*Periodica Mathematica Hungarica Vol. 2 (1–4), (1972), pp. 191–213.*

# A CLASS OF MEASURES OF INFORMATIVITY OF OBSERVATION CHANNELS

by

I. CSISZÁR (Budapest)[1]

*Dedicated to the memory of Alfréd* RÉNYI

**Summary**

A class of numerical measures of informativity of observation channels or statistical experiments is defined by the aid of $f$-divergences introduced by the author as measures of difference of two probability distributions. For observation channels with given prior probabilities, the $f$-informativity measures are generalizations of Shannon's mutual information and include Gallager's function $E_0(\varrho,\mathfrak{Q})$ appearing in the derivation of error exponent for noisy channels, as well. For observation channels without prior probabilities, the suggested informativity measures have the geometric interpretation of a radius.

The $f$-informativity defined for the Bayesian case shares several useful properties of the mutual information, such as e. g. the "data processing theorem". Its maximum with respect to all possible prior distributions is shown by a minimax argument to be just the $f$-radius, thus the latter is a generalization of channel capacity. The $f$-informativity measures can also be used to characterize the statistical sufficiency of indirect observations.

## § 1. Introduction

While Shannon's measure of the amount of information is a cornerstone of information theory, generalizations of Shannon's entropy function have also been suggested. This kind of research was initiated by A. RÉNYI's paper [17]. He introduced the concept of entropy of order $\alpha$ which shares several nice properties of Shannon's entropy. Examples of concrete problems leading to entropy of order $\alpha$ are also known, see e. g. [5], [11].

From Shannon's entropy one immediately gets a most useful measure of mutual information of two random variables. No similar measures of some practical value seem to have been obtained from generalized entropies, in

[1] This work was done while the author was visiting professor at The Catholic University of America, Washington, D.C., sponsored by National Science Foundation Grant No. GP-9396.

spite of the fact that Gallager's function $E_0(\varrho,\mathfrak{Q})$, see e.g. [9], p. 138, playing an important role in coding theory behaves in many respects like the mutual information.

This fact was one main motivation of this paper. Apparently, when generalizing mutual information, one should start from generalized $I$-divergence rather than generalized entropy. A possible way of doing this was indicated by I. VAJDA [24]; here we adopt another approach. The class of informativity measures we are going to introduce includes both Shannon's mutual information and Gallager's function (the latter aside from a scale transformation), eliminating the need of deriving their common properties—in particular those connected with their maximization for a given channel—separately.

Another aim of the author was to apply to more general experiments the measures of difference of probability distributions called $f$-divergences, introduced in [7], which proved to be useful in the case of simple alternative hypotheses (see [7], [8] and also [16]). In this respect, R. SIBSON's paper [21] should be referred to; his "information radius of order $\alpha$" defined by the aid of RÉNYI's "information gain of order $\alpha$" (see [17]) is—aside from a scale transformation—a particular case of the informativity measures defined here. SIBSON's approach was motivated, however, by statistics only (he did not even point out that his information radius of order 1 is identical with Shannon's mutual information), and essentially on a Bayesian basis; sorts of "absolute" informativity measures (which are closer to the intuitive idea of a radius) were not considered by him.

Finally, the author wants to make clear his point of view about the value of generalizations of Shannon's information measure. In force of the coding theorems of information theory, Shannon's measure of the amount of information in the very concrete sense as described by these theorems cannot be challenged. In this respect, information theorists claiming that this information measure is the only true one are right. There are, however, many problems outside the scope of the mentioned theorems where one may wish to speak of "information" in a technical sense, as e. g. the characterization of informativity of statistical experiments or observation channels. In such cases Shannon's information measure may still be useful as shown by D. V. LINDLEY [13] (further merits of this approach have been pointed out in A. RÉNYI's paper [18] initiating an extensive research, cf. e. g. [19], [23]) but there is no reason to believe that it is always the most suitable one. The distinguished role of Shannon's information measure, however, will be honored in this paper by reserving the technical term "information" for it only, while the introduced generalizations will be referred to as "informativity".

## § 2. Preliminaries.
## Definition and simple properties of f-informativity measures

Throughout this paper, the terms "probability distribution", "random variable" and "observation channel" will be abbreviated as PD, RV and OC, respectively. "Almost every" or "almost everywhere" will be abbreviated as a. e. and "if and only if" as iff.

$X$, $Y$ and $Z$ will denote sets, $Y$ and $Z$ will be considered to be equipped with $\sigma$-algebras $\mathcal{Y}$ and $\mathcal{Z}$, respectively. The measurable spaces $(Y, \mathcal{Y})$, $(Z, \mathcal{Z})$ and $(Y \times Z, \mathcal{Y} \times \mathcal{Z})$ will be referred to simply as $Y$, $Z$ and $Y \times Z$, respectively; it will always be clear from the context, whether these symbols mean just a set or, rather, a measurable space. In particular, a PD $\mathcal{P}$ on $Y$ is understood as a measure on $\mathcal{Y}$ with $\mathcal{P}(Y) = 1$.

DEFINITION 2.1. An experiment with parameter space $X$ and sample space $Y$ or an OC from $X$ to $Y$ is defined as a family $\Pi = \{\mathcal{P}_x\}_{x \in X}$ of PD's on $Y$.

In recent literature the term experiment has been used also in a more general sense, see e. g. [10]. The term OC was introduced by A. PEREZ [15] who imposed the additional condition that $\mathcal{P}_x(B)$ be $\mathcal{X}$-measurable for every $B \in \mathcal{Y}$ where $\mathcal{X}$ is a given $\sigma$-algebra of subsets of $X$.

As to the terminology, statisticians would probably prefer to speak of experiments while information theorists of OC's. In the sequel, the term OC will be used.

DEFINITION 2.2. Let $f(u)$, $u \in (0, \infty)$ be an arbitrary convex function which is strictly convex at $u = 1$. The $f$-divergence of two PD's $\mathcal{P}$ and $\mathcal{Q}$ on $Y$ is defined as

$$I_f(\mathcal{P} \| \mathcal{Q}) = \int q(y) f\left(\frac{p(y)}{q(y)}\right) \lambda(dy) \tag{2.1}$$

where $\lambda$ is some (finite or $\sigma$-finite) dominating measure and $p(y)$ and $q(y)$ are the densities of $\mathcal{P}$ and $\mathcal{Q}$ respectively, with respect to $\lambda$.

Undefined terms in (2.1) are understood as

$$\begin{gathered} f(0) = \lim_{u \downarrow 0} f(u); \quad 0 \cdot f\left(\frac{0}{0}\right) = 0; \\ 0 \cdot f\left(\frac{a}{0}\right) = \lim_{u \downarrow 0} uf\left(\frac{a}{u}\right) = a \lim_{u \downarrow 0} uf\left(\frac{1}{u}\right) \qquad (a > 0). \end{gathered} \tag{2.2}$$

This definition was introduced (with slightly different notations and without the assumption that $f$ is strictly convex at $u = 1$) in [7]. It is easy to show (see [7], [8]) that the integral (2.1) is always well defined, its value does not depend on the choice of $\lambda$ and

$$I_f(\mathcal{P} \,||\, \mathcal{Q}) \geqq f(1), \qquad \text{equality iff } \mathcal{P} = \mathcal{Q}; \tag{2.3}$$

moreover, there exists a function $\varphi(u)$ (depending on $f$) with $\lim_{u \downarrow f(1)} \varphi(u) = 0$ such that

$$|\mathcal{P} - \mathcal{Q}| \leqq \varphi(I_f(\mathcal{P} \,||\, \mathcal{Q})) \tag{2.4}$$

where $|\mathcal{P} - \mathcal{Q}|$ denotes the total variation of the signed measure $\mathcal{P} - \mathcal{Q}$, called the variation distance of $\mathcal{P}$ and $\mathcal{Q}$.

The convexity of $f(u)$ is equivalent to that of

$$\tilde{f}(u) = uf\left(\frac{1}{u}\right) \tag{2.5}$$

and

$$I_f(\mathcal{P} \,||\, \mathcal{Q}) = I_{\tilde{f}}(\mathcal{Q} \,||\, \mathcal{P}) = \int p(y) \tilde{f}\left(\frac{q(y)}{p(y)}\right) \lambda(dy). \tag{2.6}$$

From (2.1) and (2.6) obviously follows that $I_f(\mathcal{P} \,||\, \mathcal{Q})$ is a convex function of both $\mathcal{P}$ and $\mathcal{Q}$.

One indication that the integral (2.1) is a reasonable measure of how different $\mathcal{P}$ and $\mathcal{Q}$ are consists in[2] (2.3), (2.4). Another such indication is the fact that $I_f(\mathcal{P} \,||\, \mathcal{Q})$ cannot be increased by indirect observations, i.e. if the PD's $\mathcal{P}$ and $\mathcal{Q}$ are changed to $\overline{\mathcal{P}}$ and $\overline{\mathcal{Q}}$, respectively, then

$$I_f(\overline{\mathcal{P}} \,||\, \overline{\mathcal{Q}}) \leqq I_f(\mathcal{P} \,||\, \mathcal{Q}). \tag{2.7}$$

If $f$ is strictly convex and $I_f(\mathcal{P} \,||\, \mathcal{Q}) < \infty$, the necessary and sufficient condition of the equality is the sufficiency (in the Halmos-Savage sense) of the indirect observation with respect to the pair $\{\mathcal{P}, \mathcal{Q}\}$. This result has been proved in [8] for the following types of indirect observations:

(i) Reduction of the underlying $\sigma$-algebra: $\overline{\mathcal{P}}$ is the restriction of $\mathcal{P}$ to some sub-$\sigma$-algebra of $\mathcal{Y}$.

(ii) Using a statistic: $T$ is a measurable mapping of $Y$ into $Z$, $\overline{\mathcal{P}} = \mathcal{P}T^{-1}$.

