KYBERNETIKA ČÍSLO 3, ROČNÍK 6/1970

On the Reducibility of a Set of Statistical Hypotheses*

ALBERT PEREZ

This paper deals with the estimation of the Bayes risk increment implied by a reduction of the set of statistical hypotheses in a statistical decision problem.

In papers [1, 2] we have considered the problem of estimating in Shannon's information terms the loss of decision quality in a statistical decision problem implied by a reduction of the sample space sigma-algebra and/or of the parameter space sigma-algebra. In the present paper we develop further certain aspects of the parameter space (statistical hypotheses set) reduction problem taking, namely, account of a parallel reduction of the decision space. In addition, we apply some new estimates of the decision quality loss in terms of the second order information loss (cf. [3, 4]).

1. REDUCTION OF A SET OF STATISTICAL HYPOTHESES

Let us consider a set of statistical hypotheses (probability distributions)

$$S = \{P_{Y/x}, x \in X\} \tag{1}$$

on a measurable (sample) space $(Y, \mathfrak{Y})$, where the parameter space X is endowed with a sigma-algebra $\mathfrak{X}$ of its subsets such that, for every set $E \in \mathfrak{Y}$ (the sample space sigma-algebra), the function $P_{Y/x}(E)$ of x is $\mathfrak{X}$-measurable. In the terminology of Information Theory, such a set of statistical hypotheses constitutes a "channel" denoted by the triplet $(\mathfrak{X}, P_{Y/x}, \mathfrak{Y})$.

Let P_X on the measurable parameter space $(X, \mathfrak{X})$ be the a priori probability distribution of the set of statistical hypotheses. Let us, further, assume that the "loss" implied by a misidentification of the hypothesis $P_{Y/x} \in S$, respectively of

* Presented at the 37th Session of the International Statistical Institute, London 3—11 September 1969.

$x \in X$ by $d \in X$ is given by the weight (loss) function $w(x, d)$, defined on the Cartesian product $X \times X$. This function is supposed to be nonnegative and measurable with respect to the Cartesian product sigma-algebra $\mathfrak{X} \times \mathfrak{X}$.

The primary (unreduced) task is to choose (in the sequel we shall suppose that such a choice is always possible) for every observed sample value $y \in Y$ such a decision $d_y \in X$ that minimizes the "average loss" $r_y(d)$, i. e.

$$(2) \qquad r_y(d_y) = \min_{d \in X} r_y(d) = \min_{d \in X} \int_X w(x, d) \, \mathrm{d}P_{X/y}(x) \, .$$

Here $P_{X/y}$ denotes the respective conditional probability distribution on $(X, \mathfrak{X})$ given $y \in Y$. The average value of $r_y(d_y)$ with respect to the marginal probability distribution P_Y on $(Y, \mathfrak{Y})$ is the so-called Bayes risk r_0 and the function $d_y = b_0(y)$ is a Bayes (optimal) decision function.

Let us now reduce the set of statistical hypotheses by reducing the parameter space sigma-algebra $\mathfrak{X}$ to $\mathfrak{X}'$, i. e. $\mathfrak{X}' \subset \mathfrak{X}$. Note that in this case the triplet $(\mathfrak{X}', P_{Y/x}, \mathfrak{Y})$ may not be a channel and the weight function w may not be measurable with respect to $\mathfrak{X}' \times \mathfrak{X}'$.

The new (reduced) set of statistical hypotheses will be given by

$$(3) \qquad S' = \{P'_{Y/x}, x \in X\} \, ,$$

where $P'_{Y/x}(E) = \mathscr{E}_{P_X}\{P_{Y/x}(E)/x, \mathfrak{X}'\}$ is the conditional expectation (and, thus, a rounded off version) of $P_{Y/x}(E)$, $E \in \mathfrak{Y}$, with respect to P_X and to the reduced sigma-algebra $\mathfrak{X}'$. The triplet $(\mathfrak{X}', P'_{Y/x}, \mathfrak{Y})$ is now a channel. The new (reduced) decision problem considered in [1, 2] was in this case defined by the system of elements:

$$(4) \qquad \{(X, \mathfrak{X}'), (Y, \mathfrak{Y}), (P'_X, S') = (P_Y, S'_1), (X, \mathfrak{X}), w'\}$$

