Finally, lest us observe that splitting the prize has always a detrimental effect on the expected return. When $a = 1$, $b = c = 0$ (i.e. if we win a unit amount only for getting the best candidate), the expected return is $e^{-1} = 0.3679$. When we get 1/3 for stopping at any of the three top candidates, the expected return is lowest, and equals only 0.2616.

Table 1

Optimal thresholds x, y, and z, and the optimal return E for different prizes b and c

c	b	x	y	z	E
0	0	0.3679	1.0000	1.0000	0.3679
	0.1	0.3594	0.9091	1.0000	0.3464
	0.2	0.3531	0.8333	1.0000	0.3281
	0.3	0.3489	0.7692	1.0000	0.3124
	0.4	0.3468	0.7143	1.0000	0.2987
	0.5	0.3470	0.6667	1.0000	0.2868
0.1	0.1	0.3507	0.8849	0.9129	0.3323
	0.2	0.3445	0.7956	0.9129	0.3136
	0.3	0.3410	0.7258	0.9129	0.2983
	0.4	0.3402	0.6689	0.9129	0.2851
	0.45	0.3409	0.6442	0.9129	0.2793
0.2	0.2	0.3379	0.7501	0.8452	0.2985
	0.3	0.3358	0.6771	0.8452	0.2774
	0.4	0.3373	0.6208	0.8452	0.2717
0.3	0.3	0.3342	0.6236	0.7906	0.2696
	0.35	0.3364	0.5957	0.7906	0.2640
0.33	0.33	0.3367	0.5868	0.7746	0.2616

Reference

J. Gilbert and F. Mosteller, *Recognizing the maximum of a sequence*, Amer. Stat. Assoc. 61 (1966), pp. 35–73.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

REDUCIBILITY OF STATISTICAL STRUCTURES AND DECISION PROBLEMS

R. BARTOSZYŃSKI and E. PLESZCZYŃSKA

Institute of Mathematics, Polish Academy of Sciences, Warszawa, Poland,
Institute of Computer Science, Polish Academy of Sciences, Warszawa, Poland

0. Introduction

The object of this paper is to analyze the necessary conditions which must be met in order that the decision procedures applicable to one statistical decision problem could also be used in other problems obtained from the original by suitable transformations. The basic concept is that of $\mathscr{F}$-reducibility of two statistical structures (and of two statistical decision problems), where $\mathscr{F}$ is a family of mappings of the underlying sample spaces. The paper gives conditions under which proximity of measures in one structure is preserved in the reduced structure (which is a prerequisite for robustness of procedures), and under which a parameter in one structure can also serve as a parameter in the other.

Moreover, the conditions are given under which certain desirable properties of parameters are preserved in the reduced structure. Applicability of the introduced concepts to statistical analysis of stochastic processes is discussed.

1. Reducible statistical structures

By a *statistical structure* we shall mean a triple $(\mathscr{X}, \mathscr{A}, \mathscr{P})$, where $\mathscr{X}$ is an arbitrary set, $\mathscr{A}$ is a σ-field of subsets of $\mathscr{X}$, and $\mathscr{P}$ is a family of probability measures on $(\mathscr{X}, \mathscr{A})$.

Let now $(\mathscr{X}, \mathscr{A})$ and $(\mathscr{X}', \mathscr{A}')$ be two measurable spaces, and let $\mathscr{K}$ and $\mathscr{K}'$ denote respectively the classes of all probability measures on $(\mathscr{X}, \mathscr{A})$ and $(\mathscr{X}', \mathscr{A}')$. Next, let $\mathscr{F}$ be a class of $\mathscr{A}-\mathscr{A}'$ measurable mappings $f\colon \mathscr{X} \to \mathscr{X}'$. With each $f \in \mathscr{F}$ we can connect a mapping $\bar{f}\colon \mathscr{K} \to \mathscr{K}'$ by putting

$$\bar{f}P(A') = P(f^{-1}(A')), \quad A' \in \mathscr{A}'.$$

Let $\bar{f}^{-1}P' = \{P\colon \bar{f}P = P'\}$. Futhermore, for $\mathscr{P} \subset \mathscr{K}$ and $\mathscr{P}' \subset \mathscr{K}'$, let

$$\bar{\mathscr{F}}(\mathscr{P}) = \bigcup_{f \in \mathscr{F}} \{\bar{f}(\mathscr{P})\} = \{P' \in \mathscr{K}'\colon \bar{f}P = P' \text{ for some } P \in \mathscr{P} \text{ and some } f \in \mathscr{F}\},$$

and

$$\bar{\mathscr{F}}^{-1}(\mathscr{P}') = \bigcup_{f\in\mathscr{F}} \{\bar{f}^{-1}(\mathscr{P}')\}$$
$$= \{P \in \mathscr{X} : \bar{f}P = P' \text{ for some } P' \in \mathscr{P}' \text{ and some } f \in \mathscr{F}\}.$$

Let us consider two statistical structures $\mathscr{M} = (\mathscr{X}, \mathscr{A}, \mathscr{P})$ and $\mathscr{M}' = (\mathscr{X}', \mathscr{A}', \mathscr{P}')$.

DEFINITION 1. The structure $\mathscr{M}$ will be called *$\mathscr{F}$-reducible to $\mathscr{M}'$*, to be denoted by $\mathscr{M} \underset{\mathscr{F}}{\rightarrow} \mathscr{M}'$, if $\mathscr{P}' \subset \bar{\mathscr{F}}(\mathscr{P})$ and $\mathscr{P} \subset \bar{\mathscr{F}}^{-1}(\mathscr{P}')$.

