Bull. Korean Math. Soc.
Vol. 9 (1972), pp. 9—16

An extension of the properties on conditional information and entropy in probability spaces

Chun Ho Choi

1. Introduction.

Information theory is founded on mathematical statistics and probability theory. We can find this idea through the book [12] of N. Wiener and the paper [10] of C. E. Shannon. Furthermore, after their basic theorems were found by McMillan [8], information theory has been treated in the pure mathematical way. The aim of this paper is to give properties of conditional information and entropy which is an extension of Sinai's theorem.

2. Conditional information and entropy

Let ξ be a countable $\mathscr{B}$-measurable partition of X and $\mathscr{C}$ be a sub-σ-algebra, $(X, \mathscr{B}, P)$ is probability measure space. [6]

DEFINITION 1. The conditional information of ξ given $\mathscr{C}$, written $I(\xi/\mathscr{C})$ is defined by the formula

2—1 $$I(\xi/\mathscr{C}) = -\sum_{A\in\xi}\chi_A \log P_{\mathscr{C}}(A),$$

where χ_A is the characteristic function and $P_{\mathscr{C}}(A)$ is conditional probability.

DEFINITION 2. The conditional entropy of given $\mathscr{C}$, written $H(\xi/\mathscr{C})$ is defined by the formula.

2—2 $$\mathrm{H}(\xi/\mathscr{C}) = -\sum_{A\in\xi}\chi_A P_{\mathscr{C}}(A)\log P_{\mathscr{C}}(A),$$

We define the information of ξ, by $I(\xi)$ and the entropy of written $H(\xi)$ by the formula

2—3 $$I(\xi) = -\sum_{A\in\xi}\chi_A \log P(A),$$

2—4 $$H(\xi) = -\sum_{A\in\xi}\chi_A P(A)\log P(A).$$

NOTE: If ξ is finite, then $H(\xi)$ will be also finite, in fact $\infty > H(\xi) \geqq 0$, thus if we denote by Z the class of such ξ, then the finite partitions belong to Z. For two partitions ξ and η of X, we define their refinement, written $\xi\vee\eta$ to be the set of subset of X

$$\{A\cap B : A\in\xi,\ B\in\eta\}.$$

It is easily checked that if ξ and η are countable measurable partitions of X, then so is $\xi\vee\eta$.

Now, we introduce an order relation on a class of countable measurable partitions of X. We say that $\xi\leqq\eta$ if each element of ξ is a union of elements from η,

2—5 $\quad \xi\leqq\eta \quad$ is equivalent to $\quad \xi\vee\eta=\eta.$

THEOREM 1. *If ξ and η are countable and measurable partitions of X, then the following identity is valid.*

2—6 $\quad I(\xi\vee\eta/\zeta)=I(\xi/\zeta)+I(\eta/\xi\vee\zeta).$

Proof: First we compute the right hand side (RHS) :

$$\text{RHS}=-\sum_{A\in\xi}\chi_A\log P_\zeta(A)-\sum_{B\in\eta}\chi_B\log P_{\xi\vee\zeta}(B)$$

$$=-\sum_{B\in\eta}\chi_{A\wedge B}\log P_\zeta(A)-\sum_{\substack{A\in\xi\\ B\in\zeta}}\chi_{A\wedge B}\log P_{\xi\vee\zeta}(B)$$

$$=-\log P_\zeta(A)P_{\xi\vee\zeta}(B) \qquad \text{on } A\cap B.$$

For, $\quad C\in\zeta,\quad P_\zeta(A)=\dfrac{P(A\cap C)}{P(C)} \qquad$ on $A\cap C$,

$$P_{\xi\vee\zeta}(B)=\frac{P(A\cap B\cap C)}{P(A\cap C)} \qquad \text{on } A\cap B\cap C,$$

thus, $\text{RHS}=-\log\dfrac{P(A\cap B\cap C)}{P(C)} \qquad$ on $A\cap B\cap C$,

therefore

$$\text{RHS}=-\sum_{\substack{A\in\xi,B\in\eta\\ C\in\zeta}}\chi_{A\wedge B\wedge C}\log\frac{P(A\cap B\cap C)}{P(C)}=-\sum_{\substack{A\in\xi\\ B\in\eta}}\chi_{A\wedge B}(\log\sum_{C\in\zeta}\chi_C\frac{P(A\cap B\cap C)}{P(C)}).$$

