

TOWARDS A CALCULUS FOR ADMISSIBILITY

BY ANDRZEJ KOZEK¹

Polish Academy of Sciences and University of Wrocław

It is shown how the calculus can be used to characterize admissible decision rules (Pareto optimal points, efficient points). Necessary and sufficient conditions for admissibility are derived in terms of the first and the second directional derivatives of convex risk functions. In particular, the results obtained imply that if p is to be estimated in the binomial distribution $B(n, p)$, then an estimator is admissible for the quadratic loss function if and only if it fulfills some analytic conditions.

1. Introduction. In Statistical Decision Theory, risk functions characterize the quality of decision rules. A risk function can be interpreted as a collection of criterion functions. Similarly in economics, and in optimal control theory, the quality of procedures or actions is judged based on values of several criterion functions. In any case we are concerned with a function R of two arguments $z \in Z$, and $s \in S$. The set Z stands for the procedures at our disposal, and S for the set of indices of the criterion functions.

If $z_0 \in Z$, and there exists no $z_1 \in Z$ such that

$$(1.1) \quad R(z_1, s) \leq R(z_0, s)$$

holds for every $s \in S$, with strict inequality for at least one $s \in S$, then z_0 is called admissible (in statistics), or Pareto optimal or efficient (in economics).

In a more general formulation, R may be considered as a mapping from Z into a partially ordered space. If S consists of one point only, then the calculus is useful in finding the optimal z_0 . Hence a calculus has been developed for functions with values in partially ordered spaces. Athans and Geering (1973), Aubin (1971), Kutateladze (1977), Neustadt (1969), Smale (1973), Thibault (1980) and Zowe (1974) deal with this development; however we refer the reader to Achilles et al., (1979) for a more complete list of references. This approach has not received attention in mathematical statistics; we know of no paper interconnecting decision or estimation theory with this "multicriterial calculus". Unfortunately, the existing theory is weighted down by various technical assumptions which make it useless in the simplest nontrivial problems of statistics. On the other hand the "heuristic" approach to admissibility in Brown (1979) suggests that a development of a calculus for admissibility may be both useful and possible.

Here we do the first step in this direction. We derive analytic characterizations of admissible decision rules when S is (1) finite, (2) compact, (3) an arbitrary space. If S is compact, the characterization of admissible decision rules is given in terms of derivatives of the first order; in the general case the use of second derivatives is necessary. This exhibits a deep discrepancy when compared with S consisting of one point only, i.e. with the case of real convex functions. These characterizations provide an analytic counterpart to the well-known characterizations of admissibility by geometrical methods (separation of convex sets), i.e. by the Bayesian approach of Wald (1950), Arrow et al. (1953), Farrell (1968), LeCam (1974).

In Section 5 we show how the obtained results imply analytic conditions characterizing admissible estimators of p in the binomial distribution $B(n, p)$.

Received April 1981; revised February 1982.

¹ The final version of the paper was written at the University of Cologne while the author was on leave from the Polish Academy of Sciences and was sponsored by a Grant of the Alexander von Humboldt Foundation.

AMS 1970 subject classifications. Primary 62C15; secondary 90A05, 49B99.

Key words and phrases. Admissibility, binomial distribution, convex transformation, efficiency, Pareto optimality.

For the sake of simplicity we consider the case where Z is a vector space, perhaps infinite dimensional. We note here only that if Z is a convex subset of a vector space and $z_0 \in Z$, then the results stated in Sections 2 and 4 remain valid whereas in Section 3 the cone $K(z_0) = \{z \in Z: \text{there exists } \lambda > 0 \text{ such that } z_0 + \lambda z \in Z\}$ consisting of all available directions from z_0 should be considered. If z_0 is an interior point of Z , then $K(z_0)$ is the whole vector space.

2. Remarks on a calculus for convex transformations and for admissibility. Let R stand for a convex transformation from a vector space Z into an ordered vector space Y endowed with the positive convex cone C^+ . Define the subdifferential $\partial R(z_0)$ of R at the point $z_0 \in Z$ as the collection of all linear transformations $T_{z_0}: Z \rightarrow Y$ (continuous if Z and Y are topological spaces, a detail, however, of minor importance at the moment) such that

$$R(z) \geq R(z_0) + T_{z_0}(z - z_0)$$

holds for each $z \in Z$. The elements of $\partial R(z_0)$ are called subgradients of R at the point z_0 . In this section we assume that $\partial R(z) \neq \emptyset$ for each $z \in Z$. Moreover, we write $R(z_1) > R(z_2)$ or $R(z_1) \geq R(z_2)$ accordingly as $R(z_1) - R(z_2) \in C^+ \setminus \{0\}$ or $R(z_1) - R(z_2) \in C^+$. With this notation, a point $z_0 \in Z$ is called admissible if there is no $z_1 \in Z$ such that $R(z_0) > R(z_1)$. If z_0 is not admissible, then it is called inadmissible.

PROPOSITION 1. *If $z_1 \neq z_0$ and if there exists $T_{z_1} \in \partial R(z_1)$ such that*

$$(2.1) \quad T_{z_1}(z_0 - z_1) > 0,$$

then $R(z_0) > R(z_1)$ holds, i.e. z_0 is inadmissible.

PROOF. By the definition of the subdifferential we have

$$R(z_0) - R(z_1) \geq T_{z_1}(z_0 - z_1) > 0$$

and hence $R(z_0) > R(z_1)$. \square

Note that if in (2.1) we change “ $T_{z_1}(z_0 - z_1) > 0$ ” to “ $T_{z_1}(z_0 - z_1) \geq 0$ ”, then we get $R(z_0) \geq R(z_1)$. But the last inequality does not imply the inadmissibility of z_0 . Now, let us consider a proposition converse to Proposition 1.

PROPOSITION 2. *Suppose $z_\alpha = \alpha z_0 + (1 - \alpha)z_1$, with $\alpha \in (0, 1)$, and*

$$(2.2) \quad R(z_0) > R(z_\alpha) > R(z_1).$$

Then for each $T_{z_\alpha} \in \partial R_{z_\alpha}$, $T_{z_\alpha}(z_0 - z_\alpha) > 0$.