Two important special cases of reducibility are obtained if

(i) $$\mathscr{P}' = \bar{\mathscr{F}}(\mathscr{P})$$

or

(ii) $$\mathscr{P} = \bar{\mathscr{F}}^{-1}(\mathscr{P}').$$

In case (i), we shall say that the structure $\mathscr{M}'$ is *$\mathscr{F}$-derived from $\mathscr{M}$*, symbolically $\mathscr{M} \underset{\mathscr{F}}{\rightarrow\mid} \mathscr{M}'$ (for the concept of $\mathscr{F}$-derived structures in case of $\mathscr{F}$ consisting of a single function see Barra [1]). In case (ii), we shall say that the structure $\mathscr{M}$ is *$\mathscr{F}$-primitive for $\mathscr{M}'$*, symbolically $\mathscr{M} \underset{\mathscr{F}}{\mapsto} \mathscr{M}'$.

Observe that these conditions are not symmetric: if $\mathscr{M}'$ is $\mathscr{F}$-derived from $\mathscr{M}$, then $\mathscr{M}$ is, in general, not $\mathscr{F}$-primitive for $\mathscr{M}'$ (since $\mathscr{P}$ need not to be equal to $\bar{\mathscr{F}}^{-1}(\bar{\mathscr{F}}(\mathscr{P}))$). Similarly, if $\mathscr{M}$ is $\mathscr{F}$-primitive for $\mathscr{M}'$, then $\mathscr{M}'$ is not, in general, $\mathscr{F}$-derived from $\mathscr{M}$.

Before giving examples, it is worth to state the following simple properties.

Let $\mathscr{M}_i = (\mathscr{X}_i, \mathscr{A}_i, \mathscr{P}_i)$, $i = 1, 2, 3$, be three statistical structures, and let $\mathscr{F}_i$ be a class of appropriately measurable mappings $f_i\colon \mathscr{X}_i \to \mathscr{X}_{i+1}$ $(i = 1, 2)$. Let $\mathscr{F}$ be the class of mappings $f\colon \mathscr{X}_1 \to \mathscr{X}_3$ where $f = f_2 \circ f_1$, $f_1 \in \mathscr{F}_1$, $f_2 \in \mathscr{F}_2$.

PROPOSITION 1. *If $\mathscr{M}_1 \underset{\mathscr{F}_1}{\rightarrow\mid} \mathscr{M}_2$ and $\mathscr{M}_2 \underset{\mathscr{F}_2}{\rightarrow\mid} \mathscr{M}_3$, then $\mathscr{M}_1 \underset{\mathscr{F}}{\rightarrow\mid} \mathscr{M}_3$.*

PROPOSITION 2. *If $\mathscr{M}_1 \underset{\mathscr{F}_1}{\mapsto} \mathscr{M}_2$ and $\mathscr{M}_2 \underset{\mathscr{F}_2}{\mapsto} \mathscr{M}_3$, then $\mathscr{M}_1 \underset{\mathscr{F}}{\mapsto} \mathscr{M}_3$.*

PROPOSITION 3. *If either $\mathscr{M}_1 \underset{\mathscr{F}_1}{\rightarrow\mid} \mathscr{M}_2$ and $\mathscr{M}_2 \underset{\mathscr{F}_2}{\mapsto} \mathscr{M}_3$ or $\mathscr{M}_1 \underset{\mathscr{F}_1}{\mapsto} \mathscr{M}_2$ and $\mathscr{M}_2 \underset{\mathscr{F}_2}{\rightarrow\mid} \mathscr{M}_3$, then $\mathscr{M}_1$ is $\mathscr{F}$-reducible to $\mathscr{M}_3$.*

Let $\mathscr{M} = (\mathscr{X}, \mathscr{A}, \mathscr{P})$ be a statistical structure and let $\mathscr{G}$ be a family of measurable mappings $\mathscr{X} \to \mathscr{X}$. If $\mathscr{M} \underset{\mathscr{G}}{\rightarrow} \mathscr{M}$, we say that $\mathscr{M}$ is *$\mathscr{G}$-self reducible*. Similarly, if $\mathscr{P} = \bar{\mathscr{G}}(\mathscr{P})$ then $\mathscr{M}$ is $\mathscr{G}$-self derived, and if $\mathscr{P} = \bar{\mathscr{G}}^{-1}(\mathscr{P})$ then $\mathscr{M}$ is $\mathscr{G}$-self primitive. If $\mathscr{G}$ is such that each $\bar{g}$ for $g \in \mathscr{G}$ is a one-to-one mapping of $\mathscr{P}$ onto $\mathscr{P}$, then $\mathscr{M}$ is called *invariant under* $\mathscr{G}$. In particular, we have

PROPOSITION 4. *If $\mathscr{G}$ is a group and $\bar{\mathscr{G}}(\mathscr{P}) = \mathscr{P}$, then $\mathscr{M}$ is invariant under $\mathscr{G}$.*

The most important examples of the introduced concepts are in the domain of time-series analysis. In all examples below, $\mathscr{X}$, $\mathscr{X}'$, and $\mathscr{X}''$ are equal to the space $\mathbf{R}^Z$ of all doubly infinite sequences of real numbers, and $\mathscr{A}$ will be the σ-field generated by cylinders with Borel bases.

