JOURNAL OF MULTIVARIATE ANALYSIS **10**, 385–397 (1980)

An Extension of the Usual Model in Statistical Decision Theory with Applications to Stochastic Optimization Problems

E. J. BALDER

Mathematical Institute, University of Utrecht, Utrecht, the Netherlands

Communicated by G. Kallianpur

By employing fundamental results from "geometric" functional analysis and the theory of multifunctions we formulate a general model for (nonsequential) statistical decision theory, which extends Wald's classical model. From central results that hold for the model we derive a general theorem on the existence of admissible nonrandomized Bayes rules. The generality of our model makes it also possible to apply these results to some stochastic optimization problems. In an appendix we deal with the question of sufficiency reduction.

1. INTRODUCTION

In this paper we shall formulate a general model for (nonsequential) statistical decision theory, which extends the classical model introduced by Wald [19] and developed by LeCam [14, 15] and Brown [7]. To this end we employ fundamental results from "geometric" functional analysis and the theory of multifunctions that are now available. It is interesting to note that the geometrical aspects of the model, so visible in the extreme-point-role of the nonrandomized decision rules, do not figure in the standard descriptions of the theory [7, 11, 14, 15, 19], although they were used in [13]. Also, we should mention that in some respects we have not aimed for the level of generality attained in [14, 15], so as to remain in closer touch with the measure theoretical setting of the model and its statistical background. Finally, we remark that quite a number of the aspects of the model emphasized here can be extended to a sequential setup.

The organization of this paper is as follows. After the formulation of the model in Section 2 we give some applications in Sections 3 and 4. Our main

Received September 13, 1977; revised November 21, 1979.

AMS 1970 subject classification 62 C05.

Key words and phrases: Statistical decision theory, Bayes rule, measurable multifunctions, stochastic programming, sufficiency reduction.

0047-259X/80/030385–13$02.00/0

683/10/3-8

result for statistical decision theory concerns the existence of nonrandomized admissible Bayes rules, whereas in Section 4 some applications to other stochastic optimization problems are given (cf. [5, 18]). Finally, we deal with the question of reduction by sufficiency in an appendix to this paper.

2. A Model for Statistical Decision Theory

Let $(X, \mathscr{X})$ be a measurable space, the *sample* (or *observation*) *space*, and let $\mathscr{P}$ be a collection of probability measures on X, dominated by a σ-finite measure μ. The Radon–Nikodym derivative f_P of P with respect to μ can be considered as an element of $L^1(X)$, the usual L^1-space of functionals on $(X, \mathscr{X}, \mu)$. Let C denote a separable Banach lattice and M its dual space, equipped with the weak $\sigma(M, C)$-topology and the canonical (product) partial order. The norm and partial ordering relation on C will be denoted by $\| \|_C$ and $\leqslant_C$, respectively, and the duality between C and M is represented by the symbol $\langle \cdot, \cdot \rangle$ (cf. [9, Chap. V]). Let Γ be a measurable multifunction on X with nonempty closed convex extremal values in the positive part M_1^+ of the unit ball M_1 of M.[1] We refer to the text by Castaing and Valadier [8] for details on multifunctions; the reader not familiar with this notion may substitute for Γ the multifunction identically equal to M_1^+ without missing the essentials of this paper. Let us remember that a subset B_0 of a convex set B in a linear space is said to be *extremal* in B if for every $b, b' \in B$ we have that $\frac{1}{2}b + \frac{1}{2}b' \in B_0$ implies that $b, b' \in B_0$ [12, Chap. I]. In such a case we have trivially that ∂B_0, the extremal boundary (or collection of extreme points) of B_0, is contained in ∂B. An M-valued function s on X is said to be a measurable selector of Γ if s is X-measurable and $s(x) \in \Gamma(x)$ (unless mentioned otherwise, a topological space is understood to be equipped with its Borel σ-algebra). The collection of all measurable selectors of Γ is denoted by $\mathscr{S}_\Gamma$, and its quotient in $L^\infty(X; M)$ by S_Γ (cf. [8]; here $L^\infty(X; M)$ denotes the usual L^∞-space of M-valued functions on $(X, \mathscr{X}, \mu)$). The set S_Γ will be called the *class of* (*randomized*) *decision rules*, and its elements will be referred to as (randomized) decision rules. We shall motivate this nomenclature below, but wish here to draw attention to the fact that the introduction of the multifunction Γ will allow the collection of available decisions (or actions) to depend upon the observation. Consider the subset $\bar{\mathscr{S}}_\Gamma$ of $\mathscr{S}_\Gamma$ composed of those measurable selectors s of Γ for which $s(x) \in \partial M_1^+$ for every $x \in X$. Its quotient in $L^\infty(X, M)$ is denoted by $\bar{S}_\Gamma$ and will be called the class of *nonrandomized decision rules*.

Let $\{l_n\}$ denote an increasing sequence of C-valued functions on $\mathscr{P} \times X$, where for every $n \in \mathbb{N}$ l_n is such that the function $f_P\, l_n(P, \cdot)$ belongs to

[1] Note that M_1^+ is compact and metrizable [12, III.12].

