KYBERNETIKA ČÍSLO 6, ROČNÍK 6/1970

Statistical Data Reduction via Construction of Sample Space Partitions*

JAN BIAŁASIEWICZ

The statistical data reduction problem is presented as a problem of construction of sample space partitions. Then, the algorithm for synthesis of an ε-sufficient partition of a sample space is derived and its modification from the view-point of applications is formulated and discussed.

1. INTRODUCTION

Let the triple $(\Omega, \mathfrak{A}, P_\zeta)$ be a probability space: here Ω is a set whose elements are called ω's, $\mathfrak{A}$ denotes the σ-algebra of all subsets of Ω, P_ζ is a probability measure defined on the (measurable) space $(\Omega, \mathfrak{A})$. Let $\zeta(\omega)$ be the random variable corresponding to P_ζ and with range Ω. We shall call $(\Omega, \mathfrak{A}, P_\zeta)$ the *parameter space*. This name will be used also in refering simply to Ω.

We shall call $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ the *sample space*, where $(X, \mathfrak{X})$ is the measurable space of outcomes of an experiment and $P_{\xi|\omega}$ are conditional probability measures defined on $(X, \mathfrak{X})$ for each given parameter value $\omega \in \Omega$. Elements of the real space X are called x's, $\mathfrak{X}$ denotes the σ-algebra of all subsets of X. The set Ω can be also considered as an *index set* of probability measures $P_{\xi|\omega}$ on $(X, \mathfrak{X})$. $\xi(\omega)$ is a random variable defined on the space Ω and taking its values in X. The name "sample space" is also used when refering only to X, its first element.

Let Y be a proper subset of X and let $(Y, \mathfrak{Y})$ be the measurable space with $\mathfrak{Y}$ being the σ-algebra of all subsets of Y. We define the *problem of data reduction* as the problem of finding a partition $\mathscr{A}_T$ of X defined by some measurable transformation T from $(X, \mathfrak{X})$ onto $(Y, \mathfrak{Y})$. In other words, the problem of data reduction may be considered as the problem of searching the new experiment to be performed which is nothing different than the determination of a new random variable $\eta(\omega)$ defined

* Results presented have been obtained when the author was with the Department of Mathematics, Oragon State University.

on Ω which may be expressed as a following composition:

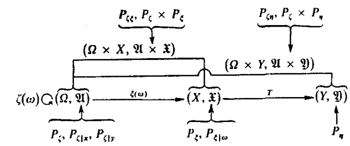
(1) $$\eta(\omega) = T \circ \xi(\omega)$$

To each point $y \in Y$ corresponds some event $A_y \in \mathfrak{X}$ such that

(2) $$Tx = y \quad \text{for all} \quad x \in A_y$$

and, of course, by definition $A_y \in \mathscr{A}_T$.

The diagram below summarizes the principle notations to be used and gives the view of their relationships, where all the probability measures are generated in a standard way, provided P_ξ, $\{P_{\xi|\omega}, \omega \in \Omega\}$, T are given and η is defined by (1).



It should be clear that usually some constraints are imposed on a class of transformations to which T belongs. These are constraints concerning preservation under transformation T of information about the unknown value of parameter ω which is incorporated in events $A \in \mathfrak{X}$. To be able to make it more clear we introduce now some additional notations. Let D be an arbitrary space of actions or decisions d, let L be a loss function defined on $\Omega \times D$, let $\mathscr{B}$ be a class of $\mathfrak{X}$-measurable decision functions δ with the range D. Further, let $\mathfrak{X}' \subset \mathfrak{X}$ be the σ-algebra generated by the partition $\mathscr{A}_T$ and let $\mathscr{B}' \subset \mathscr{B}$ be a class of $\mathfrak{X}'$-measurable decision functions δ'. We are now in position to give the following definition.

Definition 1. The space X and the partition $\mathscr{A}_T$ are said to be equally informative if there exists an element $\delta'_0 \in \mathscr{B}'$ such that

(3) $$r(P_{\zeta\xi}, \delta'_0) = \inf_{\delta \in \mathscr{B}} r(P_{\zeta\xi}, \delta)$$

where

(4) $$r(P_{\zeta\xi}, \delta) = \int_{\Omega \times X} L(\omega, \delta(x)) \, \mathrm{d}P_{\zeta\xi} \, .$$

In the sequel we shall consider partitions $\mathscr{A}_T$ which are "as informative as" X, as well as, such which are not. We remark that in general case only the latter lead to the essential data reduction. This statement is clarified later.

