

## INFORMATION AND SUFFICIENT SUB-FIELDS<sup>1</sup>

BY S. G. GHURYE<sup>2</sup>

*Indiana University*

**0. Introduction and summary.** This paper is the result of an attempt to clarify and improve some results in the theory of statistical information.

The term information is used to denote different things in different contexts. First of all, there is Shannon's information,  $-\sum p_i \log p_i$ , defined for probability distributions on a finite sample space; this measures, in an esthetically satisfactory way, the entropy or amount of uncertainty in a distribution. Then there is Wiener's information,  $\int f(x) \log f(x) dx$ , defined for an absolutely continuous distribution on the line (or in  $n$ -space); it was introduced by Wiener, with an acknowledgment to von Neumann, as a "reasonable measure" of the amount of information, having the property of being "the negative of the quantity usually defined as entropy in similar situations" ([10], p. 76). Finally, there is "information of one probability distribution  $P$  with respect to another  $Q$ ," commonly known as Kullback-Leibler information. On a finite sample space, this has the form  $\sum p_i \log (p_i/q_i) = -\sum p_i \log q_i - (-\sum p_i \log p_i)$ , and thus has some relationship to entropy; note that the second term, which is the entropy of  $\{p_i\}$ , is the minimum of the first expression over all distributions  $\{q_i\}$ . An interesting idea due to Gelfand, Kolmogorov and Yaglom [3] establishes a connection between the Kullback-Leibler information for a finite probability space and that for any space: If  $P, Q$  are probability measures on a measurable space  $(\Omega, \mathcal{F})$ ,  $P \ll Q$ , and  $\{A_i, i = 1, \dots, n\}$  is any finite measurable partition of  $\Omega$ , then the supremum of  $\sum_i \log [P(A_i)/Q(A_i)]P(A_i)$  over all finite measurable partitions is  $\int_{\Omega} \log (dP/dQ) dP$ . The only published proof of this result seems to be that due to Kallianpur [5], which uses martingale theory.

In Section 1, we shall obtain a rather simple direct proof of this result (Theorem 1.1) and extend it to the case where  $Q$  is any  $\sigma$ -finite measure (Theorem 1.2). Wiener's information is then seen to be the supremum of  $\sum \log [P(A_i)/Q(A_i)]P(A_i)$  over countable partitions, with  $Q =$  Lebesgue measure. Section 2 will be concerned with Kullback-Leibler information. We shall define conditional information relative to a sub-field, establish a relation between this conditional information and sufficiency of the sub-field (Theorem 2.2), and also show that this conditional information equals the difference between information contained in the field and that in the sub-field (Theorem 2.3). These are extensions of results obtained by Kullback and Leibler in a somewhat limited context.

**1. Integral of a convex function and amount of information.** We begin with some notation and definitions, followed by statements of two theorems and

---

Received 28 December 1967.

<sup>1</sup> Research supported in part by NSF Grant GP-6681 at Stanford University.

<sup>2</sup> Now at the University of Alberta, Edmonton, Canada.

comments regarding these; the proofs of the theorems will be given at the end of the section.

Let  $f$  be a non-negative integrable function on a  $\sigma$ -finite measure space  $(\Omega, \mathcal{F}, \mu)$  and let  $\nu(S) = \int_S f d\mu, S \in \mathcal{F}$ . Let  $\varphi$  be a continuous convex function on  $[0, \infty)$  and  $\varphi^+ = \max(\varphi, 0), \varphi^- = \min(\varphi, 0)$ .

DEFINITION 1.1. For any countable family  $\mathcal{A}$  of disjoint measurable sets, let

$$(1.1) \quad \begin{aligned} I_\mu(\varphi^\alpha; \nu, \mathcal{A}) &= \sum_{A \in \mathcal{A}} \varphi^\alpha[\nu(A)/\mu(A)]\mu(A) \\ &= \sum \varphi^\alpha[\text{av}_A f]\mu(A), \end{aligned} \quad \alpha = +, -.$$

If at least one of these two quantities is finite, let

$$(1.2) \quad I_\mu(\varphi; \nu, \mathcal{A}) = I_\mu(\varphi^+; \nu, \mathcal{A}) + I_\mu(\varphi^-; \nu, \mathcal{A}).$$

In case both terms on the right hand side are infinite,  $I_\mu(\varphi; \nu, \mathcal{A})$  is undefined.

Throughout this paper, the same symbol will be used to denote both a family of sets and the  $\sigma$ -ring generated by it. For any  $\sigma$ -field  $\mathcal{G}$ , let  $\Pi_{\mathcal{F}}(\mathcal{G})$  be the class of finite  $\mathcal{G}$ -measurable partitions of  $\Omega$ . Note that if  $\mu(\Omega) < \infty, I_\mu(\varphi; \nu, \mathcal{A})$  exists for all  $\mathcal{A} \in \Pi_{\mathcal{F}}(\mathcal{G})$ , since  $\varphi$  cannot be unbounded both below and above.

DEFINITION 1.2. For any  $\sigma$ -field  $\mathcal{G}$ , let

$$(1.3) \quad I_\mu(\varphi; \nu, \mathcal{G}) = \sup_{\mathcal{A} \in \Pi_{\mathcal{F}}(\mathcal{G})} I_\mu(\varphi; \nu, \mathcal{A}).$$

THEOREM 1.1. If  $\mu(\Omega) < \infty$ , and  $\varphi$  is continuous convex on  $[0, \infty)$ , then

$$(1.4) \quad I_\mu(\varphi; \nu, \mathcal{F}) = \int_\Omega \varphi(f) d\mu.$$

REMARK 1.1. We shall next try to remove the restriction  $\mu(\Omega) < \infty$ , imposed in Theorem 1.1. A few simple considerations immediately show that Theorem 1.1 is not true in case  $\mu(\Omega) = \infty$ . For example, if  $\varphi(0) > 0$  and  $\mu(\Omega) = \infty$ , then  $\text{av}_\Omega f = 0$  and  $\varphi(\text{av}_\Omega f)\mu(\Omega) = \infty$  irrespective of the value of  $\int_\Omega \varphi(f) d\mu$ ; similarly, if  $\varphi(0) = 0$  and  $\Omega^- = \{\omega: \varphi[f(\omega)] \leq 0\}$  has infinite measure, we have  $I_\mu(\varphi; \nu, \mathcal{F}) \geq 0$ , since we can use partitions having  $\Omega^-$  as an atom. It turns out that the theorem can be extended to  $\sigma$ -finite measure spaces if we restrict the class of partitions in (1.3) by the condition that all atoms of a partition must have finite measure. For this purpose, we replace Definition 1.2 by Definition 1.3 or its equivalent Definition 1.3'.

