Časopis pro pěstování matematiky, roč. 99 (1974), Praha

ON MEASURES OF STATISTICAL DEPENDENCE*)

JANA ZVÁROVÁ, Praha

(Received April 27, 1971)

1. INTRODUCTION

One of the most important problems of mathematical statistics is to express the strength of statistical dependence between two random variables. There have been given different sets of requirements that have to be satisfied by an adequate measure of statistical dependence. To all of these sets some requirements are common. It seems to be natural to choose a range of values of measures of statistical dependence to be in the closed interval [0, 1], to reach the lower bound 0 if and only if random variables are independent and the upper bound 1 in the case of their highest dependence. The highest dependence of random variables has been introduced in different ways by authors. For example, we can remind W. HÖFFDING's [4] and A. RÉNYI's [15] approaches to this problem. Important properties for adequate measures of statistical dependence have also been pointed out by A. PEREZ [10]. However, in practical situations, for a proper selection of an adequate measure of statistical dependence an important role is played by both the specific features of the given task and the behaviour of sample estimators of measures of statistical dependence.

In Sec. 2 of this paper a set of requirements 1 – 4 on measures of statistical dependence is given. There also the problem of the highest dependence of random variables is discussed. In Sec. 3 a class of measures of statistical dependence that satisfy the requirements 1 – 4 is found and in Sec. 4 upper bounds of such measures of statistical dependence under particular restrictions on random variables are derived. In Sec. 5 sample properties of a special class of measures of statistical dependence are examined.

---

*) A slightly different version of this paper was presented as a part of the lecture at Sixth Prague Conference on Information Theory, September 19—25, 1971.

2. PROBLEM FORMULATION

Let ξ and η be two abstract valued random variables. It is well known that to the random variables ξ and η there correspond sample probability spaces $(X, \mathscr{X}, P_\xi)$ and $(Y, \mathscr{Y}, P_\eta)$ respectively, i.e., $\xi \to (X, \mathscr{X}, P_\xi)$ and $\eta \to (Y, \mathscr{Y}, P_\eta)$. Let $(X \times Y, \mathscr{X} \times \mathscr{Y})$ be the Cartesian product of $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ and let us assume that to the abstract valued random variable (ξ, η) there corresponds a sample probability space $(X \times Y, \mathscr{X} \times \mathscr{Y}, P_{\xi\eta})$, i.e., $(\xi, \eta) \to (X \times Y, \mathscr{X} \times \mathscr{Y}, P_{\xi\eta})$. Moreover, let P_ξ and P_η be marginal probability measure of $P_{\xi\eta}$ on $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ respectively. If we consider the probability measure $P_\xi \times P_\eta$ and a measure λ on $(X \times Y, \mathscr{X} \times \mathscr{Y})$, where λ is an arbitrary dominating measure of $P_{\xi\eta}$ and $P_\xi \times P_\eta$, we shall denote by $p_{\xi\eta}(x, y) = dP_{\xi\eta}/d\lambda$ and $p_\xi(x)\, p_\eta(y) = d(P_\xi \times P_\eta)/d\lambda$ the corresponding Radon-Nikodym densities.

Further we shall denote by $e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta)$ the minimum probability of error (Bayes risk) for testing the hypothesis H_0: $P = P_\xi \times P_\eta$ against H_1: $P = P_{\xi\eta}$ in the case that the a priori probabilities of H_0 and H_1 are equal to $\frac{1}{2}$, i.e.

$$(1) \qquad e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta) = \frac{1}{2} \int_{X \times Y} \min \left[p_{\xi\eta}(x, y),\ p_\xi(x)\, p_\eta(y) \right] d\lambda\,.$$

Now we shall give some general requirements on adequate measures of statistical dependence stimulated by W. Höffding's [4] and A. Perez's [10] works. If we denote by $\delta(\xi, \eta)$ a measure of statistical dependence of random variables ξ and η, these requirements do not determine $\delta(\xi, \eta)$ uniquely, reading as follows:

1. $0 \leqq \delta(\xi, \eta) \leqq 1.$

2. a) $\delta(\xi, \eta) = 0$ if and only if ξ and η are independent;

   b) $\lim\limits_{e \uparrow 1/2} \sup\limits_{\mathscr{D}_1(e)} \delta(\xi, \eta) = 0,$

   where $\mathscr{D}_1(e) = \{(\xi, \eta) : \frac{1}{2} > e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta) \geqq e\}$;

   c) $\lim\limits_{\delta \downarrow 0} \sup\limits_{\mathscr{E}_1(\delta)} e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta) = \frac{1}{2}$,

   where $\mathscr{E}_1(\delta) = \{(\xi, \eta) : 0 < \delta(\xi, \eta) \leqq \delta\}$.

3. a) $\delta(\xi, \eta) = 1$ if and only if ξ and η are singular;

   b) $\lim\limits_{e \downarrow 0} \inf\limits_{\mathscr{D}_2(e)} \delta(\xi, \eta) = 1,$

   where $\mathscr{D}_2(e) = \{(\xi, \eta) : 0 < e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta) \leqq e\}$;

   c) $\lim\limits_{\delta \uparrow 1} \inf\limits_{\mathscr{E}_2(\delta)} e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta) = 0,$

   where $\mathscr{E}_2(\delta) = \{(\xi, \eta) : 1 > \delta(\xi, \eta) \geqq \delta\}$.

4. If $(\xi', \eta') \to (X \times Y, \mathscr{X}' \times \mathscr{Y}', P_{\xi'\eta'})$, where $\mathscr{X}' \times \mathscr{Y}' \subset \mathscr{X} \times \mathscr{Y}$ is a sub-σ algebra and $P_{\xi'\eta'}$ is the restriction of $P_{\xi\eta}$ on $\mathscr{X}' \times \mathscr{Y}'$, then

a) $\delta(\xi', \eta') \leqq \delta(\xi, \eta)$;

b) $\delta(\xi', \eta') = \delta(\xi, \eta)$ if and only if $\mathscr{X}' \times \mathscr{Y}'$ is sufficient with respect to $P_{\xi\eta}$ and $P_\xi \times P_\eta$.