PROOF. By the definition of the subdifferential we have

$$R(z_1) \geq R(z_\alpha) + T_{z_\alpha}(z_1 - z_\alpha)$$

and hence

$$0 > R(z_1) - R(z_\alpha) \geq T_{z_\alpha}(z_1 - z_\alpha) = -\frac{\alpha}{1 - \alpha} T_{z_\alpha}(z_0 - z_\alpha).$$

Thus, $T_{z_\alpha}(z_0 - z_\alpha) > 0$. \square

Now we give an example where z_0 is not admissible and the assumptions of Proposition 1 are not satisfied. In this case there does not exist $\alpha \in (0, 1)$ such that $R(z_\alpha)$ satisfies Assumption (2.2) of Proposition 2.

EXAMPLE 1. Assume that $R: R^1 \rightarrow \ell^1(m)$, where $\ell^1(m)$ is the space of m -summable sequences with weights $m(i) = 1/i^3, i = 1, 2, \dots$. Let $\ell^1(m)$ be endowed with the natural ordering and $R(z) = \{c_i(z), i = 1, 2, \dots\}$, where

$$c_i(z) = z - 1 \quad \text{if } z \geq \frac{1}{i}$$

$$c_i(z) = -(2i - 1)z + 1, \quad \text{if } z \leq \frac{1}{i}.$$

Clearly, R is convex. If $z_0 = 0, z_1 = 1$, then $R(1) = R(z_1) = (0, 0, \dots) < R(0) = R(z_0) = (1, 1, \dots)$ and $\partial R(z_1) = \{T: R^1 \rightarrow \ell^1(m), T(z) = (\alpha z, +1, +1, \dots), \alpha \in [-1, 1]\}$. Thus, for any $T_{z_1} \in \partial R(z_1)$ we have $T_{z_1}(z_0 - z_1) = (-\alpha, -1, -1, \dots)$ and hence $T_{z_1}(z_0 - z_1) \notin C^+ \setminus \{0\}$. Moreover, if $\alpha \in (0, 1)$, then $R(z_\alpha) < R(z_0)$ but $R(z_1)$ is not comparable with $R(z_\alpha)$ and $T_{z_\alpha}(z_0 - z_\alpha) \notin C^+ \setminus \{0\}$.

Example 1 can also be modified to the case where Y is a space of continuous functions on a compact extremally disconnected topological space.

PROPOSITION 3. If $Y = R^n, C^+ = \{y = (y_1, \dots, y_n) : y_i \geq 0, i = 1, 2, \dots, n\}$, and $R(z_0) > R(z')$, then there exist z_1 and $\alpha \in (0, 1)$, such that (2.2) is valid.

PROOF. Since $R(z_0) = (R_1(z_0), \dots, R_n(z_0)) > (R_1(z'), \dots, R_n(z')) = R(z')$, there exists i such that $R_i(z_0) > R_i(z')$. If $f_i(\beta)$ is given by

$$f_i(\beta) = R_i(z' + \beta(z_0 - z')),$$

then by the convexity of f_i there exists $\beta_i \in (0, 1)$ such that f_i is increasing on $[\beta_i, 1]$. For $i' \neq i$ the convexity of $f_{i'}$ implies that there exist $\beta_{i'} \in (0, 1)$ such that $f_{i'}$ is nondecreasing on $[\beta_{i'}, 1]$. Let $\beta_0 = \max(\beta_1, \dots, \beta_n), z_1 = z' + \beta_0(z_0 - z')$. For every $\alpha \in (0, 1)$ we get

$$R(z_0) > R(z_\alpha) > R(z_1). \quad \square$$

Propositions 1-3 imply immediately the following necessary and sufficient conditions for admissibility.

THEOREM 1. If R is convex, $Y = R^n, C^+ = \{(y_1, \dots, y_n) : y_i \geq 0\}$, then z_0 is admissible if and only if there are no $z_1 \in Z$ and $T_{z_1} \in \partial R(z_1)$ such that (2.1) holds. \square

The following simple proposition shows that another statement similar to the theorem converse to Proposition 1 is valid.

PROPOSITION 4. If $R(z_1) < R(z_0)$, then for each T_{z_0} from $\partial R(z_0)$ we have

$$(2.3) \quad T_{z_0}(z_1 - z_0) < 0.$$

PROOF. By the definition of the subgradient $T_{z_0} \in \partial R(z_0)$ we get

$$0 > R(z_1) - R(z_0) \geq T_{z_0}(z_1 - z_0). \quad \square$$

However the theorem converse to Proposition 4 is not valid, either. This is shown in Example 2.

EXAMPLE 2. Let $R: R^1 \rightarrow R^2$. If R^2 is endowed with the positive cone $C^+ = \{(y_1, y_2) : y_1 \geq 0, y_2 \geq 0\}$ and $R(z) = ((z + 1)^2, (z - 1)^2)$, then R is convex, and $\partial R(z_0)$ consists of the only linear transformation from R^1 into R^2 given by $T_{z_0}(z) = (2z(z_0 + 1), 2z(z_0 - 1))$. Thus, if $z_0 = -1$ and $z > -1$, then $T_{z_0}(z + 1) = (0, -4(z + 1)) \in -C^+ \setminus \{0\}$. But $z_0 = -1$ is admissible, and there exists no z such that $R(z) < R(-1)$. Similarly, if $z_0 = 1$ and $z < 1$,

then $T_{z_0}(z - 1) = (4(z - 1), 0) \in -C^+ \setminus \{0\}$. But $z_0 = 1$ is admissible, and there exists no z such that $R(z) < R(1)$.

The subgradient T_z , considered in Propositions 1-3 and in Theorem 1, is taken at a point z different from z_0 . It would be desirable to have at our disposal both necessary and sufficient conditions for admissibility where, like is done in Proposition 4, just T_{z_0} is used. But Example 2 shows that such a program is unrealizable without any additional assumptions. Therefore, both Proposition 4 and Example 2 should be compared with the results of the next section.

It is natural to call z_0 locally admissible if there exists a neighbourhood U of z_0 such that z_0 is admissible provided R is restricted to U . We shall need in the sequel the following trivial but useful lemma.

LEMMA 1. *If R is convex and $R(z_1) < R(z_0)$, then for every $z_\alpha = \alpha z_1 + (1 - \alpha)z_0$, $\alpha \in (0, 1)$*

$$R(z_\alpha) < R(z_0)$$

holds. \square

Thus, if R is convex and Z is a locally convex topological vector space, then z_0 is admissible if and only if z_0 is locally admissible.