DEFINITION 1.3. For any  $\sigma$ -field  $\mathcal{G}$ , let  $\Pi_C(\mathcal{G})$  be the class of all countable  $\mathcal{G}$ -measurable partitions whose atoms are of finite measure, and let

$$(1.5) \quad I_\mu(\varphi^\alpha; \nu, \mathcal{G}) = \sup_{\mathcal{A} \in \Pi_C(\mathcal{G})} I_\mu(\varphi^\alpha; \nu, \mathcal{A}), \quad \alpha = +, -.$$

If at least one of these two quantities is finite, let

$$(1.6) \quad I_\mu(\varphi; \nu, \mathcal{G}) = I_\mu(\varphi^+; \nu, \mathcal{G}) + I_\mu(\varphi^-; \nu, \mathcal{G}).$$

In case both terms on the right hand side are infinite,  $I_\mu(\varphi; \nu, \mathcal{G})$  is undefined.

DEFINITION 1.3'. With the above notation, let

$$(1.7) \quad I_\mu(\varphi; \nu, \mathcal{G}) = \sup_{\mathcal{A} \in \Pi_C(\mathcal{G})} I_\mu(\varphi; \nu, \mathcal{A}),$$

in case the latter quantity is defined, with the convention that if there exists an  $\mathcal{A}_0 \in \Pi_C(\mathcal{G})$  with  $I_\mu(\varphi^-; \nu, \mathcal{A}_0) > -\infty$ , then  $\sup I_\mu(\varphi; \nu, \mathcal{A}) > -\infty$ , even though  $I_\mu(\varphi; \nu, \mathcal{A})$  might be undefined for some  $\mathcal{A}$ .

**THEOREM 1.2.** *If  $\mu$  is  $\sigma$ -finite and  $\varphi$  is continuous convex on  $[0, \infty)$ , then*

$$(1.8) \quad I_\mu(\varphi; \nu, \mathcal{F}) = \int_\Omega \varphi(f) d\mu,$$

whenever the integral exists.

**REMARK 1.2.** Note that if  $\mu(\Omega) < \infty$ , we obtain the same result whether we use finite partitions or countable partitions.

**REMARK 1.3.** In these two remarks and in the next section, we shall be concerned with the function

$$(1.9) \quad \begin{aligned} \varphi_0(x) &= x \log x, & x > 0, \\ &= 0, & x = 0, \end{aligned}$$

and shall write simply  $I_\mu(\nu, \mathcal{F})$  for  $I_\mu(\varphi; \nu, \mathcal{F})$ ; we shall refer to it as the “ $\mu$ -information of  $\nu$  in  $\mathcal{F}$ ”. When  $\mu, \nu$  are both probability measures,  $I_\mu(\nu, \mathcal{F})$  is the “information of  $\nu$  with respect to  $\mu$  in  $\mathcal{F}$ ,” or Kullback-Leibler information.

**REMARK 1.4.** Theorem 1.2 provides us with one justification for regarding Wiener’s information as “a reasonable measure” of the amount of information. Suppose we are concerned with a statistical experiment whose outcome is known to be real-valued. In the absence of any additional information, it is heuristically reasonable to regard the outcome as being equidistributed on the real line in the sense that the “probability” of the outcome belonging to any Borel set is proportional to the Lebesgue measure of the set. Suppose now that some knowledge about the experiment suggests the hypothesis  $H$  that the outcome has a distribution with density function  $f$ . In actual practice, we are able to measure the outcome on countable scales of varying accuracy. Roughly speaking,  $\int \varphi_0[f(x)] dx$ , if it exists, is then the maximum amount of information that can be obtained from the experiment, by the use of countable scales of measurement, for discriminating in favor of  $H$  against the hypothesis of equidistribution. Unfortunately, Wiener’s information suffers from the serious handicap that the integral does not exist for every density function  $f$ .

**PROOF OF THEOREM 1.1.** To begin with, let us observe that the value of the integral in (1.4) is either a finite number or  $+\infty$ . This follows from the fact that, on account of convexity,  $\varphi$  has a line of support  $ax + b$ , so that

$$\int_\Omega \varphi(f) d\mu \geq a\nu(\Omega) + b\mu(\Omega).$$

As another important consequence of convexity, we have

$$(1.10) \quad I_\mu(\varphi; \nu, \mathcal{A}) \leq I_\mu(\varphi; \nu, \mathcal{B})$$

whenever  $\mathcal{A}, \mathcal{B} \in \Pi_{\mathcal{F}}(\mathcal{F})$  and  $\mathcal{A} \subset \mathcal{B}$ . This follows from the fact that every atom of  $\mathcal{A}$  is a union of atoms of  $\mathcal{B}$  and from Jensen’s inequality. One more conse-

quence of Jensen's inequality is

$$(1.11) \quad \varphi[\nu(A)/\mu(A)]\mu(A) = \varphi[\text{av}_A f(\omega)]\mu(A) \leq \int_A \varphi[f(\omega)]\mu(d\omega),$$

where  $\text{av}_A$  stands for "average over  $A$  with weight-function  $d\mu/\mu(A)$ ," and hence,

$$(1.12) \quad I_\mu(\varphi; \nu, \mathcal{A}) \leq \int_\Omega \varphi(f) d\mu.$$

Taking the supremum over all  $\mathcal{A} \in \Pi_{\mathcal{F}}(\mathcal{F})$ , we have

$$(1.13) \quad I_\mu(\varphi; \nu, \mathcal{F}) \leq \int_\Omega \varphi(f) d\mu.$$

In order to establish the equality, we shall use the following fact: If  $h$  is either non-negative and measurable or bounded and measurable, then

$$\int_\Omega h d\mu = \sup_{\mathcal{A} \in \Pi_{\mathcal{F}}(\mathcal{F})} \sum_{A \in \mathcal{A}} [\inf_A h(\omega)]\mu(A).$$

Note that the convexity of  $\varphi$  implies the existence of a partition of  $[0, \infty)$  into four intervals  $J_i, i = 1, \dots, 4$ , such that  $\varphi$  is non-negative and non-increasing on  $J_1$ , non-positive and non-increasing on  $J_2$ , non-positive and non-decreasing on  $J_3$  and non-negative and non-decreasing on  $J_4$ . Some of these intervals may, of course, be empty. Let  $K_i = f^{-1}(J_i), i = 1, \dots, 4$ .

