KYBERNETIKA — VOLUME 13 (1977), NUMBER 6

ε-Admissible Simplifications of the Dependence Structure of a Set of Random Variables*)

ALBERT PEREZ

After introducing the concept of simplification of the dependence structure of a set of random variables, the paper is concerned with the so-called ε-admissible simplifications and their information-theoretic analysis. The concept of ε-sufficiency, previously introduced by the author for characterizing reduced sample σ-algebras (statistics) in a Bayes statistical decision problem, may be considered as a special case of ε-admissibility of dependence structure simplification. The role of Shannon's information and its loss caused by a data reduction is here replaced by the concept of dependence tightness and its loss caused by a simplification of the dependence structure of a set of random variables.

1. INTRODUCTION

The handling of sets of mutually dependent random variables for theoretical or applied purposes may become so complicated that the need arises for simplifying their dependence structure.

The most simple dependence structure is that of mutually independent random variables. Therefore it may serve as a reference point for measuring the *dependence tightness* of a set of random variables. A good measure of this kind is the relative entropy (divergence) of Shannon's type of the probability measure of the initial set of random variables with respect to the corresponding product probability measure of their marginals.

Let, thus,

$$A = \{X_1, X_2, \ldots, X_N\} \tag{1.1}$$

*) This paper is an English version of the author's contribution at the Colloque International du C.N.R.S.: *Les développements récents de la Théorie de l'information et leurs applications*, E.N.S.E.T., Cachan (France), 4–8 july 1977.

be a set of (abstract valued) random variables with a joint probability measure $P_{1...N}$ and with marginals denoted by $P_1, P_2, \ldots, P_N$, respectively. By their *dependence tightness*, denoted by $I(X_1, \ldots, X_N)$ resp. by $I(P_{1...N})$, we understand the quantity defined as the relative entropy, $H(P_{1...N}, P_1 \times \ldots \times P_N)$, of $P_{1...N}$ with respect to the product measure $P_1 \times \ldots \times P_N$, i.e.

$$(2.1) \qquad I(X_1, \ldots, X_N) = I(P_{1...N}) = H(P_{1...N}, P_1 \times \ldots \times P_N) = \\ = \int \log \frac{dP_{1...N}}{dP_1 \times \ldots \times dP_N} \, dP_{1...N} \,,$$

provided that $P_{1...N}$ is absolutely continuous with respect to $P_1 \times \ldots \times P_N$. Otherwise, this entropy is defined to be equal to $+\infty$.

As a relative or generalized entropy (concept introduced and studied many years ago, for instance, in [1]) the dependence tightness is non-negative and equals zero iff the random variables in question are mutually independent. Further, as in the case of mutual information (which may be considered to be a dependence tightness of a pair of random variables), the dependence tightness of a set of random variables produced by reducing the σ-algebras corresponding to the initial random variables $X_1, \ldots, X_N$ is smaller or equal than $I(X_1, \ldots, X_N)$ with equality obtained iff the resulting Cartesian-product reduced σ-algebra is *sufficient* with respect to the pair of measures $(P_{1...N}, P_1 \times \ldots \times P_N)$. In the same context, the dependence tightness may be approximated by that of a suitable set of random variables produced by taking finite measurable partitions of the spaces corresponding to the initial random variables, i.e. may be approximated by similar quantities introduced by S. Watanabe [2] in 1960 for measuring the degree of the statistical interrelation (multivariate correlation) of a set of discrete random variables.

Obviously, it would be possible in introducing the concept of dependence tightness to apply other measures of divergence (generalized f-entropies) of $P_{1...N}$ with respect to $P_1 \times \ldots \times P_N$, similarly as in studying Data reduction problems beginning from [3] we arrived at [5]. However, as we shall see below, the Shannon's generalized entropy (corresponding to the convex function $f(u) = u \log u$) used in the definition (2.1) presents certain advantages due to its additivity properties.