Remark 1. Independence and singularity of random variables ξ and η is defined by the equality and singularity of probability measures $P_{\xi\eta}$ and $P_\xi \times P_\eta$, i.e., $P_{\xi\eta} = P_\xi \times P_\eta$ and $P_{\xi\eta} \perp P_\xi \times P_\eta$ respectively.

Further we shall discuss the problem of the highest dependence of random variables. The highest dependence given by the singularity of random variables ξ and η in the requirement 3.a) corresponds to the c-dependence introduced in [4]. The c-dependence likewise the strict dependence introduced in [15] have been defined for real valued random variables. We extend both these definitions to abstract valued random variables in the following way.

Definition 1. Random variables ξ and η are c-dependent if there exists an $A \in \mathscr{X} \times \mathscr{Y}$ such that

$$\int_A p_{\xi\eta}(x, y)\, d\lambda - \int_A p_\xi(x)\, p_\eta(y)\, d\lambda = 1\,.$$

Definition 2. Random variables ξ and η are strictly dependent if either $\xi = g(\eta)$ or $\eta = h(\xi)$, where $g(y)$ is a measurable mapping of $(Y, \mathscr{Y})$ into $(X, \mathscr{X})$ and $h(x)$ is a measurable mapping of $(X, \mathscr{X})$ into $(Y, \mathscr{Y})$.

In the following two lemmas we examine the relationship of the strict dependence and c-dependence.

Definition 3. A set (a class of P-equivalent sets) C in a probability space $(\Omega, \mathscr{A}, P)$ is an atom, if $P(C) > 0$ and for $C \supset B \in \mathscr{A}$ either $P(B) = 0$ or $P(C - B) = 0$ [7].

Lemma 1. *Let C be an atom in $(X \times Y, \mathscr{X} \times \mathscr{Y}, P_{\xi\eta})$. Then there exists an atom E in $(X, \mathscr{X}, P_\xi)$ and an atom F in $(Y, \mathscr{Y}, P_\eta)$ such that $C = E \times F\ [P_{\xi\eta}]$.*

If $P_{\xi\eta}(D) = 1$, $D \in \mathscr{X} \times \mathscr{Y}$ and there exists an atom C in $(X \times Y, \mathscr{X} \times \mathscr{Y}, P_{\xi\eta})$, then $P_\xi \times P_\eta(D) > 0$.

Proof. Let us consider the sequence $\varepsilon_n = 1/n$, $n = 1, 2, \ldots$, and let n_0 be such a positive integer that $\varepsilon_{n_0} < P_{\xi\eta}(C)$. For any $n \geqq n_0$, there exists a set $A_n = \bigcup_{i=1}^{k_n} E_{in} \times F_{in}$ such that for a fixed n the sets $E_{in} \times F_{in}$ $(i = 1, 2, \ldots, k_n)$ are disjoint and $P_{\xi\eta}(C \,\Delta A_n) < \varepsilon_n$. Since for any $n \geqq n_0$ $P_{\xi\eta}(C - A_n) = 0$, therefore

$$P_{\xi\eta}(C \cap A_n) = \sum_{i=1}^{k_n} P_{\xi\eta}(C \cap (E_{in} \times F_{in})) = P_{\xi\eta}(C)\,.$$

Moreover, for any $n \geqq n_0$ there exists a unique index i_n such that

$$P_{\xi\eta}(C) \leqq P_{\xi\eta}(E_{i_n n} \times F_{i_n n}) < P_{\xi\eta}(C) + \varepsilon_n .$$

For $n \geqq n_0$ we denote $E_n = E_{i_n n}$ and $F_n = F_{i_n n}$. For $n < n_0$ we define $E_n = X$, $F_n = Y$. Let $E_0 = \bigcap_{n=1}^{\infty} E_n, F_0 = \bigcap_{n=1}^{\infty} F_n$. Clearly $E_0 \times F_0 = \bigcap_{n=1}^{\infty} (E_n \times F_n)$ and $P_{\xi\eta}((E_0 \times F_0) - C) = 0$. Simultaneously $P_{\xi\eta}(C - (E_0 \times F_0)) \leqq \sum_{n=1}^{\infty} P_{\xi\eta}(C - (E_n \times F_n)) = 0$.

Now we shall prove that the set E_0 contains an atom E in $(X, \mathscr{X}, P_\xi)$ such that $P_{\xi\eta}(E \times F_0) = P_{\xi\eta}(C)$. Let us establish a decomposition of the set E_0 into at most countable union of disjoint atoms E'_i and their non atomic complement E' in E_0 (see [6], p. 110), i.e. $E_0 = \bigcup_{i=1}^{\infty} E'_i \cup E'$. If $P_{\xi\eta}(E'_i \times F_0) = 0$ for all $i = 1, 2, \ldots$, then $P_{\xi\eta}(E' \times F_0) = P_{\xi\eta}(C)$ and $E' \times F_0$ is an atom in $(X \times Y, \mathscr{X} \times \mathscr{I}, P_{\xi\eta})$. Let us divide E' into m_0 disjoint sets E''_j such that $P_\xi(E''_j) < P_{\xi\eta}(C)$, $j = 1, 2, \ldots, m_0$. Then $P_{\xi\eta}(E''_j \times F_0) = 0$ for $j = 1, 2, \ldots, m_0$ which is a contradiction. Therefore indeed there exists such an atom E in $(X, \mathscr{X}, P_\xi)$.

Similarly we find an atom F in $(Y, \mathscr{I}, P_\eta)$ such that $P_{\xi\eta}(E \times F) = P_{\xi\eta}(C)$. This proves the first part of the lemma.

If $P_{\xi\eta}(D) = 1$ and C is an atom in $(X \times Y, \mathscr{X} \times \mathscr{I}, P_{\xi\eta})$, then it follows from the first part of the lemma that there exist atoms E, F such that $C = E \times F\ [P_{\xi\eta}]$.

Let us denote $C^* = C \cap D \cap (E \times F)$. Since C^* is an atom in $(X \times Y, \mathscr{X} \times \mathscr{I}, P_{\xi\eta})$, therefore $C^* = C\ [P_{\xi\eta}]$. Now we shall show that $P_\xi \times P_\eta(C^*) = P_\xi(E)\, P_\eta(F)$.

