Kybernetika

Igor Vajda

On the amount of information contained in a sequence of independent observations

Kybernetika, Vol. 6 (1970), No. 4, (306)--324

Persistent URL: http://dml.cz/dmlcz/125308

Terms of use:

© Institute of Information Theory and Automation AS CR, 1970

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://project.dml.cz

KYBERNETIKA ČÍSLO 4, ROČNÍK 6/1970

On the Amount of Information Contained in a Sequence of Independent Observations

IGOR VAJDA

In the present paper basic properties of a Chernoff bound established previously are summarized and new ones are derived. The Chernoff bound is figuring as an asymptotic parameter in a formula for Shannon's information contained in a sequence of independent observations concerning a discrete parameter.

By θ we denote a random variable taking on a finite number of values $1, 2, \ldots$ and by ξ another random variable with a sample measurable space $(X, \mathscr{X})$. By $\xi_1, \xi_2, \ldots$ subsequent realizations of ξ will be denoted; they are supposed to be mutually independent for any given value of θ. Finally, $I(\theta, \xi_1, \ldots, \xi_n)$ will denote the Shannon's information contained in $(\xi_1, \xi_2, \ldots, \xi_n)$ concerning θ.

The information $I(\theta, \xi_1, \ldots, \xi_n)$ can serve as an important numerical characteristic of the following statistical problem: the statistician is interested in the value of θ which is not directly observable but he can observe the values of $\xi_1, \xi_2, \ldots, \xi_n$. It holds $I(\theta, \xi_1, \ldots, \xi_n) = 0$ iff (if and only if) the sample $(\xi_1, \xi_2, \ldots, \xi_n)$ is independent of θ. In general $I(\theta, \xi_1, \ldots, \xi_n) \in [0, H(\theta)]$, where $H(\theta)$ is the Shannon's entropy of the random variable θ; relation $I(\theta, \xi_1, \ldots, \xi_n) = H(\theta)$ holds iff for any realization of $(\xi_1, \xi_2, \ldots, \xi_n)$ the value of θ can be uniquely determined with probability 1. (Remark that the first equality holds iff θ and ξ are independent whereas the second equality holds iff there exists a deterministic relation between θ and ξ.)

It can be relatively very easily shown (cf. Th. 1 in [1]) that*

$$I(\theta, \xi_1, \ldots, \xi_n) \approx H(\theta) - \exp(-nD), \tag{1}$$

where $D \in [0, +\infty]$ depends on a conditional distribution $P_{\xi|\theta}$ of ξ only. The parameter D has been independently evaluated by A. Rényi [2] and by the author (cf. Th. 2 in [1]); it was shown that D is the Chernoff bound [3] corresponding to a Bayes testing of the simple hypotheses $H_i : \theta = i$, $i = 1, 2, \ldots$, on the basis of $(\xi_1, \xi_2, \ldots$

* We write $a_n \approx a - \lambda^n$ instead of $a_n = a - \lambda^{n+o(n)}$, $n = 1, 2, \ldots$

$\ldots, \xi_n)$. In [3], D has been interpreted as an asymptotic efficiency of the Bayes test suggested above.

Some basic properties of the parameter D were presented in [3], another ones were stated in [1], however, without explicite proofs. Moreover, it is to be noted that assertions (d), (e), and (g) in [1] hold only if the probability measures considered there are absolutely continuous (this supposition was not explicitely emphasized in [1]). Consequently, an analogical investigation of the "discontinuous" case which is very interesting too is advisable. Therefore, by the present rather review paper we are resuming the subject of [1].

In Theorems 5–7 and 9, 10 below the assertions of [1] are summarized (including the case where the probability measures mentioned above are not absolutely continuous). In Theorems 1–4 basic properties of a modified α-entropy and a modified relative Shannon's entropy are established. The modified concepts differ from non-modified ones in the "discontinuous" case mentioned above only; it seems however that they are not only more suitable than the non-modified ones when asymptotic problems of the present type are solved, but also provide tools for a more accurate analysis of such problems. (In this respect, compare, for example, (b) and (d) in [1] or (2.8) in [2] with Th. 5 below.) Finally, in Th. 8 a convergence property od D's corresponding to a sequence of sub-σ-algebras of $\mathscr{X}$ is established. Though D is a special version of the α-entropy, this property cannot be deduced directly from the semimartingale convergace theorem.

1. MODIFIED CONCEPTS OF α- AND SHANNON'S ENTROPY

Already in [3] a functional of the following form

$$\int_X p^\alpha q^{1-\alpha} \, \mathrm{d}\mu \,, \quad \alpha \in (0, 1) \,,$$

(cf. also [4]) has been investigated, where p, q are the Radon-Nikodym densities of probability distributions P, Q on $(X, \mathscr{X})$ with respect to another (dominating) probability distribution μ on $(X, \mathscr{X})$. In accordance with [4], the functional will be denoted by $H_\alpha(P, Q)$ and called, simply, α-entropy. Some basic properties of this functional can be deduced from Theorem 4.1.s in Chap. VII of [5].

Before going into a more detailed discussion of α-entropies, let us note that in the statistical model itroduced above we shall suppose that θ takes on two values 1 and 2 only and that $P[\theta = 1] = P[\theta = 2] = \frac{1}{2}$. It follows from what was said in [1, 2] that the general case where the number of possible values of θ is arbitrary finite does not present any essential new difficulty. (In the general case D is defined as the minimum of the Chernoff bounds corresponding to the pairs of hypotheses $\theta = i$, $\theta = j$ such that $P[\theta = i] > 0$, $P[\theta = j] > 0$, taken over all such pairs.) In the sequel, P or Q will be interpreted as the conditional distribution $P_{\xi|\theta=1}$ or $P_{\xi|\theta=2}$ respectively.

Thus,

$$(2) \qquad P_{\xi_1\xi_2\ldots\xi_n|\theta=1} = P \times P \times \ldots \times P(n \text{ times}),$$
$$P_{\xi_1\xi_2\ldots\xi_n|\theta=2} = Q \times Q \times \ldots \times Q(n \text{ times}).$$

In a connection with an evaluation of the parameter D, the following slightly modified concept of the α-entropy will be useful

$$(3) \qquad H'_\alpha(P, Q) = \int_{C(P,Q)} p^\alpha q^{1-\alpha}\, \mathrm{d}\mu,$$

where $C(P, Q) = \{pq > 0\} \in \mathscr{X}$ is a set of absolute continuity of P, Q.