Let us, for instance, take an arbitrary subset $E = \{X_{i_1}, X_{i_2}, \ldots, X_{i_m}\}$ of the set A of random variables given in (1.1). Then

$$(3.1) \qquad I(X_1, \ldots, X_N) = I(X_{i_1}, X_{i_2}, \ldots, X_{i_m}) + I(X_{i_{m+1}}, \ldots, X_{i_N}) + \\ + I([X_{i_1}, \ldots, X_{i_m}], [X_{i_{m+1}}, \ldots, X_{i_N}]) \,,$$

or, more compactly,

$$(4.1) \qquad I(A) = I(E) + I(A - E) + I(E, A - E) \,.$$

In words, the dependence tightness of the set A of random variables is equal to the sum of the dependence tightnesses of the set E and its complement $A - E$ and of the information (dependence tigthness) of the pair of vector random variables with their components contained in E and $A - E$, respectively.

The concept of simplification of the dependence structure of a set A of random variables will be defined in the next section. As a result of such simplification the joint probability measure $P_{1\ldots N}$ is modified to some other joint probability measure $\bar{P}_{1\ldots N}$.

In order to judge the admissibility of the simplification we apply again the relative entropy $H(P_{1\ldots N}, \bar{P}_{1\ldots N})$. We shall see that this entropy is equal to the *loss of dependence tightness* on passing from the original set of random variables to the simplified one. If this loss is no greater than a given positive ε we shall say that the simplification in question is *ε-admissible.*

In the sequel we shall compare this notion of ε-admissibility with the concept of *ε-sufficiency*, previously introduced by the author (cf. [3], [4], [5]) for characterizing reduced sample σ-algebras (statistics) in a Bayes statistical decision problem. The latter means that the *loss of Shannon's information* on the parameter r.v. contained in the sample r.v. caused by reduction is equal or less than ε, what implies that the loss of decision quality (risk increase) thus resulting is "small" the smaller is ε. If, now, instead of reduction, we simplify the dependence structure of the sets of random variables corresponding to the different statistical hypotheses by taking ε-admissible simplifications, the situation remains the same up to the fact that the loss of information is here replaced by the average loss of dependence tightness.

The concept of ε-sufficiency may be considered as a special case of ε-admissibility of dependence structure simplification.

For obtaining suitable simplifications of the dependence structure of a set of random variables a method is described which is based on forming "coalition structures" of the random variables by using relation (4.1) in such a way that the random variables in a coalition have a relatively high dependence tightness whereas the dependence between coalitions may be considered as negligible (memoryless coalition structure simplification) or may be replaced by a suitable Markov chain (Markov coalition structure simplification) in order that the total loss of dependence tightness is as small as possible.

2. DEPENDENCE STRUCTURE SIMPLIFICATION

Let us first introduce the concept of elementary simplification of the dependence structure of a set of random variables $A = \{X_1, X_2, \ldots, X_N\}$ with joint probability measure $P_{1\ldots N}$ (cf. (1.1)). For this purpose take an arbitrary proper subset $E = \{X_{i_1}, X_{i_2}, \ldots, X_{i_m}\}$ of the set A and, further, an arbitrary proper subset $F = \{X_{j_1}, X_{j_2}, \ldots, X_{j_k}\}$ of the set E. Let us denote by $P_{A-E|E}$ and by $P_{A-E|F}$ the conditional probabilities of the vector r.v. having its components in the set $A - E$, i.e. of $[X_{i_{m+1}}, \ldots, X_{i_N}]$, given the value of the vector r.v. $[X_{i_1}, \ldots, X_{i_m}]$ or the value of the vector r.v. $[X_{j_1}, \ldots, X_{j_k}]$, respectively.

Obviously, the joint probability measure $P_{1\ldots N}$, denoted more compactly by P_A, is generated by P_E and $P_{A-E|E}$. Similarly, let us denote by $\bar{P}_A$ the probability measure generated by P_E and $P_{A-E|F}$. By *elementary* (E, F)-*simplification* of the dependence structure of a set A of random variables, where $F \subset E \subset A$, we understand that the original joint probability measure P_A is replaced by $\bar{P}_A$. Let us remark that by elementary simplifications the identity of the random variables as such remains unchanged in the sense that this is the case for the corresponding marginal probabilities. The dependence tightness loss connected with the elementary (E, F)-simplification above is given by (cf. (2.1))

