KYBERNETIKA — VOLUME 23 (1987), NUMBER 2

ASYMPTOTIC BEHAVIOUR OF EMPIRICAL MULTIINFORMATION

MILAN STUDENÝ

The asymptotic behaviour of an estimator of multiinformation is investigated. It is shown that it qualitatively depends on the value of certain numerical characteristic. If this characteristic is non-zero then the estimator is asymptotically normally distributed. In the opposite case the asymptotic distribution of the estimator is the distribution of a weighted sum of squares of independent normally distributed random variables.

1. INTRODUCTION

This paper deals with the asymptotic behaviour of empirical multiinformation (it can be also called sample multiinformation). Multiinformation was studied by Perez in [6] as a generalization of the concept of the mutual information of two random variables. It enables us to characterize the level of the dependence of more than two random variables.

Given a sequence of random variables $\vartheta_n(\omega)$ and a non-degenerate probability distribution D we say that $\vartheta_n(\omega)$ has asymptotic distribution D iff there exist constants $a_n > 0$, b_n such that $a_n^{-1} . \{\vartheta_n(\omega) - b_n\} \to D$ in distribution. Correctness and uniqueness of this concept is shown in [2] (Lemma 1 of § 2 of Chapter VIII). Moreover, in case $D = N(0, 1)$ we say that $\vartheta_n(\omega)$ has asymptotically distribution $N(b_n, a_n^2)$ and that a_n^2 is its asymptotic variance.

In the present paper we shall study possibility of the estimation of multiinformation on the basis of empirical data. The investigated estimator is called the empirical multiinformation; it is simply the multiinformation of the corresponding joint empirical distribution. We are dealing with the asymptotic behaviour of this estimator. The results are derived for the case of non-degenerated finite-valued random variables. Note that this assumption seems to be essential (see remark in the third section). The obtained results are a generalization of the results reported in [5]. The asymptotic behaviour of empirical information (of another type) was studied also in [8].

In the first section we define a numerical characteristic of a probability distribution, P, denoted by $R[P]$. As it will be shown later, this number is very important for the asymptotic behaviour of the empirical multiinformation. If $R[P] > 0$ (we call it the ordinary case), then the asymptotic distribution is normal and $n^{-1} . R[P]$ is its asymptotic variance (Proposition 3, n is the sample size). If $R[P] = 0$, then the asymptotic distribution is not normal. But this case takes place only if the measure P is a "truncated product measure" (Proposition 1). Then, the asymptotic distribution is the distribution of a weighted sum of squares of independent $N(0, 1)$-distributed random variable. The weights are given as the eigenvalues of a matrix which is explicitly obtained. It depends only on the probability distribution P. Note that some of the weights can be negative and that explicit formulas for densities of such distributions are given in [4].

Throughout the paper we make the following general assumption.

[A] Let $T_1, \ldots, T_m$ be nonempty finite sets, where $m \geqq 2$. We put $T = T_1 \times \ldots \times T_m$ and consider T as a measurable space with σ-algebra of all its subsets. Let $(\Omega, \mathscr{A}, Q)$ be a probability space and
$\xi^j : (\Omega, \mathscr{A}, Q) \to (T, \exp T), \quad j = 1, 2, \ldots,$
be a sequence of i.i.d. T-valued random variables defined on Ω and, moreover, their common distribution P on T is not concentrated in a single point of T.

2. NOTATION

$\mathbb{N}$ denotes the set of natural numbers, $\mathbb{R}$ the set of real numbers; δ_{ab} is the well known Kronecker symbol; card A denotes the cardinality of a set A, $f \circ g$ the composition of functions f and g.

If $t \in T$ and $i \in \{1, \ldots, m\}$, then t_i denotes the ith component of the point $t \in T$. If P is a probability measure on T and $i \in \{1, \ldots, m\}$, then P_i denotes its marginal measure on T_i.

Let S and R be finite sets; by S-vector we shall understand a real function on S (i.e. element of $\mathbb{R}^S$), by $(S \times R)$-matrix a real function on $S \times R$ (i.e. element of $\mathbb{R}^{S \times R}$). Since later we shall multiply vectors by matrices, we shall regard them as column vectors, i.e. an S-vector is an $(S \times \{1\})$-matrix. We shall denote matrices by capital bold faces, vectors by lower case bold faces and their elements or components by lower case light faces. Further, elements of matrices and components of vectors will be denoted by the corresponding letters, i.e. matrix $\mathbf{A}$ has always elements $a(t, v)$, vector $\mathbf{f}$ has always elements $f(t)$ and so on. The transpose of the matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{\mathrm{T}}$, the identity matrix by $\mathbf{I}$. If a set is indicated by a lower index $i \in \{1, \ldots, m\}$, as for example S_i, then the corresponding S_i-vector and its components have the same lower index. So the S_i-vector $\mathbf{f}_i$ has components $f_i(s)$, $s \in S_i$. An analogous principle holds for matrices and their elements.

Finally, $N(a, b)$ denotes normal distribution with expectation a and variance b, $N(0, \boldsymbol{A})$ multivariate normal distribution with zero expectation and covariance matrix $\boldsymbol{A}$.

