KYBERNETIKA ČÍSLO 4, ROČNÍK 1/1965

Information, ε-Sufficiency and Data Reduction Problems

ALBERT PEREZ

Placing us in the frame of the Bayes model of statistical decision we try to estimate in information-theoretical terms the average (respectively, Bayes) risk change caused by a modification of the probability law in action. Especially there are given some upper estimates of the Bayes risk increase on passing from an initial decision problem to a reduced one resulting from the first by a reduction of the sample space σ-algebra as well as of the parameter space σ-algebra. The concept of *ε-sufficiency*, previously introduced by the author as a natural extension of the concept of sufficiency applied in mathematical statistics, is in a certain sense automatically involved in the above estimates as the decrease of information implied by the reduction.

CONTENTS

1. INTRODUCTION

The present paper is inscribed in my general attempt to investigate the impact of information theory in data reduction problems. This line of investigation originates essentially in my paper [1], devoted to the study of statistical decision problems from the point of view of information theory, where the idea of maximal reduction of the decision space as well as of the information source space, reduction compatible of course with a given quality of decision or transmission, plays a fundamental role throughout the paper.

A further development of these ideas was given in paper [2] by extending, namely, the application of the reduction procedure also to the sample or output signal space, We arrived, thus, at a new concept, the concept of ε-*sufficiency*, representing a natural generalization of the notion of sufficiency and of sufficient statistics so fundamental in mathematical statistics. A general exposé of my conceptual model concerning data reduction as developed in papers [1] and [2] is contained in papers [3, 4, 5].

From the intuitive point of view it is clear that at the base of every data reduction problem lies (in an explicit or implicit manner) some decision problem or some class of such problems. The tendency to reduce the data to be processed in decision making arises from the fact that the different "capacities", "memories" and "delays" at our disposal for data processing are always limited.

Obviously, every reduction of data, provided that the data are always exploited in the best way, is directed against the quality of the corresponding decision procedure. Thus, every reduction of data to be admissible must be compatible with the possibility to attain the required or prescribed quality of decision.

When the data at our disposal are insufficient for obtaining the required decision quality in spite of the optimization of the decision procedure, there arises an opposite tendency to enlarge the sphere of data (constitution). In real decision processes both the procedures of constitution and reduction are applied in an alternate sequence in order to obtain the required quality of decision on the base of the most economic set of data.

Just the same requirements and the same limitations led to the development of information theory, what is not always sufficiently appreciated in spite of the excellent example of communication theory.

Theorems of Shannon's type on the transmissibility of an information source through a communication channel represent essentially particular criteria on the possibilities of decision, i.e. on the quality of decision or transmission obtainable in the decision problem under consideration, not requiring of course the prior solution of the problem. What we expect at the first place from information theory is, namely, to formulate such criteria on the possibilities of decision which would apply to a more and more large class of types of decision problems.

As regards in particular data reduction problems, the ideal would be to be able to establish the admissibility or inadmissibility of a given version of reduction by applying on the respective decision problem the adequate criteria not requiring the prior solution of the problem. Otherwise the procedures of constitution and reduction could soon become too complicated and difficult to dominate if only because of the too great number of variants which must be usually explored.

For the present, except of some few cases, namely if the fundamental equipartition property or information stability takes place (see, for instance, [6]), the existence of such criteria is rather a postulate than the true state of affairs. However, I think that information theory is destined in its development to play a growing role in the indicated direction.

In the present paper some simple criteria of admissibility shall be presented concerning reductions of the sample space σ-algebra as well as of the parameter space σ-algebra. We are placed in the frame of the classical model of statistical decision as considered in the theory of statistical decision functions. The existence of a probability law on the Cartesian product of the two spaces above, governing the decision problem under consideration, is supposed. However, the corresponding a priori probability distribution may be, for instance, known only on a reduced parameter space σ-algebra while the corresponding system of probability distributions on the sample space is supposed to be known. Possible deviations of the probability law and of the weight function from the supposed ones could also be taken in account so that it would be possible at the same time to derive criteria concerning the stability of the applied decision procedures, but in this paper we shall not insist on this question.

It is remarkable that the concept of ε-sufficiency intervenes in the formulation of these criteria automatically and plays a fundamental role.

2. INFORMATION AND ε-SUFFICIENCY

Let us consider the classical model of statistical decision introduced in the *theory of statistical decision functions* [7, 8].

Let, thus, $(X, \mathfrak{X})$ be the *measurable parameter space*, where $\mathfrak{X}$ is the *input* or *parameter space σ-algebra*, i.e. a σ-algebra of subsets of the set X of the possible values, in general abstract, of the parameter.

Let, similarly, $(Y, \mathfrak{Y})$ be the *measurable output* or *sample space.*

Let, further, $\{P_{Y|x}, x \in X\}$ be the class we consider of the probability distributions on the measurable sample space $(Y, \mathfrak{Y})$. As ordinary, it will be supposed that for every set $F \in \mathfrak{Y}$ the function $P_{Y|x}(F)$ of $x \in X$ is $\mathfrak{X}$-measurable. In the terminology of information theory we then say that the observation *channel* $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$ is given, through which the input space $(X, \mathfrak{X})$ is observed by directly observing the output space Y at the precision or detail characterized by the output or sample space σ-algebra $\mathfrak{Y}$.

Besides the observation channel $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$, a probability distribution P_X on the measurable parameter space will be supposed to be in action, representing the so-called *a priori probability distribution* in the terminology of mathematical statistics, or the *information source* in the terminology of information theory, the latter being, of course, applied directly, i.e. without coding (other than the identical one) on the input space $(X, \mathfrak{X})$ of the channel.

The information source together with the observation channel generate on the Cartesian product $(X \times Y, \mathfrak{X} \times \mathfrak{Y})$ of the measurable input and output spaces a *double source*, i.e. a joint probability distribution P_{XY}, defined as the unique extension on all the sets of the Cartesian product σ-algebra $\mathfrak{X} \times \mathfrak{Y}$ of the set function

$$P_{XY}(E \times F) = \int_E P_{Y|x}(F)\, \mathrm{d}P_X(x)\,, \quad E \in \mathfrak{X}\,, \quad F \in \mathfrak{Y}\,. \tag{2.1}$$

In the sequel we shall suppose that always one such probability distribution is in action but not necessarily completely known.

Let, finally, $(D, \mathfrak{D})$ be the *measurable decision space* of the decision problem under consideration and $w(x, d)$, $x \in X$, $d \in D$, the *weight* or *loss function*, i.e. a $\mathfrak{X} \times \mathfrak{D}$-measurable nonnegative function serving as a measure of the "loss" implied by taking a decision d while x is the realized value of the parameter at the input. By *decision function* or *decision procedure* we understand a function $b(y)$ defined on the sample space Y and taking its values in the decision space D which is $\mathfrak{Y}$-measurable and, eventually, satisfies some further conditions. The set of all possible decision functions b will be denoted by $\mathscr{B}$.

By *risk* corresponding to the parameter value $x \in X$ and to the decision procedure $b \in \mathscr{B}$ we understand the average value of the loss corresponding to them, i.e. the function

$$R(x, b) = E_x\{w(x, b(y))\} \tag{2.2}$$

where by $E_x\{\ldots\}$ we denote the expectation of the quantity in brackets taken with respect to the probability distribution $P_{Y|x}$.

In the sequel we shall use the *Bayes principle* of ordering of the decision procedures, i.e. that based on the concept of *average risk*

$$r(P_{XY}, b) = \int R(x, b)\, \mathrm{d}P_X(x) \tag{2.3}$$

corresponding to the decision function b and to the a priori probability distribution P_X.

By *optimal* or *Bayes decision procedure* (if exists) we understand a decision function $b_0 \in \mathscr{B}$ which minimizes the average risk. By *Bayes risk* we understand the quantity

$$r_0(P_{XY}) = \inf_{b \in \mathscr{B}} r(P_{XY}, b)\,. \tag{2.4}$$

Definition of information: Let $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, P_{XY})$ be a probability space of the Cartesian product type and let P_X and P_Y be the marginal distributions induced by P_{XY} on the measurable spaces $(X, \mathfrak{X})$ and $(Y, \mathfrak{Y})$, respectively.

Then the corresponding *information* $I(P_{XY})$ is defined as follows:

$$\begin{aligned} I(P_{XY}) &= \int \log f(x, y)\, \mathrm{d}P_{XY}(x, y) && \text{if} \quad P_{XY} \ll P_X \times P_Y \\ I(P_{XY}) &= \infty && \text{if} \quad P_{XY} \not\ll P_X \times P_Y \end{aligned} \tag{2.5}$$

where $f(x, y)$ is the Radon-Nikodym density of the joint probability measure P_{XY} with respect to the product measure $P_X \times P_Y$.

In the sequel we shall suppose that the information is *finite* so that P_{XY} will be absolutely continuous with respect to $P_X \times P_Y$, i.e. $P_{XY} \ll P_X \times P_Y$. Hence the information will be expressed by the first formula (2.5). Moreover, the absolute continuity above implies the existence of a regular version of the conditional probability $P_{Y|x}$, i.e. of the channel $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$ corresponding to P_{XY} (see [9], p. 198).

Let $\mathfrak{X}' \subset \mathfrak{X}$ and $\mathfrak{Y}' \subset \mathfrak{Y}$ be sub- σ-algebras of the σ-algebras $\mathfrak{X}$ and $\mathfrak{Y}$, respectively, and let P'_{XY} be the restriction of P_{XY} on $\mathfrak{X}' \times \mathfrak{Y}'$, P'_X and P'_Y the corresponding marginal distributions and $f'(x, y)$ the density of P'_{XY} with respect to $P'_X \times P'_Y$.

For the corresponding information $I(P'_{XY})$ we have [9]

$$I(P'_{XY}) \leqq I(P_{XY}), \tag{2.6}$$

the sign of equality taking place if, and only if, the reduced σ-algebra $\mathfrak{X}' \times \mathfrak{Y}'$ is *sufficient* (in the sense of mathematical statistics) with respect to the system of measures $\{P_{XY}, P_X \times P_Y\}$, i.e. if, and only if, there exists a version of the density $f'(x, y)$ which is equal to $f(x, y)$ or, what is the same, if, and only if, $f(x, y)$ is not only $\mathfrak{X} \times \mathfrak{Y}$-measurable but also $\mathfrak{X}' \times \mathfrak{Y}'$-measurable.

In the special case where only $\mathfrak{Y}$ is reduced to $\mathfrak{Y}'$, for example, this condition for the equality is equivalent to the condition of sufficiency of $\mathfrak{Y}'$ with respect to the system $\{P_{Y|x}, x \in X, [P_X]\}$ of conditional probability distributions of the observation channel $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$ corresponding to P_{XY}.* It is well known that the reduction of the initial sample σ-algebra $\mathfrak{Y}$ to a sufficient sub-σ-algebra $\mathfrak{Y}'$ has no influence on the quality of decision obtainable. In particular, the Bayes risk r'_0 corresponding to $\mathfrak{Y}'$, i.e. taken with respect to all the decision procedures which are $\mathfrak{Y}'$-measurable, is in the case of sufficiency equal to the Bayes risk r_0, obtained before reduction.

