PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 1, Fasc. 1 (1980), p. 15-27

MINIMA OF CONVEX INTEGRAL FUNCTIONALS AND UNBIASED ESTIMATION

BY

ANDRZEJ KOZEK (WROCŁAW)

Abstract. Necessary and sufficient conditions for the optimality of unbiased estimators in case of arbitrary finite convex loss functions are given. These conditions are derived from a theorem on subdifferentials of convex integral functionals on Orlicz spaces. The results obtained provide a basic tool for problems concerning universal loss functions and considered in paper [12]. They are also related to Cramér-Rao type inequalities.

1. Introduction. The aim of this paper is to provide necessary and sufficient conditions for the attainment of a minimum by a convex integral functional over a linear manifold (Theorem 2). This theorem can be transformed into a general form of the Lehmann-Scheffé-Rao lemma which yields a basic tool for the paper [12] on universal loss functions.

Both Theorem 2 and the Lehmann-Scheffé-Rao lemma are analogous to the corresponding known theorems ([13], Theorem 5.3, [18], Theorem 1, [14], [19]). Their novelty consists in that neither assumptions of topological nature nor assumptions on the differentiability of convex integral functionals are explicitly required. That form is convenient for applications, e.g. in the estimation theory, because it needs no assumptions which are restrictive and unnatural for the considered problems. However, in the proof of Theorem 2, an appropriate Orlicz space is constructed such that the considered integral functional becomes continuous for the norm topology. This proof requires also using a theorem on decomposition of subdifferentials of convex integral functionals on Orlicz spaces (Theorem 1). An analogous theorem for Köthe spaces was proved in [3] by Clauzure, however, Orlicz spaces are not contained in the class of Köthe spaces [4]. Our proof of Theorem 1 is different from that given by Clauzure.

2. Convex integral functionals on Orlicz spaces. The theory of Orlicz spaces originated in works [16] and [17] by Orlicz and based on properties of conjugate N-functions which were first introduced and investigated in [1] by Birnbaum and Orlicz. Orlicz spaces that we shall consider represent one of the known extensions of those original results.

Let X be a Banach space and let $(T, \mathscr{A}, \mu)$ be a measure space, where T is a set, $\mathscr{A}$ a σ-field of subsets of T, and μ a σ-finite complete measure on $\mathscr{A}$.

Definition 1 (Rockafellar). A function $f\colon X\times T \to (-\infty, +\infty]$, not identically equal to $+\infty$, is called a *normal integrand* if

(a) f is $(\mathscr{B}_X \otimes \mathscr{A})$-measurable, where $\mathscr{B}_X$ stands for the σ-field of Borel subsets of X,

(b) for every $t\in T$, $f(\cdot, t)$ is lower semicontinuous on X.

If, in addition,

(c) $f(\cdot, t)$ is convex on X for each $t\in T$,

then f is called a *normal convex integrand.*

Definition 2 (Kozek [8]). $\Phi\colon X\times T \to [0, \infty]$ is called an *N-function* if Φ is a normal convex integrand and if the following conditions are fulfilled for each $t\in T$:

(d) $\Phi(0, t) \equiv 0$, $\Phi(x, t) = \Phi(-x, t)$,

(e) $\lim\limits_{\|x\|\to\infty} \Phi(x, t) = +\infty$,

(f) $\Phi(\cdot, t)$ is continuous at zero.

We shall use Φ to denote N-functions, only. Moreover, let $M_X(\mathscr{A})$ denote the set of all strongly $\mathscr{A}$-measurable functions from T into X. We shall identify functions which are equal μ-a.e. to each other.

Definition 3. An Orlicz space L_Φ is a *vector space* of functions $x(\cdot)\in M_X(\mathscr{A})$ such that

$$I_\Phi(kx(\cdot)) = \int \Phi(kx(t), t)\,\mu(dt) < \infty$$

holds for some constant $k > 0$ and for a given N-function Φ.

L_Φ can be endowed with two norm topologies and the corresponding norms N_1 and N_2 are given by

$$N_1(x(\cdot)) = \inf_{\alpha>0}\left(\frac{1}{\alpha}\big(1+I_\Phi(\alpha x(\cdot))\big)\right), \qquad x(\cdot)\in L_\Phi,$$

$$N_2(x(\cdot)) = \inf\left\{\frac{1}{\alpha}\colon I_\Phi(\alpha x(\cdot)) \leqslant 1, \alpha > 0\right\}, \qquad x(\cdot)\in L_\Phi.$$

Norms N_1 and N_2 are equivalent on L_Φ because

$$N_2(x(\cdot)) \leqslant N_1(x(\cdot)) \leqslant 2N_2(x(\cdot))$$

holds for each $x(\cdot)\in L_\Phi$. Remind that N_1 is called the *Orlicz norm* on L_Φ whereas N_2 is called the *Luxemburg norm* on L_Φ. The space L_Φ endowed with the norm topology is a Banach space ([8], Theorem 2.4).

Let us note that condition (e) in the definition of N-function is satisfied if and only if L_Φ is complete for the norm topology. Moreover, condition (f) is fulfilled if and only if L_Φ is *topologically decomposable*, i.e. if there exist sets $T_n\in\mathscr{A}$,

$$T_{n+1}\supset T_n,\quad \mu(T_n)<\infty,\quad \mu(T\setminus\bigcup_{i=1}^{\infty}T_i)=0,$$

such that an embedding of the strict inductive limit of Banach spaces $L_\infty(T_i,\mu)$ into L_Φ is continuous ($x(\cdot)\in L_\infty(T_i,\mu)$ is identified with a function defined on T which equals $x(\cdot)$ on T_i and equals zero outside of T_i) ([5], Théorème 1.1.4).

