CONTINUITY AND CHARACTERIZATION OF SHANNON-WIENER INFORMATION MEASURE FOR CONTINUOUS PROBABILITY DISTRIBUTIONS

BY SADAO IKEDA

(Received Dec. 10, 1959)

1. Introduction

C. E. Shannon [1] and N. Wiener [2] independently introduced a measure of information for probability distributions: $H=-C\sum_i p_i \log p_i$ in discrete case, and $H=-c\int_{-\infty}^{\infty} p(x)\log p(x)dx$ in continuous case, where $c>0$ is a scaling constant. Important properties of these measures are listed, for example, in C. E. Shannon and W. Weaver [3]. Characterization problem of these measures has been treated by A. I. Khinchin [4] in finite discrete case, and by H. Hatori [5] in continuous case. These authors proved the uniqueness of the above expressions under some resonable postulates.

Except for finite discrete case, the value of the information measure defined as above is not necessarily definite, and the continuity property will be complicated. In the present paper a sufficient condition for yielding the convergence of information measures will be given in the first, and the problem of characterization will be reexamined by postulating a suitable continuity assumption, in the second.

2. Continuity

Let (R, S, m) be a σ-finite measure space, i.e., let R be any abstract space, S a σ-field of subsets of R, and m a σ-finite measure on (R, S). Consider a random variable or a probability distribution, X, which is defined in the space R and is absolutely continuous with respect to the measure m. Let $p(x)$ be a generalized probability density function ($pdf.$) of X, and $D(X)=\{x\colon p(x)>0\}$. For any subsets A and B of the space R, the relation $A\subset B\,(m)$ means that the difference $A-B$ is a set of m-measure zero. If $p'(x)$ is another generalized $pdf.$ of X and $D'(X)=\{x\colon p'(x)>0\}$, then $D(X)=D'(X)(m)$. For any set E in S, $V(E)$ denotes the class of all absolutely continuous (m) probability distributions such that $D(X) \subset E(m)$.

Define, for X in $V(R)$,

$$H(X)=\int_R p(x) \log p(x)\, dm \ . \tag{2.1}$$

Let $D(X)=E$, and $E=E'+E''$ where

$$E=\{x\colon\ p(x)>1\},\ E''=\{x\colon\ 0<p(x)\leqq 1\} \ . \tag{2.2}$$

The expression (2.1) may be regarded as $H(X)=H'(X)+H''(X)$, where

$$H'(X)=\int_{E'} p(x) \log p(x)\, dm,\ \ H''(X)=\int_{E''} p(x) \log p(x)\, dm \ . \tag{2.3}$$

This definition of the Shannon-Wiener information measure is slightly different from the original one in its sign, scaling constant and basic measure space, but it will not matter much in the present study.

Let E be any set in S with positive m-measure. Consider a finite or countably infinite partition $Z=\{A_i\}\ (i=1, 2, \cdots)$ of E such that

$$E=\sum_i A_i(m),\ A_i \cap A_j=0(m),\ (i\neq j),\ 0<m(A_i)<\infty \ , \quad (i=1, 2, \cdots) \ , \tag{2.4}$$

which is called a (finite or countably infinite) m-partition of E. If a probability distribution $\overline{X}$ in $V(E)$ has a *pdf*. such as

$$\overline{p}(x)=p_i/v_i(m) \text{ on } A_i,\ (i=1, 2, \cdots),\ =0, \text{ otherwise}, \tag{2.5}$$

where $v_i=m(A_i)$, $p_i \geqq 0$, $(i=1, 2, \cdots)$, and $\sum_i p_i=1$, then $\overline{X}$ is said to be a simple probability distribution. Denote by $\overline{V}(E)$ the class of all simple probability distributions defined in E. For $\overline{X}$ in $\overline{V}(R)$ with *pdf*. such as (2.5), the definition (2.1) will become

$$H(\overline{X})=H'(\overline{X})+H''(\overline{X}) \ , \tag{2.6}$$

where

$$H'(\overline{X})=\sum{}' p_i \log\frac{p_i}{v_i} \ , \quad H''(\overline{X})=\sum{}'' p_i \log\frac{p_i}{v_i} \ ,$$

are series with positive and negative terms respectively.

First, two lemmas are stated.