It is to see at the first sight that $H'_\alpha = H_\alpha$ if P, Q are mutually absolutely continuous and

$$(4) \qquad H'_\alpha(P, Q) = H_\alpha(P, Q) \quad \alpha \neq 0, 1,$$

for every P, Q. (Let us note that, unless the contrary will be explicitely stated, we shall consider the α-entropies for $\alpha \in [0, 1]$ only.) Further, it is fruitful to notice (cf. [4, 6]) that $H_\alpha(P, Q)$ is the real restriction of

$$H(z) = \int_{-\infty}^{+\infty} \mathrm{e}^{zu}\, \mathrm{d}F(u),$$

where $z = \alpha + i\beta$ is a complex number and $F(u) = Q(\{p \leqq q \exp(u)\})$ is the distribution function of the likelihood ratio coresponding to the simple hypotheses P and Q. It follows from the theory of bilateral Laplace transform (cf. [7]) that $H_\alpha(P, Q)$ is finite for $\alpha \in [0, 1]$ and that $H_\alpha(P, Q)$ is an analytic function of α on $(0, 1)$ with derivatives (cf. [4])

$$(5) \quad \frac{\mathrm{d}^k}{\mathrm{d}\alpha^k} H_\alpha(P, Q) = \int_X p^\alpha q^{1-\alpha} \left(\log \frac{p}{q}\right)^k \mathrm{d}\mu \quad \text{for every} \quad \alpha \in (0, 1),\ k = 1, 2, \ldots$$

which, however, need not be always continuous at $\alpha = 0, 1$.

Using (4) these results may be immediately applied to H'_α; as we shall prove below (cf. Th. 2, where properties of H'_α as a function of α are summarized), $H'_\alpha(P, Q)$ is continuous on $[0, 1]$ (or, more generally, on the set J of all α for which $H'_\alpha(P, Q) < \infty$; it follows from the theory of bilateral Laplace transform that J is always an interval on the real line).

Since

$$(6) \qquad H'_\alpha(P, Q) = \mathsf{E}_Q \chi_{C(P,Q)} \left(\frac{p}{q}\right)^\alpha,$$

the semimartingale theorem cannot be applied to H'_α unless $\chi_{C(P,Q)} = 1\ [Q]$. Nevertheless we shall see that H'_α possesses all the convergence properties, which can be derived for H_α from the semimartingale convergence theorem. Of course, in view of (4), we may restrict ourselves to the case $\alpha = 0$ or 1.

Let $\mathscr{X}_1 \subset \mathscr{X}_2 \subset \ldots$ be sub-σ-algebras of $\mathscr{X}$ and let P_n, Q_n be restrictions of P, Q on $\mathscr{X}_n$, $n = 1, 2, \ldots$ Clearly, $C(P_n, Q_n) \in \mathscr{X}_n$.

We shall say that a sub-σ-algebra $\mathscr{X}_n \subset \mathscr{X}$ is $C(P, Q)$-sufficient if $P(C(P_n, Q_n)) = P(C(P, Q))$. Obviously, if $\mathscr{X}_n$ is $C(P, Q)$-sufficient, it need not be also $C(Q, P)$-sufficient, but if it is sufficient with respect to P and Q, then it is $C(P, Q)$- as well as $C(Q, P)$-sufficient.

Theorem 1. *For every* $\alpha \in [0, 1]$

$$H'_\alpha(P_1, Q_1) \geqq H'_\alpha(P_2, Q_2) \geqq \ldots \tag{7}$$

and, if $\mathscr{X}$ *is generated by the algebra*

$$\mathscr{X}_0 = \bigcup_{n=1}^{\infty} \mathscr{X}_n ,$$

then

$$\lim_n H'_\alpha(P_n, Q_n) = H'_\alpha(P, Q) . \tag{8}$$

If $\alpha \in (0, 1)$ *then* $H'_\alpha(P_n, Q_n) = H'_\alpha(P, Q)$ *iff (if and only if)* $\mathscr{X}_n$ *is sufficient with respect to P and Q. If* $\alpha = 0$ *or* 1 *then this equality holds iff* $\mathscr{X}_n$ *is* $C(Q, P)$*-sufficient or* $C(P, Q)$*-sufficient respectively.*

Proof. By (4), the assertion stated here for $\alpha \in (0, 1)$ has been proved in [5]. If $\alpha = 0, 1$ then, it may be easily deduced from the inclusion

$$C(P_n, Q_n) \supset C(P, Q) \tag{9}$$

and from the fact that $\{1 - \chi_{C(P_n, Q_n)}\}$, $n = 1, 2, \ldots$ is a semimartingale with respect to both P and Q.

To prove (9) it will suffice to prove that the conditional densities

$$p_n = \mathsf{E}(p \mid \mathscr{X}_n) , \quad q_n = \mathsf{E}(q \mid \mathscr{X}_n) \tag{10}$$

may be defined in such a way that $p(x)\, q(x) > 0$ implies $p_n(x)\, q_n(x) > 0$, $x \in X$. If $E = \{p_n = 0\} \in \mathscr{X}_n$, then the equality defining p_n implies that the set $F \subset E$ of all $x \in X$ for which $p(x) > 0$ is of P-measure zero, i.e. we may put $p_n = 1$ on F. Thus $p_n(x) = 0$ implies $p(x) = 0$ for every $x \in X$. Since we may analogically proceed with q, q_n, the implication requested above is true. (9) implies that $\{1 - \chi_{C(P_n, Q_n)}\}$, $n = 1, 2, \ldots, \infty$ is a semimartingale, Q.E.D.

Theorem 2. $H'_\alpha(P, Q)$ *is continuous convex function on* $[0, 1]$ *with*

(11)

$$\frac{\mathrm{d}^k}{\mathrm{d}\alpha^k} H'_\alpha(P, Q) = \int_{C(P,Q)} p^\alpha q^{1-\alpha} \left(\log \frac{p}{q}\right)^k \mathrm{d}\mu \quad \textit{for every} \quad k = 1, 2, \ldots \textit{ and } \alpha \in [0, 1] ,$$

 where the integrals in (11) *are finite for* $\alpha \in (0, 1)$ *and well-defined for* $\alpha = 0, 1$, *and*

$$0 \leqq H'_\alpha(P, Q) \leqq 1 ,$$

where $H'_\alpha(P, Q) = 0$ *for some* $\alpha \in [0, 1]$ (*and, consequently, for all* $\alpha \in [0, 1]$) *iff* $P \perp Q$ *and* $H'_\alpha(P, Q) = 1$ *for some* $\alpha \in (0, 1)$ (*and, consequently, for all* $\alpha \in [0, 1]$) *iff* $P = Q$. *For* $\alpha = 0$ *or* 1, $H'_\alpha(P, Q) = 1$ *iff* $Q \ll P$ *or* $P \ll Q$ *respectively*. H'_α *is strictly convex if neither* $P \perp Q$ *nor* $P = Q$.

Remark. The derivatives in (11) for $\alpha = 0$ or 1 are to be considered as those on the right or left respectively.

Proof. We shall prove firstly that H'_α is continuous on $[0, 1]$. One of the methods to prove this is to form a sequence of sub-σ-algebras $\mathscr{X}_1 \subset \mathscr{X}_2 \subset \ldots$ of $\mathscr{X}$ generated by finite measurable decompositions of X. As it was shown in [8], the decompositions may be defined in such a manner that the σ-algebra $\mathscr{X}' \subset \mathscr{X}$ generated by the corresponding algebra $\mathscr{X}_0$ (cf. Th. 1) is sufficient with respect to P and Q. Since, evidently, every $H'_\alpha(P_n, Q_n)$ is continuous and convex on $[0, 1]$, it follows from Th. 1 that $H_\alpha(P, Q)$ is a limit of continuous and uniformly converging (on $[0, 1]$) functions, i.e. it is continuous as well. The convexity will follow from (11) for $k = 2$ and the assertions following (11) can be deduced from (3) and (11).

