

ON A NEW NON-SHANNON TYPE INFORMATION INEQUALITY*

ZHEN ZHANG†

Abstract. Recently, K. Makarychev, Y. Makarychev, A. Romashchenko and N. Vereshchagin proved a new non-Shannon-type information inequality [3] involving 5 random variables which can be viewed as a natural generalization of the ZY98 inequality. In this article, we give a simple proof of a further generalized version of this new inequality and explain its information theoretical meaning by formulating two information theoretical optimization problems. The first is the approximate representation of the mutual information of two random variables (AR problem) and the second is the approximately independent coverings of the mutual information of two random variables (AIC problem). Two information quantities $coef_{ar}(Z, U)$ and $coef_{aic}(Z, U)$ are defined for a pair of random variables (Z, U) which measure the effectiveness of the optimal AR and the optimal AIC of $I(Z; U)$. This new inequality implies that the sum of these two coefficients is at least 1. This result shows that these two information theoretical optimization problems are contradictory in the sense that the solutions of these two problems can not be good (the corresponding coefficient is close to 0) simultaneously. Two regions \mathcal{R}_{ar} and \mathcal{R}_{ipc} are defined which are related to the case of the existence of independent perfect coverings and the case that the mutual information is not approximately representable, respectively. We proved that

$$\mathcal{R}_{ipc} \subset \mathcal{R}_{ar}.$$

1. Introduction. The entropy function H_{Ω_n} associated with n discrete random variables $\Omega_n = \{X_i : i = 1, 2, \dots, n\}$ is defined as a mapping

$$H_{\Omega_n} : 2^{\{1, 2, \dots, n\}} \rightarrow R^+$$

where for any $\alpha \subseteq \{1, 2, \dots, n\}, \alpha \neq \phi$,

$$H_{\Omega_n}(\alpha) \triangleq H(\{X_i : i \in \alpha\})$$

and

$$H_{\Omega_n}(\phi) \triangleq 0.$$

The Shannon's basic inequalities refer to the following fundamental properties of entropy function:

- If F is an entropy function, then for any two subsets $\alpha, \beta \in \{1, 2, \dots, n\}$,

$$(1) \quad F(\alpha \cup \beta) + F(\alpha \cap \beta) \leq F(\alpha) + F(\beta),$$

- $\alpha \subset \beta$ implies

$$(2) \quad F(\alpha) \leq F(\beta),$$

*Received on July 16, 2001; accepted for publication on November 20, 2002

†Communication Sciences Institute, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2565. E-mail: zzhang@milly.usc.edu

•

$$(3) \quad F(\phi) = 0.$$

Let \mathcal{F}_n be the set of all functions defined on $2^{\{1,2,\dots,n\}}$ taking values in $[0, \infty)$. Define

$$\Gamma_n \stackrel{\text{def}}{=} \{F \in \mathcal{F}_n : F \text{ satisfies (1),(2) and (3)}\}.$$

DEFINITION 1. A function $F \in \mathcal{F}$ is called constructible if and only if there exist n jointly distributed random variables Ω_n such that $F = H_{\Omega_n}$.

Define

$$(4) \quad \Gamma_n^* = \{F \in \mathcal{F}_n : F \text{ is constructible}\}.$$

DEFINITION 2. A function F is called asymptotically constructible if and only if $F \in \bar{\Gamma}_n^*$, the closure of the set Γ_n^* .

We have for any n , $\bar{\Gamma}_n^* \subseteq \Gamma_n$. This means that Γ_n characterizes some properties of entropy function. It is known in [1] that $\Gamma_2^* = \Gamma_2$ and $\bar{\Gamma}_3^* = \Gamma_3$. In 1998, Z. Zhang and R. W. Yeung discovered $\bar{\Gamma}_4^* \neq \Gamma_4$ in [2], which is done by proving a so-called non-Shannon-type information inequality.