3. BASIC CONCEPTS

As was mentioned in the introduction the assumption that ξ^j are finite-valued random variable seems to be essential. For example, let us consider two independent $[0, 1]$-valued random variables the distribution of which is given by the Lebesgue measure. Obviously, the multiinformation (mutual information) is zero, but on the other hand it can be proved (but we shall not perform it here) that the empirical multiinformation has expectation $\ln(n)$ where n is the sample size.

The definition of multiinformation for general case is contained in [6] (the multiinformation is called the dependence tightness there). But we shall use its version for finite measurable spaces only.

Definition 1. Let P be a probability measure on T. The *multiinformation* (of P) is the following value:

$$I[P] = \sum_{\substack{t \in T \\ P(t) > 0}} P(t) \,.\, \ln\left(P(t) \,.\, P_1^{-1}(t_1) \,.\, \ldots \,.\, P_m^{-1}(t_m)\right).$$

Further, we introduce:

$$R[P] = \sum_{\substack{t \in T \\ P(t) > 0}} P(t) \,.\, \ln^2\left(P(t) \,.\, P_1^{-1}(t_1) \,.\, \ldots \,.\, P_m^{-1}(t_m)\right) - I^2[P].$$

Proposition 1. The inequalities $I[P] \geqq 0$ and $R[P] \geqq 0$ hold for all probability measures P.

$I[P] = 0$ iff P is a product of its marginal measures.

$R[P] = 0$ iff there exists $\alpha \geqq 1$ such that for each $t \in T$ either $P(t) = 0$ or $P(t) = \alpha \,.\, P_1(t_1) \,.\, \ldots \,.\, P_m(t_m)$.

Proof. The statement about $I[P]$ can be proved by similar way as an analogous statement about mutual information (see [3]). Now we put

$$V = \{t \in T, P(t) > 0\} \quad \text{and}$$

$$b(t) = -\ln\left(P^{-1}(t) \,.\, P_1(t_1) \,.\, \ldots \,.\, P_m(t_m)\right), \qquad t \in V.$$

So, the well-known inequality

$$\left(\sum_{v \in V} b(v) \,.\, P(v)\right)^2 \leqq \sum_{v \in V} b^2(v) \,.\, P(v)$$

with equality only for b constant implies the statement about $R[P]$. $\square$

Definition 2. For every $n \in \mathbb{N}$ we define a real function on $T \times \Omega$ (see [A]):

$$p^n(t, \omega) = n^{-1} \,.\, \text{card}\,\{j \leqq n, \xi^j(\omega) = t\}, \qquad t \in T;\ \omega \in \Omega.$$

We call it the *empirical probability*. Indeed, $p^n(\cdot, \omega)$ is a probability measure on T for every $\omega \in \Omega$. So, the function $\omega \mapsto I[p^n(\cdot, \omega)]$ will be called the *empirical multiinformation* (see Definition 1). Moreover, we shall use vector notation for the empirical probability:

$$\boldsymbol{p}^n(\omega) = [p^n(t, \omega)]_{t \in T} \in \mathbb{R}^T, \quad \omega \in \Omega .$$

Analogously for $i \in \{, \dots, m\}$ and $n \in \mathbb{N}$ we define the function p_i^n on $T_i \times \Omega$ by

$$p_i^n(s, \omega) = n^{-1} . \operatorname{card} \{j \leqq n, \xi_i^j(\omega) = s\}, \quad s \in T_i; \omega \in \Omega .$$

Note that the empirical multiinformation is a random variable (it can be seen from Lemma 3).

4. AUXILIARY CONCEPTS

Definition 3. We introduce the T-vector $\boldsymbol{p} = [P(t)]_{t \in T}$ and the $(T \times T)$-matrix $\boldsymbol{A}$ (see [A]):

$$a(t, v) = P(t) . \{\delta_{tv} - P(v)\}, \quad t, v \in T.$$

Analogously we introduce the $(T_i \times T_i)$-matrix $\boldsymbol{A}_i$ by

$$a_i(r, s) = P_i(r) . \{\delta_{rs} - P_i(s)\}, \quad r, s \in T_i; \; i \in \{1, \dots, m\} .$$

It is easy to verify that

$$(1) \qquad a_i(r, s) = \sum_{\substack{t \in T \\ t_i = r}} \sum_{\substack{v \in T \\ v_i = s}} a(t, v), \quad r, s \in T_i \, ; \; i \in \{1, \dots, m\} .$$

Lemma 1. The mapping $\omega \mapsto \boldsymbol{p}^n(\omega)$ is a random T-vector on Ω with expectation $\boldsymbol{p}$ and covariance matrix $n^{-1} . \boldsymbol{A}$.

Proof. For every $j \in \mathbb{N}$ and $t \in T$ we define the function $\omega \mapsto x^j(t, \omega)$ as the composition of ξ^j and the indicator of the set $\{t\}$. [A] implies that the random vectors $\omega \mapsto \boldsymbol{x}^j(\omega) \in \mathbb{R}^T$ are i.i.d. It is easy to see that $\boldsymbol{x}^j$ has expectation $\boldsymbol{p}$ and covariance matrix $\boldsymbol{A}$ (see Definition 3). Obviously

$$(2) \qquad p^n(t, \omega) = n^{-1} . \sum_{j=1}^{n} x^j(t, \omega), \quad \omega \in \Omega \, ; \; t \in T ; \; n \in \mathbb{N} .$$

Now we can easy derive expectation and covariance matrix of $\boldsymbol{p}^n(\omega)$. □

For study of the asymptotic behaviour of the empirical multiinformation it will be useful to restrict the set Ω.