Lemma 2.1

(i) *Let $x>0, x'\geqq 0$, and $z=x'/x$. Then it holds that*

$$|x\log x - x'\log x'| \leqq |1-z|(|x\log x| + x + x') .$$

(ii) *Let* $x \geqq 0$ *and* $x' \geqq 0$. *Then, for any* $\varepsilon > 0$, *there exists a* $\delta > 0$ *such that* $|x-x'| < \delta$ *implies*

$$|x\log x - x'\log x'| \leqq \varepsilon(x+1+\varepsilon) .$$

PROOF. Evidently (i) is valid when $x'=0$. Since

$$|\log z| \leqq \begin{cases} |1-z| , & (z \geqq 1) , \\ |1-z|/z, & (0<z<1) , \end{cases}$$

it holds that

$$|x\log x - x'\log x'| \leqq |1-z||x\log x| + x'|\log z|, \quad (z>0) ,$$

$$|1-z||x\log x| + x'|\log z| \leqq \begin{cases} |1-z||x\log x| + x'|1-z| , & (z \geqq 1) , \\ |1-z||x\log x| + x|1-z| , & (0<z<1) , \end{cases}$$

Hence

$$|x\log x - x'\log x'| \leqq |1-z|(|x\log x| + x + x') , \qquad (z>0) ,$$

which completes the proof of (i).

From the mean value theorem it follows, for $x', x \geqq 1$, that

$$|x\log x - x'\log x'| \leqq |x-x'| \cdot \max(x, x') .$$

From this and the uniform continuity of the function $x\log x$ on the interval $0 \leqq x \leqq 2$, it will be easily concluded that, for any $\varepsilon>0$, there exists a $\delta>0$ (less than ε), such that $|x-x'|<\delta$ implies

$$|x\log x - x'\log x'| \leqq \max(\varepsilon \max(x, x'), \varepsilon), (x, x' \geqq 0) ,$$

which assures of the validity of (ii), since $\max(x, x') \leqq x+\varepsilon$.

Hereafter throughout the paper, the expression "condition $C(x)(m)$", as in Lemma 2.2 (a)(i) below, means that the condition $C(x)$ is valid for all x on the set interested except for a set of m-measure zero. The following is the well-known Lebesgue's convergence theorem.

LEMMA 2.2 (Lebesgue) *Let* $f(x)$ *and* $\{f_i(x)\}$, $(i=1, 2, \cdots)$ *be real valued functions defined in a* σ-*finite measure space* (R, S, m). *Then the sufficient conditions for the convergence*

$$\int_R f_i(x)\, dm \to \int f(x)\, dm, \ (i \to \infty) ,$$

are given as follows:

(a) *For the case when* $-\infty < \int_R f(x)\,dm < \infty$, *the condition is*

(i) $f_i(x) \to f(x)(i \to \infty)(m)$, *in* R, *and*

(ii) *there exists a function* $K(x)$ *such that* $|f_i(x)| \leqq K(x)(m)$ *in* R, $(i=1, 2, \cdots)$, *and* $\int_R K(X)\,dm < \infty$.

(b) *For the case* $\int_R f(x)\,dm = \infty$,

(i) $f_i(x) \to f(x)(i \to \infty)(m)$, *in* R, *and*

(ii)′ *there exists a function* $K(x)$ *such that* $f_i(x) \geqq K(x)(m)$ *in* R, $(i=1, 2, \cdots)$, *and* $\int_R K(x)\,dm > -\infty$.

(c) *For the case* $\int_R f(x)\,dm = -\infty$,

(i) $f_i(x) \to f(x)(i \to \infty)(m)$, *in* R, *and*

(ii)″ *there exists a function* $K(x)$ *such that* $f_i(x) \leqq K(x)(m)$ *in* R, $(i=1, 2, \cdots)$, *and* $\int_R K(x)\,dm < \infty$.

The following theorem states the continuity of the information measure for the probability distributions in $V(E)$ when $m(E) < \infty$.