$$(1.2) \qquad H(P_A, \bar{P}_A) = I(P_A) - I(\bar{P}_A)\,.$$

The sense of an elementary simplification consists in replacing an original conditional probability, $P_{A-E|E}$, by one of its conditional expectations, $P_{A-E|F}$,

$$(2.2) \qquad P_{A-E|F} = \mathscr{E}_{P_A}\{P_{A-E|E}/\mathfrak{F}\}\,,$$

where by $\mathfrak{F}$ we denote the σ-algebra corresponding to the set F of random variables, $\mathfrak{F} \subset \mathfrak{E}$, with respect to which the conditional expectation is taken, being, thus, $\mathfrak{F}$-measurable.

From this point of view, the concept of elementary simplification may be extended in order to concern the dependence structure of a general probability space $(X, \mathfrak{X}, P)$. Let, thus, $\mathfrak{X}'$ be a sub-σ-algebra of the σ-algebra $\mathfrak{X}$, and $\mathfrak{X}''$ a sub-σ-algebra of $\mathfrak{X}'$. Let, further, $P(\cdot/\mathfrak{X}', x)$ be the conditional probability function corresponding to P and to $\mathfrak{X}'$ (being, thus, $\mathfrak{X}'$-measurable) and $P(\cdot/\mathfrak{X}'', x)$ that corresponding to P and to $\mathfrak{X}''$. Obviously, the probability measure P is generated by its restriction P' on $\mathfrak{X}'$ and the conditional probability function $P(\cdot/\mathfrak{X}', x)$. Similarly, let us denote by $\bar{P}$ the probability measure on $\mathfrak{X}$ generated by P' and $P(\cdot/\mathfrak{X}'', x)$. By *elementary* $(\mathfrak{X}', \mathfrak{X}'')$-*simplification* of the dependence structure of the probability space $(X, \mathfrak{X}, P)$, where $\mathfrak{X}'' \subset \mathfrak{X}' \subset \mathfrak{X}$, we understand that the original probability space $(X, \mathfrak{X}, P)$ is replaced by $(X, \mathfrak{X}, \bar{P})$.

After this observation, let us introduce the concept of simplification of the dependence structure of a set A of random variables as a superposition of successive elementary simplifications. Essentially it consists of a cumulation of a certain number of compatible elementary simplifications in the following sense: After performing an elementary (E, F)-simplification of the set A, we perform a second elementary (E_1, F_1)-simplification of the set of random variables E which plays now the role of A while E_1 plays the role of E and F_1 the role of F, $F_1 \subset E_1 \subset E$. The joint probability measure of the set of random variables E, denoted as before by P_E, will be replaced by $\bar{P}_E$, generated by P_{E_1} and $P_{E-E_1|F_1}$, similarly as in the first step P_A was replaced by $\bar{P}_A$, generated by P_E and $P_{A-E|F}$. As to the probability measure of the original set A of random variables will be in this second step replaced by the joint probability measure $\bar{P}^1_A$, generated by $\bar{P}_E$ and $P_{A-E|F}$, i.e. $\bar{P}^1_A$ is generated by P_{E_1}, $P_{E-E_1|F_1}$ and $P_{A-E|F}$.

Continuing in this way, in the $(n + 1)$-step, i.e. after performing the elementary (E_n, F_n)-simplification of the set E_{n-1}, where $F_n \subset E_n \subset E_{n-1}$, the joint probability measure P_A of the original set of random variables will be replaced by $\bar{P}^n_A$, generated by $P_{E_n}, P_{E_{n-1}-E_n|F_n}, \ldots, P_{E-E_1|F_1}$ and $P_{A-E|F}$.

Hence the definition: By $(E, F; E_1, F_1; \ldots; E_n, F_n)$-*simplification* of the dependence structure of the set A of random variables, where $F_n \subset E_n \subset E_{n-1}, \ldots, F_1 \subset \subset E_1 \subset E, F \subset E \subset A$, we understand that the original joint probability measure P_A is replaced by $\bar{P}^n_A$, resulting by application of the above sequence of compatible elementary simplifications.