EXAMPLE 1. Let $\mathscr{M} = (\mathscr{X}, \mathscr{A}, \mathscr{P})$ be the white noise structure, i.e. $\mathscr{P}$ is a family of measures corresponding to the case where random variables ξ_i, $i \in Z$, defined as $\xi_i(x) = x$ for any $x \in \mathscr{X}$, are i.i.d. random variables with a (univariate) distribution belonging to some family, say Q, all members of Q being dominated by some fixed measure λ. It is assumed that $E\xi_i^2 < \infty$ for any member of Q.

Let now $\mathscr{C}$ be a class of real sequences $c = \{c_k, k = 0, 1, \ldots\}$ with $\sum c_k^2 < \infty$ and for any $t \in Z$ and any $c \in \mathscr{C}$ let $(n_0, n_1, \ldots)$ be a subsequence of integers such that the corresponding subsequence of partial sums $S_{n_j}(x; c, t) = \sum_{k=0}^{n_j} c_k x_{t-k}$ converges a.e. λ. Let $\mathscr{F}$ be a family of mappings $\mathscr{X} \to \mathscr{X}'$ defined by the formula

$$x_t' = \begin{cases} \lim\limits_{j\to\infty} S_{n_j}(x; c, t) & \text{if this series converges,} \\ 0 & \text{otherwise,} \end{cases}$$

where $c \in \mathscr{C}$.

In this way, we obtain the structure $\mathscr{M}' = (\mathscr{X}', \mathscr{A}, \mathscr{P}')$ which is $\mathscr{F}$-derived from the white noise structure $\mathscr{M}$. Depending on the choice of the class $\mathscr{C}$, one obtains here various structures of linear processes, e.g. linear autoregressive-moving average (ARMA) processes, etc. (obviously, the choices of subsequences $\{n_k\}$ have no influence on $\mathscr{M}'$).

EXAMPLE 2. Assume that $\mathscr{M}' = (\mathscr{X}', \mathscr{A}, \mathscr{P}')$ is an linear ARMA structure. Let $\mathscr{F}^*$ be a family of mappings $\mathscr{X}'' \to \mathscr{X}'$, where

$$x_t' = \sum_{k=0}^{p} (-1)^k \binom{p}{x} x_{t-k}'', \quad p = 1, 2, \ldots$$

(so that x' is obtained by taking pth differences of x''). If $\mathscr{P}'' = \mathscr{F}^{*-1}(\mathscr{P}')$, then $\mathscr{M}'' = (\mathscr{X}'', \mathscr{A}, \mathscr{P}'')$ is a structure which is $\mathscr{F}^*$-primitive for the linear ARMA structure. The structure $\mathscr{M}''$ was introduced by Box and Jenkins [3] and was called there *autoregressive-integrated moving average* (ARIMA) *structure*.

The common practical situations are formalized as those in which we can observe elements of $\mathscr{X}$ in a statistical structure $\mathscr{M} = (\mathscr{X}, \mathscr{A}, \mathscr{P})$ sampled according to some $P \in \mathscr{P}$, where $\mathscr{M}$ is defined as either $\mathscr{F}$-primitive or $\mathscr{F}$-derived from some other structure $\mathscr{M}'$. Here $\mathscr{M}'$ and $\mathscr{F}$ provide a formalization of the reality as represented by $\mathscr{M}$. For any decision problem it is therefore important to know how the "small deviations" from the model $\mathscr{M}$ influence $\mathscr{F}$ and $\mathscr{M}'$ and vice versa. This seems to be the essence of the general concept of robustness as applied to the present situation. In the sequel, we give a possible approach to this problem.

2. "Robustness" of reducible structures

From now on, we assume that $\mathscr{X}, \mathscr{X}', \ldots$ are metric spaces, with distances denoted respectively by $\varrho, \varrho', \ldots$; the σ-fields $\mathscr{A}, \mathscr{A}', \ldots$ will always be the Borel fields, all measures will be Borel measures, and all functions under consideration will be tacitly assumed appropriately measurable.

Let P_1, P_2 be two probability measures on the space $\mathscr{X}$. The *Lévy–Prohorov distance* $L(P_1, P_2)$ is defined as $L(P_1, P_2) = \inf\{\varepsilon > 0: \forall F \text{ closed}, P_1(F) \leqslant P_2(F^\varepsilon)+ +\varepsilon \text{ and } P_2(F) \leqslant P_1(F^\varepsilon)+\varepsilon\}$, where F^ε is the ε-neighbourhood of the set F. Then (see Prohorov [6]), L is a metric in the space $\mathscr{K}$ of all measures on $\mathscr{X}$, the space $\mathscr{K}$ with metric L is separable and complete, if $\mathscr{X}$ has these properties, and $L(P_n, P) \to 0$ iff $\{P_n\}$ converges weakly to P.

We shall prove the following

THEOREM 1. *Let $\mathscr{M}_i = (\mathscr{X}, \mathscr{A}, \mathscr{P}_i)$, $i = 1, 2$, be such that $\mathscr{P}_2 \subset \mathscr{P}_1^\varepsilon$, where $\mathscr{P}_1^\varepsilon$ is the ε-neighbourhood (according to metric L) of $\mathscr{P}_1$. Given a family $\mathscr{F}$ of mappings $\mathscr{X} \to \mathscr{X}'$, let $\mathscr{M}_i' = (\mathscr{X}', \mathscr{A}', \mathscr{P}_i')$ be $\mathscr{F}$-derived from $\mathscr{M}_i$.*

If all $f \in \mathscr{F}$ are uniformly equicontinuous, then there exists a function $k(\varepsilon) \to 0$ as $\varepsilon \to 0$ such that $\mathscr{P}_2 \subset (\mathscr{P}_1')^{k(\varepsilon)}$.