2. PARTITIONS WHICH ARE SUFFICIENT ON A SAMPLE SPACE

In Backwell and Girshick [1] may be found the following definition of a *sufficient partition.*

Definition 2. Let $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ be a sample space. A partition $\mathscr{A}$ of X is said to be sufficient on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ if for every bounded function f defined on X and every $A \in \mathscr{A}$, the conditional expectation of f, given A and ω

$$\mathrm{E}_{\omega}(f \mid A) = \frac{1}{P_{\xi|\omega}(A)} \int_A f \,\mathrm{d}P_{\xi|\omega}$$

is independent of ω for those $\omega \in \Omega$ for which $P_{\xi|\omega}(A) > 0$.

Using the factorization theorem (see [1] for the formulation and proof) one can prove

Theorem 1. *Let $\mathscr{A}$ be a sufficient partition on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$. Then X and $\mathscr{A}$ are equally informative.*

It follows from Theorem 1 that if $\mathscr{A}$ is a non-trivial sufficient partition (i.e. such a sufficient partition which does not exclusively consists of individual points of X), then instead of making precise measurements of the physical parameters of some objects represented by the vector $x \in X$, one can check only to which $A \in \mathscr{A}$ this vector belongs. If there exist non-trivial sufficient partitions $\mathscr{A}$ of X, the question arises how to construct the minimal sufficient partition.

The appropriate algorithm may be readily written on the basis of Lemma 8.4.1 and Lemma 8.4.3 given in [1] under the following assumptions:

(a) the sample space $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ is auch that for each $x \in X$ there exists at least one $\omega \in \Omega$ with $P_{\xi|\omega}(x) > 0$,

(b) the parameter space Ω is finite.

The assumption (a) means that the space X is such that its points really occur as results of the experiment performed. It is clear that from the view-point of applications the assumption (b) can not be considered as a restriction.

3. SUFFICIENT STATISTICS

Classically the sufficient statistic T on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ is defined as a random variable such that the partition $\mathscr{A}_T$ of X determined by T is sufficient on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$.

Proposition 1. *Let $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ be a sample space, let $(\Omega, \mathfrak{A}, P_{\zeta})$ be a parameter space, and let T be a random variable defined on X and with range $Y \ni y$. Then T is a sufficient statistic on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ if and only if for each pair (x, y) such* *that*

(5) $$y = Tx$$

the equality

(6) $$P_{\zeta|x}(B) = P_{\zeta|y}(B)$$

holds for all $B \in \mathfrak{A}$ *such that* $\int_B P_{\xi|\omega}(x)\, P_\zeta(\omega)\, d\omega > 0$.

Proof. Suppose that T is a sufficient statistic on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$ and $\mathscr{A}_T$ is the corresponding sufficient partition. Then for every $x \in A_y$ and each $A_y \in \mathscr{A}_T$

$$Tx = y$$

Define

$$s_\omega(x) = \frac{P_{\xi|\omega}(x)}{\int_\Omega P_{\xi|\omega}(x)\, P_{\zeta|\omega}(\omega)\, d\omega}$$

assuming that for each x the denominator is positive which is true by hypothesis (see Definition 2). This is also equivalent to the appropriate condition in Proposition 1. Then, from factorization theorem for sufficient statistics

$$s_\omega(x) = \frac{h(Tx, \omega)\, q(x)}{\int_\Omega h(Tx, \omega)\, q(x)\, P_\zeta(\omega)\, d\omega} = \frac{h(Tx, \omega)}{\int_\Omega h(Tx, \omega)\, P_\zeta(\omega)\, d\omega} = r_\omega(Tx)$$

and

(7) $$P_{\xi|\omega}(x) = r_\omega(Tx)\, P_\xi(x)\,.$$

Using (7) we obtain

(8) $$P_{\zeta|A}(B) = \frac{1}{P_\xi(A)} \int_A dx \int_B P_{\xi|\omega}(x)\, P_\zeta(\omega)\, d\omega = \\ = \frac{1}{P_\xi(A)} \int_A P_\xi(x)\, dx \int_B r_\omega(Tx)\, P_\zeta(\omega)\, d\omega$$

where $A \in \mathfrak{X}$, $B \in \mathfrak{A}$. $P_{\zeta|A}(B)$ may be also expressed as