3. Admissibility, compact parameter space. Let Z be a vector space, S a set and R a real function on $Z \times S$. Denote

$$R'(z, h, s) = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha} (R(z + \alpha h, s) - R(z, s))$$

whenever the limit on the right hand side of the equality exists.

THEOREM 2. *If for each $h \in Z$ there exists $s \in S$ such that $R(\cdot, s)$ attains on the line $\{z : z = z_0 + \alpha h\}$ its unique infimum at z_0 , then z_0 is admissible.*

PROOF. For each z_1 there exists $s \in S$ such that

$$R(z_0 + (z_1 - z_0), s) > R(z_0, s)$$

holds. Thus, z_0 is admissible. \square

Notice that in Theorem 2 no assumptions on the regularity of R are required. However, it is convenient to formulate a "regular" version of this Theorem which emphasizes some properties of the directional derivatives.

We recall that a function r on a vector space Z is called quasi-convex (bowl-shape) if for every $\alpha \in \mathbb{R}$ the set $\{z \in Z : r(z) \leq \alpha\}$ is convex (perhaps empty).

THEOREM 2'. *Assume that for every $h \in Z$ and $s \in S$ the function $R(z_0 + \alpha h, s)$ is differentiable with respect to α at $\alpha = 0$. If for every $h \in Z$ there exists $s \in S$ such that $R(z_0 + \alpha h, s)$ is a strictly convex function of α on a neighbourhood of 0 and $R'(z_0, h, s) = 0$, then z_0 is locally admissible. If, moreover, for every $s \in S$ $R(\cdot, s)$ is bowl-shape (quasi-convex), then z_0 is admissible. \square*

THEOREM 3. *Suppose that the following conditions are fulfilled.*

- a) *the risk function $R(z, \cdot)$ is continuous on a compact space S ,*
- b) *for each $h \in Z$ there exists $\varepsilon = \varepsilon(h)$ such that $R'(z_0 + \alpha h, h, s)$ is continuous on $(\alpha, s) \in [-\varepsilon, \varepsilon] \times S$.*

If z_0 is admissible, then for every $h \in Z$ the function $R'(z_0, h, \cdot)$ is neither in the interior of the positive cone of $C(S)$, nor in the interior of the negative cone of $C(S)$.

REMARK. Note that the conclusion of Theorem 3 says that for every $h \in Z$, $R'(z_0, h, \cdot)$ is neither positive on S , nor negative on S . Moreover, if S is connected, then for every $h \in Z$ there exists $s \in S$ such that $R'(z_0, h, s) = 0$. Clearly, in saying that a function is positive on S , we mean that it is positive at each point of S .

PROOF. If $R'(z_0, h, s) < 0$ on S , then, by the continuity of $R'(z_0 + \alpha h, h, s)$ with respect to $(\alpha, S) \in [0, \epsilon] \times S$ and by the compactness of S there exists $\epsilon' \in (0, \epsilon)$ such that

$$R'(z_0 + \alpha h, h, s) < 0$$

for each $\alpha \in (0, \epsilon')$, and for each $s \in S$. Hence, by the mean value theorem there exist $\tilde{\alpha} = \tilde{\alpha}(s)$, such that $\tilde{\alpha}(s) \in (0, \alpha)$ and

$$(3.1) \quad R(z_0, s) = R(z_0 + \alpha h, s) - \alpha R'(z_0 + \tilde{\alpha}h, s) > R(z_0 + \alpha h, s)$$

holds, which contradicts the admissibility of z_0 . Similarly, if $R'(z_0, h, s) > 0$ on S , then, by the continuity of $R'(z_0 + \alpha h, h, s)$ on $[-\epsilon, 0] \times S$, there exists $\epsilon' < \epsilon$ such that for every $\alpha \in (-\epsilon', 0)$ we have $R'(z_0 + \alpha h, h, s) > 0$ on S . Hence the inequalities in (3.1) hold, again contradicting the admissibility of z_0 . Thus, if S is connected, there exists $s \in S$ such that $R'(z_0, h, s) = 0$. \square

From the proof of Theorem 3, the following corollary is immediate.

COROLLARY 1. If the assumptions of Theorem 3 are satisfied, and $R'(z_0, h, s)$ is negative for every $s \in S$ (positive for every $s \in S$), then z_0 is inadmissible. \square

It is convenient to reformulate Theorem 3 assuming that Z is a locally convex vector space and that $R(\cdot, s)$ is Gateaux differentiable. If Z' is a dual space of Z , $z \in Z$ and $z' \in Z'$, then we denote by (z, z') the value of the functional z' at z .

THEOREM 3'. Let Z be a locally convex topological vector space and Z' the dual space of Z . Assume that S is a compact space, that $R : Z \times S \rightarrow R^1$ is Gateaux differentiable for every $s \in S$, and that the Gateaux differential $R'(z, s)$ is weakly continuous on S for every $z \in Z$. If z_0 is admissible, then for every $h \in Z$ the function $(h, R'(z_0, \cdot))$ is neither positive, nor negative on S . If, moreover, S is connected, then for every $h \in Z$ there exists $s \in S$ such that

$$(3.2) \quad (h, R'(z_0, s)) = 0. \square$$

If the assumptions of Theorem 3' are fulfilled, then $R'(z_0, \cdot)$ is a map from S into Z' , and $R'(z_0, S)$ is a star-weakly compact subset of Z' . In this case (3.2) admits geometrical interpretation formulated in Theorem 4 below, known in statistics as the Bayesian approach. The conclusion that if $R(z, \cdot)$ is continous on S (compact), then every admissible rule is necessarily a Bayes rule is well known and can be found e.g. in Wald (1950). The proof included here differs from that of Wald. If $K \subset Z'$, then by $\text{conv } K$ we denote the smallest convex set containing K , and by $\text{cl } K$ the smallest closed set containing K .

LEMMA 2. Let (Z, Z') be a pair of dual locally convex topological vector spaces.