Thus it remains to prove that the integrals in (11) are finite or well-defined respectively and that (11) holds. But, according to (5) (see also [4]), the integrals (11) are finite for every $\alpha \in (0, 1)$ and $k = 1, 2, \ldots$ Since the functions $u(\log u)^k$ are bounded from below for every $u \in (0, +\infty)$ and $k = 1, 2, \ldots$, the integrals in (11) are well-defined for $\alpha = 1$ as well. The same is true also for $\alpha = 0$ and $k = 2, 4, 6, \ldots$ If k is odd, then we can write

$$\int_{C(P,Q)} q \left(\log \frac{p}{q}\right)^k \mathrm{d}\mu = - \int_{C(P,Q)} q \left(\log \frac{q}{p}\right)^k \mathrm{d}\mu$$

so that, interchanging the role of P and Q in the case $\alpha = 1$ above we obtain the desired assertion.

Relation (11) holds for every $\alpha \in (0, 1)$ and $k = 1, 2, \ldots$ by (4) and (5). If $\alpha = 0$ and $k = 1$ (for $\alpha = 1$ as well as $k = 2, 3, \ldots$ a similar argument can be used), we can write

$$H'_0(P, Q) - H'_\alpha(P, Q) = \alpha \int_{C(P,Q)} u^{\xi(u)} \log u \, \mathrm{d}Q \quad \text{for every} \quad \alpha \in (0, 1) ,$$

where $\xi(u) \in [0, \alpha]$ is a Borel function of $u \in [0, +\infty]$ and $u = p/q$ on $C(P, Q)$. If

$$\int_{C(P,Q)} \log u \, \mathrm{d}Q$$

is finite the proof is obvious. Now, since $\log u \leqq u - 1$ for every real u, the following inequality holds

$$\int_{C(P,Q)} \log u \, \mathrm{d}Q \leqq P(C(P, Q)) - Q(C(P, Q)) < +\infty$$

and it remains to investigate the case

$$\int_{C_*} \log u \, \mathrm{d}Q = -\infty ,$$

where the set $C_* \in \mathscr{X}$ is defined by $C_* = \{u \leqq 1\} \cap C(P, Q)$. Since $\xi(u) \in [0, \alpha]$, it holds

$$u^{\xi} \log u \leqq u^{\alpha} \log u \quad \text{on} \quad C_*$$

and it remains to prove that for every $A > 0$ there exists $\alpha \in (0, 1)$ such that

$$\int_{C_*} u^{\alpha} \log u \, \mathrm{d}Q < -A .$$

If we $F_n = \{u \geqq 1/n\} \cap C_* \in \mathscr{X}$, then

$$\lim_n \int_{F_n} \log u \, \mathrm{d}Q = -\infty$$

so that, for some n,

$$\int_{F_n} \log u \, \mathrm{d}Q \leqq -2A .$$

If now $0 \leqq \alpha < \log 2/\log n$, then

$$u^{\alpha} \log u < \tfrac{1}{2} \log u \quad \text{on} \quad F_n ,$$

and we can successively write

$$\int_{C_*} u^{\alpha} \log u \, \mathrm{d}Q \leqq \int_{F_n} u^{\alpha} \log u \, \mathrm{d}Q < \tfrac{1}{2} \int_{F_n} \log u \, \mathrm{d}Q \leqq -A .$$

The same modification as above we shall also consider in connection with the generalized entropy of Shannon (or discrimination information) of P, Q introduced into the literature by S. Kullback and A. Perez, i.e. instead of

$$H(P, Q) = \int_X p \log \frac{p}{q} \, \mathrm{d}\mu$$

we shall consider

$$H'(P, Q) = \int_{C(P,Q)} p \log \frac{p}{q} \, \mathrm{d}\mu \geqq P(C) \log \frac{P(C)}{Q(C)} \geqq -\frac{1}{e} , \tag{12}$$

where C stands for $C(P, Q)$. Let us notice that $H'(P, Q)$ may take on negative values as well and that $H'(P, Q) < 0$ implies $H(P, Q) = +\infty$.

Theorem 3. *If $H'(P, Q) \leqq 0$ then $H'(Q, P) \geqq 0$ where the strict inequality holds unless either $P \perp Q$ on $\mathscr{X}$ or $P = Q$ on $C(P, Q) \cap \mathscr{X}$.*

Proof. From (12) we obtain

$$H'(P, Q) + H'(Q, P) \geqq (P(C) - Q(C)) \log \frac{P(C)}{Q(C)} \geqq 0 \tag{13}$$

so that $H'(P, Q) \leqq 0$ or <0 implies $H'(Q, P) \geqq 0$ or >0 respectively. If $H'(P, Q) = H'(Q, P) = 0$ then, by (13), $P(C) = Q(C)$ so that, according to Lemma 1.1 in [9], either $P(C) = Q(C) = 0$ (i.e. $P \perp Q$ on $\mathscr{X}$) or $P = Q$ on $C \cap \mathscr{X}$.

The following identity (14) was found for discrete distributions by A. Rényi [10] (cf. also [11]).

Theorem 4. *For every P and Q,*

$$\lim_{\alpha \to 1^-} \frac{1}{\alpha - 1} \log H'_\alpha(P, Q) = H(P, Q), \tag{14}$$

$$\lim_{\alpha \to 0^+} -\frac{1}{\alpha} \log H'_\alpha(P, Q) = H(Q, P). \tag{15}$$

Proof. We shall prove (14) only; (15) may be proved analogically. If $P \not\ll Q$, then $H(P, Q) = +\infty$ and $H'_1(P, Q) < 1$ (see Th. 2) so that (14) holds. If $P \ll Q$, then, by Th. 2, $H'_1(P, Q) = 1$ so that we can succesively write

$$\lim_{\alpha \to 1^-} \frac{1}{\alpha - 1} \log H'_\alpha(P, Q) = \frac{\lim_{\alpha \to 1^-} \frac{\mathrm{d}}{\mathrm{d}\alpha} H'_\alpha(P, Q)}{\lim_{\alpha \to 1^-} H'_\alpha(P, Q)} = \frac{H(P, Q)}{H'_1(P, Q)} = H(P, Q)$$

(cf. (11), (12)).

It is to see that (14) and (15) may be replaced by

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} H_\alpha(P, Q)\Big|_{\alpha=0} = -H(Q, P),$$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} H_\alpha(P, Q)\Big|_{\alpha=1} = H(P, Q),$$

where the derivatives are to be considered as those on the right or left respectively.
Analogical relations

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} H'_\alpha(P, Q)|_{\alpha=0} = -H'(Q, P)$$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} H'_\alpha(P, Q)|_{\alpha=1} = H'(P, Q)$$

follow immediately from (11) and (12).