(9) $$P_{\zeta|A}(B) = \frac{1}{P_\xi(A)} \int_A P_{\zeta|x}(B)\, dP_\xi(x) = \frac{1}{P_\xi(A)} \int_A P_\xi(x)\, dx \int_B P_{\zeta|x}(\omega)\, d\omega$$

Now, assuming that $A = A_y \in \mathscr{A}_T$ we conclude that for $x \in A_y$ $r_\omega(Tx)$ does not depend upon x. Moreover, the left hand sides of (8) and (9) become $P_{\zeta|y}(B)$. Next, comparing the right hand sides of (8) and (9) we conclude that $\int_B P_{\zeta|x}(\omega)\, d\omega$ does not depend upon $x \in A_y$. This means that (6) holds for all $x \in A_y$ and $B \in \mathfrak{A}$ such that $\int_B P_{\xi|\omega}(x)\, P_\zeta(\omega)\, d\omega > 0$. Conversely, suppose (6) together with (5) holds. Since

$$P_{\zeta|x}(\omega) = \frac{P_{\xi|\omega}(x)\, P_\zeta(\omega)}{P_\xi(x)}$$

and

$$P_{\zeta|y}(\omega) = \frac{P_{\eta|\omega}(y)\, P_\zeta(\omega)}{P_\eta(y)}$$

we obtain

$$P_{\xi|\omega}(x) = \frac{P_{\eta|\omega}(y)}{P_\eta(y)}\, P_\xi(x)$$

or

$$P_{\xi|\omega}(x) = \frac{P_{\eta|\omega}(Tx)}{P_\eta(Tx)}\, P_\xi(x)$$

which making appriopriate definitions is equivalent to the necessary and sufficient condition (given by factorization theorem) for a random variable T to be a sufficient statistic.

This completes the proof of the proposition.

4. PARTITIONS WHICH ARE ε-SUFFICIENT ON A SAMPLE SPACE. GENERAL CONSIDERATIONS

Let T be any measurable transformation from the measurable space $(X, \mathfrak{X})$ onto a measurable space $(Y, \mathfrak{Y})$ as stated in Introduction. We have by definition

$$P_{\zeta|y}(B) = \frac{P_{\zeta\eta}(B, y)}{P_\eta(y)} = \frac{P_{\zeta\xi}(B, A)}{P_\xi(A)}\,, \quad B \in \mathfrak{A}\,, \quad A \in \mathfrak{X} \tag{10}$$

where

$$A = T^{-1}y = \{x : Tx = y\}\,. \tag{11}$$

From (10)

$$P_{\zeta|y}(B) = \frac{P_{\zeta\xi}(B, A)}{P_\eta(y)} = = \frac{1}{P_\eta(y)} \int_A P_{\zeta|x}(B)\, dP_\xi \tag{12}$$

which is equivalent to (6), provided A is an element of a sufficient partition. Note that $[P_\eta]$ the following relations hold:

$$\frac{P_\xi(x)}{P_\eta(y)} \geqq 0\,, \quad x \in X\,, \quad y \in Y \tag{13}$$

$$\frac{1}{P_\eta(y)} \int_A \mathrm{d}P_\xi = 1\,, \quad A \in \mathfrak{X} \quad \text{and} \quad A = T^{-1}y\,. \tag{14}$$

Let h be any concave function. Then, since (13) and (14) hold, we can apply to (12) Jensen's inequality. We get then

$$h\left(\frac{1}{P_\eta(y)} \int_A P_{\zeta|x}(B)\,\mathrm{d}P_\xi\right) \geqq \frac{1}{P_\eta(y)} \int_A h(P_{\zeta|x}(B))\,\mathrm{d}P_\xi\,. \tag{15}$$

Taking the integral of the both sides of (15) over the space $\Omega \times Y$, we obtain

$$\int_\Omega \mathrm{d}\omega \int_Y P_\eta(y)\, h(P_{\zeta|y}(\omega))\,\mathrm{d}y \geqq \int_\Omega \mathrm{d}\omega \int_X P_\xi(x)\, h(P_{\zeta|x}(\omega))\,\mathrm{d}x\,. \tag{16}$$

If we denote

$$\varepsilon' = \int_\Omega \mathrm{d}\omega \int_Y P_\eta(y)\, h(P_{\zeta|y}(\omega))\,\mathrm{d}y - \int_\Omega \mathrm{d}\omega \int_X P_\xi(x)\, h(P_{\zeta|x}(\omega))\,\mathrm{d}x$$

and if we choose

$$h(P) = -P \log P \tag{17}$$

where the basis of logarithms is equal to 2, then, from (16) we get

$$\begin{aligned} &-\int_\Omega \mathrm{d}\omega \int_Y P_{\zeta\eta}(\omega, y) \log P_{\zeta|y}(\omega)\,\mathrm{d}y \,+ \\ &+ \int_\Omega \mathrm{d}\omega \int_X P_{\zeta\xi}(\omega, x) \log P_{\zeta|x}(\omega)\,\mathrm{d}x \leqq \varepsilon \end{aligned} \tag{18}$$

where $\varepsilon \geqq \varepsilon'$ is a non-negative number.