The concept of ε-sufficiency. We shall say that the reduced σ-algebra $\mathfrak{X}' \times \mathfrak{Y}'$, where $\mathfrak{X}' \subset \mathfrak{X}$ and $\mathfrak{Y}' \subset \mathfrak{Y}$, is *$\varepsilon$-sufficient* ($\varepsilon \geqq 0$) with respect to the system of measures $\{P_{XY}, P_X \times P_Y\}$ if

$$I(P_{XY}) - I(P'_{XY}) \leqq \varepsilon, \tag{2.7}$$

where as before P'_{XY} is the measure induced by P_{XY} on $\mathfrak{X}' \times \mathfrak{Y}'$ and $I(P_{XY})$ and $I(P'_{XY})$ are the informations introduced above (see, in particular, (2.6)).

This definition of ε-sufficiency, introduced first in [2], is a natural generalization of the concept of sufficiency introduced in mathematical statistics, to which it reduces for $\varepsilon = 0$. Clearly, it would be possible in defining the concept of ε-sufficiency to part from other definitions of information, namely, the *f-informations* of Rényi and Csiszár [12], which also have the property to be conserved only with respect to sufficient transformations and other analogous properties as the information defined by (2.5). However, among them only the latter has the well-known fundamental properties in information theory related to the *additivity properties of the informa-*

* By $[P_X]$ we understand as usually: *with the possible exception of a measurable set of measure P_X equal to zero.*

tion density $\log f(x, y)$ (see (2.5)) corresponding to it. In any case it would be interesting to try to improve the different criteria of admissibility of reduction by using the other definitions of information.

Let us define on the σ-algebra $\mathfrak{X} \times \mathfrak{Y}$ the probability measure $\tilde{P}^0_{XY}$ by

$$(2.8) \qquad \tilde{P}^0_{XY}(G) = \int_G f' \, \mathrm{d}P_X \times P_Y \,, \quad G \in \mathfrak{X} \times \mathfrak{Y} \,,$$

where f' is the density of P'_{XY} with respect to $P'_X \times P'_Y$.

The probability measure P_{XY} is *absolutely continuous* with respect to $\tilde{P}^0_{XY}$, defined by (2.8), i.e.

$$(2.9) \qquad P_{XY} \ll \tilde{P}^0_{XY}\,(\mathfrak{X} \times \mathfrak{Y}) \,.$$

Indeed, let B be the following set: $B = \{(x, y) : f'(x, y) = 0\}$. Then from (2.8) we obtain

$$\tilde{P}^0_{XY}(B) = P_{XY}(B) = 0 \,,$$

since on $\mathfrak{X}' \times \mathfrak{Y}'$ the two measures P_{XY} and $\tilde{P}^0_{XY}$ reduce both to P'_{XY} and $B \in \mathfrak{X}' \times \mathfrak{Y}'$. If now for some set $E \in \mathfrak{X} \times \mathfrak{Y}$ we have $P_{XY}(E) > 0$, then

$$P_{XY}(E) = P_{XY}(E - B) + P_{XY}(E \cap B) = P_{XY}(E - B) > 0 \,,$$

so that, according to $P_{XY} \ll P_X \times P_Y$, it also holds

$$P_X \times P_Y(E - B) > 0 \,.$$

But then

$$\tilde{P}^0_{XY}(E) = \tilde{P}^0_{XY}(E - B) = \int_{E-B} f' \, \mathrm{d}P_X \times P_Y > 0$$

and thus (2.9) is proved. The corresponding density is given by

$$(2.10) \qquad \frac{\mathrm{d}P_{XY}}{\mathrm{d}\tilde{P}^0_{XY}} = \frac{f}{f'} \,,$$

where f is the density of P_{XY} with respect to $P_X \times P_Y$ and f' that of P'_{XY} with respect to $P'_X \times P'_Y$.

For the *generalized entropy* $H_{\tilde{P}^0_{XY}}(P_{XY})$ of P_{XY} with respect to $\tilde{P}^0_{XY}$, defined as in [9] but with inversed sign, i.e.

$$(2.11) \qquad H_{\tilde{P}^0_{XY}}(P_{XY}) = \int \log \frac{\mathrm{d}P_{XY}}{\mathrm{d}\tilde{P}^0_{XY}} \, \mathrm{d}P_{XY} \,,$$

the following relation takes place according to (2.10)

$$(2.12) \qquad H_{\tilde{P}^0_{XY}}(P_{XY}) = \int \log \frac{f}{f'} \, \mathrm{d}P_{XY} = I(P_{XY}) - I(\tilde{P}^0_{XY}) = I(P_{XY}) - I(P'_{XY})$$

since, in particular, it holds

$$I(\tilde{P}^0_{XY}) = I(P'_{XY}) . \tag{2.13}$$

In the case of ε-sufficiency (see (2.7)) *the following two inequalities take place for the total variation of the pair of probability measures* P_{XY} *and* $\tilde{P}^0_{XY}$,

$$\int |f - f'| \, \mathrm{d}P_X \times P_Y \leqq 2\varepsilon + \Gamma \sqrt{\varepsilon} , \tag{2.14}$$

$$\int |f - f'| \, \mathrm{d}P_X \times P_Y \leqq 2\varepsilon + c \tag{2.15}$$

where Γ *and* c *are two positive universal constants.* (Γ is the constant introduced by Pinsker in [10]; Γ and c can be taken equal to 10 and $2\mathrm{e}^{-1}$ respectively. We use *natural logarithms* in calculating informations according to (2.5)).

These inequalities, proved first in [2], may be derived directly from the following more general entropy version

Theorem 2.1. *Let* ω *and* w *be two probability measures on a measurable space* $(Z, \mathfrak{Z})$ *and let* $\omega \ll w$, *the corresponding density being* f.

If $\mathfrak{Z}' \subset \mathfrak{Z}$ *is a sub-σ-algebra of the σ-algebra* $\mathfrak{Z}$ *and* f' *is the density of* ω' *with respect to* w', *where* ω' *and* w' *are the restrictions on* $\mathfrak{Z}'$ *of* ω *and* w, *respectively, then the following inequalities take place* (*natural logarithms*)

$$\int |f - f'| \, \mathrm{d}w \leqq 2 \int \log \frac{f}{f'} \, \mathrm{d}\omega + \Gamma \sqrt{\left[\int \log \frac{f}{f'} \, \mathrm{d}\omega\right]} , \tag{2.16}$$

$$\int |f - f'| \, \mathrm{d}w \leqq 2 \int \log \frac{f}{f'} \, \mathrm{d}\omega + c , \tag{2.17}$$

where Γ *and* c *are the positive constants introduced above. By defining the probability measure* $\tilde{\omega}$ *on* $\mathfrak{Z}$ *as in* (2.8), *i.e. by*

$$\tilde{\omega}(G) = \int_G f' \, \mathrm{d}w , \quad G \in \mathfrak{Z} , \tag{2.18}$$

we have $\omega \ll \tilde{\omega}$ *with* $\mathrm{d}\omega/\mathrm{d}\tilde{\omega} = f/f'$ *and the following relation takes place*

$$H_w(\omega) - H_{w'}(\omega') = \int \log f \, \mathrm{d}\omega - \int \log f' \, \mathrm{d}\omega' = \int \log \frac{f}{f'} \, \mathrm{d}\omega = H_{\tilde{\omega}}(\omega) . \tag{2.19}$$

Proof. According to [9], pp. 194–195, we have in the case of natural logarithms

$$H_{\tilde{\omega}}(\omega) = \int \log \frac{f}{f'} \, \mathrm{d}\omega = \int_A (f - f') \, \mathrm{d}w + K = 1 - \tilde{\omega}(A) + K , \tag{2.20}$$

where

$$(2.21)\qquad \begin{cases} A = \{z : f(z) > 0\}\,, \\ K = \dfrac{1}{2}\displaystyle\int_A \left(\dfrac{f}{f'} - 1\right)^2 \dfrac{d\tilde{\omega}}{h}\,. \end{cases}$$

Here $\tilde{\omega}$ is the probability measure defined by (2.18) and $h(z)$ is a real function on Z with values lying between $f(z)/f'(z)$ and 1 for $z \in A$.

Similarly, according to [9], relation (5.9′), we obtain

$$(2.22)\qquad \int_A |f - f'|\, dw \leqq \int \left|\log \frac{f}{f'}\right| d\omega + K\,,$$

so that, taking account of (2.20), we have

$$(2.23)\qquad \int |f - f'|\, dw = \int_A |f - f'|\, dw + \int_{Z-A} f'\, dw \leqq$$

$$\leqq \int \left|\log \frac{f}{f'}\right| d\omega + K + 1 - \tilde{\omega}(A) = \int \left|\log \frac{f}{f'}\right| d\omega + \int \log \frac{f}{f'}\, d\omega\,.$$

By applying the Lemma 3.1 of [6], based on the inequality (2.4.11) proved by Pinsker in [10], we can write

$$(2.24)\qquad \int \left|\log \frac{f}{f'}\right| d\omega \leqq \int \log \frac{f}{f'}\, d\omega + \Gamma \sqrt{\left(\int \log \frac{f}{f'}\, d\omega\right)}$$

and

$$(2.25)\qquad \int \left|\log \frac{f}{f'}\right| d\omega \leqq \int \log \frac{f}{f'}\, d\omega + c\,,$$

where Γ and c are the constants introduced above.

On the base of (2.24) we deduce from (2.23) that the total variation of the pair of measures ω and $\tilde{\omega}$ satisfies the inequality

$$(2.26)\qquad \int |f - f'|\, dw \leqq 2 \int \log \frac{f}{f'}\, d\omega + \Gamma \sqrt{\left(\int \log \frac{f}{f'}\, d\omega\right)} = 2H_{\tilde{\omega}}(\omega) + \Gamma \sqrt{(H_{\tilde{\omega}}(\omega))}$$

which coincides with (2.16).

Similarly, on the base of (2.25) we deduce from (2.23) that the total variation above satisfies the inequality

$$(2.27)\qquad \int |f - f'|\, dw \leqq 2 \int \log \frac{f}{f'}\, d\omega + c = 2H_{\tilde{\omega}}(\omega) + c\,,$$

which coincides with (2.17) and, thus, the theorem is proved.

Note. The estimates of the total variation of ω and $\tilde{\omega}$ given by Theorem 2.1 are a little better than the corresponding estimates we could derive by applying the Pinsker's inequality (2.3.14) in [10].

It would be possible to define the concept of ε-sufficiency of $\mathfrak{Z}'$ with respect to the system of measures $\{\omega, w\}$ by the relation

(2.28) $$H_w(\omega) - H_{w'}(\omega') = H_{\tilde{\omega}}(\omega) \leqq \varepsilon\,,$$

which coincides with (2.7) in the special case $(Z, \mathfrak{Z}) = (X \times Y, \mathfrak{X} \times \mathfrak{Y})$, $\mathfrak{Z}' = \mathfrak{X}' \times \mathfrak{Y}'$, $\omega = P_{XY}$, $w = P_X \times P_Y$. It is, of course, well known [9] and it results immediately from (2.19) that

(2.29) $$H_{w'}(\omega') \leqq H_w(\omega)\,,$$

the sign of equality taking place if, and only if, the reduced σ-algebra $\mathfrak{Z}'$ is sufficient with respect to the system of measures $\{\omega, w\}$, as in the case of (2.6).