- a) If K is a compact subset of Z' , then $0 \in \text{cl conv } K$ if and only if for every $z \in Z$ the function (z, \cdot) is neither positive, nor negative on K .
- b) If K is a compact connected subset of Z' , then $0 \in \text{cl conv } K$ if and only if for every $z \in Z$ there exists $z' \in K$ such that

$$(z, z') = 0.$$

PROOF. If $0 \in \text{cl conv } K$ then, by Proposition 1.2 in Phelps (1966), there exists a probability measure μ on K such that for every $z \in Z$

$$\int (z, z')\mu(dz') = 0.$$

Thus (z, z') cannot be positive for every $z' \in K$, nor negative for every $z' \in K$. If K is connected, then, by the continuity of (z, \cdot) there exists $z' \in K$ such that $(z, z') = 0$. On the other hand, if $0 \notin \text{cl conv } K$, then 0 and $\text{cl conv } K$ can be strictly separated. Hence, there exists $z \in Z$ and $\varepsilon > 0$ such that $(z, z') > \varepsilon$ for each $z' \in K$. \square

THEOREM 4. *If the assumptions of Theorem 3' are fulfilled then for every admissible z_0 there exists a probability measure μ on S such that*

$$(3.3) \quad \int R'(z_0, s)\mu(ds) = 0$$

holds, where the integral is in the Pettis sense, i.e. z_0 is Bayes with respect to the prior distribution μ . If, moreover, $Z = R^n$, then μ can be chosen concentrated on at most $n + 1$ points of S .

PROOF. By Lemma 2 we get $0 \in \text{cl conv } R'(z_0, S)$. Thus, there exists a probability measure μ on S such that (3.2) holds (Phelps, 1966, Proposition 1.2). If $Z = R^n$, then the last statement follows immediately from the Caratheodory theorem. \square

4. Admissibility, arbitrary parameter space. In this section we assume that Z is an arbitrary vector space and S is an arbitrary set. R is a function from $Z \times S$ into R^1 such that $R(\cdot, s)$ is convex on Z for every $s \in S$. Given $z_0, z_1 \in Z$ we denote

$$(4.1) \quad R_{z_0, z_1}(\alpha, s) = R(z_0 + \alpha z_1, s),$$

$$(4.2) \quad R'_{z_0, z_1}(\alpha, s) = \frac{\partial}{\partial \alpha} R_{z_0, z_1}(\alpha, s),$$

$$(4.3) \quad R''_{z_0, z_1}(\alpha, s) = \frac{\partial^2}{\partial \alpha^2} R_{z_0, z_1}(\alpha, s).$$

If f_1 and f_2 are two real functions on S , then we write $f_1(s) \supseteq f_2(s)$ whenever this inequality is valid for every s and the strict inequality holds for at least one $s \in S$.

The following theorem provides a characterization of inadmissible decision rules.

THEOREM 5. *Let $z_0, z_1 \in Z$ and let $R_{z_0, z_1}(\cdot, s)$, $R'_{z_0, z_1}(\cdot, s)$, and $R''_{z_0, z_1}(\cdot, s)$ given by (4.1)-(4.3) be continuous on $[0, \varepsilon]$ for every $s \in S$. Suppose that there exist $K > 0$ and $\delta > 0$ such that for each $\alpha \in [0, \delta]$ and $s \in S$*

$$(4.4) \quad |R''_{z_0, z_1}(0, s) - R''_{z_0, z_1}(\alpha, s)| \leq K |R'_{z_0, z_1}(0, s)|.$$

Then the following two statements imply each other:

(i) For some $\alpha > 0$

$$R(z_0, s) \supseteq R(z_0 + \alpha z_1, s)$$

holds;

(ii) There exists $\gamma > 0$ such that

$$(4.5) \quad R'_{z_0, z_1}(0, s) + \gamma R''_{z_0, z_1}(0, s) \leq 0.$$

PROOF. Suppose $R(z_0, s) \supseteq R(z_0 + \alpha z_1, s)$ holds. Then, by the Taylor expansion we get

$$R_{z_0, z_1}(\alpha, s) = R_{z_0, z_1}(0, s) + \alpha R'_{z_0, z_1}(0, s) + \frac{\alpha^2}{2} R''_{z_0, z_1}(\bar{\alpha}, s),$$

where $\bar{\alpha} = \bar{\alpha}(s)$ and $0 \leq \bar{\alpha} \leq \alpha$. Therefore

$$R'_{z_0, z_1}(0, s) + \frac{\alpha}{2} R''_{z_0, z_1}(\bar{\alpha}, s) \leq 0.$$

Since R is convex, R''_{z_0, z_1} is nonnegative. Thus, for a fixed $\alpha < \delta$

$$R''_{z_0, z_1}(\bar{\alpha}, s) \leq \frac{2}{\alpha} |R'_{z_0, z_1}(0, s)|$$

holds. Take now $\gamma > 0$ such that $\gamma |(2/\alpha) + K| < 1$. Then, by (4.4), we get

$$R'_{z_0, z_1}(0, s) + \gamma R''_{z_0, z_1}(0, s) \leq \left[1 - \gamma \left(\frac{2}{\alpha} + K \right) \right] R'_{z_0, z_1}(0, s) \leq 0.$$

To prove the opposite implication suppose that for some $\gamma > 0$

$$R'_{z_0, z_1}(0, s) + \gamma R''_{z_0, z_1}(0, s) \leq 0$$

holds. R_{z_0, z_1} is convex with respect to α , hence $R''_{z_0, z_1}(0, s) \geq 0$ for every $s \in S$ and hence $R'_{z_0, z_1}(0, s) \leq 0$. Moreover, we have

$$R''_{z_0, z_1}(0, s) \leq \frac{1}{\gamma} |R'_{z_0, z_1}(0, s)|.$$

Take $\alpha > 0$ such that $((1/\gamma) + K)(\alpha/2) < 1$. Then, by (4.4), we get for any $\beta \in (0, \alpha)$

$$R'_{z_0, z_1}(0, s) + \frac{\alpha}{2} R''_{z_0, z_1}(\beta, s) \leq \left[1 - \frac{\alpha}{2} \left(\frac{1}{\gamma} + K \right) \right] R'_{z_0, z_1}(0, s) \leq 0.$$

Since

$$R(z_0 + \alpha z_1, s) = R(z_0, s) + R'_{z_0, z_1}(0, s) + \frac{\alpha^2}{2} R''_{z_0, z_1}(\bar{\alpha}, s)$$

for some $\bar{\alpha} \in (0, \alpha)$, we conclude that

$$R(z_0 + \alpha z_1, s) \leq R(z_0, s).$$

Thus, $z_0 + \alpha z_1$ is better than z_0 . \square

COROLLARY 2. *Given $z_0 \in Z$, suppose that Condition (4.4) is satisfied for every $z_1 \in Z$ and $s \in S$. Then z_0 is admissible if and only if for every $\gamma > 0$, inequality (4.5) has no solution z_1 . \square*

Note that, for every quadratic loss function, condition (4.4) is trivially fulfilled for each $z_0, z_1 \in Z$. Moreover, if $R(\cdot, s)$ is strictly convex for every $s \in S$, then, in proving the admissibility of z_0 , we need to consider only those directions z_1 for which $R'_{z_0, z_1}(0, s)$ is negative for every $s \in S$. Indeed, if $R'_{z_0, z_1}(0, s_0)$ equals zero, then z_1 cannot be a direction improving z_0 .

