On Triple Mutual Information

Toru Tsujishita

Series ♯255. August 1994

HOKKAIDO UNIVERSITY
PREPRINT SERIES IN MATHEMATICS

♯ 231: A. Hoshiga, The asymptotic behaviour of radial solutions near the blow-up point to quasi-linear wave equations in two space dimensions, 8 pages. 1994.

♯ 232: Y. Giga, N. Mizoguchi, Existence of periodically evolving convex curves moved by anisotropic curvature, 12 pages. 1994.

♯ 233: C. Dohmen, Y. Giga, N. Mizoguchi, Existence of selfsimilar shrinking curves for anisotropic curvature flow equations, 13 pages. 1994.

♯ 234: T. Nakazi, M. Yamada, Invertible Toeplitz operators and uniform algebras, 14 pages. 1994.

♯ 235: C. Dohmen, A Pohozaev-type inequality for a quasilinear Haraux-Weissler equation, 6 pages. 1994.

♯ 236: T. Mikami, Large deviations for the first exit time on small random perturbations of dynamical systems, 32 pages. 1994.

♯ 237 T. Ozawa, Characterization of Trudinger's inequality, 7 pages. 1994.

♯ 238 T. Hibi, Buchsbaum complexes with linear resolutions, 12 pages. 1994.

♯ 239 Y. Giga, N. Mizoguchi, On time periodic solutions of the Dirichlet problem for degenerate parabolic equations of nondivergence type, 23 pages. 1994.

♯ 240 C. Dohmen, Existence of Fast Decaying Solutions to a Haraux-Weissler Equation With a Prescribed Number of Zeroes, 12 pages. 1994.

♯ 241 K. Sugano, Note on H-separable Frobenius extensions, 8 pages. 1994.

♯ 242 J. Zhai, Some Estimates For The Blowing up Solutions of Semilinear Heat Equations, 11 pages. 1994.

♯ 243 N. Hayashi, K. Kato and T. Ozawa, Dilation Method and Smoothing Effect of the Schrödinger Evolution Group, 10 pages. 1994.

♯ 244 D. Lehmann, T. Suwa, Residues of holomorphic vector fields relative to singular invariant subvarieties, 26 pages. 1994.

♯ 245 H. Kubo, Slowly decaying solutions for semilinear wave equations in odd space dimensions, 30 pages. 1994.

♯ 246 T. Nakazi, M. Yamada, (A_2)-Conditions and Carleson Inequalities, 27 pages. 1994.

♯ 247 N. Hayashi, K. Kato and T. Ozawa, Dilation Method and smoothing Effect of Solutions to the Benjamin-ono Equation, 17 pages. 1994.

♯ 248 H. Kikuchi, Sheaf cohomology theory for measurable spaces, 12 pages. 1994.

♯ 249 A. Inoue, Tauberian theorems for Fourier cosine transforms, 9 pages. 1994.

♯ 250 S. Izumiya, G. T. Kossioris, Singularities for viscosity solutions of Hamilton-Jacobi equations, 23 pages. 1994.

♯ 251 H. Kubo, K. Kubota, Asymptotic behaviors of radially symmetric solutions of $\Box u = |u|^p$ for super critical values p in odd space dimensions, 51 pages. 1994.

♯ 252 T. Mikami, Large Deviations and Central Limit Theorems for Eyraud-Farlie-Gumbel-Morgenstern Processes, 9 pages. 1994.

♯ 253 T. Nishimori, Some remarks in a qualitative theory of similarity pseudogroups, 19 pages. 1994.

♯ 254 T. Suwa, Residues of complex analytic foliations relative to singular invariant subvarieties, 15 pages. 1994.

On Triple Mutual Information

Toru Tsujishita

1994.4.12

Introduction

The mutual information of two random variables plays fundamental roles in many areas of applied mathematics. This quantity can be generalized to finite sets of random variables by (1). In contrast to the non-negativity of the usual mutual information, the triple mutual information can be negative as well as positive and its sign gives us a rough indication of the mode of mutual dependency among three random variables. The purpose of this note is to determine the range of the triple mutual information and to examin when the extremals are attained.

1 Statement of Results

Let (X, B, p) be a probability space. A *finite random variable* is a mesurable map defined on X with finite range space.