where P'_X is the restriction of P_X on $\mathfrak{X}'$; $S'_1 = \{P'_{X/y}, y \in Y\}$ is the converse to the set S' (cf. (3)) with $P'_{X/y}$ defined by $\mathrm{d}P'_{X/y}(x) = f'_x(y) \, \mathrm{d}P'_X(x)$, where $f'_x(y)$ is the Radon-Nikodym density of $P'_{Y/x}$ with respect to P_Y (this density exists under the assumption we make throughout the paper that the Shannon's information $I(P_X, S)$, corresponding to the a priori probability distribution P_X and to the system of statistical hypotheses S, is finite (cf. [1, 2])); the weight function $w'(., d)$ is for every $d \in X$ defined as the conditional expectation of the weight function $w(., d)$ with respect to P_X and the reduced sigma-algebra $\mathfrak{X}'$: $w'(x, d) = \mathscr{E}_{P_X}\{w(x, d)/x, \mathfrak{X}'\}$. Thus, $w'(., d)$ is measurable with respect to $\mathfrak{X}'$.

It is possible to see that the minimal average loss $r'_y(d'_y)$ (cf. (2)) and the Bayes risk r'_0 corresponding to the decision problem (4) coincide with those of the decision problem defined by the system of elements:

$$(5) \qquad \{(X, \mathfrak{X}), (Y, \mathfrak{Y}), (P_X, S') \quad \text{or} \quad (P_Y, \tilde{S}'_1), (X, \mathfrak{X}), w\} \, .$$

Here $\tilde{S}'_1 = \{\tilde{P}'_{X/y}, y \in Y\}$, where $\tilde{P}'_{X/y}$ is an extension of $P'_{X/y}$ from $\mathfrak{X}'$ to $\mathfrak{X}$ defined by the relation $d\tilde{P}'_{X/y}(x) = f'_y(x)\, dP_X(x)$ where $f'_y(x)$ is the Radon-Nikodym density of $P'_{X/y}$ with respect to P'_X (this density exists under the hypothesis made above). This extension dominates $P_{X/y}$ and conserves the information and, as a consequence, it conserves also the above risks (cf. [1, 2]).

Better estimates of the decision quality losses $r'_y(d'_y) - r_y(d_y)$ and $r'_0 - r_0$ than those obtained in [1, 2] in terms of the first order (i. e. Shannon's) generalized entropy it is possible to obtain by applying the method of the constrained extremum of the generalized f-entropy (with $f(u)$ convex function not necessarily of the Shannon's type $f(u) = u \log u$) developed in [3, 4]. Thus, restricting us to the case of the second order generalized entropy ($f(u) = u^2$), we obtain under very general conditions the following estimates:

$$(6) \qquad r'_y(d'_y) - r_y(d_y) \leqq (r'_y(w^2, d'_y) - r'^2_y(d'_y))^{1/2} \times (H_2(P_{X/y}, \tilde{P}'_{X/y}) - 1)^{1/2}$$

where $r'_y(w^2, d'_y) = \int_X w^2(x, d'_y)\, d\tilde{P}'_{X/y}(x)$ and $H_2(P_{X/y}, \tilde{P}'_{X/y})$ is the second order generalized entropy of $P_{X/y}$ with respect to $\tilde{P}'_{X/y}$. If $g_y(x)$ is the Radon-Nikodym density (it exists) of the first with respect to the second, then

$$(7) \qquad H_2(P_{X/y}, \tilde{P}'_{X/y}) = \int_X g^2_y(x)\, d\tilde{P}'_{X/y}(x),$$

$$(8) \qquad 0 \leqq r'_0 - r_0 \leqq (r'(w^2, b'_0) - r'^2_0)^{1/2} \times (H_2(P, \tilde{P}') - 1)^{1/2}$$

where P is generated by P_Y and $P_{X/y}$ and $\tilde{P}'$ by P_Y and $\tilde{P}'_{X/y}$. The second order generalized entropy $H_2(P, \tilde{P}')$ of P with respect to $\tilde{P}'$ is equal to the average value of $H_2(P_{X/y}, \tilde{P}'_{X/y})$ with respect to P_Y. The expression $H_2(P, \tilde{P}') - 1$ represents, thus, the "second order information loss" caused by the reduction above whereas in the estimates of [1, 2] was implied the "first order (i. e. Shannon's) information loss".

However, it is more adequate in defining the new (reduced) decision problem, corresponding to the reduced set S' of statistical hypotheses (cf. (3)) as well as the respective decision quality loss, to take account of a parallel reduction of the decision space D. Up to now D was taken equal to the unreduced parameter space $(X, \mathfrak{X})$.

2. PARALLEL REDUCTION OF THE DECISION SPACE

In the sequel we shall estimate the additional loss of decision quality resulting from a reduction of the decision space parallel to that of the set of statistical hypotheses from S to S' (cf. (1) and (3)).

