Six New Non-Shannon Information Inequalities

Randall Dougherty
Center for Communications Research
4320 Westerra Court
San Diego, CA 92121-1969
Email: rdough@ccrwest.org

Christopher Freiling
Department of Mathematics
California State University
5500 University Parkway
San Bernardino, CA 92407-2397
Email: cfreilin@csusb.edu

Kenneth Zeger
Dept. of Electrical and Computer Eng.
University of California, San Diego
La Jolla, CA 92093-0407
Email: zeger@ucsd.edu

***Abstract*— All unconstrained information inequalities in three or fewer random variables are known to be "Shannon-type", in that they are nonnegative linear combinations of instances of the inequality $I(A;B|C) \geq 0$. In 1998, Zhang and Yeung gave the first example of an information inequality in four variables that is not "Shannon-type". Here we give six new unconstrained non-Shannon information inequalities in four variables. The new inequalities are independent of each other and of the Zhang-Yeung inequality.**

I. Introduction

Standard information inequalities are generally "Shannon-type" inequalities, which can be derived by combining special cases of the nonnegativity of conditional mutual information. These were the only known types of information inequalities until Zhang and Yeung in 1998 published a non-Shannon-type information inequality [8]. Some other results on non-Shannon-type information inequalities have been given by Lněnička [2]. Makarychev, Makarychev, Romashchenko, and Vereshchagin [3], Zhang [6], and Zhang and Yeung [7].

II. Information Inequalities

For collections A, B, and C of jointly related discrete random variables, denote the *entropy* of A by $H(A)$, the *conditional entropy* of A given B by $H(A|B)$, the *mutual information* between random variables A and B by $I(A;B)$, and the *conditional mutual information* between random variables A and B given C by $I(A;B|C)$.

Definition II.1. Let p be a positive integer, and let $S_1, \ldots, S_k$ be subsets of $\{1, \ldots, p\}$. Let $\alpha_i \in \mathbf{R}$ for $1 \leq i \leq k$. An inequality of the form

$$\alpha_1 H(\{A_i : i \in S_1\}) + \cdots + \alpha_k H(\{A_i : i \in S_k\}) \geq 0$$

is called an *information inequality* if it holds for all jointly distributed random variables $A_1, \ldots A_p$.

As an example, taking $p = 2$, $S_1 = \{1\}$, $S_2 = \{2\}$, $S_3 = \emptyset$, $S_4 = \{1,2\}$, $\alpha_1 = \alpha_2 = 1$, and $\alpha_4 = -1$, one obtains $H(A_1) + H(A_2) - H(A_1, A_2) \geq 0$, which is an information inequality since it is always true (this can be more succinctly expressed as $I(A_1; A_2) \geq 0$).

Since all conditional entropies and all conditional mutual informations can be written as linear combinations of joint entropies, any valid linear inequality involving conditional entropies and conditional mutual informations will also be called an information inequality. The textbook [5] refers to information inequalities as "the laws of information theory".

The information inequalities $H(X) \geq 0$, $H(X|Y) \geq 0$, and $I(X;Y) \geq 0$ were originally given in 1948 by Shannon [4] and can all be obtained as special cases (e.g. see [5]) of the inequality

$$I(A;B|C) \geq 0 \tag{1}$$

or equivalently

$$H(A,C) + H(B,C) \geq H(C) + H(A,B,C). \tag{2}$$

A *Shannon-type information inequality* is any information inequality that is (or can be rearranged[1] to be) of the form

$$\sum_i \alpha_i I(A_i; B_i|C_i) \geq 0 \tag{3}$$

where each α_i is a nonnegative real number. Virtually every known result in information theory that makes use of an information inequality only makes use of Shannon-type information inequalities.

Any information inequality that cannot be expressed in the form (3) will be called a *non-Shannon-type information inequality*. It is known [5, p. 308] that all information inequalities containing three or fewer random variables are Shannon-type inequalities. The first known non-Shannon-type information inequality was published in 1998 by Zhang and Yeung and is stated in Theorem II.2 below. To date, it is the only published unconstrained non-Shannon-type information inequality for four random variables. The Zhang-Yeung inequality has recently been applied to network coding to demonstrate that Shannon-type information inequalities are in general insufficient for computing the coding capacity of a network [1].