THEOREM 1. (ZY98 Inequality) For any 4 discrete random variables X, Y, Z and U ,

$$(5) \quad I(Z; U) \leq I(Z; U|X) + I(Z; U|Y) + I(X; Y) + I(X; Z|U) + I(X; U|Z) + I(Z; U|X),$$

$$(6) \quad I(Z; U) \leq I(Z; U|X) + I(Z; U|Y) + I(X; Y) + I(Y; Z|U) + I(Y; U|Z) + I(Z; U|Y).$$

This inequality has been generalized to any number of random variables in [2]. It has been shown that there exists vectors in Γ_4 for which (6) is not satisfied. In other word, (6) is not provable by basic inequalities for the set of random variables $\{X, Y, Z, U\}$. Although (6) has been found useful in some practical problems, for instance in the theory of fault-tolerant distributed data base [4], its information theoretical meaning has not been fully understood. All well-known information inequalities such as basic inequalities and data processing inequality have clear intuitive information theoretical meanings. The intuitive meanings of these inequalities help us in their applications in various information theoretical problems. Therefore, to give ZY98 inequality and other non-Shannon-type inequalities clear information theoretical explanations is very important. This is the main goal of this paper.

Recently, K. Makarychev, Y. Makarychev, A. Romashchenko and N. Vereshchagin generalized ZY98 inequality to a new non-Shannon-type information inequality [3] involving 5 random variables.

THEOREM 2. Let $\{X, Y, Z, U, V\}$ be 5 discrete random variables. Then

$$(7) \quad I(Z; U) \leq I(Z; U|X) + I(Z; U|Y) + I(X; Y) + I(Z; V|U) + I(Z; U|V) + I(U; V|Z).$$

This inequality is further generalized to any number of random variables as follows:

THEOREM 3. Let $\{X_1, X_2, \dots, X_n, Z, U, V\}$ be $n + 3$ discrete random variables where $n \geq 2$. Then

$$(8) \quad nI(Z; U; V) \leq \sum_{i=1}^n I(Z; U|X_i) + \sum_{i=1}^n H(X_i) - H(X_1, \dots, X_n) + I(Z, U; V).$$

We will call it the new inequality in this article.

2. A Proof of a Conditional Version of the New Inequality. This new inequality has the following conditional version:

THEOREM 4. Let $\{X_1, X_2, \dots, X_n, Z, U, V, Y\}$ be $n + 4$ discrete random variables where $n \geq 2$. Then

$$(9) \quad nI(Z; U; V|Y) \leq \sum_{i=1}^n I(Z; U|X_i, Y) + \sum_{i=1}^n H(X_i|Y) - H(X_1, \dots, X_n|Y) + I(Z, U; V|Y).$$

The proof of this inequality is given below. This proof follows completely the method in the original paper [2], whereas the proof of the new inequality in [3] is quite different and much more complicated.

Proof. Let V_1 be a random variables jointly distributed with X_1, \dots, X_n, Y, Z, U . The joint distribution is given by

$$p(x_1, \dots, x_n, y, z, u, v_1) = \begin{cases} \frac{p(x_1, \dots, x_n, y, z, u)p(y, z, u, v_1)}{p(y, z, u)} & p(y, z, u) > 0, \\ 0 & p(y, z, u) = 0 \end{cases}$$

where $p(x_1, \dots, x_n, y, z, u)$ is the marginal of $p(x_1, \dots, x_n, y, z, u, v)$, the joint distribution of $X_1, \dots, X_n, Y, Z, U, V$, and $p(y, z, u, v_1)$ is identical to the marginal $p(y, z, u, v)$ of $p(x_1, \dots, x_n, y, z, u, v)$. From Lemma 2 of [2], we have for any $i : 1 \leq i \leq n$

$$I(Z; U|Y) - I(Z; U|X_i, Y) - I(Z; U|V, Y) \leq I(X_i; V_1|Y).$$

These imply that

$$\begin{aligned}
& nI(Z; U|Y) - \sum_{i=1}^n I(Z; U|X_i, Y) - nI(Z; U|V, Y) \\
& \leq \sum_{i=1}^n I(X_i; V_1|Y) \\
& = I(X^n; V_1|Y) + \sum_{i=2}^n I(X_i; X^{i-1}; V_1|Y) \\
& \leq I(Z, U; V|Y) + \sum_{i=2}^n I(X_i; X^{i-1}|Y) \\
& = I(Z, U; V|Y) + \sum_{i=1}^n H(X_i|Y) - H(X^n|Y).
\end{aligned}$$

In the proof, X^i stands for (X_1, \dots, X_i) . Rearranging the terms in the above expression, we obtain the desired inequality. \square

This inequality in the unconditional case of 5 random variables has a clear intuitive information theoretical meaning which is discussed in the following sections.