$L^1(X; C)$ for every $P \in \mathscr{P}$ (here $L^1(X; C)$ denotes the usual L^1-space of C-valued functions on $(X, \mathscr{X}, \mu)$; in our notation we shall not strictly adhere to the distinction between equivalence classes and their representants). The sequence $\{l_n\}$ will be referred to as the *loss sequence.* We shall suppose also that there exists $k_1 \in C$ for which $l_1 \geqslant_C k_1$ on $\mathscr{P} \times X$, although it is possible to relax this assumption. The *risk* $R(P, \delta)$ of a decision rule $\delta \in S_\Gamma$ when sampling takes place under the probability measure $P \in \mathscr{P}$, is defined by

$$R(P, \delta) = \lim_n \uparrow \int \langle \delta(x), l_n(P, x) \rangle \, P(dx), \tag{2.1}$$

(where we use the duality symbol introduced above). Obviously, (2.1) is well defined since the assumptions cause the limit in the right-hand side to be monotone (possibly equal to $+\infty$). The $(-\infty, +\infty]$-valued functional $R(\cdot, \delta)$ on $\mathscr{P}$ is called the *risk function* of the decision rule δ.

If we agree to measure the "effectiveness" of a statistical decision rule only through its risk function, the measure space $(X, \mathscr{X}, \mu)$ can be supposed to be complete without loosing generality (see the Appendix; note that the original σ-algebra $\mathscr{X}$ is sufficient for $\mathscr{P}$ within its own μ-completion). This will be done from now on. Important specializations of the model described above are obtained by taking C to be a lattice of functionals on a topological space A. This space A is then called the action space, and the elements of the loss sequence $\{l_n\}$ can consequently be regarded as functionals on $\mathscr{P} \times X \times A$. Before furnishing some examples, let us agree to call a functional on $X \times A$ a *normal integrand* if ϕ is $\mathscr{X} \otimes \mathscr{A}$-measurable and $\phi(x, \cdot)$ is lower semicontinuous on A for every $x \in X$ (here $\mathscr{A}$ denotes the Borel σ-algebra on A; cf. [8, Chap. VII] for details on normal integrands).

EXAMPLE 1. Let A be a compact metric space (its metric will be indicated by ρ) and let $C \equiv \mathscr{C}(A)$, the separable Banach lattice of continuous functionals on A, equipped with the supremum norm. Then M represents the collection $M(A)$ of all (signed) Radon measures on A, equipped with the vague topology. Take Γ to be the multifunction on X which is identically equal to the closed convex set $M_1^+(A)$ of all Radon probability measures on A. Note that $\partial M_1^+(A)$ consists of all Dirac probability measures on A and actually this set is homeomorphic to A by [9, Proposition 12.9, Problem 25.1]. In this case the set $\mathscr{S}_\Gamma$ consists of all transition probabilities with respect to X and A (cf. [17, Chap. III]) and the set $\bar{\mathscr{S}}_\Gamma$ is homeomorphic to the collection of all measurable functions from X into A.

Concerning the loss sequence, we shall demonstrate that, in the present case, there exists a functional l on $\mathscr{P} \times X \times A$ such that for every $P \in \mathscr{P}$ $l(P, \cdot, \cdot)$ is a normal integrand and such that for every $P \in \mathscr{P}$, $\delta \in S_\Gamma$,

$$R(P, \delta) = \int P(dx) \int \delta(x; da) \, l(P, x, a). \tag{2.2}$$

Conversely, corresponding to every functional l on $\mathscr{P} \times X \times A$, bounded below by some $k_0 \in \mathscr{C}(A)$ and such that $l(P, \cdot, \cdot)$ is a normal integrand for every $P \in \mathscr{P}$, there exists a loss sequence $\{l_n\}$ for which (2.2)—and (2.1)—hold.

Let us begin by proving the first claim. For fixed $P \in \mathscr{P}$, $n \in \mathbb{N}$ the function $l_n(P, \cdot)$ into $\mathscr{C}(A)$ can also be regarded as a functional $(x, a) \to l_n(P, x)(a)$ on $X \times A$ that is measurable in its first and continuous in its second argument, whence $\mathscr{X} \otimes \mathscr{A}$-measurable. Thus, $l \equiv \lim_n \uparrow l_n$ satisfies the claims made about l, and (2.2) follows from (2.1) by virtue of the monotone convergence theorem. The second claim is proved as follows. Rather than subtracting k_0 from l we may suppose that l is nonnegative on $\mathscr{P} \times X \times A$. We define now, in analogy to the proof of a theorem by Baire, for $n \in \mathbb{N}$, $P \in \mathscr{P}$, $x \in X$, $a' \in A$, the nonnegative functional $\tilde{l}_n$ by