Obviously, (2.12) is a special case of (2.19), (2.14) is a special case of (2.16) and (2.15) is a special case of (2.17).

As we shall see in the next sections, the concept of ε-sufficiency of a reduced σ-algebra as defined in information terms by (2.7) or, more generally, in generalized entropy terms by (2.28), intervenes in a natural manner in the estimations of the Bayes risk increase caused by the corresponding reduction. In the special case of sufficiency ($\varepsilon = 0$) this increase is zero.

3. ESTIMATION OF AVERAGE RISK CHANGE CAUSED BY A MODIFICATION OF THE PROBABILITY LAW IN A STATISTICAL DECISION PROBLEM

Let us consider, as in section 2, a classical statistical decision problem Π with input (parameter) measurable space $(X, \mathfrak{X})$, output (sample) measurable space $(Y, \mathfrak{Y})$, decision measurable space $(D, \mathfrak{D})$, probability law P_{XY} on the Cartesian product measurable space $(X \times Y, \mathfrak{X} \times \mathfrak{Y})$ of the input and output, and weight (loss) function $w(x, d)$, $x \in X$, $d \in D$. As said in section 2, the weight function is supposed to be nonnegative and $\mathfrak{X} \times \mathfrak{D}$-measurable.

Let, further, b be a decision function either of the pure type, i.e. a measurable transformation of the sample space $(Y, \mathfrak{Y})$ to the decision space $(D, \mathfrak{D})$, or of the mixed type, i.e. a system of probability measures $\{P_{D|y}, y \in Y\}$ on $(D, \mathfrak{D})$ such that, for every set $E \in \mathfrak{D}$, $P_{D|y}(E)$ is an $\mathfrak{Y}$-measurable function of $y \in Y$. In other words, a decision function of the mixed type (randomized decision procedure) is represented by a channel $(\mathfrak{Y}, P_{D|y}, \mathfrak{D})$ (see section 2). In this case, to every sample value $y \in Y$ there corresponds in general not a single decision $d \in D$ but a probability distribution $P_{D|y}$ on $(D, \mathfrak{D})$, so that the final choice of the decision d is made randomly according to $P_{D|y}$. In the sequel, the set of all possible decision functions b will be denoted by $\mathscr{B}$.

As a consequence of the application of the decision function b there is induced by P_{XY} on the Cartesian product $(X \times D, \mathfrak{X} \times \mathfrak{D})$ a probability measure which will be denoted by $P_{XY}b^{-1}$. Thus, the average risk $r(\Pi, b)$, related to the decision problem Π and to the decision function or decision procedure b, is given by

$$(3.1) \qquad r(\Pi, b) = \int_{X \times D} w(x, d)\, \mathrm{d}P_{XY}b^{-1}\,.$$

Let, now, $\tilde{\Pi}$ be a new decision problem differing from the above decision problem Π only in what concerns the probability law in action: in the place of P_{XY} we have now $\tilde{P}_{XY}$ on $\mathfrak{X} \times \mathfrak{Y}$. By applying to $\tilde{\Pi}$ the decision function $\tilde{b}$ there is induced on $(X \times Y, \mathfrak{X} \times \mathfrak{D})$ by $\tilde{P}_{XY}$ a probability measure $\tilde{P}_{XY}\tilde{b}^{-1}$ and the corresponding average risk is given by

$$(3.2) \qquad r(\tilde{\Pi}, \tilde{b}) = \int_{X \times D} w(x, d)\, \mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\,.$$

In the present section we give an upper estimate of the average risk change on passing from the decision problem Π to the decision problem $\tilde{\Pi}$ or conversely, namely, under different conditions concerning the choice of the decision functions b and $\tilde{b}$ applied in the two cases. We can, for instance, take $\tilde{b} = b$ (decision function stability question) or, more generally, $\tilde{b} = bT$, where T is a measurable one-to-one transformation of $(Y, \mathfrak{Y})$ onto $(Y, \mathfrak{Y})$ conserving, thus, the information. We can also consider the Bayes risk change on passing from Π to $\tilde{\Pi}$.

In face of such a task the direct method would be to solve in each case the corresponding decision problem. However, this method, if realizable at all, is not always economic to apply, so that every estimation of the decision possibilities (i.e. of the decision quality attainable) before beginning to solve a decision problem is always desirable.

Lemma 3.1. *For the average risks* (3.1) *and* (3.2) *the following inequalities take place*

$$(3.3) \qquad \begin{aligned} &- \sqrt{(\textstyle\int w^2\, \mathrm{d}P_{XY}b^{-1})}\,\sqrt{[2H_{P_{XY}b^{-1}}(\tilde{P}_{XY}\tilde{b}^{-1})]} \leqq \\ &\leqq r(\tilde{\Pi}, \tilde{b}) - r(\Pi, b) \leqq \sqrt{(\textstyle\int w^2\, \mathrm{d}\tilde{P}_{XY}{}^{-1})}\,\sqrt{[2H_{\tilde{P}_{XY}\tilde{b}^{-1}}(P_{XY}b^{-1})]} \end{aligned}$$

where by $H_w(\omega)$ we denote the generalized entropy of the probability measure ω with respect to the probability measure w, defined as in (2.11) *in the case $\omega \ll w$ and as $+\infty$ otherwise, i.e.*

$$(3.4) \qquad \begin{aligned} H_w(\omega) &= \int \log \frac{\mathrm{d}\omega}{\mathrm{d}w}\, \mathrm{d}\omega \quad \text{for} \quad \omega \ll w\,, \\ H_w(\omega) &= \infty \qquad\qquad \text{for} \quad \omega \not\ll w\,. \end{aligned}$$

Proof. Let us prove the second inequality (3.3). If the generalized entropy $H_{\tilde{P}_{XY}\tilde{b}^{-1}}(P_{XY}b^{-1})$ is infinite this inequality obviously holds. Let, thus, suppose that the generalized entropy above is finite. Then necessarily we have [13] $P_{XY}b^{-1} \ll \tilde{P}_{XY}\tilde{b}^{-1}$, so that, g being the corresponding density, we can write

$$(3.5)\qquad H_{\tilde{P}_{XY}\tilde{b}^{-1}}(P_{XY}b^{-1}) = \int g \log g \,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1} = \\ = 1 - \tilde{P}_{XY}\tilde{b}^{-1}(A) + \frac{1}{2}\int_A \frac{(1-g)^2}{h}\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1},$$

where, similarly as in (2.20) and (2.21), $h(x, d)$ takes its values in the interval between 1 and $g(x, d)$ provided that $(x, d) \in A$ with $A = \{(x, d) : g(x, d) > 0\}$.

Further, by taking $B = \{(x, d) : 1 - g(x, d) > 0\}$ and $A' = X \times D - A$, we obtain on the base of (3.5) successively

$$(3.6)\qquad r(\tilde{\Pi}, \tilde{b}) - r(\Pi, b) = \int_{X\times D} w(x, d)\,\mathrm{d}(\tilde{P}_{XY}\tilde{b}^{-1} - P_{XY}b^{-1}) \leqq \\ \leqq \int_B w(x, d)\,[1 - g(x, d)]\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}(x, d) \leqq \\ \leqq \sqrt{\left(\int_B w^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\right)}\sqrt{\left(\int_B (1-g)^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\right)} = \\ = \sqrt{\left(\int_B w^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\right)}\sqrt{\left[\int_{A\cap B} (1-g)^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1} + \int_{A'\cap B} (1-g)^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\right]} \leqq \\ \leqq \sqrt{\left(\int_B w^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\right)}\sqrt{\left[\int_{A\cap B} \frac{(1-g)^2}{h}\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1} + \tilde{P}_{XY}\tilde{b}^{-1}(A')\right]} \leqq \\ \leqq \sqrt{\left(\int_B w^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\right)}\sqrt{\left[\int_A \frac{(1-g)^2}{h}\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1} + 2\tilde{P}_{XY}\tilde{b}^{-1}(A')\right]} \leqq \\ \leqq \sqrt{\left(\int w^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1}\right)}\sqrt{[2H_{\tilde{P}_{XY}\tilde{b}^{-1}}(P_{XY}b^{-1})]}$$

where $h(x, d)$ is the positive function intervening in (3.5) and, thus, it does not takes values greater than 1 on the set $A \cap B$ since for $(x, d) \in B$ we have $g(x, d) > 1$. Hence, in particular, the third inequality in (3.6) is justified on account of the additional fact that by denoting by A' the complement of the set A we have

$$(3.7)\qquad \int_{A'\cap B} (1-g)^2\,\mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1} = \int_{A'\cap B} \mathrm{d}\tilde{P}_{XY}\tilde{b}^{-1} = \tilde{P}_{XY}\tilde{b}^{-1}(A')\,.$$

Thus, the second inequality (3.3) is proved and in a completely similar manner it is possible to prove the first one, so that the lemma is proved.

Lemna 3.2. *If the decision functions b and $\tilde{b}$ applied in the decision problems Π and $\tilde{\Pi}$ are related by the equality*

$$\tilde{b} = bT, \tag{3.8}$$

where T is a measurable one-to-one transformation of $(Y, \mathfrak{Y})$ onto $(Y, \mathfrak{Y})$, then for the corresponding average risks (3.1) and (3.2) the following inequalities take place

$$-\sqrt{\left(\int w^2 \, \mathrm{d}P_{XY}b^{-1}\right)}\sqrt{[2H_{P_{XY}}(\tilde{P}_{XY}T^{-1})} \leqq r(\tilde{\Pi}, bT)] - \tag{3.9}$$
$$- r(\Pi, b) \leqq \sqrt{\left(\int w^2 \, \mathrm{d}\tilde{P}_{XY}T^{-1}b^{-1}\right)}\sqrt{[2H_{\tilde{P}_{XY}T^{-1}}(P_{XY})]}\,.$$

Proof. In general, it holds (see [9] and [12]) that

$$H_{\tilde{P}_{XY}T^{-1}b^{-1}}(P_{XY}b^{-1}) \leqq H_{\tilde{P}_{XY}T^{-1}}(P_{XY})\,, \tag{3.10}$$
$$H_{P_{XY}b^{-1}}(\tilde{P}_{XY}T^{-1}b^{-1}) \leqq H_{P_{XY}}(\tilde{P}_{XY}T^{-1})\,.$$

By applying on (3.3) the inequalities (3.10) we obtain the inequalities (3.9) and the lemma is, thus, proved.