Let Y be the dual space of X. Assume that Ψ, the conjugate of Φ, is a function from $Y\times T$ into $[0,\infty]$ given by

$$\Psi(y,t)=\sup\{(x,y)-\Phi(x,t)\colon x\in X\},$$

where $(x,y)=y(x)$. If X is separable, then Ψ is an N-function ([8], Proposition 4.6) and L_Ψ is a Banach space. If, moreover, Y is separable (i.e., if Y has the Radon-Nikodym property), then L'_Φ, the dual space of L_Φ, admits a representation

$$L'_\Phi=L_\Psi\oplus\Lambda.$$

The function $y(\cdot)\in L_\Psi$ is identified here with a continuous functional $\varphi\in L'_\Phi$ given by

$$\varphi\big(x(\cdot)\big)=\int\big(x(t),y(t)\big)\,\mu(dt),\qquad x(\cdot)\in L_\Phi.$$

Elements of Λ are called *singular functionals.* For each $\varphi\in\Lambda$ there exists a decreasing sequence of sets $\{A_k\}$,

$$A_k\in\mathscr{A},\qquad \mu(\bigcap_{k=1}^{\infty}A_k)=0,$$

such that $\varphi\big(1_{A_k^-}(\cdot)\,x(\cdot)\big)=0$ for every k and for each $x(\cdot)\in L_\Phi$, A_k^- being the complement of A_k ([5], Corollaire 1.4.6 and Théorème 1.5.2, [10], Theorem 2.2 and Proposition 2.1).

If f is a normal convex integrand, then I_f given by

$$I_f\big(x(\cdot)\big)=\int f\big(x(t),t\big)\,\mu(dt),\qquad x(\cdot)\in L_\Phi, \tag{1}$$

is a convex functional on L_Φ. If t is fixed and $x_0\in\operatorname{dom}f(\cdot,t)$, then $\partial f(x_0,t)$ consists of all functionals $y\in Y$ such that the inequality

$$f(x,t)\geqslant f(x_0,t)+(x-x_0,y)$$

holds for every $x \in X$. Then $\partial f(x_0, t)$ is called a *subdifferential of* $f(\cdot, t)$ at x_0 and elements of $\partial f(x_0, t)$ are called *subgradients of* $f(\cdot, t)$ at x_0. Similarly, if $x_0(\cdot) \in L_\Phi$ and $x_0(\cdot) \in \operatorname{dom} I_f$, then $\partial I_f(x_0(\cdot))$ is called the *subdifferential of* I_f at $x_0(\cdot)$ and it consists of all functionals $\varphi \in L'_\Phi$ such that the inequality

$$I_f(x(\cdot)) \geqslant I_f(x_0(\cdot)) + \varphi(x(\cdot) - x_0(\cdot))$$

holds for every $x(\cdot) \in L_\Phi$.

Denote by $Df(x_0(\cdot))$ the set of all elements $y(\cdot)$ of L_Ψ such that $y(t) \in \partial f(x_0(t), t)$ μ-a.e. Moreover, let $K_f(x_0(\cdot))$ be the set of all singular functionals $\varphi \in \Lambda$ such that $\varphi(x(\cdot) - x_0(\cdot)) \leqslant 0$ for each $x(\cdot) \in \operatorname{dom} I_f$.

We prove a theorem on representation of the subdifferential of I_f. This theorem extends Theorems 3.1 and 3.2 in [10].

THEOREM 1. *Let X and $Y = X'$ be separable Banach spaces and let f be a normal convex integrand on $X \times T$. If I_f is a convex functional on L_Φ given by* (1) *and $x_0(\cdot) \in \operatorname{dom} I_f$, then*

$$\partial I_f(x_0(\cdot)) = Df(x_0(\cdot)) + K_f(x_0(\cdot))$$

and $\partial I_f(x_0(\cdot))$ is empty if $Df(x_0(\cdot))$ is empty.

If, moreover, $x_0(\cdot) \in \operatorname{int} \operatorname{dom} I_f(x_0(\cdot))$, then $\partial I_f(x_0(\cdot))$ is non-empty, $K_f(x_0(\cdot)) = \{0\}$ and each $\mathscr{A}$-measurable function $y(\cdot)$ such that $y(t) \in \partial f(x_0(t), t)$ is an element of L_Ψ, and hence $y(\cdot) \in Df(x_0(\cdot))$.

Proof. It is clear that $0 \in K_f(x_0(\cdot))$. Thus $\partial I_f(x_0(\cdot))$ is empty if and only if $Df(x_0(\cdot))$ is empty. If $y(t) \in \partial f(x_0(t), t)$ for every $t \in T$, then

$$f(x, t) \geqslant f(x_0(t), t) + (x - x_0(t), y(t))$$

holds for each $x \in X$ and $t \in T$. Thus, if $y(\cdot) \in Df(x_0(\cdot))$ and $\varphi \in K_f(x_0(\cdot))$, then the definitions of $Df(x_0(\cdot))$ and $K_f(x_0(\cdot))$ imply that $y(\cdot) + \varphi \in \partial I_f(x_0(\cdot))$.