Let us assume that $P_\xi \times P_\eta(C^*) < P_\xi(E)\, P_\eta(F)$. Then there exists a countable union of disjoint rectangles $E_i \times F_i$ $(i = 1, 2, ..)$ such that $C^* \subset \bigcup_{i=1}^{\infty} E_i \times F_i \subset E \times F$ and simultaneously

$$P_\xi \times P_\eta(C^*) \leqq \sum_{i=1}^{\infty} P_\xi(E_i)\, P_\eta(F_i) < P_\xi(E)\, P_\eta(F) . \tag{2}$$

Moreover, $P_{\xi\eta}(C^*) = \sum_{i=1}^{\infty} P_{\xi\eta}(E_i \times F_i) = P_{\xi\eta}(E \times F)$. In view of the fact that $E \times F$ is an atom in $(X \times Y, \mathscr{X} \times \mathscr{I}, P_{\xi\eta})$, there exists a unique index i_0 such that $P_{\xi\eta}(E_{i_0} \times F_{i_0}) = P_{\xi\eta}(C^*)$. Therefore $P_\xi(E_{i_0}) \geqq P_{\xi\eta}(C^*) > 0$, $P_\eta(F_{i_0}) \geqq P_{\xi\eta}(C^*) > 0$ and since E and F are atoms, it follows $P_\xi(E_{i_0}) = P_\xi(E)$, $P_\eta(F_{i_0}) = P_\eta(F)$. However, this contradicts the second part of inequality (2). Consequently $P_\xi \times P_\eta(C^*) = P_\xi(E)\, P_\eta(F)$ and $P_\xi \times P_\eta(D) > 0$.

Corollary 1. *If C is an atom in $(X \times Y, \mathscr{X} \times \mathscr{I}, P_{\xi\eta})$ and ξ and η are strictly dependent then ξ and η are not c-dependent.*

Lemma 2. *The strict dependence $\eta = h(\xi)$ or $\xi = g(\eta)$ implies the c-dependence if and only if there are no atoms in $(Y, \mathscr{I}, P_\eta)$ and $(X, \mathscr{X}, P_\xi)$, respectively.*

Proof. Let $\eta = h(\xi)$ and let ξ and η be c-dependent. Then for $D = \{(x, y) : : y = h(x)\}$ it is $P_{\xi\eta}(D) = 1$ and $P_\xi \times P_\eta(D) = \int_X P_\eta(h(x))\, dP_\xi = 0$. Therefore there are no atoms in $(Y, \mathscr{I}, P_\eta)$.

If $\eta = h(\xi)$ and there are no atoms in $(Y, \mathscr{I}, P_\eta)$, then $P_{\xi\eta}(D) = P_{\xi\eta}\{(x, y) : : y = h(x)\} = 1$ and $P_\xi \times P_\eta(D) = \int_X P_\eta(h(x))\, dP_\xi = 0$. Then we can see that ξ and η are c-dependent.

The proof for the strict dependence $\xi = g(\eta)$ is similar.

Corollary 2. *If there are no atoms in $(X, \mathscr{X}, P_\xi)$ and $(Y, \mathscr{I}, P_\eta)$ then ξ and η are strictly dependent if and only if they are c-dependent.*

Remark 2. We can notice that ξ and η are c-dependent if and only if there exists a function $k(x, y)$ such that $k(x, y) = 0[P_{\xi\eta}]$ and $k(x, y) \neq 0[P_\xi \times P_\eta]$. It seems to us that there are no reasons to restrict ourselves to strict dependences with $k(x, y) = y - h(x)$ or $k(x, y) = x - g(y)$.

Now we will state some problems that arise in this field.

Problem 1. Are there any measures of statistical dependence $\delta(\xi, \eta)$ that satisfy all the requirements 1–4?

Problem 2. What are upper bounds of $\delta(\xi, \eta)$ and a lower bound of $e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta)$ figuring in the requirements 2 and 3 attainable under particular restrictions on random variables ξ and η?

Problem 3. What are the sample properties of adequate measures $\delta(\xi, \eta)$?

In the following sections we try to answer at least partly all these questions.

3. f-INFORMATIONAL MEASURES OF STATISTICAL DEPENDENCE

In the sequel we shall be interested in measures of statistical dependence that are based on the notion of f divergence of two probability measures (called also generalized f-entropy [9], [11], [13]) introduced by I. Csiszar in [1]. The most important properties of f-divergences are based on the convexity of a function $f(u)$ defined on $[0, \infty)$, where the following conventions are observed:

(3) $$f(0) = \lim_{u \downarrow 0} f(u)\,, \quad 0 f\left(\frac{0}{0}\right) = 0$$

and

$$0 f\left(\frac{v}{0}\right) = v f_\infty \text{ where } v > 0 \text{ and } f_\infty = \lim_{u \uparrow \infty} \frac{f(u)}{u}\,.$$

For the sake of simplicity we shall denote $f_1 = f(1)$ and $f_2 = f_0 + f_\infty$, where $f_0 = f(0)$.

If we consider two probability measures $P_{\xi\eta}$ and $P_\xi \times P_\eta$ on $(X \times Y, \mathscr{X} \times \mathscr{Y})$ then in this special case the f-divergence of $P_{\xi\eta}$ and $P_\xi \times P_\eta$ is defined by

$$\text{(4)} \qquad D_f(P_{\xi\eta}, P_\xi \times P_\eta) = \int_{X \times Y} f\left(\frac{p_{\xi\eta}(x, y)}{p_\xi(x)\, p_\eta(x)}\right) p_\xi(x)\, p_\eta(y)\, d\lambda\,.$$

According to the notation in [9] we shall call $[D_f(P_{\xi\eta}, P_\xi \times P_\eta) - f_1]$ the generalized f-information. However, considering the fact that the additive constant $-f_1$ is irrelevant in all what follows, for the purpose of this paper we denote

$$\text{(5)} \qquad I_f(\xi, \eta) = D_f(P_{\xi\eta}, P_\xi \times P_\eta)$$

and also call it the f-information.