Theorem 3.1. *Let the weight function $w(x, d)$ be uniformly bounded, i.e. $w(x, d) \leqq \leqq w_0 < \infty$, $x \in X$, $d \in D$.*

If the decision functions b and $\tilde{b}$ applied in the decision problems Π and $\tilde{\Pi}$ are related by (3.8), i.e. if $\tilde{b} = bT$, where T is a measurable one-to-one transformation of $(Y, \mathfrak{Y})$ onto $(Y, \mathfrak{Y})$, then for the corresponding average risks (3.1) and (3.2) the following inequalities take place

$$-\sqrt{[2w_0 r(\Pi, b)\, H_{P_{XY}}(\tilde{P}_{XY}T^{-1})]} \leqq r(\tilde{\Pi}, bT) - \tag{3.11}$$
$$- r(\Pi, b) \leqq \sqrt{[2w_0 r(\tilde{\Pi}, bT)\, H_{\tilde{P}_{XY}T^{-1}}(P_{XY})]}\,.$$

If $r_0(\Pi) = \inf_{b \in \mathscr{B}} r(\Pi, b)$ and $r_0(\tilde{\Pi}) = \inf_{b \in \mathscr{B}} r(\tilde{\Pi}, b)$ are the Bayes risks corresponding to the decision problems Π and $\tilde{\Pi}$, respectively, then it holds

$$-\sqrt{[2w_0 r_0(\Pi) \inf_T H_{P_{XY}}(\tilde{P}_{XY}T^{-1})]} \leqq r_0(\tilde{\Pi}) - r_0(\Pi) \leqq \tag{3.12}$$
$$\leqq \sqrt{[2w_0 r_0(\tilde{\Pi}) \inf_T H_{\tilde{P}_{XY}T^{-1}}(P_{XY})]}\,.$$

Proof. Since by hypothesis $w(x, d) \leqq w_0$, $x \in X$, $d \in D$, it follows that

$$\int w^2 \, \mathrm{d}P_{XY}b^{-1} \leqq w_0 r(\Pi, b)\,, \tag{3.13}$$

$$\int w^2 \, \mathrm{d}\tilde{P}_{XY}T^{-1}b^{-1} \leqq w_0 r(\tilde{\Pi}, bT)\,.$$

By applying on (3.9) of Lemma 3.2 the inequalities (3.13) we immediately deduce the inequalities (3.11).

Let us now prove the second inequality (3.12). From the second inequality (3.11) it follows that for every decision function $b \in \mathscr{B}$ for which $r(\tilde{\Pi}, bT) > 0$ it holds

$$\frac{r(\tilde{\Pi}, bT) - r(\Pi, b)}{\sqrt{[r(\tilde{\Pi}, bT)]}} \leqq \sqrt{[2H_{\tilde{P}_{XY}T^{-1}}(P_{XY})\, w_0]}\,. \tag{3.14}$$

This is, in particular, the case if $r_0(\tilde{\Pi}) > 0$. If, on the contrary, $r_0(\tilde{\Pi}) = 0$, then the second inequality (3.12) is trivially fulfilled.

Let us, thus, assume that $r_0(\tilde{\Pi}) > 0$. Then it holds

$$\frac{r_0(\tilde{\Pi}) - r(\Pi, b)}{\sqrt{[r_0(\tilde{\Pi})]}} \leqq \frac{r(\tilde{\Pi}, bT) - r(\Pi, b)}{\sqrt{[r(\tilde{\Pi}, bT)]}} \tag{3.15}$$

for every $b \in \mathscr{B}$, since the function $f(x) = (x - a)/\sqrt{x}$ $(a \geqq 0,\ x > 0)$ is an increasing function of x. Indeed,

$$f'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x - a}{\sqrt{x}}\right) = \frac{\sqrt{x}}{x}\,\frac{x + a}{2x} > 0 \quad \text{for} \quad x > 0\,. \tag{3.16}$$

Combining the inequalities (3.14) and (3.15) we obtain

$$\frac{r_0(\tilde{\Pi}) - r(\Pi, b)}{\sqrt{[r_0(\tilde{\Pi})]}} \leqq \sqrt{[2w_0 H_{\tilde{P}_{XY}T^{-1}}(P_{XY})]} \tag{3.17}$$

for every decision function $b \in \mathscr{B}$ and for every transformation T of the above type. The second inequality (3.12) is an immediate consequence of (3.17). In a similar manner it is possible to prove the first inequality (3.12) and this completes the proof of the theorem.

Theorem 3.1, especially by taking in (3.11) the transformation T equal to the "identical" transformation so that $\tilde{b} = bT = b$, may be applied to problems of stability or sensibility of a decision procedure with respect to different deviations of the probability law from the supposed one. In the sequel we shall apply this theorem and, in particular, relation (3.12) in data reduction problems.

4. REDUCTION OF THE OUTPUT (SAMPLE SPACE) σ-ALGEBRA

Let us consider, as in section 3, the statistical decision problem Π, corresponding to the probability space $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, P_{XY})$, and suppose that the output (sample space) σ-algebra $\mathfrak{Y}$ is reduced to a σ-algebra $\mathfrak{Y}' \subset \mathfrak{Y}$. Let P'_{XY} be the restriction on the Cartesian product σ-algebra $\mathfrak{X} \times \mathfrak{Y}'$ of the probability measure P_{XY} and consider the statistical decision problem Π' resulting from Π and corresponding to the reduced probability space $(X \times Y, \mathfrak{X} \times \mathfrak{Y}', P'_{XY})$.

$(D, \mathfrak{D})$ being the measurable decision space (the same for both the problems) and $w(x, d)$, $x \in X$, $d \in D$, being the weight function (the same for both the problems) let $r(\Pi, b)$ be the average risk (defined as in (3.1)) corresponding to Π and to a decision function $b \in \mathscr{B}$, where $\mathscr{B}$ is the set of all possible $\mathfrak{Y}$-measurable decision functions and let $r(\Pi', b')$ be the average risk corresponding to Π' and to a decision function $b' \in \mathscr{B}'$, where $\mathscr{B}'$ is the set of all possible $\mathfrak{Y}'$-measurable decision functions.

Let, further, $r_0(\Pi) = \inf_{b\in\mathscr{B}} r(\Pi, b)$ be the Bayes risk corresponding to the problem Π and $r_0(\Pi') = \inf_{b'\in\mathscr{B}'} r(\Pi', b')$ that corresponding to the problem Π'.

Our task is to give an upper estimate of the Bayes risk change $r_0(\Pi') - r_0(\Pi)$ resulting from the reduction of the sample σ-algebra $\mathfrak{Y}$ to $\mathfrak{Y}' \subset \mathfrak{Y}$.

It is, however, impossible to apply directly in the present case the inequalities (3.12) of Theorem 3.1, for instance, since the probability measure P'_{XY} is defined on the smaller σ-algebra $\mathfrak{X} \times \mathfrak{Y}'$ and not on the initial σ-algebra $\mathfrak{X} \times \mathfrak{Y}$, as do P_{XY} and $\tilde{P}_{XY}$.

It is, thus, in the general case necessary to search for such an extension $\tilde{P}_{XY}$ of the measure P'_{XY} to $\mathfrak{X} \times \mathfrak{Y}$ that the reduced σ-algebra $\mathfrak{Y}'$ is *sufficient* with respect to the system of conditional probability distributions $\{\tilde{P}_{Y|x}, x \in X, [P_X]\}$ induced by $\tilde{P}_{XY}$ on $\mathfrak{Y}$ which is $\mathfrak{X}$-measurable in the sense that, for every set $E \in \mathfrak{Y}$, the function $\tilde{P}_{Y|x}(E)$ of $x \in X$ is $\mathfrak{X}$-measurable. In other words, $\tilde{P}_{XY}$ *must be such that on the reduced σ-algebra* $\mathfrak{X} \times \mathfrak{Y}'$ *it holds* $\tilde{P}'_{XY} = P'_{XY}$ *and that the latter is sufficient* (see section 2) *with respect to the system* $\{\tilde{P}_{XY}, \tilde{P}_X \times \tilde{P}_Y\}$ *of measures*, where $\tilde{P}_X = P_X$ and $\tilde{P}_Y$ are the marginal distributions induced by $\tilde{P}_{XY}$ on $\mathfrak{X}$ and $\mathfrak{Y}$, respectively.

As we have seen before in section 2, this sufficiency takes place if, and only if, the information $I(\tilde{P}_{XY})$ corresponding to the probability space $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, \tilde{P}_{XY})$ is equal to the information $I(\tilde{P}'_{XY}) = I(P'_{XY})$, corresponding to the reduced probability space $(X \times Y, \mathfrak{X} \times \mathfrak{Y}', \tilde{P}'_{XY}) = (X \times Y, \mathfrak{X} \times \mathfrak{Y}', P'_{XY})$, i.e.

$$I(\tilde{P}_{XY}) = I(\tilde{P}'_{XY}) = I(P'_{XY}) . \tag{4.1}$$

The condition of sufficiency above or, equivalently, the condition of conservation of information (4.1) is necessary (at least in some cases; see [8], [11]), and sufficient for the conservation of the Bayes risk on passing from the decision problem Π' to the decision problem $\tilde{\Pi}$ resulting from Π' by extending $\mathfrak{Y}'$ to $\mathfrak{Y}$ and P'_{XY} to $\tilde{P}_{XY}$, so that in particular also $\mathscr{B}'$ is extended to $\mathscr{B}$, i. e. for the validity of the equality

$$(4.2) \qquad r_0(\tilde{\Pi}) = r_0(\Pi') .$$

Provided that $I(P'_{XY}) < \infty$, the general form of an extension $\tilde{P}_{XY}$ of P'_{XY} to $\mathfrak{X} \times \mathfrak{Y}$ satisfying (4.1) is given by the following theorem.

Theorem 4.1. *Let $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, \omega)$ be a probability space of the Cartesian product type and let ω' be the restriction of ω to the sub-σ-algebra $\mathfrak{X}' \times \mathfrak{Y}'$, where $\mathfrak{X}' \subset \mathfrak{X}$ and $\mathfrak{Y}' \subset \mathfrak{Y}$.*

Let, further, $I(\omega')$ be the information corresponding to the reduced probability space $(X \times Y, \mathfrak{X}' \times \mathfrak{Y}', \omega')$ and suppose that $I(\omega') < \infty$, so that $\omega' \ll \mu' \times \nu'$, where μ' and ν' are the marginal measures induced by ω' on $\mathfrak{X}'$ and $\mathfrak{Y}'$, respectively.

Then the general form of every extension $\tilde{\omega}$ of ω' from $\mathfrak{X}' \times \mathfrak{Y}'$ to $\mathfrak{X} \times \mathfrak{Y}$ conserving the information, i.e. for which it holds $I(\tilde{\omega}) = I(\omega')$, is given by

$$(4.3) \qquad \tilde{\omega}(G) = \int_G f' \, \mathrm{d}\tilde{\mu} \times \tilde{\nu} , \quad G \in \mathfrak{X} \times \mathfrak{Y} ,$$

where $f' = \mathrm{d}\omega'/(\mathrm{d}\mu' \times \nu')$ is the Radon-Nikodym density of ω' with respect to the product measure $\mu' \times \nu'$ and $\tilde{\mu}$ and $\tilde{\nu}$ are arbitrary extensions from $\mathfrak{X}'$ to $\mathfrak{X}$ of the measure μ' and from $\mathfrak{Y}'$ to $\mathfrak{Y}$ of the measure ν', respectively.