Given any partition  $\mathcal{A}$  of  $\Omega$ , let  $\mathcal{A}_i = \mathcal{A} \cap K_i, i = 1, \dots, 4$ . Since  $\{\mathcal{A}_i, i = 1, \dots, 4\}$  is a finer partition of  $\Omega$  than  $\mathcal{A}$ , we have

$$(1.14) \quad I_\mu(\varphi; \nu, \mathcal{A}) \leq I_\mu[\varphi; \nu, \{\mathcal{A}_i, i = 1, \dots, 4\}] = \sum_i I_\mu(\varphi; \nu, \mathcal{A}_i).$$

Hence,

$$I_\mu(\varphi; \nu, \mathcal{F}) = \sup_{\mathcal{A}} \sum_i I_\mu(\varphi; \nu, \mathcal{A}_i) = \sum_i \sup_{\mathcal{A}_i} I_\mu(\varphi; \nu, \mathcal{A}_i).$$

Now notice that if  $A \in \mathcal{A}_1 \cup \mathcal{A}_2$ , we have

$$\inf_A \varphi[f(\omega)] = \varphi[\sup_A f(\omega)] \leq \varphi[\text{av}_A f(\omega)],$$

and if  $A \in \mathcal{A}_3 \cup \mathcal{A}_4$ , we have

$$\inf_A \varphi[f(\omega)] = \varphi[\inf_A f(\omega)] \leq \varphi[\text{av}_A f(\omega)].$$

But  $\mathcal{A}_i$  is a finite measurable partition of  $K_i$  and  $\varphi[f(\omega)] \geq 0$  for  $\omega \in K_1 \cup K_4$ . Therefore,

$$(1.15) \quad \int_{K_1 \cup K_4} \varphi(f) d\mu = \sup_{\mathcal{A}} \sum_{A \in \mathcal{A}_1 \cup \mathcal{A}_4} [\inf_A \varphi(f)]\mu(A) \leq \sup_{\mathcal{A}} I_\mu(\varphi; \nu, \mathcal{A}_1 \cup \mathcal{A}_4).$$

Next suppose  $\varphi$  to be bounded below. Then  $\varphi$  is a bounded function on  $K_2 \cup K_3$ . Hence,

$$(1.16) \quad \int_{K_2 \cup K_3} \varphi(f) d\mu = \sup_{\mathcal{A}} \sum_{A \in \mathcal{A}_2 \cup \mathcal{A}_3} [\inf_A \varphi(f)]\mu(A) \leq \sup_{\mathcal{A}} I_\mu(\varphi; \nu, \mathcal{A}_2 \cup \mathcal{A}_3).$$

Thus, in case  $\varphi$  is bounded below, we have

$$(1.17) \quad \int_{\Omega} \varphi(f) d\mu \leq I_{\mu}(\varphi; \nu, \mathfrak{F}).$$

Inequalities (1.13) and (1.17) prove the theorem in case  $\varphi$  is bounded below.

Finally, suppose  $\varphi$  is unbounded below; then it is bounded above, and the integral is finite. We shall now see that  $I_{\mu}(\varphi; \nu, \mathfrak{F})$  is arbitrarily close to the integral. For this purpose, let  $-\alpha x + \beta (\alpha > 0)$  be a line of support for  $\varphi(x)$  at  $x = c_0$  for some  $c_0$  such that  $\varphi(c_0) < 0$ ; then  $\varphi(x) \geq -\alpha x + \beta$  for  $c > c_0$ . Let  $B(c) = \{\omega: f(\omega) > c\}$  and note that the finiteness of  $\nu(\Omega)$  and  $\int \varphi(f) d\mu$  and the fact that  $\varphi$  is non-increasing permit us to choose  $c$  so that  $|\int_{B(c)} \varphi(f) d\mu|$  and  $\nu[B(c)]$  are arbitrarily small. Further, for  $c > c_0$ , we have

$$(1.18) \quad \begin{aligned} |\varphi(\text{av}_{B(c)} f)|\mu[B(c)] &\leq [\alpha \text{av}_{B(c)} f] + |\beta| \mu[B(c)] \\ &= \alpha \nu[B(c)] + |\beta| \mu[B(c)] \leq (\alpha + |\beta|/c) \nu[B(c)]. \end{aligned}$$

Now, consider any partition of the form  $\mathfrak{A} = \{B(c), \mathfrak{A}_0\}$ , where  $\mathfrak{A}_0$  is a partition of  $\Omega - B(c)$ . On  $\Omega - B(c)$ ,  $\varphi(f)$  is bounded and hence, as in (1.16), we have

$$(1.19) \quad \int_{\Omega - B(c)} \varphi(f) d\mu \leq \sup_{\mathfrak{A}_0} I_{\mu}(\varphi; \nu, \mathfrak{A}_0).$$

Hence,

$$(1.20) \quad \begin{aligned} I_{\mu}(\varphi; \nu, \mathfrak{F}) &\geq \sup_{\mathfrak{A}_0} I_{\mu}(\varphi; \nu, \mathfrak{A}_0) + \varphi(\text{av}_{B(c)} f) \mu[B(c)] \\ &\geq \int_{\Omega} \varphi(f) d\mu - \int_{B(c)} \varphi(f) d\mu + \varphi(\text{av}_{B(c)} f) \mu[B(c)]. \end{aligned}$$

By (1.18), we can choose  $c$  so that the last two terms are arbitrarily small. Hence, (1.20) together with (1.13) completes the proof.

PROOF OF THEOREM 1.2. To begin with, observe that we have  $\mu(A) < \infty$  for every atom of  $\mathfrak{A} \in \Pi_c(\mathfrak{F})$ , and hence  $\varphi(\text{av}_A f) \leq \int_A \varphi(f) d\mu$  by Jensen's inequality. Consequently,

$$(1.21) \quad I_{\mu}(\varphi; \nu, \mathfrak{A}) \leq \int_{\Omega} \varphi(f) d\mu.$$

Now suppose  $\mu(\Omega) < \infty$ , and let  $\mathfrak{A}'$  be a finite partition  $\subset \mathfrak{A}$ . Then Jensen's inequality yields

$$(1.22) \quad I_{\mu}(\varphi; \nu, \mathfrak{A}') \leq I_{\mu}(\varphi; \nu, \mathfrak{A}).$$

Hence, (1.8) follows from (1.21) and Theorem 1.1.