In statistics some f-informations have been frequently used for measuring statistical dependence between two random variables. The most important of them are *Pearson's mean square contingency*

$$\text{(6)} \qquad \chi^2 = \int_{X \times Y} \frac{[p_{\xi\eta}(x, y) - p_\xi(x)\, p_\eta(y)]^2}{p_\xi(x)\, p_\eta(y)}\, d\lambda$$

with $f(u) = (1 - u)^2$, *Shannon's information*

$$\text{(7)} \qquad I = \int_{X \times Y} p_{\xi\eta}(x, y) \log \frac{p_{\xi\eta}(x, y)}{p_\xi(x)\, p_\eta(y)}\, d\lambda$$

with $f(u) = u \log u$ and *Höffding's coefficient of statistical dependence*

$$\text{(8)} \qquad \gamma = \frac{1}{2} \int_{X \times Y} \left| p_{\xi\eta}(x, y) - p_\xi(x)\, p_\eta(y) \right| d\lambda$$

with $f(u) = \frac{1}{2}|1 - u|$. Moreover, γ and $e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta)$ are tied together by the relation [19]

$$\text{(9)} \qquad e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta) = \tfrac{1}{2}(1 - \gamma)\,.$$

One of further measures of statistical dependence based on the notion of f-information is *Hellinger's integral*

$$\text{(10)} \qquad h = -\int_{X \times Y} [p_{\xi\eta}(x, y)\, p_\xi(x)\, p_\eta(y)]^{1/2}\, d\lambda$$

with $f(u) = -\sqrt{u}$.

The adequacy of f-information with $f(u) = u \log u$ for measuring statistical dependence has been already discussed in [9], [10]. The relationship of f-informations with convex functions $f(u)$ satisfying (3) to the requirements in [15] has been examined in [2].

Now we shall try to give some statements concerning the behaviour of measures of statistical dependence based on f-informations with respect to the requirements 1–4. We strongly rely on the results of I. Csiszar [1], [2], A. Perez [9], [11] and I. Vajda [19], [20] that systematically examined properties of f-divergences.

Let us denote by F the class of convex functions $f(u)$ defined on $[0, \infty)$ and satisfying the conventions (3) and let $\tilde{F}$ be a subclass of F such that every $f(u) \in \tilde{F}$ is strictly convex with $f_2 < \infty$.

Theorem 1. *For every* $f(u) \in \tilde{F}$

$$\delta_f(\xi, \eta) = \frac{I_f(\xi, \eta) - f_1}{f_2 - f_1} \tag{11}$$

satisfies all the requirements 1–4.

Proof. The satisfaction of the requirements 1, 2, 3 follows directly from the reuslts in [20], [12] and 4 from [1].

Remark 3. We can notice that the function $f(u) = -\sqrt{u}$ (Hellinger's integral h is based on it) satisfies the assumptions of Theorem 1.

Remark 4. If $f(u) \in F$ with $f_2 < \infty$ is not strictly convex, we cannot quarantee that $\delta_f(\xi, \eta)$ given by (11) satisfies the requirements 2.a), 2.c) and 4.b). However, for the function $f(u) = \frac{1}{2}|1 - u|$ (minimüm probability of error $e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta)$ and Höffding's coefficient of statistical dependence γ are based on it) that is not strictly convex with $f_2 < \infty$, we can state the following obvious lemma.

Lemma 3. *Höffding's coefficient of statistical dependence* $\gamma = 0$ (i.e. $e_{1/2}(P_{\xi\eta}, P_\xi \times P_\eta) = \frac{1}{2}$) *if and only if* ξ *and* η *are independent.*

From Lemma 3 it follows that Höffding's coefficient of statistical dependence satisfies moreover the requirement 2.a) and it obviously satisfies also the requirement 2.c).

Theorem 2. *For every* $f(u) \in F$ *with* $f_2 = \infty$

$$\delta_f(\xi, \eta) = \varphi[I_f(\xi, \eta)] \tag{12}$$

satisfies the requirements 1, 3.b) *and* 4.a). *The function* $\varphi(t)$ *in* (12) *is an arbitrary real function defined and increasing on* $[f_1, \infty]$, *with* $\varphi(\infty) = \lim_{t \uparrow \infty} \varphi(t)$, *that is mapping the closed interval* $[f_1, \infty]$ *onto the closed interval* $[0, 1]$.

Proof follows from the results in [20], [1].

This sort of transformations of f-informations with $f_2 = \infty$ has already been used in statistics. For example, the transformation of Pearson's mean square contingency χ^2 by the function $\varphi_1(t) = \sqrt{(t/(1+t))}$ gives the contingency coefficient $\sqrt{(\chi^2/(1+\chi^2))}$ [3]. The transformation of Shannon's information I by the function $\varphi_2(t) = \sqrt{(1 - e^{-2t})}$ gives the informational coefficient of correlation $\sqrt{(1 - e^{-2I})}$ [5].

We can find many other functions $\varphi(t)$ that are increasing on $(f_1, \infty]$ and mapping this interval onto $[0, 1]$. However, the functions $\varphi_1(t)$ and $\varphi_2(t)$ are mapping χ^2 and I respectively onto the closed interval $[0, 1]$ in such a way that in the case of Gaussian distribution $P_{\xi\eta}$ with the coefficient of correlation ϱ

$$\sqrt{\frac{\chi^2}{1+\chi^2}} = \sqrt{(1 - e^{-2I})} = |\varrho| . \tag{13}$$

This property for adequate measures of statistical dependence has been required in [15].

Further, $\delta_f(\xi, \eta)$ given by (11) and (12) will be called f-informational measures of statistical dependence. We can see that f-informational measures of statistical dependence are even symmetrical, i.e. $\delta_f(\xi, \eta) = \delta_f(\eta, \xi)$. The symmetry for adequate measures of statistical dependence has been required in [15]. However, it remains an open problem whether the symmetry of measures of statistical dependence is a useful property in general. In some cases asymmetrical measures of statistical dependence seem to be much preferable [10].