COROLLARY 1. *Let ξ and η be as in the theorem* (1),

2—7 $\quad H(\xi\vee\eta/\zeta)=H(\xi/\zeta)+H(\eta/\xi\vee\zeta).$

COROLLARY 2. *Given ξ and η as in theorem* (1),

2—8 $\quad I(\xi\vee\eta)=I(\xi)+I(\eta/\xi),$

2—9 $\quad H(\xi\vee\eta)=H(\xi)+H(\eta/\xi).$

COROLLARY 3. *If ξ, η and ζ are in the theorem* (1) *and $\xi\geqq\eta$,* then

2—10 $\quad I(\xi/\zeta)\geqq I(\eta/\zeta),\qquad I(\xi)\geqq I(\eta),$

2—11 $H(\xi/\zeta) \geqq H(\eta/\zeta), \quad H(\xi) \geqq H(\eta).$

Proof: $I(\xi \vee \eta/\zeta) = I(\xi/\eta \vee \zeta) + I(\eta/\zeta)$

since $\xi \vee \eta = \xi, \quad I(\xi/\eta \vee \zeta) \geqq 0,$

so we have

$$I(\xi/\zeta) \geqq I(\eta/\zeta).$$

We write $\bar{\xi}$ for the σ-algebra generated by ξ.

PROPOSITION (1). $\bar{\xi} \subset \mathscr{C}$ if and only if

2—12 $I(\bar{\xi}/\mathscr{C}) = 0, \qquad H(\bar{\xi}/\mathscr{C}) = 0.$

Proof: $I(\bar{\xi}/\mathscr{C}) = -\log P_{\mathscr{C}}(\bar{\xi}) = -\log \dfrac{P(\mathscr{C} \cap \bar{\xi})}{P(\mathscr{C})} = -\log \dfrac{P(\mathscr{C})}{P(\mathscr{C})} = 0.$

Suppose, two sub-σ-algebra $\mathscr{C}_1$ and $\mathscr{C}_2$ of $\mathscr{B}$ such that $\mathscr{C}_1 \subset \mathscr{C}_2$ for each $c_1 \in \mathscr{C}_1$, $c_2 \in \mathscr{C}_2$, we assume that exists $c_2 \in \mathscr{C}_2$ such that $c_1 \subset c_2$ and $c_2 \notin c_1$. In this situation,

since $P_{C_1}(y_i) \leqq P_{C_2}(y_i)$ implies $I(y_i/c_1) \geqq I(y_i/c_2)$,

we can see immediately that.

2—13 $H(y_i/c_1) \geqq H(y_i/c_2),$

where ξ is a countable $\mathscr{B}$-measurable partition of X and $y_i \in X_\xi$.

Consider a set $B \in \mathscr{B}$ such that $B = B_1 \cup B_2$, then we have

$$P_B(y_i) = P_{B_1}(y_i) + P_{B_2}(y_i) \geqq P_{B_1}(y_i) P_{B_2}(y_i),$$

therefore

2—14 $I(y_i/B) \leqq I(y_i/B_1) + I(y_i/B_2),$

$$H(y_i/B) \leqq H(y_i/B_1) + H(y_i/B_2).$$

THEOREM 2. (Entropy theorem)