Obviously, $(A - E, E - E_1, E_1 - E_2, \ldots, E_{n-1} - E_n, E_n)$ represents a partition of the set A of random variables. In terms of this partition, by consecutive application of the relation (4.1) one obtains the following expression for the dependence tightness of this set with the original joint probability measure P_A,

$$\begin{aligned}(3.2) \qquad I(P_A) = I(A) = I(A - E) + I(E - E_1) + I(E_1 - E_2) + \ldots \\ \ldots + I(E_{n-1} - E_n) + I(E_n) + I(E, A - E) + I(E_1, E - E_1) + \\ + I(E_2, E_1 - E_2) + \ldots + I(E_n, E_{n-1} - E_n),\end{aligned}$$

since

$$\begin{aligned}I(A) = I(A - E) + I(E) + I(E, A - E), \\ I(E) = I(E - E_1) + I(E_1) + I(E_1, E - E_1), \ldots \\ \ldots, I(E_{n-1}) = I(E_{n-1} - E_n) + I(E_n) + I(E_n, E_{n-1} - E_n).\end{aligned}$$

Theorem 1.2. The loss of dependence tightness caused by the simplification above of the dependence structure of the set A of random variables which changes the original joint probability measure P_A to $\bar{P}^n_A$ is given by (cf. (3.2))

$$\begin{aligned}(4.2) \qquad I(P_A) - I(\bar{P}^n_A) = I(E, A - E) - I(F, A - E) + \\ + I(E_1, E - E_1) - I(F_1, E - E_1) + I(E_2, E_1 - E_2) - I(F_2, E_1 - E_2) + \ldots \\ \ldots + I(E_n, E_{n-1} - E_n) - I(F_n, E_{n-1} - E_n)\end{aligned}$$

and may be considered as the sum of dependence tightness losses produced by the corresponding sequence of elementary simplifications, i.e.

$$\begin{aligned}(5.2) \qquad I(P_A) - I(\bar{P}^n_A) = H(P_A, \bar{P}^n_A) = H(P_A, \bar{P}_A) + H(P_E, \bar{P}_E) + \\ + H(P_{E_1}, \bar{P}_{E_1}) + \ldots + H(P_{E_{n-1}}, \bar{P}_{E_{n-1}}) = \\ = I(P_A) - I(\bar{P}_A) + I(P_E) - I(\bar{P}_E) + \ldots + I(P_{E_{n-1}}) - I(\bar{P}_{E_{n-1}}).\end{aligned}$$

Proof. The relation (4.2) results from the fact that $I(\bar{P}^n_A)$ has an expression differing from the expression (3.2) of $I(P_A)$ only as it concerns the informations (dependence tightnesses) of pairs of vector r.v. there figuring: instead of E we have F since the conditional probability $P_{A-E|E}$ is replaced by $P_{A-E|F}, \ldots$, instead of E_n we have F_n since the conditional probability $P_{E_{n-1}-En/En}$ is replaced in constructing $\bar{P}^n_A$ by $P_{E_{n-1}-E_n|F_n}$.

As to the first equality (5.2), it results similarly as (1.2) from (2.1) applied for P_A and for $\bar{P}^n_A$, respectively, taking account of the fact that in both cases the product measure of the marginals is the same and that

$$\log \frac{\mathrm{d}P_A}{\mathrm{d}\bar{P}^n_A} = \log \frac{\mathrm{d}P_A}{\mathrm{d}P_1 \times \ldots \times \mathrm{d}P_N} - \log \frac{\mathrm{d}\bar{P}^n_A}{\mathrm{d}P_1 \times \ldots \times \mathrm{d}P_N}, \tag{6.2}$$

while all the integrations are performed (due to the construction of $\bar{P}^n_A$) with respect to P_A, The other two equalities are obtained in a similar manner. In particular,

$$H(P_A, \bar{P}^n_A) = H(P_A, \bar{P}_A) + H(\bar{P}_A, \bar{P}^1_A) + \ldots + H(\bar{P}^{n-1}_A, \bar{P}^n_A) \tag{7.2}$$

and $H(\bar{P}_A, \bar{P}^1_A) = H(P_E, \bar{P}_E), \ldots, H(\bar{P}^{n-1}_A, \bar{P}^n_A) = H(P_{E_{n-1}}, \bar{P}_{E_{n-1}})$. Thus, the theorem is proved.