$$\tilde{l}_n(P, x, a') = \inf_{a \in A} [n\rho(a, a') + l(P, x, a)] \tag{2.3}$$

Fix $n \in \mathbb{N}$, $P \in \mathscr{P}$. Trivially, $\tilde{l}_n(P, x, \cdot)$ is a continuous functional for every $x \in X$. Also, for every $a' \in A$ we have that $\{x \in X \mid \tilde{l}_n(P, x, a') < \beta\} = \mathrm{proj}_X\{(x, a) \mid n\rho(a, a') + l(P, x, a) < \beta\}$, for every $\beta \in \mathbb{R}$, so by a well-known projection theorem $\tilde{l}_n(P, \cdot, a')$ is $\mathscr{X}$-measurable [8, Theorem III.23]. Also, as in the proof of Baire's theorem, one shows easily that lower semicontinuity of l implies that $l(P, x, a) = \lim_n \uparrow l_n(P, x, a)$ for every $P \in \mathscr{P}$, $x \in X$, $a \in A$ (cf. [4, Theorem 1] for an abstract setting of this result). Finally, we set $l_n = \inf(\tilde{l}_n, n)$, $n \in \mathbb{N}$, and conclude that $\{l_n\}$ can indeed be considered as a loss sequence for which (2.1) and (2.2) hold. Let us note that in establishing this correspondence the compactness of the space A has not really been used. Finally, it should be mentioned that Brown has shown, by a slightly more careful argument, that the completeness condition on $(X, \mathscr{X}, \mu)$ used in the argument here can be dispensed with [7, Theorem 3.10].

In the following example we show that the introduction of a multifunction can mean more than a mathematical embellishment of the theory.

EXAMPLE 2. Let A be a locally compact metric space which is countable at infinity. Let Δ be a measurable multifunction from X into A with nonempty compact values and define, for $x \in X$, the subset $\Gamma(x)$ of $M_1^+(A)$ by $\Gamma(x) \equiv \{\nu \mid \nu(\Delta(x)) = 1\}$. Then the multifunction is measurable by [8, Theorems III.30, IV.12] with nonempty closed convex extremal values. By means of an Alexandroff compactification (addition of a point ∞) A can be regarded as an open set in the compact metrizable space $\hat{A} \equiv A \cup \{\infty\}$ [6, Chap. 9, Sect. 2]. Of course, compact subsets of A are compact in $\hat{A}$. Since the inclusion ι of subsets of A into $\hat{A}$ is continuous, the multifunction $\iota \circ \Delta$ is measurable from X into $\hat{A}$ with nonempty compact values [8, Definition III.1] (instead of $\iota \circ \Delta$ we shall continue to write Δ without risking much confusion). Consequently, Γ can also be regarded as a measurable multifunction from X into the set $M_1^+(\hat{A})$ of Radon probability measures on $\hat{A}$. Clearly, as far as the class of decision rules is concerned, we have now regained the structure of Example 1. Note that the class $\mathscr{S}_\Gamma$ is homeomorphic to the class of all measurable selectors of the multifunction Δ. Given a nonnegative loss functional l on $\mathscr{P} \times X \times A$ which is such that $l(P, \cdot, \cdot)$ is a normal integrand for every $P \in \mathscr{P}$, we can extend l to $\mathscr{P} \times X \times \hat{A}$ as follows. Define for $P \in \mathscr{P}$, $x \in X$ $\hat{l}(P, x, a) \equiv l(P, x, a)$ if $a \in \Delta(x)$ and $\equiv +\infty$ elsewhere; since the graph $gph\Delta$ of the multifunction Δ is $\mathscr{X} \otimes \hat{\mathscr{A}}$-measurable by [8, Theorem III.30], $\hat{l}(P, \cdot, \cdot)$ is $\mathscr{X} \otimes \hat{\mathscr{A}}$-measurable (of course, $\hat{\mathscr{A}}$ indicates the Borel σ-algebra on $\hat{A}$). Also, the extension preserves the lower semicontinuity in the last argument, so the structure of Example 1 has been regained completely. In conclusion, we observe that it would have been possible to set out with l defined only on $\mathscr{P} \times gph\Delta$.

EXAMPLE 3. (cf. [7, 14, 19]). As Example 2, only suppose now that Δ is identically equal to A on X and that the loss functional l is such that for every $P \in \mathscr{P}$, $x \in X$ $\bar{l}(P, x, \cdot)$ tends to $+\infty$ (i.e., given any $\beta \in \mathbb{R}$ there exists a compact $K \subset A$ such that $l(P, x, a) > \beta$ for $a \notin K$). This time, define for $P \in \mathscr{P}$, $x \in X$, $\bar{l}(P, x, a) \equiv l(P, x, a)$ if $a \in A$ and $\equiv +\infty$ if $a = \infty$. This extension $\bar{l}$ of l is such that $\bar{l}(P, \cdot, \cdot)$ is a normal integrand on $X \times \hat{A}$ for every $P \in \mathscr{P}$. In addition, define Γ to be identically equal to $M_1^+(\hat{A})$. A slight disadvantage of this extension is, naturally, that $\mathscr{S}_\Gamma$ now also contains "extraordinary" decision rules δ for which for some $x \in X$ the support of $\delta(x)$ may contain ∞. Such rules would be statistically meaningless. This hardly matters since corresponding to every "extraordinary" decision rule δ there always exists an "ordinary" decision rule δ' such that $R(P, \delta') \leqslant R(P, \delta)$ for every $P \in \mathscr{P}$.