Conversely, let $\varphi \in \partial I_f(x_0(\cdot))$. Then, in view of the decomposition of L'_Φ, we have $\varphi = y(\cdot) + \varphi'$, where $y(\cdot) \in L_\Psi$ and $\varphi' \in \Lambda$.

We show that $y(t) \in \partial f(x_0(t), t)$ μ-a.e. and that $\varphi' \in K_f(x_0(\cdot))$. Suppose to the contrary that the set

$$\hat{T} = \{t \in T\colon\ y(t) \notin \partial f(x_0(t), t)\}$$

is not μ-null. By Theorems I.3.B.4 and I.3.B.5 in [6] (p. 8 and 9) the graph of $y(\cdot)$ belongs to $\mathscr{B}_Y \otimes \mathscr{A}$. Hence and since the graph of the multifunction $t \to \partial f(x_0(t), t)$ is $(\mathscr{B}_Y \otimes \mathscr{A})$-measurable (it is equal to the set $\{(y, t)\colon f(x_0(t), t) + f^*(y, t) - (x_0(t), y) \leqslant 0\}$, where $f^*(\cdot, t)$ is the conjugate of $f(\cdot, t)$), $\hat{T}$ is $\mathscr{A}$-measurable ([20], Theorem 4.2g v, vii). Therefore, $\mu(\hat{T}) \geqslant \varepsilon > 0$ holds. We can assume $\mu(\hat{T}) < \infty$, for – otherwise – we can take instead of $\hat{T}$ any $\mathscr{A}$-measurable subset of $\hat{T}$ of a finite μ-measure. Let us consider a multifunction

$$t \to \{x \in X\colon f(x, t) - (x - x_0(t), y(t)) < f(x_0(t), t)\}, \qquad t \in \hat{T}.$$

$(\mathscr{B}_X \otimes \mathscr{A})$-measurability of functions inside of the brackets yields $(\mathscr{B}_X \otimes \mathscr{A})$-measurability of the graph of this multifunction. Thus, by Theorem 5.10 in [20], there exists an $\mathscr{A}$-measurable selector $x(\cdot)$ which satisfies the condition

$$f(\bar{x}(t), t) < f(x_0(t), t) + (\bar{x}(t) - x_0(t), y(t))$$

μ-a.e. for $t \in \hat{T}$. Let us take numbers N and n such that

$$\mu\{t \in \hat{T}\colon \|\bar{x}(t)\| \leqslant N \text{ and } f(\bar{x}(t), t) \geqslant -N\} > \frac{3}{4}\varepsilon,$$

$$\mu(A_n \cap \hat{T}) < \frac{1}{4}\varepsilon \quad \text{and} \quad \mu(T_n^- \cap \hat{T}) < \frac{1}{4}\varepsilon,$$

where A_n and T_n are elements of the sequences of sets characterizing φ' and the decomposability of L_Φ, respectively. Now, we put

$$\tilde{T} = \{t \in \hat{T} \cap T_n \cap A_n^-\colon \|\bar{x}(t)\| \leqslant N \text{ and } f(\bar{x}(t), t) \geqslant -N\}.$$

Clearly, $\mu(\tilde{T}) > \frac{1}{4}\varepsilon$. Define $\tilde{x}_0(\cdot)$ by

$$\tilde{x}_0(t) = \begin{cases} x_0(t) & \text{if } t \notin \tilde{T}, \\ \bar{x}(t) & \text{if } t \in \tilde{T}. \end{cases}$$

Then, $\tilde{x}_0(\cdot) \in L_\Phi$. Moreover, $\tilde{x}_0(\cdot) \in \operatorname{dom} I_f$ and we have

$$\begin{aligned} -\infty &< I_f(\tilde{x}_0(\cdot)) \\ &= I_f(x_0(\cdot)) + \int_{\tilde{T}} \big(f(\tilde{x}_0(t), t) - f(x_0(t), t)\big)\,\mu(dt) \\ &< I_f(x_0(\cdot)) + \int (\tilde{x}_0(t) - x_0(t), y(t))\,\mu(dt) \\ &= I_f(x_0(\cdot)) + \int (\tilde{x}_0(t) - x_0(t), y(t))\,\mu(dt) + \varphi'(\tilde{x}_0(\cdot) - x_0(\cdot)) < \infty. \end{aligned}$$

Hence, $y(\cdot) + \varphi'$ cannot be an element of $\partial I_f(x_0(\cdot))$ and this yields a contradiction.

Finally, suppose that $\varphi' \notin K_f(x_0(\cdot))$. Then there exists $x(\cdot) \in \operatorname{dom} I_f$ such that $\varphi'(x(\cdot) - x_0(\cdot)) \geqslant \varepsilon > 0$. Let us take n such that

$$\Big|\int_{A_n} (x(t) - x_0(t), y(t))\,\mu(dt)\Big| < \frac{\varepsilon}{3}$$

and

$$\Big|\int_{A_n} \big(f(x_0(t), t) - f(x(t), t)\big)\,\mu(dt)\Big| < \frac{\varepsilon}{3},$$

where A_n is an appropriate element of the sequence of sets characterizing φ'.