Definition 3. The measurable transformation T from the measurable space $(X, \mathfrak{X})$ onto a measurable space $(Y, \mathfrak{Y})$ is ε-sufficient if the inequality (18) holds with $\varepsilon > 0$. The partition $\mathscr{A}_T$ of X determined by such T is said to be ε-sufficient on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega})$.

Remark 1. It is obvious that if and only if $\varepsilon = 0$ in (18) then T is sufficient.

The concept of ε-sufficient transformation was in the different way first introduced by Perez [2] and was studied by him in [3, 4, 5]. The equivalence of both definitions is straightforward.

Equality (3) defines the Bayes risk. As stated by Theorem 1 the Bayes risk remains unchanged if X is replaced by its sufficient partition $\mathscr{A}_T$. This is, however, no longer the case if $\mathscr{A}_T$ is an ε-sufficient partition. This means that in the case of ε-sufficient data reduction we do need an estimate of the Bayes risk increase. Such an estimate is given by the Perez's theorem to be found in [3]. See also Perez [4, 5] for further considerations.

5. ε-SUFFICIENT DATA REDUCTION. CONSTRUCTIVE RESULTS

The class of sample spaces to be considered in this section is that with finite parameter space Ω; a member of this class is denoted by $(X, \mathfrak{X}, \Omega, P_{\xi}|_{\omega_i})$. Let M be the number of points ω_i in Ω. We will give an algorithm for constructing an ε-sufficient partition. This partition turns out to be finite. The considerations of this section extend the earlier results of the author presented in [6] and [7].

Let us assume, for a moment without any motivation, that in the case of finite parameter space it is possible for any positive value ε to construct an ε-sufficient partition of X which is finite. This implies the finiteness of the space Y. Let K be the number of elements A_j in $\mathscr{A}_T$ (or the number of elements y_j in Y). With these assumptions we can replace the inequality (18) by

$$(19) \qquad -\sum_{\Omega}\sum_{j=1}^{K} P_{\zeta\xi}(\omega, A_j)\log P_{\zeta|A_j}(\omega) + \sum_{\Omega}\int_X P_{\zeta|x}(\omega)\log P_{\zeta|x}(\omega)\,\mathrm{d}P_{\xi}(x) \leqq \varepsilon\,.$$

Now, we give an information-theoretic interpretation of the problem of searching an ε-sufficient transformation in the case considered. Let $(\Omega, X, \mathfrak{X}, P_{\xi|\omega_i})$ and $(\Omega, \mathfrak{A}, P_{\zeta})$ be a semicontinuous channel and a source of information, respectively. The average amount of information per transmission received through the channel is given by

$$R = H(\Omega) - H(\Omega \mid X)$$

where

$$H(\Omega) = -\sum_{\Omega} P_{\zeta}(\omega)\log P_{\zeta}(\omega)\,,$$

$$H(\Omega \mid X) = -\sum_{\Omega}\int_X P_{\zeta|x}(\omega)\log P_{\zeta|x}(\omega)\,\mathrm{d}P_{\xi}(x)\,.$$

It was proved by Feinstein [8] that for any $\varepsilon > 0$ one can replace the semicontinuous channel defined above by a discrete one which assures the decrease of the average amount of information per transmission not greater than ε. This means that one can find the transformation T defined above for which

$$(20) \qquad R - R' \leqq \varepsilon$$

with

$$R' = H(\Omega) - H(\Omega \mid \mathscr{A}_T)$$

where

$$H(\Omega \mid \mathscr{A}_T) = -\sum_{\Omega} \sum_{j=1}^{K} P_{\zeta\xi}(\omega, A_j) \log P_{\zeta|A_j}(\omega) .$$

Clearly, (20) is equivalent to (19). This leads us to the following assertion:

Theorem 2. *If the parameter space Ω is finite, then for any positive ε there exists an ε-sufficient measurable transformation T from the measurable space $(X, \mathfrak{X})$ onto a finite measurable space $(Y, \mathfrak{X})$. This means that the corresponding ε-sufficient partition of X is finite.*

We propose an algorithm for constructing an ε-sufficient partition $\mathscr{A}_T$ of X This algorithm is based on the proof of Feinstein's theorem, given in [8].