Lemma 2. There exists a measurable set $\Omega' \subset \Omega$ such that $Q(\Omega') = 1$ and for every $\omega \in \Omega'$, $n \in \mathbb{N}$ and $t \in T$ it holds:

$$P(t) = 0 \quad \text{implies} \quad p^n(t, \omega) = 0 .$$

Proof. We put

$$\Omega' = \bigcap_{\substack{t \in T \\ P(t)=0}} \bigcap_n {}^t\Omega^n ,$$

where ${}^t\Omega^n = \{\omega \in \Omega, p^n(t, \omega) = 0\}$.

According to Lemma 1 all sets ${}^t\Omega^n$ are measurable. Further

$$Q(\Omega \setminus {}^t\Omega^n) = Q\{\omega, \exists j \leqq n\ \xi^j(\omega) = t\} \leqq \sum_{j=1}^{n} Q\{\omega, \xi^j(\omega) = t\} = 0$$

since P is the distribution of ξ^j and $P(t) = 0$. Hence all sets ${}^t\Omega^n$ have $Q({}^t\Omega^n) = 1$. $\square$

Definition 4. The symbol Ω' will denote the set from Lemma 2. Further, we denote

$$V = \{t \in T, P(t) > 0\} , \quad V_i = \{s \in T_i, P_i(s) > 0\} , \quad i \in \{1, \dots, m\} .$$

According to Lemma 2 it is possible to interpret $\boldsymbol{p}^n(\omega)$ as a random V-vector defined on Ω'. Further, the lemma says:

$$P_i(s) = 0 \quad \text{implies} \quad p_i^n(s, \omega) = 0 , \quad n \in \mathbb{N} ; \ \omega \in \Omega' ; \ s \in T_i ; \ i \in \{1, \dots, m\} .$$

So, $\boldsymbol{p}_i^n(\omega)$ is a random V_i-vector defined on Ω'. In addition, $\boldsymbol{p}$, $\boldsymbol{A}$ and $\boldsymbol{A}_i$ from Definition 3 are V-vector, $(V \times V)$-matrix and $(V_i \times V_i)$-matrices respectively. Since we are interested in the asymptotic behaviour of a random variable, the restriction to Ω' is not essential.

Definition 5. *Normalized empirical probability* is a real function on $V \times \Omega'$:

$$f^n(t, \omega) = \sqrt{n} \, . \, \{p^n(t, \omega) - P(t)\} , \qquad t \in V ; \ \omega \in \Omega' ; \ n \in \mathbb{N} .$$

Analogously we define:

$$f_i^n(s, \omega) = \sqrt{n} \, . \, \{p_i^n(s, \omega) - P_i(s)\} , \quad s \in V_i ; \ \omega \in \Omega' ; \ n \in \mathbb{N} ; \ i \in \{1, \dots, m\} .$$

The corresponding random V-vector (resp. V_i-vector) is $\boldsymbol{f}^n(\omega)$, (resp. $\boldsymbol{f}_i^n(\omega)$).

The following proposition summarizes properties of $\boldsymbol{f}^n(\omega)$ which we shall use later.

Proposition 2. The normalized empirical probability $\boldsymbol{f}^n(\omega)$ is a random V-vector on Ω' with zero expectation and covariance matrix $\boldsymbol{A}$ (see Definitions 3 and 4); $\boldsymbol{f}_i^n(\omega)$ is a random V_i-vector on Ω'. Further, the following formulas hold:

$$\sum_{t \in V} f^n(t, \omega) = 0 , \qquad \omega \in \Omega' ; \ n \in \mathbb{N} \tag{3}$$

$$f_i^n(s, \omega) = \sum_{\substack{t \in V \\ t_i = s}} f^n(t, \omega) , \qquad \omega \in \Omega' ; \ n \in \mathbb{N} ; \ s \in V_i ; \ i \in \{1, \dots, m\} . \tag{4}$$

Finally, $\boldsymbol{f}^n(\omega) \to N(0, \boldsymbol{A})$ in distribution and $\boldsymbol{f}_i^n(\omega) \to N(0, \boldsymbol{A}_i)$ in distribution.

Proof. The first part is a consequence of Lemma 1. Lemma 2 implies $\sum_{t \in V} p^n(t, \omega) =$

$= \sum_{t \in T} p^n(t, \omega) = 1$ for every $\omega \in \Omega'$ and $n \in \mathbb{N}$. Hence (3) is evident. Since $p_i^n(\cdot, \omega)$ is the marginal measure of $p^n(\cdot, \omega)$ on T_i, we conclude, using Lemma 2 that

$$p_i^n(s, \omega) = \sum_{\substack{t \in V \\ t_i = s}} p^n(t, \omega) \quad \text{for every} \quad \omega \in \Omega' , \ s \in V_i , \ n \in \mathbb{N} .$$

Analogous expression holds for $P_i(s)$, hence (4).