Now, let $\tilde{x}(\cdot)$ be given by

$$\tilde{x}(t) = \begin{cases} x_0(t) & \text{if } t \notin A_n, \\ x(t) & \text{if } t \in A_n. \end{cases}$$

Clearly, $\tilde{x}(\cdot) \in L_\Phi$ and $\tilde{x}(\cdot) \in \operatorname{dom} I_f$. Moreover, we have

$$I_f(x_0(\cdot)) - I_f(\tilde{x}(\cdot)) + \int (\tilde{x}(t) - x_0(t), y(t))\,\mu(dt) + \varphi'(\tilde{x}(\cdot) - x_0(\cdot)) > \frac{\varepsilon}{3}$$

and, therefore, $y(\cdot) + \varphi'$ cannot be an element of $\partial I_f(x_0(\cdot))$.

So, we conclude that $y(\cdot) \in Df(x_0(\cdot))$ and $\varphi' \in K_f(x_0(\cdot))$.

If $x_0(\cdot) \in \operatorname{int} \operatorname{dom} I_f$, then I_f is continuous at $x_0(\cdot)$, and hence I_f is subdifferentiable at $x_0(\cdot)$ ([15], Propositions 5f and 10c), i.e., $\partial I_f(x_0(\cdot)) \neq \emptyset$. The equality $K_f(x_0(\cdot)) = \{0\}$ follows now trivially from the definition of $K_f(x_0(\cdot))$.

We show that every $\mathscr{A}$-measurable function $y(\cdot)$ such that $y(t) \in \partial f(x_0(t), t)$ is an element of L_Ψ. To this end it is enough to prove that $\int (x(t), y(t))\,\mu(dt)$ is finite for each $x(\cdot) \in L_\Phi$ ([8], Proposition 4.4). Indeed, by the convexity of f, for each $t \in T$, $x \in X$ and $\lambda > 0$ we have

$$\lambda^{-1}\big(f(x_0(t) + \lambda x, t) - f(x_0(t), t)\big) \geqslant (x, y(t))$$

and

$$\lambda^{-1}\big(f(x_0(t) - \lambda x, t) - f(x_0(t), t)\big) \geqslant -(x, y(t)).$$

Thus, if $x(\cdot) \in L_\Phi$ and λ is small enough, then we get

$$\int (x(t), y(t))\,\mu(dt) \leqslant \lambda^{-1}\big(I_f(x_0(\cdot) + \lambda x(\cdot)) - I_f(x_0(\cdot))\big) < \infty$$

and

$$-\int (x(t), y(t))\,\mu(dt) \leqslant \lambda^{-1}\big(I_f(x_0(\cdot) - \lambda x(\cdot)) - I_f(x_0(\cdot))\big) < \infty.$$

Hence

$$\left|\int (x(t), y(t))\,\mu(dt)\right| < \infty \quad \text{for every } x(\cdot) \in L_\Phi.$$

This completes the proof of the theorem.

3. Minima of convex integral functionals. This section contains a detailed discussion of some problems related to the attainment of a minimum by a convex integral functional on a linear manifold. The optimality in the theory of unbiased estimation may be interpreted as a common minimum of integral functionals called a *risk function* over a linear manifold called a *set of unbiased estimators.*

Denote by f a normal convex integrand on $X \times T$ (Definition 1) and let $x_i(\cdot)$, $i = 0, 1$, be $\mathscr{A}$-measurable functions from T into X (X and $Y = X'$ are assumed to be separable Banach spaces). Moreover, let

$$I_f(x_1(\cdot) + \lambda x_0(\cdot)) = \int_T f(x_1(t) + \lambda x_0(t), t)\,\mu(dt).$$

If $x_0(\cdot)$ and $x_1(\cdot)$ are fixed, we write

$$i_f(\lambda) = I_f\big(x_1(\cdot)+\lambda x_0(\cdot)\big).$$

Clearly, $i_f\colon \boldsymbol{R} \to \bar{\boldsymbol{R}}$ is convex. In the sequel we assume that $i_f(0) = I_f\big(x_1(\cdot)\big) \in \boldsymbol{R}$ and that $f(\cdot, t)$ is continuous at $x_1(t)$ for each $t \in T$. Thus $\partial f\big(x_1(t), t\big)$ is non-empty and star-weakly compact for every $t \in T$ ([15], Section 10c).

We shall discuss in terms of subgradients of $f(\cdot, t)$, only, both sufficient and necessary conditions for the inequality

$$I_f\big(x_1(\cdot)\big) \leqslant I_f\big(x_1(\cdot)+\lambda x_0(\cdot)\big) \tag{2}$$

to hold for each $\lambda \in \boldsymbol{R}$ and each $x_0(\cdot)$ from a given set E of functions. If I_f is considered as a convex functional on a topological vector space, the necessary and sufficient conditions for (2) to hold are well known ([7], p. 30). Here we avoid assumptions of topological nature on I_f at least in formulations of theorems. This is convenient for purposes of the theory of unbiased estimation (see the Lehmann-Scheffé-Rao lemma given in the next section). None the less, we shall use arguments of topological nature in proofs of the theorems.