(iii) $T = \{T_y(\,\cdot\,)\}_{y \in Y}$ is an OC from $Y$ to $Z$ such that $T_y(C)$ is $\mathcal{Y}$-measurable for every $C \in \mathcal{Z}$;
$\overline{\mathcal{P}}(C) = \int T_y(C)\, \mathcal{P}(dy) \quad (C \in \mathcal{Z})$.

[2] This interpretation could be made even more attractive by restricting attention to functions $f$ vanishing at $u = 1$; the additive constant $f(1)$ is, however, rather harmless and some calculations will be simpler using a function $f$ with $f(1) \neq 0$.

Here we shall need this result also for the following generalization of case (iii):

(iv) $\mathcal{Y}_0 \subset \mathcal{Y}$ is a $\sigma$-ideal and $T_y(C)$ $(y \in Y,\ C \in \mathcal{Z})$ is a function with values in $[0, 1]$ such that

(a) $T_y(C)$ is $\mathcal{Y}$-measurable for every $C \in \mathcal{Z}$;

(b) $T_y(Z) = 1$ for every $y \in Y$ and, for any fixed family of pairwise disjoint sets $C_k \in \mathcal{Z}$, the set $\{y: \sum_k T_y(C_k) \neq T_y(\bigcup_k C_k)\}$ belongs to $\mathcal{Y}_0$;

then for any PD $\mathcal{P}$ on $Y$ such that PD $\mathcal{P}(B_0) = 0$ for every $B_0 \in \mathcal{Y}_0$ and such that there exists a PD $\mathcal{P}^\star$ on $Y \times Z$ with $\mathcal{P}^\star(B \times C) = \int_B T_y(C)\,\mathcal{P}(dy)$ for $B \in \mathcal{Y}$, $C \in \mathcal{Z}$ (such PD's $\mathcal{P}$ will be called admissible) $\overline{\mathcal{P}}$ is defined as the marginal of $\mathcal{P}^\star$ on $Z$:

$$\overline{\mathcal{P}}(C) = \int T_y(C)\,\mathcal{P}(dy) \qquad (C \in \mathcal{Z}). \tag{2.8}$$

The proof of (2.7) and of the condition of equality for indirect observations of type (iv) is exactly the same as for those of type (iii); here by sufficiency of an indirect observation of type (iv) with respect to an arbitrary OC, $\Pi = \{\mathcal{P}_x\}_{x \in X}$ from $X$ to $Y$ (such that the $\mathcal{P}_x$'s are admissible) we mean the existence of a function $S_z(B)$, $\mathcal{Z}$-measurable for every $B \in \mathcal{Y}$ and satisfying $\int_C S_z(B)\,\mathcal{P}_x(dz) = \mathcal{P}_x^\star(B \times C)$ for every $x \in X$, $B \in \mathcal{Y}$, $C \in \mathcal{Z}$.

REMARK 2.1. The importance of considering statistical operations defined by functions $T_y(C)$ with the properties (a) and (b) above has been revealed by the study of N. MORSE and M. SACKSTEDTER [14] of the problem of statistical isomorphism, see also [20].

When measuring the distance of two PD's by their $f$-divergence, the "$f$-radius" of a set $\Pi = \{\mathcal{P}_x\}_{x \in X}$ of PD's on $Y$ i.e. $\inf_{\mathcal{Q}} \sup_{x \in X} I_f(\mathcal{P}_x \,||\, \mathcal{Q})$ is a measure of how different PD's are contained in $\Pi$; this may be considered as a measure of informativity (or, rather, of potential informativity) of $\Pi$. If a prior PD $W$ is given on $X$ (defined on a $\sigma$-algebra of subsets of $X$ with respect to which $I_f(\mathcal{P}_x \,||\, \mathcal{Q})$ is measurable for each PD $\mathcal{Q}$ on $Y$), one can consider $\inf_{\mathcal{Q}} \int I_f(\mathcal{P}_x \,||\, \mathcal{Q})\,W(dx)$ as a measure of informativity of the OC $\Pi$, with prior PD $W$.

In this paper, attention will be restricted to the case that $X$ is a finite set, $X = \{1, \ldots, m\}$ say. Then a PD on $X$ is given by $W = \{w_1, \ldots, w_m\}$. The author intends to return to the general case in another paper.

DEFINITION 2.3. The $f$-informativity of an OC $\Pi = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$ from $X = \{1, \ldots, m\}$ to $Y$ with prior PD $W = \{w_1, \ldots, w_m\}$ on $X$ is defined as

$$I_f(\Pi, W) = \inf_{\mathcal{Q}} \sum_{i=1}^{m} w_i I_f(\mathcal{P}_i \,||\, \mathcal{Q}) \tag{2.9}$$

and the absolute $f$-informativity or $f$-radius of $\Pi$ is defined as

$$\varrho_f(\Pi) = \inf_{\mathcal{Q}} \max_{1 \leq i \leq m} I_f(\mathcal{P}_i \| \mathcal{Q}) \tag{2.10}$$

where the infimum is meant with respect to all PD's $\mathcal{Q}$ on $Y$.

From (2.3), (2.9) and (2.10) obviously follows

$$f(1) \leq I_f(\Pi, W) \leq \varrho_f(\Pi) \tag{2.11}$$

where the first inequality is strict unless all $\mathcal{P}_i$'s with $w_i > 0$ are identical, cf. (2.4).

Let $p_i(y)$ denote the density of $\mathcal{P}_i$, $i = 1, \dots, m$, with respect to a common dominating measure $\lambda$; write

$$Y^+ = \{y : \max_{1 \leq i \leq m} p_i(y) > 0\}; \quad Y^{++} = \{y : \min_{1 \leq i \leq m} p_i(y) > 0\}. \tag{2.12}$$

Note that $\varrho_f(\Pi)$ is always finite if $f(0) < \infty$; in fact, then $I_f(\mathcal{P}_i \| \mathcal{Q}) < \infty$, $i = 1, \dots, m$ e. g. for $\mathcal{Q} = \frac{1}{m}(\mathcal{P}_1 + \dots + \mathcal{P}_m)$. If $f(0) = \infty$, a necessary condition of $\varrho_f(\Pi) < \infty$ consists in $\lambda(Y^{++}) > 0$, see (2.12); this is sufficient, too, if $\tilde{f}(0) < \infty$ because in that case $I_f(\mathcal{P}_i \| \mathcal{Q}) < \infty$, $i = 1, \dots, m$ e. g. for the PD $\mathcal{Q}$ with $\lambda$-density $q(y) = c \min_{1 \leq i \leq m} p_i(y)$, where $c$ is a proper constant. Finally, if $f(0) = \tilde{f}(0) = \infty$, a necessary (but not sufficient) condition of $\varrho_f(\Pi) < \infty$ is $\lambda(Y^+ \setminus Y^{++}) = 0$, i.e. the mutual equivalence of the PD's $\mathcal{P}_i$ $i = 1, \dots, m$.

Lemma 2.1. *Both in* (2.9) *and* (2.10), *the infimum may be restricted to* $\mathcal{Q} \ll \mathcal{P}_1 + \dots + \mathcal{P}_m$.

Proof. For any $\mathcal{Q}$ with density $q(y)$ (having chosen $\lambda$ so that $\mathcal{Q} \ll \lambda$), consider the PD $\mathcal{Q}_1$ with density $q_1(y) = q(y) + \mathcal{Q}(Y \setminus Y^+)\, p_1(y)$ if $y \in Y^+$ and $q_1(y) = 0$ if $y \notin Y^+$ (see (2.12)). Then $\mathcal{Q}_1 \ll \mathcal{P}_1 + \dots + \mathcal{P}_m$ and as a particular case of (2.7)—easily checked directly, as well—we have $I_f(\mathcal{P}_i \| \mathcal{Q}_1) \leq I_f(\mathcal{P}_i \| \mathcal{Q})$ $i = 1, \dots, m$.

Our $f$-informativity measures are compatible with the classical concept of informativity of experiments.

A finite experiment $\Pi = \{\mathcal{P}_1, \dots, \mathcal{P}_m\}$ is said to be more informative than another $\Pi' = \{\mathcal{P}'_1, \dots, \mathcal{P}'_m\}$ iff for any loss function, every loss vector attainable by some decision function in $\Pi'$ is also attainable in $\Pi$. D. Blackwell [2] attributes this definition to Bohnenblust, Shapley and Sherman (unpublished work). Blackwell [3] has proved that $\Pi$ is more informative than $\Pi'$ iff after reduction to standard experiments (as defined by him), $\Pi'$ arises from $\Pi$—in our terminology—by an indirect observation of type (iii). Since standard reduction is a sufficient indirect observation of type (ii), from BLACKWELL's theorem and Proposition 2.1 below it follows that if $\Pi$ is more informative than $\Pi'$ in the classical sense, we have $I_f(\Pi, W) \geqq I_f(\Pi', W)$ for any prior PD $W$ and also $\varrho_f(\Pi) \geqq \varrho_f(\Pi')$.

PROPOSITION 2.1. *Let $\Pi = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$ be an* OC *from $X = \{1, \ldots, m\}$ to $Y$ and consider an indirect observation from $Y$ to $Z$ (of any type* (i)—(iv)*) changing $\Pi$ to $\overline{\Pi} = \{\overline{\mathcal{P}}_1, \ldots, \overline{\mathcal{P}}_m\}$. Then for arbitrary convex $f$ and any prior* PD *$W$ on $X$*

$$I_f(\overline{\Pi}, W) \leqq I_f(\Pi, W); \qquad \varrho_f(\overline{\Pi}) \leqq \varrho_f(\Pi). \tag{2.13}$$

COROLLARY. If the indirect observation is sufficient with respect to $\Pi$ (2.13) holds with equalities.

