PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 1, Fasc. 1 (1980), p. 29-47

ON TWO NECESSARY σ-FIELDS AND ON UNIVERSAL LOSS FUNCTIONS

BY

ANDRZEJ KOZEK (WROCŁAW)

Abstract. It is shown that the class of best unbiased estimators can be characterized by two necessary σ-fields $\mathscr{S}$ and $\mathscr{U}$. The "large" σ-field $\mathscr{S}$ is a makeshift of the minimal sufficient σ-field whereas the "small" σ-field $\mathscr{U}$ is a makeshift of the maximal complete σ-field. Each estimator which is best unbiased for a strictly convex loss function is $\mathscr{S}$-measurable. Every $\mathscr{U}$-measurable estimator is best unbiased for arbitrary convex loss. Relations of properties of $\mathscr{S}$ and $\mathscr{U}$ with the structure of the class of best unbiased estimators and with properties of universal loss functions are investigated.

1. Introduction. Let $(T, \mathscr{A}, \mathscr{P})$ be a statistical space. Since the fundamental paper [11] of Lehmann and Scheffé was published in 1950 it has been known that if the minimal sufficient and complete σ-field $\mathscr{M}$ exists, then (in view of the Rao-Blackwell theorem) for each estimable function of parameters there exists a unique $\mathscr{M}$-measurable estimator which is uniformly best unbiased for an arbitrary convex loss function. In 1957 Bahadur [1] investigated the structure of estimators which are best unbiased for quadratic loss functions. He proved, making no assumptions on the existence of the minimal sufficient and complete σ-field, that there exists a necessary σ-field $\mathscr{U}$ such that each $\mathscr{U}$-measurable estimator has the uniformly minimum variance. In 1970 Padmanabhan [14] and in 1972 Strasser [21] noted that the $\mathscr{U}$-measurable estimators are uniformly best unbiased for an arbitrary convex loss function. In 1970-1974 Padmanabhan [14], Schmetterer [18]-[20], Strasser [20], [21], Linnik [7], [12], Klebanov [5]-[7] and Ruhin [7], [12] proved several characterizations of $\mathscr{U}$-measurability. Their theorems state that if some additional conditions are satisfied and an estimator is uniformly best unbiased for a strictly convex loss function of a special form, then this estimator is uniformly best unbiased for each convex loss function. Loss functions having such a characteristic peculiarity have been called *universal loss functions* (see [5] and [6]). Moreover, there was a hope that each strictly convex loss function is universal or, at least, that universal loss functions include a large class of convex loss functions (cf. [21], Remark 4.9, and [18], Section 5). It is true that the situation was not too clear and that the Bahadur σ-field $\mathcal{U}$ had very interesting properties.

In 1978 Bednarek-Kozek and Kozek [2] showed that there exist statistical spaces for which very natural strictly convex loss functions are not universal. On the other hand, Proposition 3.5 in [8] suggested that the maximal necessary σ-field $\mathcal{S}$ is important for the theory of unbiased estimation.

In Section 2 we show that in the general case, where the minimal sufficient and complete σ-field does not exist, two necessary σ-fields $\mathcal{S}$ and $\mathcal{U}(\mathcal{E})$ describe the structure of the class of best unbiased estimators. The "large" σ-field $\mathcal{S}$ is the greatest necessary σ-field for $(T, \mathcal{A}, \mathcal{P})$ and is related to sufficiency. Each estimator which is uniformly best unbiased for a strictly convex loss function is $\mathcal{S}$-measurable (Theorem 1). The "small" σ-field $\mathcal{U}(\mathcal{E})$ depends on the class $\mathcal{E}$ of considered estimators and is related to the completeness. An estimator in $\mathcal{E}$ is $\mathcal{U}(\mathcal{E})$-measurable if and only if it is uniformly best unbiased for an arbitrary convex loss function (Theorem 2). If $\mathcal{E}$ is the class of all square integrable estimators, then $\mathcal{U}(\mathcal{E})$ coincides with the original Bahadur's σ-field. If the minimal sufficient and $\mathcal{E}$-complete σ-field exists, then both σ-fields $\mathcal{S}$ and $\mathcal{U}(\mathcal{E})$ coincide. Theorem 2 was known previously, at least in the case of real-valued estimators and in connection with universal loss functions ([20], Sätze 1 and 2; see also [5]-[7], [12], [14], [18]-[21]). Our proof is valid for estimators with values in a Banach space and differs from the previous ones.

Moreover, we answer the following questions concerning relations between the considered necessary σ-fields $\mathcal{S}$ and $\mathcal{U}(\mathcal{E})$ and the class of estimators which are best unbiased for a strictly convex loss function:

1. Is σ-field $\mathcal{S}$ sufficient whenever $\mathcal{U}(\mathcal{E}) = \mathcal{E}$? The answer is negative (Example 1).

2. Does there exist for each $\mathcal{S}$-measurable estimator $x_1(\cdot) \in \mathcal{E}$ a strictly convex loss function such that $x_1(\cdot)$ is uniformly best unbiased? The answer is negative (Example 2).

3. Suppose that for a given statistical space and for each $\mathcal{S}$-measurable estimator $x(\cdot) \in \mathcal{E}$ there exists a strictly convex weakly differentiable loss function for which $x(\cdot)$ is uniformly best unbiased. Is it true that $\mathcal{U}(\mathcal{E}) = \mathcal{E}$? If $X = \boldsymbol{R}$, the answer is affirmative (Theorem 3).

In Section 3 we consider universal loss functions for a given statistical space (in Linnik's sense) and universal loss functions which are not related to any particular statistical space. We show (Theorem 5) that the class of loss functions universal in the second sense coincides with the known class of universal loss functions considered by Klebanov [5], [6], Schmetterer [18]-[20] and Strasser [20], [21]. Our proof is valid, however, only if values of considered estimators are in $\boldsymbol{R}$, so the problem remains open in the general case of Banach spaces. Moreover, we answer the following problem posed by Klebanov:

4. Suppose that in a given statistical space every strictly convex loss function is universal in Linnik's sense. Does there exist a minimal sufficient and complete σ-field? The answer is negative (Examples 1 and 3).

Finally, note that the Lehmann-Scheffé-Rao lemma proved in [9] provides a basic tool for problems considered in the present paper.

2. Necessary σ-fields $\mathscr{U}(\mathscr{E})$ and $\mathscr{S}$. We denote by $(T, \mathscr{A}, \mathscr{P})$ a statistical space, where T is a set, $\mathscr{A}$ is a σ-field of subsets of T, and $\mathscr{P}$ is a class of probability measures on $\mathscr{A}$. A subset A of T is called *$\mathscr{P}$-null* if there exists an $A_0 \in \mathscr{A}$ such that $A \subset A_0$ and $P(A_0) = 0$ for every $P \in \mathscr{P}$. If $P \in \mathscr{P}$ and $\mathscr{B}$ is a σ-subfield of $\mathscr{A}$, then we denote by $\mathscr{B}_P$ the completion of $\mathscr{B}$ with respect to P. A σ-field $\mathscr{B}$ is called *$\mathscr{P}$-complete* if

$$\mathscr{B} = \bigcap_{P \in \mathscr{P}} \mathscr{B}_P.$$

All σ-fields we shall consider (thus $\mathscr{A}$ itself also) are assumed to be $\mathscr{P}$-complete. We use the notation $\mathrm{E}_P^{\mathscr{B}} x(\cdot)$ both for a version of the conditional expectation of $x(\cdot)$ given a σ-field $\mathscr{B}$ and for the class of all such versions. So, we write $x(\cdot) = \mathrm{E}_P^{\mathscr{B}} x(\cdot)$ or $x(\cdot) \in \mathrm{E}_P^{\mathscr{B}} x(\cdot)$, respectively. $\mathbf{1}_A(\cdot)$ stands for the indicator function of the set A. If $\{t \in T \colon x_1(t) \neq x_2(t)\}$ is a $\mathscr{P}$-null set, we write $x_1(\cdot) = x_2(\cdot)$ $\mathscr{P}$-a.e.

We assume that the set X of decisions for a statistician is a separable Banach space and that Y, the dual space of X, is also separable. Certainly, the most important case is that where X is the space of real numbers $\boldsymbol{R}$. However, the majority of arguments used in the paper remain valid with no change in a more general situation. Moreover, this somewhat more general framework leads to no complication of reasoning and formulas. So, it seems more convenient to present the results of the paper in the general case. The assumptions on the separability of X and Y are imposed in order to avoid difficulties as to measurabilities of functions with values in X and Y. This does not exclude the most important infinite-dimensional spaces useful, e.g., in the estimation of density functions of probability measures on Borel subsets of $\boldsymbol{R}^n$.