The statement concerning $\mathbf{f}^n(\omega)$ is a consequence of the multivariate central limit theorem (see [2] or [7]). Indeed, (2) says $\mathbf{p}^n(\omega) = n^{-1} . \sum_{j=1}^{n} \mathbf{x}^j(\omega)$, where $\mathbf{x}^j$ are i.i.d. By Lemma 2 it is possible to understand $\mathbf{x}^j$ as V-vectors on Ω'.
Now we introduce the $(V_i \times V)$-matrix $\mathbf{K}_i$ by

$$k_i(s, v) = \delta_{sv_i} , \quad s \in V_i \, ; \ v \in V .$$

According to (4) $\mathbf{f}_i^n(\omega) = \mathbf{K}_i . \mathbf{f}^n(\omega)$ and hence we derive (see [1]) that $\mathbf{f}_i^n(\omega) \to$ $\to N(0, \mathbf{K}_i . \mathbf{A} . \mathbf{K}_i^{\mathrm{T}})$ in distribution. However (1) implies $\mathbf{K}_i . \mathbf{A} . \mathbf{K}_i^{\mathrm{T}} = \mathbf{A}_i$. □

5. ORDINARY CASE

Proposition 3. Let assumption [A] hold and $R[P] > 0$ (see Definition 1). Then the empirical multiinformation (see Definition 2) has asymptotically normal distribution $N(I[P], n^{-1} . R[P])$, i.e.

$$\sqrt{n} . R[P]^{-1/2} . \{I[p^n(\cdot, \omega)] - I[P]\} \to N(0, 1) \quad \text{in distribution} .$$

Proof. Let us introduce $\mathcal{O} = \langle 0, \infty)^V$ and a real function $\mathcal{I}$ defined on $\mathcal{O}$:

$$\mathcal{I}(x) = \sum_{t \in V} x_t . \ln \left(x_t . \left(\prod_{i=1}^{m} r_{it}(x)\right)^{-1}\right) , \quad x \in \mathcal{O} ,$$

where

$$r_{it}(x) = \sum_{\substack{v \in V \\ v_i = t_i}} x_v , \quad i \in \{1, \ldots, m\} \, ; \ t \in V .$$

Note that $0 . \ln(\cdot)$ is defined by 0. We see that $I[p^n(\omega)]$ is the composition of $\mathbf{p}^n(\omega)$ and $\mathcal{I}$ (since $\omega \in \Omega'$, see Definition 4). We compute that

$$\frac{\partial}{\partial x_v} \ln r_{it}(y) = \delta_{t_i v_i} . r_{iv}^{-1}(y) , \quad v, t \in V ; \ i \in \{1, \ldots, m\} \, ; \ y \in (0, \infty)^V$$

and further

$$\frac{\partial}{\partial x_v} \mathcal{I}(y) = \ln y_v - \sum_{i=1}^{m} \ln r_{iv}(y) + (1 - m) , \quad v \in V ; \ y \in (0, \infty)^V .$$

So, $\mathbf{p}$ (see Definition 3) is an inner point of $\mathcal{O}$ and $\mathcal{I}$ has the Frechet differential at $\mathbf{p}$.

Finally, we introduce the V-vector $\boldsymbol{g}$ by:

$$g(t) = \frac{\partial}{\partial x_t} \mathscr{I}(\boldsymbol{p}) = \ln \left(P(t) . P_1^{-1}(t_1) . \dots . P_m^{-1}(t_m)\right) + (1 - m), \quad t \in V .$$

Since $\boldsymbol{g}^{\mathrm{T}} . \boldsymbol{A} . \boldsymbol{g} = R[P] > 0$ (as it can be seen from Definitions 1 and 3) and $\sqrt{n} . \{\boldsymbol{p}^n(\omega) - \boldsymbol{p}\} \to N(0, \boldsymbol{A})$ in distribution (see Proposition 2) it is possible to use the known theorem (see [7], Theorem II in 6.a.2). So $\sqrt{n} . \{\mathscr{I} \circ \boldsymbol{p}^n(\omega) - \mathscr{I} \circ \boldsymbol{p}\} \to$ $\to N(0, R[P])$ in distribution. □

6. CASE OF "TRUNCATED PRODUCT MEASURE"

In this case we can describe the asymptotic behaviour of the empirical multiinformation by means of eigenvalues of the matrix defined below.

Definition 6. We introduce the $(V \times V)$-matrix $\boldsymbol{E}$ by

$$e(t, v) = \tfrac{1}{2} . \delta_{tv} - \tfrac{1}{2} P(v) . \sum_{i=1}^{m} \delta_{t_i v_i} . P_i^{-1}(t_i), \qquad t, v \in V .$$

It is easy to verify that

$$\sum_{v \in V} e(t, v) = \tfrac{1}{2}(1 - m) \quad \text{for every} \quad t \in V . \tag{5}$$

In particular,

$$\sum_{v \in V} \{e(t, v) - e(z, v)\} . y = 0 \quad \text{for every } t,\ z \in V \ \text{ and } \ y \in \mathbb{R} . \tag{6}$$

Proposition 4. Under assumption [A] let us suppose $R[P] = 0$. Then $n . \{I[p^n(\cdot, \omega)] - I[P]\}$ tends in distribution to a random variable

$$\varrho(\omega) = \sum_{j=1}^{\operatorname{card}(V)-1} \lambda_j . t_j^2(\omega) ,$$

where $\{t_j(\omega),\ j = 1, \dots, \operatorname{card}(V) - 1\}$ are independent $N(0, 1)$-distributed random variables and λ_j are the eigenvalues of $\boldsymbol{E}$ (see Definition 6) with the exception that the multiplicity of $\lambda_0 = \tfrac{1}{2}(1 - m)$ must be reduced by 1. Moreover, all eigenvalues of $\boldsymbol{E}$ are real.