PROOF. For any PD $\mathcal{Q} \ll \mathcal{P}_1 + \ldots + \mathcal{P}_m$, we have from (2.7) $I_f(\overline{\mathcal{P}}_i \,\|\, \overline{\mathcal{Q}}) \leqq$ $\leqq I_f(\mathcal{P}_i \,\|\, \mathcal{Q})$, $i = 1, \ldots, m$, whence (2.13) follows by Definition 2.3 and Lemma 2.1. The role of the assumption $\mathcal{Q} \ll \mathcal{P}_1 + \ldots + \mathcal{P}_m$ is to ensure, in case of indirect observations of type (iv), that $\mathcal{Q}$ be admissible for the indirect observation if the $\mathcal{P}_i$'s are.

The Corollary is immediate, since sufficiency implies that $\Pi$, too, is obtainable from $\overline{\Pi}$ by an indirect observation (in general of type (iv), see the paragraph after (2.8)).

If $f(u) = u \log_2 u$, $I_f(\mathcal{P} \,\|\, \mathcal{Q})$ reduces to KULLBACK's $I$-divergence (see e. g. [12])

$$I(\mathcal{P} \,\|\, \mathcal{Q}) = \int p(y) \log_2 \frac{p(y)}{q(y)} \lambda(dy) \tag{2.14}$$

and $I_f(\Pi, W)$ is just the Shannon mutual information

$$I(\Pi, W) = \sum_{i=1}^{m} w_i I(\mathcal{P}_i \,\|\, \mathcal{Q}^*), \quad \mathcal{Q}^* = \sum_{i=1}^{m} w_i \mathcal{P}_i \tag{2.15}$$

between the input and output of the given OC; this follows from the easily checked identity of F. TOPSOE [22]

$$\sum_{i=1}^{m} w_i I(\mathcal{P}_i \,\|\, \mathcal{Q}) = \sum_{i=1}^{m} w_i I(\mathcal{P}_i \,\|\, \mathcal{Q}^*) + I(\mathcal{Q}^* \,\|\, \mathcal{Q}). \tag{2.16}$$

DEFINITION 2.4. Let $\xi$ be a RV with values in a finite set, say $X = \{1, \ldots, m\}$, and $\eta$ a RV with values in $Y$. Then the $f$-informativity of $\eta$ with respect to $\xi$ is defined as

$$I_f(\xi; \eta) = I_f(\Pi_{\eta/\xi}, W_\xi) \tag{2.17}$$

where $\Pi_{\eta/\xi} = \{\mathcal{P}_i\}_{1 \leqq i \leqq m}$, $\mathcal{P}_i(B) = \mathsf{P}\{\eta \in B / \xi = i\}$ $(B \in \mathcal{Y})$ and $W_\xi = \{w_1, \ldots, w_m\}$, $w_i = \mathsf{P}\{\xi = i\}$, $1 \leqq i \leqq m$.

From (2.11) follows that $f(1) \leqq I_f(\xi;\eta)$ with equality iff $\xi$ and $\eta$ are independent RV's, supporting the interpretation of $I_f(\xi;\eta)$ as a measure of informativity (see footnote [2]).

Another useful property of $I_f(\xi;\eta)$ common with Shannon's mutual information $I(\xi;\eta)$ is the validity of the so-called data processing theorem.

PROPOSITION 2.2. *Let the* RV*'s* $\tau, \xi, \eta, \zeta$ *form a Markov chain (in the indicated order) where* $\tau$ *and* $\xi$ *have a finite number of possible values while the state spaces* $Y$ *and* $Z$ *of* $\eta$ *and* $\zeta$, *respectively, are arbitrary. Then for any convex* $f$

$$I_f(\tau;\zeta) \leqq I_f(\xi;\eta). \tag{2.18}$$

PROOF. Consider the indirect observation of type (iv) defined by $\mathcal{Y}_0 = \{B_0 : \mathsf{P}\{\eta \in B_0\} = 0\}$, $T_y(C) = \mathsf{P}\{\zeta \in C \mid \eta = y\}$. From the Markov property follows that if we denote by $\mathcal{P}_i$ and $\bar{\mathcal{P}}_i$ the conditional distribution given $\xi = i$ of $\eta$ and $\zeta$, respectively, then $\mathcal{P}_i$ and $\bar{\mathcal{P}}_i$ are connected by (2.8), $i = 1, \ldots, m$ (we assume, without any loss of generality, that $\xi$ takes on the values $1, \ldots, m$ with probabilities $w_i > 0$, $i = 1, \ldots, m$). Thus in force of (2.17) and the first inequality in (2.13) we have

$$I_f(\xi;\zeta) \leqq I_f(\xi;\eta). \tag{2.19}$$

Now suppose that $\tau$ takes on the values, $1, \ldots, l$ with probabilities $\mathsf{P}\{\tau = h\} = v_h > 0$, $h = 1, \ldots, l$. Set $r_{hi} = \mathsf{P}\{\xi = i \mid \tau = h\}$ and let $\mathcal{P}_h^0$ denote the conditional distribution of $\zeta$ given $\tau = h$. Then $\mathcal{P}_h^0 = \sum_{i=1}^{m} r_{hi} \bar{\mathcal{P}}_i$ thus, by convexity,

$$I_f(\mathcal{P}_h^0 \| \mathcal{Q}) \leqq \sum_{i=1}^{m} r_{hi} I_f(\bar{\mathcal{P}}_i \| \mathcal{Q}). \tag{2.20}$$

Multiplying both sides of (2.20) by $v_h$ and summing for $1 \leqq h \leqq l$ we obtain

$$\sum_{h=1}^{l} v_h I_f(\mathcal{P}_h^0 \| \mathcal{Q}) \leqq \sum_{i=1}^{m} w_i I_f(\bar{\mathcal{P}}_i \| \mathcal{Q}); \tag{2.21}$$

this, in view of Definitions 2.3 and 2.4 implies

$$I_f(\tau,\zeta) \leqq I_f(\xi;\zeta). \tag{2.22}$$

completing the proof.

EXAMPLE 2.1. An important special case is $f(u) = -u^\alpha$, $0 < \alpha < 1$; in this case, we shall write $I_\alpha$ instead of $I_f$. We obtain

$$I_\alpha(\mathcal{P} \| \mathcal{Q}) = -\int p^\alpha(y)\, q^{1-\alpha}(y)\, \lambda(dy) \tag{2.23}$$

and

$$I_\alpha(\Pi, W) = -[\int(\sum_{i=1}^{m} w_i p_i^\alpha(y))^{1/\alpha}\lambda(dy)]^\alpha \tag{2.24}$$

where $p_i(y)$ is the density of $\mathcal{P}_i$ with respect to some dominating measure $\lambda$. In fact, for the PD $\mathcal{Q}^*$ with density

$$q^*(y) = \frac{(\sum_{i=1}^{m} w_i p_i^\alpha(y))^{1/\alpha}}{\int(\sum_{i=1}^{m} w_i p_i^\alpha(y))^{1/\alpha}\lambda(dy)}, \tag{2.25}$$

$\sum_{i=1}^{m} w_i I_\alpha(\mathcal{P}_\alpha \| \mathcal{Q}^*)$ is equal to the right hand side of (2.24) while for an arbitrary PD $\mathcal{Q}$ on $Y$ with density $q(y)$ from (2.23) and (2.25) follows

$$\sum_{i=1}^{m} w_i I_\alpha(\mathcal{P}_i \| \mathcal{Q}) = [\int(\sum_{i=1}^{m} w_i p_i^\alpha(y))^{1/\alpha}\lambda(dy)]^\alpha \cdot I_\alpha(\mathcal{Q}^* \| \mathcal{Q}); \tag{2.26}$$

here the second factor of the right hand side is $\geq -1$, by (2.3). Note that (2.24) is equivalent with SIBSON's formula for "information radius of order $\alpha$", see [21]. The identity (2.26) appears, essentially, also in [1]. Of course, (2.24) can be interpreted also as the formula of $I_\alpha(\xi;\eta)$—the $\alpha$-informativity of a RV $\eta$ with respect to a RV $\xi$ with values in a finite set.

One obtains similar formulas also for $f(u) = u^\alpha$, $\alpha > 1$ or $\alpha < 0$. The case $\alpha = 2$ might be expected to be the most interesting since $I_2(\mathcal{P} \| \mathcal{Q}) = \int \frac{p^2(y)}{q(y)}\lambda(dy)$ is just the $\chi^2$-divergence of $\mathcal{P}$ and $\mathcal{Q}$, aside from the additive constant $f(1) = 1$.

The $\alpha$-informativity (2.24) is in a one-to-one functional relationship with Gallager's function $E_0(\varrho, W)$ (with $\alpha = 1/1 + \varrho$, see e. g. [9] p. 138 and p. 322) which plays a fundamental role in coding theory.

In information theory, the maximization of Shannon's mutual information and of Gallager's function is an important task. The relevant theorems (see e. g. [9], theorems 4.5.1 and 5.6.5) may be given the equivalent formulations that for $f(u) = u \log_2 u$ and $f(u) = -u^\alpha$ $(0 < \alpha < 1)$, respectively

$$\max_W I_f(\Pi, W) = \varrho_f(\Pi), \tag{2.27}$$

and for the maximizing prior PD $W^* = \{w_1^*, \ldots, w_m^*\}$ we have $I_f(\mathcal{P}_i \| \mathcal{Q}^*) \leq \varrho_f(\Pi)$, $1 \leq i \leq m$, with equality if $w_i^* > 0$. In particular, for $f(u) = u \log_2 u$, $\varrho_f(\Pi)$ equals the capacity of the OC $\Pi$.

In the next section we shall show that these results carry over to the general case almost completely, see Theorem 3.2 and its Corollary.