Proof. It was pointed out to the authors by V. M. Zolotarev that this theorem (as well as the following one) follows from his theorems 1 and 12 (see Zolotarev [8]) and theorem of Strassen (Strassen, [7]). Nevertheless in this case it is simpler (and useful for further reference) to give an independent proof, which uses more direct methods.

Under the assumption of uniform equicontinuity, there exists a function, say $r(\varepsilon)$, with $r(\varepsilon) \to 0$ as $\varepsilon \to 0$, such that, for every $f \in \mathscr{F}$, $\varrho(x, y) < \varepsilon \Rightarrow \varrho'(f(x), f(y)) < r(\varepsilon)$.

We shall show first that for any $f \in \mathscr{F}$ and $H \subset \mathscr{X}'$ we have

$$f^{-1}(H)^\varepsilon \subset f^{-1}(H^{r(\varepsilon)}). \tag{2.1}$$

Indeed, $x \in f^{-1}(H)^\varepsilon \Rightarrow \exists y \in f^{-1}(H), \varrho(x, y) < \varepsilon \Rightarrow f(y) \in H, \varrho(x, y) < \varepsilon \Rightarrow \varrho'(f(x), f(y)) < r(\varepsilon) \Rightarrow f(x) \in H^{r(\varepsilon)} \Rightarrow x \in f^{-1}(H^{r(\varepsilon)})$.

Assume now that $L(P_1, P_2) < \varepsilon$, i.e. $P_1(F) \leqslant P_2(F^\varepsilon)+\varepsilon$ for any closed $F \subset \mathscr{X}$. For any fixed $f \in \mathscr{F}$ let $P_i' = \bar{f}P_i$, $i = 1, 2$. Then for any closed $H \subset \mathscr{X}'$ we can write (since $f^{-1}(H)$ is closed): $P_1'(H) = P_1(f^{-1}(H)) \leqslant P_2(f^{-1}(H)^\varepsilon)+\varepsilon \leqslant P_2(f^{-1}(H^{r(\varepsilon)}))+\varepsilon = P_2'(H^{r(\varepsilon)})+\varepsilon$, which completes the proof since the right-hand side does not depend on the choice of $f \in \mathscr{F}$.

For $\mathscr{F}$-primitive structures, let us first consider the case where $\mathscr{F} = \{f\}$ consists of a single function. We shall find conditions on f under which the inverse images of two measures in $\mathscr{K}'$ which are close one another are also close. Clearly, one has to restrict the considerations to such measures which have nonempty inverse images under f.

We shall prove

THEOREM 2. *Let $L(P_1', P_2') < \varepsilon$ and assume that for some $P_1, P_2 \in \mathscr{K}$ we have $\bar{f}P_i = P_i'$, $i = 1, 2$. If f is one-to-one and f^{-1} is uniformly continuous, then there exists $k(\varepsilon)\searrow 0$ as $\varepsilon \searrow 0$ such that $L(P_1, P_2) \leqslant \max(\varepsilon, k(\varepsilon))$.*

Proof. Put $k(\varepsilon) = \sup\limits_{\varrho'(x,y)<\varepsilon} \varrho(f^{-1}(x), f^{-1}(y))$. By assumption, $k(\varepsilon) \searrow 0$ as $\varepsilon \searrow 0$. We shall first prove the inclusion

$$f^{-1}(f(H)^\varepsilon) \subset H^{k(\varepsilon)}, \tag{2.2}$$

valid for any $H \subset \mathscr{X}$. Indeed, we have

$$\begin{aligned} x \in f^{-1}(f(H)^\varepsilon) &\Rightarrow f(x) \in f(H)^\varepsilon \\ &\Rightarrow \exists z \in f(H),\ \varrho'(f(x), z) < \varepsilon \\ &\Rightarrow \exists y \in H,\ f(y) = z,\ \varrho'(f(x), f(y)) < \varepsilon \\ &\Rightarrow y \in H,\ \varrho(x, y) < k(\varepsilon) \Rightarrow x \in H^{k(\varepsilon)}. \end{aligned}$$

If $L(P_1', P_2') < \varepsilon$, then $P_1'(F) \leqslant P_2'(F^\varepsilon)+\varepsilon$ for all closed $F \subset \mathscr{X}'$. We have then for all H closed in $\mathscr{X}$ $P_1(H) \leqslant P_1(f^{-1}(f(H))) = P_1'(f(H)) \leqslant P_2'(f(H)^\varepsilon)+\varepsilon = P_2(f^{-1}(f(H)^\varepsilon))+\varepsilon \leqslant P_2(H^{k(\varepsilon)})+\varepsilon$. This yields $L(P_1, P_2) \leqslant \max(k(\varepsilon), \varepsilon)$, which proves the theorem.

The difficulties connected with passing from one function f to a family $\mathscr{F}$ of functions for the case of primitive structures are caused by the fact that two measures in $\mathscr{K}'$, even if they have nonempty inverse images under $\mathscr{F}$, may involve entirely different subsets of $\mathscr{F}$.

THEOREM 3. *Suppose that all functions in $\mathscr{F}$ are one-to-one and write*

$$r_{\mathscr{F}}(\varepsilon) = \sup_{\varrho'(x,y)<\varepsilon} \sup_{f,g\in\mathscr{F}} \varrho(f^{-1}(x), g^{-1}(y)). \tag{2.3}$$

If P_1', P_2' have nonempty inverse images $\bar{\mathscr{F}}^{-1}(P_1')$ and $\bar{\mathscr{F}}^{-1}(P_2')$, and $L(P_1', P_2') < \varepsilon$, then $L(P_1, P_2) < \max(\varepsilon, r_{\mathscr{F}}(\varepsilon))$ for all $P_i \in \bar{\mathscr{F}}^{-1}(P_i')$, $i = 1, 2$.