The proof is based on the following three lemmas.

Lemma 3. For every $\omega \in \Omega'$ and $n \in \mathbb{N}$

$$n . \{I[p^n(\cdot, \omega)] - I[P]\} = V^n(\omega) + U^n(\omega) - \sum_{i=1}^{m} U_i^n(\omega) ,$$

where

$$V^n(\omega) = n . \sum_{t \in V} \{p^n(t, \omega) - P(t)\} . \ln \left(P(t) . P_1^{-1}(t_1) . \dots . P_m^{-1}(t_m)\right)$$

and (cf. Definition 5)

$$U^n(\omega) = \tfrac{1}{2}\sum_{t\in V} P^{-1}(t) \,.\, \{f^n(t, \omega)\}^2 + n^{-1/2} \sum_{t\in V} \Phi^n(t, \omega) \,.\, \{f^n(t, \omega)\}^3$$

$$U_i^n(\omega) = \tfrac{1}{2}\sum_{s\in V_i} P_i^{-1}(s) \,.\, \{f_i^n(s, \omega)\}^2 + n^{-1/2} \sum_{s\in V_i} \Phi_i^n(s, \omega) \,.\, \{f_i^n(s, \omega)\}^3$$

where $\Phi^n(t, \omega)$ and $\Phi_i^n(t, \omega)$ are measurable functions bounded by constants which are independent of $\omega \in \Omega'$ and $n \in \mathbb{N}$.

Proof. Since ω and n are fixed we shall write $p(t)$ instead of $p^n(t, \omega)$ and $p_i(s)$ instead of $p_i^n(s, \omega)$. Further we denote $h(u) = u \,.\, \ln(u)$ for $u > 0$; $h(0) = 0$.

I) We introduce $L = \{t \in V, p(t) > 0\}$ and put

$$k(t) = \ln\left(p(t) \,.\, p_1^{-1}(t_1) \,.\, \ldots \,.\, p_m^{-1}(t_m)\right), \qquad t \in L$$

$$K(t) = \ln\left(P(t) \,.\, P_1^{-1}(t_1) \,.\, \ldots \,.\, P_m^{-1}(t_m)\right), \qquad t \in V.$$

Since $L \subset V$ (see Lemma 2) it holds

$$I[p] - I[P] = \sum_{t\in L} p(t) \,.\, k(t) - \sum_{t\in V} P(t) \,.\, K(t) =$$

$$= \sum_{t\in V} \{p(t) - P(t)\} \,.\, K(t) + \sum_{t\in L} p(t) \,.\, \{k(t) - K(t)\}\,.$$

The first term is $n^{-1} \,.\, V^n(\omega)$, the second one can be written as

$$\sum_{t\in L} p(t) \,.\, \ln\left(P^{-1}(t) \,.\, p(t)\right) - \sum_{i=1}^{m} \sum_{t\in L} p(t) \,.\, \ln\left(P_i^{-1}(t_i) \,.\, p_i(t_i)\right).$$

Since $\omega \in \Omega'$ it holds $p(t) = 0$ for every $t \in V \setminus L$. Thus

$$\sum_{t\in L} p(t) \,.\, \ln\left(P^{-1}(t) \,.\, p(t)\right) = \sum_{t\in V} P(t) \,.\, h\left(P^{-1}(t) \,.\, p(t)\right) \tag{7}$$

$$\sum_{t\in L} p(t) \,.\, \ln\left(P_i^{-1}(t_i) \,.\, p_i(t_i)\right) = \sum_{s\in V_i} P_i(s) \,.\, h\left(P_i^{-1}(s) \,.\, p_i(s)\right). \tag{8}$$

To show (8) we divide L into groups with the same ith components and compute sums over these groups. We denote the expression in (7) as $n^{-1} \,.\, U^n(\omega)$, the expression in (8) as $n^{-1} \,.\, U_i^n(\omega)$.

II) Using the Taylor expansion of h at the point $u = 1$ we see, that there exists a Borel measurable function τ: $[0, \infty) \to [0, 1)$ such that

$$h(u) = (u - 1) + \tfrac{1}{2}(u - 1)^2 - \tfrac{1}{2}\,\tau(u) \,.\, (u - 1)^3, \qquad u \geqq 0\,. \tag{9}$$

Indeed, for the remainder we have (Cauchy's form)

$$\mathscr{R}_2^h(u) = \tfrac{1}{2}(u - \sigma)^2 \,.\, (u - 1) \,.\, h'''(\sigma) \quad \text{where} \quad u < \sigma < 1 \quad \text{or} \quad 1 < \sigma < u\,.$$

Putting $\theta = (u - 1)^{-1} \,.\, (\sigma - 1)$ and $\tau(u) = (1 - \theta)^2 \,.\, \{1 + \theta \,.\, (u - 1)\}^{-2}$ it is easy to see that $0 < \theta < 1$, $1 - \theta \leqq 1 + \theta \,.\, (u - 1)$ and thus $0 < \tau(u) \leqq 1$.
By substituting (9) in (7) we obtain:

$$U^n(\omega) = n \,.\, \sum_{t\in V} \{p(t) - P(t)\} + \tfrac{1}{2} n \sum_{t\in V} P^{-1}(t) \,.\, \{p(t) - P(t)\}^2 -$$

$$- \tfrac{1}{2} n \,.\, \sum_{t\in V} \tau\left(P^{-1}(t) \,.\, p(t)\right) \,.\, P^{-2}(t) \,.\, \{p(t) - P(t)\}^3\,.$$