Proof. The conditions $I(\tilde{\omega}) = I(\omega')$ and $I(\omega') < \infty$ imply that also $I(\tilde{\omega}) < \infty$ and, consequently (see section 2), $\tilde{\omega} \ll \tilde{\mu} \times \tilde{\nu}$. Let $\tilde{f}(x, y)$, $x \in X$, $y \in Y$, be the corresponding density.

The condition $I(\tilde{\omega}) = I(\omega')$ further implies that the density $\tilde{f}(x, y)$ is $\mathfrak{X}' \times \mathfrak{Y}'$-measurable (see section 2) and, thus, there exists such a version of $\tilde{f}(x, y)$ which satisfies the equality

$$(4.4) \qquad \tilde{f}(x, y) = f'(x, y) , \quad x \in X , \quad y \in Y ,$$

due to the fact that, by hypothesis, the restriction of $\tilde{\omega}$ on $\mathfrak{X}' \times \mathfrak{Y}'$ coincides with ω'.

From (4.4) it immediately follows that the general form of $\tilde{\omega}$ is given by (4.3), where f' is the density of ω' with respect to $\mu' \times \nu'$ and $\tilde{\mu}$ and $\tilde{\nu}$ are arbitrary extensions of μ' and ν' on $\mathfrak{X}$ and $\mathfrak{Y}$, respectively.

Indeed, if $\tilde{\omega}$ is of the form (4.3), then its restriction on $\mathfrak{X}' \times \mathfrak{Y}'$ coincides with ω' and, due to the $\mathfrak{X}' \times \mathfrak{Y}'$-measurability of the density $\mathrm{d}\tilde{\omega}/(\mathrm{d}\tilde{\mu} \times \tilde{\nu}) = \tilde{f} = f'[\tilde{\omega}]$, its information $I(\tilde{\omega})$ is equal to $I(\tilde{\omega}') = I(\omega')$ and, thus, (4.1) is satisfied. This completes the proof of the theorem.

Lemma 4.1. *Conserving the notations and definitions of Theorem* 4.1, *let us suppose that $I(\omega) < \infty$, from which it follows that also $I(\omega') < \infty$.*

Then the relation of absolute continuity

$$(4.5) \qquad \omega \ll \tilde{\omega}$$

 holds if, and only if,

$$\mu \ll \tilde{\mu} \quad and \quad \nu \ll \tilde{\nu}, \tag{4.6}$$

where μ and ν are the marginals of ω on $\mathfrak{X}$ and $\mathfrak{Y}$, respectively, and $\tilde{\omega}$ is given by (4.3).

The density of ω with respect to $\tilde{\omega}$ is then given by

$$\frac{d\omega}{d\tilde{\omega}} = \frac{d\omega}{d\tilde{\mu} \times \tilde{\nu}} \Big/ \frac{d\omega'}{d\mu' \times \nu'}. \tag{4.7}$$

Proof. Let, as in proving (2.9), B be the set $\{(x, y) : f'(x, y) = 0\}$. Then from (4.3) we obtain

$$\tilde{\omega}(B) = \omega(B) = 0$$

since on $\mathfrak{X}' \times \mathfrak{Y}'$ the two measures ω and $\tilde{\omega}$ reduce both to ω' and $B \in \mathfrak{X}' \times \mathfrak{Y}'$. If now for some set $E \in \mathfrak{X} \times \mathfrak{Y}$ we have $\omega(E) > 0$, then

$$\omega(E) = \omega(E - B) + \omega(E \cap B) = \omega(E - B) > 0,$$

so that, according to the relation $\omega \ll \mu \times \nu \ll \tilde{\mu} \times \tilde{\nu}$, following from $I(\omega) < \infty$ and from our hypothesis (4.6), it also holds

$$\tilde{\mu} \times \tilde{\nu}(E - B) > 0.$$

But then

$$\tilde{\omega}(E) = \tilde{\omega}(E - B) = \int_{E-B} f' \, d\tilde{\mu} \times \tilde{\nu} > 0$$

and, thus, (4.5) is proved, the corresponding density being given by (4.7). On the other hand, the necessity of the condition (4.6) follows directly from the fact that (4.5) implies (4.6). Thus, the lemma is proved.

Theorem 4.2. *Conserving the notations and definitions of Theorem* 4.1, *let us suppose that $I(\omega) < \infty$, from which it follows that also $I(\omega') < \infty$.*

Let us denote by $\tilde{\omega}_0$ that probability measure on $\mathfrak{X} \times \mathfrak{Y}$ which results from (4.3) *by taking $\tilde{\mu} = \mu$ and $\tilde{\nu} = \nu$, where μ and ν are the marginals of ω on $\mathfrak{X}$ and $\mathfrak{Y}$, respectively*, i.e.

$$\tilde{\omega}_0(G) = \int_G f' \, d\mu \times \nu, \quad G \in \mathfrak{X} \times \mathfrak{Y}. \tag{4.8}$$

Let us further denote by $\tilde{\Omega}$ the set of all the probability measures $\tilde{\omega}$ on $\mathfrak{X} \times \mathfrak{Y}$ of the form (4.3), *i.e. the set of all the probability measures on $\mathfrak{X} \times \mathfrak{Y}$ having a restriction on $\mathfrak{X}' \times \mathfrak{Y}'$ equal to ω' and conserving the information $I(\omega')$ of ω': $I(\tilde{\omega}) = I(\tilde{\omega}') = I(\omega')$.*

Then for the generalized entropy $H_{\tilde{\omega}_0}(\omega)$ of ω with respect to $\tilde{\omega}_0$ (see definition (3.4)) *the following relation takes place*

$$H_{\tilde{\omega}_0}(\omega) = I(\omega) - I(\omega') = \min_{\tilde{\omega}\in\tilde{\Omega}} H_{\tilde{\omega}}(\omega)\,. \tag{4.9}$$

Proof. For every pair of extensions $\tilde{\mu}$, $\tilde{\nu}$ of μ', ν', through which is defined by (4.3) some $\tilde{\omega} \in \tilde{\Omega}$ having $\tilde{\mu}$ and $\tilde{\nu}$ as marginals, there are two possibilities: either (4.6) holds, i.e. $\mu \ll \tilde{\mu}$ and $\nu \ll \tilde{\nu}$, or not. According to Lemma 4.1, in the first case we have $\omega \ll \tilde{\omega}$ with

$$\frac{d\omega}{d\tilde{\omega}} = \frac{d\omega}{d\tilde{\mu} \times \tilde{\nu}} \Big/ \frac{d\omega'}{d\mu' \times \nu'} \tag{4.10}$$

and in the second case we have $\omega \not\ll \tilde{\omega}$, hence $H_{\tilde{\omega}}(\omega) = \infty$.

Let us calculate the generalized entropy $H_{\tilde{\omega}}(\omega)$ in the first case. Besides (4.10) it holds

$$\frac{d\omega}{d\tilde{\mu} \times \tilde{\nu}} = \frac{d\omega}{d\mu \times \nu}\frac{d\mu \times \nu}{d\tilde{\mu} \times \tilde{\nu}} = \frac{d\omega}{d\mu \times \nu}\frac{d\mu}{d\tilde{\mu}}\frac{d\nu}{d\tilde{\nu}}, \tag{4.11}$$

the existence of the densities $d\omega/(d\mu \times \nu)$ and $d\omega'/(d\mu' \times \nu')$ being assured on the base of the assumption $I(\omega) < \infty$ implying $I(\omega') < \infty$.

From (4.10) and (4.11) we derive

$$\begin{aligned} H_{\tilde{\omega}}(\omega) &= \int \log\frac{d\omega}{d\tilde{\omega}}\, d\omega = \int \log\frac{d\omega}{d\tilde{\mu} \times \tilde{\nu}}\, d\omega - \int \log\frac{d\omega'}{d\mu' \times \nu'}\, d\omega = \\ &= \int \log\frac{d\omega}{d\mu \times \nu}\, d\omega + \int \log\frac{d\mu}{d\tilde{\mu}}\, d\mu + \int \log\frac{d\nu}{d\tilde{\nu}}\, d\nu - \int \log\frac{d\omega'}{d\mu' \times \nu'}\, d\omega' = \\ &= I(\omega) - I(\omega') + H_{\tilde{\mu}}(\mu) + H_{\tilde{\nu}}(\nu) \geqq I(\omega) - I(\omega') = H_{\tilde{\omega}_0}(\omega)\,. \end{aligned} \tag{4.12}$$

Thus, (4.9) holds and the theorem is proved.

Let us now return to our question concerning the general form of an extension $\tilde{P}_{XY}$ of P'_{XY} to $\mathfrak{X} \times \mathfrak{Y}$ satisfying the relation (4.1) and thus assuring the equality (4.2). The following corollary of Theorems 4.1 and 4.2 gives an answer not only to this question but also to the question: *what is the "better" among these extensions from the point of view of applying inequalities* (3.12) *of Theorem* 3.1 *in estimating the Bayes risk change* $r_0(\Pi') - r_0(\Pi) = r_0(\tilde{\Pi}) - r_0(\Pi)$ *resulting from the reduction of the sample σ-algebra* $\mathfrak{Y}$ *to* $\mathfrak{Y}' \subset \mathfrak{Y}$.

Corollary 4.1. *Conserving the notations and definitions used for the formulation of the main problem of the present section, the general form of an extension* $\tilde{P}_{XY}$ *of* P'_{XY} *to* $\mathfrak{X} \times \mathfrak{Y}$ *satisfying the relation* (4.1), *i.e.* $I(\tilde{P}_{XY}) = I(\tilde{P}'_{XY}) = I(P'_{XY})$, *is given by*

$$\tilde{P}_{XY}(G) = \int_G f'\, dP_X \times \tilde{P}_Y\,, \qquad G \in \mathfrak{X} \times \mathfrak{Y}\,, \tag{4.13}$$

where $f' = \mathrm{d}P'_{XY}/(\mathrm{d}P_X \times P'_Y)$ *is the Radon-Nikodym density of* P'_{XY} *with respect to the product measure* $P_X \times P'_Y$ *and* $\tilde{P}_Y$ *is an arbitrary extension from* $\mathfrak{Y}'$ *to* $\mathfrak{Y}$ *of the measure* P'_Y. *The existence of the density* f' *is assured by the assumption that* $I(P'_{XY}) < \infty$.

If, moreover, $I(P_{XY}) < \infty$, *then*

$$\min_{\tilde{P}_{XY} \in \tilde{\mathscr{P}}} H_{\tilde{P}_{XY}}(P_{XY}) = H_{\tilde{P}^0_{XY}}(P_{XY}) = I(P_{XY}) - I(P'_{XY}) \tag{4.14}$$

where $\tilde{\mathscr{P}}$ *is the set of all the probability measures* $\tilde{P}_{XY}$ *of the type* (4.13) *and* $\tilde{P}^0_{XY}$ *is that special* $\tilde{P}_{XY}$ *which results form* (4.13) *by taking* $\tilde{P}_Y = P_Y$.