THEOREM 2.1 *Let* (R, S, m) *be a* σ*-finite measure space, and let the probability distributions* X, $\{X_i\}$, $(i=1, 2, \cdots)$ *be defined in* $V(E)$, *where* $0 < m(E) < \infty$. *Then the condition*

$$d(X, X_i) = \operatorname*{ess\cdot sup}_{x \in E} |\, p(x) - p_i(x) \,| \to 0 \,, \quad (i \to \infty) \tag{2.7}$$

implies that

$$H(X_i) \to H(X) \,, \quad (i \to \infty) \,. \tag{2.8}$$

PROOF. Let X_E be the uniform distribution on E with *pdf.* such as

$$p_E(x) = 1/m(E)(m), \text{ on } E \,,$$
$$= 0 \,, \text{ otherwise} \,.$$

Then it will easily be verified that $H(X) \geqq H(X_E)$. Hence the following two cases may be considered:

$$-\infty < H(X) < \infty \,, \tag{1°}$$

$$H(X) = \infty \,. \tag{2°}$$

First, the theorem will be proved in the case (1°). Put, for brevity, $f(x) = p(x) \log p(x)$ and $f_i(x) = p_i(x) \log p_i(x)$, $(i = 1, 2, \cdots)$. From (ii) of

Lemma 2.1 and the condition (2.7) it follows that, for any $\varepsilon>0$, there exists a positive integer N such that $i\geqq N$ implies

$$(2.9)\qquad |f(x)-f_i(x)|\leqq\varepsilon(p(x)+1+\varepsilon)(m),\ \text{on } E,$$

from which we obtain, for $i\geqq N$,

$$(2.10)\qquad |f_i(x)|\leqq|f(x)|+\varepsilon(p(x)+1+\varepsilon)(m),\ \text{on } E,$$

where the right-hand member of the above expression has the properties of $K(x)$ in Lemma 2.2 (a) by virtue of (1°) and the finiteness of $m(E)$. The inequality (2.9) above implies that

$$(2.11)\qquad f_i(x)\to f(x)(i\to\infty)(m),\ \text{in } E,$$

and the conditions (i) and (ii) of Lemma 2.2 (a) are satisfied. Since $H(X)=\int_E f(x)\,dm$ and $H(X_i)=\int_E f_i(x)\,dm$, we have (2.8), which completes the proof in case (1°).

Next, we consider the case (2°). Since (2.9) is also true in this case it holds, for $i\geqq N$, that

$$(2.12)\qquad f_i(x)\geqq f(x)-\varepsilon(p(x)+1+\varepsilon)(m),\ \text{on } E,$$

where the integral of the right-hand member over E is equal to ∞. The pointwise convergence (2.11) is valid too, therefore, our theorem follows from Lemma 2.2 (b) in case (2°). The proof of the theorem is then completed.

If we remove the condition of finiteness of $m(E)$, then an additional condition to (2.7) will be required, which may easily be illustrated by simple examples.

THEOREM 2.2 *Let the measure space (R, S, m) be σ-finite, X the probability distribution in $V(R)$ with $D(X)=E$, and $\{X_i\}$, $(i=1, 2, \cdots)$ the sequence in $V(E)$. If the conditions*

$$(2.13)\qquad d(X, X_i)=\operatorname*{ess\cdot sup}_{x\in E}|p(x)-p_i(x)|\to 0,\ (i\to\infty),$$

and, for $E_i=D(X_i)$ and $q_i(x)=p_i(x)/p(x)(m)$ on E_i, $(i=1, 2, \cdots)$,

$$(2.14)\qquad h_i(X, X_i)=\operatorname*{ess\cdot sup}_{x\in E_i}|1-q_i(x)|\to 0,\ (i\to\infty)$$

are satisfied, then

$$(2.15)\qquad H(X_i)\to H(X),\ (i\to\infty).$$

PROOF. Since $m(E)$ is not necessarily finite, the following four cases may be considered:

(1°) $$-\infty < H(X) < \infty ,$$

(2°) $$H(X) = \infty ,$$

(3°) $$H(X) = -\infty ,$$

(4°) $$H(X) = \infty - \infty .$$

As in the proof of the preceding theorem, $f(x)$ and $f_i(x)$ denote $p(x) \log p(x)$ and $p_i(x) \log p_i(x)$ respectively. In the first, we prove the theorem in the case (1°). From (i) of Lemma 2.1 and the condition (2.14) it follows that, for any $\varepsilon > 0$, there exists a positive integer N such that $i \geqq N$ implies

(2.16) $$|1 - q_i| < \varepsilon\ (m) , \text{ on } E_i ,$$

and

(2.17) $$|f(x) - f_i(x)| \leqq \varepsilon(|f(x)| + p(x) + p_i(x))(m), \text{ on } E_i$$

hence it holds that, for $i \geqq N$,

(2.18) $$|f(x) - f_i(x)| < \varepsilon(|f(x)| + (2+\varepsilon)p(x))\ (m) , \text{ on } E_i .$$

It follows from (2.18) that, for $i \geqq N$,

(2.19) $$|f_i(x)| \leqq (1+\varepsilon)\,|f(x)| + \varepsilon(2+\varepsilon)p(x)\ (m) , \text{ on } E .$$

Since (2.13) implies the pointwise convergence of $f_i(x)$ to $f(x)$ as $i \to \infty$, for almost all (m) x on E, and the right-hand member of (2.19) is integrable over E, we obtain (2.15) by virtue of Lemma 2.2 (a).