The proof is analogous to that of Theorem 2, except that inclusion (2.2) must be replaced by $f^{-1}(g(H)^\varepsilon) \subset H^{k(\varepsilon)}$, valid for all $f, g \in \mathscr{F}$ and all $H \subset \mathscr{X}$. Indeed, we have

$$\begin{aligned} x \in f^{-1}(g(H)^\varepsilon) &\Rightarrow f(x) \in g(H)^\varepsilon \\ &\Rightarrow \exists z \in g(H),\ \varrho'(f(x), z) < \varepsilon \\ &\Rightarrow \exists y \in H,\ g(y) = z,\ \varrho'(f(x), g(y)) < \varepsilon \\ &\Rightarrow y \in H,\ \varrho(f^{-1}(f(x)), g^{-1}(g(y))) \leqslant r_{\mathscr{F}}(\varepsilon) \\ &\Rightarrow y \in H,\ \varrho(x, y) \leqslant r_{\mathscr{F}}(\varepsilon) \Rightarrow x \in H^{r_{\mathscr{F}}(\varepsilon)}. \end{aligned}$$

Naturally, except for the case where all functions in $\mathscr{F}$ coincide, we shall have $r_{\mathscr{F}}(\varepsilon) > \eta$ for all $\varepsilon > 0$ and some positive η.

As a corollary, one can formulate

THEOREM 4. *If all functions in $\mathscr{F}$ are one-to-one and $(\mathscr{X}, \mathscr{A}, \mathscr{P}_i)$ are $\mathscr{F}$-primitive for $(\mathscr{X}', \mathscr{A}', \mathscr{P}_i')$, $i = 1, 2$, then for any $\varepsilon > 0$ the condition $\mathscr{P}_1' \subset (\mathscr{P}_2')^\varepsilon$ implies $\mathscr{P}_1 \subset \mathscr{P}_2^{\max(\varepsilon, r_{\mathscr{F}}(\varepsilon))}$.*

These theorems show the essential difference of conditions for using the $\mathscr{F}$-derived and $\mathscr{F}$-primitive structures. In some oversimplification, the $\mathscr{F}$-derived structure is useful for constructing robust procedures if the images of measures which are close one to another are also close (compare e.g. Hampel [4]). For $\mathscr{F}$-primitive structures the situation is much more complex, and if $\mathscr{F}$ consists of more than one function, the inverse images of "small" neighbourhoods need not be "small". As will be shown in the subsequent sections, these facts are of some importance for parameters of statistical structures and for decision problems.

3. Parameters of reducible statistical structures

DEFINITION 2. Given $\mathscr{M} = (\mathscr{X}, \mathscr{A}, \mathscr{P})$ and a set Γ, any function $\varphi\colon \mathscr{P} \to \Gamma$ will be called a *parameter* of $\mathscr{M}$ (with values in Γ).

Consider now a family $\mathscr{F}$ of mappings $f\colon \mathscr{X} \to \mathscr{X}'$ and the associated family $\bar{\mathscr{F}}$ of mappings $\bar{f}\colon \mathscr{K} \to \mathscr{K}'$. Let $\mathscr{M}' = (\mathscr{X}', \mathscr{A}', \mathscr{P}')$ be the structure $\mathscr{F}$-derived from $\mathscr{M}$.

DEFINITION 3. We say that $\mathscr{F}$ *preserves parameter* φ *of* $\mathscr{M}$ *in the* $\mathscr{F}$*-derived structure* $\mathscr{M}'$ if $\forall P' \in \mathscr{P}'$, $\forall \bar{f}_1, \bar{f}_2 \in \bar{\mathscr{F}}$, $\forall P_1, P_2 \in \mathscr{P}$

$$\bar{f}_1 P_1 = P' = \bar{f}_2 P_2 \Rightarrow \varphi(P_1) = \varphi(P_2). \tag{3.1}$$

To grasp the intuition behind the concept of preserving parameters, it is best to consider an example of a mapping which does not have this property. Let $\mathscr{X} = R^2$, and let $\mathscr{P}$ be the class of all bivariate normal distributions of the form $N(m, m, \sigma_1, \sigma_2, \varrho)$. Then the mapping φ which assigns the mean to each member of $\mathscr{P}$ is a parameter of $\mathscr{M} = (R^2, \mathscr{A}, \mathscr{P})$. Suppose that $\mathscr{F}$ consists of a single function $f\colon R^2 \to R$, defined as $f(x, y) = x-y$. Such mapping violates condition (3.1). As a consequence, if the observation structure is $\mathscr{M}'$, one cannot make any inference about the mean.

Thus, condition (3.1) is essentially that of preserving "identifiability" of φ as considered in statistical literature.

If (3.1) holds, one can define the parameter $\varphi' = \varphi'[\varphi]$ of $\mathscr{M}'$ with the same set of values Γ as φ, by putting $\varphi'(P') = \gamma$ if for some $P \in \mathscr{P}$ and $f \in \mathscr{F}$ we have $\varphi(P) = \gamma$ and $\bar{f}P = P'$. The existence of such P and f follows from the assumption that $\mathscr{M}'$ is $\mathscr{F}$-derived from $\mathscr{M}$. Condition (3.1) guarantees that this assignment does not depend on the choice of P and f.

