THE COPY LEMMA AND NON-SHANNON INFORMATION INEQUALITIES

MICHAEL TANG

ABSTRACT. An *information inequality* is a linear inequality involving the entropies $H(X)$ of (sets of) random variables. As shown by Zhang and Yeung [2], not all information inequalities are "Shannon-type" inequalities, ones that can be derived by repeatedly applying a basic result called Shannon's inequality. Dougherty, Freiling, and Zeger [1] used a technique called the "copy lemma" to generate a list of 214 *non*-Shannon-type information inequalities, all of a particular special form. In this semi-expository paper, we introduce the relevant concepts to the study of information inequalities, and make some observations about Doughtery, Freiling, and Zeger's work that may explain the occurrence of this special form.

1. PRELIMINARIES

We first recall some basic definitions.

Definition 1.1. Let X be a random variable taking values in a finite set $\mathcal{X}$, and let $p(x) := \Pr[X = x]$ for $x \in \mathcal{X}$. The *entropy* of X is defined by

$$H(X) := \sum_x -p(x) \log p(x).$$

(We may assume that $p(x) > 0$ for all $x \in \mathcal{X}$; otherwise we may shrink the set $\mathcal{X}$.)

The entropy of a random variable X measures, in some sense, its "uncertainty": how difficult it is for one to "guess" the value of a given sample of X.

Some important notions related to entropy are as follows:

Definition 1.2. The *mutual information* of two random variables A and B is given by

$$I(A; B) := H(A) + H(B) - H(AB),$$

and the *conditional mutual information* of A and B *given* a random variable C is

$$I(A; B|C) := H(AC) + H(BC) - H(ABC) - H(C).$$

(For convenience, we will often use concatenation to denote subsets in this paper. For example, $H(AB)$ in the above means $H(A, B)$, where A, B is considered as a joint random variable.)

The mutual information $I(A; B)$ is often interpreted as "how much A *knows* about B," or vice versa. For example, if B is a *function* of A, then $H(AB) = H(A)$; thus, $I(A; B) = H(A) + H(B) - H(AB) = H(B)$. That is, A knows "everything" about B. On the other hand, if A and B are independent random variables, it can be checked that $H(AB) = H(A) + H(B)$, that is, $I(A; B) = 0$.

Observe that if C is taken to be trivial (that is, C takes only one value), then the formula for $I(A;B|C)$ reduces to the formula for $I(A;B)$. This agrees with our intuition that if C is trivial, then we are not "given" anything, so we should just get the mutual information between A and B.

A fundamental result in information theory is the following inequality:

Proposition 1.3 (Shannon's inequality)**.** *For all random variables X, Y, Z we have*

$$I(X;Y|Z) \geq 0.$$

Thus, (conditional) mutual information is always nonnegative: the case mentioned above where A and B are independent is the "worst" it can get.

By adding together (nonnegatively weighted) applications of Shannon's inequality, we can build other true inequalities involving entropy; such inequalities, as we will see, are collectively called "Shannon-type inequalities."

2. Information inequalities

Motivated by the existence of Shannon's inequality, we might seek to study the possible inequalities that can be written down about entropies of variables. This leads to the following definition:

Definition 2.1. An *information inequality* on (four) variables A, B, C, D is a *true* inequality of the form

$$0 \leq \sum_{\emptyset \neq X \subseteq ABCD} C_X H(X),$$

with real coefficients C_X.

(It isn't hard to generalize this definition to n variables, but the case of four variables is the one we are interested in.) For example, Shannon's inequality, applied in the form $I(A,B;C) \geq 0$, is an information inequality, because it can be rewritten in terms of entropies as

$$-H(C) + H(AC) + H(BC) - H(ABC) \geq 0.$$

Thus $C_C = -1$, $C_{AC} = 1$, $C_{BC} = 1$, $C_{ABC} = -1$, and all other $C_X = 0$ in this case.

In working with (four-variable) information inequalities, it can be helpful to identify an information inequality $0 \leq \sum C_X HX$ with its "coefficient vector"

$$(C_X) = (C_A, C_B, C_C, C_D, C_{AB}, \ldots, C_{ABCD}) \in \mathbb{R}^{15}.$$

In this way, each information inequality corresponds to a point in the vector space $\mathbb{R}^{15}$. However, it is clear that not every point corresponds to a (valid) information inequality. (As a simple example, if a point p gives rise to an information inequality that is not always an equality, such as Shannon's inequality, then clearly the point $-p$ does not also give rise to an information inequality.) Thus, the following notion naturally arises:

Definition 2.2. For four variables, "*information inequality space*" is the subset M_4 of $\mathbb{R}^{15}$ consisting of all points $(C_A, C_B, \ldots, C_{ABCD})$ which define information inequalities.

Instead of M_4, one might take a more direct approach and try to understand the space of entropies themselves: that is, the space of all vectors of the form

$$(H(A), H(B), \ldots, H(ABCD)) \in \mathbb{R}^{15}.$$

By definition, this latter space, which is denoted Γ_4^* in [1], is the "nonnegative dual" of M_4: the set of all points in $\mathbb{R}^{15}$ whose inner product with any point in M_4 is nonnegative. It is an open problem to understand the structure of the space Γ_4^* (or, equivalently, M_4). (Note that the notation M_4 is not particularly standard.)

As mentioned before, one class of information inequalities are those that are derived from multiple applications of Shannon's basic inequality. In symbols:

Definition 2.3. A *Shannon-type inequality* is an information inequality of the form

$$\sum_{i=1}^{k} \alpha_i I(X_i; Y_i | Z_i) \geq 0,$$

where $\alpha_i \geq 0$ and $X_i, Y_i, Z_i \subseteq ABCD$.