A collection of information inequalities is said to be an *independent set* if none of its inequalities can be expressed as a nonnegative linear combination of the set's other inequalities (including versions obtained by substitution of variables) and Shannon-type inequalities. Note that, in particular, any

[1] We allow replacement of 0 by $H(\emptyset)$. This seemingly trivial technicality is needed, for example, in order to be able to assert that $I(A;B) \geq 0$ is of the form $I(A;B|\emptyset) \geq 0$.

inequality in an independent set is a non-Shannon-type information inequality.

In this paper, we present six new non-Shannon-type information inequalities. These form an independent set of inequalities when taken together with the Zhang-Yeung non-Shannon-type inequality.

Let $n \geq 2$, and let $s_1, s_2, \ldots, s_{2^n-1}$ be the list of nonempty subsets of $\{1, \ldots n\}$, where s_i contains j if and only if the jth bit from the right, in the binary representation of i, is 1. For any collection of random variables, $X = \{X_1, \ldots X_n\}$, define the subset of random variables

$$Y_i(X) = \{X_j : j \in s_i\},$$

the list of entropies

$$h(X) = (H(Y_1(X)), \ldots, H(Y_{2^n-1}(X))),$$

and the set of entropy lists (or "entropic" vectors)

$$\Gamma_n^* = \{h(X) : X \text{ is a collection of } n \text{ jointly distributed random variables}\}.$$

Also, let Γ_n denote the set of all points in $\mathbf{R}^{2^n-1}$ which satisfy every Shannon-type inequality (where the coordinates represent the 2^n-1 entropies of the subsets of n random variables).

As an example, taking X_1 and X_2 to be independent uniform binary random variables gives $H(X_1) = H(X_2) = 1$ and $H(X_1, X_2) = 2$, so $(1, 1, 2) \in \Gamma_2^*$.

We use the symbols $\subseteq$ and $\subset$ to denote subsets and proper subsets, respectively. Clearly $\Gamma_n^* \subseteq \Gamma_n$. The Shannon outer bound, Γ_n, is known to be an unbounded convex set with planar boundaries. The closure, $\bar{\Gamma}_n^*$, of Γ_n^*, is known [5, p. 306] to be a convex cone in $\mathbf{R}^{2^n-1}$. Also, it has been known that $\Gamma_2^* = \bar{\Gamma}_2^* = \Gamma_2$, $\Gamma_3^* \subset \bar{\Gamma}_3^* = \Gamma_3$, and $\Gamma_4^* \subset \bar{\Gamma}_4^* \subset \Gamma_4$. The proper inclusion $\bar{\Gamma}_4^* \subset \Gamma_4$ is a consequence of the following result of Zhang and Yeung.

Theorem II.2. *[8] The following is a 4-variable non-Shannon-type information inequality:*

$$2I(C;D) \leq I(A;B)+I(A;C,D)+3I(C;D|A)+I(C;D|B).$$

It was previously unknown whether Shannon-type information inequalities together with the Zhang-Yeung non-Shannon-type inequality completely determine the space $\bar{\Gamma}_4^*$. Our results here demonstrate that the space $\bar{\Gamma}_4^*$ is not so determined, but is indeed a bit more complicated.

The Zhang-Yeung inequality yields a new outer bound $\Gamma_4^{(1)}$ satisfying $\bar{\Gamma}_4^* \subseteq \Gamma_4^{(1)} \subset \Gamma_4$. The set $\Gamma_4^{(1)}$ is formed by "chopping off" 12 pieces of Γ_4 by planar cuts in $\mathbf{R}^{15}$. Our results in this paper, given in the next section, yield an improved outer bound, $\Gamma_4^{(2)}$, satisfying $\bar{\Gamma}_4^* \subseteq \Gamma_4^{(2)} \subset \Gamma_4^{(1)} \subset \Gamma_4$. Thus, in particular, we establish that $\bar{\Gamma}_4^* \neq \Gamma_4^{(1)}$.

It can be shown that the cone Γ_4 has 41 extremal rays, the cone $\Gamma_4^{(1)}$ has 89 extremal rays, and the cone $\Gamma_4^{(2)}$ has 299 extremal rays.

III. New Inequalities

The following theorem summarizes our main results.