To construct an appriopriate set $\mathscr{A}_T = \{A_j\}$ we explicitly put

$$-\log P_{\zeta|x}(\omega_i) = 0 \quad \text{for} \quad P_{\zeta|x}(\omega_i) = 0 \,, \quad \omega_i \in \Omega \,. \tag{21}$$

Then, denoting by m any positive integer, we define the following sets:

$$\Lambda_m(\omega_i) = \{x : -\log P_{\zeta|x}(\omega_i) < m\} \,, \quad \omega_i \in \Omega \,, \tag{22}$$

$$\Lambda_m = \bigcap_{\Omega} \Lambda_m(\omega_i) \,. \tag{23}$$

Algorithm 1

1° Find the smallest subscript $m = m_0$ such that

$$- \sum_{\Omega} P_{\zeta\xi}(\omega_i, X \setminus \Lambda_{m_0}) \log P_{\zeta\xi}(\omega_i, X \setminus \Lambda_{m_0}) + P_{\xi}(X \setminus \Lambda_{m_0}) \log P_{\xi}(X \setminus \Lambda_{m_0}) \leqq \gamma \tag{24}$$

where

$$P_{\zeta\xi}(\omega_i, X \setminus \Lambda_{m_0}) = P_{\xi|\omega_i}(X \setminus \Lambda_{m_0}) \, P_{\zeta}(\omega_i) \,, \tag{25}$$

$$P_{\xi}(X \setminus \Lambda_{m_0}) = \sum_{\Omega} P_{\zeta\xi}(\omega_i, X \setminus \Lambda_{m_0}) \tag{26}$$

and γ is a positive number such that

$$\gamma < \varepsilon \tag{27}$$

where ε is a positive number chosen before.

As a result of this step of the algorithm we obtain the set Λ_{m_0}.

2° Choose the smallest positive integer n such that

$$\frac{1}{n} P_{\xi}(\Lambda_{m_0}) \leqq \varepsilon - \gamma \,. \tag{28}$$

3° Find

$$P_{\max} = \sup_{\Omega \times \Lambda m_0} P_{\zeta|x}(\omega_i) \tag{29}$$

and then

$$k_{\min} = [1 - n \log P_{\max}] . \tag{30}$$

4° For each $\omega_i \in \Omega$ construct the following sequence of sets

$$\Lambda_{k,\omega_i} = \{x : 2^{-k/n} < P_{\zeta|x}(\omega_i) \leqq 2^{-(k-1)/n}\} \cap \Lambda_{m_0} \tag{31}$$

$$k = k_{\min}, \ldots, nm_0 ,$$

$$\Lambda_{0,\omega_i} = \{x : P_{\zeta|x}(\omega_i) = 0\} \cap \Lambda_{m_0} . \tag{32}$$

5° Construct all the following sets:

$$\Lambda(k_1, k_2, \ldots, k_M) = \bigcap_{i=1}^{M} \Lambda_{k_i,\omega_i} \tag{33}$$

where

$$k_i = 0, k_{\min}, k_{\min} + 1, \ldots, nm_0 . \tag{34}$$

It is proved in the sequel that as a result of this step we obtain the set

$$\mathscr{A}_T = \{A_j\} = \{\{\Lambda(k_1, k_2, \ldots, k_M)\} , \quad X \setminus \Lambda_{m_0}\} . \tag{35}$$

Now we make some remarks which will be found helpful in proving that the formulated algorithm possesses the desired properties.

Remark 2. The sequence Λ_m is a non-decreasing sequence of sets. Therefore

$$\lim \Lambda_m = \bigcup_{m=1}^{\infty} \Lambda_m = X . \tag{36}$$

Remark 3. It follows from relations (29), (30), (31) and (32) that

$$\bigcup_{k=0,k_{\min},\ldots,nm_0} \Lambda_{k,\omega_i} = \Lambda_{m_0} , \quad \omega_i \in \Omega , \tag{37}$$

$$\bigcup_{k=0,k_{\min},\ldots,nm_0} \Lambda_{k,\omega_i} = \emptyset , \quad \omega_i \in \Omega . \tag{38}$$

The same assertions are clearly true for the sets $\Lambda(k_1, k_2, \ldots, k_M)$.