Denote by $\mathscr{E}$ a set of considered estimators, i.e., a class of $\mathscr{A}$-measurable functions from T into X which are Bochner P-integrable for every $P \in \mathscr{P}$. We assume that $\mathscr{E}$ fulfils the following conditions:

E1. $\mathscr{E}$ is a vector space.

E2. Each $\mathscr{A}$-measurable function which takes values in a compact subset of X is an element of $\mathscr{E}$.

E3. If $x(\cdot)\in\mathscr{E}$ and $\{t\in T\colon x_1(t)\neq x(t)\}$ is a $\mathscr{P}$-null set, then $x_1(\cdot)\in\mathscr{E}$.

E4. If $x(\cdot)\in\mathscr{E}$ and $\mathscr{B}$ is a sufficient σ-field for $(T,\mathscr{A},\mathscr{P})$, then $\mathrm{E}^{\mathscr{B}}x(\cdot)\in\mathscr{E}$.

E5. If $x(\cdot)\in\mathscr{E}$ and $\varphi(\cdot)$ is a real bounded $\mathscr{A}$-measurable function, then $\varphi(\cdot)x(\cdot)\in\mathscr{E}$.

E6. If $x(\cdot)\in\mathscr{E}$, $x\in X$ and $y\in Y$, then $\big(x(\cdot),y\big)x\in\mathscr{E}$.

If $x(\cdot)\in\mathscr{E}$ and $\mathrm{E}_P x(\cdot)=0$ for every $P\in\mathscr{P}$, then $x(\cdot)$ is called an *unbiased estimator of zero*. The set of all unbiased estimators of zero in $\mathscr{E}$ will be denoted by $\mathscr{E}_0$. Moreover, if $\mathscr{C}$ is a σ-subfield of $\mathscr{A}$, we denote by $M(\mathscr{C})$ the set of all estimators from $\mathscr{E}$ which are $\mathscr{C}$-measurable.

Now, we define two σ-fields which will be investigated in the sequel.

Definition 1. $\mathscr{S}$ is the intersection of all σ-subfields of $\mathscr{A}$ which are $\mathscr{P}$-complete and sufficient for $(T,\mathscr{A},\mathscr{P})$.

Clearly, $\mathscr{S}$ is the greatest necessary σ-field for $(T,\mathscr{A},\mathscr{P})$.

Definition 2. $\mathscr{U}(\mathscr{E})$ is the class of subsets of T given by

$$\mathscr{U}(\mathscr{E})=\Big\{A\in\mathscr{A}\colon \int_A x_0(t)\,P(dt)=0 \text{ for every } x_0(\cdot)\in\mathscr{E}_0 \text{ and each } P\in\mathscr{P}\Big\}.$$

It is easy to see that $\mathscr{U}(\mathscr{E})$ is a σ-field (cf. the proof of Theorem 7 in [17]). The σ-field $\mathscr{U}(\mathscr{E})$ is called *universal* [6]. It was considered first by Bahadur [1] for $\mathscr{E}$ consisting of all square integrable functions and later by Schmetterer [17], [20], Strasser [20], [21], and Klebanov [6]. The definition of $\mathscr{U}(\mathscr{E})$ given above is adopted from [20], [21].

The following two simple propositions characterize the σ-field $\mathscr{U}(\mathscr{E})$ (cf. also [1], [5], [6], [20], [21]).

PROPOSITION 1. *$\mathscr{U}(\mathscr{E})$ is the greatest σ-field in the class of all $\mathscr{P}$-complete σ-fields $\mathscr{C}$ such that*

$$\mathrm{E}_P^{\mathscr{C}}x_0(\cdot)=0 \tag{2.1}$$

holds P-a.e. for every $P\in\mathscr{P}$ and $x_0(\cdot)\in\mathscr{E}_0$.

Proof. We know that $\mathscr{U}(\mathscr{E})$ is a σ-field. $\mathscr{P}$-completeness of $\mathscr{U}(\mathscr{E})$ is obvious and, moreover, the equality $\mathrm{E}_P^{\mathscr{U}(\mathscr{E})}x_0(\cdot)=0$ P-a.e. follows from the definitions of $\mathscr{U}(\mathscr{E})$ and conditional expectation. Assume that $\mathscr{C}$ has property (2.1) and let $A\in\mathscr{C}$. Then

$$\int_A x_0(t)\,P(dt)=0$$

holds for each $x_0(\cdot)\in\mathscr{E}_0$ and $P\in\mathscr{P}$. Hence $A\in\mathscr{U}(\mathscr{E})$ and $\mathscr{C}\subset\mathscr{U}(\mathscr{E})$. This proves the proposition.

Definition 3. A σ-field $\mathscr{C}$ is called *$\mathscr{E}$-complete* if $M(\mathscr{C})\cap\mathscr{E}_0$ consists of estimators equal to 0 $\mathscr{P}$-a.e.

PROPOSITION 2. *$\mathscr{U}(\mathscr{E})$ is $\mathscr{P}$-complete.*

Proof. Let $x_0(\cdot)\in M(\mathscr{U}(\mathscr{E}))\cap\mathscr{E}_0$. Clearly, $x_0(\cdot)\in \mathrm{E}_P^{\mathscr{U}(\mathscr{E})}x_0(\cdot)$ for each $P\in\mathscr{P}$. In view of Proposition 1 we have $0\in \mathrm{E}_P^{\mathscr{U}(\mathscr{E})}x_0(\cdot)$ for every $P\in\mathscr{P}$. Thus, the set $\{t\in T\colon x_0(t)\neq 0\}$ does not depend on $P\in\mathscr{P}$ and is P-null for each $P\in\mathscr{P}$. This proves the proposition.

There exists an important relevance of the σ-fields $\mathscr{S}$ and $\mathscr{U}(\mathscr{E})$ to the theory of unbiased estimation with convex loss functions. This relevance is given in Theorems 1 and 2.

A loss function $L(\cdot,\cdot)$ is assumed to be a function from $X\times\mathscr{P}$ into $[0,\infty)$ such that $L(\cdot, P)$ is convex and lower semicontinuous for every $P\in\mathscr{P}$. Recall that this implies that $L(\cdot, P)$ is continuous on X.

Definition 4. An estimator $x_1(\cdot)\in\mathscr{E}$ is called *uniformly best unbiased* (or, simply, *best unbiased*) for a loss function L if

$$R_L(x_1(\cdot), P)\leqslant R_L(x_1(\cdot)+x_0(\cdot), P) \tag{2.2}$$

holds for every $x_0(\cdot)\in\mathscr{E}_0$ and each $P\in\mathscr{P}$, where

$$R_L(x(\cdot), P) = \int L(x(t), P)\,P(dt). \tag{2.3}$$

THEOREM 1. *If $x_1(\cdot)\in\mathscr{E}$ is best unbiased for a strictly convex loss function L and the risk function $R_L(x_1(\cdot), P)$ is finite for every $P\in\mathscr{P}$, then $x_0(\cdot)$ is $\mathscr{S}$-measurable.*

Proof. Let $\mathscr{B}$ be an arbitrary $\mathscr{P}$-complete and sufficient σ-subfield of $\mathscr{A}$. Let $\mathrm{E}^{\mathscr{B}}x_1(\cdot)$ be a version of $\mathrm{E}_P^{\mathscr{B}}x_1(\cdot)$ which does not depend on P. By Jensen's inequality for conditional expectations (see [10] and [15]) we have

$$L(\mathrm{E}^{\mathscr{B}}x_1(\cdot)(t), P)\leqslant \mathrm{E}_P^{\mathscr{B}}L(x_1(\cdot), P)(t)\quad P\text{-a.e.} \tag{2.4}$$

for each $P\in\mathscr{P}$. However, since $x_1(\cdot)$ is best unbiased, the equality

$$R_L(x_1(\cdot), P) = R_L(\mathrm{E}^{\mathscr{B}}x_1(\cdot), P) \tag{2.5}$$

is valid for every $P\in\mathscr{P}$. Inequality (2.4) together with (2.5) imply

$$\mathrm{E}_P^{\mathscr{B}}L(x_1(\cdot), P)(t) = L(\mathrm{E}^{\mathscr{B}}x_1(\cdot)(t), P)\quad P\text{-a.e.}$$

for every $P\in\mathscr{P}$. The strict convexity of $L(\cdot, P)$ for each $P\in\mathscr{P}$ implies $x_1(t) = \mathrm{E}^{\mathscr{B}}x_1(\cdot)(t)$ P-a.e. for every $P\in\mathscr{P}$ [10] (for $X = \mathbf{R}^n$, see [15]). The set $\{t\in T\colon x_1(t)\neq \mathrm{E}^{\mathscr{B}}x_1(\cdot)(t)\}$ is $\mathscr{A}$-measurable and does not depend on P. Since $\mathscr{B}$ is $\mathscr{P}$-complete, this implies that $x_1(\cdot)$ is $\mathscr{B}$-measurable. Note that this conclusion is valid for an arbitrary $\mathscr{P}$-complete and sufficient σ-field $\mathscr{B}$. Hence $x_1(\cdot)$ is $\mathscr{S}$-measurable.