In other words, a Shannon-type inequality is a nonnegative linear combination of applications of Shannon's inequality – an element of the "nonnegative span" of the possible applications of Shannon's inequality.

A reasonable guess might be that the Shannon-type inequalities are the *only* information inequalities. In other words, one might hypothesize that the possible applications of Shannon's inequality "nonnegatively span" the space M_4. It took until the mid-1990s for a counterexample to be discovered, by Zhang and Yeung [2]. They discovered the following inequality, which cannot be written as a linear combination of applications of Shannon's inequality:

Theorem 2.4 (Zhang-Yeung inequality)**.** *For all random variables A, B, C, D, we have*

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

(Here is one reason why the case of four variables is interesting. It turns out that in three or fewer variables, all information inequalities are Shannon-type, so this is the first "non-Shannon" case.)

The Zhang-Yeung inequality was discovered and proved by applying a certain result in entropy theory called the "copy lemma." This work has since been superseded and generalized by Dougherty, Freiling, and Zeger [1], who used the copy lemma together with a large computer search to generate 214 new non-Shannon inequalities. We now examine their work in more detail.

3. The copy lemma and non-Shannon inequalities

The *copy lemma* is the following result, which promises the existence of a random variable satisfying certain conditions:

Lemma 3.1 (Copy lemma)**.** *Let A, B, C, D be random variables. Then there is a random variable R with the following properties:*

(i) The joint random variables (A, B, C) and (A, B, R) have the same distribution.
(ii) $I(CD; R|AB) = 0$.

We say that R is a "$D-$copy of C over AB."

By using the copy lemma, Dougherty, Freiling, and Zeger [1] generated 214 new non-Shannon inequalities with a computer search:

Theorem 3.2 (Dougherty, Freiling, Zeger 2011)**.** *The inequality*

$$\begin{aligned} aI(A;B) \leq\ & bI(A;B|C) + cI(A;C|B) + dI(B;C|A) + eI(A;B|D) \\ & + fI(A;D|B) + gI(B;D|A) + hI(C;D) + iI(C;D|A) \end{aligned}$$

is non-Shannon for each of the following tuples (a,b,c,d,e,f,g,h,i):

$$\begin{gathered} (2,4,2,1,3,1,0,2,0), (2,3,3,1,5,2,0,2,0), (3,6,3,1,6,3,0,3,0), \\ (2,4,2,1,2,0,0,2,3), (2,3,3,2,2,0,0,2,0), (4,6,4,3,4,2,1,4,0), \\ (2,5,2,1,2,0,0,2,0), (2,4,3,1,2,0,0,2,0), (2,4,1,2,2,3,0,2,0), \\ (3,7,4,1,4,1,0,3,0), (4,6,11,3,6,2,0,4,0), \ldots \textit{[203 more]} \end{gathered}$$

In their paper, Dougherty, Freiling, and Zeger wondered why they had only found inequalities of the special form

$$\begin{aligned} aI(A;B) \leq\ & bI(A;B|C) + cI(A;C|B) + dI(B;C|A) + eI(A;B|D) \\ & + fI(A;D|B) + gI(B;D|A) + hI(C;D) + iI(C;D|A), \end{aligned}$$

with 9 parameters. Indeed, a generic information inequality on A, B, C, D would have $2^4 - 1 = 15$ parameters, one for each nonempty subset of $ABCD$. To conclude this note, we present one possible answer.

4. Results

We have found 6 "nice" constraints that completely characterize the form of the above inequalities.

Theorem 4.1. *An information inequality $0 \leq \sum_X C_X H(X)$ can be written in the form*

$$\begin{aligned} aI(A;B) \leq\ & bI(A;B|C) + cI(A;C|B) + dI(B;C|A) + eI(A;B|D) \\ & + fI(A;D|B) + gI(B;D|A) + hI(C;D) + iI(C;D|A) \end{aligned}$$

if and only if the following six constraints hold:

$$C_{ABCD} = \sum_X C_X = \sum_{A \in X} C_X = \sum_{B \in X} C_X = \sum_{C \in X} C_X = \sum_{D \in X} C_X = 0.$$

Proof. The theorem is essentially a statement in linear algebra over $\mathbb{R}^{15}$. To prove it, it suffices to check that (i) the six constraints are linearly independent; (ii) the nine terms $I(A;B)$, $I(A;B|C)$, etc. are linearly independent; (iii) all these nine terms satisfy the six constraints. □

Given the natural form of these constraints, a good direction for future work would be to determine whether or not the method of [1] necessarily produces inequalities $0 \leq \sum_X C_X H(X)$ that satisfy these constraints. It is also an open question, to the best of our knowledge, whether there are any non-Shannon inequalities not of this form that are not easily derived from the 214 above inequalities.

5. Acknowledgments

The author wishes to thank Prof. Andreas Blass for his guidance and discussions. He also wishes to thank the University of Michigan mathematics REU program for the opportunity to do this work, and for being accommodating with regards to the difficult circumstances of the past summer.

References

[1] R. Dougherty, C. Freiling, and K. Zeger, "Non-Shannon Information Inequalities in Four Random Variables", available online at `https://arxiv.org/abs/1104.3602v1`, April, 2011.

[2] Zhen Zhang and Raymond W. Yeung, "On characterization of entropy func- tion via information inequalities," IEEE Trans. Information Theory 44 (1998) 1440-452.