Next let  $\mu(\Omega) = \infty$  and suppose  $\int_{\Omega} \varphi(f) d\mu = -\infty$ . Then (1.8) follows immediately from (1.21). Thus it remains to consider only the case  $\int_{\Omega} \varphi(f) d\mu > -\infty$ .

For any  $A \in \mathfrak{F}$ , let  $A^{\alpha} = A \cap \Omega^{\alpha}$ ,  $\alpha = +, -$ , and for any  $\mathfrak{A} \in \Pi_c(\mathfrak{F})$ , let  $\mathfrak{A}^{\alpha} = \mathfrak{A} \cap \Omega^{\alpha}$ ,  $\alpha = +, -$ . Let  $\mathfrak{B} = \{B_n, n = 1, 2, \dots\} \in \Pi_c(\mathfrak{F})$ . Then  $\int_{\Omega^-} \varphi(f) d\mu = \sum_n \int_{B_n^-} \varphi(f) d\mu$ . Since  $\mu(B_n^-) < \infty$ , we have  $\sup_{\mathfrak{A}} I_{\mu}(\varphi; \nu, \mathfrak{A} \cap B_n^-) = \int_{B_n^-} \varphi(f) d\mu$ ,  $n = 1, 2, \dots$ , and  $I_{\mu}(\varphi; \nu, \mathfrak{A}^-) \leq I_{\mu}(\varphi; \nu, \mathfrak{A} \cap \mathfrak{B}^-) = \sum_n I_{\mu}(\varphi; \nu, \mathfrak{A} \cap B_n^-)$ . Hence  $\sup_{\mathfrak{A}^-} I_{\mu}(\varphi; \nu, \mathfrak{A}^-) = \int_{\Omega^-} \varphi(f) d\mu$ , which is assumed finite. Therefore, there exists an  $\mathfrak{A} \in \Pi_c$  such that  $I_{\mu}(\varphi; \nu, \mathfrak{A}^-) > -\infty$ , and so  $I_{\mu}(\varphi; \nu, \mathfrak{F})$  is defined and  $> -\infty$ . Now, let  $\mathfrak{A} \in \Pi_c(\mathfrak{F})$  with  $I_{\mu}(\varphi; \nu, \mathfrak{A}^-) > -\infty$ .

Then

$$(1.23) \quad I_\mu(\varphi; \nu, \mathfrak{A}) \leq I_\mu(\varphi; \nu, \mathfrak{A}^+) + I_\mu(\varphi; \nu, \mathfrak{A}^-).$$

Consequently,

$$(1.24) \quad I_\mu(\varphi; \nu, \mathfrak{F}) = \sup I_\mu(\varphi; \nu, \mathfrak{A}^+) + \sup I_\mu(\varphi; \nu, \mathfrak{A}^-).$$

We have already seen that the last quantity =  $\int_{\Omega^-} \varphi(f) d\mu$  and by a similar argument, we deduce that  $\sup I_\mu(\mathfrak{A}^+) = \int_{\Omega^+} \varphi(f) d\mu$ , thus completing the proof.

**2. Conditional information and sufficiency.** In this section, we shall be concerned only with the function  $\varphi_0$  as defined in (1.9) and mainly with the case in which  $\mu$  and  $\nu$  are probability measures, i.e., with Kullback-Leibler information. It was proved by Kullback and Leibler in a somewhat limited context that if  $\mathfrak{G}$  is a sub- $\sigma$ -field of  $\mathfrak{F}$ , then  $I_\mu(\nu, \mathfrak{F}) \geq I_\mu(\nu, \mathfrak{G})$  with equality holding in case  $I_\mu(\nu, \mathfrak{F})$  is finite if and only if  $\mathfrak{G}$  is a sufficient sub-field of  $\mathfrak{F}$  for the pair of probability measures  $\mu, \nu$ . It is easy to construct examples in which  $I_\mu(\nu, \mathfrak{F}) = I_\mu(\nu, \mathfrak{G}) = \infty$ , but  $\mathfrak{G}$  is not a sufficient sub-field. It might be of interest to obtain a relation between sufficiency and the amount of information which holds without the restriction that  $I_\mu(\nu, \mathfrak{F})$  be finite.

For this purpose, we study the notion of conditional information which has already been defined by Kullback ([7]) for the kind of probability spaces he dealt with. In order to obtain an appropriate definition of the conditional information  $I_\mu(\nu, \mathfrak{F} | \mathfrak{G})$  in general, we shall first establish an important property of conditional information when the fields involved are countable; this is done in Theorem 2.1 in which, for completeness and for future use, we permit  $\mu$  to be  $\sigma$ -finite. Theorem 2.2 states that  $I_\mu(\nu, \mathfrak{F} | \mathfrak{G}) = 0$  if and only if  $\mathfrak{G}$  is a sufficient sub-field of  $\mathfrak{F}$ ; finally, Theorem 2.3 states that  $I_\mu(\nu, \mathfrak{F}) = I_\mu(\nu, \mathfrak{G}) + I_\mu(\nu, \mathfrak{F} | \mathfrak{G})$ . These results extend to fairly general situations the corresponding results first obtained by Kullback and Leibler. Proofs of all these results are given at the end of the section.