According to (3) the first term is zero. Putting $\Phi^n(t, \omega) = -\frac{1}{2}\tau(P^{-1}(t) . p^n(t, \omega)) . P^{-2}(t)$ we obtain the first formula. Analogously it is possible to prove the second one. □

Lemma 4. Let $\mathbf{A}$ and $\mathbf{C}$ be real symmetric $(k \times k)$-matrices with $\mathbf{A}$ positive definite. Let us consider the quadratic form H on $\mathbb{R}^k$ defined by $H(\mathbf{x}) = \mathbf{x}^T . \mathbf{C} . \mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^k$. Let $\mathbf{X}(\omega)$ be a k-dimensional random vector with distribution $N(0, \mathbf{A})$. Then all eigenvalues of $\mathbf{C} . \mathbf{A}$ are real and we can describe the distribution of $H \circ \mathbf{X}(\omega)$ as the distribution of $\varrho(\omega) = \sum_{j=1}^{k} \lambda_j . t_j^2(\omega)$, where λ_j are the eigenvalues of $\mathbf{C} . \mathbf{A}$ and $\{t_j(\omega), j = 1, \ldots, k\}$ are independent $N(0, 1)$-distributed random variables.

Proof. Since $\mathbf{A}$ is positive definite there exists a real nonsingular matrix $\mathbf{Q}$ such that $\mathbf{A} = \mathbf{Q} . \mathbf{Q}^T$. It is possible to choose $\mathbf{Q}$ in such a way that $\mathbf{G} = \mathbf{Q}^T . \mathbf{C} . \mathbf{Q}$ is a real diagonal matrix. In the opposite case we find an orthogonal matrix $\mathbf{F}$ such that $\mathbf{F}^T . \mathbf{G} . \mathbf{F}$ is diagonal and use $\mathbf{Q} . \mathbf{F}$ instead of $\mathbf{Q}$. It is easy to see that $\mathbf{G}$ and $\mathbf{C} . \mathbf{A}$ are similar, hence they have the same spectrum. Since $\mathbf{G}$ has its eigenvalues on a diagonal, the eigenvalues are real. Let us put $\mathbf{Y}(\omega) = \mathbf{Q}^{-1} . \mathbf{X}(\omega)$. It is easy to verify that $\mathbf{Y}(\omega)$ is an $N(0, \mathbf{I})$-distributed random vector. Moreover

$$H \circ \mathbf{X}(\omega) = \mathbf{X}^T(\omega) . \mathbf{C} . \mathbf{X}(\omega) = \mathbf{Y}^T(\omega) . \mathbf{G} . \mathbf{Y}(\omega) = \sum_{j=1}^{k} \lambda_j . \mathbf{Y}_j^2(\omega) ,$$

where λ_j are the diagonal elements of $\mathbf{G}$, i.e. the eigenvalues of $\mathbf{C} . \mathbf{A}$. □

Lemma 5. Given $z \in V$ we put $K = V \setminus \{z\}$ (see Definition 4). Note that $K \neq \emptyset$ (P is non-degenerate – see [A]). Let us introduce the $(K \times K)$-matrix $\mathbf{D}$ by (see Definition 6):

$$d(t, v) = e(t, v) - e(z, v) , \quad t, v \in K . \tag{10}$$

Then $\mathbf{D}$ has the same spectrum as $\mathbf{E}$ with exception that the multiplicity of $\lambda_0 = \frac{1}{2}(1 - m)$ must be reduced by 1. (But (5) implies that λ_0 corresponds to constant eigenvector.)

Proof. Given $\lambda \neq \frac{1}{2}(1 - m)$ and $s \in \mathbb{N}$ we consider the following statements:
[B] λ is an eigenvalue of $\mathbf{E}$ with multiplicity at least s
[C] λ is an eigenvalue of $\mathbf{D}$ with multiplicity at least s.

I) [B] implies [C].

Indeed, let $[f_i(t)]_{t \in V}$ $i = 1, \ldots, s$ be a system of linearly independent eigenvectors of $\mathbf{E}$ corresponding to λ. By adding the constant vector $f_{s+1}(t) \equiv 1$ we preserve linear independence, because $[f_{s+1}(t)]_{t \in V}$ corresponds to λ_0 (see (5)). We put:

$$g_i(t) = f_i(t) - f_i(z), \quad t \in V;\ i = 1, \ldots, s.$$

Let us suppose that $\sum_{i=1}^{s} c_i . g_i(t) = 0$ for every $t \in K$. If we put $c_{s+1} = -\sum_{i=1}^{s} c_i . f_i(z)$, then

$$\sum_{i=1}^{s+1} c_i . f_i(t) = \sum_{i=1}^{s} c_i . f_i(t) - \sum_{i=1}^{s} c_i . f_i(z) . 1 = \sum_{i=1}^{s} c_i . g_i(t) = 0$$

for every $t \in K$. This equality also holds for $t = z$. Linear independence of $[f_i(t)]_{t\in V}$, $i = 1, \ldots, s + 1$, implies $c_1 = \ldots = c_{s+1} = 0$. So, we have proved that $[g_i(t)]_{t\in K}$, $i = 1, \ldots, s$, are linearly independent. In particular, they are nonzero. Since $[f_i(t)]_{t\in V}$, $i = 1, \ldots, s$, are eigenvectors of **E** corresponding to λ we can take the respective equations and subtract from them the equation for $t = z$:

$$\sum_{v\in V} \{e(t, v) - e(z, v)\} \cdot f_i(v) = \lambda \cdot \{f_i(t) - f_i(z)\} = \lambda \cdot g_i(t) \quad \text{for} \quad t \in K \quad \text{and} \quad i = 1, \ldots, s.$$