If S is not compact, then, contrary to the assertion of Corollary 1, the condition

$$R'_{z_0, z_1}(0, s) < 0 \quad \text{for each } s \in S$$

does not imply that z_1 is a direction improving z_0 . For example, if $X \sim N(\theta, 1)$ and $R(d(\cdot), \theta) = E_\theta \{d(X) - \theta\}^2$, then $R'_{X, -X}(0, \theta) = -2E_\theta X(X - \theta) = -2$. However, the estimators $\alpha X, 0 < \alpha < 1$, do not improve the estimator $d(X) = X$.

5. Admissibility of estimators of p in the binomial distribution $B(n, p)$. In this Section we give a simple application of the presented theory. We consider the estimation of p in the binomial distribution $B(n, p)$ assuming that the quadratic loss function is used. We show how Theorems 2' and 3' imply a complete analytic description of the class of all

admissible estimators of p . This completes the previous results on this subject and yields the analytic counterpart for the Bayesian description in Johnson (1971). Clearly, both approaches must be equivalent. At the end of this Section we include a simple proof of this equivalence.

It is convenient to collect in Theorem 6 those immediate consequences of Theorems 2' and 3' that will be used in this section.

THEOREM 6. *Let S be a compact connected space, $Z = R^k$ and $R: R^k \times S \rightarrow R^1$. Assume that $R(\cdot, s)$ is differentiable on R^k for each $s \in S$ and that $\text{grad}_z R(\mathbf{z}, s)$ is continuous on $R^k \times S$.*

(a) *Given $\mathbf{z}_0 \in R^k$, if for every $\mathbf{h} \in R^k$ there exists $s \in S$ such that $R(\cdot, s)$ is strictly convex, and*

$$(5.1) \quad (\mathbf{h}, \text{grad}_z R(\mathbf{z}_0, s)) = 0$$

holds, then \mathbf{z}_0 is admissible.

(b) *If \mathbf{z}_0 is admissible, then for each $\mathbf{h} \in R^k$ there exists $s \in S$ such that equality (5.1) is valid. \square*

If we estimate p in the binomial distribution $B(n, p)$ and squared error loss is used, then the risk function is of the form

$$(5.2) \quad R(\mathbf{z}, p) = \sum_{i=0}^n (z_i - p)^2 \binom{n}{i} p^i (1-p)^{n-i},$$

where $p \in [0, 1]$, and $\mathbf{z} = (z_0, z_1, \dots, z_n)$ represents an estimator. Suppose that $\mathbf{z} = (z_0, z_1, \dots, z_n)$ is admissible. Then, clearly, $z_i \in [0, 1]$ for each i . By Theorem 6, part (b), for every $\mathbf{h} = (h_0, h_1, \dots, h_n) \in R^{n+1}$ there exists $p \in [0, 1]$ such that

$$\sum_{i=0}^n h_i (z_i - p) p^i (1-p)^{n-i} = 0$$

holds. Since $z_i = z_i p + z_i (1-p)$, we get

$$\sum_{i=0}^n h_i (z_i - p) p^i (1-p)^{n-i} = h_0 z_0 (1-p)^{n+1} + \sum_{j=1}^n (h_{j-1} (z_{j-1} - 1) + h_j z_j) p^j (1-p)^{n+1-j} + h_n (z_n - 1) p^{n+1} = (\mathbf{g}(p), \mathbf{A}\mathbf{h}),$$

where

$$\mathbf{g}(p) = ((1-p)^{n+1}, \dots, p^i (1-p)^{n+1-i}, \dots, p^{n+1})',$$

$$\mathbf{A} = \mathbf{A}(\mathbf{z}) = (a_{ij}), i = 0, 1, \dots, n+1, j = 0, 1, \dots, n, a_{jj} = z_j, a_{j+1,j} = z_j - 1, a_{ij} = 0$$

if $i \notin \{j, j+1\}, j = 0, 1, \dots, n$. The parentheses (\cdot, \cdot) stand for the usual scalar product.

Let $H = \{\mathbf{A}\mathbf{h} : \mathbf{h} \in R^{n+1}\}$. Clearly, H can be considered as a subspace of R^{n+2} . If $z_i \in (0, 1), i = 0, 1, \dots, n$, then $\dim H = n + 1$. Moreover, if $\mathbf{b} = (b_0, b_1, \dots, b_{n+1})'$, where

$$(5.3) \quad \begin{aligned} b_0 &= (1 - z_0)(1 - z_1) \dots (1 - z_n) \\ b_1 &= z_0(1 - z_1) \dots (1 - z_n) \\ &\vdots \\ b_i &= z_0 z_1 \dots z_{i-1} (1 - z_i) \dots (1 - z_n) \\ &\vdots \\ b_{n+1} &= z_0 z_1 \dots z_n, \end{aligned}$$

then \mathbf{b} is perpendicular to H .

Denote by cone $\mathbf{g}([0, 1])$ the convex cone spanned on the set $\{\mathbf{g}(p) : p \in [0, 1]\}$. Clearly, cone $\mathbf{g}([0, 1])$ is closed in R^{n+2} .