A very simple and well-known sufficient condition for inequality (2) to hold is given in the following

PROPOSITION 1. *Let $x_1(\cdot)$ and $x_0(\cdot)$ be given $\mathscr{A}$-measurable functions from T into X and let $I_f\big(x_1(\cdot)\big) \in \boldsymbol{R}$. If $y(\cdot)$ is an $\mathscr{A}$-measurable function from T into Y such that $y(t) \in \partial f\big(x_1(t), t\big)$ μ-a.e. for $t \in T$ and*

$$\int \big(x_0(t), y(t)\big)\mu(dt) = 0 \tag{3}$$

holds, then inequality (2) *is valid for each $\lambda \in \boldsymbol{R}$.*

Let us note that $i_f(\lambda)$ may be here equal to $+\infty$ for $\lambda \neq 0$. Proposition 1 implies trivially the following

COROLLARY 1. *Let $I_f\big(x_1(\cdot)\big) \in \boldsymbol{R}$ and let E be a class of measurable functions. If for each $x_0(\cdot) \in E$ there exists a measurable function $y_{x_0}(\cdot)$ such that $y_{x_0}(t) \in \partial f\big(x_1(t), t\big)$ μ-a.e. and*

$$\int \big(x_0(t), y_{x_0}(t)\big)\mu(dt) = 0 \tag{3'}$$

holds, then inequality (2) *is valid for each $\lambda \in \boldsymbol{R}$ and $x_0(\cdot) \in E$.*

Clearly, it may happen that there exists a function $y(\cdot)$ such that $y(t) \in \partial f\big(x_1(t), t\big)$ μ-a.e. and $y(\cdot)$ fulfils condition (3′) for each $x_0(\cdot) \in E$. Then (2) holds for each $x_0(\cdot) \in \operatorname{Lin} E$. For instance, if $f(\cdot, t)$ is weakly (Gateaux) differentiable for every $t \in T$, then $\partial f(x, t)$ contains only one element $f'(x, t)$ and, therefore, each function $y_{x_0}(\cdot)$ such that $y_{x_0}(t) \in \partial f\big(x_1(t), t\big)$ equals $f'\big(x_1(t), t\big)$ μ-a.e.

Now, we discuss necessary conditions for inequality (2) to hold. Suppose that $x_1(\cdot)$ and $x_0(\cdot)$ are given and that $i_f(\lambda) = I_f\big(x_1(\cdot)+\lambda x_0(\cdot)\big)$ is of the form

$$i_f(\lambda) = \begin{cases} +\infty & \text{if } \lambda \neq 0, \\ a,\, a \in \boldsymbol{R}, & \text{if } \lambda = 0. \end{cases}$$

Then condition (3) need not be fulfilled and no characterization of the integral in (3) is possible. Namely, if $y(t) \in \partial f(x_1(t), t)$, then the following cases are possible: the integral $\int \big(x_0(t), y(t)\big)\, \mu(dt)$ may be equal to $+\infty$ or to $-\infty$, it may be finite and different from zero, it may be equal to zero and, finally, it may be not well defined. It is not difficult to give simple examples for each of these cases.

If $y(t) \in \partial f(x_1(t), t)$ and $i_f(\lambda)$ is finite for $\lambda \in [0, \varepsilon)$, then it is easy to infer from the definition of the subdifferential that

$$\int \big(x_0(t), y(t)\big)\, \mu(dt) < +\infty.$$

Similarly, if $i_f(\lambda)$ is finite for $\lambda \in (-\varepsilon, 0]$, then

$$\int \big(x_0(t), y(t)\big)\, \mu(dt) > -\infty.$$

Now, we discuss the regular case where $i_f(\lambda)$ is finite for $\lambda \in (-\varepsilon, \varepsilon)$, $\varepsilon > 0$. We start with the following

PROPOSITION 2. *Let $x_1(\cdot)$ and $x_0(\cdot)$ be given $\mathscr{A}$-measurable functions from T into X and let $f(\cdot, t)$ be finite and continuous at $x_1(t)$ for each $t \in T$. If inequality* (2) *holds and if $i_f(\lambda)$ is finite for $\lambda \in (-\varepsilon, \varepsilon)$, then there exists a measurable function $y(\cdot)$ such that $y(t) \in \partial f(x_1(t), t)$ for each $t \in T$ and*

$$\int \big(x_0(t), y(t)\big)\, \mu(dt) = 0$$

holds.

Proof. First, we note that the function

$$\lambda \to \frac{1}{\lambda}\Big(f\big(x_1(t)+\lambda x_0(t), t\big) - f\big(x_1(t), t\big)\Big)$$

is non-decreasing. Moreover, let $f'\big(x_1(\cdot), \cdot; x_0(\cdot)\big)$ and $f'\big(x_1(\cdot), \cdot; -x_0(\cdot)\big)$ be given by

$$f'\big(x_1(t), t; x_0(t)\big) = \lim_{n\to\infty} n\left(f\left(x_1(t)+\frac{1}{n}\, x_0(t), t\right) - f\big(x_1(t), t\big)\right)$$

and

$$f'\big(x_1(t), t; -x_0(t)\big) = \lim_{n\to\infty} n\left(f\left(x_1(t)-\frac{1}{n}\, x_0(t), t\right) - f\big(x_1(t), t\big)\right),$$