4. UPPER BOUNDS OF f-INFORMATIONAL MEASURES OF STATISTICAL DEPENDENCE

In Sec. 3 we introduce a class of f-informational measures of statistical dependence and have not put any restrictions on random variables (ξ, η) under consideration. In some cases we can a priori restrict the investigated class of random variables (ξ, η) and then it may happen that the highest dependence defined by the requirement 3.a) in Sec. 3 never can occur. According to Lemma 1, this arises in all cases when there exists an atom in $(X \times Y, \mathscr{X} \times \mathscr{Y}, P_{\xi\eta})$ and, consequently, in the case when we consider a class of random variables $(\xi, \eta) \to (X \times Y, \tilde{\mathscr{X}} \times \tilde{\mathscr{Y}}, P_{\xi\eta})$, where $\tilde{\mathscr{X}}$ and $\tilde{\mathscr{Y}}$ are σ-algebras generated by measurable decompositions $D_X = (X_1, X_2, \ldots, X_r)$ of $(X, \mathscr{X})$ and $D_Y = (Y_1, Y_2, \ldots, Y_s)$ of $(Y, \mathscr{Y})$ respectively. Therefore it seems to be useful to ask for attainable upper bounds of $\delta_f(\xi, \eta)$ with respect to an a priori restricted class of random variables (ξ, η). Owing to the relations (11), (12) it is sufficient to solve this problem for f-informations $I_f(\xi, \eta)$. In the sequel we use the notation introduced above.

Theorem 3. *Let* $\xi \to (X, \tilde{\mathscr{X}}, P_\xi)$ *and* $\eta \to (Y, \tilde{\mathscr{Y}}, P_\eta)$ *be two random variables and let* $(\xi, \eta) \to (X \times Y, \tilde{\mathscr{X}} \times \tilde{\mathscr{Y}}, P_{\xi\eta})$ *be a random variable with marginal probability* *measures* P_ξ *and* P_η *on* $(X, \widetilde{\mathscr{X}})$ *and* $(Y, \widetilde{\mathscr{Y}})$ *respectively. Let us denote by* $p_{ij} = \mathrm{P}_{\xi\eta}(X_i \times Y_j)$, $p_{i\cdot} = \mathrm{P}_\xi(X_i)$, $p_{\cdot j} = \mathrm{P}_\eta(X_j)$ *for* $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$ *and assume* $p_{i\cdot} > 0$ *for* $i = 1, 2, \ldots, r$, $p_{\cdot j} > 0$ *for* $j = 1, 2, \ldots, s$. *Then*

$$\mathrm{I}_f(\xi, \eta) \leqq \min\left[\mathrm{H}_f(\xi), \mathrm{H}_f(\eta)\right], \tag{14}$$

where

$$\mathrm{H}_f(\xi) = \sum_{i=1}^{r} p_{i\cdot}^2 f\left(\frac{1}{p_{i\cdot}}\right) + f(0)\,(1 - \sum_{i=1}^{r} p_{i\cdot}^2) \tag{15}$$

and

$$\mathrm{H}_f(\eta) = \sum_{j=1}^{s} p_{\cdot j}^2 f\left(\frac{1}{p_{\cdot j}}\right) + f(0)\,(1 - \sum_{j=1}^{s} p_{\cdot j}^2)\,. \tag{16}$$

Proof. Let us consider two measurable spaces $(I, \mathscr{B}, \mu)$ and $(R, \mathscr{R}, \nu)$, where $I = [0, 1]$, $\mathscr{B}$ is the σ-algebra of Borel sets in I and μ is Lebesgue measure, $R = \{1, 2, \ldots, r\}$, $\mathscr{R}$ is the σ-algebra of all subsets in R and ν is the counting measure. Let us divide I into r intervals $J_i = [a_i, b_i)$, $i = 1, 2, \ldots, (r - 1)$, $J_r = [a_r, b_r]$, where $a_i = \sum_{k=0}^{i-1} p_{k\cdot}$, $b_i = \sum_{k=0}^{i} p_{k\cdot}$, $p_{0\cdot} = 0$ and let $g_i(t)$ denote the density of uniform distribution on J_i, $i = 1, 2, \ldots, r$. A measurable decomposition $D_{J_i} = (E_{i1}, E_{i2}, \ldots, E_{is})$ of $(J_i, \mathscr{B}_i)$, $\mathscr{B}_i = J_i \cap \mathscr{B}$ into s parts is done in such a way that

$$\int_{E_{ij}} g_i(t)\,\mathrm{d}\mu(t) = \frac{p_{ij}}{p_{i\cdot}}, \quad i = 1, 2, \ldots, r,\ j = 1, 2, \ldots, s\,.$$

If we denote $E_j = \bigcup_{i=1}^{r} E_{ij}$, $j = 1, 2, \ldots, s$, then $D_E = (E_1, E_2, \ldots, E_s)$ is a measurable decomposition of $(I, \mathscr{B})$ and $S(D_E)$ denotes the minimum σ-algebra generated by D_E.

We can define a probability measure $\widetilde{\mathrm{P}}_{\xi\eta}$ on $(R \times I, \mathscr{R} \times \mathscr{B})$ in the following way: $\mathrm{d}\widetilde{\mathrm{P}}_{\xi\eta}(i, t) = p_{i\cdot}\, g_i(t)\, \mathrm{d}[\nu \times \mu]$. Then marginal probability measures of $\widetilde{\mathrm{P}}_{\xi\eta}$ on $(I, \mathscr{B})$ and $(R, \mathscr{R})$ are $\mathrm{d}\widetilde{\mathrm{P}}_\eta(t) = h(t)\,\mathrm{d}\mu$, where $h(t) = \sum_{i=1}^{r} p_{i\cdot}\, g_i(t)$ and $\widetilde{\mathrm{P}}_\xi(i) = p_{i\cdot}$ respectively. If we denote by $\mathrm{P}_{\xi\eta}$ the restriction of $\widetilde{\mathrm{P}}_{\xi\eta}$ on $(R \times I, \mathscr{R} \times S(D_E))$, we can see that $\mathrm{P}_{\xi\eta}(i, E_j) = p_{ij}$ for $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. Then it follows from Theorem 3 in [1]