To prove the theorem in the remaining cases, we define, for E' and E'' given in (2.2),

(2.20) $$E'_i = E' \cap E_i , \quad E''_i \cap E_i , \quad (i=1, 2, \cdots) ,$$

and

(2.21) $$H'(X_i) = \int_{E'_i} f_i(x)\, dm , \quad H''(X_i) = \int_{E''_i} f_i(x)\, dm , \quad (i=1, 2, \cdots) ,$$

Then, since $E_i \subset E(m)$, $(i=1, 2, \cdots)$, it is clear that

(2.22) $$E_i = E'_i + E''_i(m) , \quad H(X_i) = H'(X_i) + H''(X_i) , \quad (i=1, 2, \cdots) .$$

Our definition of H' and $H''(X)$ in (2.3) implies that $H'(X) \geqq 0$ and $H''(X) \leqq 0$. In the present theorem, case (2°) occurs when $H'(X) = \infty$

and $H''(X)>-\infty$, case (3°) when $H'(X)<\infty$ and $H''(X)=-\infty$, and case (4°) when $H'(X)=\infty$ and $H''(X)=-\infty$. In each of these cases, in order to prove the theorem it will be sufficient to show that

$$H'(X_i)\to H'(X)\,,\quad H''(X_i)\to H''(X)\,,\quad (i\to\infty)\,. \tag{2.23}$$

Now, consider the case (2°). Put

$$\begin{aligned} &p'(x)=\begin{cases}p(x)\,, & (\text{on } E')\,,\\ 0\,, & (\text{on } R-E')\end{cases} \qquad p''(x)=\begin{cases}p(x)\,, & (\text{on } E'')\,,\\ 0\,, & (\text{on } R-E'')\,,\end{cases}\\ &p_i'(x)=\begin{cases}p_i(x)\,, & (\text{on } E_i')\,,\\ 0\,, & (\text{on } R-E_i')\,,\end{cases} \qquad p_i''(x)=\begin{cases}p_i(x)\,, & (\text{on } E_i'')\,,\\ 0\,, & (\text{on } R-E_i'')\,,\end{cases}\\ &\qquad\qquad\qquad\qquad\qquad (i=1,2,\cdots)\,. \end{aligned} \tag{2.24}$$

Since $m(E')\leqq 1$, the functions $p'(x)$ and $\{p_i'(x)\}$, $(i=1,2,\cdots)$ satisfy the conditions of Theorem 2.1 except the fact that they are not necessarily the probability density functions. It will be noticed that the proof of Theorem 2.1 in case (2°) does not require the property that the integrals of *pdf.*'s of X and $\{X_i\}$ are all equal to unity. Hence the proof similar to that in case (2°) of the preceding theorem will assure of the validity of the first half of (2.23).

The truncation (2.24) will reserve all the properties of $p(x)$ and $\{p_i(x)\}$, $(i=1,2,\cdots)$, indispensable in the proof of the present theorem for case (1°), therefore, we shall be able to confirm the validity of the latter half of (2.23) by the proof parallel to that in case (1°) of the present theorem.

In the case (3°), the first half of (2.23) will be proved in the similar way as the case (1°) of Theorem 2.1. In the following, we shall prove the latter half of (2.23).

As in the proof in case (1°) of the present theorem, it is easy to see that, for any $\varepsilon>0$, there exists a positive integer N such that $i\geqq N$ implies

$$|f(x)-f_i(x)|\leqq\varepsilon(|f(x)|+(2+\varepsilon)p(x))(m),\ \text{on } E_i''\,, \tag{2.25}$$

which leads to

$$f_i(x)\leqq(1-\varepsilon)f(x)+\varepsilon(2+\varepsilon)p(x)(m),\ \text{on } E_i''\,, \tag{2.26}$$

because $f(x)\leqq 0(m)$, on E_i''. Since $f_i(x)=0\,(m)$, on the set $E''-E_i''$, it follows from (2.26) that, for any $\varepsilon>0$, there exists a positive integer N such that $i\geqq N$ implies

(2.27) $f_i(x) \leqq \max(0, (1-\varepsilon)f(x)+\varepsilon(2+\varepsilon)p(x))(m)$, on E''.