respectively. By the monotone convergence theorem and (2) we have

$$0 \leqslant \int f'\big(x_1(t), t; x_0(t)\big)\, \mu(dt) < \infty \tag{4a}$$

and

(4b) $$0 \leqslant \int f'(x_1(t), t; -x_0(t))\,\mu(dt) < \infty.$$

Moreover,

$$f'(x_1(t), t; x_0(t)) = \sup\{(x_0(t), y) : y \in \partial f(x_1(t), t)\}$$

and

$$f'(x_1(t), t; -x_0(t)) = \sup\{-(x_0(t), y) : y \in \partial f(x_1(t), t)\}$$

([15], p. 65). The multifunction $t \to \partial f(x_1(t), t)$ is measurable and sets $\partial f(x_1(t), t)$ are star-weakly compact. Since Y is separable, the considered multifunction is of Suslin type and, therefore, there exist $\mathscr{A}$-measurable selectors $y^-(\cdot)$ and $y^+(\cdot)$ of $\partial f(x_1(\cdot), \cdot)$ such that

$$f'(x_1(t), t; x_0(t)) = (x_0(t), y^+(t))$$

and

$$f'(x_1(t), t; -x_0(t)) = -(x_0(t), y^-(t))$$

([20], Theorem 9.1). Therefore, (4a) and (4b) imply that there exists an $\alpha \in [0, 1]$ such that

$$\int (x_0(t), \alpha y^+(t) + (1-\alpha) y^-(t))\,\mu(dt) = 0.$$

Clearly, by the convexity of $\partial f(x_1(t), t)$ we have

$$\alpha y^+(t) + (1-\alpha) y^-(t) \in \partial f(x_1(t), t) \quad \mu\text{-a.e.}$$

Thus, to complete the proof of the proposition it is enough to put $y(\cdot) = \alpha y^+(\cdot) + (1-\alpha) y^-(\cdot)$.

Remark. If f does not depend on t, then the assertion of Proposition 2 reduces to the following one:

There exists a measurable selector $y(\cdot)$ of $\partial f(x_1(\cdot))$ such that

$$\int (x_0(t), y(t))\,\mu(dt) = 0$$

holds.

If f is weakly differentiable, then the mapping $x \to f'(x)$ is weakly continuous ([15], p. 80). Thus $y(t) = f'(x_1(t))$, where $f'(\cdot)$ is Borel measurable. So, it is interesting to ask whether $y(\cdot)$ admits a representation of the form $y(t) = v(x_1(t))$, where $v(\cdot)$ is a Borel measurable selector of the multifunction $x \to \partial f(x)$. Let us note that a Castaing representation of Borel measurable selectors of the multifunction $x \to \partial f(x)$ exists whenever f is, e.g., continuous and convex on X. None the less, the following example shows that the answer to this problem is, in general, negative.

Example. Let $T = \{1, 2, 3\}$ and let P be a probability measure on 2^T given by $P(1) = P(2) = P(3) = \frac{1}{3}$. Let $X = \boldsymbol{R}$, $f(x, t) = |x|$ for each $t \in T$

and let $x_1(\cdot)$ and $x_0(\cdot)$ be given by

$$x_1(1) = x_1(2) = 0, \quad x_1(3) = 1, \quad x_0(1) = -1, \quad x_0(2) = 1, \quad x_0(3) = \tfrac{3}{4}.$$

Then

$$i_f(\lambda) = I_f\big(x_1(\cdot)+\lambda x_0(\cdot)\big) = \tfrac{2}{3}|\lambda|+\tfrac{1}{3}|1+\tfrac{3}{4}\lambda| \quad \text{and} \quad i_f(0) \leqslant i_f(\lambda)$$

for each $\lambda \in \boldsymbol{R}$. Moreover, $\partial f\big(x_1(1)\big) = \partial f\big(x_1(2)\big) = [-1, 1]$ and $\partial f\big(x_1(3)\big) = \{1\}$. Thus, if $y(\cdot)$ is a function from T into $Y = \boldsymbol{R}$ such that condition (3) is fulfilled and $y(t) \in \partial f\big(x_1(t)\big)$ for each $t \in T$, then

$$\int x_0(t)\, y(t)\, P(dt) = \tfrac{1}{3}\big(-y(1)+y(2)+\tfrac{3}{4}\big) = 0.$$

Therefore, $y(1) = \frac{3}{4}+y(2)$ and $y(1) \neq y(2)$. So, $y(\cdot)$ cannot be a function of $x_1(\cdot)$ because $x_1(1) = x_1(2) = 0$ holds.

If instead of a single function $x_0(\cdot)$ we have a set of functions E, then Proposition 2 applied to each particular function $x_0(\cdot) \in E$ yields immediately the following

COROLLARY 2. *Let functions $x_1(\cdot)$ and $x_0(\cdot)$ $(x_0(\cdot) \in E)$ be measurable and let $f(\cdot, t)$ be finite and continuous at $x_1(t)$ for each $t \in T$. Assume, moreover, that inequality* (2) *holds for $x_0(\cdot) \in E$ and that $I_f\big(x_1(\cdot)+\lambda x_0(\cdot)\big)$ is finite for $\lambda \in (-\varepsilon, \varepsilon)$, $\varepsilon > 0$, $x_0(\cdot) \in E$ (ε may depend on $x_0(\cdot)$). Then for each $x_0(\cdot) \in E$ there exists a measurable function $y_{x_0}(\cdot)$ such that $y_{x_0}(t) \in \partial f\big(x_1(t), t\big)$ μ-a.e. for $t \in T$ and*

$$\int \big(x_0(t), y_{x_0}(t)\big)\, \mu(dt) = 0$$

holds.