EXAMPLE 2.2. Let $X = Y = \{1, \ldots, m\}$ and let the OC $\Pi$ be symmetric in the sense that any permutation of the rows of the matrix $(p_{ik})$ (where $\mathcal{P}_i = \{p_{i1}, \ldots, p_{im}\}$) is equivalent to some permutation of its columns and conversely. Then $I_f(\Pi,W)$ is a symmetric concave function of $W$, thus it is maximized for $W^* = \left\{\frac{1}{m}, \ldots, \frac{1}{m}\right\}$; furthermore, $\sum_{i=1}^{m} w_i^* I_f(\mathcal{P}_i \| \mathcal{Q})$ is a symmetric convex function of $\mathcal{Q}$ minimized for $\mathcal{Q}^* = \left\{\frac{1}{m}, \ldots, \frac{1}{m}\right\}$. Thus in this particular case

$$\max_W I_f(\Pi, W) = \varrho_f(\Pi) = I_f(\mathcal{P}_i \| \mathcal{Q}^*) = \frac{1}{m} \sum_{k=1}^{m} f(mp_{ik}). \tag{2.28}$$

In information theory, the entropy of a RV $\xi$ is sometimes defined as the mutual information $I(\xi;\xi)$. This suggests the following

DEFINITION 2.5. The $f$-entropy of a RV $\xi$ with values in a finite set is defined as

$$H_f(\xi) = I_f(\xi;\xi). \tag{2.29}$$

Here we suppose that $f(0) < \infty$, because else $H_f(\xi)$ would be infinite for every non-trivial RV $\xi$.

Instead of the $f$-entropy of a RV, one can also speak of the $f$-entropy of a finite PD $W = \{w_1, \ldots, w_m\}$. Let $\Pi_0$ denote the OC $\{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$ from $\{1, \ldots, m\}$ to itself where $\mathcal{P}_i$ denotes the PD concentrated at $i$. Then one may write

$$H_f(W) = I_f(\Pi_0, W). \tag{2.30}$$

Obviously, $H_f(\xi) = H_f(W_\xi) \geq f(1)$; the inequality is strict unless $\xi$ is constant with probability 1.

PROPOSITION 2.3. *For any* RV'*s* $\xi$ *and* $\eta$ *with values in finite sets we have*

$$I_f(\xi;\eta) \leq H_f(\xi), \quad I_f(\xi;\eta) \leq H_f(\eta); \tag{2.31}$$

*moreover, the first inequality holds for an arbitrary* RV $\eta$, *as well. Furthermore, for any* $W = \{w_1, \ldots, w_m\}$

$$H_f(W) \leq H_f\left(\frac{1}{m}, \ldots, \frac{1}{m}\right) = \frac{1}{m} f(m) + \left(1 - \frac{1}{m}\right) f(0). \tag{2.32}$$

PROOF. (2.31) is an immediate consequence of Proposition 2.2 and Definition 2.5, while (2.32) is just the particular case $\Pi = \Pi_0$ of (2.28), see (2.30).

EXAMPLE 2.3. In the case $f(u) = -u^\alpha$ $(0 < \alpha < 1)$ we have, specializing (2.24) to $\Pi = \Pi_0$,

$$H_\alpha(W) = -\left(\sum_{k=1}^{m} w_k^{1/\alpha}\right)^\alpha. \tag{2.33}$$

It is interesting to note that while our $I_\alpha(\mathcal{P} \,||\, \mathcal{Q})$ — see (2.23) — corresponds to RÉNYI's information gain of order $\alpha$, the associated entropy (2.33) corresponds to the entropy of order $\frac{1}{\alpha}$ in the sense of RÉNYI (cf. [17]); here correspondence means a one-to-one functional relationship.

## § 3. The main theorems

We shall show that in the definitions of $f$-informativity with prior PD and of absolute $f$-informativity, see (2.9), (2.10), the inf can be replaced by min; if $f$ is strictly convex, the minimum is attained for a unique PD $\mathcal{Q}^*$ on $Y$. Moreover, the absolute $f$-informativity of an OC equals its maximum $f$-informativity with prior PD, for all possible prior PD's $W$.

A heuristic application of the Lagrange multipliers method suggests that $\sum_{i=1}^{m} w_i I_f(\mathcal{P}_i \,||\, \mathcal{Q}) = \sum_{i=1}^{m} w_i \int p_i(y) \tilde{f}\left(\frac{q(y)}{p_i(y)}\right) \lambda(dy)$ (see (2.5), (2.6)) is minimized for $\mathcal{Q} = \mathcal{Q}^*$ if

$$\sum_{i=1}^{m} w_i \tilde{f}'\left(\frac{q^*(y)}{p_i(y)}\right) = c. \tag{3.1}$$

If $\tilde{f}$ is not everywhere differentiable, one may guess that (3.1) should hold with $\leq$ for the left and with $\geq$ for the right derivatives.

The following Lemma asserts, essentially, that this condition can be fulfilled.

LEMMA 3.1. *Let $g(u)$, $u \in (0, \infty)$ be a non-decreasing left continuous function (not identically constant) and let*

$$\bar{g}(u) = \lim_{v \downarrow u} g(v) \qquad u \in (0, \infty) \tag{3.2}$$

*denote its right continuous pair. Let $\Pi = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$ be an* OC *from $X = \{1, \ldots, m\}$ to $Y$ and $W = \{w_1, \ldots, w_m\}$ a* PD *on $X$. Then—with the notation* (2.12)*—there exists a* PD *$\mathcal{Q}^*$ on $Y$ with density $q^*(y)$ satisfying* (i) *and* (ii) *below*[3] *provided in the case $g(\infty) = \infty$ that $\lambda(Y^{++}) > 0$:*

[3] Here, as well as in the sequel, all densities are understood with respect to a common dominating measure $\lambda$ and all statements concerning densities are meant $\lambda$-a.e.

(i) *$q^*(y)$ is positive on $Y^{++}$ and vanishes outside $Y^+$; if $g(\infty) = \infty$ then $q^*(y)$ vanishes outside $Y^{++}$ and if $g(\infty) < \infty$, $g(0) = -\infty$ then $q^*(y)$ is positive on $Y^+$;*

(ii) *there exists a (finite) constant c such that*

$$\sum_{i=1}^{m} w_i\, g\left(\frac{q^*(y)}{p_i(y)}\right) \leq c \leq \sum_{i=1}^{m} w_i\, \bar{g}\left(\frac{q^*(y)}{p_i(y)}\right) \quad \text{if} \quad q^*(y) > 0 \tag{3.3}$$

*and if $q^*(y) = 0$, the left hand side of (3.3) is $\geq c$ whenever it does not contain infinite terms of different signs.*

Here we understand

$$g(0) = \bar{g}(0) = \lim_{u \downarrow 0} g(u)\,, \quad g(\infty) = \bar{g}(\infty) = \lim_{u \uparrow \infty} g(u) \tag{3.4}$$

and

$$g\left(\frac{a}{0}\right) = \bar{g}\left(\frac{a}{0}\right) = g(\infty) \text{ for every } a \geq 0. \tag{3.5}$$

PROOF. Let $T$ denote the set of all $m$-tuples of non-negative or positive numbers $t_1, \ldots, t_m$ according as $g(\infty) < \infty$ or $g(\infty) = \infty$, respectively. Consider the functions of $c \in (-\infty, g(\infty))$

$$Z_{t_1,\ldots,t_m}(c) = \min\left\{u: \sum_{i=1}^{m} w_i\, \bar{g}\left(\frac{u}{t_i}\right) \geq c\right\}, \ (t_1, \ldots, t_m) \in T \tag{3.6}$$

(in view of the right continuity of $\bar{g}$, one may write min rather than inf). Then $Z_{t_1,\ldots,t_m}$ is a non-negative, finite valued, non-decreasing left continuous function of $c \in (-\infty, g(\infty))$ for any fixed $(t_1, \ldots, t_m) \in T$, and $z_{t_1,\ldots,t_m}(c) > 0$ iff

$$c > \lim_{u \downarrow 0} \sum_{i=1}^{m} w_i\, \bar{g}\left(\frac{u}{t_i}\right) = \begin{cases} \sum\limits_{i:\,t_i>0} w_i\, g(0) + \sum\limits_{i:\,t_i=0} w_i\, g(\infty) & \text{if } g(\infty) < \infty \\ g(0) & \text{if } g(\infty) = \infty. \end{cases} \tag{3.7}$$

Let us denote the limit in (3.7) by $a_{t_1,\ldots,t_m}$. Then the right continuous pair of $z_{t_1,\ldots,t_m}$ in the sense of (3.2) is given by

$$\bar{Z}_{t_1,\ldots,t_m}(c) = \begin{cases} \max\left\{u: \sum\limits_{i=1}^{m} w_i\, g\left(\frac{u}{t_i}\right) \leq c\right\} & \text{if } a_{t_1,\ldots,t_m} \leq c < g(\infty) \\ 0 & \text{if } c < a_{t_1,\ldots,t_m}. \end{cases} \tag{3.8}$$

In particular, if $\bar{Z}_{t_1,\ldots,t_m}(c) > 0$ then

$$\sum_{=1}^{m} w_i\, g\left(\frac{\bar{Z}_{t_1,\ldots,t_m}(c)}{t_i}\right) \leq c \leq \sum_{i=1}^{m} w_i\, \bar{g}\left(\frac{Z_{t_1,\ldots,t_m}(c)}{t_i}\right). \tag{3.9}$$

Now define for $(p_1(y), \ldots, p_m(y)) \in T$ and $c < g(\infty)$

$$q_c(y) = Z_{p_1(y), \ldots, p_m(y)}(c) \tag{3.10}$$
$$\bar{q}_c(y) = \bar{Z}_{p_1(y), \ldots, p_m(y)}(c)$$

while for $(p_1(y), \ldots, p_m(y)) \notin T$ we set $q_c(y) = \bar{q}_c(y) = 0$.