Let ξ be a countable $\mathscr{B}$-measurable partition of X and $H(\xi) < \infty$. If a sequence $\mathscr{C}_1$, $\mathscr{C}_2$ of sub-σ-algebra of $\mathscr{B}$ is $\mathscr{C}_n \uparrow \mathscr{C}$,

then

$$I(\xi/\mathscr{C}_n) \uparrow I(\xi/\mathscr{C}),$$

2—15 $H(\xi/\mathscr{C}_n) \uparrow H(\xi/\mathscr{C}).$

3. McMillan's theorem

In order to prove the McMillan's theorem we need notations. let T be measurable, measure preserving transformation in the preceeding section and let ξ be a countable $\mathfrak{B}$-measurable partition of X, and let $\mathfrak{S}$ be a sub-σ-algebra of $\mathfrak{B}$. We introduce the following notations:

$$\bigvee_{i=1}^{n} T^{-i}(\xi)=T^{-1}(\xi)\vee T^{-2}(\xi)\vee\cdots\cdots\cdots\cdots\vee T^{-n}(\xi).$$

McMillan's theorem was proved by him in [8] for the case of a finite partition and L_1-convergence.

The theorem was extended to the case of almost everywhere convergence by Breiman in [1] and by Chung in his paper [3] proved the theorem for the case of a countable partition ξ with $H(\xi)<\infty$.

In this section, we prove this theorem by using Ergodic theorem.

THEOREM 3. (Ergodic theorem)

Let $f(x)\in L_1(x)$ and T be as above, then there exists $f^*(x)\in L_1(x)$ suuch that

(i) $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x))\to f^*(x)$

(ii) $f^*(T(x))=f^*(x)$

(iii) $\int_X f^*(x)dm(x)=\int_X f(x)dm(x)$

THEOREM 4. Suppose ξ is a countable $\mathfrak{B}$-measurable partition of X such that

$$H(\xi)<\infty, \text{ if } h(\xi, T)=\lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n-1}T^{-i}(\xi)),$$

then

$$h(\xi, T)=\lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n-1}T^{-i}(\xi))=\lim_{n\to\infty}H(\xi/\bigvee_{i=1}^{n}T^{-i}(\xi)).$$

Proof:

$$I(\bigvee_{i=0}^{n-1}T^{-i}(\xi))=I(\xi\bigvee_{i=1}^{n-1}T^{-i}(\xi))=I(\bigvee_{i=1}^{n-1}T^{-i}(\xi))+I(\xi/\bigvee_{i=1}^{n-1}T^{-i}(\xi)),$$

therefore

$$I(\xi/\bigvee_{i=1}^{n-1}T^{-i}(\xi))=I(\bigvee_{i=0}^{n-1}T^{-i}(\xi))-I(\bigvee_{i=1}^{n-1}T^{-i}(\xi)),$$

$$I(\xi/\bigvee_{i=2}^{n-1}T^{-i}(\xi)=I(\bigvee_{i=1}^{n-1}T^{-i}(\xi))-I(\bigvee_{i=2}^{n-1}T^{-i}(\xi)),$$

$$I(\xi/T^{-(n-1)}(\xi))=I(\bigvee_{i=n-2}^{n-1}T^{-i}(\xi))-I(T^{-(n-1)}(\xi))=I(\bigvee_{i=n-2}^{n-1}T^{-i}(\xi))-I(\xi),$$

$$\sum_{k=1}^{n-1}I(\xi/\bigvee_{i=k}^{n-1}T^{-i}(\xi))=I(\bigvee_{i=0}^{n-1}T^{-i}(\xi))-I(\xi),$$

thus

$I(\xi/\bigvee_{i=1}^{k}T^{-i}(\xi))$ is decreasing by k, so it will have a limiting value.