For the sake of simplicity, we shall here restrict us to the special case where the measurable parameter space is the Cartesian product of two measurable spaces $(X_1, \mathfrak{X}_1)$ and $(X_2, \mathfrak{X}_2) : (X, \mathfrak{X}) = (X_1 \times X_2, \mathfrak{X}_1 \times \mathfrak{X}_2)$, and the reduced set S' of statistical hypotheses corresponds to the reduced sigma-algebra $\mathfrak{X}' = \mathfrak{X}_1 \times X_2$.

Thus, for every set $E \in \mathfrak{Y}$,

(9) $$P'_{Y/x}(E) = \int_{X_2} P_{Y/x}(E)\, \mathrm{d}P_{X_2/x_1}(x_2) = P_{Y/x_1}(E)$$

where P_{X_2/x_1} is the conditional probability distribution on $(X_2, \mathfrak{X}_2)$ given $x_1 \in X_1$ corresponding to $P_X = P_{X_1 \times X_2}$.

Similarly, $P'_X = P_{X_1}$ where P_{X_1} is the marginal probability distribution induced by P_X on $(X_1, \mathfrak{X}_1)$, and the rounded off weight function $w'(x, d)$ is for every $d \in X_1 \times X_2$ given by

(10) $$w'(x, d) = \int_{X_2} w(x_1, x_2; d)\, \mathrm{d}P_{X_2/x_1}(x_2) = w'(x_1, d)\,.$$

In the sequel we give two different approaches in defining the new (reduced) decision problem.

First procedure

The new (reduced) decision problem is defined by the following system of elements:

(11) $$\{(X_1, \mathfrak{X}_1), (Y, \mathfrak{Y}), P_{X_1}, S' = \{P_{Y/x_1}, x_1 \in X_1\}, D = (X_1, \mathfrak{X}_1), w''\}$$

(as in (4) and (5), the first element represents the measurable parameter space, the second element represents the measurable sample space, the third element represents the a priori probability distribution, the fourth element represents the set of statistical hypotheses (channel), the fifth element represents the decision space, and the sixth element represents the weight function). Here w'' is defined in a natural way as a rounded off version of $w'(x_1; d_1, d_2)$ with respect to $d_2 \in X_2$ (cf. (10)):

(11') $$w''(x_1, d_1) = \int_{X_2} w'(x_1; d_1, d_2)\, \mathrm{d}P_{X_2/d_1}(d_2)$$

for x_1 and d_1 of X_1 and x_2 and d_2 of X_2.

The following theorem permits us to estimate the additional decision quality losses $r''_y(d''_y) - r'_y(d'_y)$ and $r''_0 - r'_0$ on passing from the decision problem (4) or (5) to the decision problem (11). Here $r''_y(d''_y) = r''_y(d''_{1y})$ and $r'_y(d'_y)$ are the minimal average losses corresponding to the decision problems (11) and (4) or (5), respectively, and r''_0 and r'_0 are the corresponding Bayes risks (cf. (2)).

Theorem 1. *If the initial weight function $w(x_1, x_2; d_1, d_2)$ is a metric on $X_1 \times X_2$ and if for $x_1 \in X_1$, $x_2 \in X_2$, $d_1 \in X_1$* and *$d_2 \in X_2$,*

(12) $$w(x_1, x_2; d_1, d_2) \geqq w(d_1, x_2; d_1, d_2)\,,$$

then it holds

$$w''(x_1, d_1) \leqq 2w'(x_1; d_1, d_2) \tag{13}$$

and, as a consequence, it results

$$r_y''(d_{1y}'') - r_y'(d_y') \leqq r_y'(d_y'), \tag{14}$$

$$r_0'' - r_0' \leqq r_0'. \tag{15}$$

Proof. Let us prove (13). On the base of our hypotheses,

$$\begin{aligned} w(x_1, x_2; d_1, d_2) &\leqq w(x_1, x_2; d_1, d_2') + w(d_1, d_2'; d_1, d_2) = \\ &= w(x_1, x_2; d_1, d_2') + w(d_1, d_2; d_1, d_2') \leqq \\ &\leqq w(x_1, x_2; d_1, d_2') + w(x_1, d_2; d_1, d_2'). \end{aligned}$$

The last inequality is namely obtained on the base of (12).