In the case of a general probability space $(X, \mathfrak{X}, P)$, the simplification of its dependence structure may also be conceived as resulting from a sequence of compatible elementary simplifications defined as above by the triplets of σ-algebras $(\mathfrak{X}''_1 \subset \subset \mathfrak{X}'_1 \subset \mathfrak{X})$, $(\mathfrak{X}''_2 \subset \mathfrak{X}'_2 \subset \mathfrak{X}'_1), \ldots, (\mathfrak{X}''_n \subset \mathfrak{X}'_n \subset \mathfrak{X}'_{n-1})$. We shall not insist in this paper on this question.

3. ε-ADMISSIBILITY AND ε-SUFFICIENCY IN BAYES STATISTICAL DECISION PROBLEMS

As said in the Introduction, a dependence structure simplification, changing the original probability measure P to $\bar{P}$, is *ε-admissible* if the relative entropy (divergence) $H(P, \bar{P})$ of P with respect to $\bar{P}$, resp. if the loss $I(P) - I(\bar{P})$ of dependence tightness, is no greater than a given positive ε.

In order to understand the significance of this concept let us place, for instance, in the frame of a Bayes statistical decision problem. Let, thus, consider a system $\{P_u, u \in U\}$ of statistical hypotheses-probability measures on a sample measurable space $(X, \mathfrak{X})$. Let $(U, \mathfrak{U})$ be the measurable parameter space.

$$P_Q(\cdot) = \int_U P_u(\cdot)\, \mathrm{d}Q(u) \tag{1.3}$$

is the marginal sample probability measure corresponding to the a priori probability measure Q on $(U, \mathfrak{U})$. Similarly, by QP we shall denote the probability measure on $(U \times X, \mathfrak{U} \times \mathfrak{X})$ generated by Q and $\{P_u, u \in U\}$. Note that, by hypothesis, for every set $E \in \mathfrak{X}$ the u-function $P_u(E)$ is $\mathfrak{U}$-measurable.

Let us, now, for every $u \in U$ consider a dependence structure simplification of P_u to $\bar{P}_u$ and denote by $Q\bar{P}$ the probability measure on $(U \times X, \mathfrak{U} \times \mathfrak{X})$ generated by Q and $\{\bar{P}_u, u \in U\}$. The question arises what is the risk increase if instead of applying an optimal (Bayes resp. epsilon-Bayes) decision function b_0 for discerning the statistical hypotheses P_u, $u \in U$, one applies for the same purpose a decision function $\bar{b}_0$ which is optimal for discerning the simplified statistical hypotheses $\bar{P}_u$, $u \in U$.

In engineering or medical applications, for instance, there is very often the tendency to simplify at the extreme the dependence structure of the observed random vector by taking its components conditionally independent between them for every statistical hypothesis.

A similar question of risk increase was considered by the author in studying Data reduction problems in statistical decision, (see, for instance, references [3], [4], [5], [6]), and there were obtained some interesting estimates of this risk increase in terms of generalized f-entropies (f-divergences) of the probability measure QP before reduction with respect to the probability measure $Q\tilde{P}$ after reduction. All these estimates may be used directly in estimating the risk increase considered above which is due to the dependence structure simplification of the statistical hypotheses. It is sufficient to replace $Q\tilde{P}$ by $Q\bar{P}$. We shall not give here examples of such estimates but we shall concentrate our attention on the relative Shannon's entropy $H(QP, Q\bar{P})$ knowing that the smaller is this entropy the smaller is the risk increase above, obtaining the value zero if $H(QP, Q\bar{P}) = 0$, i.e. iff $QP = Q\bar{P}$.