Let  $\Omega, \mathfrak{F}, \mu, \nu$  be as in Section 1, and let  $\mathfrak{G}$  be a  $\sigma$ -field  $\subset \mathfrak{F}$ . For any  $\sigma$ -field  $\mathfrak{H}$ , we shall denote the class of  $\mathfrak{H}$ -measurable partitions simply by  $\Pi(\mathfrak{H})$  and shall understand that  $\Pi = \Pi_c$  if  $\mu(\Omega) = \infty$ , whereas if  $\mu(\Omega) < \infty$ , one may take  $\Pi = \Pi_f$ . Let  $A, B$  represent typical atoms respectively of  $\mathfrak{A}, \mathfrak{B} \in \Pi(\mathfrak{F})$ , and let  $\mathfrak{A} \vee \mathfrak{B} = \{A \cap B, A \varepsilon \mathfrak{A}, B \varepsilon \mathfrak{B}\}$ . Then

$$(2.1) \quad I_\mu(\nu, \mathfrak{A}) = \sum \varphi_0[\nu(A)/\mu(A)]\mu(A) = \sum \log [\nu(A)/\mu(A)]\nu(A),$$

and

$$(2.2) \quad I_\mu(\nu, \mathfrak{A} \vee \mathfrak{B}) = I_\mu(\nu, \mathfrak{B}) + I_\mu(\nu, \mathfrak{A} | \mathfrak{B}),$$

where

$$(2.3) \quad I_\mu(\nu, \mathfrak{A} | \mathfrak{B}) = \sum_B \{ \sum_A \varphi_0[\nu(A | B)/\mu(A | B)]\mu(A | B) \} \nu(B).$$

Here we have used the notation of conditional probability and denoted  $\mu(A \cap B)/\mu(B)$  by  $\mu(A | B)$  when  $\mu(B) > 0$ . Note that if  $\mathfrak{G} \subset \mathfrak{F}$  is a  $\sigma$ -field and

$F \in \mathcal{F}$ , then  $\mu(F \cap G)$ ,  $G \in \mathcal{G}$ , is a measure on  $\mathcal{G}$  which is absolutely continuous with respect to  $\mu$  and has a Radon-Nikodym derivative  $\mu^{\mathcal{G}}(F, \omega)$  with respect to  $\mu$  on  $\mathcal{G}$ . Hence,  $\mu(A | B)$  is the value of  $\mu^{\mathcal{B}}(A, \omega)$  for  $\omega \in B$ ; note that  $\mu^{\mathcal{B}}$  and  $\nu^{\mathcal{B}}$  are probability measures on  $\mathcal{A}$  for almost every  $\omega$  [ $\nu$ ]. The expression in  $\{ \}$  in (2.3) is then the  $\mu^{\mathcal{B}}$ -information of  $\nu^{\mathcal{B}}$  in  $\mathcal{A}$ . Denoting it by  $I_{\mu}^{\mathcal{B}}(\nu, \mathcal{A}, \omega)$ , or simply  $I_{\mu}^{\mathcal{B}}(\nu, \mathcal{A})$ , and using  $\varepsilon_{\nu}$  to denote the integral with respect to  $\nu$  over  $(\Omega, \mathcal{B})$ , we can write (2.3) as

$$(2.4) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{B}) = \varepsilon_{\nu} I_{\mu}^{\mathcal{B}}(\nu, \mathcal{A}) = \int_{\Omega} \sum \varphi_0[\nu^{\mathcal{B}}(A)/\mu^{\mathcal{B}}(A)]\mu^{\mathcal{B}}(A) d\nu.$$

Notice that, for almost every  $\omega$ ,  $I_{\mu}^{\mathcal{B}}(\nu, \mathcal{A}, \omega)$  is defined and non-negative even if  $\mu$  is not totally finite. The same holds true if  $\mathcal{B}$  is replaced by any  $\sigma$ -field  $\mathcal{G} \subset \mathcal{F}$ , and we have

$$(2.5) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{G}) = \varepsilon_{\nu} I_{\mu}^{\mathcal{G}}(\nu, \mathcal{A})$$

defined and non-negative for every  $\mathcal{A} \in \Pi(\mathcal{F})$ . Properties of this conditional information analogous to those of conditional entropy can be easily derived (for a good discussion of conditional entropy, see Billingsley [1]). One such property which we shall need is monotonicity in  $\mathcal{A}$ ; this follows from the fact that, for almost every  $\omega$ ,  $I_{\mu}^{\mathcal{G}}(\nu, \mathcal{A})$  is the information of one probability measure with respect to another. Hence,  $\mathcal{A}, \mathcal{B} \in \Pi(\mathcal{F})$  and  $\mathcal{A} \supset \mathcal{B}$  imply  $I_{\mu}^{\mathcal{G}}(\nu, \mathcal{A}, \omega) \geq I_{\mu}^{\mathcal{G}}(\nu, \mathcal{B}, \omega)$  a.e. on  $(\Omega, \mathcal{G}, \nu)$ . Consequently,

$$(2.6) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{G}) \geq I_{\mu}(\nu, \mathcal{B} | \mathcal{G}) \quad \text{for } \mathcal{A}, \mathcal{B} \in \Pi(\mathcal{F}), \quad \mathcal{A} \supset \mathcal{B}.$$

There is an important monotonicity property possessed by conditional entropy which conditional information does not have; this is monotonicity in the conditioning field. Whereas in the case of entropy,  $\mathcal{G} \subset \mathcal{F}$  implies  $H(\mathcal{A} | \mathcal{G}) \geq H(\mathcal{A} | \mathcal{F})$ , it is not true that  $I_{\mu}(\nu, \mathcal{A} | \mathcal{G}) \geq I_{\mu}(\nu, \mathcal{A} | \mathcal{F})$ ; an inequality of this type which does hold is given by

**THEOREM 2.1.** *If  $(\Omega, \mathcal{F}, \mu)$  is a  $\sigma$ -finite measure-space,  $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \Pi(\mathcal{F})$  and  $\mathcal{A} \supset \mathcal{B} \supset \mathcal{C}$ , then*

$$(2.7) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{C}) = I_{\mu}(\nu, \mathcal{A} | \mathcal{B}) + I_{\mu}(\nu, \mathcal{B} | \mathcal{C}) \geq I_{\mu}(\nu, \mathcal{A} | \mathcal{B}).$$

We shall now use these considerations to define conditional information  $I_{\mu}(\nu, \mathcal{F} | \mathcal{G})$  when  $\mathcal{F}, \mathcal{G}$  are  $\sigma$ -fields with  $\mathcal{G} \subset \mathcal{F}$ . Relations (2.6) and (2.7) suggest two ways of extending the notion of conditional information from finite or countable fields to arbitrary  $\sigma$ -fields; these are stated now.