2. D-DIVERGENCE

Now our attention will be paid to $D(P, Q)$ which is a parameter of convergente in (1). The fact that $D(P, Q)$, as it will be defined in this section, is identical with that of the formula (1) will be proved, for the sake of completeness, in the following section.

Let us put* (cf. (e) in [1])

$$D(P, Q) = \sup_{\alpha\in[0,1]} - \log H_\alpha(P, Q) = -\log \min_{\alpha\in[0,1]} H'_\alpha(P, Q) . \tag{16}$$

According to (4) and Th. 2, the minimum in (16) exists and the second equality holds. Th. 2 also implies the following two theorems (cf. (b) and (d) in [1]).

Theorem 5. *$H'(P, Q)$, $H'(Q, P) > 0$ iff*

$$D(P, Q) = -\log H'_\alpha(P, Q) , \tag{17}$$

for $\alpha \in (0, 1)$ which is a unique solution of the equation

$$\int_{C(P,Q)} p^\alpha q^{1-\alpha} \log \frac{p}{q} \, \mathrm{d}\mu = 0 , \tag{18}$$

$H'(P, Q) \leqq 0$ iff

$$D(P, Q) = -\log H'_1(P, Q), \tag{19}$$

and $H'(Q, P) \leqq 0$ iff

$$D(P, Q) = -\log H'_0(P, Q) .$$

According to Th. 3, $H'(P, Q)$, $H'(Q, P) \leqq 0$ iff $H'(P, Q) = H'(Q, P) = 0$ which is equivalent to $P \perp Q$ or $P = Q$. By Th. 2, both later conditions imply $H'_0(P, Q) = H'_1(P, Q)$ so that Th. 5 is self-consistent. Let us recall that $H'(P, Q) \neq H(P, Q)$,

* By log we denote in this paper the natural logarithm.

 i.e., particularly, $H'(P, Q) < 0$, may appear only if $P \not\ll Q$, so that if $P \equiv Q$, $P \neq Q$, then (17) is true.

Theorem 6. *$D(P, Q)$ is symmetric non-negative extended real valued function of P, Q. $D(P, Q) = 0$ iff $P = Q$ and $D(P, Q) = +\infty$ iff $P \perp Q$.*

The symmetry stated in this theorem follows from (16) and from the identity $H'_\alpha(Q, P) = H'_{1-\alpha}(P, Q)$ which is true for every $\alpha \in [0, 1]$.
In [3] it was proved that (cf. (f) in [1])

$$D(\prod_{i=1}^{n} P_i, \prod_{i=1}^{n} Q_i) \leqq \prod_{i=1}^{n} D(P_i, Q_i)$$

and

$$D(\prod_{i=1}^{n} P_i, \prod_{i=1}^{n} Q_i) = n\, D(P, Q) \quad \text{if} \quad P_i = P,\ Q_i = Q,\ i = 1, 2, \ldots$$

Th. 1 together with Th. 2 (cf. (11), (12)) yields the following result (cf. (g) in [1]).

Theorem 7. *If P', Q' are restrictions of P, Q on a sub-σ-algebra $\mathscr{X}'$ of $\mathscr{X}$, then*

$$D(P', Q') \leqq D(P, Q) \tag{21}$$

where the sign of equality holds iff $\mathscr{X}'$ is sufficient with respect to P, Q or $C(P, Q)$-sufficient or $C(Q, P)$-sufficient depending on whether $H'(P, Q)$, $H'(Q, P) > 0$ or $H'(P, Q) \leqq 0$ or $H'(Q, P) \leqq 0$ respectively.
The following assertion is new.

Theorem 8. *If $\mathscr{X}_n$, P_n, Q_n are defined as in Th. 1, then*

$$D(P_1, Q_1) \leqq D(P_2, Q_2) \leqq \ldots \tag{22}$$

and if $\mathscr{X}$ is generated by the algebra $\mathscr{X}_0$ (cf. Th. 1), *then*

$$\lim_n D(P_n, Q_n) = D(P, Q)\,. \tag{23}$$

Proof. According to (16) and Th. 7,

$$-\log H'_\alpha(P_n, Q_n) \leqq D(P_n, Q_n) \leqq D(P, Q)$$

where α is defined by $D(P, Q) = -\log H'_\alpha(P, Q)$. Now it remains to apply Th. 1.

Next we shall prove that $D(P, Q)$ as defined by (16) is identical with that defined by a different manner in (3.2) of [1]. The definition (3.2) was merely based on the concept of generalized Shannon's entropy. As a by-product the inequality $D(P, Q) \leqq \min [H(P, Q), H(Q, P)]$ will be obtained. This result becomes evident if compared with the Chernoff-Stein asymptotical formulas for the power of Neyman-Pearson tests of $\theta = 1(2)$ against $\theta = 2(1)$ based on $\xi_1, \xi_2, \ldots, \xi_n$. In these formulas $H(P, Q)$

(or $H(Q, P)$) is figuring in the exponent of convergence analogically as $D(P, Q)$ in (1). For a deeper insight into these questions we refer to [13, 11] (cf. also the following formula (36)).

Let P and Q be arbitrary fixed probability measures and denote by $\mathscr{P}$ or $\mathscr{Q}$ the set of all probability measures R on $(X, \mathscr{X})$ such that

$$H(R, P) \geqq H(R, Q) \tag{24}$$

or

$$H(R, P) \leqq H(R, Q) \tag{25}$$

respectively. The definition we beared in mind above is as follows:

$$D(P, Q) = \min\left[\inf_{\mathscr{P}} H(R, P), \inf_{\mathscr{Q}} H(R, Q)\right]. \tag{26}$$

The next our aim will be to prove and precise (26)*.

Let $\mathscr{P}_0 \subset \mathscr{P}$ or $\mathscr{Q}_0 \subset \mathscr{Q}$ denote the subclasses of all R such that $H(R, P) < +\infty$ or $H(R, Q) < +\infty$ respectively and let $\mathscr{R}$ stands for the set of all measures R dominated by μ and concentrated on $C(P, Q)$, i.e. $R(C(P, Q)) = 1$.

Lemma 1. *$\mathscr{P}_0 \cup \mathscr{Q}_0 \subset \mathscr{R}$, i.e. if $R \in \mathscr{P} \cup \mathscr{Q} - \mathscr{R}$, then $H(R, P) = H(R, Q) = +\infty$.*

Proof. Let $R \in \mathscr{P} - \mathscr{R}$ and let us distinguish two alternate cases: $R \not\ll \mu$, $R(C) = 1$ and $R \ll \mu$, $R(X - C) > 0$, where, here and in the sequel, C denotes $C(P, Q)$. If $R \not\ll \mu$, then also $R \not\ll P$ so that, by the definition of the generalized Shannon's entropy, $H(R, P) = +\infty$. If $R(X - C) > 0$, then either $R \not\ll P$ or $R \not\ll Q$. The first case we have just investigated above and if $R \not\ll Q$, then $H(R, Q) = +\infty$. This together with the condition (24) for $R \in \mathscr{P}$ implies $H(R, P) = +\infty$, Q.E.D.