Remark. It is interesting to compare Theorem 1 with Proposition 3.5 from [8] assuming that the Lehmann-Scheffé-Rao lemma is applicable. Suppose first that $x_1(\cdot)$ is best unbiased for a strictly convex and weakly differentiable L. Then $L'(\cdot, P)$ is star-weakly continuous ([13], p. 80), and hence $(\mathscr{B}_X, \mathscr{B}_Y)$-measurable, where $\mathscr{B}_X$ and $\mathscr{B}_Y$ are σ-fields of Borel subsets of X and Y, respectively. On the other hand, $x_1(\cdot)$ is $\mathscr{S}$-measurable, and so is $L'(x_1(\cdot), P)$. Thus, if L is strictly convex and weakly differentiable, Theorem 1 implies Proposition 3.5 of [8]. Conversely, Proposition 3.5 in [8] asserts that $L'(x_1(\cdot), P)$ is, for each $P \in \mathscr{P}$, P-equivalent to an $\mathscr{S}$-measurable function. By the strict convexity of $L(\cdot, P)$ the transformation $L'(\cdot, P)$ is one-to-one and, therefore, the image of a Borel set in X is a Borel set in Y ([4], p. 131). So, we can conclude that $x_1(\cdot)$ is $\mathscr{S}$-measurable, i.e., Proposition 3.5 of [8] implies Theorem 1. However, if L is neither strictly convex nor differentiable, then for each $P \in \mathscr{P}$ there exists $y_P(\cdot)$ such that

$$y_P(t) \in \partial L(x_1(t), P) \quad \text{and} \quad \int (x_0(t), y_P(t)) P(dt) = 0.$$

The Example in [9], Section 3, shows that, in general, $y_P(\cdot)$ is not a function of $x_1(\cdot)$. None the less, Proposition 3.5 in [8] implies that for each $P \in \mathscr{P}$ the function $y_P(\cdot)$ is P-equivalent to an $\mathscr{S}$-measurable function. Clearly, if $L(\cdot, P)$ is not strictly convex, then $x_1(\cdot)$ may be best unbiased for L, notwithstanding $x_1(\cdot) \notin M(\mathscr{S})$. However, the discussion above suggests that there exists $x_2(\cdot) \in M(\mathscr{S})$ such that $\mathrm{E}_P x_1(\cdot) = \mathrm{E}_P x_2(\cdot)$ and $x_2(\cdot)$ is best unbiased for L. This is the case where $\mathscr{S}$ is a minimal sufficient σ-field, however, we do not know the answer in the general case.

Remark. It is easy to see that an argument analogous to that given in the proof of Theorem 1 shows that the following statement is true:
If $L(\cdot, P)$ is strictly convex, $R_L(x_1(\cdot), P)$ is finite for every $P \in \mathscr{P}$ and $x_1(\cdot)$ is admissible, then $x_1(\cdot)$ is $\mathscr{S}$-measurable.

THEOREM 2. *An estimator $x_1(\cdot) \in \mathscr{E}$ is best unbiased for each convex loss function L if and only if it is $\mathscr{U}(\mathscr{E})$-measurable.*

Proof. Sufficiency. In view of Proposition 1 and Jensen's inequality for conditional expectations [10], we have

$$\mathrm{E}_P^{\mathscr{U}(\mathscr{E})} L(x_1(\cdot) + x_0(\cdot), P)(t) \geqslant L(x_1(t), P) \quad P\text{-a.e.}$$

for every $P \in \mathscr{P}$ and $x_0(\cdot) \in \mathscr{E}_0$. Now, integration of both sides of the inequality shows that the estimator $x_1(\cdot)$ is best unbiased.

Necessity. Inequality (2.2) holds for every loss function L, thus also for loss functions $L_{y,c}$ which are of the form

$$L_{y,c}(x, P) = \max\{0, (x - g(P), y) - c\},$$

where $y \in Y$, $c \in \boldsymbol{R}$, $g(\cdot)$ is a function from $\mathscr{P}$ into X and $(x, y) = y(x)$. For simplicity we can take $g(P) \equiv 0$. Thus we have

$$L_{y,c}(x) = \max\{0, (x, y) - c\}. \tag{2.6}$$

For a given $y \in Y$, $c \in \boldsymbol{R}$ and $x_0(\cdot) \in \mathscr{E}_0$ we put

$$A = A(\beta) = A(y, c, \beta) = \{t \in T\colon (x_1(t) + \beta x_0(t), y) > c\}, \quad \beta \in \boldsymbol{R},$$

and

$$B = B(c) = B(c, y) = \{t \in T\colon (x_1(t), y) > c\}.$$

Then, by (2.3), we have

$$R_{L_{y,c}}(x_1+\beta x_0, P) - R_{L_{y,c}}(x_1, P) = \int_{A\cap B^-} (x_1(t), y)\,P(dt) -$$

$$- \int_{B\cap A^-} (x_1(t), y)\,P(dt) + \beta \int_A (x_0(t), y)\,P(dt) - cP(A) + cP(B)$$

$$\leqslant \beta \int_A (x_0(t), y)\,P(dt),$$

where B^- and A^- denote the complements of B and A, respectively.

Moreover, we have

$$\lim_{\beta\to 0^+} \mathbf{1}_{A(\beta)}(t) = \mathbf{1}_B(t) + \mathbf{1}_{A_1}(t) \quad \text{and} \quad \lim_{\beta\to 0^-} \mathbf{1}_{A(\beta)}(t) = \mathbf{1}_B(t) + \mathbf{1}_{A_2}(t),$$

where $A_1 = \{t \in T\colon (x_0(t), y) > 0,\ (x_1(t), y) = c\}$ and $A_2 = \{t \in T\colon (x_0(t), y) < 0,\ (x_1(t), y) = c\}$.

Given $P \in \mathscr{P}$ we assume that c satisfies the following condition:

$$P(\{t \in T\colon (x_1(t), y) = c\}) = 0. \tag{2.7}$$

Then

$$\lim_{\beta\to 0} \mathbf{1}_{A(\beta)}(t) = \mathbf{1}_B(t)$$

holds P-a.e. Therefore, if $\int_B (x_0(t), y)\,dP \neq 0$, then the integrals

$$\int_{A(\beta)} (x_0(t), y)\,dP \quad \text{and} \quad \int_B (x_0(t), y)\,dP$$

have the same sign provided $|\beta|$ is sufficiently small. Thus, we can choose β such that

$$\beta \int_{A(\beta)} (x_0(t), y)\,dP < 0$$

holds. This, however, contradicts inequality (2.2). Therefore

$$\Big(\int_{B(c)} x_0(t)\,dP, y\Big) = \int_{B(c)} (x_0(t), y)\,dP = 0 \tag{2.8}$$

holds for each $c \in \boldsymbol{R}$ such that condition (2.7) is satisfied. Denote by $C(P)$ the class of all numbers c satisfying (2.7). The function $(x_1(\cdot), y)$ is P-integrable, thus the set $\boldsymbol{R}\backslash C(P)$ is at most countable. Therefore, if $c' \notin C(P)$, we can choose numbers c_n $(n = 1, 2, \ldots)$ such that $c_n \to c'$, $c_n > c'$ and $c_n \in C(P)$. Since (2.8) holds for each c_n, it follows from the dominated convergence theorem that (2.8) is valid also for c'. Hence equality (2.8) holds for every

$c \in \boldsymbol{R}$, $P \in \mathscr{P}$, $y \in Y$ and $x_0(\cdot) \in \mathscr{E}_0$. A standard argument shows that the class of all sets D for which

$$\Big(\int_D x_0(t)\, dP, y\Big) = 0$$

for every $y \in Y$, $P \in \mathscr{P}$ and $x_0(\cdot) \in \mathscr{E}_0$ forms a σ-field and contains sets $B(c)$. Since X is a separable Banach space, the Borel σ-fields spanned over norm-open and over weakly-open subsets of X coincide. Hence

$$\int_D x_0(t)\, P(dt) = 0$$

for each $D \in x_1^{-1}(\mathscr{B}_X)$, $x_0(\cdot) \in \mathscr{E}_0$ and $P \in \mathscr{P}$. Let $\mathscr{C} = x_1^{-1}(\mathscr{B}_X)$. The last equality implies $E_P^{\mathscr{C}} x_0(\cdot) = 0$ P-a.e. for each $P \in \mathscr{P}$ and $x_0(\cdot) \in \mathscr{E}_0$. In view of Proposition 1 we obtain $\mathscr{C} \subset \mathscr{U}(\mathscr{E})$, i.e., $x_1(\cdot)$ is $\mathscr{U}(\mathscr{E})$-measurable.