EXAMPLE 4. Consider the following estimation problem (cf. [19, Chap. I]). Let $\mathscr{P}$ be indexed by a set Θ of real parameters, $\mathscr{P} = \{P_\theta \mid \theta \in \Theta\}$, where P_θ stands for the normal distribution with mean θ and variance 1 on the real line X. In case we use the loss functional $l(\theta, x, a) \equiv (\theta - a)^2$, where the action variable a is an element of the real line A, we are in the setting of Example 3. On the other hand, if we adopt the loss function $l'(\theta, x, a) \equiv \min(10^3, (\theta - a)^2)$ (in a situation where errors beyond a certain magnitude do not matter anymore) and if we also define the multifunction Δ by $\Delta(x) \equiv [x - 10^6, x + 10^6]$, $x \in X$, to restrict our possible estimates after the observation, we are in the setting of Example 2.

Remark 1. Proceeding in a slightly different way, the embedding procedure of Example 2 can also be performed in case the space A is only metric and separable. In this approach one compactifies within the Hilbert cube. Details are left to the reader (cf. [10, Théorèmes III.58, 59]). For an even more general approach—although formulated for optimal control problems—we refer to [20, 21].

Remark 2. Although, as we have argued, completeness of $(X, \mathscr{X}, \mu)$ can always be presupposed w.l.o.g., it should be pointed out that all results of this paper can also be proven without using the completeness assumption, cf. [7, 13, 21].

3. Decision Theoretical Results

In the following we shall demonstrate how some of the main results of statistical decision theory follow from known results in geometric functional analysis and the theory of multifunctions. To begin with, let us equip the space $L^\infty(X; M)$ with the weak $\sigma(L^\infty(X; M), L^1(X; C))$-topology (we note that this makes $L^\infty(X; M)$ into a Hausdorff locally convex space). All topological statements regarding subsets of $L^\infty(X; M)$ will be made with respect to this topology. As an immediate consequence of the topologization we have the following (cf. [14, Theorem 2], [19, Theorem 3.1]).

THEOREM 1. (i) *The class S_Γ of randomized decision rules is a nonempty compact convex set in $L^\infty(X; M)$.* (ii) *For every $P \in \mathscr{P}$ the functional $R(P, \cdot)$ is lower semicontinuous and affine on S_Γ.* (iii) *If the σ-algebra $\mathscr{X}$ is countably generated, S_Γ is metrizable.*

Proof. (i) Follows immediately from [8, Theorem V.1]. (ii) Follows directly from (2.1) and the definition of the weak topology on $L^\infty(X; M)$. (iii) Follows from the definition of the topology on $L^\infty(X; M)$ and the separability of the space $L^1(X; C)$, which is a consequence of the additional assumption [12, III.12F].

Let us agree to say that a subset F of S_Γ has property D if $\delta \in F$, $\delta' \in S_\Gamma$ and $R(\cdot, \delta') \leqslant R(\cdot, \delta)$ imply $\delta' \in F$ (here $\leqslant$ denotes the canonical product ordering for functionals on $\mathscr{P}$). In statistical decision theory this property is not unfamiliar; a decision rule $\bar{\delta} \in S_\Gamma$ is said to be *admissible* if the set $\{\delta \in S_\Gamma | R(\cdot, \delta) = R(\cdot, \bar{\delta})\}$ has property D. The following consequence of Theorem 1 is, in essence, Wald's minimal complete class result. We include its proof, because it establishes a certain line of argument that will be continued below.

COROLLARY 2. *Corresponding to every decision rule $\delta \in S_\Gamma$ there exists an admissible decision rule $\bar{\delta}$ such that $R(\cdot, \bar{\delta}) \leqslant R(\cdot, \delta)$.*

Proof. Consider the collection $\mathscr{F}_0$ of all nonempty closed subsets of S_Γ

that have property D. Equip $\mathscr{F}_0$ with the partial order by inclusion. From Theorem 1 it follows immediately, by the finite intersection property, that $\mathscr{F}_0$ is inductive. Let $\bar{\delta} \in S_\Gamma$; evidently, the set $\bar{F}$ consisting of those $\delta \in S_\Gamma$ for which $R(\cdot, \delta) \leqslant R(\cdot, \bar{\delta})$ has property D. Also, by Theorem 1 (ii), $\bar{F}$ is closed. By Zorn's lemma $\bar{F}$ contains a minimal element F_0 of $\mathscr{F}_0$. Consider two arbitrary elements δ', δ'' of F_0 and suppose that for some $P_0 \in \mathscr{P}$ $R(P_0, \delta') < R(P_0, \delta'')$. Define $\alpha \equiv \inf(R(P_0, \delta) | \delta \in F_0)$ and $F_1 \equiv \{\delta \in F_0 | R(P_0, \delta) = \alpha\}$. By Theorem 1 F_1 is nonempty and closed; also, it has property D and is strictly contained in F_0. Hence we have a contradiction and conclude that for all δ', $\delta'' \in F_0$ $R(\cdot, \delta') = R(\cdot, \delta'')$. This means that every $\bar{\delta} \in F_0$ is admissible and the proof is finished.