Lemma 2. *If $R \in \mathscr{R}$, then $R \in \mathscr{P}_0$ or $\mathscr{Q}_0$ iff*

$$\int_C r \log \frac{q}{p} \, d\mu \geqq 0 \quad or \quad \leqq 0 \tag{27}$$

and $R \in \mathscr{P}_0 \cap \mathscr{Q}_0$ iff

$$\int_C r \log \frac{q}{p} \, d\mu = 0\,,$$

where $r = dR/d\mu$.

Proof. If $R \in \mathscr{P}_0 \subset \mathscr{R}$, then $r = dR/d\mu$ exists by Lemma 1. It follows from the definition of C that the integral in (27) or (28) exists. The remainder is clear.

* During a preparation of this manuscript for printing, A. Rényi has published analogical definition of D in the printed version [2] of his lecture.

Lemma 3. *For every $R \in \mathscr{R}$ and $\alpha \in [0, 1]$*

$$H(R, P) \geqq (1 - \alpha) \int_C r \log \frac{q}{p} \, d\mu - \log H'_\alpha(P, Q) \tag{29}$$

with equality iff

$$r = \begin{cases} (H'_\alpha(P, Q))^{-1} \, p^\alpha q^{1-\alpha} & \text{on } C \\ 0 & \text{out of } C \, . \end{cases} \tag{30}$$

Proof. S. Kullback proved in Chap. 3 of his book [12] that for every extended real-valued measurable statistic T defined on $(X, \mathscr{X})$, for every real τ and non-negative β, and for every $R \ll \mu$ the following inequality holds

$$H(R, P) \geqq \tau \int_X rT \, d\mu + \log \beta + 1 - \beta \int_X \exp(\tau T) \, d\mu$$

if only the corresponding integrals exist and that the equality takes place iff $r = \exp(\tau T)$. Putting $\tau = \alpha - 1$,

$$T = \begin{cases} \log \dfrac{p}{q} & \text{on } C \, , \\ -\infty & \text{out of } C \, , \end{cases}$$

and $\beta = (H'_\alpha(P, Q))^{-1}$ we obtain (29). The rest of the proof is now clear.

On the basis given by these lemmas, (26) can be easily proved. Let us consider firstly the case where $H'(P, Q)$, $H'(Q, P) > 0$. Here $P(E \cap C)$, $Q(C - E) > 0$ and, since P, Q are absolutely continuous on $C \cap \mathscr{X}$, $P(C - E)$, $Q(E \cap C) > 0$, where $E = \{\log p/q \geqq 0\} \in \mathscr{X}$. It is easy to see that these facts enable us to argue that $\mathscr{P}_0 \cap \mathscr{Q}_0 \neq 0$. Further, the definition of $\mathscr{P}_0$, $\mathscr{Q}_0$ yields

$$\inf_{\mathscr{P}} H(R, P) = \inf_{\mathscr{P}_0} H(R, P) \, , \quad \inf_{\mathscr{Q}} H(R, Q) = \inf_{\mathscr{Q}_0} H(R, Q) \, . \tag{31}$$

However, we shall prove more, namely,

$$\inf_{\mathscr{P}_0} H(R, P) = \inf_{\mathscr{P}_0 \cap \mathscr{Q}_0} H(R, P) = \inf_{\mathscr{P}_0 \cap \mathscr{Q}_0} H(R, Q) = \inf_{\mathscr{Q}_0} H(R, Q) \, . \tag{32}$$

Theorem 9. *If $H'(P, Q)$, $H'(Q, P) > 0$, then $\mathscr{P}_0 \cap \mathscr{Q}_0 \neq 0$, (32) holds and*

$$D(P, Q) = \inf_{\mathscr{P}_0 \cap \mathscr{Q}_0} H(R, P) = H(R, P) \, ,$$

where $R \in \mathscr{P}_0 \cap \mathscr{Q}_0$ is uniquely $[\mu]$ defined by (30) for $\alpha \in (0, 1)$ given by (18).

Proof. Let α in Lemma 3 be defined by (18) and let $R \in \mathscr{P}_0 \cap \mathscr{Q}_0$ be arbitrary. Then, by Lemmas 3, 2 and Theorem 5, $H(R, P) \geqq D(P, Q)$ with equality iff (30) holds. Q.E.D.

Theorem 10. *If $H'(P, Q) \leqq 0$ and P, Q are not mutually singular, then $\mathscr{P}_0 \neq 0$ and*

$$D(P, Q) = \inf_{\mathscr{P}_0} H(R, P) = H(R, P) \leqq \inf_{\mathscr{Q}} H(R, Q),$$

where $R \in \mathscr{P}_0$ is defined uniquely $[\mu]$ by (30) for $\alpha = 1$.

Proof. If P and Q are not singular, then $P(C) > 0$ and r defined by (30) for $\alpha = 1$ is a probability density function. By Lemma 2, $R \in \mathscr{R}$ given by r belongs to $\mathscr{P}$ (and, consequently, to $\mathscr{P}_0$) iff $H'(P, Q) \leqq 0$. The equalities in Th. 10 now follow from Lemma 3 (for $\alpha = 1$) and from (19). As to the inequality, let us notice that, replacing P and Q in Lemma 3, we may write

$$H(R, Q) \geqq \int_C r \log \frac{p}{q} \, \mathrm{d}\mu + D(P, Q)$$

for every $R \in \mathscr{R}$, where $D(P, Q)$ is defined by (19) again. But (27) and Lemma 2 imply that the integral is non-negative for any $R \in \mathscr{Q}$, i.e. the inequality is true.

Th. 9 and Th. 10 imply the following

Corollary. *The relation (26) holds. If $D(P, Q) < +\infty$, then the minimum in (26) is attained on $R \in \mathscr{R}$ defined by (30) for appropriately defined $\alpha \in [0, 1]$.*

Since $P \in \mathscr{Q}$, $Q \in \mathscr{P}$, (26) implies the following inequality:

(33) $$D(P, Q) \leqq \min\,[H(P, Q), H(Q, P)] .$$

3. TOTAL VARIATION

In [1] an estimate of $D(P, Q)$ in terms of a more simple functional $V(P, Q)$ was given. $V(P, Q)$ was denoting the total variation of P and Q (cf. (h) in [1]). The total variation is defined by

(34) $$V(P, Q) = \int_X |p - q| \, \mathrm{d}\mu = 2 \sup_{E \in \mathscr{X}} [P(E) - Q(E)] = 2[P(F) - Q(F)] ,$$

where $F = \{p \geqq q\} \in \mathscr{X}$. The estimate was of the following form*

(35) $$-\frac{1}{2} \log\left(1 - \frac{V^2(P, Q)}{4}\right) \leqq D(P, Q) \leqq -\log\left(1 - \frac{V(P, Q)}{2}\right).$$

The right hand inequality follows directly from the following formula (36) and from the inequality

$$1 - \tfrac{1}{2}V(P^n, Q^n) \geqq (1 - \tfrac{1}{2}V(P, Q))^n \quad n = 1, 2, \ldots,$$

* My thanks are due to Prof. O. Kraft for calling my attention to the fact that this estimate occurs also in Ch. Kraft, Univ. California Publ. Statist. *2* (1955), 125–142 (added in proof).