The *entropy* of a finite random variable$f$ is defined by

$$H(f) := \sum_{v \in \mathrm{Image}(f)} p(f^{-1}v) \log p(f^{-1}v).$$

The multiple mutual information is defined by

$$I(f_1, \cdots, f_n) := \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 \leq \cdots \leq i_k} H(f_{i_1}, \cdots, f_{i_k}), \tag{1}$$

where $H(f_{i_1}, \cdots, f_{i_k})$ denotes the entropy of the finite random variable $f_{i_1} \times \cdots \times f_{i_k}$.

For example,

$$I(f_1, f_2) = H(f_1) + H(f_2) - H(f_1, f_2),$$

$$\begin{aligned} I(f_1, f_2, f_3) &= H(f_1) + H(f_2) + H(f_3) \\ &\quad - H(f_1, f_2) - H(f_2, f_3) - H(f_3, f_1) + H(f_1, f_2, f_3). \end{aligned}$$

The quantity $I(f_1, f_2)$ is the mutual information of two random variables f_1, f_2 and the following are some of its fundamental properties (cf. [1]):

Proposition 1

$$0 \leq I(f_1, f_2) \leq \min\{\, H(f_1), H(f_2) \,\}, \tag{2}$$

$$I(f_1, f_2) = 0 \Leftrightarrow f_1 \textit{ and } f_2 \textit{ are independent}, \tag{3}$$

$$I(f_1, f_2) = H(f_1) \Leftrightarrow f_2 = \varphi \circ f_1 \textit{ for some map } \varphi. \tag{4}$$

The main result is the following theorem which asserts similar properties for the triple mutual information funciton.

Theorem 1

(i) $\quad -\min_{i,j} H(f_i | f_j) \leq I(f_1, f_2, f_3) \leq \min_i H(f_i)$

(ii) $\quad I(f_1, f_2, f_3) = H(f_i)$

$$\Leftrightarrow \; f_i = \varphi_j(f_j) = \varphi_k(f_k) \textit{ a.e. for some maps } \varphi_j \textit{ and } \varphi_k.$$

(iii) $$I(f_1, f_2, f_3) = -H(f_i|f_j)$$
$$\Leftrightarrow \begin{cases} f_i = \varphi(f_j, f_k) \textit{ a.e. for a map } \varphi, \\ f_i \textit{ and } f_k \textit{ are independent} \end{cases}$$

Here $\{ i, j, k \} = \{ 1, 2, 3 \}$.

The avarage conditional entorpy function $H(f|g)$ of finite random variables , used in the statement, is defined as follows:

$$H(f|g) := \sum_{z \in \mathrm{Image}(g)} p(g^{-1}z) H(f|g^{-1}(z)),$$

Here, for an event $E \subset X$ with $p(E) > 0$, $H(f|E)$ denotes the conditional entropy of f defined by

$$H(f|E) := \sum_{x \in \mathrm{Image}(f)} p(x|E) \log p(x|E),$$

with $p(x|E) := p(f^{-1}x \cap E)/p(E)$.

2 properties of avarage conditional entropy

The following proposition states important properties of the avarage conditional entropy function, which we need in the proof of the main theorem. See [1] for the proof.

Proposition 2

$$H(f_1|f_2) = H(f_1, f_2) - H(f_2), \tag{5}$$

$$0 \leq H(f_1|f_2) \leq H(f_1), \tag{6}$$

$$H(f_1|f_2) = 0 \; \Leftrightarrow \; f_1 = \varphi \circ f_2 \text{ a.e. } \textit{for some } \varphi, \tag{7}$$

$$H(f_1|f_2) = H(f_1) \Leftrightarrow f_1 \textit{ and } f_2 \textit{ are independent,} \tag{8}$$

$$I(f_1, f_2) = H(f_1) - H(f_1|f_2). \tag{9}$$

We denote

$$H(f_1, \cdots, f_n | g_1, \cdots, g_m) := H(f_1 \times \cdots \times f_n | g_1 \times \cdots \times g_n)$$

and we generalize (1) as

$$I(f_1, \cdots, f_n | g_1, \cdots, g_m) := \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 \leq \cdots \leq i_k} H(f_{i_1}, \cdots, f_{i_k} | g_1, \cdots, g_m) \ . \tag{10}$$

The following lemma follows immediately from the definitions by simple calculation.

Lemma 3

$$I(f_1, f_2, f_3) = I(f_1, f_2) - I(f_1, f_2 | f_3), \tag{11}$$

$$I(f_1, f_2, f_3) = H(f_1) - H(f_1|f_2) - H(f_1|f_3) + H(f_1|f_2, f_3). \tag{12}$$

We need also the following lemma in the proof of the main theorem.