**DEFINITION 2.1.** For  $\sigma$ -fields  $\mathcal{F}, \mathcal{G}$  with  $\mathcal{G} \subset \mathcal{F}$ , let  $I_{\mu}(\nu, \mathcal{A} | \mathcal{G})$  be as defined by (2.5), and let

$$(2.8) \quad I_{\mu}(\nu, \mathcal{F} | \mathcal{G}) = \sup_{\mathcal{A} \in \Pi(\mathcal{F})} I_{\mu}(\nu, \mathcal{A} | \mathcal{G}).$$

The quantity so defined will be called "the conditional  $\mu$ -information of  $\nu$  in  $\mathcal{F}$  relative to  $\mathcal{G}$ ."

Alternatively, we could set

$$(2.9) \quad I_{\mu}'(\nu, \mathcal{F} | \mathcal{G}) = \inf_{\mathcal{B} \in \Pi(\mathcal{G})} \sup_{\mathcal{A} \in \Pi(\mathcal{F})} I_{\mu}(\nu, \mathcal{A} | \mathcal{B}),$$

and define  $I_\mu'(\nu, \mathfrak{F} | \mathcal{G})$  as the conditional information. Under usual conditions, the two quantities will be seen to be equal; the reason for preferring (2.8) to (2.9) is that it leads immediately to

**THEOREM 2.2.** *Let  $M$  be a dominated family of probability measures on  $(\Omega, \mathfrak{F})$  and let  $\mathcal{G}$  be a  $\sigma$ -field  $\subset \mathfrak{F}$ . Then  $\mathcal{G}$  is a sufficient sub- $\sigma$ -field of  $\mathfrak{F}$  for the family  $M$  if and only if  $I_\mu(\nu, \mathfrak{F} | \mathcal{G}) = 0$  for all  $\mu, \nu \in M$ .*

Finally, it remains to extend the relation (2.2) to the case of arbitrary  $\sigma$ -fields; this is done, under the slight restriction of separability of  $\mathcal{G}$ , in

**THEOREM 2.3.** *Let  $\nu$  be a probability measure on a probability space  $(\Omega, \mathfrak{F}, \mu)$ ,  $\nu \ll \mu$ , and let  $\mathcal{G}$  be a separable sub- $\sigma$ -field of  $\mathfrak{F}$ . If  $I_\mu(\nu, \mathfrak{F})$  is finite, then*

$$(2.10) \quad I_\mu(\nu, \mathfrak{F} | \mathcal{G}) = I_\mu(\nu, \mathfrak{F}) - I_\mu(\nu, \mathcal{G}).$$

**REMARK 2.1.** Combining Theorems 2.2 and 2.3, we now see that if  $I_\mu(\nu, \mathfrak{F})$  is finite and  $\mathcal{G}$  is separable, then  $\mathcal{G}$  is a sufficient sub-field of  $\mathfrak{F}$  for the dominated family  $M$  of probability measures if and only if  $I_\mu(\nu, \mathfrak{F}) = I_\mu(\nu, \mathcal{G})$  for all  $\mu, \nu \in M$ .

**REMARK 2.2.** Note that  $\mathfrak{B} \in \Pi(\mathcal{G})$  implies  $I_\mu(\nu, \mathfrak{F} | \mathfrak{B}) = I_\mu(\nu, \mathfrak{F}) - I_\mu(\nu, \mathfrak{B})$ , and taking the infimum over  $\Pi(\mathcal{G})$  we get from (2.9)

$$(2.11) \quad I_\mu'(\nu, \mathfrak{F} | \mathcal{G}) = I_\mu(\nu, \mathfrak{F}) - I_\mu(\nu, \mathcal{G}).$$

Consequently, when the assumptions of Theorem 2.3 are satisfied, we have

$$(2.12) \quad I_\mu(\nu, \mathfrak{F} | \mathcal{G}) = I_\mu'(\nu, \mathfrak{F} | \mathcal{G}).$$

**REMARK 2.3.** Finally, suppose that  $(\Omega, \mathfrak{F})$  is a Lusin space (Blackwell [2]). Then, for every  $\omega' \in \Omega$ ,  $\mu^{\mathcal{G}}(\cdot, \omega')$  and  $\nu^{\mathcal{G}}(\cdot, \omega')$  are probability measures on  $(\Omega, \mathfrak{F})$ . Suppose further that, for every  $\omega' \in \Omega$ ,  $\nu^{\mathcal{G}}(\cdot, \omega') \ll \mu^{\mathcal{G}}(\cdot, \omega') \ll \mu$  on  $(\Omega, \mathfrak{F})$ . Then there exist functions  $h(\omega, \omega')$  and  $m(\omega, \omega')$  which are measurable on  $(\Omega \times \Omega, \mathfrak{F} \times \mathcal{G})$  such that, for each  $\omega'$ ,  $h$  and  $m$  are respectively the Radon-Nikodym derivatives  $d\nu^{\mathcal{G}}/d\mu^{\mathcal{G}}$  and  $d\mu^{\mathcal{G}}/d\mu$  (Meyer [8], p. 154). We then have

$$(2.13) \quad I_\mu(\nu, \mathfrak{F} | \mathcal{G}) = \int_{\Omega \times \Omega} \varphi_0(h)m d(\mu \times \nu)_{\mathfrak{F} \times \mathcal{G}},$$

which is a generalization of the formula given by Kullback ([7], p. 13).

**PROOF OF THEOREM 2.1.** Since  $\mathfrak{A} \vee \mathfrak{B} = \mathfrak{A} \vee \mathfrak{C} = \mathfrak{A}$ , the assertion is an immediate consequence of (1.10) and (2.2) if  $I_\mu(\nu, \mathfrak{A})$  is finite. It is proven in general by direct computation. For this we first note that every atom of  $\mathfrak{C}$  is a union of atoms of  $\mathfrak{B}$ , and each atom of  $\mathfrak{B}$  is a union of atoms of  $\mathfrak{A}$ . For simplicity of notation, we shall denote typical atoms of  $\mathfrak{A}, \mathfrak{B}, \mathfrak{C}$  respectively by  $A, B, C$  without suffixes. Let

$$(2.14) \quad \begin{aligned} J(\mathfrak{A}, \mathfrak{C}) &= \{ \sum_A \varphi_0[\nu(A | C)/\mu(A | C)]\mu(A | C) \} \nu(C) \\ &= \{ \sum_{\{A: A \subset C\}} \varphi_0[(\nu(A)/\nu(C))/(\mu(A)/\mu(C))]\mu(A)/\mu(C) \} \nu(C) \\ &= \sum_{\{A: A \subset C\}} \{ \log [\nu(A)/\mu(A)] - \log [\nu(C)/\mu(C)] \} \nu(A) \\ &= I_\mu(\nu, \mathfrak{A} \cap C) - \log [\nu(C)/\mu(C)]\nu(C), \end{aligned}$$