If $\varepsilon<1$, then the right-hand member of (2.27) is integrable over E''. Clearly $f_i(x)$ converges to $f(x)$ for almost all (m) x in E'' as $i \to \infty$. Hence we obtain the latter half of (2.23) by virtue of Lemma 2.2 (c).

For the case (4°), the proofs of the first and second half of (2.23) will be essentially included in those of the present theorem in the cases (2°) and (3°) above, respectively.

Thus the proof of our theorem has been completed.

As an application of Theorem 2.2 above, we prove the following.

COROLLARY 2.1. *Let (R, S, m) be a σ-finite measure space, and let X be in $V(R)$. Then there exists a sequence $\{X_i\}$, $(i=1, 2, \cdots)$ in $\overline{V}(R)$ such that $H(X_i) \to H(X)$ as $i \to \infty$.*

PROOF. Let $E=D(X)$, and $E'+E''=E$ as in (2.2). By constructing the product partition we can get a sequence $\{Z_i\}$, $(i=1, 2, \cdots)$ of m-partitions of E, which satisfies the following three conditions:

(i) $Z_i=Z_i'+Z_i''$, where Z_i' and Z_i''

are the m-partitions of E' and E'', respectively, that is, if $Z_i=\{A_{ij}\}$, $Z_i'=\{A_{ij}'\}$ and $Z_{ij}''=\{A_i''j\}$, $(j=1, 2, \cdots)$, then $\{A_{ij}\}=\{A_{ij}'\} \cup \{A_{ij}''\}$, $(i=1, 2, \cdots)$, such that

(ii) $$\operatorname*{ess\cdot var}_{x\in A_{ij}} p(x) \leqq 1/2^i, \qquad (i, j=1, 2, \cdots),$$

where $\operatorname*{ess\cdot var}_{x\in A} p(x) = \operatorname*{ess\cdot sup}_{x\in A} p(x) - \operatorname*{ess\cdot inf}_{x\in A} p(x)$, and simultaneously

(iii) $$\operatorname*{ess\cdot var}_{x\in A_{ij}} \log p(x) \leqq 1/2^i, \qquad (i, j=1, 2, \cdots).$$

For each of the m-partitions $\{Z_i\}$, $(i=1, 2, \cdots)$, we define the following simple *pdf*.

(2.28) $$p_i(x)=\begin{cases} P(A_{ij})/m(A_{ij}), \text{ on } A_{ij}, & (j=1, 2, \cdots), \\ 0, \text{ otherwise}, \end{cases}$$

where $P(A)=\int_A P(x)\,dm$, and let X_i be the probability distribution with $p_i(x)$ as its *pdf*. Then X_i is in $\overline{V}(E)$, or more precisely, $D(X_i)=E(m)$.

Now, it will be shown that the sequence $\{X_i\}$ $(i=1, 2, \cdots)$ satisfies the conditions (2.13) and (2.14) of the preceding theorem. Put

(2.29) $$u_{ij}=\operatorname*{ess\cdot inf}_{x\in A_{ij}} \log p(x), \quad U_{ij}=\operatorname*{ess\cdot sup}_{x\in A_{ij}} \log p(x), \qquad (i, j=1, 2, \cdots).$$

Then it becomes

(2.30) $$\operatorname*{ess\cdot inf}_{x\in A_{ij}} p(x)=e^{u_{ij}},\quad \operatorname*{ess\cdot sup}_{x\in A_{ij}} p(x)=e^{U_{ij}},\qquad (i,j=1,2,\cdots)$$

or

(2.31) $$e^{u_{ij}}\leqq p(x)\leqq e^{U_{ij}}\,(m)\,,\quad \text{on } A_{ij}\,,\qquad (i,j=1,2,\cdots)\,.$$

Integrating (2.31), over A_{ij}, dividing by $m(A_{ij})$, we obtain the following inequality

(2.32) $$e^{u_{ij}}\leqq P(A_{ij})/m(A_{ij})\leqq e^{U_{ij}}\,,\qquad (i,j=1,2,\cdots)\,.$$