3. Perfect Covering and Perfect Representation of Mutual Information between Two Random Variables. Let Z and U be two random variables. A *perfect covering* of the mutual information between Z and U is a random variable V jointly distributed with Z and U having the property that

$$(10) \quad I(Z; U|V) = 0.$$

The Venn diagram in Figure 1 shows an explanation of this concept.

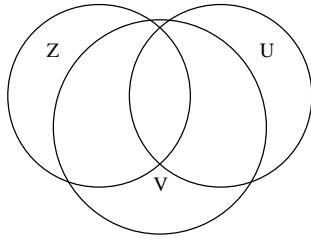


FIG. 1. *Illustration of perfect covering*

For instance, Z is a perfect covering of the mutual information between Z and U because $I(Z; U|Z) = 0$.

A random variable is called a *perfect representation* of the mutual information between Z and U iff

1) V is a perfect covering of the mutual information between Z and U ;

2)

$$(11) \quad H(V|Z) = H(V|U) = 0.$$

By Venn diagram in Figure 2, this is the case where the region corresponding to V is the intersection of the region for Z and U .

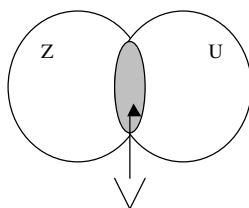


FIG. 2. *Illustration of perfect representation*

This notion has been used by P. Gács and J. Körner in [6], although they did not introduce the name of perfect representation, as a necessary and sufficient condition for the common information of two random variables defined in their paper to be equal to the mutual information .

As an example for the existence of perfect representation of the mutual information of two random variables, let X, Y, W be three independent random variables, and let

$$Z = (X, W) \quad U = (Y, W).$$

Then $V = W$ is a perfect representation of the mutual information between Z and U . Equation (11) implies that V is both a function of Z and a function of U .

Perfect covering exists for any pair of random variables (Z, U) while perfect representation does not exist for most pairs of random variables.

Another concept associated with perfect covering is whether there exist two *independent perfect coverings* of the mutual information between two random variables

Z and U , i.e., whether there exist two random variables X and Y such that

$$I(Z;U|X) = 0, \quad I(Z;U|Y) = 0, \quad I(X;Y) = 0.$$

The following is a non-trivial example of the existence of independent perfect coverings. Let the joint probability mass function of two binary random variables Z and U be

$$P(Z = 0, U = 0) = \frac{1}{2}, \quad P(Z = 0, U = 1) = \frac{1}{4},$$

$$P(Z = 1, U = 0) = \frac{1}{4}, \quad P(Z = 1, U = 1) = 0.$$

Define the joint distribution of Z, U, X, Y by the following conditional probabilities

$$P(X = 0, Y = 1|Z = 0, U = 0) = \frac{1}{2}, \quad P(X = 1, Y = 0|Z = 0, U = 0) = \frac{1}{2},$$

$$P(X = 1, Y = 1|Z = 1, U = 0) = 1, \quad P(X = 0, Y = 0|Z = 0, U = 1) = 1.$$

Then the joint distribution of X, Y is uniform which implies that

$$I(X;Y) = 0.$$

We also have that $p(Z = 0|X = 0) = 1$, $P(U = 0|X = 1) = 1$, $P(Z = 0|Y = 0) = 1$ and $P(U = 0|Y = 1) = 1$. These properties imply that

$$I(Z;U|X) = I(Z;U|Y) = 0.$$

For a given pair of random variables Z and U , whether there exists a pair of independent perfect coverings of their mutual information is an interesting problem. The following definition seems fundamental for this concept. Given r_1, r_2 satisfying $0 < r_1 < 1$ and $\max\{1 - r_1, r_1\} \leq r_2 < 1$, does there exist a pair of random variables (Z, U) satisfying the following conditions?