and define  $J(\mathfrak{A}, B), J(\mathfrak{B}, C)$  similarly. Note that all of these are non-negative and that

$$(2.15) \quad I_\mu(\nu, \mathfrak{A} | \mathfrak{C}) = \sum_C J(\mathfrak{A}, C) = \sum_C \{I_\mu(\nu, \mathfrak{A} \cap C) - \log [\nu(C)/\mu(C)]\nu(C)\}$$

and

$$(2.16) \quad \begin{aligned} I_\mu(\nu, \mathfrak{B} | \mathfrak{C}) &= \sum_C J(\mathfrak{B}, C) \\ &= \sum_C \{I_\mu(\nu, \mathfrak{B} \cap C) - \log [\nu(C)/\mu(C)]\nu(C)\}. \end{aligned}$$

Now,  $I_\mu(\nu, \mathfrak{A} \cap C) \geq I_\mu(\nu, \mathfrak{B} \cap C)$  since  $\mathfrak{A} \supset \mathfrak{B}$ . Hence,  $I_\mu(\nu, \mathfrak{B} \cap C) = \infty$  for some  $C$  implies that  $I_\mu(\nu, \mathfrak{B} | \mathfrak{C}) = I_\mu(\nu, \mathfrak{A} | \mathfrak{C}) = \infty$ , so that (2.7) holds. Suppose therefore that  $I_\mu(\nu, \mathfrak{B} \cap C) < \infty$  for all  $C \in \mathfrak{C}$ . Then

$$(2.17) \quad \begin{aligned} I_\mu(\nu, \mathfrak{A} | \mathfrak{B}) &= \sum_B J(\mathfrak{A}, B) = \sum_C \sum_{\{B: B \subset C\}} J(\mathfrak{A}, B) \\ &= \sum_C \sum_{\{B: B \subset C\}} \{I_\mu(\nu, \mathfrak{A} \cap B) - \log [\nu(B)/\mu(B)]\nu(B)\} \\ &= \sum_C \{I_\mu(\nu, \mathfrak{A} \cap C) - I_\mu(\nu, \mathfrak{B} \cap C)\}. \end{aligned}$$

Again if  $I_\mu(\nu, \mathfrak{A} \cap C) = \infty$  for some  $C$ , then  $I_\mu(\nu, \mathfrak{A} | \mathfrak{C}) = I_\mu(\nu, \mathfrak{A} | \mathfrak{B}) = \infty$ , so that (2.7) holds. Hence let us also assume that  $I_\mu(\nu, \mathfrak{A} \cap C) < \infty$  for all  $C$ . The relation (2.7) then follows from (2.15), (2.16), (2.17) by noting that  $I_\mu(\nu, \mathfrak{A} | \mathfrak{C}), I_\mu(\nu, \mathfrak{A} | \mathfrak{B})$  and  $I_\mu(\nu, \mathfrak{B} | \mathfrak{C})$  are sums of series with non-negative terms and the typical term of the first is the sum of the corresponding terms of the second and third series.

PROOF OF THEOREM 2.2. Since  $M$  is a dominated family,  $\mathfrak{G}$  is a sufficient sub-field if and only if it is pair-wise sufficient (Halmos and Savage [4]); and, by definition of sufficiency,  $\mathfrak{G}$  is a sufficient sub- $\sigma$ -field of  $\mathfrak{F}$  for the pair  $(\mu, \nu)$  of probability measures if and only if, for every  $F \in \mathfrak{F}, \mu^{\mathfrak{G}}(F, \omega) = \nu^{\mathfrak{G}}(F, \omega)$  for almost every  $\omega$ .

Now suppose  $\mathfrak{G}$  is a sufficient sub-field. Then, for every  $\mathfrak{A} \in \Pi(\mathfrak{F}), I_\mu^{\mathfrak{G}}(\nu, \mathfrak{A}, \omega) = 0$  for almost every  $\omega$ . Thus, by (2.5),  $I_\mu(\nu, \mathfrak{A} | \mathfrak{G}) = 0$  and hence  $I_\mu(\nu, \mathfrak{F} | \mathfrak{G}) = 0$ .

Conversely, if  $I_\mu(\nu, \mathfrak{F} | \mathfrak{G}) = 0$ , then  $I_\mu(\nu, \mathfrak{A} | \mathfrak{G}) = 0$  for every  $\mathfrak{A} \in \Pi(\mathfrak{F})$ . But this implies that, for every  $\mathfrak{A} \in \Pi(\mathfrak{F}), I_\mu^{\mathfrak{G}}(\nu, \mathfrak{A}, \omega) = 0$  for almost every  $\omega$ , since this quantity is almost everywhere non-negative. But the  $\mu^{\mathfrak{G}}$ -information of  $\nu^{\mathfrak{G}}$  contained in the finite field  $\mathfrak{A}$  is zero if and only if the two measures agree on  $\mathfrak{A}$ ; i.e., if and only if  $\mu^{\mathfrak{G}}(F, \omega) = \nu^{\mathfrak{G}}(F, \omega)$  for almost every  $\omega$ .

PROOF OF THEOREM 2.3. Since  $\mathfrak{G}$  is separable, there exists an increasing sequence  $\{\mathfrak{B}_n\}$  of finite partitions such that  $\mathfrak{G}$  is the  $\sigma$ -field generated by  $\cup \mathfrak{B}_n$ . For each  $F \in \mathfrak{F}$ , we have  $\{\mu^{\mathfrak{B}_n}(F), \nu^{\mathfrak{B}_n}(F)\} \rightarrow \{\mu^{\mathfrak{G}}(F), \nu^{\mathfrak{G}}(F)\}$  a.e.  $[\nu]$  (Billingsley [1]).