Other classical results of statistical decision theory, such as the completeness of the class of Bayes rules in the wide sense and the existence of Bayes rules, follow equally directly from Theorem 1. In constrast, the following important result is not represented in the standard literature on the subject.

THEOREM 3. *The class* $\bar{S}_\Gamma$ *of nonrandomized decision rules is the extremal boundary* ∂S_Γ *of the class* S_Γ *of randomized decision rules.*

Proof. By definition of Γ the set $\Gamma(x)$ is extremal in M_1^+ for every $x \in X$. Thus $\partial\Gamma(x)$ is contained in ∂M_1^+ for every $x \in X$, and this implies trivially that $\partial\Gamma(x) = \Gamma(x) \cap \partial M_1^+$ for every $x \in X$. Denote by $\mathscr{S}_{\partial\Gamma}$ the collection of those $\delta \in \mathscr{S}_\Gamma$ for which $\delta(x) \in \partial\Gamma(x)$ for every $x \in X$ and denote its quotient in $L^\infty(X; M)$ by $S_{\partial\Gamma}$. By the above $\bar{\mathscr{S}}_\Gamma = \mathscr{S}_{\partial\Gamma}$ and by [8, Theorem IV.15] $\partial S_\Gamma = S_{\partial\Gamma}$, so the proof is finished.

Let us now consider an application of Theorem 3. To this end we shall introduce a few more assumptions that set the stage for some Bayesian considerations. Suppose that the σ-algebra $\mathscr{X}$ is countably generated and that the set $\mathscr{P}$ of probability measures has been equipped with a σ-algebra Σ such that $P \to P(X_0)$ is Σ-measurable for every $X_0 \in \mathscr{X}$. Also, suppose that the elements of the loss sequence $\{l_n\}$ are $\Sigma \otimes \mathscr{X}$-measurable. By [16, VIII.10] there exists a $\Sigma \otimes \mathscr{X}$-measurable version of the functional $(P, x) \mapsto f_P(x)$. A positive measure on $\mathscr{P}$ will be called a *prior* (*measure*). For every prior τ the *Bayes risk* (τ, δ) of the decision rule $\delta \in S_\Gamma$ is defined by

$$r(\tau, \delta) = \int R(P, \delta)\, \tau(dP),$$

and in view of our assumptions this definition makes sense. A decision rule $\delta \in S_\Gamma$ is said to be a *Bayes rule with respect to the prior* τ if $r(\tau, \delta) \leqslant r(\tau, \delta')$ for every $\delta' \in S_\Gamma$. The following result generalizes [19, Theorem 3.5].

THEOREM 4. *Corresponding to every prior τ there exists a nonrandomized admissible Bayes rule.*

Proof. Let τ be an arbitrary prior. Consider the collection $\mathscr{F}_1$ of all nonempty closed extremal subsets F of S_Γ that have property D, equipped with the partial order by inclusion. It is easy to check that $\mathscr{F}_1$ is inductive (cf. the proof of Theorem 2). Since the collection of all Bayes rules with respect to τ belongs to $\mathscr{F}_1$ (note that, in view of Theorem 1 (iii) and Fatou's lemma, the functional $r(\tau, \cdot)$ is lower semicontinuous on S_Γ), it must contain a minimal element F_0 of $\mathscr{F}_1$. Completely analogous to the proof of Theorem 2 we show that F_0 is such that for some $\delta' \in S_\Gamma$ $F_0 = \{\delta \in S_\Gamma | R(\cdot, \delta) = R(\cdot, \delta')\}$. By [12, Lemma II.13] F_0 contains an extreme point, which, by extremality of F_0, is also an extreme point of S_Γ. Since all elements of F_0 are clearly admissible, the proof is finished by Theorem 3.

Other applications of Theorem 3 to the classical decision theoretical framework will follow elsewhere. The remainder of this paper is devoted to the application of the above results to some stochastic optimization problems.

4. SOME APPLICATIONS TO STOCHASTIC OPTIMIZATION PROBLEMS

A special case of the following optimization problem was studied in [5]. We consider a measurable space $(\Omega, \mathscr{F})$ and a measurable stochastic process (y_t), $0 \leqslant t \leqslant 1$ defined on it. We shall equip the time interval with its Borel σ-algebra $\mathbb{B}[0, 1]$ and the Lebesgue measure λ. Let $(\mathscr{F}_t)_t$ be an increasing family of sub-σ-algebras of $\mathscr{F}$ such that the process $(y_t)_t$ is adapted to it (cf. [10]). Let $\mathscr{Q}$ be a collection of probability measures on $(\Omega, \mathscr{F})$, dominated by a σ-finite measure ν. Let us suppose that the process $(y_t)_t$ is governed by an unknown element of $\mathscr{Q}$. Let A be a compact metric space of control points, equipped with its Borel σ-algebra $\mathscr{A}$, and let c be a nonnegative instantaneous cost functional defined on $\mathscr{Q} \times [0, 1] \times \Omega \times A$ which is such that $((t, \omega), a) \mapsto c(Q, t, \omega, a)$ is a normal integrand on $([0, 1] \times \Omega) \times A$ for every $Q \in \mathscr{Q}$.