By subtracting the formula (6) for $y = f_i(z)$ and considering $g_i(z) = 0$ we can easily derive that $\sum_{v\in K} d(t, v) \cdot g_i(v) = \lambda \cdot g_i(t)$ for $t \in K$ and $i = 1, \ldots, s$. So, we have proved that $[g_i(t)]_{t\in K}$, $i = 1, \ldots, s$, are linearly independent eigenvectors of **D** corresponding to λ.

II) [C] implies [B].

Indeed, let $[g_i(t)]_{t\in K}$, $i = 1, \ldots, s$, be linearly independent eigenvectors of **D** corresponding to λ. We put $g_i(z) = 0$. Since $\lambda \neq \frac{1}{2}(1 - m)$ we can define:

$$f_i(z) = \{\lambda - \tfrac{1}{2}(1 - m)\}^{-1} \cdot \sum_{v\in K} e(z, v) \cdot g_i(v), \quad i = 1, \ldots, s,$$

(11)
$$f_i(t) = g_i(t) + f_i(z), \quad t \in V; \ i = 1, \ldots, s.$$

Let us suppose

(12)
$$\sum_{i=1}^{s} c_i \cdot f_i(t) = 0, \quad t \in V.$$

We substitute (11) in (12). Using (12) for $t = z$ it follows that $\sum_{i=1}^{s} c_i \cdot g_i(t) = 0$ for $t \in K$. So, $[f_i(t)]_{t\in V}$, $i = 1, \ldots, s$, are linearly independent. Now we can rewrite the definition of $f_i(z)$:

$$\sum_{v\in V} e(z, v) \cdot g_i(v) = \{\lambda - \tfrac{1}{2}(1 - m)\} \cdot f_i(z) \quad \text{for} \quad i = 1, \ldots, s.$$

Using (11) and (5) for $t = z$ we derive from it:

(13)
$$\sum_{v\in V} e(z, v) \cdot f_i(v) = \lambda \cdot f_i(z), \quad i = 1, \ldots, s.$$

Our assumptions concerning $[g_i(t)]_{t\in V}$ imply that

$$\sum_{v\in V} \{e(t, v) - e(z, v)\} \cdot g_i(v) = \lambda \cdot g_i(t) \quad \text{for} \quad t \in V \quad \text{and} \quad i = 1, \ldots, s.$$

We add up (6) to it, where $y = f_i(z)$ and using (11) get:

$$\sum_{v\in V} \{e(t, v) - e(z, v)\} \cdot f_i(v) = \lambda \cdot g_i(t) \quad \text{for} \quad t \in V \quad \text{and} \quad i = 1, \ldots, s.$$

Finally we add up (13). So, we have proved that $[f_i(t)]_{t\in V}$ are eigenvalues of **E** corresponding to λ.

III) So, I) and II) together yield the proof. □

Proof of Proposition 4. Let us introduce the $(V \times V)$-matrix $\mathbf{B}$ by

$$b(t, v) = \tfrac{1}{2} \,.\, \delta_{tv} \,.\, P^{-1}(t) - \tfrac{1}{2} \,.\, \sum_{i=1}^{m} \delta_{t_i v_i} \,.\, P_i^{-1}(t_i), \qquad t, v \in V. \tag{14}$$

Let us further introduce the sequence of random variables (see Definition 5):

$$k^n(\omega) = \sum_{t \in V} \sum_{v \in V} b(t, v) \,.\, f^n(t, \omega) \,.\, f^n(v, \omega), \qquad \omega \in \Omega' \,;\; n \in \mathbb{N} \,. \tag{15}$$

(I) $n \,.\, \{I[p^n(\cdot, \omega)] - I[P]\} - k^n(\omega) \to 0$ in probability. According to the Proposition 1 there exists a constant $\alpha \geqq 1$ such that for every $t \in V$ $\alpha = P(t) \,.\, P_1^{-1}(t_1) \,. \dots$ $\dots \,.\, P_m^{-1}(t_m)$. Hence in Lemma 3 $V^n(\omega) = 0$ for $\omega \in \Omega'$ and $n \in \mathbb{N}$ (we have used (3)). According to Proposition 2 $f^n(t, \omega)$ tends in distribution (for fixed $t \in V$). So (see Lemma 3) $\Phi^n(t, \omega) \,.\, \{f^n(t, \cdot)\}^3$ is a stochastically bounded sequence (see [2] or [7]). It implies (see Lemma 3)

$$U^n(\omega) - \tfrac{1}{2} \sum_{t \in V} P^{-1}(t) \,.\, \{f^n(t, \omega)\}^2 \to 0 \quad \text{in probability} \,.$$

Similarly we can derive that

$$U_i^n(\omega) - \tfrac{1}{2} \sum_{s \in V_i} P_i^{-1}(s) \,.\, \{f_i^n(s, \omega)\}^2 \to 0 \quad \text{in probability} \,.$$

But using (4) we easily verify that

$$\sum_{s \in V_i} P_i^{-1}(s) \,.\, \{f_i^n(s, \omega)\}^2 = \sum_{t \in V} \sum_{v \in V} \delta_{t_i v_i} \,.\, P_i^{-1}(t_i) \,.\, f^n(t, \omega) \,.\, f^n(v, \omega) \,.$$

Combining these facts we get the desired statement (I).