We are now in a position to formulate the main result of this section.

Theorem 4.3. *Let* Π *and* Π' *be the two decision problems formulated at the beginning of the present section, the second resulting from the first by a reduction of the sample* σ-*algebra* $\mathfrak{Y}$ *to* $\mathfrak{Y}' \subset \mathfrak{Y}$.

Let the weight function w *be uniformly bounded by* w_0, *i.e.* $w(x, d) \leqq w_0 < \infty$, $x \in X$, $d \in D$, *and suppose that the information* $I(P_{XY})$ *is finite.*

If $r_0(\Pi)$ *and* $r_0(\Pi')$ *are the Bayes risks corresponding to the decision problems* Π *and* Π', *respectively, then*

$$0 \leqq r_0(\Pi') - r_0(\Pi) \leqq \sqrt{\{2w_0 r_0(\Pi') [I(P_{XY}) - I(P'_{XY})]\}} \,. \tag{4.15}$$

In other words, if the reduced σ-*algebra* $\mathfrak{Y}'$ *is* ε-*sufficient with respect to the system* $\{P_{Y|x}, x \in X, [P_X]\}$ *of conditional probability distributions, corresponding to* P_{XY} *(see section* 2), *then*

$$0 \leqq r_0(\Pi') - r_0(\Pi) \leqq \sqrt{[2w_0 \varepsilon\, r_0(\Pi')]} \,. \tag{4.16}$$

Proof. Defining the probability measure $\tilde{P}^0_{XY}$ as in Corollary 4.1, we obtain that simultaneously it holds

(i) $$I(\tilde{P}^0_{XY}) = I(P'_{XY}) \,,$$

(ii) $$r_0(\tilde{\Pi}^0) = r(\Pi') \,, \quad (\tilde{\Pi}^0 = \tilde{\Pi} \text{ for } \tilde{P}_{XY} = \tilde{P}^0_{XY}) \,,$$

(iii) $$H_{\tilde{P}^0_{XY}}(P_{XY}) = I(P_{XY}) - I(P'_{XY}) \,,$$

according to Corollary 4.1, since by hypothesis $I(P_{XY}) < \infty$. As we have seen (see relations (4.1) and (4.2)), the equality (ii) is an immediate consequence of equality (i).

The relation (4.15) follows immediately from (ii), (iii) and Theorem 3.1, second inequality (3.12), where we replace $\tilde{P}_{XY}$ by $\tilde{P}^0_{XY}$ and we take as T the "identical" transformation.

The relation (4.16) is derived from (4.15) on the base of the inequality $I(P_{XY}) - I(P'_{XY}) \leqq \varepsilon$, implied by the definition of ε-sufficiency we suppose. Thus, the theorem is proved.

Remark 1. If, instead of the assumption that the weight function w is uniformly bounded, we suppose that

$$(4.17)\qquad \int w^2 \, \mathrm{d}Q_{XD} \leqq k \left[\int w \, \mathrm{d}Q_{XD} \right]^2$$

for every probability measure Q_{XD} on $\mathfrak{X} \times \mathfrak{D}$ of the type $Q_{XD} = P_{XY}b^{-1}$ resp. $\tilde{P}_{XY}b^{-1}$, where k is a given constant (obviously, $k \geqq 1$), then from Lemma 3.2 it would be possible to derive, by following a similar procedure as that used in proving Theorem 3.1, the inequalities

$$(4.18)\qquad \begin{aligned} - r_0(\Pi) \sqrt{[2k \inf_T H_{P_{XY}}(\tilde{P}_{XY}T^{-1})]} \leqq r_0(\tilde{\Pi}) - r_0(\Pi) \leqq \\ \leqq r_0(\tilde{\Pi}) \sqrt{[2k \inf_T H_{\tilde{P}_{XY}T^{-1}}(P_{XY})]}\,, \end{aligned}$$

instead of (3.12). Indeed, the role of (3.13) is here played by (4.17) and the role of the function $f(x)$ by the function $g(x) = (x - a)/x \, (a \geqq 0, \, x > 0)$, which is also an increasing function of x since, as in the case of (3.16), we have

$$(4.19)\qquad g'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x - a}{x}\right) = \frac{a}{x^2} \geqq 0\,.$$

In the place of Theorem 4.3, relation (4.15), it would be possible then to derive from (4.18) the relation

$$(4.20)\qquad 0 \leqq r_0(\Pi') - r_0(\Pi) \leqq r_0(\Pi') \sqrt{\{2k[I(P_{XY}) - I(P'_{XY})]\}}\,.$$

Remark 2. In [2] we have used the inequalities (2.14) and (2.15), derived as a special case of Theorem 2.1, for proving the following inequalities ($\Delta I = I(P_{XY}) - I(P'_{XY})$)

$$(4.21)\qquad \begin{cases} r_0(\Pi') - r_0(\Pi) \leqq w_0 \,.\, (2\Delta \; + \Gamma \sqrt{\Delta I}) \\ r_0(\Pi') - r_0(\Pi) \leqq w_0 \,.\, (2\Delta I + c) \end{cases}$$

under the assumption that the information $I(P_{XY})$ is finite and that the weight function w is uniformly bounded by w_0.

The proof of (4.21) is made as that of Theorem 4.3. Namely in the place of P'_{XY} we use its extension $\tilde{P}^0_{XY}$ on $\mathfrak{X} \times \mathfrak{Y}$ for which (i), (ii) and (iii) hold. Based on inequalities (2.16) and (2.17) of Theorem 2.1 and further on (3.10) we then successively obtain, for every decision function $b \in \mathscr{B}$,

$$(4.22)\qquad \begin{aligned} |r(\tilde{\Pi}^0, b) - r(\Pi, b)| &\leqq \int_{X \times D} w(x, d) \, \mathrm{d}|\tilde{P}^0_{XY}b^{-1} - P_{XY}b^{-1}| \leqq \\ &\leqq w_0 \int \mathrm{d}|\tilde{P}^0_{XY}b^{-1} - P_{XY}b^{-1}| \leqq \\ &\leqq w_0(2H_{\tilde{P}^0_{XY}b^{-1}}(P_{XY}b^{-1}) + \Gamma \sqrt{[H_{\tilde{P}^0_{XY}b^{-1}}(P_{XY}b^{-1})]}) \leqq \\ &\leqq w_0(2H_{\tilde{P}^0_{XY}}(P_{XY}) + \Gamma \sqrt{[H_{\tilde{P}^0_{XY}}(P_{XY})]}) = w_0(2\Delta I + \Gamma \sqrt{\Delta I}) \end{aligned}$$

and, similarly,

$$|r(\tilde{\Pi}^0, b) - r(\Pi, b)| \leqq w_0(2\Delta I + c)\,. \tag{4.23}$$

The inequalities (4.21) follow directly from (4.22) and (4.23), respectively. Obviously, these estimates are less satisfactory than that given by Theorem 4.3, since in particular $\Gamma \doteq 10$ for natural logarithms we use throughout the paper.

5. REDUCTION OF THE INPUT (PARAMETER SPACE) σ-ALGEBRA

Let us consider, as in sections 3 and 4, the statistical decision problem Π, corresponding to the probability space $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, P_{XY})$ and to the weight function w, and suppose that the input (parameter space) σ-algebra $\mathfrak{X}$ is reduced to a σ-algebra $\mathfrak{X}' \subset \mathfrak{X}$. This situation may arise, for instance, in the case we wish to "eliminate" the influence of a "nuisance parameter" and to conserve only the presence of the "essential parameter" (represented by the reduced σ-algebra $\mathfrak{X}'$). A similar situation may arise if we wish to "compress" (by *precoding*, see [1]) a given information source in such an extent that the rounded off source thus resulting contains just the essential to be transmitted from the point of view of the receiver, i.e. from the point of view of the decision or transmission problem under consideration.

Let P'_{XY} be the restriction on the Cartesian product σ-algebra $\mathfrak{X}' \times \mathfrak{Y}$ of the probability measure P_{XY}.

Let P_X and P'_X be the corresponding marginal distributions on $\mathfrak{X}$ and $\mathfrak{X}'$, respectively, and $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$ and $(\mathfrak{X}', P'_{Y|x}, \mathfrak{Y})$ the corresponding channels, i.e. the systems of conditional probability distributions corresponding to P_{XY} and P'_{XY}, respectively. The existence of these channels is here assured by $I(P_{XY}) < \infty$ (see p. 301).

Obviously, it holds

$$P'_{Y|x}(.) = \underset{P_X}{E}\{P_{Y|x}(.) \mid x, \mathfrak{X}'\}\,[P_x]\,, \tag{5.1}$$

where $\underset{P_X}{E}\{. \mid x, \mathfrak{X}'\}$ denotes the conditional expectation with respect to P_X and to the reduced σ-algebra $\mathfrak{X}'$. The channel $(\mathfrak{X}', P'_{Y|x}, \mathfrak{Y})$ is, thus, a *rounded off version of the initial channel* $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$.

In a similar way we define the *rounded off version* w' *of the weight function* w with respect to P_X and to the reduced σ-algebra $\mathfrak{X}'$, i.e.

$$w'(x, .) = \underset{P_X}{E}\{w(x, .) \mid x, \mathfrak{X}'\}\,. \tag{5.2}$$

The weight function $w'(x, d)$ being $\mathfrak{X}' \times \mathfrak{D}$-measurable, we can define the *rounded off decision problem* Π' corresponding to the initial decision problem Π as follows: measurable sample space $(Y, \mathfrak{Y})$ and decision space $(D, \mathfrak{D})$ those of Π; measurable parameter space $(X, \mathfrak{X}')$; system of probability measures that represented by the *rounded off channel* (see (5.1)); weight function w' the *rounded off weight function* given by (5.2); a priori probability measure P'_X, generating with the rounded off channel $(\mathfrak{X}', P'_{Y|x}, \mathfrak{Y})$ the restriction P'_{XY} of P_{XY} on the reduced σ-algebra $\mathfrak{X}' \times \mathfrak{Y}$; decision function space $\mathscr{B}$ that of Π.

Let $r_0(\Pi) = \inf_{b \in \mathscr{B}} r(\Pi, b)$ be the Bayes risk corresponding to the problem Π and $r_0(\Pi') = \inf_{b \in \mathscr{B}} r(\Pi', b)$ that corresponding to the problem Π'.

As in section 4, our task is to give an upper estimate of the Bayes risk change $r_0(\Pi') - r_0(\Pi)$ resulting from the reduction of the parameter σ-algebra $\mathfrak{X}$ to $\mathfrak{X}' \subset \mathfrak{X}$.

As in section 4, it is also here impossible to apply directly such inequalities as the inequalities (3.12) of Theorem 3.1, for instance, since the probability measure P'_{XY} is defined on the smaller σ-algebra $\mathfrak{X}' \times \mathfrak{Y}$ and not on the initial σ-algebra $\mathfrak{X} \times \mathfrak{Y}$, as do P_{XY} and $\tilde{P}_{XY}$. Moreover, the weight functions w and w' are, in general, different, namely if w is not measurable with respect to the reduced σ-algebra $\mathfrak{X}' \times \mathfrak{D}$.