Theorem 1.3. If the dependence structure simplifications of the different statistical hypotheses in the Bayes statistical decision problem considered above are ε-admissible $[Q]$, then

$$H(QP, Q\bar{P}) \leqq \varepsilon\,. \tag{2.3}$$

If the sample random variable is a vector one and if the simplifications above concern the dependence structure of the set of its components, then

$$H(QP, Q\bar{P}) = \int_U [I(P_u) - I(\bar{P}_u)]\,\mathrm{d}Q(u)\,, \tag{3.3}$$

i.e. it is equal to the average dependence tightness loss.

Proof. Inequality (2.3) results from the relation

$$H(QP, Q\bar{P}) = \int_U H(P_u, \bar{P}_u)\,\mathrm{d}Q(u) \tag{4.3}$$

since the ε-admissibility assumption implies $H(P_u, \bar{P}_u) \leqq \varepsilon$, $u \in U$, [Q].

The equality (3.3) is a consequence of (4.3) and of the assumption concerning the character of the simplifications which implies that (cf. (5.2) of Theorem 1.2) $H(P_u, \bar{P}_u) = I(P_u) - I(\bar{P}_u)$. Thus, the theorem is proved.

Let us, now, recall the concept of ε-sufficiency introduced in [3]. We say that the reduced σ-algebra $\mathfrak{X}' \subset \mathfrak{X}$ is *ε-sufficient* with respect to Q and the system $\{P_u, u \in U\}$ of statistical hypotheses if the loss of Shannon's information on passing from $\mathfrak{X}$ to $\mathfrak{X}'$ is equal or smaller than $\varepsilon > 0$, i.e. if (denoting by η the parameter r.v. and by ξ resp. ξ' the sample r.v. before resp. after reduction, and by P'_u and P'_Q the restrictions of P_u and P_Q on $\mathfrak{X}'$)

$$(5.3)\qquad I(\eta, \xi) - I(\eta, \xi') = \int \log \frac{\mathrm{d}P_u}{\mathrm{d}P_Q}\, \mathrm{d}P_u\, \mathrm{d}Q(u) - \int \log \frac{\mathrm{d}P'_u}{\mathrm{d}P'_Q}\, \mathrm{d}P'_u\, \mathrm{d}Q(u) \leqq \varepsilon\,.$$

More generally, if $\tilde{P}_u$ is any extension of P'_u from $\mathfrak{X}'$ to $\mathfrak{X}$, $u \in U$, and if $Q\tilde{P}$ is generated by Q and $\{\tilde{P}_u, u \in U\}$, we say that $\mathfrak{X}'$ is ε-sufficient if $H(QP, Q\tilde{P}) \leqq \varepsilon$. One finds (see [3]) that for a given Q the best extension of P'_u to $\tilde{P}_u$ in the sense of minimizing the above entropy is obtained by taking

$$(6.3)\qquad \mathrm{d}\tilde{P}_u(x) = \mathrm{d}P'_u(x)\, \mathrm{d}P_Q(x/\mathfrak{X}', x)$$

where $P_Q(\cdot/\mathfrak{X}', x)$ is the conditional probability function corresponding to P_Q and to $\mathfrak{X}'$, i.e. for which it holds

$$(7.3)\qquad \mathrm{d}P_Q(x) = \mathrm{d}P'_Q(x)\, \mathrm{d}P_Q(x/\mathfrak{X}', x)\,.$$

In this case one easily obtains

$$(8.3)\qquad H(QP, Q\tilde{P}) = I(\eta, \xi) - I(\eta, \xi') =$$
$$= \textit{loss of Shannon's information on passing from } \mathfrak{X} \textit{ to } \mathfrak{X}'\,.$$

It is instructive to compare (3.3) of Theorem 1.3 with (8.3).

4. ε-SUFFICIENCY AS SOME KIND OF ε-ADMISSIBILITY OF DEPENDENCE STRUCTURE SIMPLIFICATION

Placing us in the same frame as in Section 3 we remark that $\tilde{P}_u$ as defined by (6.3) cannot be in general considered as resulting from some dependence structure simplification of P_u since, by definition, one should find a sub-σ-algebra $\mathfrak{X}''_u$ of $\mathfrak{X}'$ such that $P_u(\cdot/\mathfrak{X}''_u, x) = P_Q(\cdot/\mathfrak{X}', x)$, $u \in U$, $[Q]$. Thus, in general, this "individualistic" (with respect to each statistical hypothesis) approach does not permit to conceive data reduction as some kind of dependence structure simplification.