(II) Thus, we are now interested in the asymptotic behaviour of $k^n(\omega)$. We choose a fixed $z \in V$, put $K = V \setminus \{z\}$ and introduce the $(K \times K)$-matrix $\mathbf{C}$ by

$$c(t, v) = b(t, v) - b(t, z) - b(z, v) + b(z, z) \,, \qquad t, v \in K \,. \tag{16}$$

We express $f^n(z, \omega)$ according to (3), substitute it in (15) and get:

$$k^n(\omega) = \sum_{t \in K} \sum_{v \in K} c(t, v) \,.\, f^n(t, \omega) \,.\, f^n(v, \omega) \,.$$

So, $k^n(\omega) = H \circ \mathbf{f}^n(\omega)$, where $\mathbf{f}^n(\omega)$ is taken as a K-vector and H is a quadratic form on $\mathbb{R}^K$ defined by $H(\mathbf{x}) = \mathbf{x}^T \,.\, \mathbf{C} \,.\, \mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^K$. According to Proposition 2 $\mathbf{f}^n(\omega) \to$ $\to N(0, \mathbf{A})$ in distribution ($\mathbf{A}$ is a $(K \times K)$-matrix here), hence $H \circ \mathbf{f}^n(\omega)$ converges in distribution to $N(0, \mathbf{A})$ transformed by H (see [1] Theorem 5.1). But this distribution is described in Lemma 4. Since the $(K \times K)$-matrix $\mathbf{A}$ has dominant diagonal, it is nonsingular and hence positive definite.

(III) It remains to characterize the spectrum of $\mathbf{D} = \mathbf{C} \,.\, \mathbf{A}$. For t, $r \in V$ fixed we shall express its element $d(t, r)$. We substitute (16) in the definition of $d(t, r)$ and use the following formula: $\sum_{v \in V} a(v, r) = 0$ for $r \in V$. So we get:

$$d(t, r) = \sum_{v \in V} b(t, v) \,.\, a(v, r) - \sum_{v \in V} b(z, v) \,.\, a(v, r) \,, \qquad t, r \in K \,. \tag{17}$$

Now we compute $2 . \sum_{v\in V} b(t, v) . a(v, r)$ for $t \in V$ and $r \in K$. After substitution (see (14) and Definition 3) we get:

$$\sum_{v\in V} \delta_{tv} . P^{-1}(t) . \delta_{vr} . P(r) - \sum_{v\in V} \sum_{i=1}^{m} \delta_{t_i v_i} . P_i^{-1}(t_i) . \delta_{vr} . P(r) -$$

$$- \sum_{v\in V} \delta_{tv} . P^{-1}(t) . P(v) . P(r) + \sum_{v\in V} \sum_{i=1}^{m} \delta_{t_i v_i} . P_i^{-1}(t_i) . P(v) . P(r) .$$

Obviously the first term is δ_{tr}, the second one is $-P(r) . \sum_{i=1}^{m} \delta_{t_i r_i} . P_i^{-1}(t_i)$, the third one is $-P(r)$. Further

$$\sum_{v\in V} \sum_{i=1}^{m} \delta_{t_i v_i} . P_i^{-1}(t_i) . P(v) = \sum_{i=1}^{m} P_i^{-1}(t_i) . \sum_{v\in V} \delta_{t_i v_i} . P(v) =$$

$$= \sum_{i=1}^{m} P_i^{-1}(t_i) . P_i(t_i) = m .$$

It means that the fourth term is $m . P(r)$. Together we obtain (see Definition 6):

$$\sum_{v\in V} b(t, v) . a(v, r) = e(t, r) + \tfrac{1}{2}(m - 1) . P(r) \qquad \text{for } t \in V \text{ and } r \in K .$$

Thus (17) implies (10). By Lemma 5 it completes the proof. □

(Received February 18, 1986.)

REFERENCES

[1] P. Billingsley: Convergence of Probability Measures. Wiley, New York 1968.

[2] W. Feller: An Introduction to Probability Theory and Its Applications. Wiley, New York 1966.

[3] R. G. Galager: Information Tneory and Relaible Communication. Wiley, New York 1968.

[4] J. Gurland: Distribution of definite and of indefinite quadratic forms. Ann. Math. Statist. *26* (1955), 123—127.

[5] Z. K. Lomicki and S. Z. Zaremba: The asymptotic distributions of estimators of the amount of transmitted information. Inform. and Control *2* (1959), 3, 260—284.

[6] A. Perez: ε-admissible simplifications of the dependence structure of a set of random variables. Kybernetika *13* (1977), 6, 439—449.

[7] A. C. Rao: Linear Statistical Inference and Its Applications. Wiley, New York 1973.

[8] J. Zvárová: The asymptotic distribution of sample information measure of dependence (in Czech). Kybernetika *5* (1969), 1, 50—59.

RNDr. Milan Studený, 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.