LEMMA 3. *If $z_i \in (0, 1)$, $i = 0, 1, \dots, n$, then the following statements imply each other.*

(i) *for every $\mathbf{h} \in R^{n+1}$ there exists $p \in (0, 1)$ such that*

$$(\mathbf{g}(p), \mathbf{A}\mathbf{h}) = 0,$$

(ii) $\mathbf{b} \in$ cone $\mathbf{g}([0, 1])$, *where \mathbf{b} is given by (5.3),*

(iii) $(\mathbf{b}, \mathbf{d}) \geq 0$ *whenever $\mathbf{d} = (d_0, d_1, \dots, d_{n+1})'$, and $\sum_{i=0}^{n+1} d_i g_i(p)$ is nonnegative on $[0, 1]$.*

PROOF. (ii) \Rightarrow (i). Since

$$\text{cone } \mathbf{g}([0, 1]) = \cup_{\lambda \geq 0} \lambda \text{ conv } \mathbf{g}([0, 1]),$$

where conv $\mathbf{g}([0, 1])$ stands for the convex hull spanned on the set $\mathbf{g}([0, 1])$, there exists a positive number λ and a probability measure m concentrated on $[0, 1]$ such that

$$(5.4) \quad \mathbf{b} = \lambda \int_0^1 \mathbf{g}(p)m(dp).$$

Let us note that m cannot be concentrated on $\{0, 1\}$ for that would contradict the assumption that $z_i \in (0, 1)$ for each i . The vector \mathbf{b} is perpendicular to H . Hence

$$(5.5) \quad (\mathbf{b}, \mathbf{A}\mathbf{h}) = \lambda \int_0^1 (\mathbf{g}(p), \mathbf{A}\mathbf{h})m(dp) = 0$$

holds for every $\mathbf{h} \in R^{n+1}$. Now, observe that $(\mathbf{g}(p), \mathbf{A}\mathbf{h})$ is continuous on $[0, 1]$ and its integral equals zero. This implies the existence of $p \in (0, 1)$ such that $(\mathbf{g}(p), \mathbf{A}\mathbf{h}) = 0$ holds.

(i) \Rightarrow (ii). Suppose that $\mathbf{b} \notin$ cone $\mathbf{g}([0, 1])$. Then \mathbf{b} and cone $\mathbf{g}([0, 1])$ can be separated by a hyperplane, i.e. there exists $\mathbf{c} \in R^{n+2}$ such that

$$(5.6) \quad (\mathbf{c}, \mathbf{y}) \leq 0 \quad \text{for } \mathbf{y} \in \text{cone } \mathbf{g}([0, 1])$$

and

$$(5.7) \quad (\mathbf{c}, \mathbf{b}) > 0.$$

Since the components of $\mathbf{g}(p)$ and \mathbf{b} are nonnegative and these vectors are different from $0 \in R^{n+2}$, we infer from (5.6) that \mathbf{c} cannot be represented in the form $\mathbf{c} = \gamma\mathbf{b}$, where γ is a positive number. Take $\alpha \in R^1$ and $\mathbf{h} \in R^{n+1}$ such that

$$\mathbf{c} = \alpha\mathbf{b} + \mathbf{A}\mathbf{h}.$$

Then $(\mathbf{c}, \mathbf{b}) = \alpha(\mathbf{b}, \mathbf{b})$; hence $\alpha > 0$ and $\mathbf{h} \neq 0$. Furthermore, we infer from (5.6) that

$$0 \geq \alpha(\mathbf{g}(p), \mathbf{b}) + (\mathbf{g}(p), \mathbf{A}\mathbf{h}) \geq \alpha\varepsilon + (\mathbf{g}(p), \mathbf{A}\mathbf{h}),$$

where $\varepsilon = \inf\{(\mathbf{g}(p), \mathbf{b}), p \in [0, 1]\} > 0$. Thus, $(\mathbf{g}(p), \mathbf{A}\mathbf{h}) < 0$ holds for every $p \in [0, 1]$ in contradiction with our assumption.

The equivalence of statements (ii) and (iii) is straightforward and well-known (see Karlin and Studden, 1966, Chapter 3, Theorem 9.1; Krein and Nudelman, 1973, Chapter 3, Theorem 1.1). \square

THEOREM 7. *Let $\mathbf{z} = (z_0, z_1, \dots, z_n)$, $z_i \in (0, 1)$, $i = 0, 1, \dots, n$, and let $\mathbf{b} = (b_0, b_1, \dots, b_{n+1})'$ be given by (5.3).*

(a) *If $n = 2\nu - 1$ and $\nu \in \{1, 2, \dots\}$, then \mathbf{z} is admissible if and only if the quadratic*

forms

$$(5.8) \quad \sum_{i,j=0}^{\nu} b_{i+j} \alpha_i \alpha_j \quad \text{and} \quad \sum_{i,j=0}^{\nu-1} b_{i+j+1} \beta_i \beta_j$$

are nonnegative definite.

(b) If $n = 2\nu$ and $\nu \in \{1, 2, \dots\}$, then \mathbf{z} is admissible if and only if the quadratic forms

$$(5.9) \quad \sum_{i,j=0}^{\nu} b_{i+j+1} \alpha_i \alpha_j \quad \text{and} \quad \sum_{i,j=0}^{\nu} b_{i+j} \beta_i \beta_j$$

are nonnegative definite.

PROOF. Since the proofs of both parts of Theorem 7 are analogous, we shall prove part (a), only.

Necessity. Let \mathbf{z} be admissible. Then, by Lemma 3, $(\mathbf{b}, \mathbf{d}) \geq 0$ whenever $\sum_{i=0}^{n+1} d_i g_i(p)$ is nonnegative on $[0, 1]$. From the Lukacs-Markov Theorem (see Krein and Nudelman, 1973, Theorem 2.2, Chapter 3; Karlin and Studden, 1966, Chapter 4, Section 1) it follows that every nonnegative polynomial of $g_i(p)$, $i = 0, 1, \dots, n + 1$, is of the form

$$\begin{aligned} W(p) &= \left\{ \sum_{k=0}^{\nu} \alpha_k p^k (1-p)^{\nu-k} \right\}^2 + (1-p)p \left\{ \sum_{k=0}^{\nu-1} \beta_k p^k (1-p)^{\nu-k-1} \right\}^2 \\ &= \sum_{i,j=0}^{\nu} g_{i+j}(p) \alpha_i \alpha_j + \sum_{i,j=0}^{\nu-1} g_{i+j+1}(p) \beta_i \beta_j. \end{aligned}$$

Thus, the two quadratic forms in (5.8) are nonnegative definite.