$$\mathrm{I}_f(\xi, \eta) = \mathrm{D}_f(\mathrm{P}_{\xi\eta}, \mathrm{P}_\xi \times \mathrm{P}_\eta) \leqq \mathrm{D}_f(\widetilde{\mathrm{P}}_{\xi\eta}, \widetilde{\mathrm{P}}_\xi \times \widetilde{\mathrm{P}}_\eta) =$$

$$= \int_{R\times I} f\left(\frac{p_{i\cdot}\, g_i(t)}{p_{i\cdot}\, h(t)}\right) p_{i\cdot}\, h(t)\,\mathrm{d}[\nu \times \mu] = \sum_{i=1}^{r} p_{i\cdot}\,.$$

$$\cdot \sum_{j=1}^{s} \int_{E_j} f\left(\frac{p_{i\cdot}\, g_i(t)}{p_{i\cdot}\, h(t)}\right) p_{i\cdot}\, h(t)\,\mathrm{d}[\nu \times \mu] = \sum_{i=1}^{r} p_{i\cdot}\,.$$

$$\cdot \sum_{k=1}^{r} p_{k\cdot} f\left(\frac{\delta_{ik}}{p_{k\cdot}}\right) = \sum_{i=1}^{r} p_{i\cdot}^2 f\left(\frac{1}{p_{i\cdot}}\right) + f(0)\,(1 - \sum_{i=1}^{r} p_{i\cdot}^2) = \mathrm{H}_f(\xi)\,,$$

where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$ is the Kronecker symbol. Similarly we get $I_f(\xi, \eta) \leqq H_f(\eta)$. Therefore $I_f(\xi, \eta) \leqq \min [H_f(\xi), H_f(\eta)]$.

Corollary 3. *If* $r = s$, $p_{ii} = p_{i\cdot} = p_{\cdot i} = 1/r$, $i = 1, 2, \ldots, r$, *then*

$$I_f(\xi, \eta) = H_f(\xi) = \frac{f(r)}{r} + f(0)\frac{r-1}{r}. \tag{17}$$

Theorem 4. *Let us consider two random variables* $\xi \to (X, \tilde{\mathscr{X}}, P_\xi)$ *and* $\bar{\xi} \to \to (X, \tilde{\mathscr{X}}, P)$, *where* $P_\xi(X_i) = p_{i\cdot}$, $P(X_i) = 1/r$ *for* $i = 1, 2, \ldots, r$. *If* $g(u) = = [f(u) - f(0)]/u$ *is a concave function, then*

$$H_f(\xi) \leqq H_f(\bar{\xi}) = \frac{f(r)}{r} + f(0)\frac{r-1}{r}. \tag{18}$$

Proof. Applying Jensen's inequality we get

$$H_f(\xi) = \sum_{i=1}^{r} p_{i\cdot}^2 f\left(\frac{1}{p_{i\cdot}}\right) + f(0)\,(1 - \sum_{i=1}^{r} p_{i\cdot}^2) = f(0) +$$

$$+ \sum_{i=1}^{r} p_{i\cdot} \left[\frac{f\left(\frac{1}{p_{i\cdot}}\right) - f(0)}{\frac{1}{p_{i\cdot}}}\right] \leqq \frac{f(r)}{r} + f(0)\frac{r-1}{r}.$$

Remark 5. We can see that Theorem 4 holds for example for the following functions: $f(u) = u \log u$, $f(u) = |1 - u|$, $f(u) = (1 - u)^2$ and $f(u) = -u^\alpha$, $\alpha \in (0, 1)$.

Remark 6. Shannon's inequality follows from Theorem 3 for $f(u) = u \log u$.

Corollary 4. *If* $(\xi, \eta) \to (X \times Y, \tilde{\mathscr{X}} \times \tilde{\mathscr{Y}}, P_{\xi\eta})$, $\bar{\xi} \to (X, \mathscr{X}, P)$ *and* $\bar{\eta} \to (Y, \tilde{\mathscr{Y}}, Q)$ *are random variables, where* $P(X_i) = 1/r$ for $i = 1, 2, \ldots, r$ *and* $Q(Y_j) = 1/s$ *for* $j = 1, 2, \ldots, s$ *and* $g(u) = [f(u) - f(0)]/u$ *is a concave function, then*

$$I_f(\xi, \eta) \leqq \min [H_f(\bar{\xi}), H_f(\bar{\eta})]. \tag{19}$$

We see that Corollary 4 enables us to estimate upper bounds of $I_f(\xi, \eta)$ for (ξ, η) with unknown marginal probability distributions and Theorem 3 for (ξ, η) with a priori given marginal distributions. However, there are some cases when we can evaluate the maximum of $I_f(\xi, \eta)$ over all random variables $(\xi, \eta) = (X \times Y, \tilde{\mathscr{X}} \times \tilde{\mathscr{Y}}, P_{\xi\eta})$ with a priori given marginal probability distributions P_ξ and P_η, directly.

Lemma 4. *Let* $\xi \to (X, \tilde{\mathscr{X}}, P)$ *with* $r = 2$, $P(X_1) = p > 0$, $P(X_2) = (1 - p) > 0$ *and* $\bar{\eta} \to (Y, \tilde{\mathscr{Y}}, Q)$ *where* $Q(Y_j) = 1/s$, $j = 1, 2, \ldots, s$. *Let us denote by* $\mathscr{C}$ *a class of* *random variables* $(\xi, \eta) = (X \times Y, \tilde{\mathscr{X}} \times \tilde{\mathscr{Y}}, P_{\xi\eta})$ *with marginal probability measures* $P_\xi = P$ *and* $P_\eta = Q$ *on* $(X, \tilde{\mathscr{X}})$ *and* $(Y, \tilde{\mathscr{Y}})$ *respectively. Then the maximum of* $I_f(\xi, \eta)$ *over* $\mathscr{C}$ *is equal to*

$$\max_{\mathscr{C}} I_f(\xi, \eta) = k\varphi\left(\frac{1}{s}\right) + \varphi\left(p - \frac{k}{s}\right) + (s - k - 1)\,\varphi(0)\,, \tag{20}$$

where

$$\varphi(x) = \frac{p}{s} f\left(\frac{sx}{p}\right) + \frac{1-p}{s} f\left(\frac{1-sx}{1-p}\right)$$

and $k = [sp]$.