Then, since $\sum_{i=1}^{m} w_i\, g\left(\frac{u}{t_i}\right) \geq g\left(\frac{u}{\max\limits_{1\leq i\leq m} t_i}\right)$, the first inequality of (3.9) implies

$$\bar{q}_c(y) \leq K \max_{1\leq i\leq m} p_i(y) \tag{3.11}$$

for some (finite) constant $K$; thus $\bar{q}_c(y)$ and even more $q_c(y) \leq \bar{q}_c(y)$ are integrable for all $c < g(\infty)$.

By dominated convergence follows that $\int q_c(y)\,\lambda(dy)$ is a left continuous function of $c \in (-\infty, g(\infty))$ and $\int \bar{q}_c(y)\,\lambda(dy)$ is its right continuous pair. Hence follows, under the condition

$$\int q_c(y)\,\lambda(dy) > 1 \text{ for some } c < g(\infty) \tag{3.12}$$

the existence of $c_1 \in [g(0), g(\infty))$ (or $c_1 \in (-\infty, g(\infty))$ if $g(0) = -\infty$) with

$$\int q_{c_1}(y)\,\lambda(dy) \leq 1 \leq \int \bar{q}_{c_1}(y)\,\lambda(dy). \tag{3.13}$$

In fact, (3.7) and (3.10) imply $q_{g(0)}(y) = 0$ for all $y \in Y$ if $g(0) > -\infty$ while $\lim\limits_{c \downarrow -\infty} q_c(y) = 0$ in all cases, see (3.6), (3.10).

From (3.13) one concludes that for some $d \in [0, 1]$

$$q^*(y) = d q_{c_1}(y) + (1 - d)\,\bar{q}_{c_1}(y) \tag{3.14}$$

is the density of a PD on $Y$. Since $q_{c_1}(y) \leq q^*(y) \leq \bar{q}_{c_1}(y)$, $q^*(y)$ satisfies (3.3) with $c = c_1$. If $q^*(y) = 0$ then $q_{c_1}(y) = 0$, i.e. either $(p_1(y), \ldots, p_m(y)) \notin T$ or else (3.7) does not hold for $c = c_1$, $t_i = p_i(y)$, $i = 1, \ldots, m$. This just means, in view of $g(\infty) > c_1$ and the convention (3.5) that the second statement in (ii) of the Lemma is also satisfied.

Furthermore, if $y \notin Y^+$ or $g(\infty) = \infty$ and $y \in Y^{++}$ (see (2.12)) then $(p_1(y), \ldots, p_m(y)) \notin T$ thus $q_c(y) = 0$ for all $c < g(\infty)$, while from (3.7) and (3.10) follows $q_c(y) > 0$ for all $c \in (g(0), g(\infty))$ if $y \in Y^{++}$ and in the case $g(\infty) < \infty$, $g(0) = -\infty$ also if $y \in Y^+$. This shows that $q^*(y)$ defined by (3.14) satisfies also the statements (i) of the Lemma except, conceivably, if $c_1$ (determined by (3.13)) equals $g(0) > -\infty$. The latter case, however, can obtain only if $\bar{z}_{t_1, \ldots, t_m}(g(0)) > 0$ for some $(t_1, \ldots, t_m)$, i.e. (see (3.8)) if $g(u)$ is constant in some neighbourhood of 0; in that case, (3.8), (3.10) and (2.12) imply $\bar{q}_{g(0)}(y) > 0$ for all $y \in Y^{++}$, thus, on account of (3.14) $q^*(y) > 0$ on $Y^{++}$ also if $c_1 = g(0)$.

To complete the proof, the fulfillment of condition (3.12) has to be examined. If $g(u)$ is not constant in any neighbourhood of infinity, (3.6) implies $\lim_{c \uparrow g(\infty)} z_{t_1,\dots,t_m}(c) = \infty$ for all $(t_1, \dots, t_m) \in T$. In view of (3.10), we thus have $\lim_{c \uparrow g(\infty)} q_c(y) = \infty$ on $Y^{++}$ and in the case $g(\infty) < \infty$ also on $Y^+$; therefore, by monotone convergence, (3.12) holds (recall that in the case $g(\infty) = \infty$ we have assumed $\lambda(Y^{++}) > 0$). If $g(u)$ is constant in some neighbourhood of infinity, set $u_0 = \inf\{u : g(u) = g(\infty)\}$. Then (3.6) yields

$$\lim_{c \uparrow g(\infty)} Z_{t_1,\dots,t_m}(c) = \min\left\{u : \sum_{i=1}^{m} w_i \bar{g}\left(\frac{u}{t_i}\right) = g(\infty)\right\} = u_0 \cdot \max_{1 \leq i \leq m} t_i$$

and we arrive at $\lim_{c \uparrow g(\infty)} \int q_c(y)\,\lambda(dy) = u_0 \cdot \int \max_{1 \leq i \leq m} p_i(y)\,\lambda(dy)$.

This limit may happen to be $\leq 1$, but in that case one can choose $d \geq u_0$ such that the integral of $q^*(y) = d \cdot \max_{1 \leq i \leq m} p_i(y)$ equals 1; then this $q^*(y)$ trivially satisfies (i) and (ii), with $c = g(\infty)$.

THEOREM 3.1. *For an arbitrary convex function $f(u)$, OC $\Pi = \{\mathcal{P}_1, \dots, \mathcal{P}_m\}$ from $X = \{1, \dots, m\}$ to $Y$ with $\varrho_f(\Pi) < \infty$ and prior PD $W = \{w_1, \dots, w_m\}$ on $X$, there exists a PD $\mathcal{Q}^*$ on $Y$ minimizing $\sum_{i=1}^{m} w_i I_i(\mathcal{P} \| \mathcal{Q})$; if $w_i > 0$, $i = 1, \dots, m$, such a minimizing PD is any $\mathcal{Q}^*$ with the properties* (i), (ii) *in Lemma* 3.1, *the left derivative of $\tilde{f}$ playing the role of $g$.*

*If the function $f(u)$ is strictly convex, the minimizing PD $\mathcal{Q}^*$ is unique, moreover, there exists a function $\psi(u)$ (depending on $f$, $\Pi$ and $W$) with $\lim_{u \downarrow I_f(\Pi, W)} \psi(u) = 0$ such that for any PD $\mathcal{Q}$ on $Y$*

$$|\mathcal{Q} - \mathcal{Q}^*| \leq \psi\Big(\sum_{i=1}^{m} w_i I_f(\mathcal{P}_i \| \mathcal{Q})\Big). \tag{3.15}$$

PROOF. If $g(u)$ denotes the left derivative of the convex function $\tilde{f}(u) = uf\left(\frac{1}{u}\right)$, we have $g(\infty) = \lim_{u \uparrow \infty} \frac{\tilde{f}(u)}{u} = f(0)$. Thus, if $g(\infty) = \infty$, the condition $\lambda(Y^{++}) > 0$ of Lemma 3.1 is implied by the assumption $\varrho_f(\Pi) < \infty$, c. f. the paragraph after (2.12).

Let $\mathcal{Q}$ and $\mathcal{Q}^*$ be two PD's on $Y$ where $\mathcal{Q}^*$ satisfies (i) and (ii) of Lemma 3.1 and $\max_{1 \leq i \leq m} I_f(\mathcal{P}_i \| \mathcal{Q}) < \infty$; then, if $f(0) = g(\infty) = \infty$, both $q^*(y)$ and the density $q(y)$ of $\mathcal{Q}$ vanish outside $Y^{++}$ (see footnote[3]).

Since $\tilde{f}$ is convex, we have

$$\tilde{f}(u) \geq \tilde{f}(u_0) + \hat{g}(u_0)(u - u_0) \qquad (u, u_0 \in (0, \infty)) \tag{3.16}$$

for any function $\hat{g}(u)$ with $g(u) \leq \hat{g}(u) \leq \bar{g}(u)$; of course, $\hat{g}(u) = \tilde{f}'(u)$ at each point $u$ where $\tilde{f}(u)$ is differentiable.

From (3.16) follows

$$r\tilde{f}\left(\frac{s}{r}\right) \geq r\tilde{f}\left(\frac{t}{r}\right) + \hat{g}\left(\frac{t}{r}\right)(s-t) \tag{3.17}$$

for any positive (finite) numbers $r, s, t$; using the conventions (2.2), (3.4) and (3.5), inequality (3.17) is easily checked to hold also if one, two or all three of the numbers $r, s, t$, are 0.

Substituting $r = p_i(y)$, $s = q(y)$, $t = q^*(y)$, multiplying by $w_i$ and summing for $i = 1, \ldots, m$:

$$\begin{aligned} \sum_{i=1}^{m} w_i p_i(y) \tilde{f}\left(\frac{q(y)}{p_i(y)}\right) &\geq \sum_{i=1}^{m} w_i p_i(y) \tilde{f}\left(\frac{q^*(y)}{p_i(y)}\right) + \\ &+ \sum_{i=1}^{m} w_i \hat{g}_y\left(\frac{q^*(y)}{p_i(y)}\right)\big(q(y) - q^*(y)\big). \end{aligned} \tag{3.18}$$

Since (3.17) is valid for any version of $\hat{g}$ within the specified range, for different $y$'s different versions may be taken, as indicated in (3.18). In particular, in force of (3.3), one may choose $\hat{g}_y$ such that

$$\sum_{i=1}^{m} w_i \hat{g}_y\left(\frac{q^*(y)}{p_i(y)}\right) = c \text{ if } q^*(y) > 0. \tag{3.19}$$

Let us remark that the right hand side of (3.18) contains no term $-\infty$ (see footnote [3]). In fact, $\hat{g}_y\left(\frac{q^*(y)}{p_i(y)}\right) = \infty$ for some $1 \leq i \leq m$ iff $g(\infty) = \infty$, $y \notin Y^{++}$, in which case, however, $q^*(y) = 0$; on the other hand, $\hat{g}_y\left(\frac{q^*(y)}{p_i(y)}\right) = -\infty$ for some $1 \leq i \leq m$ iff $g(0) = -\infty$, $q^*(y) = 0$, $y \in Y^+$, possible only if $g(\infty) = \infty$, $y \in Y^+ \quad Y^{++}$ (by Lemma 3.1, (i)), in which case, however, $q(y) = 0$, too.