COROLLARY 1. *The σ-field $\mathscr{U}(\mathscr{E})$ is a σ-subfield of $\mathscr{S}$.*

Proof. By Theorem 2 every estimator $x(\cdot)$ which is $\mathscr{U}(\mathscr{E})$-measurable is best unbiased for each convex loss function, hence also for a strictly convex one. Thus, if a strictly convex function on the Banach space X exists, then Theorem 1 implies that $x(\cdot)$ is $\mathscr{S}$-measurable. Thus the inclusion $\mathscr{U}(\mathscr{E}) \subset \mathscr{S}$ holds. To complete the proof it is enough to note that on each separable Banach space there exists a strictly convex function. This is an immediate consequence of the existence of a strictly convex norm $|||\cdot|||$ on X which is equivalent to the original one (see [3], Corollary 3.1, p. 179). For instance, the function $L(x) = |||x||| - \ln(1 + |||x|||)$ is strictly convex and dominated by $c\|x\|$, where c is a constant. Hence the risk R_L is finite on $\mathscr{E}$. This completes the proof of the corollary.

Denote by E_{opt} the class of estimators in $\mathscr{E}$ which are best unbiased for a strictly convex loss function, i.e.

$$E_{\text{opt}} = \{x(\cdot) \in \mathscr{E}\colon \text{there exists a strictly convex loss function } L \text{ such that } x(\cdot) \text{ is best unbiased for } L\}.$$

In the sequel we shall be concerned with relations between E_{opt} and the σ-fields $\mathscr{U}(\mathscr{E})$ and $\mathscr{S}$. Note that if $\mathscr{S} = \mathscr{U}(\mathscr{E})$, then E_{opt} consists of all $\mathscr{S}$-measurable estimators. The equality $M(\mathscr{S}) = E_{\text{opt}}$ holds, e.g., if $\mathscr{S}$ is $\mathscr{E}$-complete and minimal sufficient. However, the converse implication does not hold. In Example 1 below we construct a statistical space $(T, \mathscr{A}, \mathscr{P})$, where $\mathscr{S}$ is not sufficient even though $\mathscr{U}(\mathscr{E}) = \mathscr{S}$. This is a slight modification of a well-known example of Pitcher [16].

Example 1. Let $T = (-1, 0) \cup (0, 1)$. Assume that $\mathscr{A}$ is the σ-field of Borel subsets of T. A set $\mathscr{P}$ consisting of probability measures on $\mathscr{A}$ is given by

$$\mathscr{P} = \{P_t\colon t \in (0, 1)\} \cup \{P_f\colon f \in L_1^+(0, 1)\},$$

where P_t is a probability measure on $\mathscr{A}$ such that $P_t(\{t\}) = P_t(\{-t\}) = 1/2$ and P_f is a probability measure on $\mathscr{A}$ with density function f (with respect to the Lebesgue measure restricted to $\mathscr{A}$) such that $f(t) = 0$ for $t \in (-1, 0)$. The unique $\mathscr{P}$-null subset of T is the empty set. Hence $\mathscr{A}$ is $\mathscr{P}$-complete. It is known that σ-fields $\mathscr{A}_s$ are sufficient for $\mathscr{P}$, where $s \in (0, 1)$ and

$$\mathscr{A}_s = \{A \in \mathscr{A}\colon A \supset \{-s, s\} \text{ or } A \cap \{-s, s\} = \emptyset\}$$

(see [16]). Let

$$\mathscr{A}_0 = \bigcap_{s \in (0,1)} \mathscr{A}_s.$$

It is easy to see that $\mathscr{A}_0$ consists of all Borel subsets A such that $A = -A$ and that $\mathscr{A}_0$ is not sufficient [16]. Let $\mathscr{E}$ be the set of all bounded $\mathscr{A}$-measurable estimators from T into $\boldsymbol{R}$. The set $\mathscr{E}_0$ of unbiased estimators of zero consists of functions satisfying the conditions $x_0(t) = -x_0(-t)$, $t \in (0, 1)$, and $x(t) = 0$ a.e. with respect to the Lebesgue measure. If $A \in \mathscr{A}_0$, then

$$\int_A x_0(t) P(dt) = 0$$

for every $x_0(\cdot) \in \mathscr{E}_0$ and $P \in \mathscr{P}$. Thus $\mathrm{E}_P^{\mathscr{A}_0} x_0(\cdot) = 0$ holds for each $P \in \mathscr{P}$ and $x_0(\cdot) \in \mathscr{E}_0$. By Proposition 1 we obtain $\mathscr{U}(\mathscr{E}) \supset \mathscr{A}_0$. On the other hand, $\mathscr{A}_0 \supset \mathscr{S}$ and Corollary 1 implies $\mathscr{U}(\mathscr{E}) = \mathscr{A}_0 = \mathscr{S}$.

Example 2, which we are going to construct, shows that, in general, sets $M(\mathscr{S}) \backslash E_{\mathrm{opt}}$ and $E_{\mathrm{opt}} \cap \big(M(\mathscr{S}) \backslash M(\mathscr{U}(\mathscr{E}))\big)$ are not empty, i.e. there may exist both $\mathscr{S}$-measurable estimators which are not elements of E_{opt} and $\mathscr{S}$-measurable estimators which are not $\mathscr{U}(\mathscr{E})$-measurable but belong to E_{opt}.

Example 2. Let $X = \boldsymbol{R}$, $T = \{1, 2, 3\}$, $\mathscr{A} = 2^T$ and $\mathscr{P} = \{P_1, P_2\}$, where

$$P_1 = (P_1(1), P_1(2), P_1(3)) = \left(\frac{1}{2}, \frac{1}{6}, \frac{1}{3}\right)$$

and

$$P_2 = (P_2(1), P_2(2), P_2(3)) = \left(\frac{1}{2}, \frac{2}{5}, \frac{1}{10}\right).$$

Assume that $\mathscr{E}$ consists of all $\mathscr{A}$-measurable functions. Thus $\mathscr{E}$ can be identified with $\boldsymbol{R}^3$. It is easy to see that $\mathscr{E}_0$ consists of estimators of the form λx_0, where $\lambda \in \boldsymbol{R}$ and $x_0 = (x_0(1), x_0(2), x_0(3)) = (-1, 1, 1)$. Therefore, $\mathscr{E}_0$ corresponds to the one-dimensional subspace of $\boldsymbol{R}^3$ spanned on the vector $(-1, 1, 1)$. By the factorization theorem, $\mathscr{A}$ is a minimal sufficient σ-field. Therefore, $\mathscr{A} = \mathscr{S}$ and each estimator in $\mathscr{E}$ is $\mathscr{S}$-measurable. If $x_0(\cdot) \in \mathscr{E}_0$, $x_0(\cdot) \neq 0$ and $A \in \mathscr{U}(\mathscr{E})$, then $\mathbf{1}_A(\cdot) x_0(\cdot) \in \mathscr{E}_0$. So $\mathbf{1}_A(\cdot) x_0 = \lambda x_0(\cdot)$ holds and, since $x_0(t) \neq 0$ for every $t \in T$, we obtain $\mathbf{1}_A(\cdot) \equiv \lambda$. Hence we get