From the condition (ii) above, (2.28), (2.31) and (2.32), it follows that

(2.33) $$|\,p(x)-p_i(x)\,|\leqq 1/2^i\,(m),\quad \text{on } E\,,\qquad (i=1,2,\cdots)\,,$$

from which the condition (2.13) follows. On the other hand, by the condition (iii), (2.31) and (2.32), we obtain

$$|\,1-q_i(x)\,|=|\,(p(x)-p_i(x))/p(x)\,|\leqq(e^{U_{ij}}-e^{u_{ij}})/e^{u_{ij}}\,(m)\,,$$

and

$$(e^{U_{ij}}-e^{u_{ij}})/e^{u_{ij}}=e^{U_{ij}-u_{ij}}-1\leqq e^{1/2^i}-1\leqq 1/2^{i-1}\,,\quad \text{on } A_{ij}\,,$$
$$(i,j=1,2,\cdots)\,,$$

hence it holds that

(2.34) $$|\,1-q_i(x)\,|\leqq 1/2^{i-1}\,(m)\,,\quad \text{on } E\,,\qquad (i=1,2,\cdots)\,,$$

which implies (2.14).

Then, our result follows from Theorem 2.2

3. Characterization

For a σ-finite measure space (R, S, m), let $M(S)$ denote the range of m-measure, i. e., $M(S)=\{m(E): E\in S\}$. Let W be the real line, and W^+ the non-negative half of W. Throughout the present section we assume that the space (R, S, m) satisfies the condition

(3.1) $$M(S)=W^+\,.$$

The product measure space of two σ-finite measure spaces (R, S, m) and (R', S', m') will be denoted by $(R\times R', S\times S', m\times m')$, where $R\times R'=\{(x, x'): x\in R, x'\in R'\}$, $S\times S'=B(\{A\times A': A\in S, A'\in S'\})$, i.e. $S\times S'$ is the smallest σ-field containing all the product sets, and $m\times m'$ is the product measure on $(R\times R', S\times S')$, for which $m\times m'(A\times A')=m(A)m'(A')$.

The product measure space is σ-finite if the component spaces are all σ-finite, and it will be noticed that, if all the component spaces satisfy the condition (3.1), so does also the product space.

Let $W^* = W \cup \{\infty, -\infty, \infty-\infty\}$. Addition and substitution between the elements of W and the additional elements $\infty, -\infty, \infty-\infty$, will be defined as usual, for example, $w \pm \infty = \pm\infty$, $w \pm(-\infty) = \mp\infty$, $w \pm (\infty-\infty) = \infty-\infty$, and commutative law will be permitted. Now, for each of the spaces satisfying (3.1), we consider a function $H(X)$ defined in $V(R)$ and ranges in W^*. Of course, it depends upon the space (R, S, m) for which it is considered, but we do not indicate it explicitly.

We postulated some assumptions on the properties of H.

ASSUMPTIONS

(I) For the uniform distribution

(i) $-\infty < H(X_E) < \infty$, for any $E(0 < m(E) < \infty)$, and any (R, S, m).

(ii) For any X_E in (R, S, m) and $X_{E'}$ in (R', S', m'), $m(E) = m'(E')$ implies $H(X_E) = H(X_{E'})$.

(iii) In (ii) above, $m(E) > m'(E')$ implies $H(X_E) < H(X_{E'})$.

(II) Let X be defined in $\bar{V}(R)$ and X' in $\bar{V}(R')$. If at least one of the three members $H(X, X')$, $H(X)$, and $H_X(X')$ is finite, then it holds that

$$H(X, X') = H(X) + H_X(X') , \tag{3.2}$$

where $H(X, X)$ denotes the value of H for the joint probability distribution (X, X') in $\bar{V}(R \times R')$, and $H_X(X') = \int_R H(X' \mid X)\, p(x)\, dm$, for the value $H(X' \mid x)$ corresponding to the conditional probability distribution of X' given $X = x$.