In order to overcome these difficulties we proceed in a similar way as in section 4. We need for such an extension $\tilde{P}_{XY}$ of the measure P'_{XY} to $\mathfrak{X} \times \mathfrak{Y}$ that the reduced σ-algebra $\mathfrak{X}' \times \mathfrak{Y}$ is *sufficient* with respect to the system $\{\tilde{P}_{XY}, \tilde{P}_X \times \tilde{P}_Y\}$ of measures, where $\tilde{P}_X$ and $\tilde{P}_Y = P_Y$ are the marginal measures corresponding to $\tilde{P}_{XY}$ and, *moreover*, where both these marginal measures (and not only the second) coincide with the corresponding marginal measures P_X and P_Y of P_{XY}. Without the latter condition the information $I(\tilde{P}_{XY})$, corresponding to the new decision problem $\tilde{\Pi}$ derived from Π by only replacing P_{XY} by $\tilde{P}_{XY}$, still remains equal to that of Π', i.e. $I(\tilde{P}_{XY}) = I(P'_{XY})$, but in general this fact alone is not sufficient here (as opposed to section 4, where (4.1) implies (4.2)) for the equality of the corresponding Bayes risks, i.e. for the validity of the relation $r_0(\tilde{\Pi}) = r_0(\Pi')$.

However, the latter relation is valid if the supplementary condition above is satisfied as it results from the following theorem.

Theorem 5.1. *The general form of an extension $\tilde{P}_{XY}$ of P'_{XY} from $\mathfrak{X}' \times \mathfrak{Y}$ to $\mathfrak{X} \times \mathfrak{Y}$ satisfying the relation*

$$I(\tilde{P}_{XY}) = I(P'_{XY}) \quad (I(P'_{XY}) < \infty) \tag{5.3}$$

is given by

$$\tilde{P}_{XY}(G) = \int_G f' \, d\tilde{P}_X \times P_Y, \quad G \in \mathfrak{X} \times \mathfrak{Y}, \tag{5.4}$$

where $f' = dP'_{XY}/dP'_X \times P_Y$ is the Radon-Nikodym density of P'_{XY} with respect to the product measure $P'_X \times P_Y$ and $\tilde{P}_X$ is an arbitrary extension from $\mathfrak{X}'$ to $\mathfrak{X}$ of the measure P'_X. The existence of the density f' is assured by the assumption that $I(P'_{XY}) < \infty$.

If, moreover, $I(P_{XY}) < \infty$, then

$$\min_{\tilde{P}_{XY} \in \tilde{\mathscr{P}}} H_{\tilde{P}_{XY}}(P_{XY}) = H_{\tilde{P}^0_{XY}}(P_{XY}) = I(P_{XY}) - I(P'_{XY}) \tag{5.5}$$

where $\tilde{\mathscr{P}}$ is the set of all the probability measures $\tilde{P}_{XY}$ of the type (5.4) and $\tilde{P}^0_{XY}$ is that special $\tilde{P}_{XY}$ which results from (5.4) by taking $\tilde{P}_X = P_X$, i.e.

$$\tilde{P}^0_{XY}(G) = \int_G f' \, \mathrm{d}P_X \times P_Y\,, \quad G \in \mathfrak{X} \times \mathfrak{Y}\,. \tag{5.6}$$

For the average risk $r(\tilde{\Pi}^0, b)$ of the decision problem $\tilde{\Pi}^0 = \tilde{\Pi}$ defined above for the special case we take $\tilde{P}_{XY} = \tilde{P}^0_{XY}$ it moreover holds

$$r(\tilde{\Pi}^0, b) = r(\Pi', b) \tag{5.7}$$

where b is any decision function belonging to $\mathscr{B}$ and $r(\Pi', b)$ is the average risk corresponding to the decision problem Π'.

For the Bayes risks in particular it holds

$$r_0(\tilde{\Pi}^0) = r_0(\Pi')\,. \tag{5.8}$$

Proof. The proof of (5.4) under (5.3) results immediately from Theorem 4.1. The proof of (5.5) results from Theorem 4.2.

It remains, thus, to prove (5.7) under (5.6) since (5.8) is an immediate consequence of (5.7).

We have

$$\begin{aligned} r(\tilde{\Pi}^0, b) &= \int w(x, b(y)) \, \mathrm{d}\tilde{P}^0_{XY}(x, y) = \\ &= \int w(x, b(y)) f'(x,y) \, \mathrm{d}P_X \times P_Y(x,y) = \\ &= \int w'(x, b(y)) f'(x, y) \, \mathrm{d}P'_X \times P_Y(x, y) = \\ &= \int w'(x, b(y)) \, \mathrm{d}P'_{XY}(x, y) = r(\Pi', b)\,, \end{aligned} \tag{5.9}$$

where w' is the rounded off weight function defined by (5.2) and corresponding to the rounded off decision problem Π'. Thus, the theorem is proved.

Let us now formulate the main theorem of this section.

Theorem 5.2. *Let Π be the initial decision problem and Π' the rounded off decision problem formulated at the beginning of the present section, the second resulting from the first by a reduction of the parameter σ-algebra $\mathfrak{X}$ to $\mathfrak{X}' \subset \mathfrak{X}$.*

Let the weight function w of Π be uniformly bounded by w_0, i.e. $w(x, d) \leqq w_0$, $x \in X$, $d \in D$, and suppose that the information $I(P_{XY})$ is finite.

If $r_0(\Pi)$ and $r_0(\Pi')$ are the Bayes risks corresponding to the decision problems Π and Π', respectively, then

$$r_0(\Pi') - r_0(\Pi) \leqq \sqrt{\{2 w_0 r_0(\Pi') \left[I(P_{XY}) - I(P'_{XY})\right]\}}\,. \tag{5.10}$$

In other words, if the reduced σ-algebra $\mathfrak{X}' \times \mathfrak{Y}$ is ε-sufficient with respect to the system $\{P_{XY}, P_X \times P_Y\}$ of probability measures (see section 2), then

$$r_0(\Pi') - r_0(\Pi) \leqq \sqrt{[2w_0\varepsilon\, r_0(\Pi')]}\,. \tag{5.11}$$

Proof. Defining the probability measure $\tilde{P}^0_{XY}$ as in Theorem 5.1 by (5.6), we obtain that simultaneously it holds

(i) $$r_0(\tilde{\Pi}^0) = r_0(\Pi')\,,$$

(ii) $$H_{\tilde{P}^0_{XY}}(P_{XY}) = I(P_{XY}) - I(P'_{XY})\,,$$

according to (5.5) and (5.8) of Theorem 5.1, since by hypothesis $I(P_{XY}) < \infty$. The relation (5.10) follows immediately from (i), (ii) and Theorem 3.1, second inequality (3.12), where we replace $\tilde{P}_{XY}$ by $\tilde{P}^0_{XY}$ and we take as T the "identical" transformation. The relation (5.11) is derived from (5.10) on the base of the inequality $I(P_{XY}) - - I(P'_{XY}) \leqq \varepsilon$, implied by the definition of ε-sufficiency we suppose. Thus, the theorem is proved.

Remark 1. As in Remark 1 of section 4, we can derive from the assumption (4.17) a relation completely analogous to (4.20).

Similarly, by a reasoning analogous to that used in Remark 2 of section 4, it is possible to prove the corresponding version of inequalities (4.21), (4.22) and (4.23).

Remark 2. It may happen that the probability law P_{XY} of the decision problem Π is only partially known in the sense, for instance, that only the channel $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$ and the restriction P''_X of P_X on a smaller σ-algebra $\mathfrak{X}'' \subset \mathfrak{X}$ (not necessarily coinciding with $\mathfrak{X}'$) are known. Let in this case $\mathscr{P}_{XY}$ be the set of all the probability measures Q_{XY} on $\mathfrak{X} \times \mathfrak{Y}$ generated (as in (2.1)) by the channel $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$, on the one hand, and by all the extensions $\tilde{P}_X$ of the probability measure P''_X from $\mathfrak{X}''$ to $\mathfrak{X}$, on the other hand. Then in the place of inequality (5.10) we can use the inequality

$$r_0(\Pi') - r_0(\Pi) \leqq \sqrt{\{2w_0 r_0(\Pi') \sup_{Q_{XY} \in \mathscr{P}_{XY}} [I(Q_{XY}) - I(Q'_{XY})]\}} \tag{5.12}$$

which obviously holds since $P_{XY} \in \mathscr{P}_{XY}$.

It is clear that we can write down similar inequalities for all the cases considered in Remark 1 and that the set $\mathscr{P}_{XY}$ in (5.12) may, in the general case, represent all we know about the probability law P_{XY}. A necessary and sufficient condition for the validity in general of (5.12) is that $P_{XY} \in \mathscr{P}_{XY}$.

6. SIMULTANEOUS REDUCTION OF THE INPUT AND OUTPUT σ-ALGEBRAS

Let us consider, as in sections 3, 4 and 5, the statistical decision problem Π, corresponding to the probability space $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, P_{XY})$ and to the weight function w, and suppose that simultaneously the input (parameter space) σ-algebra $\mathfrak{X}$ is reduced to a σ-algebra $\mathfrak{X}' \subset \mathfrak{X}$ and the output (sample space) σ-algebra $\mathfrak{Y}$ is reduced to a σ-algebra $\mathfrak{Y}' \subset \mathfrak{Y}$.

Let P'_{XY} be the restriction on the Cartesian product σ-algebra $\mathfrak{X}' \times \mathfrak{Y}'$ of the probability measure P_{XY}.

Let P_X and P'_X and P_Y and P'_Y be the corresponding marginal measures on $\mathfrak{X}$ and $\mathfrak{X}'$ and $\mathfrak{Y}$ and $\mathfrak{Y}'$, respectively, and $(\mathfrak{X}, P_{Y|x}, \mathfrak{Y})$ and $(\mathfrak{X}', P''_{Y|x}, \mathfrak{Y}')$ the corresponding channels, i.e. the systems of conditional probability measures corresponding to P_{XY} and P'_{XY}, respectively.

Obviously, $(\mathfrak{X}', P''_{Y|x}, \mathfrak{Y}')$ is a restriction on $\mathfrak{Y}'$ of the *rounded off channel* $(\mathfrak{X}', P'_{Y|x}, \mathfrak{Y})$ defined by (5.1).

Let, futher, w' be the *rounded off weight function* defined by (5.2) as the conditional expectation of the weight function w with respect to P_X and to the reduced σ-algebra $\mathfrak{X}'$.

Define the reduced decision problem Π' corresponding to the initial decision problem Π as follows: measurable parameter space $(X, \mathfrak{X}')$; measurable sample space $(Y, \mathfrak{Y}')$; measurable decision space $(D, \mathfrak{D})$ that of Π; decision function space $\mathscr{B}'$ that of all $\mathfrak{Y}'$-measurable decision functions b'; weight function w' the rounded off weight function given by (5.2); system of probability measures that represented by the channel $(\mathfrak{X}', P''_{Y|x}, \mathfrak{Y}')$; a priori probability measure P'_X.