In force of (3.19) and the second statement of Lemma 3.1, (ii), the integral of the second sum on the right hand side of (3.18) is $\geq 0$, thus from (3.18) and (2.6) follows

$$\sum_{i=1}^{m} w_i I_f(\mathcal{P}_i \,||\, \mathcal{Q}) \geq \sum_{i=1}^{m} w_i I_f(\mathcal{P}_i \,||\, \mathcal{Q}^*), \tag{3.20}$$

for any PD $\mathcal{Q}$ on $Y$ with $\max\limits_{1 \leq i \leq m} I_f(\mathcal{P}_i \,||\, \mathcal{Q}) < \infty$. This proves, in particular, that in the case $w_i > 0$, $i = 1, \ldots, m$ the PD $\mathcal{Q}^*$ minimizes $\sum\limits_{i=1}^{m} w_i I_f(\mathcal{P}_i \,||\, \mathcal{Q})$. If, however, $w_i = 0$ for some $i \in \{1, \ldots, m\}$, the $\mathcal{P}_i$'s with $w_i = 0$ may be omitted and the existence of a minimizing $\mathcal{Q}^*$ still follows (although this $\mathcal{Q}^*$ need not have the properties stated in Lemma 3.1 for the original $\mathcal{P}_i$'s).

If $f$ is strictly convex so is $\tilde{f}$, too, and

$$\alpha_\delta(u_0) = \min\{\tilde{f}(u_0+\delta) - \tilde{f}(u_0) - \delta\,\hat{g}(u_0), \tilde{f}(u_0) - \delta\,\hat{g}(u_0) - \tilde{f}(u_0-\delta)\} > 0 \tag{3.21}$$

for every $u_0 \geq \delta > 0$. Since $\alpha_\delta(u_0)$ is a lower semi-continuous function of $u_0$ (continuous at the points where $\tilde{f}$ is differentiable) we have for every $K > \delta > 0$

$$\varepsilon_{\delta,K} = \frac{1}{\delta} \min_{\delta \leq u_0 \leq K} \alpha_\delta(u_0) > 0. \tag{3.22}$$

From the convexity of $\tilde{f}$ follows that for $\delta \leq u_0 \leq K$, $|u - u_0| \geq \delta$ the term $\varepsilon_{\delta,K}|u - u_0|$ may be added to the right hand side of (3.16). Thus instead of (3.20) one arrives at

$$\begin{aligned}\sum_{i=1}^m w_i I_f(\mathcal{P}_i \| \mathcal{Q}) &\geq \sum_{i=1}^m w_i I_f(\mathcal{P}_i \| \mathcal{Q}^*) + \\ &+ \varepsilon_{\delta,K} \cdot \sum_{i=1}^m w_i \int\limits_{\substack{\delta p_i(y) \leq q^*(y) \leq K p_i(y) \\ |q(y)-q^*(y)| \geq \delta p_i(y)}} |q(y) - q^*(y)|\,\lambda(dy).\end{aligned} \tag{3.23}$$

Taking into account that $\int\limits_{|q(y)-q^*(y)|<\delta p_i(y)} |q(y) - q^*(y)|\,\lambda(dy) < \delta$ and denoting by $w_0$ the smallest $w_i > 0$, the second sum on the right-hand side of (3.23) is bounded from above by

$$(1-\delta)\,w_0 \cdot \int\limits_{\delta \min_{1\leq i\leq m} p_i(y) \leq q(y) \leq K \max_{1\leq i\leq m} p_i(y)} |q(y) - q^*(y)|\,\lambda(dy). \tag{3.24}$$

But since $q^*(y)$ satisfies (3.3) (the left derivative of $\tilde{f}$ playing the role of $g$), there exist $0 < \delta < K$ such that

$$\delta \min_{1\leq i\leq m} p_i(y) \leq q^*(y) \leq K \max_{1\leq i\leq m} p_i(y) \quad \text{for a.e. } y \in Y.$$

Then the integral in (3.24) equals $|\mathcal{Q} - \mathcal{Q}^*|$ and thus (3.23) and (3.24) give rise to

$$|\mathcal{Q} - \mathcal{Q}^*| \leq \frac{1}{\varepsilon_{\delta,K}(1-\delta)\,w_0} \left(\sum_{i=1}^m w_i I_f(\mathcal{P}_i \| \mathcal{Q}) - \sum_{i=1}^m w_i I_f(\mathcal{P}_i \| \mathcal{Q}^*)\right). \tag{3.25}$$

This completes the proof for the case $w_i > 0$ $i = 1, \ldots, m$, while in the opposite case the $\mathcal{P}_i$'s with $w_i = 0$ may be omitted and the statement still follows.

It will be convenient to introduce the notation

$$I_f'(\Pi, W) = \inf_{\mathcal{Q} \in \mathfrak{M}} \sum_{i=1}^m w_i I_f(\mathcal{P}_i \| \mathcal{Q}) \tag{3.26}$$

where $\mathfrak{M}$ denotes the class of PD's $\mathcal{Q}$ on $Y$ with

$$I_f(\mathcal{P}_i \| \mathcal{Q}) < \infty, \quad i = 1, \ldots, m.$$

LEMMA 3.2. *$I'_f(\Pi, W)$ is a continuous concave function of the* PD *$W$, for any convex function $f$ and* OC *$\Pi = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$, and $I'_f(\Pi, W) \geqq I_f(\Pi, W)$ with equality if $w_i > 0$, $i = 1, \ldots, m$. Furthermore, either of the conditions*

(i) $f(0) < \infty$

(ii) $\tilde{f}(0) < \infty$; *the $\mathcal{P}_i$'s are mutually equivalent measures*

(iii) *there exists a constant $K$ such that*

$$\max_{1 \leqq i \leqq m} p_i(y) \leqq K \min_{1 \leqq i \leqq m} p_i(q) \qquad \lambda - \text{a.e.}$$

*implies $I'_f(\Pi, W) = I_f(\Pi, W)$ for all* PD*'s $W = \{w_1, \ldots, w_m\}$.*

PROOF. $I'_f(\Pi, W)$ is the infimum of linear functions of $W$, hence it is a continuous concave function. The second statement is obvious, while the last one follows from the possibility of substituting any PD $\mathcal{Q}$ on $Y$ satisfying $I_f(\mathcal{P}_i \| \mathcal{Q}) < \infty$ for some $i \in \{1, \ldots, m\}$ by a PD $\mathcal{Q}_1$ so that $I_f(\mathcal{P}_i \| \mathcal{Q}_1) < I_f(\mathcal{P}_i \| \mathcal{Q}) + \varepsilon$ for each $i \in \{1, \ldots, m\}$ with $I_f(\mathcal{P}_i \| \mathcal{Q}) < \infty$ (where $\varepsilon > 0$ is arbitrary) in such a way that $\mathcal{Q}_1$ have a density $q_1(y)$ satisfying $q_1(y) \geqq K_1 \max_{1 \leqq i \leqq m} p_i(y)$ if $f(0) < \infty$ and $q_1(y) \leqq K_2 \min_{1 \leqq i \leqq m} p_i(y)$ or $K_1 \max_{1 \leqq i \leqq m} p_i(y) \leqq q_1(y) \leqq K_2 \min_{1 \leqq i \leqq m} p_i(y)$, respectively, if $f(0) = \infty$ but condition (ii) or (iii) is valid,[4] ensuring in all three cases, $\mathcal{Q}_1 \in \mathfrak{M}$.

THEOREM 3.2. *For an arbitrary convex function $f$ and* OC *$\Pi = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$ from $X = \{1, \ldots, m\}$ to $Y$ we have*

$$\varrho_f(\Pi) = \sup_W I_f(\Pi, W). \tag{3.27}$$

*If $\varrho_f(\Pi) < \infty$ and $f$ is strictly convex, the necessary and sufficient condition for a* PD *$W^* = \{w_1^*, \ldots, w_m^*\}$ on $X$ to maximize $I_f(\Pi, W)$ consists in*

$$I_f(\mathcal{P}_i \| \mathcal{Q}^*) \leqq K, \qquad i = 1, \ldots, m, \tag{3.28}$$

*with equality whenever $w_i^* > 0$, where $\mathcal{Q}^*$ denotes the unique* PD *on $Y$ maximizing $\sum_{i=1}^m w_i^* I_f(\mathcal{P}_i \| \mathcal{Q})$ and $K$ is a constant. Moreover, in that case $K = \varrho_f(\Pi)$, and $\mathcal{Q}^*$ is the unique* PD *on $Y$ minimizing $\max_{1 \leqq i \leqq m} I_f(\mathcal{P}_i \| \mathcal{Q})$.*

COROLLARY. If $f$ is strictly convex and either of the conditions (i)—(iii) of Lemma 3.2 is fulfilled, there exists a unique PD $\mathcal{Q}^*$ on $Y$ minimizing $\max_{1 \leqq i \leqq m} I_f(\mathcal{P}_i \| \mathcal{Q})$ and it corresponds to a PD $W^*$ on $X = \{1, \ldots, m\}$ maximizing $I_f(\Pi, W)$.

[4] Note that in this case $\mathcal{Q}$ must be absolutely continuous with respect to the $\mathcal{P}_i$'s.

REMARK 3.1. Since $\varrho_f(\Pi) = \inf_{\mathcal{Q}} \max_{1 \leq i \leq m} I_f(\mathcal{P}_i \,||\, \mathcal{Q})$ is interpreted as the "radius" of the set of PD's $\Pi = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$, the above $\mathcal{Q}^*$ may be given the geometric interpretation of the "centre" of the set $\Pi$. By another method it is possible to show the existence of such a "centre" $\mathcal{Q}^*$ even if there is no $W^*$ maximizing $I_f(\Pi, W)$ but in that case $\mathcal{Q}^*$ does not correspond to any PD $W$ on $X$ in the sense of Theorem 3.1 (see Example 3.1 below). The relevance of the "centre" of $\Pi$ for coding theory (for $f(u) = -u^\alpha$, $0 < \alpha < 1$) is most clearly seen from [1].