Sufficiency. If both forms in (5.8) are nonnegative definite, then, by Lemma 3, for every $\mathbf{h} \in R^{n+1}$ there exists $p \in (0, 1)$ such that $(\mathbf{g}(p), \mathbf{A}\mathbf{h}) = 0$. Equivalently, for each $\mathbf{h} \in R^{n+1}$ there exists $p \in (0, 1)$ such that

$$(\mathbf{h}, \text{grad}_{\mathbf{z}} R(\mathbf{z}, p)) = 0.$$

Since $R(\cdot, p)$ is strictly convex on R^{n+1} for $p \in (0, 1)$, we infer from Theorem 6, part (a), that \mathbf{z} is admissible. \square

Now, we are going to derive a characterization of the entire class of admissible estimators of p . We apply here in fact the method of sequences of priors and sample spaces used in Hsuan (1979) and Brown (1981). First we note that if $\mathbf{z} = (z_0, z_1, \dots, z_n)$ is admissible, then by Theorem 11 of Karlin and Rubin (1956), $z_i \leq z_{i+1}$, $i = 0, 1, \dots, n - 1$. In our case, if $z_i > z_{i+1}$ holds for some $i < n$, then a direct calculation, as well as the use of Corollary 1, shows that the estimator $\mathbf{z}' = (z_0, z_1, \dots, z_{i-1}, z_i + h_i, z_{i+1} + h_{i+1}, z_{i+2}, \dots, z_n)$ is better than \mathbf{z} provided

$$h_{i+1} = - \frac{(i+1)h_i(1-z_i)}{(n-i)z_i}$$

and h_i is negative and near to zero. Moreover, it is known that if $\mathbf{z} = (0, \dots, 0, z_r, \dots, z_s, 1, \dots, 1)$, and if $\mathbf{z}' = (z'_0, z'_1, \dots, z'_n)$ is not worse than \mathbf{z} , then $z'_0 = z'_1 = \dots = z'_{r-1} = 0$ and $z'_{s+1} = \dots = z'_n = 1$ hold (cf. Johnson, 1971). It is also trivial and known (cf. Johnson, 1971) that the estimators of the form $(0, 0, \dots, z_i, 1, \dots, 1)$, $z_i \in (0, 1)$, $i \in \{0, 1, \dots, n\}$, are admissible. Finally, we note that $\mathbf{z} = (0, \dots, 0, z_r, \dots, z_s, 1, \dots, 1)$, $r < s$, $z_i \in (0, 1)$ for $i \in \{r, r + 1, \dots, s\}$, is admissible with respect to R given by (5.2) if and only if $\tilde{\mathbf{z}} = (z_r, z_{r+1}, \dots, z_s)$ is admissible with respect to the function

$$R_1(\tilde{\mathbf{z}}, p) = \sum_{i=0}^{s-r} (z_{r+i} - p)^2 b_i p^i (1-p)^{s-r-i},$$

where $b_i = \binom{n}{r+i}$ (cf. also Theorem 3.1 in Brown, 1981). From Theorem 6 and from the proof of Theorem 7 it follows that the form of the positive coefficients b_i has no influence on the admissibility of \mathbf{z} . Hence, using $s - r$ instead of n , we can apply to \mathbf{z} the characterization of admissible estimators given in Theorem 7.

THEOREM 8. *An estimator $\mathbf{z} = (z_0, z_1, \dots, z_n)$ is admissible under the quadratic loss function for estimating $p \in [0, 1]$ in the binomial distribution $B(n, p)$ if and only if it is of one of the following forms*

- (a) $\mathbf{z} = (0, 0, \dots, 0)$, or $\mathbf{z} = (0, 0, \dots, 0, 1, \dots, 1)$, or $\mathbf{z} = (1, 1, \dots, 1)$, or
- (b) $\mathbf{z} = (0, \dots, 0, z_i, 1, \dots, 1)$ and $z_i \in (0, 1), i \in \{0, 1, \dots, n\}$,

or

- (c) $\mathbf{z} = (0, \dots, 0, z_r, \dots, z_s, 1, \dots, 1), r < s, r, s \in \{0, 1, \dots, n\}, z_i \in (0, 1)$ for $i \in \{r, r + 1, \dots, s\}$, and $\tilde{\mathbf{z}} = (\tilde{z}_0, \tilde{z}_1, \dots, \tilde{z}_{r-s})$ fulfills the characterization given in Theorem 7 with $r - s$ instead of n , and $\tilde{z}_i = z_{r+i}, i \in \{0, 1, \dots, s - r\}$. \square

REMARK 1. A direct and simple calculation shows that Theorem 8 implies that if $n = 1$, then $\mathbf{z} = (z_0, z_1)$ is admissible if and only if $0 \leq z_0 \leq z_1 \leq 1$. Similarly, if $n = 2$, then $\mathbf{z} = (z_0, z_1, z_2)$ is admissible if and only if $0 \leq z_0 \leq z_1 \leq z_2 \leq 1$. If $n > 2$, then there exist nondecreasing and increasing sequences z_0, z_1, \dots, z_n such that $\mathbf{z} = (z_0, z_1, \dots, z_n)$ are not admissible.

For example, if $n = 3$ and $0 < z_0 < z_1 < z_2 < z_3 < 1$, then $\mathbf{z} = (z_0, z_1, z_2, z_3)$ is admissible if and only if

$$z_2(z_1 - z_0)(z_3 - z_2)(1 - z_1) \geq z_0(1 - z_3)(z_2 - z_1)^2.$$

REMARK 2. In the search for the prior distribution with respect to which a given admissible estimator is Bayes, one can follow the patterns given in Karlin and Studden (1966, Chapter IV, Section 2). For example, it is possible to show that the “upper” and “lower” prior distributions are concentrated on zeros of some determinants related to the quadratic forms given by (5.8) and (5.9) and functions $g_i(p)$ (cf. also Remark 4 below).

REMARK 3. The set of admissible estimators of p is a closed subset of R^{n+1} (cf. Johnson, 1971, or Theorem 8). Hence, and from Ferguson (1967), Exercise 1.8.8, it follows that if $X \sim B(n, p)$, then every linear estimator $AX + B$ which takes values in $[0, 1]$ is admissible.