However, the latter is acheived if we consider dependence structure simplifications of QP rather than of the inidivudal P_u's. Indeed, it holds

Theorem 1.4. The probability measure $Q\tilde{P}$ on $(U \times X, \mathfrak{U} \times \mathfrak{X})$, generated by Q and $\{\tilde{P}_u, u \in U\}$ as defined by (6.3), may be obtained by dependence structure simplification of QP.

Proof. For QP it holds

$$\mathrm{d}QP(u, x) = \mathrm{d}QP'(u, x)\,\mathrm{d}QP(x/\mathfrak{U} \times \mathfrak{X}', u, x)\,, \tag{1.4}$$

where QP' denotes the restriction of QP on $\mathfrak{U} \times \mathfrak{X}'$ and is generated by Q and $\{P'_u, u \in U\}$, i.e. $\mathrm{d}QP'(u, x) = \mathrm{d}Q(u)\,\mathrm{d}P'(x)$. As to $QP(\cdot/\mathfrak{U} \times \mathfrak{X}', u, x)$ is the conditional probability function of QP which is $\mathfrak{U} \times \mathfrak{X}'$-measurable.

Let us, now, take as sub-σ-algebra of $\mathfrak{U} \times \mathfrak{X}'$ the σ-algebra $\{\emptyset, U\} \times \mathfrak{X}'$ and consider the following dependence structure simplification of QP defined by

$$\mathrm{d}\overline{QP}(u, x) = \mathrm{d}QP'(u, x)\,\mathrm{d}QP(x/\{\emptyset, U\} \times \mathfrak{X}', u, x)\,. \tag{2.4}$$

Obviously, $QP(\cdot/\{\emptyset, U\} \times \mathfrak{X}', u, x) = P_Q(\cdot/\mathfrak{X}', x)$ so that $\overline{QP} = Q\tilde{P}$. The theorem is, thus, proved.

5. A COALITION METHOD FOR OBTAINING SUITABLE SIMPLIFICATIONS OF THE DEPENDENCE STRUCTURE OF A SET OF RANDOM VARIABLES

Placing us in the frame of Section 2, we may see that any dependence structure simplification, conceived as a sequence of elementary simplifications applied one after another on the result obtained by the preceding ones, may be viewed as some $(E, F; E_1, F_1; \ldots; E_n, F_n)$-simplification. Whatever be the case, the final result of a simplification concerning the dependence structure of a set A of random variables may be described by a partition (coalition structure) of the set A in some n disjoint sets $G_1, G_2, \ldots, G_n$, and by a sequence of $n - 1$ subsets $F_1, F_2, \ldots, F_{n-1}$ of the set A such that, for some permutation $G_{i_1}, \ldots, G_{i_n}$ of the sequence $G_1, \ldots, G_n$, it holds

$$F_1 \subset G_{i_1},\ F_2 \subset G_{i_1} \cup G_{i_2}, \ldots, F_{n-1} \subset G_{i_1} \cup G_{i_2} \cup \ldots \cup G_{i_{n-1}}\,. \tag{1.5}$$

The probability measure $\bar{P}_A$ of the simplified dependence structure is then that generated by $P_{G_{i1}}, P_{G_{i2}|F_1}, P_{G_{i3}|F_2}, \ldots, P_{G_{in}|F_{n-1}}$.