Lemma 4

$$0 \leq I(f_1, f_2 | f_3) \leq \min\{ H(f_1|f_3), H(f_2|f_3) \}, \tag{13}$$

$$I(f_1, f_2 | f_3) = H(f_1 | f_3) \Leftrightarrow f_1 = \varphi(f_2, f_3) \text{ a.e. } \textit{for some map } \varphi. \tag{14}$$

Note that the lemma implies that, when the maximum is attained, the equality $H(f_1|f_3) = H(f_2|f_3)$ holds.

Proof. Using (5) we have

$$\begin{aligned} I(f_1, f_2|f_3) &= H(f_1|f_3) + H(f_2|f_3) - H(f_1, f_2|f_3) \\ &= H(f_1|f_3) + H(f_2, f_3) - H(f_1, f_2, f_3) \\ &= H(f_1|f_3) - H(f_1|f_2, f_3) \\ &\le H(f_1|f_3). \end{aligned}$$

The equality holds if and only if $H(f_1|f_2, f_3) = 0$, which means $f_1 = \varphi(f_2, f_3)$ a.e. for some map φ by (7). **q.e.d.**

3 Proof of Theorem 1

Proof of (i). By (11) and (1)

$$I(f_1, f_2, f_3) \le I(f_1, f_2) \le \min\{\, H(f_1), H(f_2) \,\}$$

$$I(f_1, f_2, f_3) \ge -I(f_1, f_2|f_3) \ge -\min\{\, H(f_1|f_3), H(f_2|f_3) \,\}.$$

Since the left hand side is symmetric, we obtain (i).

Proof of (ii). Suppose

$$I(f_1, f_2, f_3) = H(f_1).$$

Then by (11),

$$I(f_1, f_2) - I(f_1, f_2|f_3) = H(f_1).$$

Hence by (9)

$$H(f_1|f_2) = -I(f_1, f_2|f_3).$$

Since the left hand sideis nonnegative and the right hand sideis nonpositive, we obtain

$$\begin{aligned} H(f_1|f_2) &= 0, && (15)\\ I(f_1, f_2|f_3) &= 0. && (16) \end{aligned}$$

The equality (15) implies

$$f_1 = \varphi_2 \circ f_2 \quad \text{a.e.} \tag{17}$$

for a map φ_2. The other equality implies

$$H(f_1, f_2|f_3) = H(f_1|f_3) + H(f_2|f_3).$$

from which we obtain, by (15),

$$H(f_1, f_2, f_3) - H(f_1, f_3) - H(f_2, f_3) + H(f_3) = 0. \tag{18}$$

Now (17) implies by virtue of (5) and (7),

$$H(f_1, f_2, f_3) - H(f_2, f_3) = 0,$$

whence together with (18) we obtain $H(f_1|f_3) = 0$, which means

$$f_1 = \varphi_3 \circ f_3 \quad \text{a.e.}$$

for some map φ_3.

Conversely, suppose

$$f_1 = \varphi_2 \circ f_2, \qquad f_3 = \varphi_3 \circ f_3 \text{ a.e.}$$

for some maps φ_2 and φ_3. Then by 7

$$H(f_1|f_2) = H(f_1|f_3) = H(f_1|f_2, f_3) = 0.$$

Hence, by (12), we obtain $I(f_1, f_2, f_3) = H(f_1)$.

Proof of (iii). We my assume $i = 1$ and $j = 2$ for simplity. Suppose first

$$I(f_1, f_2, f_3) = -H(f_1|f_2). \tag{19}$$

By (12)

$$I(f_1, f_2, f_3) + H(f_1|f_2) \begin{aligned}[t] &= H(f_1) - H(f_1|f_3) + H(f_1|f_2, f_3) \\ &= I(f_1, f_3) + H(f_1|f_2, f_3) \end{aligned} \tag{20}$$

Hence (19) holds only if

$$H(f_1|f_2, f_3) = 0 \text{ and } I(f_1, f_3) = 0, \tag{21}$$

which mean that

$$\begin{cases} f_1 = \varphi(f_2, f_3) \text{ for some map } \varphi \\ \\ f_1 \text{ and } f_3 \text{ are independent.} \end{cases} \tag{22}$$

Conversely suppose (22) holds. Then (21) holds. Then by (20) we obtain

$$I(f_1, f_2, f_3) = -H(f_1|f_2).$$

This completes the proof of the main theorem.

References

[1] D. Welsh, *Codes and cryptography* Oxford Univ Press, 1988.