By integrating the first and the last member of the above relation with respect to x_2 by P_{X_2/x_1} we obtain, according to (10),

$$w'(x_1; d_1, d_2) \leqq w'(x_1; d_1, d_2') + w(x_1, d_2; d_1, d_2').$$

By a second integration with respect to d_2 by P_{X_2/d_1} we obtain, according to (11'), (10),

$$w''(x_1, d_1) \leqq w'(x_1; d_1, d_2') + w'(x_1; d_1, d_2') = 2w'(x_1; d_1, d_2'), \quad d_2' \in X_2.$$

Thus, the inequality (13) is proved. As to the inequality (14), it results from the fact that

$$r_y''(d_{1y}'') \leqq r_y''(d_{1y}'),$$

where d_{1y}' is the first component of $d_y' = (d_{1y}', d_{2y}')$, minimizing $r_y'(d)$, whereas d_{1y}'' minimizes $r_y''(d_1)$, $d_1 \in X_1$. But, according to (13),

$$\begin{aligned} r_y''(d_{1y}') &= \int_{X_1} w''(x_1; d_{1y}') \, \mathrm{d}P_{X_1/y}(x_1) \leqq \\ &\leqq \int_{X_1} 2w'(x_1; d_{1y}', d_{2y}') \, \mathrm{d}P_{X_1/y}(x_1) = 2r_y'(d_y'). \end{aligned}$$

Combining this result with the above inequality, we obtain the inequality (14). The inequality (15) results immediately from the inequality (14) by integrating with respect to P_Y. Thus, the theorem is proved.

The new (reduced) decision problem is defined by the following system of elements:

$$\{(X_1, \mathfrak{X}_1), (Y, \mathfrak{Y}), P_{X_1}, S' = \{P_{Y/x_1}, x_1 \in X_1\}, \quad D = \{(x_1, d_2(x_1)), x_1 \in X_1, d_2(x_1) \in X_2\}, w'\}, \tag{16}$$

or, equivalently, by

$$\{(X_1 \times X_2, \mathfrak{X}_1 \times \mathfrak{X}_2), (Y, \mathfrak{Y}), (P_Y, \tilde{S}'_1) = \text{as in (5)}, D = \text{as in (16)}, w\}, \tag{17}$$

where the decision space $D \subset X_1 \times X_2$, of power equal to the power of X_1, i. e. the function $d_2(x_1)$ of $x_1 \in X_1$, is to be chosen in a manner to minimize the corresponding Bayes risk r''_0 of the decision problem (16) or (17).

Theorem 2. *Under the assumptions of Theorem* 1 *concerning the weight function* w, *there always exists a decision space* D *of the above type such that the loss of decision quality on passing from the decision problem* (4) *or* (5), *with Bayes risk* r'_0, *to the decision problem* (16) *or* (17), *with Bayes risk* $r''_0(D)$, *satisfies the inequality*

$$r''_0(D) - r'_0 \leqq r'_0 . \tag{18}$$

Proof. Let b be a decision function for the decision problem (5) and take for the decision problem (17) a decision function b' such that

$$b'(y) = (x_1, d_2(x_1)) \quad \text{for} \quad y \in b^{-1}(x_1 \times X_2) \tag{19}$$

with $d_2(x_1) \in X_2$ to be chosen later.

By taking $P_Y b^{-1} = Q_X$ for the corresponding average risks $r'(b)$ and $r''(b')$ we have:

$$r'(b) = \int_{X_1}\int_{X_2}\int_{X_1}\int_{X_2} w(x'_1, x'_2; x_1, x_2)\, d\tilde{P}'_{X/b^{-1}(x_1,x_2)}(x'_1, x'_2)\, dQ_{X_1}(x_1)\, dQ_{X_2/x_1}(x_2),$$

$$r''(b') = \int_{X_1}\int_{X_2}\int_{X_1}\int_{X_2} w(x'_1, x'_2; x_1, d_2(x_1))\, d\tilde{P}'_{X/b^{-1}(x_1,x_2)}(x'_1, x'_2)\, dQ_{X_1}(x_1)\, dQ_{X_2/x_1}(x_2).$$

According to the metric hypothesis concerning w, it holds

$$w(x'_1, x'_2; x_1, d_2(x_1)) \leqq w(x'_1, x'_2; x_1, x_2) + w(x_1, x_2; x_1, d_2(x_1))$$

so that by integrating as above we obtain

$$r''(b') \leqq r'(b) + \tag{20}$$

$$+ \int_{X_1}\int_{X_2}\int_{X_1}\int_{X_2} w(x_1, x_2; x_1, d_2(x_1))\, d\tilde{P}'_{Y/b^{-1}(x_1,x_2)}(x'_1, x'_2)\, dQ_{X_1}(x_1)\, dQ_{X_2/x_1}(x_2).$$