Note that P_A is correspondingly generated by:

$$P_{G_{i1}}, P_{G_{i2}|G_{i1}}, P_{G_{i3}|G_{i1} \cup G_{i2}}, \ldots, P_{G_{in}|G_{i1} \cup G_{i2} \cup \ldots \cup G_{in-1}}\,.$$

The dependence tightness $I(\bar{P}_A)$ is given by

$$I(\bar{P}_A) = I(G_{i_1}) + \ldots + I(G_{i_n}) + I(F_1, G_{i_2}) + \ldots + I(F_{n-1}, G_{i_n})\,. \tag{2.5}$$

A *memoryless coalition structure simplification* is obtained if

$$F_1 = \emptyset, \ldots, F_{n-1} = \emptyset \tag{3.5}$$

 and, thus,

(4.5) $$I(\bar{P}_A) = I(G_1) + \ldots + I(G_n).$$

A *Markov* (first order) *coalition structure simplification* is obtained if

(5.5) $$F_1 = G_{i_1}, \ldots, F_{n-1} = G_{i_{n-1}}$$

and, thus,

(6.5) $$I(\bar{P}_A) = I(G_1) + \ldots + I(G_n) + I(G_{i_1}, G_{i_2}) + \ldots + I(G_{i_{n-1}}, G_{i_n}).$$

For obtaining ε-admissible simplifications suitable from the point of view of handling with them one may try to find memoryless ones with the maximal n compatible with the given ε. If this maximal $n = 1$, for instance, and, thus, non interesting, one may try to construct Markov coalition structure simplifications. For a given coalition structure $(G_1, \ldots, G_n)$ it is then desirable to take such a permutation $G_{i_1}, \ldots, G_{i_n}$ that the corresponding dependence tightness, given by (6.5), be maximal in order that the dependence tightness loss $H(P_A, \bar{P}_A) = I(P_A) - I(\bar{P}_A)$ be minimal. For doing this it is possible to proceed as follows.

For every $i_1 = 1, 2, \ldots, n$, take $i_2 \neq i_1$ such that $I(G_{i_1}, G_{i_2}) = \max_{j \neq i_1} I(G_{i_1}, G_j)$; further take i_3 such that $I(G_{i_2}, G_{i_3}) = \max_{j \neq i_1, i_2} I(G_{i_2}, G_j)$, and so on. Let

(7.5) $$S(i_1) = I(G_{i_1}, G_{i_2}) + \ldots + I(G_{i_{n-1}}, G_{i_n})$$

and take I_1 such that

(8.5) $$S(I_1) = \max_{i_1 = 1, \ldots, n} S(i_1).$$

The loss minimizing permutation is then given by $G_{I_1}, \ldots, G_{I_n}$ constructed as above.

However, the main question: how to choose the coalition structure $G_1, \ldots, G_n$, remains open. In the sequel we shall indicate a method which may be useful in searching for coalition structures maximizing $I(G_1) + \ldots + I(G_n)$ for different n's (cf. (4.5) and (6.5)). This method, inspired by coalition game theory, proceeds as follows: G_1 is taken such that, for a given positive parameter c which in the general case may depend on $|G_1|$ (= number of random variables contained in the set G_1), the quantity

(9.5) $$h(G_1) = \frac{I(G_1) - c\, I(G_1, A - G_1)}{|G_1|}$$

is maximal. Remark that by taking G_1 in a memoryless coalition structure simplification, there is a gain of dependence tightness $I(G_1)$ but, at the same time, a definite loss of dependence tightness $I(G_1, A - G_1)$ which in the Markov case (cf. (6.5)) may be partially recuperated. The parameter c is weigting the compromise we make between this gain and this loss per random variable.

Similarly, G_2 is taken as a subset of $A - G_1$ which maximizes

$$h(G_2) = \frac{I(G_2) - c\, I(G_2, A - G_1 - G_2)}{|G_2|} \tag{9.10}$$

and so on. Thus, the number n of coalitions also results and depends on the parameter c. A relative program was constructed [7].

(Received July 28, 1977.)

REFERENCES

[1] A. Perez: 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.

[2] S. Watanabe: Information theoretical analysis of multivariate correlation. IBM J. Res. Develop. *4* (1960), 66–82.

[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. In: Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (1965), Prague 1967, 55–87.

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

[6] A. Perez: Risk estimates in terms of generalized f-entropies. Proceedings of the Colloquium on Information Theory, Debrecen (Hungary), 1967, 299–315.

[7] O. Kříž: A program on optimal coalition structure search, Institute of Information Theory and Automation, 1977, Prague.

Dr. Albert Perez, DrSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.