Theorem III.1. *The following are 4-variable non-Shannon-type information inequalities:*

(i)
$$2I(A;B) \leq 3I(A;B|C) + 3I(A;C|B) + 3I(B;C|A) + 2I(A;D) + 2I(B;C|D).$$

(ii)
$$2I(A;B) \leq 4I(A;B|C) + I(A;C|B) + 2I(B;C|A) + 3I(A;B|D) + I(B;D|A) + 2I(C;D).$$

(iii)
$$2I(A;B) \leq 3I(A;B|C) + 2I(A;C|B) + 4I(B;C|A) + 2I(A;C|D) + I(A;D|C) + 2I(B;D) + I(C;D|A).$$

(iv)
$$2I(A;B) \leq 5I(A;B|C) + 3I(A;C|B) + I(B;C|A) + 2I(A;D) + 2I(B;C|D).$$

(v)
$$2I(A;B) \leq 4I(A;B|C) + 4I(A;C|B) + I(B;C|A) + 2I(A;D) + 3I(B;C|D) + I(C;D|B).$$

(vi)
$$2I(A;B) \leq 3I(A;B|C) + 2I(A;C|B) + 2I(B;C|A) + 2I(A;B|D) + I(A;D|B) + I(B;D|A) + 2I(C;D).$$

Furthermore, these six inequalities, together with the inequality in Theorem II.2, form an independent set of information inequalities.

Each of the non-Shannon-type information inequalities in Theorem III.1 can be expressed in the form given in Definition II.1, solely in terms of entropy functions. For example, the first inequality in Theorem III.1 can equivalently be written as:

$$\begin{aligned} &- 3H(A) - 5H(B) + 8H(A,B) - 3H(C) + 6H(A,C) \\ &+ 6H(B,C) - 9H(A,B,C) - 2H(A,D) + 2H(B,D) \\ &+ 2H(C,D) - 2H(B,C,D) \geq 0. \end{aligned}$$

Next, we will provide a proof sketch of the first of the six inequalities in Theorem III.1. Due to space limitations and to provide clarity, we will omit many laborious details of the proof as well as the proofs of the other five parts of the theorem.

Definition III.2. Given jointly distributed random variables A and S and collections B and D of random variables, S is said to be a *D-copy of A over B* if the following two conditions are satisfied:

(C1) $I(S; A, D|B) = 0$.

(C2) The joint probability distributions of (S, B) and (A, B) are equal.

We note that condition (C1) is sometimes written as $S \to B \to (A, D)$. The following lemma is based on a technique used in [8].

Lemma III.3. *[5, Lemma 14.8] Given jointly distributed random variable A and collections of random variables B and D, there exists a random variable S, jointly related to A, B, and D, such that S is a D-copy of A over B.*

The next lemma is used in the proof of Theorem III.1(i).

Lemma III.4. *The following is a 5-variable information inequality:*

$$\begin{aligned}&I(A;B)\\ &\le I(A,R;D) + I(D;R|B,C) + I(B;C|A,R) + I(B;C|D)\\ &\quad + I(A;C|B,R) + I(A;B|C,R) + I(A;B|C).\end{aligned}$$

Proof. By expanding mutual informations into entropies and cancelling terms, one can verify the following 6-variable identity:

$$\begin{aligned}&I(A;B)\\ &\quad + I(D;S|A,B,C,R) && (4)\\ &\quad + I(D;S|C) && (5)\\ &\quad + I(D;S|B) && (6)\\ &\quad + I(D;R|B,C,S) && (7)\\ &\quad + I(B,C;S|A,D,R) && (8)\\ &\quad + I(B;C|D,S) && (9)\\ &\quad + I(B;C|A,R,S) && (10)\\ &\quad + I(A;S|B,C,D,R) && (11)\\ &\quad + I(A;S|C,R) && (12)\\ &\quad + I(A;S|B,R) && (13)\\ &\quad + I(A,R;D|S) && (14)\\ &= I(A,R;D) + I(D;R|B,C) + I(B;C|A,R) + I(B;C|D)\\ &\quad + I(A;C|B,R) + I(A;B|C,R) + I(A;B|C)\\ &\quad + 3I(A,D;S|B,C,R) && (15)\\ &\quad - (H(S) - H(A)) && (16)\\ &\quad + (H(S,B) - H(A,B)) && (17)\\ &\quad + (H(S,C) - H(A,C)) && (18)\\ &\quad - (H(S,B,C) - H(A,B,C)) && (19)\\ &\quad + (H(S,B,R) - H(A,B,R)) && (20)\\ &\quad + (H(S,C,R) - H(A,C,R)) && (21)\\ &\quad - 2(H(S,B,C,R) - H(A,B,C,R)). && (22)\end{aligned}$$

Each of the terms in lines (4)–(14) is a conditional mutual information and is therefore nonnegative. Thus, if the terms in (4)–(14) are erased and the "=" is replaced by "$\le$", then we obtain a 6-variable Shannon-type inequality. By Lemma III.3 we may choose S to be a D-copy of A over (B, C, R). Then, the term in line (15) is zero by condition (C1), and each of the terms in lines (16)–(22) equals zero by condition (C2). □

We note (without proof) that the inequality in Lemma III.4 is a 5-variable non-Shannon-type information inequality.