$\mathscr{U}(\mathscr{E}) = \{\emptyset, T\}$. Therefore, the set of $\mathscr{U}(\mathscr{E})$-measurable functions can be identified with the one-dimensional subspace of $\boldsymbol{R}^3$ spanned on the vector $(1, 1, 1)$. Let $x_1(\cdot) \in \mathscr{E}$ and let $L(x, P)$ be a strictly convex loss function. In view of the Lehmann-Scheffé-Rao lemma [9] the estimator $x_1(\cdot)$ is best unbiased for L if and only if

$$\int x_0(t)\, y_P(t)\, P(dt) = 0 \tag{2.9}$$

holds for $P \in \{P_1, P_2\}$, where $y_P(t)$ is an element of $\partial L(x_1(t), P)$. Define vectors v_1 and v_2 as follows:

$$v_1 = \big(y_{P_1}(1), y_{P_1}(2), y_{P_1}(3)\big)$$

and

$$v_2 = \big(y_{P_2}(1), y_{P_2}(2), y_{P_2}(3)\big).$$

Equality (2.9) can be interpreted in the following way: the vectors v_1 and v_2 are perpendicular in $\boldsymbol{R}^3$ to vectors

$$u_1 = \big(P_1(1)x_0(1), P_1(2)x_0(2), P_1(3)x_0(3)\big) = \left(-\frac{1}{2}, \frac{1}{6}, \frac{1}{3}\right)$$

and

$$u_2 = \big(P_2(1)x_0(1), P_2(2)x_0(2), P_2(3)x_0(3)\big) = \left(-\frac{1}{2}, \frac{2}{5}, \frac{1}{10}\right),$$

respectively. It is easy to check that

$$u_1^{\perp} = \operatorname{Lin}\{(1, 1, 1), (-1, 5, -4)\} \quad \text{and} \quad u_2^{\perp} = \operatorname{Lin}\{(1, 1, 1), (1, 2, -3)\},$$

where $\operatorname{Lin}\{\ldots\}$ stands for the linear space spanned by vectors indicated in the brackets. Since $L(\cdot, P)$ is strictly convex, $y_P(t_1) < y_P(t_2)$ holds whenever $x_1(t_1) < x_1(t_2)$. Moreover, since $v_1 \in u_1^{\perp}$ and $v_2 \in u_2^{\perp}$, the ranks of the components of v_1, v_2 and x_1 are the same and equal to $(2, 3, 1)$ or $(2, 1, 3)$. Therefore, if $x_1(\cdot)$ is an estimator such that $\big(x_1(1), x_1(2), x_1(3)\big)$ has the rank vector different both from $(2, 3, 1)$ and $(2, 1, 3)$ (e.g., $x(t) = t$), then there is no strictly convex loss function $L(\cdot, P)$ such that (2.9) is satisfied. Consequently, there is no strictly convex loss function such that $x_1(\cdot)$ is best unbiased for L, i.e. $x_1(\cdot)$ is not an element of E_{opt}.

On the other hand, an estimator $x_2(\cdot)$ given by $x_2(1) = 2$, $x_2(2) = 3$, $x_2(3) = 1$ is best unbiased for the loss function $L(x, P)$ given by

$$L(x, P) = W\big(x - \mathrm{E}_p x_2(\cdot)\big), \tag{2.10}$$

where

$$W(x) = \int_0^x w(t)\, dt,$$

$$w(t) = \begin{cases} t & \text{if } t \in [0, 1], \\ 7t-6 & \text{if } t \in [1, 7/6], \\ 16t-33/2 & \text{if } t \in [7/6, \infty), \end{cases} \tag{2.11}$$

$$w(-t) = w(t) \quad \text{if } t \in [0, \infty).$$

To this end it is enough to verify that $x_2(\cdot)$ satisfies condition (2.9) provided the loss function L is given by (2.10)-(2.11). We omit the easy calculations.

The situation illustrated by Example 2 seems to be typical for the general case. The existence of $\mathcal{S}$-measurable estimators which are not elements of E_{opt} can be interpreted in the following way: the class of strictly convex loss functions is too small in order to label each $\mathcal{S}$-measurable estimator best unbiased for a strictly convex loss.

Now, we shall be concerned with the problem whether $\mathcal{U}(\mathcal{E}) = \mathcal{S}$ whenever each $\mathcal{S}$-measurable estimator is best unbiased for a strictly convex loss function. If the loss functions are weakly differentiable, then Theorem 3 gives a partial answer for X being a Banach space and the complete one for $X = \boldsymbol{R}$.

THEOREM 3. *Assume that for each set $A \in \mathcal{S}$ there exists a strictly convex and weakly differentiable loss function L such that*

(a) $L(x, P) \geqslant c_P \|x\|$ *whenever* $\|x\| \geqslant r_P$, $c_P > 0$, $r_P > 0$;

(b) *for each $x \in X$ the estimator $x \cdot \mathbf{1}_A(\cdot)$ is best unbiased for L;*

(c) *for each $x_0(\cdot) \in \mathcal{E}_0$ and $P \in \mathcal{P}$ the risk $R_L(\lambda x_0(\cdot), P)$ is finite for some positive $\lambda = \lambda(x_0(\cdot), P)$.*

Then $\mathcal{U}(\mathcal{E}) = \mathcal{S}$.

Proof. Given $x \in X$ and $A \in \mathcal{S}$ we put $x_1(\cdot) = x \cdot \mathbf{1}_A(\cdot)$. By our assumption $x_1(\cdot)$ is best unbiased for L. It is easy to see that the assumptions of the Lehmann Scheffé-Rao lemma (part (b)) [9] are satisfied. So

$$\int \big(x_0(t), L'(x_1(t), P)\big) P(dt) = 0 \tag{2.12}$$

holds for every $x_0(\cdot) \in \mathcal{E}_0$ and $P \in \mathcal{P}$. For each $x_0(\cdot) \in \mathcal{E}_0$ we have

$$\int_A x_0(t) P(dt) = -\int_{A^-} x_0(t) P(dt).$$

Thus from (2.12) we obtain, for each $x_0(\cdot) \in \mathcal{E}_0$, $x \in X$ and $P \in \mathcal{P}$,

$$\Big(\int_A x_0(t) P(dt), L'(x, P) - L'(0, P)\Big) = 0. \tag{2.13}$$

Assumption (a) implies that $L^*(\cdot, P)$, the conjugate of $L(\cdot, P)$, is finite on $K_Y(0, c_P)$, the ball in Y centred at zero and of radius c_P. The range of $L'(\cdot, P)$ contains $\operatorname{int}\operatorname{dom} L^*(\cdot, P)$, and hence it contains $\operatorname{int} K_Y(0, c_P)$.

So (2.13) implies that

$$\int_A x_0(t)\,P(dt) = 0$$

for every $x_0(\cdot)\in\mathscr{E}_0$ and $P\in\mathscr{P}$. By the definition of $\mathscr{U}(\mathscr{E})$ we get $A\in\mathscr{U}(\mathscr{E})$. This proves the theorem.

Remark. The proof of Theorem 3 is adopted from [17], Theorem 7 (cf. also [6] and [21]).

COROLLARY 2. *Let $X = \boldsymbol{R}$ and assume that for each $\mathscr{S}$-measurable estimator $x_1(\cdot)$ there exists a strictly convex and differentiable loss function L such that $x_1(\cdot)$ is best unbiased for L and $R_L(\lambda x_0(\cdot), P)$ is finite for every $x_0(\cdot)\in\mathscr{E}_0$, $P\in\mathscr{P}$ and for some positive λ, $\lambda = \lambda(x_0(\cdot), P)$. Then $\mathscr{U}(\mathscr{E}) = \mathscr{S}$.*

Proof. Let us consider estimators of the form $x_1(\cdot) = \mathbf{1}_A(\cdot)$, where $A\in\mathscr{S}$. Repeating the argument given in the proof of Theorem 3, we obtain equality (2.13) which is now of the form

$$\big(L'(1,P)-L'(0,P)\big)\int_A x_0(t)\,P(dt) = 0.$$

Since $L(\cdot, P)$ is strictly convex, $L'(1,P) > L'(0,P)$ and we conclude that

$$\int_A x_0(t)\,P(dt) = 0$$

for each $x_0(\cdot)\in\mathscr{E}_0$ and $P\in\mathscr{P}$. Hence $A\in\mathscr{U}(\mathscr{E})$.

3. Universal loss functions. By Theorem 1 we know that if $x(\cdot)\in\mathscr{E}$ is best unbiased for a strictly convex loss function $L(x, P)$, then $x(\cdot)$ is $\mathscr{S}$-measurable. It is also important to know conditions which guarantee a best unbiased estimator to be $\mathscr{U}(\mathscr{E})$-measurable. Then, by Theorem 2, the $\mathscr{U}(\mathscr{E})$-measurable estimator is best unbiased for each convex loss function. The first sufficient conditions for $\mathscr{U}(\mathscr{E})$-measurability of estimators were given by Bahadur [1]. Other conditions were given in [17], Theorem 7, and – after 1970 – in [5], [6], [12], [18]-[21]. In all theorems of this type it has been assumed that an estimator $x_1(\cdot)$ is best unbiased for a strictly convex differentiable loss function of a specific form. Assuming, moreover, various additional conditions on $x_1(\cdot)$, on the loss function and on the class $\mathscr{E}$ of considered estimators to be satisfied it has been proved that $x_1(\cdot)$ is $\mathscr{U}(\mathscr{E})$-measurable. It is convenient for us to recall in Theorem 4 a typical result of this kind. In the proof of Theorem 4 we use the arguments of Schmetterer and Strasser [20], Satz 2. However, the present proof is given for the general case where X is a Banach space whereas the earlier ones were formulated for $X = \boldsymbol{R}$.