A *nonanticipative strategy* is defined to be a measurable function from $[0, 1] \times \Omega$ into A, where $[0, 1] \times \Omega$ is equipped with the adapted sub-σ-algebra $\mathscr{B}$ of $\mathbb{B}[0, 1] \times \mathscr{F}$, consisting of those B whose section B_t belongs to $\mathscr{F}_t$ for every $t \in [0, 1]$. The collection of all nonanticipative strategies is denoted by Π. The expected cost $E(Q, \pi)$ of operating with the nonanticipative strategy $\pi \in \Pi$ under the probability measure $Q \in \mathscr{Q}$ is defined by

$$E(Q, \pi) = \int Q(d\omega) \int \lambda(dt)\, c(Q, t, \omega, \pi(t, \omega)). \tag{4.1}$$

A strategy $\bar{\pi} \in \Pi$ is said to be *admissible* if for every $\pi \in \Pi$ $E(\cdot, \pi) \leqslant E(\cdot, \bar{\pi})$ implies $E(\cdot, \pi) = E(\cdot, \bar{\pi})$. Of course, in case the set $\mathscr{Q}$ is a singleton, the notion of admissibility contracts to that of optimality used in [5]. The following result generalizes the main result of [5] in several ways (and makes the argument in [5] regarding conditional expectation more precise).

THEOREM 5. *There exists an admissible nonanticipative strategy.*

Proof. As was demonstrated in Example 1, there exists an increasing sequence $\{c_n\}$, converging pointwise to c and such that $c_n(Q, \cdot)$ is (a representant of) an element in $L^1(X; C(A))$ for every $n \in \mathbb{N}$, $Q \in \mathscr{Q}$. Here we set $X = [0, 1] \times \Omega$. For every $n \in \mathbb{N}$, $Q \in \mathscr{Q}$ the conditional expectation $\bar{c}_n(Q, \cdot)$ of $c_n(Q, \cdot)$ with respect to $\mathscr{B}$ and $\lambda \times Q$ is well defined as an element of the L^1-space of $\mathscr{C}(A)$-valued functions on $(X, \mathscr{B}, \lambda \times v)$ (e.g., [8, Proposition VIII.32]). From the definition of conditional expectation it follows trivially that for every $Q \in \mathscr{Q}$ and every transition probability δ with respect to $(X, \mathscr{B})$ and $(A, \mathscr{A})$

$$E(Q, \delta) = \lim_n \uparrow \int Q(d\omega) \int \lambda(dt) \langle \bar{c}_n(Q, t, \omega), \delta(t, \omega) \rangle,$$

where the left-hand side represents the trivial extension of (4.1) to transition probabilities ("randomized nonanticipative strategies"), and $\langle .,. \rangle$ the duality between $\mathscr{C}(A)$ and $M_1^+(A)$ (cf. Example 1). It will be clear that the result follows now from an application of Theorem 4 with $\mathscr{P} = \lambda \times \mathscr{Q}$, Γ identically equal to $M_1^+(A)$ and τ equal to the zero measure on $\mathscr{P}$ (i.e., $\imath(\tau, \cdot)$ identically equal to zero). Note that here, as in Section 1, the σ-algebra $\mathscr{B}$ can be supposed complete with respect to $\lambda \times v$ without loss of generality.

To obtain another application we turn to a standard problem studied in the area of stochastic programming with recourse. We shall give a slight extension of the main result in [18] and derive this in a different and more direct fashion. We work within a setting which is a particular case of Example 2. Namely, let $(X, \mathscr{X}, P)$ be a complete probability space, A a compact set in a metrizable locally convex vector space and Δ a measurable multifunction from X into A with nonempty closed convex values. In terms of Example 2, $\hat{A}$ can be set equal to A and the definition of the multifunction Γ is obvious. Thus, the collection $\mathscr{S}_\Gamma$ consists of those transition probabilities δ from $(X, \mathscr{X})$ into $(A, \mathscr{A})$ for which $\delta(x; \Delta(x)) = 1$ for every $x \in X$. (To connect our set-up with that of [18], note that one can take $\mathscr{X}$ to be a suitable nonanticipative σ-algebra in the two-stage setting of [18].) For $i = 0, \ldots, n$, let l_i be a normal integral on the graph of Δ (cf. Example 2), which is convex in its second argument. Also, we shall assume that l_0 is bounded below and that $l_1,\dots, l_n$ are bounded on $gph\,\Delta$. The *primal optimization problem* (01) is defined by

$$(01)\colon \inf\left(\int P(dx) \int \delta(x; da)\, l_0(x, a) \mid \delta \in G\right),$$

where the set G is defined to be the quotient of those nonrandomized $\delta \in \mathcal{S}_\Gamma$ that satisfy

$$\int \delta(x; da)\, l_i(x, a) \leqslant 0 \; P\text{-a.e.} \qquad \text{for} \quad i = 1,\dots, n. \tag{4.2}$$