For each atom  $A$  of a finite  $\mathfrak{F}$ -measurable partition  $\mathfrak{A}$  of  $\Omega$ , let  $M_A = \{\omega: \mu^{\mathfrak{G}}(A, \omega) > 0\}, N_A = \{\omega: \nu^{\mathfrak{G}}(F, \omega) > 0\}$ . We may assume that  $M_A^c \cap N_A^c \in \mathfrak{B}_n$  for all  $A \in \mathfrak{A}$  and all  $n$ ; since if this is not the case for some  $n$ , we can replace  $\mathfrak{B}_n$  by a finer partition  $\mathfrak{B}'_n$  obtained by sub-dividing each atom of  $\mathfrak{B}_n$  into its intersections with the  $M_A^c \cap N_A^c, A \in \mathfrak{A}$ , and their complements, which are all

$\mathcal{G}$ -measurable. Then we have:  $\mu^{\mathcal{G}}(A, \omega) = 0 = \nu^{\mathcal{G}}(A, \omega)$  for  $\omega \in M_A^c \cap N_A^c$ ; and since  $\{\mu^{\mathcal{B}_n}(A), \nu^{\mathcal{B}_n}(A)\} = \mathcal{E}[\{\mu^{\mathcal{G}}(A), \nu^{\mathcal{G}}(A)\} | \mathcal{B}_n]$ , also  $\mu^{\mathcal{B}_n}(A, \omega) = 0 = \nu^{\mathcal{B}_n}(A, \omega)$  a.e.  $[\nu]$  on  $M_A^c \cap N_A^c$ . Further,  $\mu(M_A^c \cap A) = \int_{M_A^c} \mu^{\mathcal{G}}(A) d\mu = 0 \Rightarrow \nu(M_A^c \cap A) = 0$ ; i.e.,  $\int_{M_A^c} \nu^{\mathcal{G}}(A) d\nu = 0$ , and hence,  $\nu(M_A^c \cap N_A) = 0$ . Finally,

$$\nu^{\mathcal{B}_n}(A, \omega) / \mu^{\mathcal{B}_n}(A, \omega) \rightarrow \nu^{\mathcal{G}}(A, \omega) / \mu^{\mathcal{G}}(A, \omega) \text{ a.e. } [\nu] \text{ on } M_A.$$

Now, let  $h_n(A, \omega) = \varphi_0[\nu^{\mathcal{B}_n}(A, \omega) / \mu^{\mathcal{B}_n}(A, \omega)] \mu^{\mathcal{B}_n}(A, \omega)$  and  $h(A, \omega) = \varphi_0[\nu^{\mathcal{G}}(A, \omega) / \mu^{\mathcal{G}}(A, \omega)] \mu^{\mathcal{G}}(A, \omega)$ . From the preceding considerations, we have

$$h_n(A, \omega) \rightarrow h(A, \omega) \text{ a.e. on } (\Omega, \mathcal{G}, \nu) \text{ for every } A \in \mathcal{A},$$

and hence,  $I_{\mu}^{\mathcal{B}_n}(\nu, \mathcal{A}, \omega) = \sum_{\alpha} h_n(A, \omega) \rightarrow \sum_{\alpha} h(A, \omega) = I_{\mu}^{\mathcal{G}}(\nu, \mathcal{A}, \omega)$  a.e. on  $(\Omega, \mathcal{G}, \nu)$ . From (2.3) and (2.4), we then have

$$(2.18) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{B}_n) = \int I_{\mu}^{\mathcal{B}_n}(\nu, \mathcal{A}) g_n d\mu,$$

where  $g_n = \mathcal{E}\{f | \mathcal{B}_n\}$  is a uniformly integrable martingale converging to  $g = \mathcal{E}\{f | \mathcal{G}\} = d\nu/d\mu[\mathcal{G}]$ . Now,  $I_{\mu}^{\mathcal{B}_n}(\nu, \mathcal{A}) g_n$  is a  $\mathcal{B}_n$ -measurable function whose value on a typical atom  $B^n$  of  $\mathcal{B}_n$  is

$$\sum_A \varphi_0[\nu(A \cap B^n) / \mu(A \cap B^n)] \mu(A \cap B^n) / \mu(B^n) - \varphi_0[\nu(B^n) / \mu(B^n)].$$

Using Doob's application of martingale theory to the theory of derivatives, as done by Kallianpur ([5], p. 269), we see that  $I_{\mu}(\nu, \mathcal{A} | \mathcal{B}_n)$  is the integral of a uniformly integrable sequence which converges a.e. to  $I_{\mu}^{\mathcal{G}}(\nu, \mathcal{A})g$ . Therefore, we have

$$(2.19) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{B}_n) \rightarrow \int I_{\mu}^{\mathcal{G}}(\nu, \mathcal{A})g d\mu = I_{\mu}(\nu, \mathcal{A} | \mathcal{G}).$$

On the other hand, from (2.2) we have

$$(2.20) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{B}_n) = I_{\mu}(\nu, \mathcal{A} \vee \mathcal{B}_n) - I_{\mu}(\nu, \mathcal{B}_n) \rightarrow I_{\mu}(\nu, \mathcal{A} \vee \mathcal{G}) - I_{\mu}(\nu, \mathcal{G}).$$

Combining (2.19) and (2.20), we get

$$(2.21) \quad I_{\mu}(\nu, \mathcal{A} | \mathcal{G}) = I_{\mu}(\nu, \mathcal{A} \vee \mathcal{G}) - I_{\mu}(\nu, \mathcal{G}),$$

and taking the supremum over all  $\mathcal{A} \in \Pi(\mathcal{F})$ , we obtain (2.10).

REFERENCES

[1] BILLINGSLEY, P. (1965). *Ergodic Theory and Information*. Wiley, New York.  
 [2] BLACKWELL, D. (1956). On a class of probability spaces. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **2** 1-6. Univ. of California Press.  
 [3] GEL'FAND, I. M., KOLMOGOROV, A. N. and YAGLOM, A. M. (1956). On the general definition of the amount of information. *Dokl. Akad. Nauk SSSR* **111** 745-748.  
 [4] HALMOS, P. R. and SAVAGE, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225-241.

- [5] KALLIANPUR, G. (1960). On the amount of information contained in a  $\sigma$ -field. *Contributions to Probability and Statistics—Essays in Honor of Harold Hotelling*. Stanford University Press.
- [6] KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22** 79–86.
- [7] KULLBACK, S. (1959). *Information Theory and Statistics*. Wiley, New York.
- [8] MEYER, P. A. (1966). *Probability and Potentials*. Blaisdell, New York.
- [9] SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379–423, 623–656.
- [10] WIENER, N. (1948). *Cybernetics*. Wiley, New York.