PROOF. Consider the zero-sum two-person game where the sets of pure strategies of the first and second player are $X$ and $\mathfrak{M}$ respectively, where $\mathfrak{M}$ is the same as in (3.26), and the payoff function is $I_f(\mathcal{P}_i \,||\, \mathcal{Q})$. Since $X$ is a finite set and the payoff function is convex, (3.27) follows from a variant of the minimax theorem. A direct proof, along the lines of [4], pp. 43—50 is as follows:

Let $S$ be the set of all points in $m$-space of form $(I_f(\mathcal{P}_1 \,||\, \mathcal{Q}), \ldots, I_f(\mathcal{P}_m \,||\, \mathcal{Q}))$ with $\mathcal{Q} \in \mathfrak{M}$. Let $S^*$ be the convex hull of $S$ and let $T$ denote the set of all points $(t_1, \ldots, t_m)$ with $t_i < \varrho_f(\Pi)$, $i = 1, \ldots, m$. Since $I_f(\mathcal{P} \,||\, \mathcal{Q})$ is convex in $\mathcal{Q}$, we have $\sum_{k=1}^{n} \lambda_k I_f(\mathcal{P}_i \,||\, \mathcal{Q}_k) \geqq I_f(\mathcal{P}_i \,||\, \sum_{k=1}^{n} \lambda_k \mathcal{Q}_k)$, $i = 1, \ldots, m$ for $\lambda_k \geqq 0$, $k = 1, \ldots, n$, $\sum_{k=1}^{n} \lambda_k = 1$, thus $S^* \cap T \neq \emptyset$ would imply $S \cap T \neq \emptyset$, a contradiction (see (2.10)). $S^*$ and $T$, being disjoint convex sets in $m$-space, have a separating hyperplane (cf. e. g. [4], p. 35). In view of the definition of $T$, the existence of a hyperplane separating $S$ and $T$ means just the existence of a PD $W^* = \{w_1^*, \ldots, w_m^*\}$ on $X$ such that $\sum_{i=1}^{m} w_i^* I_f(\mathcal{P}_i \,||\, \mathcal{Q}) \geqq \varrho_f(\Pi)$ for all $\mathcal{Q} \in \mathfrak{M}$, i.e., see (3.26),

$$I_f'(\Pi, W^*) \geqq \varrho_f(\Pi). \tag{3.29}$$

But from Lemma 3.2 follows $\sup_W I_f'(\Pi, W) = \sup_W I_f(\Pi, W)$ thus, on account of (2.11), from (3.29) one concludes

$$\sup_W I_f(\Pi, W) = I_f'(\Pi, W^*) = \varrho_f(\Pi). \tag{3.30}$$

To prove the second statement of Theorem 3.2, note first that (3.28) implies $K \geq \varrho_f(\Pi)$, see (2.10), while the additional assumption $I_f(\mathcal{P} \,||\, \mathcal{Q}) = K$ for $w_i^* > 0$ and (2.11) imply $K = I_f(\Pi, W^*) \leq \varrho_f(\Pi)$. Thus the sufficiency part and $K = \varrho_f(\Pi)$ are proved.

Suppose now that $W^* = \{w_1^*, \ldots, w_m^*\}$ maximizes $I_f(\Pi, W)$; then by (3.27) $I_f(\Pi, W^*) = \varrho_f(\Pi)$. Consider a sequence $\mathcal{Q}_1, \mathcal{Q}_2, \ldots$ of PD's on $Y$ such that

$$\max_{1 \leq i \leq m} I_f(\mathcal{P}_i \,||\, \mathcal{Q}_n) \to \varrho_f(\Pi). \tag{3.31}$$

Then

$$\lim_{n\to\infty} \sum_{i=1}^{m} w_i^* I_f(\mathcal{P}_i \,||\, \mathcal{Q}_n) = \varrho_f(\Pi) = I_f(\Pi, W^*). \tag{3.32}$$

If $f$ is strictly convex, there exists a unique PD $\mathcal{Q}^*$ on $Y$ such that $\sum_{i=1}^{m} w_i^* I_f(\mathcal{P} \,||\, \mathcal{Q}^*) = I_f(\Pi, W^*) = \varrho_f(\Pi)$, and (3.32) implies

$$\lim_{n\to\infty} |\mathcal{Q}_n - \mathcal{Q}^*| = 0 \tag{3.33}$$

according to Theorem 3.1. We claim that (3.33) implies

$$\varliminf_{n\to\infty} I_f(\mathcal{P}_i \,||\, \mathcal{Q}_n) \geqq I_f(\mathcal{P}_i \,||\, \mathcal{Q}^*), \qquad i = 1, \ldots, m. \tag{3.34}$$

In fact, let $\lambda$ be a common dominating measure of the considered PD's; then (3.33) means that the density $q^*(y)$ of $\mathcal{Q}_n$ converges to $q^*(y)$ in $L^1(\lambda)$. Now (3.34) follows from the Fatou lemma provided that $f$ is non-negative (note that $q_n(y) \xrightarrow{\lambda} q^*(y)$ implies $q_n(y) f\left(\frac{p_i(y)}{q_n(y)}\right) \xrightarrow{\lambda} q^*(y) f\left(\frac{p_i(y)}{q^*(y)}\right)$ even in the case $f(0) = \infty$, since then $q_n(y) = 0$ whenever $p_i(y) = 0$, being $\varrho_f(\Pi) < \infty$). But this can always be achieved by adding $a + bu$ to $f(u)$, which amounts to adding the constant $a + b$ to $I_f(\mathcal{P} \,||\, \mathcal{Q})$ thus does not affect the validity of (3.34).

(3.31) and (3.34) prove the necessity of (3.28) (with $K = \varrho_f(\Pi)$) and since $\sum_{i=1}^{m} w_i^* I_f(\mathcal{P}_i \,||\, \mathcal{Q}^*) = \varrho_f(\Pi)$, in (3.28) the equality must hold whenever $w_i^* > 0$. In particular, we have also shown that if $I_f(\Pi, W)$ is maximized for some PD $W^*$ on $X = \{1, \ldots, m\}$, the corresponding PD $\mathcal{Q}^*$ on $Y$ minimize $\max_{1 \leqq i \leqq m} I_f(\mathcal{P}_i \,||\, \mathcal{Q})$. The uniqueness of the minimizing $\mathcal{Q}^*$ follows from the strict convexity of $I_f(\mathcal{P}_i \,||\, \mathcal{Q})$ in $\mathcal{Q}$, implied by the strict convexity of $f$.

The Corollary follows from the fact that under either of the conditions (i)–(iii) of Lemma 3.2, the function $I_f(\Pi, W)$ is continuous, thus it is maximized for some $W^*$.

Note that while there may exist several PD's $W^*$ on $X$ maximizing $I_f(\Pi, W)$, our result implies that for any of them, $\sum_{i=1}^{m} w_i^* I_f(\mathcal{P}_i \,||\, \mathcal{Q})$ is minimized for the same $\mathcal{Q}^*$.

In order to strengthen Proposition 2.1 to a characterization of sufficiency of indirect observations with respect to an OC $\Pi = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$ let us formulate two conditions, one for the function $f$ and one for the OC and the indirect observation.

(A) $f(0) < \infty$ and $\lim_{u \uparrow \infty} (f(u) - uf'(u)) = -\infty$;

(B) The PD's $\bar{\mathcal{P}}_1, \dots, \bar{\mathcal{P}}_m$ on $Z$ obtained by the indirect observation are mutually equivalent measures.

THEOREM 3.3 *Let $f(u)$ be a strictly convex function and let an indirect observation*[5] *from $Y$ to $Z$ change $\Pi = \{\mathcal{P}_1, \dots, \mathcal{P}_m\}$ to $\bar{\Pi} = \{\bar{\mathcal{P}}_1, \dots, \bar{\mathcal{P}}_m\}$. Then, if $I_f(\bar{\Pi}, W) < \infty$ and either of the conditions* (A) *and* (B) *is fulfilled, the indirect observation is sufficient with respect to the* OC $\Pi$ *iff for some prior* PD $W = \{w_1, \dots, w_m\}$ *with $w_i > 0$, $i = 1, \dots, m$,*

$$I_f(\bar{\Pi}, W) = I_f(\Pi, W). \tag{3.35}$$

Note that $I_f(\bar{\Pi}, W) < \infty$ is certainly true under (A), and also under (B) if $\tilde{f}(0) < \infty$, cf. the paragraph after (2.12).

PROOF. On account of the Corollary of Proposition 2.1, we have to prove only that (3.35) implies sufficiency.

Let $\mathcal{Q}^*$ be the (unique) PD on $Y$ with

$$\sum_{i=1}^{m} w_i I_f(\mathcal{P}_i \| \mathcal{Q}^*) = I_f(\Pi, W), \tag{3.36}$$

existing by Theorem 3.1.