(cf. (37)) which is the proof of Th. 1 in [1] based on. The left hand inequality may be proved by a method indicated in [14] (cf. the proof of the inequality (15) in [14]; in this proof it is indifferent whether the measures P, Q are discrete or not), but here another idea will be used.

Let $\mathscr{X}'$ be the sub-σ-algebra of $\mathscr{X}$ consisting of two elements $F, X - F \in \mathscr{X}$, where F is defined as in (34) and let P', Q' be reductions of P, Q on $\mathscr{X}'$. Then, by Th. 7, $D(P', Q') \leqq D(P, Q)$, where

$$D(P', Q') = -\log \inf_{\alpha \in (0,1)} = \psi_\alpha(U, V) \quad \text{for} \quad U = Q(F)\,, \quad V = V(P, Q)\,,$$

and where

$$\psi_\alpha(U, V) = \left(\frac{V}{2} + U\right)^\alpha U^{1-\alpha} + \left(1 - \frac{V}{2} - U\right)^\alpha (1 - U)^{1-\alpha}\,,$$

$$0 \leqq U \leqq 1 - \frac{V}{2}\,, \quad 0 \leqq V \leqq 2\,.$$

Thus it remains to prove that

$$\sup_{U \in [0, 1 - V/2]} \inf_{\alpha \in (0,1)} \psi_\alpha(U, V) \leqq \sqrt{\left(1 - \frac{V^2}{4}\right)} \quad \text{for every} \quad V \in (0, 2)\,,$$

or

$$\sup_{U \in [0, 1 - V/2]} \psi_{1/2}(U, V) \leqq \sqrt{\left(1 - \frac{V^2}{4}\right)}\,.$$

But, however, $\psi_{1/2}(U, V)$ is strictly concave function of U on the interval $[0, 1 - V/2]$ with maximum on $U_0 = \frac{1}{2}(1 - V/2)$, for any $V \in (0, 2)$ so that the desired result follows from this identity:

$$\psi_{1/2}(U_0, V) = \sqrt{\left(1 - \frac{V^2}{4}\right)}\,.$$

The main idea of [1, 2] was based on the fact that a relation between the variation $V(P^n, Q^n)$ and the quantity $H(\theta) - I(\theta, \xi_1, \ldots, \xi_n)$ (cf. (3) and the assumption following it) exists. This relation is represented by a both-sides estimate which is "best possible", i.e. for any value V of $V(P^n, Q^n)$, $V \in [0, 2]$, one can find two nonnegative numbers $L_n(V)$, $U_n(V)$ such that

$$L_n(V) \leqq H(\theta) - I(\theta, \xi_1, \ldots, \xi_n) \leqq U_n(V)\,, \quad n = 1, 2, \ldots$$

provided that (3) and other related assumptions hold and, moreover, both the bounds considered here are attainable, for any $n = 1, 2, \ldots$

We do not aim to discuss this relation explicitely here; it will be important for us only that on the base of such an estimate one can argue that (1) holds iff

$$(36) \qquad 2 - V(P^n, Q^n) \approx \exp\left(-n\, D(P, Q)\right)$$

for $D = D(P, Q)$, where

(37) $\quad P^n = P \times P \ldots \times P(n \text{ times}), \quad Q^n = Q \times Q \times \ldots \times Q(n \text{ times}).$

But, as it was shown in Th. 1 of [1], one can very easily show that (36) always holds for some $D(P, Q)$.

Unfortunately, these considerations do not yield that $D(P, Q)$ figuring here satisfies (16) for every P, Q. However, this statement together with (36) has been proved firstly by H. Chernoff [3]. For the sake of completeness we next reproduce the proof of Chernoff in a slightly modified way using the definition (26) instead of (16) (cf. also Sanov [15]).

Let us suppose, firstly, that $P = (p_1, p_2, \ldots, p_s)$, $Q = (q_1, q_2, \ldots, q_s)$ are two discrete distributions, i.e. that

$$P[\xi = i \mid \theta = 1] = p_i, \quad P[\xi = i \mid \theta = 2] = q_i, \quad i = 1, 2, \ldots, s,$$

(cf. (2)), where

$$\sum_{i=1}^{s} p_i = \sum_{i=1}^{s} q_i = 1.$$

It follows from (34) that

$$1 - \tfrac{1}{2}V(P^n, Q^n) = \sum_{j_1, j_2, \ldots, j_s} \min\left[p(j_1, \ldots, j_s), q(j_1, \ldots, j_s)\right] =$$
$$= \sum_{A_n} p(j_1, \ldots, j_s) + \sum_{B_n} q(j_1, \ldots, j_s),$$

where

$$p(j_1, \ldots, j_s) = \frac{n!}{\prod_{i=1}^{s} j_i!} \prod_{i=1}^{s} p_i^{j_i}, \quad q(j_1, \ldots, j_s) = \frac{n!}{\prod_{i=1}^{s} j_i!} \prod_{i=1}^{s} q_i^{j_i}$$

and

$$A_n = \{j_1, j_2, \ldots, j_s : j_i \geqq 0, \sum_{i=1}^{s} j_i = n, \prod_{i=1}^{s} p_i^{j_i} \leqq \prod_{i=1}^{s} q_i^{j_i}\},$$

$$B_n = \{j_1, j_2, \ldots, j_s : j_i \geqq 0, \sum_{i=1}^{s} j_i = n, \prod_{i=1}^{s} p_i^{j_i} > \prod_{i=1}^{s} q_i^{j_i}\}.$$

Let us now denote by $\mathscr{P} \cup \mathscr{Q}$ the set of all discrete probability distributions $R = (r_1, r_2, \ldots, r_s)$, where $\mathscr{P}$, $\mathscr{Q}$ are defined by (24), (25), and let $\mathscr{R}' \subset \mathscr{P} \cup \mathscr{Q}$ be the set of all R such that for every r_i there exists an integer k such that $r_i = k/n$. Let us put $\mathscr{P}' = \mathscr{P} \cap \mathscr{R}'$, $\mathscr{Q}' = \mathscr{Q} \cap \mathscr{R}'$. Clearly,

$$p(n\tilde{r}_1, \ldots, n\tilde{r}_s) + q(n\tilde{\tilde{r}}_1, \ldots, n\tilde{\tilde{r}}_s) \leqq 1 - \tfrac{1}{2}V(P^n, Q^n) \leqq$$
$$\leqq \operatorname{card}(A_n \cup B_n)\left[p(n\bar{r}_1, \ldots, n\bar{r}_s) + q(n\bar{\bar{r}}_1, \ldots, n\bar{\bar{r}}_s)\right],$$

where $\bar{R} = (\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_s) \in \mathscr{P}'$, $\bar{\bar{R}} = (\bar{\bar{r}}_1, \bar{\bar{r}}_2, \ldots, \bar{\bar{r}}_s) \in \mathscr{Q}'$ are choosen according to

