function γ. By the RB-condition, there exists an estimator $\tilde{\gamma}^*$ depending only on $T = x_1$, such that

$$R_\theta \tilde{\gamma}^* \leqslant R_\theta \gamma^*, \quad \theta \in [0,1].$$

The last inequality for $\theta = 1$ gives

$$\int\limits_{R^1} w(\gamma^*(0,s), \gamma(1)) dG(s) \geqslant w(\tilde{\gamma}^*(0), \gamma(1)).$$

By the arbitrariness of $\gamma(1)$ and of the c.d.f. $G(s)$, the last inequality implies that the family $\{w(s,t), t \in R^1\}$ is reducing. $\square$

We assume that a loss function $w(s,t)$ satisfies one of the conditions a) or b) from Theorem 2.1. From the proven theorems it follows that the following implications hold

CI-condition $\Rightarrow$ LR-condition $\Leftrightarrow$ RB-condition $\Rightarrow$ S-condition.

Now, we consider several examples.

Example 2.1. The Laplace loss function

$$w(s,t) = |s - t|$$

satisfies the LR, RB and S conditions, but does not satisfy the CI-condition.

This fact follows from Theorems 2.1–2.4 and Example 1.5.

Example 2.2. The loss function

$$w(s,t) = \begin{cases} 0 & \text{for } |s-t| < a, \\ 1 & \text{for } |s-t| \geqslant a, \end{cases}$$

where $a > 0$ is a given number, does not satisfy the S-condition. (This loss function corresponds to a confidence interval of length $2a$.)

This statement follows from Theorem 2.3 and Example 1.6.

Note that the quadratic loss function satisfies all conditions: CI, LR, RB and S. In this aspect it has more positive properties than the Laplace function.

Let us consider in more detail the case of a loss function depending on difference of arguments.

Theorem 2.5. *Let $v(x)$ ($v(0) = 0$) be positive, locally integrable function on R^1, such that for each $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that*

$$\inf\{v(s) : |s| \geq \epsilon\} \geq \delta.$$

Then, the loss function $w(s,t) = v(s-t)$ satisfies at least one of the conditions LR, RB, *or* S *if and only if the function v is convex.*

Proof. The proof follows immediately from Theorems 2.2–2.4 and Theorem 1.3. □

Theorem 2.6. *Let $v(x)$ ($v(0) = 0$) be positive, locally integrable function on R^1, strongly increasing for $z \geqslant 0$. The loss function $w(s,t) = v(s-t)$ satisfies the condition of the full information if and only if the function $v(z)$ is strongly convex.*

Proof. The proof follows directly from Theorems 2.1 and 1.4. □

3. Concluding remarks

We saw that the use of loss functions depending on the difference of its arguments together with some desirable properties leads to the class of convex (strictly convex) loss functions. In other words, it leads to ill-posedness of the problem of statistical estimation. Therefore we have to use bounded loss functions, which are non-invariant with respect to the shifts of their arguments. Some classes of such loss functions were given in [3]. Let us note, that most of the results of this paper were obtained by L. B. Klebanov in more restrictive form (see [3]).

References

1. D. Blackwell, *Conditional expectation and unbiased sequential estimation.* — Ann. Math. Statist. **18(1)** (1947), 105–110.
2. A. M. Kagan, Yu. V. Linnik, and C. R. Rao, *Characterization Problems of Mathematical Statistics.* Wiley, J., New York (1973).
3. A. V. Kakosyan, L. B. Klebanov, J. A. Melamed, *Choice of a Model in Statistical Estimation Theory.* Metzniereba, Tbilisi (1986).

4. L. B. Klebanov, Yu. V. Linnik, and A. L. Rukhin, *Unbiased estimation and matrix loss functions.* — Dokl. Akad. Nauk SSSR **200(5)** (1971), 1024–1025.
5. L. B. Klebanov, *The characterization of loss functions in the theory of estimation.* — Zap. Nauchn. Semin. LOMI **53**, (1975) 130–141.
6. L. B. Klebanov, *Characterization problems of loss functions in the theory of estimation.* Proceedings of the Russian-Japaneese Symphosium in Probability Theory, Taskient (1975), pp. 28–29.
7. L. B. Klebanov, *On use of randomization in the theory of statistical estimation.* 12th European meeting of Statisticians, Abstracts, Varna, Bulgaria 1979, p. 129.
8. L. B. Klebanov, *On the problem of model selection in the theory of statistical estimation.* Proceedings of the 3rd International Vilnius Conference on Probability Theory and Mathematical Statistics, Vilnius, **1** (1981), pp. 231–232.
9. L. B. Klebanov, *On the problem of chosing a loss function in the theory of statistical estimation.* — Teor. Veroyatnost. Primenen. **29(4)** (1984), 812.
10. N. A. Lebedev, Yu. V. Linnik, and A. L. Rukhin, *Convex and monotone loss functions in statistics.* — Trudy Mat. Inst. Steklov **62**, (1971) 291–299.
11. E. L. Lehmann, *Testing Statistical Hyphotheses.* Wiley, New York (1959).
12. Yu. V. Linnik and A. L. Rukhin, *Matrix loss functions admitting the rao-blackwellization.* — Sankhyā A **34(1)**, (1972) 1–4.
13. C. R. Rao, *Information and accuracy attainable in estimation of statistical parameters..* — Bull. Cal. Math. Soc. **37**, (1945) 81–91.

С.-Петербургский
государственный университет
архитектуры и строительства

Поступило 15 декабря 2000 г.