Therefore

$$I(\xi,\ T)=\lim_{n\to\infty}\frac{1}{n}I(\bigvee_{i=0}^{n-1}T^{-i}(\xi))-\lim_{n\to\infty}\frac{1}{n}I(\xi)+\lim_{n\to\infty}I(\xi/\bigvee_{i=k}^{n-1}T^{-i}(\xi))$$
$$=\lim_{n\to\infty}I(\xi/\bigvee_{i=k}^{n-1}T^{-i}(\xi)),$$

thus

$$h(\xi,\mathrm{T})=\lim_{n\to\infty}\frac{1}{n}H(\bigvee_{i=0}^{n-1}T^{-i}(\xi))=\lim_{n\to\infty}H(\xi/\bigvee_{i=1}^{n}T^{-i}(\xi)).$$

THEOREM 5. If ξ is a finite measurable partition generated by ξ, then there is a finite sub-partition η of ξ_0 such that

$$H(\xi/\eta)<\varepsilon.$$

Proof: Let A_1, A_2, A_3, …… be fibers of ξ and each of them has positive measure and

$$\phi(t)=-t\log t \qquad (0\leqq t\leqq 1)$$

is continuous function, where $\phi(0)=\phi(1)=0$, then for δ_0 $(0<\delta_0<1)$, we have

$$\phi\ (t)<\frac{\varepsilon}{\gamma.}$$

If B_1, B_2, B_3, …… be fibers of an another partition η such that $P(A_i/B_j)<\delta_0$, then

$$H(\xi/\eta)=-\sum P(B_j)P(A_i/B_j)\log P(A_i/B_j)$$
$$=\sum P(B_j)\phi(P(A_i/B_j))\leqq P(B_j)\cdot\frac{\varepsilon}{\gamma}<\frac{\varepsilon}{\gamma}<\varepsilon,$$

therefore $H(\xi/\eta)<\varepsilon$, *i. e.*, $H(\xi/\eta)\to 0$,

THEOREM 6. Let ξ and η are two countable $\mathfrak{B}$-measurable partitions of X, then we have $h(\xi,\ T)\leqq h\ (\eta,\ T)+\mathrm{h}(\xi/\eta)$.

Proof:

$$H(\bigvee_{i=0}^{n-1}T^{-i}(\xi))\leqq H(\bigvee_{i=0}^{n-1}T^{-i}(\xi)\vee\bigvee_{j=0}^{n-1}T^{-j}(\eta))$$
$$=H(\bigvee_{j=0}^{n-1}T^{-j}(\eta))+H(\bigvee_{i=0}^{n-1}T^{-i}(\xi)/\bigvee_{j=0}^{n-1}T^{-j}(\eta)),$$

$$H(\bigvee_{i=0}^{n-1}T^{-i}(\xi)/\bigvee_{j=0}^{n-1}T^{-j}(\eta))\leqq H(\xi/\bigvee_{j=0}^{n-1}T^{-j}(\eta))+H(T^{-1}(\xi)/\bigvee_{j=0}^{n-1}T^{-j}(\xi))$$
$$\leqq\sum_{i=0}^{n-1}H(T^{-i}(\xi)/T^{-j}(\eta)),=n\,H(\xi/\eta),$$

hence,

$$H(\bigvee_{i=0}^{n-1} T^{-i}(\xi)) \leqq H(\bigvee_{j=0}^{n-1} T^{-j}(\eta)) + nH(\xi/\eta),$$

therefore,

$$\frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}(\xi)) \leqq \frac{1}{n} H(\bigvee_{j=0}^{n-1} T^{-j}(\eta)) + H(\xi/\eta),$$

thus,

$$h(\xi, T) \leqq h(\eta, T) + H(\xi/\eta).$$

COROLLARY 4.

$$h(\xi, T) = h(\eta, T).$$

Proof: By the theorem(8) and (7),

$$h(\xi, T) \leqq h(\eta, T) + H(\xi, T).$$

We can make η such that

$$H(\xi/\eta) < \varepsilon,$$

therefore

$$h(\xi, T) \leqq h(\eta, T) + \varepsilon, \text{ i.e., } h(\xi, \mathrm{T}) = h(\eta, \mathrm{T}).$$

THEOREM 7. (Sinai's theorem)