Let Θ be a set of parameters and let $\{K_0(\cdot,\theta)\colon \theta\in\Theta\}$ be a class of convex functions of the form

$$K_0(x,\theta) = C_1(\theta)\,W(x)+(x, y_\theta)+C_2(\theta), \tag{3.1}$$

where $C_1(\theta) > 0$, $C_2(\theta) \in \boldsymbol{R}$, $y_\theta \in Y$ and $W(\cdot)$ is a strictly convex finite non-negative and weakly differentiable function on X. Every function π from Θ_1, $\Theta_1 \subset \Theta$, onto a set $\mathscr{P}$ of probability measures is called a *parametrization* of $\mathscr{P}$.

THEOREM 4. *Let $(T, \mathscr{A}, \mathscr{P})$ be a statistical space and let $\mathscr{E}$ be a set of estimators. Assume that $x_1(\cdot) \in \mathscr{E}$ takes values in a compact subset of X and that the condition*

$$R_L(x_1(\cdot)+\lambda x_0(\cdot), P) = \int L(x_1(t)+\lambda x_0(t), t) P(dt) < \infty \tag{3.2}$$

is satisfied for every $x_0(\cdot) \in \mathscr{E}_0$, $P \in \mathscr{P}$ and $\lambda \in (-\varepsilon, \varepsilon)$. Assume further that the loss function $L(x, P)$ admits a representation

$$L(x, \pi(\theta)) = K_0(x, \theta), \quad \theta \in \Theta_1, \tag{3.3}$$

where $\Theta_1 \subset \Theta$, π is a parametrization of $\mathscr{P}$ and K_0 is of the form (3.1). *If $x_1(\cdot)$ is best unbiased for $L(x, P)$, then $x_1(\cdot)$ is $\mathscr{U}(\mathscr{E})$-measurable.*

Proof. First, we note that if for a given parametrization π the loss function L is of the form (3.3) and K_0 is of the form (3.1), then the partial orderings in $\{x_1(\cdot)+\mathscr{E}_0\}$ induced by L and W are the same. Therefore, $x_1(\cdot)$ is best unbiased for W.

By the Lehmann-Scheffé-Rao lemma, the equality

$$\int \big(x_0(t), W'(x_1(t))\big) P(dt) = 0$$

holds for every $P \in \mathscr{P}$ and $x_0(\cdot) \in \mathscr{E}_0$. In particular, since $\big(x_0(\cdot), y\big) x \in \mathscr{E}_0$, we have

$$\int \big(x, W'(x_1(t))\big)\big(x_0(t), y\big) P(dt) = 0$$

for every $P \in \mathscr{P}$, $x \in X$ and $y \in Y$. Note that $x_1(\cdot)$ takes values in a compact set and W' is a continuous function from X into Y endowed with admissible topologies ([13], p. 79). Therefore, $\big(x, W'(x_1(\cdot))\big)$ is a bounded function and $\big(x, W'(x_1(\cdot))\big)\big(x_0(\cdot), y\big) x \in \mathscr{E}_0$. So, by the Lehmann-Scheffé-Rao lemma, we obtain

$$\int \big(x, W'(x_1(t))\big)^n \big(x_0(t), y\big) P(dt) = 0, \quad n = 1, 2, 3, \ldots,$$

for every $x \in X$, $y \in Y$, $x_0(\cdot) \in \mathscr{E}_0$ and $P \in \mathscr{P}$. Similarly as in [20] we obtain

$$\int_{x_1(\cdot)\in B} x_0(t) P(dt) = 0$$

for $B \in \mathscr{B}_X$ provided $W'(B)$ is a Borel subset of Y for every $B \in \mathscr{B}_X$. Since $W'(\cdot)$ is continuous, it is Borel measurable. Moreover, by the strict convexity of W the mapping $W'\colon X \to Y$ is one-to-one. Thus, by [4], Theorem III, 7.2, we have $W'(B) \in \mathscr{B}_Y$ for each $B \in \mathscr{B}_X$. This completes the proof of Theorem 4.

Given a set of parameters Θ, denote by $\{K(\cdot,\theta)\colon \theta\in\Theta\}$ a class of convex finite lower semicontinuous and non-negative functions on X. If π is a parametrization from Θ_1, $\Theta_1\subset\Theta$, into a class of probability measures $\mathscr{P}$, then it is possible to create convex loss functions of the form

$$L(x,\pi(\theta)) = K(x,\theta), \qquad \theta\in\Theta_1.$$

Thus, it is convenient to call the function $K\colon X\times\Theta\to[0,\infty)$, described above, a *parametric loss function* or, simply, a *loss function* if there is no danger of confusion.

Definition 5. A convex parametric loss function K is called *universal* if Theorem 4 remains true whenever K is used in (3.3) instead of K_0.

Clearly, Definition 5 implies that each K_0 of the form (3.1) is universal. The first known universal parametric loss function was the most popular quadratic one $K(x,\theta)=(x-\theta)^2$, where $X=\boldsymbol{R}$ and $\theta\in\boldsymbol{R}$ (see [1]). Note that Klebanov [6], Theorem 7, proved that if $X=\boldsymbol{R}$, then a natural class of convex but not strictly convex parametric loss functions is not universal.

There was a hope that every strictly convex weakly differentiable parametric loss function is universal or, at least, that the class of universal loss functions is large (cf. [18], Section 5, and [21], Remark 4.9). Recently it has appeared that very natural classes of strictly convex differentiable parametric loss functions are not universal [2]. Now, we shall prove that, at least in the case $X=\boldsymbol{R}$, every universal loss function is of the form (3.1).

First, however, we note that a strictly convex loss function $K(x,\theta)$ is of the form (3.1) if and only if for each $\theta_1,\theta_2\in\Theta$

$$K(x,\theta_2) = C'(\theta_1,\theta_2)K(x,\theta_1)+(x,y_{\theta_1,\theta_2})+C''(\theta_1,\theta_2), \tag{3.4}$$

where $C'(\theta_1,\theta_2)>0$, $C''(\theta_1,\theta_2)\in\boldsymbol{R}$ and $y_{\theta_1,\theta_2}\in Y$. For $X=\boldsymbol{R}$, (3.4) can be rewritten in the form

$$K(x,\theta_2) = \alpha_1 K(x,\theta_1)+\alpha_2 x+\alpha_3, \tag{3.5}$$

where

$$\alpha_1=\alpha_1(\theta_1,\theta_2)>0, \qquad \alpha_2=\alpha_2(\theta_1,\theta_2)\in\boldsymbol{R} \quad \text{and} \quad \alpha_3=\alpha_3(\theta_1,\theta_2)\in\boldsymbol{R}.$$

THEOREM 5. *Let $X=\boldsymbol{R}$ and let K_1 and K_2 be two strictly convex functions on X which do not admit representation* (3.5). *Then there exists a statistical space $(T,\mathscr{A},\{P_1,P_2\})$ such that for each class $\mathscr{E}$ of estimators there exists a bounded estimator $x_1(\cdot)$ which is not $\mathscr{U}(\mathscr{E})$-measurable and is best unbiased for the loss function L given by*

$$L(x,P)=\begin{cases} K_1(x) & \text{if } P=P_1,\\ K_2(x) & \text{if } P=P_2.\end{cases}$$