REMARK 4. We show that the analytic conditions of Theorem 7 are equivalent to Johnson’s Bayesian description. (C. Hipp and a referee independently noted, proved and generously communicated this equivalence to the author. Their proofs differ slightly. Here we include the somewhat simpler proof of C. Hipp.) Let $\mathbf{z} = (z_0, z_1, \dots, z_n)$ with $z_1 \in (0, 1)$ be admissible. Then it is Bayes for some prior on $[0, 1]$ and

$$z_k = \frac{\int_0^1 p^{k+1}(1-p)^{n-k}m(dp)}{\int_0^1 p^k(1-p)^{n-k}m(dp)}.$$

Since $z_n < 1$, we have $m(1) < 1$. Notice that

$$\frac{z_k}{1 - z_k} = \frac{\int_0^1 p^{k+1}(1-p)^{n-k}m(dp)}{\int_0^1 p^k(1-p)^{n-k+1}m(dp)}.$$

and for $k = 0, 1, \dots, n$

$$b'_{k+1} = \frac{z_0}{1 - z_0} \frac{z_1}{1 - z_1} \dots \frac{z_k}{1 - z_k} = \frac{\int_0^1 p^{k+1}(1-p)^{n-k}m(dp)}{\int_0^1 (1-p)^{n+1}m(dp)}.$$

Let $m = m' + m''$, where m' and m'' are orthogonal and m'' is concentrated on $\{1\}$ (we admit $m''(1) = 0$, too). We put

$$dm_1 = (1-p)^{n+1} dm' / \int_0^1 (1-p)^{n+1} dm \quad \text{and} \quad dm_2 = dm'' / \int_0^1 (1-p)^{n+1} dm.$$

Then

$$b'_{k+1} = \int_0^1 \left(\frac{p}{1-p} \right)^{k+1} dm_1(p), \quad k = 0, 1, \dots, n-1,$$

and

$$b'_{n+1} = \int_0^1 \left(\frac{p}{1-p} \right)^{n+1} dm_1(p) + m_2(1).$$

Finally, denote by m'_1 the measure on $[0, \infty)$ induced by the mapping $p \rightarrow p/(1-p)$ and m_1 . Thus, we have

$$b'_{k+1} = \int_0^\infty t^{k+1} dm'_1(t), \quad k = 0, 1, \dots, n-1,$$

and

$$b'_{n+1} = \int_0^\infty t^{n+1} dm'_1(t) + m_2(1).$$

Now, by Theorem V.10.1 in Karlin and Studden (1966), we infer the necessity of Conditions (5.8) and (5.9). Proceeding in the reverse direction, we obtain the desired equivalence.

Acknowledgment. The author wishes to thank B. Bednarek-Kozek, T. Bednarski, M. M. Bock, C. Hipp, O. Krafft, M. Musiela and R. Zmysłony for discussions and comments on the ideas and problems related to the content of this paper. Moreover, he is very grateful to C. Hipp for the provision of the content of Remark 4 and to the referees for pointing out some references and making various detailed comments which have considerably improved this paper.

REFERENCES

- ACHILLES, A., ELSTER, K. H. and NEHSE, R. (1979). Bibliographie zur Vektoroptimierung (Theorie und Anwendungen). *Math. Operationsforsch. Statist., Ser. Optimization* **10** 277-321.
- ARROW, K. J., BARANKIN, E. W. and BLACKWELL, D. (1953). Admissible points of convex sets. In: Kuhn, H. W. and Tucker, A. W. (eds.): Contributions to the theory of games. *Ann. Math. Studies* **28** 87-91. Princeton Univ., Princeton, New Jersey.
- ATHANS, M. and GEERING, H. P. (1973). Necessary and sufficient conditions for differentiable nonscalarvalued functions to attain extrema. *IEEE Trans. Automatic Control* **AC-18** **2** 132-139.
- AUBIN, J.-P. (1971). A Pareto minimum principle. In: Kuhn, H. W. and Szegö, G. P. (eds.): *Differential Games and Related Topics* 147-175. North-Holland, Amsterdam.
- BROWN, L. D. (1979). A heuristic method for determining admissibility of estimators—with applications. *Ann. Statist.* **7** 960-994.
- BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample space. *Ann. Statist.* **9** 1289-1300.
- FARRELL, R. H. (1968). Towards a theory of generalized Bayes tests. *Ann. Math. Statist.* **39** 1-22.
- FERGUSON, T. S. (1967). *Mathematical Statistics. A Decision Theoretic Approach*. Academic, New York.
- HSUAN, F. C. Y. (1979). A stepwise Bayesian procedure. *Ann. Statist.* **7** 860-868.
- JOHNSON, B. McK. (1971). On admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.* **42** 1579-1587.

- KARLIN, S. and RUBIN, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Ann. Math. Statist.* **27** 272-299.
- KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers. Wiley, New York.
- KREIN, M. G. and NUDELMAN, A. A. (1973). *The Problem of Markov's Moments and Extremal Problems*. Nauka, Moscow (in Russian).
- KUTATELADZE, S. S. (1977). Changes of variables in the Young transformation. *Dokl. Akad. Nauk USSR* **233** 1039-1041 (in Russian).
- LECAM, L. (1974). *Notes on Asymptotic Methods in Statistical Decision Theory*. Centre de Recherches Mathématiques, Université de Montreal.
- NEUSTADT, L. W. (1969). A general theory of extremals. *J. Comp. System Science* **3** 57-92.
- PHELPS, R. R. (1966). *Lecture Notes in Choquet's Theorem*. Van Nostrand, Princeton.
- SMALE, S. (1973). Global analysis and economics I. Pareto optimum and a generalization of Morse theory. In: Peixoto, M. (ed.): *Dynamical Systems*, pages 531-544. Academic, New York.
- THIBAUT, L. (1979). Subdifferentiability of compactly Lipschitzian mappings. Third Symposium on Operations Research (Univ. Mannheim, Mannheim, 1978), Section I, 637-645, Operations Research Verfahren, 31, Hain, Meisenheim.
- WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.
- ZOWE, J. (1974). Subdifferentiability of convex functions with values in an ordered vector space. *Math. Scand.* **34** 69-83.

INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
PL-51-617 WROCZAW,
KOPERNIKA 18
POLAND

INSTITUTE OF MATHEMATICS
UNIVERSITY OF WROCZAW
PL-50-384 WROCZAW
PL. GRUNWALDZKI 2/4
POLAND