For notational convenience we define, for η in $L^1(X)$, $\delta \in S_\Gamma$

$$q_i(\eta, \delta) = \int P(dx)\, \eta(x) \int \delta(x; da)\, l_i(x, a).$$

Define $Y \equiv (L^1_+(X))^n$ and let 1_X denote the characteristic function of the set X. We define for $y = (y_1,\dots, y_n) \in Y$ and $\delta \in S_\Gamma$

$$q(y, \delta) = q_0(1_X, \delta) + \sum_{i=1}^{n} q_i(y_i, \delta).$$

The *dual objective functional* g is defined by $g(y) = \inf(q(y, \delta) \mid \delta \in S_\Gamma)$, $y \in Y$ and the *dual optimization problem* (02) by

$$(02)\colon \sup(g(y) \mid y \in Y)$$

THEOREM 6. *The primal optimization problem* (01) *has an optimal solution and the following duality relation holds*:

$$\min(01) = \sup(02).$$

Proof. Denote by $\tilde{G}$ the quotient of those $\delta \in \mathcal{S}_\Gamma$ that satisfy (4.2) and define the optimization problem ($\tilde{0}1$) by replacing G with $\tilde{G}$ in the definition of (01). For $\delta \in S_\Gamma$ we have that $\sup(\sum_{i=1}^{n} q_i(y_i, \delta) \mid y \in Y) = 0$ if $\delta \in \tilde{G}$, $= +\infty$ otherwise. By Theorem 1 and Examples 1, 2 the functional figuring in the above statement is lower semicontinuous on S_Γ, so we conclude that $\tilde{G}$ is a compact subset of S_Γ. Therefore $\inf(\tilde{0}1)$ is attained. By the same results, for every $y \in Y$ the functional $q(y, \cdot)$ is lower semicontinuous and affine on the compact convex set S_Γ. Also, the functional $q(\cdot, \delta)$ is easily checked to be L^1-continuous and affine an Y for every $\delta \in S_\Gamma$. By a well-known minimax theorem [2, Théorème 1.4] it follows that $\min(\tilde{0}1) = \sup(02)$. Since $\tilde{G}$ contains G, the proof will clearly be finished it for every $\tilde{\delta} \in \tilde{G}$ there exists $\delta \in G$ such that $q_0(1_X, \delta) \leqslant q_0(1_X, \tilde{\delta})$. This will now be shown to follow from the convexity assumptions on l_i and Δ(cf. [11, 2.8]). Let $\tilde{\delta} \in \mathscr{S}_\Gamma$ be a representant of an element in $\tilde{G}$. By [1, Sect. 2] there exists for every $x \in X$ an element $r(x) \in \Delta(x)$ which is the barycenter of the probability measure $\tilde{\delta}(x)$, and by [1, Proposition I.2.2] we have for every $x \in X$, $i = 0, 1, \ldots, n$ that $\int \tilde{\delta}(x; da)\ l_i(x, a) \geqslant l_i(x, r(x))$. Moreover the selection $x \mapsto r(x)$ of Δ is measurable since it is the composition of $x \mapsto \tilde{\delta}(x)$—measurable from X into $M_1^+(A)$ by definition—and $\tilde{\delta}(x) \to r(x)$, which is continuous from $M_1^+(A)$ into A by [9, Proposition 26.3]. It is now trivial to finish the proof.

APPENDIX: SUFFICIENCY REDUCTION

For the model of Section 2 we shall now formulate a general result on the reduction by means of sufficiency; cf. [3, 11]. Let $\mathscr{Y}$ denote a sub-σ-algebra of $\mathscr{X}$ which is sufficient for $\mathscr{P}$, i.e., for every $B \in \mathscr{X}$ there exists an $\mathscr{Y}$-measurable function ϕ_B such that for every $P \in \mathscr{P}$ ϕ_B is a version of the conditional expectation of the characteristic function 1_B with respect to $\mathscr{Y}$ and P [3]. Denote by $S_\Gamma(\mathscr{Y})$ the quotient of the collection of all $\mathscr{Y}$-measurable selectors of Γ. We have the following result.

THEOREM. *Suppose that the multifunction Γ is $\mathscr{Y}$-measurable and that every element l_n, $n \in \mathbb{N}$, of the loss sequence is such that $l_n(P, \cdot)$ is $\mathscr{Y}$-measurable for every $P \in \mathscr{P}$. Then there corresponds to every decision rule $\delta \in S_\Gamma$ a decision rule $\tilde{\delta} \in S_\Gamma(\mathscr{Y})$ such that $R(\cdot, \tilde{\delta}) = R(\cdot, \delta)$.*

Proof. Observe that the sufficiency notion can easily be implemented in the proof of [8, Proposition VIII.32] to yield that, corresponding to the given δ(which is, a fortiori, an element of $L^1(X; M)$, the L^1-space of M-valued functions on $(X, \mathscr{X}, \mu)$), there exists $\tilde{\delta} \in L^1(X; M)$ which has a $\mathscr{Y}$-measurable representant and is such that

$$\text{for every } P \in \mathscr{P}, \qquad B \in \mathscr{Y} \int_B \tilde{\delta}\, dP = \int_B \delta\, dP. \tag{A1}$$