Suppose now that we have a parameter $\varphi'\colon \mathscr{P}' \to \Gamma$ in the structure $\mathscr{M}'$.

DEFINITION 4. We say that $\mathscr{F}$ *preserves the parameter* φ' *of* $\mathscr{M}'$ *in the* $\mathscr{F}$*-primitive structure* $\mathscr{M}$ if $\forall P \in \mathscr{P}$ $\forall f_1, f_2 \in \mathscr{F}$

$$\bar{f}_1 P, \bar{f}_2 P \in \mathscr{P}' \Rightarrow \varphi'(\bar{f}_1 P) = \varphi'(\bar{f}_2 P). \tag{3.2}$$

If (3.2) holds, one can define the parameter $\varphi = \varphi[\varphi']$ of $\mathscr{M}$ by putting $\varphi(P) = \gamma$ if for some f we have $\varphi'(\bar{f}P) = \gamma$. Again, the existence of such f is implied by the assumption that $\mathscr{M}$ is $\mathscr{F}$-primitive for $\mathscr{M}'$, and condition (3.2) guarantees that the definition of φ is unambiguous since the value $\varphi(P)$ does not depend on the choice of f.

We have the following simple

PROPOSITION 5. *If the parameter* φ *of* $\mathscr{M}$ *is one-to-one and* $\mathscr{F}$ *preserves* φ *in the* $\mathscr{F}$*-derived structure* $\mathscr{M}'$*, then* $\bar{\mathscr{F}}$ *consists of one-to-one functions only.*

For the case of "multidimensional" parameters, the situation is as follows. Suppose $\varphi\colon \mathscr{P} \to \Gamma_1$ and $\psi\colon \mathscr{P} \to \Gamma_2$ are two parameters of $\mathscr{M}$, and let $h = (\varphi, \psi)\colon \mathscr{P} \to \Gamma_1 \times \Gamma_2$. We have an immediate consequence of Definitions 3 and 4:

PROPOSITION 6. *Parameters* φ *and* ψ *are preserved in* $\mathscr{F}$*-derived structure* $\mathscr{M}'$ *iff* h *has this property. The same holds if* φ', ψ', *and* h' *are parameters of the* $\mathscr{F}$*-primitive structure* $\mathscr{M}'$.

An important situation in which there is a "natural candidate" for a parameter occurs when we have a structure $\mathscr{M}'$ $\mathscr{F}$-derived from $\mathscr{M}$ such that $\forall f, g \in \mathscr{F}$, $\forall P_1, P_2 \in \mathscr{P}$

$$\bar{f}P_1 = \bar{g}P_2 \Rightarrow f = g. \tag{3.3}$$

In this case we may put $\Gamma = \mathscr{F}$ (i.e. members of the family $\mathscr{F}$ serve as values of parameters) and for $P' \in \mathscr{P}'$ define $\varphi(P') = f$ if $P' = \bar{f}P$ for some $P \in \mathscr{P}$. Condition (3.3) guarantees that φ is well defined.

If all $P \in \mathscr{P}$ are dominated by some σ-finite measure λ, one can introduce a natural equivalence among functions $f\colon \mathscr{X} \to \mathscr{X}'$ by putting $f \sim g$ if $\lambda\{x\colon f(x) \neq g(x)\} = 0$. In such a case, one may weaken condition (3.3) by requiring that $\bar{f}P_1 = \bar{g}P_2$ implies $f \sim g$, and define parameter φ with values in $\mathscr{F}/\sim$.

An analogous condition for the possibility of using elements of the family $\mathscr{F}$ as values of parameters in case of $\mathscr{F}$-primitive structures can be formulated as: for all $f \in \mathscr{F}$ and $P \in \mathscr{P}$, if $\bar{f}P \in \mathscr{P}'$ then $\forall g \in \mathscr{F}$ $[\bar{g}P \in \mathscr{P}' \Rightarrow f = g]$. In this case, we can assign to P the unique f for which $\bar{f}P \in \mathscr{P}'$.

From now on, we assume that all spaces of values of parameters are metric spaces with the metrics denoted by ϱ, if necessary with identifying subscripts.

DEFINITION 5. The parameter $\varphi\colon \mathscr{P} \to \Gamma$ is said to be *sensitive* if for any $\gamma \in \Gamma$ and any $\varepsilon > 0$ there exists a function $q(\varepsilon) \searrow 0$ as $\varepsilon \searrow 0$ such that whenever $\varrho(\gamma, \gamma^*) < \varepsilon$ then $\varphi^{-1}(\gamma^*) \subset [\varphi^{-1}(\gamma)]^{q(\varepsilon)}$ (where the neighbourhood on the right-hand side is induced by metric L).

Since the bound $q(\varepsilon)$ is assumed to be independent of γ, one can equivalently say (see Bednarek-Kozek and Kozek [2]) that for any sets $A, B \subset \Gamma$ with $A \subset B^{\varepsilon}$ we have $\varphi^{-1}(A) \subset \varphi^{-1}(B)^{q(\varepsilon)}$. We have then

THEOREM 5. *Let* $\mathscr{M} \underset{\mathscr{F}}{\to\mid} \mathscr{M}'$ *and let* $\varphi\colon \mathscr{P} \to \Gamma$ *be a parameter in* $\mathscr{M}$ *preserved in* $\mathscr{M}'$*. If* φ *is sensitive and all functions in* $\mathscr{F}$ *are uniformly equicontinuous, then* $\varphi'[\varphi]$ *is also sensitive.*