Remark 4. The set $\mathscr{A}_T$ given by (35) is a partition of the space X (it is motivated by Remarks 2 and 3).

Remark 5. Since $0 < \gamma < \varepsilon$ and γ may assume any value from this interval, then for the fixed value ε we can obtain uncountably many partitions $\mathscr{A}_T$ of the space X.

We formulate now the theorem concerning Algorithm 1.

Theorem 3. *Let* $(X, \mathfrak{X}, \Omega, P_{\xi|\omega_i})$ *be a given sample space with the parameter space* Ω *which consists of M points, and let also the a priori probability measure* $P_\zeta(\omega)$ *be given. Then a partition* $\mathscr{A}_T = \{A_j\}$ *of the space X obtained as a result of Algorithm 1 is ε-sufficient.*

Proof. Taking into account Remark 2 and $\lim_{a\to 0} a \log a = 0$ we conclude that it is possible to find $m = m_0$ such that for the chosen positive $\gamma < \varepsilon$ and any $\varepsilon < 0$ we will have

$$(39) \qquad -\sum_{\omega_i \in \Omega} P_{\zeta\xi}(\omega_i, X \setminus \Lambda_{m_0}) \log P_{\zeta|X \setminus \Lambda m_0}(\omega_i) =$$

$$= -\sum_{\omega_i \in \Omega} P_{\zeta\xi}(\omega_i, X \setminus \Lambda_{m_0}) \log P_{\zeta\xi}(\omega_i, X \setminus \Lambda_{m_0})$$

$$+ P_\xi(X \setminus \Lambda_{m_0}^r) \log P_\xi(X \setminus \Lambda_{m_0}) \leqq \gamma$$

where the set Λ_{m_0} is defined in the step 1° of Algorithm 1.

If $r > 0$, then

$$(40) \qquad 2^{-r/n} < P_{\zeta|x}(\omega_i) \leqq 2^{-(r-1)/n}$$

for all x belonging to any $\Lambda(k_1, k_2, \ldots, k_M)$ for which $k_i = r$, where n is defined by (28) and the sets $\Lambda(k_1, k_2, \ldots, k_M)$ are defined by (33). Since

$$(41) \qquad P_{\zeta|\Lambda(k_1,k_2,\ldots,k_M)}(\omega_i) = \frac{1}{P_\xi(\Lambda(k_1, k_2, \ldots, k_M))} \int_{\Lambda(k_1,k_2,\ldots,k_M)} P_{\zeta|x}(\omega_i) \, \mathrm{d}P_\xi$$

the same inequality should be true for $P_{\zeta|\Lambda(k_1,k_2,\ldots,k_M)}(\omega_i)$ except those cases when $P_\xi(\Lambda(k_1, k_2, \ldots, k_M)) = 0$, which means that $P_{\zeta\xi}(\omega_i, \Lambda(k_1, k_2, \ldots, k_M)) = 0$.

Further, if $k_i = 0$, then

$$(42) \qquad P_{\zeta\xi}(\omega_i, \Lambda(k_1, k_2, \ldots, k_M)) = 0$$

for all corresponding sets $\Lambda(k_1, k_2, \ldots, k_M)$.

Define on Λ_{m_0} for each ω_i the following function (recall Remark 3):

$$(43) \qquad g(\omega_i, x) = \begin{cases} -\log P_{\zeta|\Lambda(k_1,k_2,\ldots,k_M)}(\omega_i) & \text{if } x \in \Lambda(k_1, k_2, \ldots, k_M) \\ \quad \text{such that } P_{\zeta\xi}(\omega_i, \Lambda(k_1, k_2, \ldots, k_M)) > 0\,, & \\ \text{any value at all other points of } \Lambda_{m_0}\,. & \end{cases}$$