Let us now define $d_2(x_1)$ by the relation*

$$(21)\quad \int_{X_2} w(x_1, x_2; x_1, d_2(x_1))\, \mathrm{d}Q_{X_2/x_1}(x_2) = \min_{x'_2 \in X_2} \int_{X_2} w(x_1, x_2; x_1, x'_2)\, \mathrm{d}Q_{X_2/x_1}(x_2)\,.$$

On the base of (21), the metric property of w and the property (12), the inequality (20) successively gives

$$(22)\qquad r''(b') \leqq r'(b) +$$

$$+ \int_{X_1}\int_{X_2}\int_{X_1}\int_{X_2} w(x_1, x_2; x_1, x'_2)\, \mathrm{d}\tilde{P}'_{X/b^{-1}(x_1,x_2)}(x'_1, x'_2)\, \mathrm{d}Q_{X_1}(x_1)\, \mathrm{d}Q_{X_2/x_1}(x_2) \leqq$$

$$\leqq r'(b) + \int_{X_1}\int_{X_2}\int_{X_1}\int_{X_2} w(x_1, x_2; x'_1, x'_2)\, \mathrm{d}\tilde{P}'_{X/b^{-1}(x_1,x_2)}(x'_1, x'_2)\, \mathrm{d}Q_{X_1}(x_1)\, \mathrm{d}Q_{X_2/x_1}(x_2) =$$

$$= r'(b) + \int_{X_1}\int_{X_2}\int_{X_1}\int_{X_2} w(x'_1, x'_2; x_1, x_2)\, \mathrm{d}\tilde{P}'_{X/b^{-1}(x_1,x_2)}(x'_1, x'_2)\, \mathrm{d}Q_{X_1}(x_1)\, \mathrm{d}Q_{X_2/x_1}(x_2) =$$

$$= 2r'(b)\,.$$

By taking as b the Bayes decision function b'_0 of the decision problem (5), let $(b'_0)'$ denote the corresponding decision function b' defined by (19), and let D be chosen according to $d_2(x_1)$ as defined by (21) for $Q_X = P_Y b_0'^{-1}$. On the base of (22) we can then write

$$r''_0(D) \leqq r''((b'_0)') \leqq 2r'(b'_0) = 2r'_0$$

and, thus, the theorem is proved.

Theorems of the type above, combined with inequalities of the type (6) and (8), permit us to obtain estimates of the decision quality loss $r''_0 - r_0$ implied by a given reduction of the set of statistical hypotheses (and a parallel reduction of the decision space) and, thus, the study of the reducibility of such a set (compatible with a given decision quality loss) may be facilitated.

(Received January 29, 1970.)

* In the general case, instead of *min* it should be taken *inf* by adding an arbitrarily small positive ε.

[1] A. Perez: Information and ε-Sufficiency. Paper No. 41 presented at the 35th Session of the International Statistical Institute, Belgrad 1965.

[2] A. Perez: Information Theory Methods in Reducing Complex Decision Problems. In: Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (1965). Prague 1967, 55—87.

[3] A. Perez: Information-Theoretic Risk Estimates in Statistical Decision. Kybernetika *3* (1967), 1, 1—21.

[4] A. Perez: Sur l'énergie informationnelle de M. Octav Onicescu. Rev. Roum. Math. Pures et Appl. *XII* (1967), 9, 1341—1347.

VÝTAH

O reduktibilitě souboru statistických hypotéz

ALBERT PEREZ

V [1, 2] byl uvažován problém odhadu pomocí Shannonovy informace snížení rozhodovací kvality, vyplývající z redukce σ-algebry výběrového nebo parametrového prostoru v statistickém rozhodování.

V tomto článku je uvažováno o některých dalších aspektech problému redukce parametrového prostoru, resp. souboru statistických hypotéz. Zejména je vzata v úvahu paralelní redukce prostoru rozhodnutí.

Odhad celkového snížení kvality rozhodování, resp. celkového zvýšení Bayesova rizika při redukci prostoru parametrů a paralelní redukci prostoru rozhodnutí se provádí ve dvou etapách. V první etapě (§ 1) se odhaduje zvýšení $r_0' - r_0$ Bayesova rizika (viz (6) a (8)), způsobené redukcí parametrového prostoru, a to pomocí tzv. ztráty informace druhého řádu (viz (7) a odkazy [3, 4]).

V druhé etapě (§ 2) se odhaduje za určitých předpokladů, týkajících se zejména váhové funkce (viz Theorem 1 a Theorem 2), dodatečné zvýšení $r_0'' - r_0'$ Bayesova rizika způsobené paralelní redukcí prostoru rozhodnutí. Ukazuje se, že toto poslední zvýšení nepřesahuje hodnotu r_0' Bayesova rizika, odpovídajícího pouhé redukci parametrového prostoru.

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