Proof. Let us consider random variables $(\xi, \eta) = (X \times Y, \tilde{\mathscr{X}} \times \mathscr{Y}, P_{\xi\eta})$ with marginal probability measures $P_\xi = P$ and $P_\eta = Q$ and denote $P_{\xi\eta}(X_1 \times Y_j) = z_j$, $P_{\xi\eta}(X_2 \times Y_j) = (1/s) - z_j$ for $j = 1, 2, \ldots, s$, $\mathbf{z} = (z_1, z_2, \ldots, z_s)$, $0 \leqq z_j \leqq 1/s$, $\sum_{j=1}^{s} z_j = p$. Then

$$I_f(\xi, \eta) = D_f(P_{\xi\eta}, P \times Q) = \sum_{j=1}^{s} \left[\frac{p}{s} f\left(\frac{sz_j}{p}\right) + \frac{1-p}{s} f\left(\frac{1-sz_j}{1-p}\right)\right] =$$
$$= \sum_{j=1}^{s} \varphi(z_j) = \Phi(\mathbf{z}).$$

In view of the convexity of $f(u)$ we can see that $\Phi(\mathbf{z})$ is a convex function. Then by Theorem 4d in [21], $\Phi(\mathbf{z})$ reaches its maximum value at the point $\mathbf{z}^* = (z_1^*, z_2^*, \ldots, z_s^*)$, where $z_1^* = z_2^* = \ldots = z_k^* = 1/s$, $z_{k+1}^* = p - 1/s$, $z_{k+2}^* = z_{k+3}^* = \ldots = z_s^* = 0$. Putting $P_{\xi\eta}(X_1 \times Y_j) = z_j^*$, $P_{\xi\eta}(X_2 \times Y_j) = (1/s) - z_j^*$, $j = 1, 2, \ldots, s$, we obtain in this case

$$I_f(\xi, \eta) = D_f(P_{\xi\eta}, P \times Q) = k\,\varphi\left(\frac{1}{s}\right) + \varphi\left(p - \frac{k}{s}\right) + (s - k - 1)\,\varphi(0)\,,$$

where $k = [sp]$, which proves (20).

The following Theorem shows the relationship of the strict dependence of random variables ξ and η to the attainability of the upper bounds $H_f(\xi)$ and $H_f(\eta)$ given by (15) and (16) respectively.

Theorem 5. *Under the assumptions of Theorem* 3, *the strict dependence* $\xi = g(\eta)$ *or* $\eta = h(\xi)$ *implies* $I_f(\xi, \eta) = H_f(\xi)$ *and* $I_f(\xi, \eta) = H_f(\eta)$, *respectively. If, moreover,* $f(u)$ *is a strictly convex function, then* $I_f(\xi, \eta) = H_f(\xi)$ *if and only if* $\xi = g(\eta)$ *and* $I_f(\xi, \eta) = H_f(\eta)$ *if and only if* $\eta = h(\xi)$.

Proof. Let us consider $\xi = g(\eta)$. Then for every i 1, 2, …, $(i = r)$ there exist numbers $p_{\cdot i_1}, p_{\cdot i_2}, \ldots, p_{\cdot i_{n(i)}}$ such that $\sum_{k=1}^{n(i)} p_{\cdot i_k} = p_{i\cdot}$ and $P_{\xi\eta}(X_i \times Y_j) = 0$ for $j \neq i_k$

$(k = 1, 2, \ldots, n(i))$, $P_{\xi\eta}(X_i \times Y_j) = p_{\cdot i_k}$ for $j = i_k$ $(k = 1, 2, \ldots, n(i))$, for $i = 1, 2, \ldots$ $\ldots, r$, $j = 1, 2, \ldots, s$. Hence

$$I_f(\xi, \eta) = \sum_{i=1}^{r} \sum_{k=1}^{n(i)} p_{i\cdot}\, p_{\cdot i_k} f\left(\frac{p_{\cdot i_k}}{p_{i\cdot}\, p_{\cdot i_k}}\right) + f(0)\,(1 - \sum_{i=1}^{r} \sum_{k=1}^{n(i)} p_{i\cdot}\, p_{\cdot i_k}) = H_f(\xi)\,.$$

Let $f(u)$ be a strictly convex function and $I_f(\xi, \eta) = H_f(\xi)$. Let us assume that $\xi = g(\eta)$ does not hold. Then there exists $j(1 \leqq j \leqq s)$ such that $1 > p_{ij}/p_{\cdot j} \geqq 0$ for $i = 1, 2, \ldots, r$ and $\sum_{i=1}^{r} p_{ij}/p_{\cdot j} = 1$. Without loss of generality we can put $j = s$ and assume $p_{is}/p_{\cdot s} > 0$ for $i = 1, 2, \ldots, m$, $m \geqq 2$ and $p_{is}/p_{\cdot s} = 0$ for $i = (m + 1)$, $(m + 2), \ldots, r$. Let us consider two measurable spaces $(I, \mathscr{B}, \mu)$ and $(R, \mathscr{R}, \nu)$, where $I = [0, 1]$, $\mathscr{B}$ is the σ-algebra of Borel sets and μ is Lebesque measure, $R = \{1, 2, \ldots, r\}$, $\mathscr{R}$ is the σ-algebra of all subsets of R and ν is the counting measure. Let us establish a measurable decomposition D_J of $(I, \mathscr{B})$, $D_J = (\tilde{J}_1, \tilde{J}_2, \ldots, \tilde{J}_{s+m-1})$, where $\tilde{J}_j = [a_j, b_j)$, $j = 1, 2, \ldots, (s + m - 2)$ and $J_{(s+m-1)} = [a_{(s+m-1)}, b_{(s+m-1)}]$, $a_j = \sum_{k=0}^{j-1} p_{\cdot k}$, $b_j = \sum_{k=0}^{j} p_{\cdot k}$ for $j = 1, 2, \ldots, (s - 1)$, $a_j = \sum_{k=0}^{s-1} p_{\cdot k} + \sum_{l=0}^{j-s} p_{ls}$, $b_j = \sum_{k=0}^{s-1} p_{\cdot k} + \sum_{l=0}^{j-s+1} p_{ls}$ for $j = s, (s + 1), \ldots, (s + m - 1)$, and $p_{\cdot 0} = p_{0s} = 0$. Let us denote by $S(D_J)$ the σ-algebra generated by D_J and define a probability measure $\tilde{P}_{\xi\eta}$ on $(R \times I, \mathscr{R} \times S(D_J))$ in the following way:

$$\tilde{P}_{\xi\eta}(i, \tilde{J}_j) = p_{ij} \text{ for } i = 1, 2, \ldots, r, j = 1, 2, \ldots, (s - 1)\,,$$

$$\tilde{P}_{\xi\eta}(i, \tilde{J}_{(s+t-1)}) = \delta_{it} p_{is} \quad \text{for} \quad i = 1, 2, \ldots, r,\ t = 1, 2, \ldots, m\,.$$

At the same time P_ξ and $\tilde{P}_\eta$ denote the corresponding marginal probability measures on $(R, \mathscr{R})$ and $(I, S(D_J))$ respectively. Let us establish another measurable decomposition $D_J = (J_1, J_2, \ldots, J_s)$ of $(I, \mathscr{B})$, where $J_i = \tilde{J}_i$, $i = 1, 2, \ldots, (s - 1)$ and $J_s = \bigcup_{i=s}^{s+m-1} \tilde{J}_i$ and denote by $P_{\xi\eta}$ the restriction of $\tilde{P}_{\xi\eta}$ on $(R \times I, \mathscr{R} \times S(D_J))$. Owing to the fact that σ-algebra $\mathscr{R} \times S(D_J)$ is not sufficient with respect to $\tilde{P}_{\xi\eta}$ and $P_\xi \times \tilde{P}_\eta$, Theorem 3 in [1] and Theorem 3 imply $H_f(\xi) = I_f(\xi, \eta) = D_f(P_{\xi\eta}, P_\xi \times P_\eta) < D_f(\tilde{P}_{\xi\eta}, P_\xi \times \tilde{P}_\eta) \leqq H_f(\xi)$ which leads to contradiction.

Remark 8. For $f(u) = u \log u$ Theorem 5 gives the known result for Shannon's information.

5. α-INFORMATIONAL MEASURES OF STATISTICAL DEPENDENCE

In Sec. 3 we met with two important subclasses of f-informations that led to interesting measures of statistical dependence. The first one is given by $f(u) = |1 - u|^\alpha$, $\alpha \geqq 1$ and such f-informations we call α-informations [20]. To this subclass total variation with $f(u) = |1 - u|$ and Pearson's mean square contingency with $f(u) = (1 - u)^2$ belong. To the second subclass with $f(u) = \operatorname{sign}(\alpha - 1)\, u^\alpha$, $\alpha > 0$, [2], [11], [18], Hellinger's integral with $f(u) = -\sqrt{u}$ belongs and Shannon's information can be derived by [11]

$$\lim_{\alpha \downarrow 1} \frac{I_{u^\alpha}(\xi, \eta) - 1}{\alpha - 1} = I \tag{21}$$

and

$$\lim_{\alpha \uparrow 1} \frac{I_{-u^\alpha}(\xi, \eta) + 1}{1 - \alpha} = I$$

or [16]

$$\lim_{\alpha \to 1} \frac{1}{\alpha - 1} \log |I_{\operatorname{sign}(\alpha-1)u^\alpha}(\xi, \eta)| = I\,, \tag{22}$$

where $(1/(\alpha - 1)) \log |I_{\operatorname{sign}(\alpha-1)u^\alpha}(\xi, \eta)|$ is the so called Rényi's information of order α.

However, it seems that an important role for measuring the statistical dependence can be played by f-informations with $f(u) = -u^\alpha$, $\alpha \in (0, 1)$, that satisfy all the requirements 1–4. Moreover, the function $f(u)/u$ is a concave function and Theorem 4 holds. In the sequel, $\delta_\alpha(\xi, \eta)$ will denote f-informational measures of statistical dependence with $f(u) = -u^\alpha$, $\alpha \in (0, 1)$, i.e. $\delta_\alpha(\xi, \eta) = 1 + I_{-u^\alpha}(\xi, \eta)$, and $\delta_\alpha(\xi, \eta)$ will be called α-informational measure of statistical dependence. In the case of Gaussian distribution $P_{\xi\eta}$ with the coefficient of correlation ϱ the relationship of $\delta_\alpha(\xi, \eta)$ to ϱ is expressed by

$$\delta_\alpha(\xi, \eta) = 1 - \frac{(1 - \varrho^2)^{(1-\alpha)/2}}{[1 - \varrho^2(\alpha - 1)^2]^{1/2}}\,. \tag{23}$$

In the first subclass of f-informations with $f(u) = |1 - u|^\alpha$, $\alpha \geqq 1$, sample properties for $\alpha = 2$ have already been investigated in [8]. In this Section we shall be interested in sample properties of f-informations with $f(u) = -u^\alpha$, $\alpha \in (0, 1)$ under the hypothesis that ξ and η are independent.

Let $\xi = (X, \mathscr{X}, P_\xi)$ and $\eta = (Y, \mathscr{Y}, P_\eta)$ be two random variables and let $(\xi, \eta) = (X \times Y, \mathscr{X} \times \mathscr{Y}, P_{\xi\eta})$ be a random variable with marginal probability measures P_ξ and P_η on $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ respectively. Let us denote $p_{ij} = P_{\xi\eta}(X_i \times Y_j)$, $p_{i.} = P_\xi(X_i)$, $p_{.j} = P_\eta(Y_j)$ and assume $p_{i.} > 0$, $p_{.j} > 0$ for $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, s$. Let us have n independent realizations of (ξ, η), i.e., (x_t, y_t), $t = 1, 2, \ldots, n$ from a sample space $(X \times Y)$. Let $\hat{p}_{ij} = n_{ij}/n$, $\hat{p}_{i.} = n_{i.}/n$, $\hat{p}_{.j} = n_{.j}/n$ be sample