Proof of Theorem III.1(i): By expanding mutual informations into entropies and cancelling terms, one can verify the following 5-variable identity:

$$\begin{aligned}&2I(A;B)\\ &\quad + 2I(C;R|A) && (23)\\ &\quad + 3I(C;R|B) && (24)\\ &\quad + I(D;R|B) && (25)\\ &\quad + I(D;R|C) && (26)\\ &\quad + 3I(D;R|A,B,C) && (27)\\ &\quad + 2I(B,C;R|A,D) && (28)\\ &\quad + I(B;C|D,R) && (29)\\ &\quad + I(A;D|R) && (30)\\ &\quad + 2I(A;R|B,C,D) && (31)\\ &\quad + [I(A,R;D) + I(D;R|B,C) + I(B;C|A,R) && (32)\\ &\quad + I(B;C|D) + I(A;C|B,R) + I(A;B|C,R) && (33)\\ &\quad + I(A;B|C) - I(A;B)] && (34)\\ &= 3I(A;B|C) + 3I(A;C|B) + 3I(B;C|A)\\ &\quad + 2I(A;D) + 2I(B;C|D)\\ &\quad + 7I(C,D;R|A,B) && (35)\\ &\quad - (H(R) - H(C)) && (36)\\ &\quad + 3(H(R,A) - H(C,A)) && (37)\\ &\quad + 3(H(R,B) - H(C,B)) && (38)\\ &\quad - 5(H(R,A,B) - H(C,A,B)). && (39)\end{aligned}$$

Each of the terms in lines (23)–(31) is a conditional mutual information and is therefore nonnegative. Thus, if the terms in (23)–(31) are erased and the "=" is replaced by "$\le$", then we obtain a 5-variable Shannon-type inequality. The expression spanning lines (32)–(34) is nonnegative by Lemma III.4. By Lemma III.3 we may choose R to be a D-copy of C over (A, B). Then, the term in line (35) is zero by condition (C1), and each of the terms in lines (36)–(39) equals zero by condition (C2). □

The fact that the six inequalities in the theorem statement and the Zhang-Yeung inequality form an independent set of information inequalities can be verified (somewhat laboriously) by finding, for each of the seven inequalities, a point in Γ_4 which does not satisfy the inequality, but which does satisfy each of the other six inequalities (including substituted forms). For the inequality given in Theorem III.1(i), such a point is $(3, 3, 4, 4, 6, 6, 7, 4, 7, 6, 7, 5, 7, 7, 7)$.

Corollary III.5. $\bar{\Gamma}_4^* \neq \Gamma_4^{(1)}$.

ACKNOWLEDGMENT

This work was supported by the Institute for Defense Analyses and the National Science Foundation. We thank Raymond Yeung and Ying-On Yan for their efforts in making available the Information Theoretic Inequality Prover (ITIP). The ITIP software facilitated various computations for this paper. We also thank Raymond Yeung for clarifying the Abstract.

(Submitted to ISIT06 on January 9, 2006.)

REFERENCES

[1] R. Dougherty, C. Freiling, and K. Zeger, "Matroids, networks, and non-Shannon information inequalities", *IEEE Transactions on Information Theory*, (submitted January 2006).
[available on-line at: http://code.ucsd.edu/zeger/pubs.html]

[2] R. Lněnička, "On the tightness of the Zhang-Yeung inequality for Gaussian vectors", *Communications in Information and Systems*, vol. 3, no. 1, pp. 41-46, June 2003.

[3] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies", *Communications in Information and Systems*, vol. 2, no. 2, pp. 147-166, December 2002.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[5] R. W. Yeung, *A First Course in Information Theory*, Kluwer, 2002.

[6] Z. Zhang, "On a new non-Shannon type information inequality", *Communications in Information and Systems*, vol. 3, no. 1, pp. 47-60, June 2003.

[7] Z. Zhang and R. W. Yeung, "A non-Shannon-type conditional inequality of information quantities", *IEEE Transactions on Information Theory*, vol. 43, pp. 1982-1985, November 1997.

[8] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities", *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1440-1452, July 1998.