Theorem 6.1. *Let Π and Π' be the decision problems introduced above, the second resulting from the first by a simultaneous reduction of the parameter σ-algebra $\mathfrak{X}$ to $\mathfrak{X}' \subset \mathfrak{X}$ and of the sample σ-algebra $\mathfrak{Y}$ to $\mathfrak{Y}' \subset \mathfrak{Y}$.*

Let the weight function w of Π be uniformly bounded by w_0, i.e. $w(x, d) \leqq w_0$, $x \in X$, $d \in D$, and suppose that the information $I(P_{XY})$ is finite.

If $r_0(\Pi) = \inf_{b \in \mathscr{B}} r(\Pi, b)$ and $r_0(\Pi') = \inf_{b' \in \mathscr{B}'} r(\Pi', b')$ are the Bayes risks corresponding to the decision problems Π and Π', respectively, then

$$r_0(\Pi') - r_0(\Pi) \leqq \sqrt{\{2w_0 r_0(\Pi') [I(P_{XY}) - I(P'_{XY})]\}}\,. \tag{6.1}$$

In other words, if the reduced σ-algebra $\mathfrak{X}' \times \mathfrak{Y}'$ is ε-sufficient with respect to the system $\{P_{XY}, P_X \times P_Y\}$ of probability measures (see section 2), *then*

$$r_0(\Pi') - r_0(\Pi) \leqq \sqrt{[2w_0 r_0(\Pi')\,\varepsilon]}\,. \tag{6.2}$$

Proof. Let us define the probability measure $\tilde{P}^0_{XY}$ by

$$\tilde{P}^0_{XY}(G) = \int_G f' \,\mathrm{d}P_X \times P_Y\,, \quad G \in \mathfrak{X} \times \mathfrak{Y}\,, \tag{6.3}$$

where $f' = \mathrm{d}P'_{XY}/(\mathrm{d}P'_X \times P'_Y)$, and let $\tilde{P}^{0\prime}_{XY}$ be its restriction on $\mathfrak{X} \times \mathfrak{Y}'$.

According to Theorem 4.1, it holds

$$I(\tilde{P}^0_{XY}) = I(\tilde{P}^{0\prime}_{XY}) = I(P'_{XY}) \tag{i}$$

and according Theorem 4.2, it holds

$$H_{\tilde{P}^0{}_{XY}}(P_{XY}) = I(P_{XY}) - I(P'_{XY})\,. \tag{ii}$$

According to the second equality (i) and Theorem 5.1, relation (5.8), where only $\mathfrak{Y}$ is replaced by $\mathfrak{Y}'$ and $\mathscr{B}$ by $\mathscr{B}'$, we have

$$r_0(\Pi') = r_0(\tilde{\Pi}^{0\prime}), \tag{6.4}$$

where $r_0(\tilde{\Pi}^{0\prime})$ is the Bayes risk of the decision problem $\tilde{\Pi}^{0\prime}$ corresponding to the Cartesian product probability space $(X \times Y, \mathfrak{X} \times \mathfrak{Y}, \tilde{P}^{0\prime}_{XY})$, to the weight function w and to the decision function space $\mathscr{B}'$.

According to the first equality (i), implying the sufficiency of $\mathfrak{Y}'$ with respect to the system of conditional probability distributions $\{\tilde{P}^0_{Y|x}, x \in X, [\tilde{P}^0_X = P_X]\}$ corresponding to $\tilde{P}^0_{XY}$, we deduce that

$$r_0(\tilde{\Pi}^{0\prime}) = r_0(\tilde{\Pi}^0), \tag{6.5}$$

where $r_0(\tilde{\Pi}^0)$ is the Bayes risk of the decision problem $\tilde{\Pi}^0$ resulting from Π by only replacing P_{XY} by $\tilde{P}^0_{XY}$.

By applying Theorem 3.1, second inequality (3.12), where we replace $\tilde{P}_{XY}$ by $\tilde{P}^0_{XY}$ and we take as T the "identical" transformation, we obtain

$$r_0(\tilde{\Pi}^0) - r_0(\Pi) \leqq \sqrt{[2w_0 r_0(\tilde{\Pi}^0)\, H_{\tilde{P}^0_{XY}}(P_{XY})]}\,. \tag{6.6}$$

On the base of (ii), (6.4), (6.5) and (6.6) we obtain (6.1). The relation (6.2) is derived from (6.1) on the base of the inequality $I(P_{XY}) - I(P'_{XY}) \leqq \varepsilon$, implied by the definition of ε-sufficiency we suppose. Thus, the theorem is proved.

Remark 1. It is analogous to Remarks 1 and 2 of section 4.

Remark 2. If P_{XY} is only partially known in the sense, for instance, of Remark 2 of section 5, and if $\mathscr{P}_{XY}$ is the smallest set we know of probability measures on $\mathfrak{X} \times \mathfrak{Y}$ containing P_{XY}, then we can use in the place of (6.1) the relation

$$r_0(\Pi') - r_0(\Pi) \leqq \sqrt{\{2w_0 r_0(\Pi') \sup_{Q_{XY} \in \mathscr{P}_{XY}} [I(Q_{XY}) - I(Q'_{XY})]\}}\,, \tag{6.7}$$

where Q'_{XY} is the restriction of Q_{XY} on $\mathfrak{X}' \times \mathfrak{Y}'$. Relation (6.7) results immediately from (6.1) on the base of the assumption that $P_{XY} \in \mathscr{P}_{XY}$.

(Received February 3rd, 1965.)

REFERENCES

[1] Perez A.: Sur la théorie de l'information et la discernabilité dans les problèmes de décision statistique. In: Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (1959). Prague 1960, 413—497.

[2] Perez A.: Information Theory in Data Reduction Problems. Paper presented at the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, 1962.

[3] Perez A.: Theory of Information and Cybernetics, Part II (in Czech). In: Kybernetika a její využití — Transactions of the First Conference on Cybernetics in Czechoslovakia (1962). Prague 1965.

[4] Perez A. and Tondl L.: On the Role of Information Theory in Certain Scientific Procedures. In: Information and Prediction (Dockx and Bernays, Brussels 1962), Academic Press, New York 1965.
[5] Perez A.: Contributions de la théorie de l'information en cybernétique. Paper presented at the IVth International Congress of Cybernetics, Namur 1964. To appear in the Transactions of the Congress.
[6] Perez A.: Extensions of Shannon-McMillan's Limit Theorem to more general Stochastic Processes. In: Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (1962). Prague 1964, 545—574.
[7] Wald A.: Statistical Decision Functions. New York 1950.
[8] Blackwell D., Girschick M. A.: Theory of Games and Statistical Decisions. New York 1955.
[9] Perez A.: Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie des martingales. In: Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (1956). Prague 1957, 183—208.
[10] Pinsker M. S.: Information and Information Stability of Random Variables and Processes. Moscow 1960.
[11] Perez A.: Transformation ou σ-algèbre suffisante et minimum de la probabilité d'erreur. Czechosl. Mathem. Journ., Vol. 7 (1957), 115—123.
[12] Csiszár I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, Vol. VIII, Series A, Fasc. 1—2 (1963), 85—108.
[13] Hájek J.: A property of J — divergences of marginal probability distributions. Czechosl. Mathem. Journ., Vol. 8 (1958), 460—463.

VÝTAH

Informace, ε-suficientnost a problémy redukce dat

ALBERT PEREZ

Ve složitých rozhodovacích problémech, typických pro kybernetiku, se zvlášť ostře projevuje omezenost různých „kapacit", „pamětí" a „lhůt", jež jsou k disposici pro zpracování údajů nejrozmanitější povahy za účelem adekvatního rozhodování. Redukční tendence, které z toho nutně vyplývají, mají být zaměřeny k nalezení co nejúspornějšího souboru údajů (tzv. souboru podstatných parametrů), který je přípustnou redukcí, tj. slučitelný s požadovanou kvalitou rozhodování, ovšem za předpokladu, že tato kvalita je vůbec za daných podmínek dosažitelná.

V rámci Bayesova modelu statistického rozhodování [7], [8] je v tomto článku formulováno několik informačněteoretických kritérií, která dovolují posoudit, zda prozkoumaná varianta redukce je přípustná, aniž by bylo nutno předem řešit odpovídající variantu rozhodovacího problému. Poslední postup by totiž vedl, za podmínek, kdy se uvažuje o velkém počtu takových variant, k nepřekonatelným obtížím.

V § 1 (*Úvod*) je dán stručný přehled o vzniku a vývoji problematiky redukce dat z hlediska teorie informace a naznačeny sledované cíle.

V § 2 (*Informace a ε-suficientnost*) je zaveden klasický model statistického rozhodování a definován pojem informace (viz. (2.5)). Dále je pomocí pojmu informace zaveden pojem ε-suficientnosti (ε-postačitelnosti, viz (2.7)) jako přirozené rozšíření pojmu postačitelnosti z matematické statistiky. Při postačitelné redukci informace se zachová. Při ε-postačitelné redukci informace klesá nejvíce o ε. Nerovnosti (2.14) a (2.15) jsou speciální případy teorému 2.1.

V § 3 (*Odhad změny průměrného rizika způsobené modifikací pravděpodobnostního zákona v statistickém rozhodovacím problému*) základem je lemma 3.1, která přes nerovnosti (3.10) pro zobecněné entropie, dovoluje dokázat teorém 3.1, dále systematicky používaný.

V § 4 (*Redukce σ-algebry výběrového prostoru*) je dán odhad zvýšení Bayesova rizika $r_0(\Pi') - r_0(\Pi)$ při přechodu z rozhodovacího problému Π k rozhodovacímu problému Π' s redukovaným výstupním (výběrovým) prostorem (viz zejména teorém 4.3, kde v nerovnostech (4.15) a (4.16) důležitou roli hraje pojem ε-postačitelnosti, tj. snížení informace, které z redukce vyplývá).

V § 5 (*Redukce σ-algebry parametrového prostoru* či *prostoru hypotéz*) roli problému Π' hraje jakýsi „zaokrouhlený" rozhodovací problém vyplývající z problému Π po redukci vstupního (parametrového) prostoru. Teorém 5.2 je analogický teorému 4.3.

V § 6 (*Simultánní redukce vstupní a výstupní σ-algebry*) k odhadu zvýšení Bayesova rizika, vyplývajícího z redukce jak vstupního tak i výstupního prostoru, slouží zejména teorém 6.1, analogický teorémům 4.3 a 5.2. Je uvažován také případ neúplné znalosti působícího pravděpodobnostního zákona P_{XY} (viz např. (6.7)).

Ve všech zde uvažovaných případech redukce, informačněteoretický pojem ε-postačitelnosti vystupuje jaksi automaticky při hledání nejlepšího (v jistém smyslu) odhadu (viz zejména teorém 4.2, vztah (4.9)).

Dr. Albert Perez, DrSc., Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.