Proof. Let us put $T=\{1,2,3\}$ and $\mathscr{A}=2^T$. Then every class $\mathscr{E}$ of estimators satisfying condition E2 coincides with the class of all functions from T into $\boldsymbol{R}$. Let $v_1(\cdot)$ and $v_2(\cdot)$ be arbitrary selectors of multifunctions $x \to \partial K_1(x)$ and $x \to \partial K_2(x)$, respectively. Since K_1 and K_2 do not satisfy condition (3.5), functions $v_1(\cdot)$, $v_2(\cdot)$ and $1(\cdot)$ ($1(x) = 1$ for each $x \in \boldsymbol{R}$) are linearly independent and there exist points x_1, x_2, x_3, $x_1 < x_2 < x_3$, such that vectors

$$v_1 = \big(v_1(x_1), v_1(x_2), v_1(x_3)\big),$$

$$v_2 = \big(v_2(x_1), v_2(x_2), v_2(x_3)\big) \quad \text{and} \quad v_3 = (1, 1, 1)$$

are linearly independent. Indeed, otherwise, for every three different points x_1, x_2, x_3 we would have

$$\big(v_2(x_1), v_2(x_2), v_2(x_3)\big) = \gamma_1\big(v_1(x_1), v_1(x_2), v_1(x_3)\big) + \gamma_2(1, 1, 1),$$

where γ_1 and γ_2 are suitably chosen coefficients. However, $v_1(\cdot)$ is increasing and, therefore, vectors $\big(v_1(x_1), v_1(x_2)\big)$ and $(1, 1)$ are independent. So, if x is different from x_1 and x_2 and

$$\big(v_2(x_1), v_2(x_2), v_2(x)\big) = \gamma_3\big(v_1(x_1), v_1(x_2), v_1(x)\big) + \gamma_4(1, 1, 1),$$

then equalities $\gamma_1 = \gamma_3$ and $\gamma_2 = \gamma_4$ must hold. Hence $v_2(x) = \gamma_1 v_1(x) + \gamma_2$ holds for every $x \in \boldsymbol{R}$. Since any selector of the subdifferential mapping may be used in the integral representation of a convex function, we conclude that K_1 and K_2 fulfil condition (3.5). This, however, contradicts our assumption.

Define an estimator $x_1(\cdot)$ by

$$x_1(1) = x_1, \quad x_1(2) = x_2, \quad x_1(3) = x_3.$$

Functions $v_1(\cdot)$ and $v_2(\cdot)$ are increasing, so we have

$$v_i\big(x_1(1)\big) < v_i\big(x_1(2)\big) < v_i\big(x_1(3)\big), \quad i = 1, 2.$$

Now, let $u_1 = (u_{11}, u_{12}, u_{13})$ and $u_2 = (u_{21}, u_{22}, u_{23})$ be vectors in $\boldsymbol{R}^3$ defined by

$$u_1 = v_1 \wedge v_3 \quad \text{and} \quad u_2 = v_2 \wedge v_3, \tag{3.6}$$

where $\wedge$ denotes the outer product in $\boldsymbol{R}^3$. Clearly, u_1 and u_2 are linearly independent and, moreover, the components of u_1 and u_2 satisfy the following inequalities:

$$\begin{aligned} u_{11} &< 0, \quad u_{12} > 0, \quad u_{13} < 0, \\ u_{21} &< 0, \quad u_{22} > 0, \quad u_{23} < 0. \end{aligned} \tag{3.7}$$

Let P_1 and P_2 be two probability measures on $\mathscr{A}$ given by

$$P_j(t) = |u_{jt}| / \sum_{i=1}^{3} |u_{ji}|, \quad t \in T,\ j = 1, 2. \tag{3.8}$$

Let p_1 and p_2 be vectors in $\boldsymbol{R}^3$ given by

$$p_1 = \big(P_1(1), P_1(2), P_1(3)\big) \quad \text{and} \quad p_2 = \big(P_2(1), P_2(2), P_2(3)\big).$$

Since u_1 and u_2 are linearly independent, so are p_1 and p_2. Thus P_1 and P_2 are two different probability measures on $\mathscr{A}$. Consequently, the minimal sufficient σ-field exists, equals $\mathscr{S}$ and is different from the trivial σ-field $\{\emptyset, T\}$. Clearly, the estimators from $\mathscr{E}$ may be identified with vectors in $\boldsymbol{R}^3$. Since p_1 and p_2 are independent, the space $\mathscr{E}_0$ consisting of all unbiased estimators of zero corresponds to a one-dimensional subspace of $\boldsymbol{R}^3$. Note that if $x_0(1) = -1, x_0(2) = 1$ and $x_0(3) = -1$, then the equality $\mathrm{E}_{P_i} x_0(\cdot) = 0$, $i = 1, 2$, follows easily from (3.6)-(3.8). Hence $\mathscr{E}_0 = \{\lambda x_0(\cdot): \lambda \in \boldsymbol{R}\}$.

Let $A \in \mathscr{U}(\mathscr{E})$. By the definition of $\mathscr{U}(\mathscr{E})$ the estimator $1_A(\cdot)x_0(\cdot)$ is also an element of $\mathscr{E}_0$. However, this is possible only if $A = T$ or $A = \emptyset$. Thus $\mathscr{U}(\mathscr{E}) = \{\emptyset, T\}$. It is clear that the estimator $x_1(\cdot)$ defined above is not $\mathscr{U}(\mathscr{E})$-measurable.

We prove that $x_1(\cdot)$ is best unbiased for L. In view of the Lehmann-Scheffé-Rao lemma it is enough to show that the condition

$$\int v_i\big(x_1(t)\big)x_0(t)P_i(dt) = 0 \tag{3.9}$$

is satisfied for $i = 1, 2$.

For every $i = 1, 2$ the vector $\big(x_0(1)P_i(1), x_0(2)P_i(2), x_0(3)P_i(3)\big)$ is parallel to u_i. Therefore, $x_1(\cdot)$ fulfils condition (3.9). Thus the estimator $x_1(\cdot)$ is best unbiased for L, however, it is not $\mathscr{U}(\mathscr{E})$-measurable. This completes the proof of Theorem 5.

Apart from the universal parametric loss functions we shall also consider universal loss functions for a statistical space $(T, \mathscr{A}, \mathscr{P})$ introduced in [5] and [6] and named by Linnik.

Definition 6. Let $(T, \mathscr{A}, \mathscr{P})$ be a statistical space, $X = \boldsymbol{R}^n$ and let $\mathscr{E}$ be the class of all estimators which are square P-integrable for each $p \in \mathscr{P}$. A convex loss function $L: X \times \mathscr{P} \to [0, \infty)$ is called *universal* for $(T, \mathscr{A}, \mathscr{P})$ (or *universal in Linnik's sense*) if each bounded estimator which is best unbiased for L is $\mathscr{U}(\mathscr{E})$-measurable.

In the rest of this section we shall assume that $(T, \mathscr{A}, \mathscr{P})$, X and $\mathscr{E}$ satisfy conditions formulated in Definition 6. In the classical case, where the minimal sufficient and $\mathscr{E}$-complete σ-field exists, we have $\mathscr{S} = \mathscr{U}(\mathscr{E})$ and $\mathscr{S}$ is sufficient. Then, in view of Theorem 1, each strictly convex loss function is universal in Linnik's sense. Clearly, every strictly convex loss function is universal in Linnik's sense if $\mathscr{S} = \mathscr{U}(\mathscr{E})$ and the existence of the minimal sufficient σ-field is not necessary for this (cf. Example 1, Section 2). However, the converse implication is not valid. Namely, in Example 3 we shall indicate a statistical space where each strictly convex loss function is universal in Linnik's sense and, simultaneously, $\mathscr{U}(\mathscr{E})$ is a proper σ-subfield of $\mathscr{S}$.

This is a slight modification of an example considered by Lehmann and Scheffé ([11], Example 3.5).

Example 3. Let $X = \boldsymbol{R}$ and let $\mathscr{A}$ be the σ-field of Lebesgue measurable subsets of $T = \boldsymbol{R}^n$. Let $\mathscr{P} = \{P_\theta \colon \theta \in \boldsymbol{R}\}$, where P_θ is the product of n identical rectangular distributions concentrated on the interval $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$. Thus, if λ_n is the Lebesgue measure on $\boldsymbol{R}^n$, then $dP_\theta/d\lambda_n(x_1, \ldots, x_n) = 1$ if $x_i \in [\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ for $i = 1, 2, \ldots, n$ and $dP_\theta/d\lambda_n = 0$ otherwise. A reduction by sufficiency ($t_1 = \min\{x_i\}$, $t_2 = \max\{x_i\}$) leads to the statistical space $(T', \mathscr{A}', \mathscr{P}')$, where $T' = \{(t_1, t_2) \in \boldsymbol{R}^2 \colon t_1 \leqslant t_2 \leqslant t_1 + 1\}$, $\mathscr{A}'$ is the σ-field of Lebesgue measurable subsets of T', and if $P'_\theta \in \mathscr{P}'$, then

$$dP'_\theta/d\lambda_2(t_1, t_2) = \begin{cases} \text{const}\,(t_2 - t_1)^{n-2} & \text{if } \theta - \frac{1}{2} < t_1 \leqslant t_2 < \theta + \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

Let us assume that the set $\mathscr{E}$ of considered estimators consists of all bounded $\mathscr{A}'$-measurable functions on T'.