(III) Let X be in $V(R)$ with $D(X) = E$. For a sequence X_i, $(i = 1, 2, \cdots)$ with $D(X_i) = E\,(m)$ for all i, define $q_i(x) = p_i(x)/p(x)$, $(i = 1, 2, \cdots)$, on E. Then, if the conditions

$$d(X, X_i) = \operatorname*{ess\cdot sup}_{x \in E} | p(x) - p_i(x) | \to 0 , \quad (i \to \infty) , \tag{3.3}$$

$$h(X, X_i) = \operatorname*{ess\cdot sup}_{x \in E} | 1 - q_i(x) | \to 0 , \quad (i \to \infty) , \tag{3.4}$$

are satisfied simultaneously, it holds that

$$H(X_i) \to H(X) , \quad (i \to \infty) . \tag{3.5}$$

When $m(E) < \infty$, the condition (3.4) is not required.

The definition of the uniform distribution will be found in the beginning of the proof of Theorem 2.1 in the preceding section. Assumption (I) states that for the uniform distributions, the function H does not depend on the distribution range itself, but does only on the m-measure of it, hence we can write $H(v)$ instead of $H(X_E)$ if $m(E)=v$. $H(v)$ becomes a monotone decreasing function in v. Consider two mutually independent uniform distributions, X_E and $X_{E'}$, in (R, S, m) and (R', S', m') respectively, with $m(E)=m'(E')=1$. Then the joint distribution $(X_E, X_{E'})$ becomes a uniform distribution on $E\times E'$ in the product space $(R\times R', S\times S', m\times m')$, with $m\times m'(E\times E')=1$. Hereby it is easy to obtain

$$H(v)=0\ , \quad \text{if } v=1\ , \tag{3.6}$$

by virtue of (I) and (II).

Now, we shall prove the uniqueness of the expression (2.1) under the assumptions (I) to (III). First, we state the following.

LEMMA 3.1 *For the uniform distributions, it holds, under the assumptions* (I) *and* (II), *that*

$$H(v)=c\log\frac{1}{v}\ , \tag{3.7}$$

where c is a positive constant.

PROOF. Considering k mutually independent uniform distributions on E, we obtain from (II)

$$H(v^k)=H(v)\ , \tag{3.8}$$

where $v=m(E)$. Solving this functional equation, we have, for $v>1$,

$$H(v)=c'\log\frac{1}{v}\ , \tag{3.9}$$

and for $v<1$,

$$H(v)=c''\log\frac{1}{v}\ , \tag{3.10}$$

where c' and c'' are the positive constants.

Let $v'>1, v''<1$ and $v'v''>1$. Then, considering two mutually independent uniform distributions, we get by assumption (II)

$$H(v'v'')=H(v')+H(v'')\ , \tag{3.11}$$

hence, from (3.9), (3.10) and (3.11) it follows that $c'=c''$. From this and (3.6) we obtain (3.7), which completes the proof of our lemma.

LEMMA 3.2 *Let $Z=\{A_i\}$, $(i=1, 2, \cdots)$ be an m-partition of E in (R, S, m), and X_Z a probability distribution with pdf. such as*

$$p_Z(x)=\begin{cases} p_i/v_i\ , & \text{on } A_i, \qquad (i=1, 2, \cdots)\ , \\ 0\ , & \textit{otherwise}\ , \end{cases} \tag{3.12}$$

where $p_i \geqq 0$, $v_i=m(A_i)$, $(i=1, 2, \cdots)$, and $\sum_i p_i=1$. Then, under the assumptions (I) *and* (II), *it holds that*

$$H(X_Z)=c\sum_i p_i \log \frac{p_i}{v_i}\ , \tag{3.13}$$

where c is a positive constant.

PROOF. We consider a sequence $Z^*=\{A_j^*\}$, $(j=1, 2, \cdots)$, of disjoint subsets of R, with

$$m(A_j^*)=p_j/v_j\ , \qquad (j=1, 2, \cdots)\ . \tag{3.14}$$

Denote by X_{Z^*} the probability distribution such that, if X_Z falls in A_i, then X_{Z^*} is uniformly distributed on A_j^* for all $i=1, 2, \cdots$. The joint distribution of X_Z and X_{Z^*} will become

$$p(x, x^*)=\begin{cases} 1\ , & \text{on } A_i \times A_i^*\ , \qquad (i=1, 2, \cdots)\ , \\ 0\ , & \text{otherwise}\ , \end{cases} \tag{3.15}$$

hence, it is uniformly distributed on the set $F=\sum A_i \times A_i^*$ in the product measure space $(R\times R, S\times S, m\times m)$, with $m\times m(F)=1$. Therefore, it follows from (3.6) that