From (3.35), (3.36) and Definition 2.4 follows that the inequalities (see (2.7))

$$I_f(\bar{\mathcal{P}}_i \| \bar{\mathcal{Q}}^*) \leqq I_f(\mathcal{P}_i \| \mathcal{Q}^*) \tag{3.37}$$

hold with the equality sign (when considering an indirect observation of type (iv), it should be noted that $\mathcal{Q}^* \ll \mathcal{P}_1 + \dots + \mathcal{P}_m$ is admissible if the $\mathcal{P}_i$'s are). This means that the indirect observation is sufficient with respect to each pair $\{\mathcal{P}_i, \mathcal{Q}^*\}$, $i = 1, \dots, m$, i.e., there exist $Z$-measurable functions $S_z^i(B)$, $i = 1, \dots, m$, satisfying

$$\int_C S_z^i(B)\, \bar{\mathcal{P}}(dz) = \mathcal{P}_i^\star(B \times C) \qquad (B \in \mathcal{Y}, C \in \mathcal{Z}) \tag{3.38}$$

and

$$\int_C S_z^i(B)\, \bar{\mathcal{Q}}^*(dz) = \mathcal{Q}^{*\star}(B \times C) \qquad (B \in \mathcal{Y}, C \in \mathcal{Z}) \tag{3.39}$$

(for indirect observations of type (i) or (ii), the right hand sides become $\mathcal{P}_i(B \cap C)$ and $\mathcal{Q}^*(B \cap C)$ or $\mathcal{P}_i(B \cap T^{-1}C)$ and $\mathcal{Q}^*(B \cap T^{-1}C)$, respectively). Since (3.39) implies $S_z^i(B) = S_z^j(B)$ $\bar{\mathcal{Q}}^*$ — a. e. for every $i, j \in \{1, \dots, m\}$,

[5] Of either type (i)—(iv) described in § 2, from (2.7) until (2.8).

the possibility of dropping the upper index $i$ in (3.38) i.e. the sufficiency of the indirect observation with respect to $\Pi$ follows if

$$\bar{\mathcal{P}}_i \ll \bar{\mathcal{Q}}^*, \quad i = 1, \ldots, m. \tag{3.40}$$

Note that (3.35), (3.36) and (3.37)—the latter with equalities—imply that $\bar{\mathcal{Q}}^*$ minimizes $\sum_{i=1}^{m} w_i I_f(\bar{\mathcal{P}}_i \,||\, \mathcal{Q})$ (where $\mathcal{Q}$ ranges over the PD's on $Z$). Thus, by Theorem 3.1, $\bar{\mathcal{Q}}^*$ is the (unique) PD corresponding to the $\bar{\mathcal{P}}_i$'s in the sense of Lemma 3.1, applied for $Z$ instead of $Y$. Since condition (A) is equivalent to $g(\infty) < \infty$, $g(0) = -\infty$, where $g(u)$ denotes the left derivative of $\tilde{f}(u) = uf\left(\frac{1}{u}\right)$, both from this condition and from (B) follows (3.40), in force of Lemma 3.1, (i). The proof is complete.

REMARK 3.2. If neither (A) nor (B) is fulfilled, (3.35) does not necessarily imply sufficiency. In fact, if $f(0) = \infty$ (as e. g. in Example 3.1 below) then $I_f(\Pi, W)$ is determined by the restrictions of the densities $p_i(y)$ to $Y^{++}$. Thus if $T$ is a mapping of $Y$ into itself with $Ty = y$ for $y \in Y^{++}$ and $Ty = y_0 \notin Y^{++}$ for all $y \notin Y^{++}$, the indirect observation of type (ii) defined by $T$ satisfies (3.35) while it is obviously not sufficient provided that there are at least two $i$'s with $\mathcal{P}_i(Y^{++}) < 1$.

EXAMPLE 3.1. For $f(u) = -\log_2 u$, $\tilde{f}(u) = u \log_2 u$ we have $I_f(\mathcal{P}_i \,||\, \mathcal{Q}) = I(\mathcal{Q} \,||\, \mathcal{P}_i)$, see (2.14), thus $I_f(\Pi, W) = \inf_{\mathcal{Q}} \sum_{i=1}^{m} w_i I(\mathcal{Q} \,||\, \mathcal{P}_i)$. By Theorem 3.1, there exists a unique minimizing PD $\mathcal{Q}^*$ (provided that $\lambda(Y^{++}) > 0$) with density $q^*(y)$ satisfying (3.1) for $y \in Y^{++}$ and vanishing for $y \notin Y^{++}$, see (2.12). Since $\tilde{f}'(u) = \log_2 u + \log_2 e$, (3.1) reduces to $\sum_{i=1}^{m} w_i \log_2 \frac{q^*(y)}{p_i(y)} = c_1$ whence

$$q^*(y) = c_2 \prod_{i=1}^{m} p_i^{w_i}(y). \tag{3.41}$$

Substituting $\mathcal{Q} = \mathcal{Q}^*$ into $\sum_{i=1}^{m} w_i I(\mathcal{Q} \,||\, \mathcal{P}_i)$, we eventually obtain

$$I_f(\Pi, W) = -\log_2 \int \prod_{i=1}^{m} p_i^{w_i}(y)\, \lambda(dy). \tag{3.42}$$

This can be interpreted also as a formula for $I_f(\xi\,;\eta)$, see Definition 2.5. As $f(0) = \infty$, the $f$-entropy is not defined.

The $f$-radius $\varrho_f(\Pi) = \inf_{\mathcal{Q}} \max_{1 \le i \le m} I_f(\mathcal{P}_i \,||\, \mathcal{Q}) = \inf_{\mathcal{Q}} \max_{1 \le i \le m} I(\mathcal{Q} \,||\, \mathcal{P}_i)$ and the minimizing PD $\mathcal{Q}^*$ can be obtained in principle (see Theorem 3.2) by maximizing (3.42) for $W$ and substituting the resulting $W^*$ (provided that it exists) into (3.41). Of course, explicit formulas are not to be hoped for, except in very special cases. Note that for $\Pi = \{\mathcal{P}_1, \mathcal{P}_2\}$ we obtain

$$\varrho_f\{\mathcal{P}_1, \mathcal{P}_2\} = \sup_{0 \leq \alpha \leq 1} \left\{ -\log_2 \int p_1^{\alpha}(y) p_2^{1-\alpha}(y)\, \lambda(dy) \right\} \tag{3.43}$$

i.e. the Chernoff information number, see [6].

It is interesting to note that in this example $\max\limits_{W} I_f(\Pi, W)$ need not exist; e. g. if $\Pi = \{\mathcal{P}_1, \mathcal{P}_2\}$ where $\mathcal{P}_i$ is the uniform distribution on the interval $[0, i]$, $i = 1, 2$, in (3.43) the sup cannot be replaced by max. In that particular case $\max\limits_{1 \leq i \leq 2} I_f(\mathcal{P}_i \,\|\, \mathcal{Q})$ is minimized for $\mathcal{Q}^* = \mathcal{P}_1$, but there exist no PD $W = \{w_1, w_2\}$ for which $\sum\limits_{i=1}^{2} w_i I_f(\mathcal{P}_i \,\|\, \mathcal{Q})$ is minimized for $\mathcal{Q}^* = \mathcal{P}_1$.

## REFERENCES

[1] U. Augustin, Error estimates for low rate codes, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** (1969), 61—88.
[2] D. Blackwell, Comparison of experiments, *Proc. 2nd Berkeley Sympos. Math. Statist. and Probability*, Vol. 1, Berkeley, 1951, 93—102.
[3] D. Blackwell, Equivalent comparison of experiments, *Ann. Math. Statist.* **24** (1953), 265—272.
[4] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, New York, 1954.
[5] L. L. Campbell, A coding theorem and Rényi's entropy, *Information and Control* **8** (1965), 423—429.
[6] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations, *Ann. Math. Statist.* **23** (1952), 493—507.
[7] I. Csiszár, Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten, *Magyar. Tud. Akad. Mat. Kutató Int. Közl.* **8** (1963), 85—108.
[8] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, *Studia Sci. Math. Hungar.* **2** (1967), 299—318.
[9] R. G. Gallager, *Information Theory and Reliable Communication*, New York, 1968.
[10] H. Heyer, Erschöpftheit und Invarianz beim Vergleich von Experimenten, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **12** (1969), 21—55.
[11] F. Jelinek, Buffer overflow in variable length coding of fixed rate sources, *IEEE Trans. Information Theory* **14** (1968), 490—501.
[12] S. Kullback, *Information Theory and Statistics*, New York, 1959.
[13] D. V. Lindley, On a measure of information provided by an experiment, *Ann. Math. Statist.* **27** (1956), 986—1005.
[14] N. Morse and R. Sackstedter, Statistical isomorphism, *Ann. Math. Statist.* **37** (1966), 203—214.
[15] A. Perez, Sur la théorie de l'information dans le cas d'un alphabet abstrait, *Trans. First Prague Confer. on Information Theory Statist. Decision Functions, Random Process*, Prague, 1957, 209—243.
[16] A. Perez, Information-theoretic risk estimates in statistical decision, *Kybernetika* (Prague) **3** (1967), 1—21.
[17] A. Rényi, On measures of entropy and information, *Proc. 4th Berkeley Sympos. Math. Statist. and Probability*, Vol. I, Berkeley, 1960, 547—561.
[18] A. Rényi, On the amount of information concerning an unknown parameter in a sequence of observations, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **9** (1964), 617—624.

[19] A. RÉNYI, On some problems of statistics from the point of view of information theory, *Proc. Colloq. on Information Theory*, Debrecen, 1967, 343—347.
[20] R. SACKSTEDTER, A note on statistical equivalence, *Ann. Math. Statist.* **38** (1967), 787—794.
[21] R. SIBSON, Information radius, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** (1969), 149—161.
[22] F. TOPSOE, An information theoretical identity and a problem involving capacity, *Studia Sci. Math. Hungar.* **2** (1967), 291—292.
[23] I. VAJDA, On the convergence of information contained in a sequence of observations, *Proc. Colloq. on Information Theory*, Vol. 2, Debrecen, 1967, 489—501.
[24] I. VAJDA, A contribution to the informational analysis of pattern, *Proc. Fourth Hawaii Internat. Confer. on System Sci.*, Western Period. Co.

*(Received December 2, 1970)*

MTA MATEMATIKAI KUTATÓ INTÉZETE,
BUDAPEST, V., REÁLTANODA U. 13–15.
HUNGARY