Remark. If E is not convex, then the dependence of the subgradients $y_{x_0}(\cdot)$ on the direction $x_0(\cdot) \in E$ cannot be avoided.

Now, we are interested in the following problem: when can we replace in the assertion of Corollary 2 "the collection $\{y_{x_0}(\cdot)\colon x_0(\cdot) \in E\}$" by "a subgradient $y(\cdot)$ independent of $x_0(\cdot) \in E$"? If such a replacement is justified, then condition (3) is fulfilled for each $x_0(\cdot) \in \operatorname{Lin} E$ and, moreover, in view of Proposition 1 condition (2) is satisfied *a posteriori* for each $x_0(\cdot) \in \operatorname{Lin} E$. Hence the assumption that E is a linear space is not restrictive for our purposes. Let us note that if $f(\cdot, t)$ is weakly differentiable at each point $x_1(t)$, $t \in T$, then $\partial f\big(x_1(t), t\big)$ contains only a unique element and, therefore, functions $y_{x_0}(\cdot)$ are equal to each other μ-a.e. However, if $f(\cdot, t)$ is not weakly differentiable, we are unable to prove this stronger version of Corollary 2 unless a nice topology in E is available. Therefore, in the proof of Theorem 2 given in the sequel we construct an appropriate Orlicz space L_Φ. None the less, in Theorem 2 we do not assume explicitly the existence of any topology in E.

THEOREM 2. *Let f be a normal convex integrand on $X \times T$, $x_1(\cdot)$ an $\mathcal{A}$-measurable function from T into X, and E a vector space of $\mathcal{A}$-measur-*

able functions from T into X. Assume, moreover, that $f(\cdot, t)$ is continuous at $x_1(t)$ for each $t \in T$. If

$$I_f(x_1(\cdot)) \leqslant I_f(x_1(\cdot)+\lambda x_0(\cdot))$$

for each $x_0(\cdot) \in E$ and $\lambda \in \boldsymbol{R}$ and if $I_f(x_1(\cdot)+\lambda x_0(\cdot))$ is finite for every $x_0(\cdot) \in E$ and $\lambda \in (-\varepsilon, \varepsilon)$, $\varepsilon = \varepsilon(x_0(\cdot)) > 0$, then there exists an $\mathscr{A}$-measurable function $y(\cdot)$: $T \to Y$ such that $y(t) \in \partial f(x_1(t), t)$ μ-a.e. and

$$\int (x_0(t), y(t))\, \mu(dt) = 0$$

holds for each $x_0(\cdot) \in E$.

Proof. First, we define an Orlicz space L_Φ appropriate for our purposes. We put

$$\Phi(x, t) = \max\{f(x_1(t)+x, t)-f(x_1(t), t),\ f(x_1(t)-x, t)-f(x_1(t), t),\ \|x\|\}.$$

It is easy to see that Φ is an N-function and, moreover, that

$$f(x_1(t)+x, t) \leqslant \Phi(x, t)+f(x_1(t), t).$$

A modular I_Φ given by

$$I_\Phi(x(\cdot)) = \int \Phi(x(t), t)\, \mu(dt), \qquad x(\cdot) \in L_\Phi,$$

is continuous at zero, and $\{x(\cdot) \in L_\Phi : I_\Phi(x(\cdot)) \leqslant 1\}$ is a unit ball in L_Φ endowed with the Luxemburg norm topology ([2], Theorem 2.10). Thus, $\tilde{I}_f$ given by

$$\tilde{I}_f(x(\cdot)) = \int f(x_1(t)+x(t), t)\, \mu(dt), \qquad x(\cdot) \in L_\Phi,$$

is bounded from above on a neighbourhood of zero in L_Φ. Let us note that the condition $I_f(x_1(\cdot)+\lambda x_0(\cdot)) \in \boldsymbol{R}$ for $\lambda \in (-\varepsilon, \varepsilon)$ implies that each function $x_0(\cdot) \in E$ is an element of L_Φ. Thus, $\tilde{I}_f$ is continuous at $0 \in L_\Phi$ ([2], Theorem 2.10) and attains at zero its minimum over the subspace E (not necessarily closed). In view of Theorem 2.5b in [9] there exists a linear functional $\varphi \in L'_\Phi$ such that $\varphi \in \partial \tilde{I}_f(0)$ and $\varphi(x_0(\cdot)) = 0$ for each $x_0(\cdot) \in E$. Since $0 \in \operatorname{int}\operatorname{dom} \tilde{I}_f$, we infer from Theorem 1 that φ admits an integral representation. Thus, there exists $y(\cdot): T \to Y$ such that $y(t) \in \partial f(x_1(t), t)$ μ-a.e. and

$$\int (x_0(t), y(t))\, \mu(dt) = 0$$

holds for every $x_0(\cdot) \in E$. This completes the proof of the theorem.

4. Lehmann-Scheffé-Rao lemma. The terminology commonly used in statistics and appearing in the following lemma is given at the beginning of Section 2 in [12].

LEHMANN-SCHEFFÉ-RAO LEMMA. *Let $(T, \mathscr{A}, \mathscr{P})$ be a statistical space, X a Banach space with a separable dual Y, L: $X \times \mathscr{P} \to [0, \infty)$ a convex* *loss function, $\mathscr{E}$ a set of estimators, $\mathscr{E}_0$ the vector space of unbiased estimators of zero, and let R_L stand for the risk function corresponding to L. Assume, moreover, that $x_1(\cdot)\in\mathscr{E}$ is a given estimator and that $\mathscr{A}$ is $\mathscr{P}$-complete* (cf. [12]).