$$p(n\bar{r}_1, \ldots, n\bar{r}_s) = \max_{R \in \mathscr{P}'} p(nr_1, \ldots, nr_s),$$

$$q(n\bar{\bar{r}}_1, \ldots, n\bar{\bar{r}}_s) = \max_{R \in \mathscr{Q}'} q(nr_1, \ldots, nr_s),$$

and $\tilde{R} = (\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_s) \in \mathscr{P}'$, $\tilde{\tilde{R}} = (\tilde{\tilde{r}}_1, \tilde{\tilde{r}}_2, \ldots, \tilde{\tilde{r}}_s) \in \mathscr{Q}' - \mathscr{P}'$ are arbitrary. Since card $(A_n \cup B_n) \leqq n^s$ and since, according to the formula of Stirling,

$$p(nr_1, \ldots, nr_s) = \exp(-n\, H(R, P) + o(n)),$$

$$q(nr_1, \ldots, nr_s) = \exp(-n\, H(R, Q) + o(n)),$$

we can write

$$1 - \tfrac{1}{2}V(P^n, Q^n) \leqq \exp(-n\, D(P, Q) + o(n)) \tag{39}$$

(cf. the inclusions $\mathscr{P}' \subset \mathscr{P}$, $\mathscr{Q}' \subset \mathscr{Q}$ and (26)). If $R = (r_1, r_2, \ldots, r_s) \in \mathscr{P} \cup \mathscr{Q}$ is such that

$$D(P, Q) = \min(H(R, P), H(R, Q)) \tag{40}$$

then, in view of that one can always choose $\tilde{R} \in \mathscr{P}'$, $\tilde{\tilde{R}} \in \mathscr{Q}' - \mathscr{P}'$ such that $|\tilde{r}_i - r_i| \leqq 1/n$, $|\tilde{\tilde{r}}_i - r_i| \leqq 1/n$, $i = 1, 2, \ldots, s$, i.e.

$$H(\tilde{R}, P) \leqq H(R, P) + o(n),$$

$$H(\tilde{\tilde{R}}, Q) \leqq H(R, Q) + o(n),$$

the following inequality can be written

$$1 - \tfrac{1}{2}V(P^n, Q^n) \geqq \exp(-n\, H(R, P) + o(n)) + \exp(-n\, H(R, Q) + o(n)).$$

Thus, by (40),

$$1 - \tfrac{1}{2}V(P^n, Q^n) \geqq \exp(-n\, D(P, Q) + o(n)).$$

This together with (39) implies that (36) holds for $D(P, Q)$ defined by (26) (and, consequently, by (16)) provided that P, Q are discrete distributions with a finite number of atoms.

To prove that (36) holds for $D(P, Q)$ given by (16) (or (26)) for every P, Q let us denote by $D^*(P, Q)$ the quantity figuring in the exponent of (36) in order to distinguish it from that given by (16) and denote, further, by P_s, Q_s restrictions of P, Q on a sub-σ-algebra $\mathscr{X}_s \subset \mathscr{X}$ generated by a decomposition of X consisting of s sets from $\mathscr{X}$. Since, evidently, $V(P_s^n, Q_s^n) \leqq V(P^n, Q^n)$ for every $n = 1, 2, \ldots$, on the base of (36) and on the base of the result we have proved above one can argue that

$$D(P_s, Q_s) \leqq D^*(P, Q). \tag{41}$$

As it is proved in [8], for every P, Q there exists a sequence $\mathscr{X}_1 \subset \mathscr{X}_2 \subset \ldots$ of sub-σ-algebras such that the σ-algebra $\mathscr{X}' \subset \mathscr{X}$ generated by

$$\bigcup_{s=1}^{\infty} \mathscr{X}_s$$

is sufficient with respect to P and Q, so that, by Th. 7 and 8,

$$\lim_s D(P_s, Q_s) = D(P, Q)$$

and, by (51),

$$D(P, Q) \leqq D^*(P, Q)\,. \tag{42}$$

To prove that the strict inequality cannot appear here we shal need the following fact, which follows from what we have already proved for discrete distributions and from results of Sec. 2: If P, Q are two discrete distributions, then

$$P^n(F_n) \approx \exp\left(-n\, D(P, Q)\right) \tag{43}$$

if $H'(Q, P) > 0$ (cf. (36)) and

$$1 - \tfrac{1}{2}V(P^n, Q^n) = P^n(F_n) + Q^n(X^n - F_n) \tag{44}$$

(cf. (34)), where

$$F_n = \{\prod_{i=1}^{n} p(x_i) \leqq \prod_{i=1}^{n} q(x_i)\} \in \mathscr{X}^n\,. \tag{45}$$

Finally, we shall use the fact that $1 - \frac{1}{2}V(P^n, Q^n) \approx \exp\left(-n\, D(P, Q)\right)$ holds for $D(P, Q)$ evaluated by (16) if we replace P, Q by arbitrary totally finite discrete measures $(p_1, p_2, \ldots, p_s)$, $(q_1, q_2, \ldots, q_s)$. Indeed, in what preceeds the norming conditions $P(X) = Q(X) = 1$ never have been used.

Put $p_i = P(E_i)$, $q_i = p_i \exp\left(-\varepsilon i\right)$ for $E_i = \{q \exp \varepsilon(i - 1) < p < q \exp \varepsilon i\}$ for every $i = 0, \pm 1, \pm 2, \ldots$ and $\varepsilon > 0$. It is easy to see that $q^\alpha < p^\alpha \exp \alpha\, \varepsilon(1 - i)$ for $x \in E_i$, $i = 0, \pm 1, \ldots$ so that

$$H'_{1-\alpha}(P, Q) < \exp \varepsilon\alpha \sum_{i=-\infty}^{+\infty} p_i \exp\left(-\alpha\varepsilon i\right) \quad \text{for every} \quad \alpha \in [0, 1]\,.$$

Thus, for appropriately chosen ε and r we can write (cf. (16))

$$\sum_{i=-r}^{r} p_i \exp\left(-\alpha_* \varepsilon i\right) > \exp\left(-D(P, Q) - \delta\right),$$

where $\delta > 0$ is an arbitrary number given in advance and $\alpha_* \in [0, 1]$ is minimizing the sum

$$\sum_{i=-r}^{r} p_i \exp\left(-\alpha\varepsilon i\right) \quad \text{on} \quad \alpha \in [0, 1]\,,$$

 i.e.