$$H(X_Z, X_{Z^*})=0\ . \tag{3.16}$$

On the other hand, from (3.14) and Lemma 3.1 we obtain

$$H(X_{Z^*}\,|\,x) = -c \log \frac{p_i}{v_i}\ , \qquad \text{if } x \in A_i\ , \qquad (i=1, 2, \cdots)\ ,$$

from which it follows that

$$H_{X_Z}(X_{Z^*})=-c\sum_i p_i \log \frac{p_i}{v_i}\ . \tag{3.17}$$

By assumption (II) we have

$$H(X_Z, X_{Z^*})=H(X_Z)+H_{X_Z}(X_{Z^*})\ , \tag{3.18}$$

which implies (3.13), by (3.16) and (3.17).

THEOREM 3.1 *Let X be any probability distribution in $V(R)$. Then, under the assumptions* (I), (II) *and* (III) *it holds that*

$$H(X)=c\int_R p(x)\log p(x)\,dm\ , \tag{3.19}$$

where c is a positive constant.

PROOF. As in the proof of Corollary 2.1, we define a sequence of simple probability distributions $\{X_i\}$, $(i=1, 2, \cdots)$, with *pdf.*'s such as, for the m-partitions $Z_i=\{A_{ij}\}$, $(i, j=1, 2, \cdots)$, in that proof,

$$p_i(x)=\begin{cases} p_{ij}/v_{ij}, & \text{on } A_{ij}, \quad (j=1, 2, \cdots)\ , \\ 0\ , & \text{otherwise}\ , \end{cases} \tag{3.20}$$

where $p_{ij}=P(A_{ij})$ and $v_{ij}=m(A_{ij})$. Then, we know that

$$d(X, X_i)\to 0\ ,\quad h(X, X_i)\to 0\ ,\qquad (i\to\infty)\ , \tag{3.21}$$

and

$$\sum_j p_{ij}\log\frac{p_{ij}}{v_{ij}}\to\int_R p(x)\log p(x)\,dm\ ,\qquad (i\to\infty)\ . \tag{3.22}$$

But, by Lemma 3.2 it holds that

$$H(X_i)=c\sum_j p_{ij}\log\frac{p_{ij}}{v_{ij}}\ ,\qquad (i=1, 2, \cdots)\ , \tag{3.23}$$

and, by the assumption (III), the convergences in (3.21) imply that

$$H(X_i)\to H(X)\ ,\qquad (i\to\infty)\ . \tag{3.24}$$

From (3.22), (3.23) and (3.24), it follows that

$$H(X)=c\int_R p(x)\log p(x)\,dm\ , \tag{3.25}$$

where c is a positive constant, which proves the theorem.

In characterizing the information measure for finite discrete case, A. I. Khinchin assumed the usual continuity property, which played an important role in his characterization procedure. The finite discrete case will correspond to the case of simple probability distributions, associated with finite m-partitions of equal weights, i.e., the partitions $Z=\{A_i\}$, $(i=1, 2, \cdots, n)$, such that $m(A_1)=m(A_2)=\cdots=(A_n)$. Then, the assumption (III) in the case where $m(E)<\infty$, will be equivalent to the continuity assumption by A. I. Khinchin.

In continuous case, our present assumptions differ from those by H. Hatori, in the class to which the addition property (3.2) is assumed, and in the presence of continuity assumption. That is, we assumed (3.2) only for the class of all simple distributions, while he required the validity of (3.2) on the class of all absolutely continuous distributions. Essentially, under the assumption (I), the postulate III by H. Hatori [5] will be equivalent to our assumptions (II) and (III).

The author expresses his thanks to Mr. Masashi Okamoto for his valuable advices.

DEPT. OF MATM., SCHOOL OF SCI. AND ENG.
NIHON UNIVERSITY

REFERENCES

[1] Shannon, C. E., "A mathematical theory of communication," *Bell System Tech. Jour.*, Vol. 27 (1948).
[2] Wiener, N., *Cybernetics*, John Wiley & Sons, New York, 1948.
[3] Shannon, C. E., and Weaver, W., *The Mathematical Theory of Communication*, Univ. of Illinois Press, Urbana, 1949.
[4] Khinchin, A. I., "The entropy concept in probability theory," *Uspekhi Matematicheskikh Nauk*, Vol. 8 (1955), (in Russian).
[5] Hatori, H., "A note on the entropy of a continuous distribution," *Kodai Math. Sem. Rep.* Vol. 10 (1958).