Proof. Let $q(\varepsilon)$ be the function appearing in the definition of sensitivity, and let $k(\varepsilon)$ be the function appearing in the assertion of Theorem 1. Write $\varphi' = \varphi'[\varphi]$; let $\varrho(\gamma, \gamma^*) < \varepsilon$ and suppose that P_1', P_2' are such that $\varphi'(P_1') = \gamma$, $\varphi'(P_2') = \gamma^*$. The last conditions imply that there exist $P_1, P_2 \in \mathscr{P}$ such that $\varphi(P_1) = \gamma$, $\varphi(P_2) = \gamma^*$, and there exist $f_1, f_2 \in \mathscr{F}$ with $\bar{f}_1 P_1 = P_1', \bar{f}_2 P_2 = P_2'$. From the assumption of sensitivity of φ it follows that $L(P_1, P_2) < q(\varepsilon)$. By Theorem 1 we obtain therefore $L(P_1', P_2') < k(q(\varepsilon))$, which was to be proved.

For the $\mathscr{F}$-primitive structures, arguing in a similar way, one can obtain the following

THEOREM 6. *Let $\mathscr{M} \underset{\mathscr{F}}{\mapsto} \mathscr{M}'$ and let $\varphi'\colon \mathscr{P}' \to \Gamma$ be a parameter in $\mathscr{M}'$ which is preserved in $\mathscr{M}$. Assume that φ' is sensitive, and all functions in $\mathscr{F}$ are one-to-one. Let $\varphi = \varphi[\varphi']$. If $\varrho(\gamma, \gamma^*) < \varepsilon$, then for all $P_1, P_2 \in \mathscr{P}$ with $\varphi(P_1) = \gamma$, $\varphi(P_2) = \gamma^*$ we have*

$$L(P_1, P_2) \leqslant \max\big(q(\varepsilon), r_{\mathscr{F}}(q(\varepsilon))\big),$$

where $r_{\mathscr{F}}(\varepsilon)$ is defined by (2.3).

We have therefore

COROLLARY 1. *Under the assumptions of Theorem* 6, *if $\mathscr{F} = \{f\}$ with f^{-1} uniformly continuous, then $\varphi[\varphi']$ is sensitive.*

4. Reducible decision problems

In classical schemes of statistical decision problems, the latter is defined (neglecting the questions of measurability, essential in randomized procedures) as a triple $(\mathscr{M}, D, c)$, where $\mathscr{M}$ is a statistical structure, D is the set of decisions, and $c\colon \mathscr{P} \times \times D \to R$ is a cost (loss) function.

DEFINITION 6. The problem $(\mathscr{M}, D, c)$ is $\mathscr{F}$-reducible to the problem $(\mathscr{M}', D', c')$ if $\mathscr{M} \underset{\mathscr{F}}{\to} \mathscr{M}'$ and for any $f \in \mathscr{F}$ there exists a mapping $h = h(f)\colon D \to D'$ such that h is one-to-one, and

$$\forall P \in \mathscr{P}\ \forall d \in D\ c(P, d) = c'(\bar{f}P, h(d)).$$

The $\mathscr{F}$-derived and $\mathscr{F}$-primitive decision problems are defined as above with the only change that one requires either $\mathscr{M} \underset{\mathscr{F}}{\to\!|}\, \mathscr{M}'$ or $\mathscr{M} \underset{\mathscr{F}}{\mapsto} \mathscr{M}'$.

Observe that if $\mathscr{G}$ is a group of (measurable) mappings $\mathscr{X} \to \mathscr{X}$ such that $(\mathscr{M}, D, c)$ is $\mathscr{G}$-self reducible and $\bar{\mathscr{G}}(\mathscr{P}) = \mathscr{P}$ then the problem $(\mathscr{M}, D, c)$ is invariant under $\mathscr{G}$ in the sense defined for instance by Lehmann [5].

A typical example of applicability of the introduced concepts is the problem of point estimation of a parameter, involving either $\mathscr{F}$-primitive or $\mathscr{F}$-derived structures.

Suppose that we have two structures $\mathscr{M} = (\mathscr{X}, \mathscr{A}, \mathscr{P})$ and $\mathscr{M}' = (\mathscr{X}', \mathscr{A}', \mathscr{P}')$ and a parameter $\varphi\colon \mathscr{P} \to \Gamma$. We consider the statistical decision problem $(\mathscr{M}, \Gamma, c)$, where $c(P, \gamma)$ is the loss function.

Assume that $\mathscr{M}'$ is $\mathscr{F}$-derived from $\mathscr{M}$, and that condition (3.1) holds, so that one can define the parameter $\varphi' = \varphi'[\varphi]$ in $\mathscr{M}'$. Then we can choose c' in such a way that the problem $(\mathscr{M}', \Gamma, c')$ is $\mathscr{F}$-derived from $(\mathscr{M}, \Gamma, c)$ The function c' is obtained by putting $c'(P', \gamma) = c(P, \gamma)$ where $P' = \bar{f}P$ for some $f \in \mathscr{F}$ (so that the function h in Definition 6 is the identity for any f).

The same construction can be carried out for $\mathscr{F}$-primitive structures when condition (3.2) holds.

5. Discussion

The concepts introduced in this paper aim at clarifying the situation existing in the domain of statistical decision problems applied to stochastic processes, especially in time series analysis. The subject is, of course, far from being exhausted, and its further development (to be published in subsequent paper) concerns robustness of decision problems, observability, and the problems of prediction.

The relevance of the results seems to lie in providing a unifying tool suitable for the analysis of all situations which appear in statistical practice.