Let us show that $\tilde{\delta} \in S_\Gamma$. By [8, Theorem III.15] we have that for every $k \in C$ the functional $x \to x^*(k \mid \Gamma(x)) \equiv \sup(\langle k, m \rangle \mid m \in \Gamma(x))$ is $\mathscr{Y}$-measurable. Let $\{k_n\}$ denote a dense sequence in C and let $\bar{\mu}$ denote a finite measure, equivalent to μ, which is an "infinite convex combination" of elements in $\mathscr{P}$ [17, Problem IV.1.3]. By (A1) we have that for every $n \in \mathbb{N}$, $B \in \mathscr{Y}$,

$$\int_B \langle k_n, \delta(x) \rangle\, \bar{\mu}(dx) = \int_B \langle k_n, \tilde{\delta}(x) \rangle\, \bar{\mu}(dx)$$

$$\leqslant \int_B x^*(k_n \mid \Gamma(x))\, \bar{\mu}(dx).$$

So we may conclude that for μ-almost every $x \in X$ it is true that

$$\langle k, \tilde{\delta}(x) \rangle \leqslant s^*(k \mid \Gamma(x)) \qquad \text{for every } k \in C. \tag{A2}$$

Let x be an element of X for which (A2) holds. If $\tilde{\delta}(x) \notin \Gamma(x)$ it would be possible to separate $\{\tilde{\delta}(x)\}$ and $\Gamma(x)$ strictly, according to the Hahn–Banach theorem. This is immediately seen to contradict (A2). We conclude that $\tilde{\delta} \in S_\Gamma(\mathscr{Y})$. The equality of the risk functions follows from using (A1) in a simple approximation argument.

Note added in proof. The Theorem in the Appendix seems only to be valid under the additional hypothesis that M_1^+ be separable for the dual norm topology. Otherwise, one cannot conclude that the weakly measurable elements of S_Γ are also strongly measurable—i.e., belong to $L^1(X; M)$—as is required in [8, Proposition VIII.32]. The additional hypothesis places the Theorem outside the reach of most applications in statistical decision theory. Nevertheless, this has no consequences for the remainder of the paper, as appears from Remark 2.

Acknowledgments

Inspiring discussions with Professor L. M. LeCam have had a great influence upon the efforts of the author. The author is further indebted to Dr. A. Verbeek and a referee for some helpful comments.

References

[1] Alfsen, E. M. (1971). *Compact Convex Sets and Boundary Integrals*. Springer–Verlag, Berlin.

[2] Auslender, A. (1972). *Problèmes de Minimex via l'Analyse Convexe et les Inégalités Variationelles*. Springer Lecture Notes in Economics and Mathematical Systems No. 77. Springer-Verlag, Berlin.

[3] Bahadur, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.

[4] Balder, E. J. (1977). An extension of duality-stability relations to nonconvex optimization problems. *SIAM J. Control Optimization* **15** 329–343.

[5] Benes, V. E. (1970). Existence of optimal strategies based on specified information, for a class of stochastic decision problems. *SIAM J. Control* **8** 179–188.

[6] Bourbaki, N. (1958). *Éléments de Mathématique, Livre III*. Hermann, Paris.

[7] Brown, L. D. (1977). Closure theorems for sequential-design processes. In *Statistical Decision Theory and Related Topics-II* (S. S. Gupta, Ed.), pp. 57–91. Academic Press, New York.

[8] Castaing, C., and Valadier, M. (1977). *Convex Analysis and Measurable Multifunctions*. Springer Lecture Notes in Mathematics No. 580. Springer-Verlag, Berlin.

[9] Choquet, G. (1969). *Lectures on Analysis*. Benjamin, Reading, Mass.

[10] Dellacherie, C., and Meyer, P. A. (1975). *Probabilités et Potentiel*. Hermann, Paris.

[11] Ferguson, T. S. (1967). *Mathematical Statistics, a Decision Theoretic Approach*. Academic Press, New York.

[12] HOLMES, R. B. (1975). *Geometric Functional Analysis and its Applications*. Springer-Verlag, Berlin.

[13] KIRSCHNER, H. P. (1976). On the risk-equivalence of two methods of randomization in statistics. *J. Multivar. Anal.* **6** 159–166.

[14] LECAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.

[15] LECAM, L. (1974). *Notes on Asymptotic Methods in Statistical Decision Theory*. Centre de Recherches Mathématiques, Université de Montréal.

[16] MEYER, P. A. (1966). *Probabilités et Potentiel*. Hermann, Paris.

[17] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden–Day, San Fransisco.

[18] ROCKAFELLAR, R. T., AND WETS, R. J.-B. (1976). Stochastic convex programming: basic duality. *Pacific J. Math.* **62** 173–195.

[19] WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.

[20] BALDER, E. J. (1979). On a useful compactification for optimal control problems. *J. Math. Anal. Appl.* **72** 391–398.

[21] BALDER, E. J. (1979). Relaxed inf-compactness for variational problems by Hilbert cube compactification. *J. Math. Anal. Appl.*, in press.