Let T has the inverse and $\bigvee_{n=-\infty}^{\infty} T^n(\xi) = \mathfrak{B}$, then

$$h(T) = h(\xi, T),$$

where

$$h(T) = \sup h(\xi, T).$$

Proof: If η is any finite subfield of $\mathfrak{B}$ then

$$h(\eta, T) \leqq h(\xi, T).$$

Let $\xi_n = \bigvee_{k=-n}^{n} T^k(\xi)$. By the theorem (4), $h(\xi_n, T) = h(\xi, T)$ and the theorem (6), we have

$$h(\eta, T) \leqq h(\xi_n, T) + H(\eta/\xi_n)$$
$$= h(\xi, T) + H(\eta/\xi_n).$$

Using theorem (5), we can prove that

$$\lim_{n\to\infty} H(\eta/\xi_n) = 0,$$

$$h(T) = \sup h(\xi_n, T) = h(\xi, \mathrm{T}).$$

Let us assume that G_1 and G_2 are σ-subfields of $\mathfrak{B}$ and write $G_1 \cong G_2$ to indicate that every set in G_1 differs by a set of measure 0 from some set in G_2.

MAIN THEOREM.

Let $\{g_n\}$ be a nondecreasing sequence of fields.

If $\bigvee_{n=1}^{\infty} \bigvee_{i=0}^{\infty} T^{-i} g_n \cong \mathfrak{B}$,

then $h(T)=\lim_{n\to\infty} \sup_{\xi\subset g_n} h(\xi, T)$.

Proof: If G_n is the field generated by $\bigcup_{i=0}^{n} T^{-i} g_n$ and $\mathfrak{B}_o=\bigcup_{n=1}^{\infty} G_n$,

then every set in $\mathfrak{B}$ differs by a set measure 0 from some set in the σ-field generated by $\mathfrak{B}_0$. Furthermore it follows by theorem (5) and (7)

$$h(T)=\sup_{\xi\subset\mathfrak{B}_0} h(\xi, T)$$

If $\eta\subset\mathfrak{B}_0$, then η is contained in G_n for some n and hence has atoms $B_1, B_2, \cdots, B_k$ of the form

$$B_u=\bigcup_{v=1}^{l} \bigcap_{i=0}^{n} T^{-i} G_{iuv}, \qquad u=1, 2, \cdots\cdots, k$$

with $G_{iuv}\in g_n$. If η is the field generated by G_{iuv} then $\eta\equiv G_{iuv}$, $\eta\subset g_n$ and

$$\eta\subset \bigvee_{i=0}^{\infty} T^{-i}\xi,$$

therefore

$$h(\eta, T)\leqq h(\bigvee_{i=0}^{n} T^{-i}(\xi), T)=h(\xi, T)\leqq \sup_{\xi\subset g_n} h(\xi, T),$$

thus,

$$h(T)=\sup_{\xi\subset\mathfrak{B}_0} h(\xi, T).$$

References

[1] Breiman, L., *The individual ergodic theorem of information theory*, Ann. Math. Stat., 28(1957), 809—811.

[2] Billingsley, P., *Ergodic theory and information*, (1965) Wiley, New York.

[3] Chung, K. L., *A note on the ergodic theorem of information theory*, Ann. Math. Stat., 32(1961), 612—614.

[4] Fano, R. M., *Transmission of information*, (1961) Massachusetts Institute of Technolagy.

[5] Feisein, A., *Foundation of information theory*, (1958) McGraw-Hill.

[6] Halmos, P. R., *Measure theory*, (1960) New York, D. Van Nostrand.

[7] Loeve, M., *Probability theory*, (1955) New York, D. Van Nostrand.

[8] McMillan, B., *The basic theorem of information theory*, Ann. Math. Stat., 24(1953), 196—219.

[9] Parry, W., *Generators and strong generators in ergodic theory*, Bull. Amer. Math. Soc., 72(1966), 294—296.

[10] Shannon, C., *A mathematical theory of communication.*, B. S. T. J. 27 (379—423) (1948).

[11] Sinai, Ja. G., *On the concept of entropy for dynamical systems*, Sov. Math., 124(1959), 768—771.

[12] Wiener, N., *The extrapolation interpolation and smoothing of stationary time series*, (1949) John Wiley, New York.

Hanyang University