To borrow for a while the traditional cybernetic terminology, a typical situation in analysis of a stochastic process is that we can observe an "output" process and should like to make inference about "input" process or vice versa. The connection between "output" and "input" is given by one (usually unknown) member of the family $\mathscr{F}$. The transformation of "input" process into "output" need not be treated literally, in the sense that it need not have any counterparts in the modelled reality: it may be a conceptual transformation of a process, real or hypothesized, which leads to a new process more amenable to analysis.

Typically, we want to make some inference, or generally, decisions, in a "difficult" statistical problem, for which there are no established procedures with known properties. This problem is then reduced so that it becomes an $\mathscr{F}$-primitive for an "easy" statistical problem (i.e. such problem that the "appropriate" procedures for it are known). The relevance of conditions such as for instance (3.2) lies in showing how the choice of $\mathscr{F}$ must be constrained in order for reducibility of structures to carry over to reducibility of statistical problems. Furthermore, concepts such as separativeness help to realize that the "good" properties of decision rules concerning $\mathscr{M}'$ (at least in the case of parameter estimation) are not automatically guaranteed in the case of the $\mathscr{F}$-primitive structure $\mathscr{M}$, even if (3.2) holds. A common practice may be roughly described as follows. Suppose that δ' is a decision rule, "good" in the problem $(\mathscr{M}', D', c')$. Then, as a decision rule in the $\mathscr{F}$-primitive problem $(\mathscr{M}, D, c)$ one is temped to take $\delta(x)$ defined as $h^{-1}\delta'(f(x))$, where f is that member of $\mathscr{F}$ which "occurs in reality" and $h = h(f)$ is the function $D \to D'$ appearing in Definition 6. Even if f were known, there is no reason to assume that δ' is a good procedure. The situation is still more complicated by the fact that f, and hence also h, are unknown and must be somehow estimated. In effect, what is usually recommended as a decision rule, is $\hat{h}^{-1}\delta'(\hat{f}(x))$, where $\hat{f}$ is some estimator of f and $\hat{h} = h[\hat{f}]$. For such a rule to be meaningful, let alone good, f it self must be a parameter of the $\mathscr{F}$-primitive structure (this imposes additional constraints on $\mathscr{F}$).

Without checking the conditions of the type presented in the paper, the recommendation of such a decision rule as described above is based on act of faith only.

Acknowledgements. The authors wish to thank their colleagues, especially Dr Jacek Koronacki and Dr Andrzej Kozek for many helpful discussions and suggestions which led to substantial improvements of the paper.

References

[1] J. R. Barra, *Notions fondamentales de statistique mathématique*, Dunod, Paris 1971.

[2] B. Bednarek-Kozek and A. Kozek, *Stability, sensitivity and sensitivity of characterizations*, Preprint No 99 of the Math. Inst. of the Polish Academy of Sciences; also this volume pp. 39–64.

[3] G. E. P. Box, G. M. Jenkins, *Time series analysis, forecasting and control*, Holden Day 1970.

[4] F. R. Hampel, *A general qualitative definition of robustness*, Ann. Math. Statist. 42 (1971). pp. 1887–1896.

[5] E. L. Lehmann, *Testing statistical hypotheses*, J. Wiley, New York 1959.

[6] Ju. Prohorov, *Convergence of random processes and limit theorems in probability theory* (in Russian), Teor. Veroyatnost. i Primenen. 1 (1956), pp. 177–238.

[7] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. 36 (1965), pp. 423–439.

[8] V. M. Zolotarev, *On continuity of stochastic sequences generated by recurrent procedures* (in Russian), Teor. Veroyatnost. i Primenen. 20 (1976), pp. 834–847.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

STABILITY, SENSITIVITY AND SENSITIVITY OF CHARACTERIZATIONS

BOGUSŁAWA BEDNAREK-KOZEK

*Institute of Mathematics, Wroclaw University, Wroclaw, Poland**

and

ANDRZEJ KOZEK

*Institute of Mathematics, Polish Academy of Sciences, Wroclaw Branch, Wroclaw, Poland***

0. Introduction

The term "stability" has a long history. It has been used by Lagrange, Poisson, Poincaré, and Liapunov in problems of mechanics. Ulam [14] discussed the notion of the stability of mathematical theorems from a rather general point of view: "When is it true that by changing 'a little' the hypothesis of a theorem one can still assert that the thesis of the theorem remains true or 'approximately' true?". Ulam decided not to formulate a generally applicable definition of stability and we do not try to do it here, either. However, a review of theorems on the stability shows that there are groups of problems in which the stability can be treated from the same point of view. Here we restrict ourselves to three types of "stabilities" which are related to some properties of transformations of metric spaces. We call them $\delta(\varepsilon)$-*stability*, $\delta^{-1}(\varepsilon)$-*sensitivity* and $\gamma(\varepsilon)$-*sensitivity of characterizations*, respectively. The present paper is strongly inspired by lectures of V. M. Zolotarev given in 1976 in Varna and Warsaw on his approach published in papers [15], [16] and [17]. In particular we use the set-theoretical model of $\gamma(\varepsilon)$-sensitivity of characterizations given in [16] and [17].

Let (X, ϱ_X) and (Y, ϱ_Y) be metric spaces and let f be a function from X into Y. We are concerned with functions which have one of the following properties:

I. $\delta(\varepsilon)$-*stability*

$$f(C^{\delta(\varepsilon)}) \subset f(C)^{\varepsilon]}$$

(intuitively: *similar reasons have similar consequences*);

* Plac Grunwaldzki 2/4, 50-384 Wrocław.

** Kopernika 18, 51-617 Wrocław.