Now, hold $\theta = \theta_0$ fixed and consider the Hilbert space H of (equivalence classes of) functions square integrable with respect to P'_{θ_0} and defined on

$$T(\theta_0) = \{(t_1, t_2) \in \boldsymbol{R}^2 \colon \theta_0 - \tfrac{1}{2} \leqslant t_1 \leqslant t_2 \leqslant \theta_0 + \tfrac{1}{2}\}.$$

Let $\{e_1, e_2, \ldots\}$ be an orthonormal base in H consisting of bounded functions on $T(\theta_0)$ and such that $e_1(t) \equiv 1$, $t \in T(\theta_0)$. Let $e \in H$ be a bounded function such that e is orthogonal in H to e_1. Now, let $x_0(\cdot,\cdot)$ be a function on T' defined by

$$x_0(t_1 + n, t_2 + n) = \begin{cases} x_0(t_2 + n, t_1 + n + 1) = e(t_1, t_2) & \text{if } (t_1, t_2) \in \operatorname{int} T(\theta_0), \\ x_0(t_2 + n, t_1 + n + 1) = 0 & \text{if } (t_1, t_2) \in T(\theta_0) \setminus \operatorname{int} T(\theta_0), \end{cases}$$

where $n \in \{\ldots, -1, 0, 1, 2, \ldots\}$. It is easy to see (cf. [11], p. 326) that $x_0(\cdot,\cdot)$ is an unbiased estimator of zero, i.e., $x_0(\cdot,\cdot) \in \mathscr{E}_0$. So, to each bounded function in H orthogonal to e_1 there corresponds an unbiased estimator $x_0(\cdot,\cdot)$ of zero in $\mathscr{E}$ such that $x_0(t_1, t_2) = e(t_1, t_2)$ for P_{θ_0}-a.e. $(t_1, t_2) \in T(\theta_0)$. In particular, to each $e_i(\cdot)$, $i = 2, 3, \ldots$, there corresponds an element of $\mathscr{E}_0$.

Suppose that $x_1(\cdot,\cdot)$ is best unbiased for a strictly convex loss function L. Since $\mathscr{E}$ consists of bounded functions and L is finite (hence continuous on X), the risk function $R_L(\cdot, P')$ is finite for each $x(\cdot,\cdot) \in \mathscr{E}$ and each $P' \in \mathscr{P}'$. So, by the Lehmann-Scheffé-Rao lemma, for each $\theta \in \boldsymbol{R}$ there exists a function $y_\theta(\cdot,\cdot)$ such that $y_\theta(t_1, t_2) \in \partial L\big(x_1(t_1, t_2), P'\big)$ and

$$\int_{T'} x_0(t_1, t_2)\, y_\theta(t_1, t_2)\, dP'(t_1, t_2) = 0 \tag{3.10}$$

holds for each $x_0(\cdot,\cdot) \in \mathscr{E}_0$ and each $\theta \in \boldsymbol{R}$. So, if $\theta = \theta_0$, it follows from (3.10) that the function y_{θ_0} restricted to $T(\theta_0)$ is orthogonal to each e_i,

$i = 2, 3, 4, \ldots$ Therefore, we infer that $y_{\theta_0}(\cdot, \cdot)$ is constant P'_{θ_0}-a.e. on $T(\theta_0)$. Moreover, $L(\cdot, P'_{\theta_0})$ is strictly convex, and hence

$$\partial L(x', P'_{\theta_0}) \cap \partial L(x'', P'_{\theta_0}) = \emptyset$$

whenever $x' \neq x''$. So, $y_\theta(t_1, t_2) \in \partial L(x_1(t_1, t_2), P'_{\theta_0})$ and $y_\theta(t_1, t_2) = \text{const}$ P'_{θ_0}-a.e. on $T(\theta_0)$ imply that $x_1(t_1, t_2) = \text{const}\ P'_{\theta_0}$-a.e. on $T(\theta_0)$. Repeating this argument for other θ's we obtain $x_1(\cdot, \cdot) = \text{const}\ \lambda_2$-a.e. on T'.

Hence, in view of Corollary 1, we conclude that $\mathscr{U}(\mathscr{E})$ is the σ-field spanned on measurable subsets of T' which are of λ_2-measure zero. Moreover, by the factorization theorem, $\mathscr{A}'$ is the minimal sufficient σ-field, and therefore $\mathscr{S} = \mathscr{A}'$. So, we have proved that each strictly convex loss function is universal for $(T', \mathscr{A}', \mathscr{P}')$ and, simultaneously, $\mathscr{U}(\mathscr{E})$ is a proper σ-subfield of $\mathscr{S}$.

Acknowledgment. The author would like to thank Professor L. Schmetterer for discussions on problems related to universal loss functions and for his interest in this work during the author's stay in Vienna realizing an exchange between Austrian and Polish Academies of Sciences.

REFERENCES

[1] R. R. Bahadur, *On unbiased estimates of uniformly minimum variance*, Sankhyā 19 (1957), p. 211-224.

[2] B. Bednarek-Kozek and A. Kozek, *Two examples of strictly convex non-universal loss functions*, Inst. of Math. Polish Acad. of Sci., Preprint No 133, March 1978 (submitted for publication).

[3] C. Bessaga and A. Pełczyński, *Selected topics in infinite-dimensional topology*, PWN – Polish Scientific Publishers, Warszawa 1975.

[4] J. Hoffmann-Jørgensen, *The theory of analytic spaces*, Matematisk Institut, Aarhus Universitet, Various Publication Series No 10 (1970).

[5] Л. Б. Клебанов, *„Универсальные" функции потерь и несмещенные оценки*, Докл. Акад. наук СССР 203 (1972), p. 1249-1251.

[6] — *Несмещенные оценки и выпуклые функции потерь*, Зап. научн. сем. ЛОМИ 43 (1974), p. 40-52.

[7] L. B. Klebanov, Ju. V. Linnik and A. L. Ruhin, *Unbiased estimation and matrix loss functions*, Dokl. Akad. Nauk. SSSR 200 (5) (1971), p. 1024-1025; Soviet Math. Dokl. 12 (1971), p. 1526-1528.

[8] A. Kozek, *Efficiency and Cramér-Rao type inequalities for convex loss functions*, J. Multivariate Anal. 7 (1977), p. 89-106.

[9] — *Minima of convex integral functionals and unbiased estimation*, this fascicle, p. 15-27.

[10] — and Z. Suchanecki, *Conditional expectations of selectors and Jensen's inequality*, Inst. of Math. Polish Acad. of Sci., Preprint No 141, June 1978 (to appear in Lecture Notes, Proc. 1978 Wisła Conference).

[11] E. L. Lehmann and H. Scheffé, *Completeness, similar regions, and unbiased estimation*, Part 1, Sankhyā 10 (1950), p. 305-340.

[12] Ju. V. Linnik and A. L. Ruhin, *Convex loss functions in the theory of unbiased*

estimation, Dokl. Akad. Nauk SSSR 198 (1971), p. 527-529; Soviet Math. Dokl. 12 (1971), p. 839-842.

[13] J. J. Moreau, *Fonctionnelles convexes*, Mimeographed lecture notes, Séminaire sur les Équations aux Derivées Partielles, Collège de France, 1966-1967.

[14] A. R. Padmanabhan, *Some results on minimum variance unbiased estimation*, Sankhyā A 32 (1970), p. 107-114.

[15] J. Pfanzagl, *Convexity and conditional expectations*, Ann. Probability 2 (1974), p. 490-494.

[16] T. S. Pitcher, *Sets of measures not admitting necessary and sufficient statistics or subfields*, Ann. Math. Statist. 28 (1957), p. 267-268.

[17] L. Schmetterer, *On unbiased estimation*, ibidem 31 (1960), p. 1154-1163.

[18] — *On the theory of unbiased estimation*, p. 313-317 in: Proc. of the Symp. to Honour Jerzy Neyman, PWN — Polish Scientific Publishers, Warszawa 1977.

[19] — *Some results on unbiased estimation*, p. 489-503 in: Trans. of the 7th Prague Conf. and on the 1974 EMS, Vol. B, Academia, Prague 1977.

[20] — und H. Strasser, *Zur Theorie der erwartungstreuen Schätzungen*, Anz. Österreich. Akad. Wiss. Math. — Naturwiss. Kl. 6 (1974), p. 59-66.

[21] H. Strasser, *Sufficiency and unbiased estimation*, Metrika 19 (1972), p. 98-114.

Institute of Mathematics
Polish Academy of Sciences
ul. Kopernika 18, 51-617 Wrocław, Poland

Received on 26. 4. 1979