$$D(\tilde{P}, \tilde{Q}) \leqq D(P, Q) + \delta, \tag{46}$$

where $\tilde{P}$, $\tilde{Q}$ are totally finite (discrete) measures defined on by the following Radon-Nikodym densities (with respect to the dominating measure μ):

$$\tilde{p}(x) = \frac{p_i}{\mu(E_i)}, \quad q_i(x) = \frac{q_i}{\mu(E_i)} \quad \text{for} \quad x \in E_i, \; i = 0, \pm 1, \ldots, \pm r,$$

and

$$\tilde{p}(x) = 1, \quad \tilde{q}(x) = 0 \quad \text{otherwise}.$$

It follows from the definition of E_i that $p/q \leqq p_i/q_i$ on E_i (if these ratios exist), $i = 0, \pm 1, \ldots$, so that $\tilde{F}_n \subset F_n$ for

$$\tilde{F}_n = \{ \prod_{i=1}^{n} \tilde{p}(x_i) \leqq \prod_{i=1}^{n} \tilde{q}(x_i) \} \in \mathscr{X}^n$$

and for F_n defined by (45) and, consequently,

$$\tilde{P}^n(F_n) = P^n(\tilde{F}_n) \leqq P^n(F_n). \tag{47}$$

In the case we have considered $H'(\tilde{Q}, \tilde{P}) > 0$, so that, according to what was said in a remark above,

$$\tilde{P}^n(\tilde{F}_n) \approx \exp(-n\, D(\tilde{P}, \tilde{Q}))$$

(cf. (43)), i.e.

$$-\frac{1}{n} \log \tilde{P}^n(\tilde{F}_n) = D(\tilde{P}, \tilde{Q}) \leqq D(P, Q) + \delta \tag{48}$$

(cf. (46)). On the other hand, taking into account (44) and (36), we can write

$$-\frac{1}{n} \log P^n(F_n) \geqq D^*(P, Q).$$

This together with (47) and (48) yields the inequality

$$D^*(P, Q) \leqq D(P, Q) + \delta.$$

Since δ may be chosen arbitrarily small, the desired equality between $D^*(P, Q)$ and $D(P, Q)$ is proved.

We remark that A. Rényi, using a more accurate relation

$$2 - V(P^n, Q^n) = O\left(\frac{1}{\sqrt{n}} \exp\left[-n\, D(P, Q)\right]\right)$$

following from a more general result of R. R. Bahadur and R. Ranga Rao [16], stated in [2] the following sharpening of (1):

$$I(\theta, \xi_1, \ldots, \xi_n) = H(\theta) - O\left(\frac{1}{\sqrt{n}} \exp\left[-n\, D(P, Q)\right]\right).$$

(Received December 18, 1969.)

REFERENCES

[1] I. Vajda: On the convergence of information contained in a sequence of observations. Proc. Coll. on Inf. Th., Debrecen (Hungary), Budapest 1969.

[2] A. Rényi: On some basic problems of statistics from the point of view of information theory. Proc. Coll. on Inf. Th., Debrecen (Hungary), Budapest 1969.

[3] H. Chernoff: A measure of efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Stat. *23* (1952), 493—507.

[4] I. Vajda: Limit theorems for total variation of Cartesian product measures. Studia Sci. Math. Hungarica (in print).

[5] J. L. Doob: Stochastic processes. J. Willey, N. Y. 1953.

[6] L. H. Koopmans: Asymptotic rate of discrimination for Markov processes. Ann. Math. Stat. *31* (1960), 982—994.

[7] S. Bochner: Lectures on Fourier integrals. Princeton Univ. Press, 1959.

[8] A. Perez: Notions généralisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales. Trans. First Prague Conf. on Inf. Th., Prague 1957.

[9] I. Csiszár: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungarica *2* (1967), 299—318.

[10] A. Rényi: On measures of entropy and information. Proc. 4th Berkeley Symp. on Prob. and Math. Stat., Berkeley, Vol. I, 547—561.

[11] O. Kraft, D. Plachky: Bounds for the power of likelihood ratio tests and their asymptotic properties (Preliminary report, University of Münster).

[12] S. Kullback: Information theory and statistics. Willey, N. Y. 1959.

[13] H. Chernoff: Large sample theory: Parametric case. Ann. Math. Stat *27* (1956), 1—22.

[14] I. Vajda: Accumulation of information in case that sample variables depend on sample size. Studia Sci. Math. Hungarica (in print).

[15] Sanov: On large deviations probabilities of random variables. Mat. Sbornik, N. S. *42* (1957), 11—44.

[16] R. R. Bahadur, R. Ranga Rao: On deviations of the sample mean. Ann. Math. Stat. *31* (1960), 1015—1027.

O množství informace obsažené v posloupnosti nezávislých pozorování

IGOR VAJDA

Nechť θ značí náhodnou veličinu nabývající hodnot $1, 2, \ldots$ a ξ jinou náhodnou veličinu s měřitelným výběrovým prostorem $(X, \mathscr{X})$. Nechť dále $\xi_1, \xi_2, \ldots$ jsou postupné realizace veličiny ξ, o kterých budeme předpokládat, že jsou navzájem nezávislé pro každou hodnotu θ. Nechť nakonec $I(\theta, \xi_1, \ldots, \xi_n)$ je množství Shannonovy informace o veličině θ obsažené v $(\xi_1, \xi_2, \ldots, \xi_n)$.

Je známo, že $I(\theta, \xi_1, \ldots, \xi_n) \in [0, H(\theta)]$, kde $H(\theta)$ je entropie veličiny θ, a že informace nabývá hodnot 0 resp. $H(\theta)$ právě když θ a $(\xi_1, \xi_2, \ldots, \xi_n)$ jsou nezávislé resp. deterministicky závislé. Je tedy informace jakožto míra statistické závislosti mezi θ a $(\xi_1, \xi_2, \ldots, \xi_n)$ důležitou číselnou charakteristikou statistického problému, který spočívá ve stanovení neznámé hodnoty parametru θ pouze na základě znalosti hodnoty náhodného výběru $(\xi_1, \xi_2, \ldots, \xi_n)$.

Poměrně velmi snadno (viz věta 1 v [1]) lze dokázat, že existuje parametr $D \in [0, +\infty]$ závislý toliko na podmíněné distribuci $P_{\xi|\theta}$ pro který platí vztah (1). Explicitní analytický výraz pro D byl nezávisle nalezen a současně publikován v referátech A. Rényiho [2] a autora [1]. Jak bylo možné intuitivně očekávat, D je totožné s tzv. Chernoffovou mezí [3], příslušnou Bayesovu testu ke stanovení správné hypotézy $\theta = i$, $i = 1, 2, \ldots$ na základě $(\xi_1, \xi_2, \ldots, \xi_n)$.

Předložená práce shrnuje vlastnosti parametru D odvozené v pracích [1, 2, 3] a dále je prohlubuje. Ve větách 5 až 7 a 9 a 10 jsou v poněkud zobecněné podobě shrnuty a dokázány ty vlastnosti parametru D, které v [1] byly vysloveny bez důkazu. Věty 1 až 4 stanoví vlastnosti modifikované α-entropie a modifikované relativní Shannonovy entropie (nazývané též diskriminační informace). Obě modifikované entropie jsou ve většině případů totožné s nemodifikovanými, avšak v jistých rovněž velmi významných případech se tyto pojmy liší. Jejich zavedení umožňuje nejen formální zjednodušení úvah, ale poskytuje též možnost přesněji popsat a jemněji klasifikovat statistické problémy uvažovaného typu. Nakonec, ve větě 8 je stanovena jistá konvergenční vlastnost funkcionálu D, která neplyne přímo z konvergenčních vlastností α-entropií.

Ing. Igor Vajda, CSc., Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.