(a) *If there exist $\mathscr{A}$-measurable functions $y_P: T\to Y$ such that $y_P(t)\in \partial L(x_1(t), P)$ P-a.e. for every $P\in\mathscr{P}$ and if*

$$\int (x_0(t), y_P(t))\, P(dt) = 0$$

holds for each $x_0(\cdot)\in\mathscr{E}_0$ and $P\in\mathscr{P}$, then $x_1(\cdot)$ is uniformly best unbiased for L.

(b) *If $x_1(\cdot)$ is uniformly best unbiased for L and if the risk function $R_L(x_1(\cdot)+\lambda x_0(\cdot), P)$ is finite for each $x_0(\cdot)\in\mathscr{E}_0$, $P\in\mathscr{P}$ and $\lambda\in(-\varepsilon,\varepsilon)$, $\varepsilon = \varepsilon(x_0(\cdot), P) > 0$, then there exist $\mathscr{A}$-measurable functions $y_P(\cdot)$ such that $y_P(t)\in\partial L(x_1(t), P)$ P-a.e. and*

$$\int (x_0(t), y_P(t))\, P(dt) = 0$$

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

It is easy to see that this lemma is, for each fixed $P\in\mathscr{P}$, a particular case of Proposition 1 and Theorem 2.

For the case of a quadratic loss function the Lehmann-Scheffé-Rao lemma was given in 1950 by Lehmann and Scheffé [13] and in 1952 by Rao [18]. Some extensions of these original results can be found in [9], [11], [14] and [19]. The lemma has important consequences for the theory of unbiased estimation with convex loss functions. For instance, it provides a basic tool for problems related to universal loss functions. Another consequence is an equivalence of two optimal properties of estimators: optimality in the class of unbiased estimators and efficiency with respect to an inequality of Cramér-Rao type. Following [11] and using the Lehmann-Scheffé-Rao lemma instead of Lemma 4.1 in [11] one can easily derive a stronger and more convenient version of Theorem 2 in [11].

THEOREM 3. *Let $x_1(\cdot)\in\mathscr{E}$ and let $R_L(x_1(\cdot)+\lambda x_0(\cdot), P)$ be finite for each $x_0(\cdot)\in\mathscr{E}_0$, $P\in\mathscr{P}$ and $\lambda\in(-\varepsilon,\varepsilon)$, $\varepsilon = \varepsilon(x_0(\cdot), P) > 0$. Then $x_1(\cdot)$ is uniformly best unbiased if and only if there exists a Cramér-Rao type inequality such that $x_1(\cdot)$ is efficient with respect to it.*

REFERENCES

[1] Z. Birnbaum und W. Orlicz, *Über die Verallgemeinerung des Begriffes der zueinander konjugierten Potenzen*, Studia Math. 3 (1931), p. 1-67.

[2] A. Brønsted, *Conjugate convex functions in topological vector spaces*, Mat.-Fys. Medd. Danske Vid. Selsk. 34 (2) (1964), p. 1-27.

[3] P. Clauzure, *Sur la décomposition d'un sous-gradient d'une intégrale fonctionnelle convexe défini sur un espace de Köthe normé*, C. R. Acad. Sci. Paris, Sér. A, 283 (1976), p. 177-180.

[4] — *Exemple d'espace d'Orlicz qui n'est pas un espace de Köthe*, Séminaire d'Analyse Convexe, Montpellier, 8 (2) (1978), Exp. No 11.

[5] E. Giner, *Espaces intégraux de type Orlicz, dualité, compacités, convergence en mesure, applications à l'optimisation*, Thèse de Doctorat és-mathématiques pures et appliquées, Univ. des Sci. et Techn. du Languedoc, Perpignan 1977.

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

[7] R. B. Holmes, *A course on optimization and best approximation*, Springer-Verlag, Berlin — New York 1972.

[8] A. Kozek, *Orlicz spaces of functions with values in Banach spaces*, Comment. Math. (Prace Mat.) 19 (1977), p. 259-288.

[9] — *On the theory of estimation with convex loss functions*, p. 177-202 in: Proc. of the Symp. to Honour Jerzy Neyman, PWN — Polish Scientific Publishers, Warszawa 1977.

[10] — *Convex integral functionals on Orlicz spaces*, Inst. of Math. Polish Acad. of Sci., Preprint No 89, May 1976, to appear in Comment. Math. (Prace Mat.).

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

[12] — *On two necessary σ-fields and on universal loss functions*, this fascicle, p. 29-47.

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

[14] 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.

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

[16] W. Orlicz, *Über eine gewisse Klasse von Räumen vom Typus B*, Bull. Acad. Polon. Sci. et Lettr., Sér. A (1932), p. 207-220.

[17] — *Über Räume* (L^M), ibidem (1936), p. 93-107.

[18] C. R. Rao, *Some theorems on minimum variance estimation*, Sankhyā 12 (1952), p. 27-42.

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

[20] D. H. Wagner, *Survey of measurable selection theorems*, SIAM J. Control Optimization 15 (1977), p. 859-903.

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

Received on 26. 4. 1979