Thus

$$(44) \qquad |-\log P_{\zeta|x}(\omega_i) - g(\omega_i, x)| \leqq \frac{1}{n} \; [P_{\zeta\xi}]$$

 on A_{m_0}, $i = 1, 2, \ldots, M$. Since $-\log P_{\zeta|x}(\omega_i)$ and $g(\omega_i, x)$ are positive $[P_{\zeta\xi}]$, we have

$$(45)\qquad \int_{Am_0} \log P_{\zeta|x}(\omega_i)\, \mathrm{d}P_{\zeta\xi}(\omega_i, x) \geqq -\int_{Am_0} g(\omega_i, x)\, \mathrm{d}P_{\zeta\xi}(\omega_i, x) - $$

$$- \frac{1}{n} P_{\zeta\xi}(\omega_i, A_{m_0}) .$$

But

$$(46)\qquad \int_{Am_0} \varrho(\omega_i, x)\, \mathrm{d}P_{\zeta\xi}(\omega_i, x) =$$

$$= - \sum_{(k_1, k_2, \ldots, k_M)} P_{\zeta\xi}(\omega_i, A(k_1, k_2, \ldots, k_M)) \log P_{\zeta|A(k_1, k_2, \ldots, k_M)}(\omega_i) .$$

So that, taking into account the definition of conditional entropy $H(\Omega \mid X)$ given above, and Eqs. (45) and (46), we obtain

$$(47)\qquad H(\Omega \mid X) \geqq - \sum_{i=1}^{M} \int_{Am_0} \log P_{\zeta|x}(\omega_i)\, \mathrm{d}P_{\zeta\xi}(\omega_i, x)$$

$$\geqq - \sum_{i=1}^{M} \{ \sum_{(k_1, \ldots, k_M)} P_{\zeta\xi}(\omega_i, A(k_1, k_2, \ldots, k_M)) \log P_{\zeta|A(k_1, k_2, \ldots, k_M)}(\omega_i) -$$

$$- \frac{1}{n} P_{\zeta\xi}(\omega_i, A_{m_0}) \} .$$

Now, taking into account (28) and (39), we obtain from (47) the following inequalities:

$$(48)\quad H(\Omega \mid X) \geqq - \sum_{i=1}^{M} \sum_{(k_1, \ldots, k_M)} P_{\zeta\xi}(\omega_i, A(k_1, k_2, \ldots, k_M)) \log P_{\zeta|A(k_1, k_2, \ldots, k_M)}(\omega_i)$$

$$- \frac{1}{n} P_{\xi}(A_{m_0}) - \sum_{i=1}^{M} P_{\zeta\xi}(\omega_i, X \setminus A_{m_0}) \log P_{\zeta|X \setminus Am_0}(\omega_i) - \gamma$$

$$\geqq - \sum_{i=1}^{M} \sum_{Aj \in \mathscr{A}_T} P_{\zeta\xi}(\omega_i, A_j) \log P_{\xi|A_j}(\omega_i) - \varepsilon$$

where $\mathscr{A}_T$ is defined by (35). One can very easily see that the inequality (48) is equivalent to the inequality (20). This proved ε-sufficiency of the partition $\mathscr{A}_T$ of X.

This completes the proof of the theorem.

One can easily see that since $0 < \gamma < \varepsilon$ there exists some optimal γ which minimizes

$$(49)\qquad \Delta = nm_0 - [1 - n \log P_{\max}] ,$$

i.e., γ corresponding to the minimal amount of the computational work to be done in steps 4° and 5° of Algorithm 1. The problem of this optimization, however, is in fact not very important from the view-point of applications of ε-sufficient data reduction.

In practical cases one will:

1. Restrict the computation to some bounded space X.
2. Assume some "regular" partition $\mathscr{S} = \{S\}$ of the space X, where the sets $S \in \mathscr{S}$ are "sufficiently small".
3. Assign the values $P_{\zeta|S}(\omega_i)$, $i = 1, 2, \ldots, M$. found experimentally, to all points $x \in S$. This means that one will have in X only points x such that either $P_{\zeta|x}(\omega_i) = 0$ or $P_{\zeta|x}(\omega_i) > 0$ with the condition

$$P_{\zeta|x}(\omega_i) > \varrho \tag{50}$$

fulfilled for every x and each ω_i for which $P_{\zeta|x}(\omega_i) \neq 0$, where $\varrho > 0$ is small.

4. Construct $\mathscr{A}_T = \{A_j\}$, where every $A_j \in \mathscr{A}_T$ will consist of at least one set $S \in \mathscr{S}$, according to Algorithm 2, which is given below and is the obvious modification of Algorithm 1.

Algorithm 2

1° Find the smallest positive value of $P_{\zeta|S}(\omega_i)$, $i = 1, 2, \ldots, M$,

$$P_{\min} = \inf_{\substack{\Omega \times \mathscr{S} \\ P_{\zeta|S}(\omega_i) \neq 0}} P_{\zeta|S}(\omega_i)$$

and then choose

$$m_0 = [-\log P_{\min}] + 1 .$$

2° Choose the smallest positive integer n such that

$$\frac{1}{n} \leqq \varepsilon$$

where ε is a positive number chosen before.

3° Proceed as in Algorithm 1 with x replaced by S.

4° Proceed as in Algorithm 1 with x replaced by S and the operations of intersection with Λ_{m_0} deleted.

5° Proceed as in Algorithm 1.

As a result of Algorithm 2 one obtains a partition

$$\mathscr{A}_T = \{\Lambda(k_1, k_2, \ldots, k_M)\} .$$

We remark that such a partition will be ε-sufficient with respect to the computed probability measures $P_{\zeta|S}(\omega_i)$, $i = 1, 2, \ldots, M$, $S \in \mathscr{S}$.

Even if $\{P_{\xi|\omega_i}\}$ and P_ζ are exactly known one can assume some "convenient" partition $\mathscr{S}$ of X and compute "how sufficient" is this partition, i.e., one can compute the number

$$\varepsilon_{\mathscr{S}} = -\sum_{i=1}^{M} \sum_{S_j \in \mathscr{S}} P_{\zeta\xi}(\omega_i, S_j) \log P_{\zeta|\mathscr{S}_j}(\omega_i) + \sum_{i=1}^{M} \int_X \log P_{\zeta|x}(\omega_i) \, \mathrm{d}P_{\zeta\xi}(\omega_i, x)$$

Then, if a partition $\mathscr{A}_T$ should be ε-sufficient one can find

$$\varepsilon_{\mathscr{A}_T|\mathscr{S}} = \varepsilon - \varepsilon_{\mathscr{S}}$$

provided $\varepsilon > \varepsilon_{\mathscr{S}}$. Further, using Algorithm 2 one can find $\mathscr{A}_T$ being an $\varepsilon_{\mathscr{A}_T|\mathscr{S}}$-sufficient partition with respect to $\mathscr{S}$ and being ε-sufficient on $(X, \mathfrak{X}, \Omega, P_{\xi|\omega_i})$.

(Received July 10, 1969.)

REFERENCES

[1] D. Blackwell, M. A. Girshick: Theory of Games and Statistical Decisions. John Wiley and Sons, Inc., N. Y. 1954.

[2] A. Perez: Notions generalisees d'incertitude, d'entropie et d'information du point de vue de la theorie de martingales. Trans. 1st Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague 1957, 183—208.

[3] A. Perez: Information, ε-sufficiency and Data Reduction Problems. Kybernetika *1* (1965), 4, 297—323.

[4] A. Perez: Information Theory Methods in Reducing Complex Decision Problems. Trans. 4th Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, Prague 1967, 55—87.

[5] A. Perez: Information-theoretic Risk Estimates in Statistical Decisions. Kybernetika 3 (1967), 1, 1—21.

[6] J. Białasiewicz: Data Reduction Problems in Pattern Recognition. Proceedings of IFAC Symposium on Technical and Biological Problems in Control, held in Yerevan, September 1968.

[7] J. Białasiewicz: On Sufficiency, ε-sufficiency and Data Reduction Algorithms. Department of Mathematics, Oragon State University, Techn. Report No. 42, December 1968.

[8] A. Feinstein: Foundations of Information Theory. McGraw-Hill, N. Y. 1958.

Dr. Jan Bialasiewicz, Industrial Institute for Automation and Measurements, Al. Jerozolimskie 202, Warszawa, Poland.

Statistická redukce dat pomocí konstrukce rozkladů výběrového prostoru

JAN BIAŁASIEWICZ

V článku je problém statistické redukce dat formulován jako problém konstrukce rozkladů výběrového prostoru. Uvažují se suficientní a ε-suficientní rozklady. Je uvedena nová definice suficientní statistiky, ze které plyne definice ε-suficientního rozkladu, jež je ekvivalentní definici Perezově. Nová definice suficientní statistiky umožnila dokázat, že pro konečný parametrový prostor je problém syntézy ε-suficientního rozkladu ekvivalentní problému redukce polospojitého kanálu na diskrétní kanál, nemá-li pokles průměrné informace na přenos překročit ε. To umožnilo odvodit algoritmus pro syntézu ε-suficientního rozkladu výběrového prostoru inspirovaný prací Feinsteina [8]. Je předložena a studována modifikace tohoto algoritmu.

Dr. Jan Białasiewicz, Industrial Institute for Automation and Measurements, Al. Jerozolimskie 202, Warszawa, Poland.