- 1) There exists a pair of independent perfect coverings for the mutual information of Z, U ,
- 2)

$$\frac{H(Z)}{H(Z) + H(U)} = r_1,$$

- 3)

$$\frac{H(Z, U)}{H(Z) + H(U)} = r_2.$$

Define

$$\mathcal{R}_{ipc} = \{(r_1, r_2) : \text{the answer to the above question is "yes"}\}.$$

The determination of \mathcal{R}_{ipc} is crucial for the study of the concept of independent perfect coverings of mutual information. The example above shows that $r_1 = \frac{1}{2}, r_2 = \frac{1.5}{2 \log 3 - \frac{4}{3}}$ is in \mathcal{R}_{ipc} .

The region \mathcal{R}_{ipc} has the following basic properties:

1. \mathcal{R}_{ipc} is convex,
2. $(r_1, r_2) \in \mathcal{R}_{ipc}$ and $1 > r_3 > r_2$ imply that $(r_1, r_3) \in \mathcal{R}_{ipc}$,
3. the region \mathcal{R}_{ipc} is symmetric with respect to $r_1 = \frac{1}{2}$.

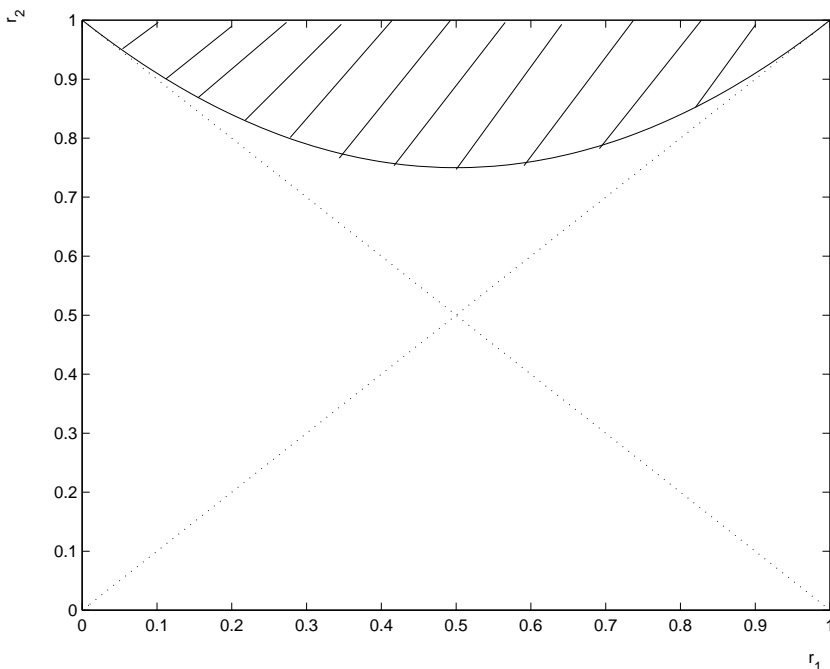


FIG. 3. Illustration of the Region \mathcal{R}_{ipc}

4. Approximate Representation and Coverings of Mutual Information between Two Random Variables.

4.1. Approximate Representation of Mutual Information. While the independent perfect coverings and perfect representation of mutual information are two very restrictive concepts, we may want to relax the requirements slightly so that broader concepts can be introduced.

For a pair of random variables Z and U , the effectiveness of a third random variable V as an approximate covering of $I(Z; U)$ can be measured by the conditional mutual information $I(Z; U|V)$. To define the concept of *approximate representation* of the mutual information, we may relax the requirements of (10) to

$$H(V|Z, U) = 0$$

and use $H(V|Z)$, $H(U|Z)$ to measure the effectiveness of V as a function of Z and of U , respectively. Then the total effectiveness of V as an approximate representation of $I(Z;U)$ can be measured by

$$\begin{aligned} & I(Z;U|V) + H(V|Z) + H(V|U) \\ &= I(Z;U|V) + H(V|Z) - H(V|Z,U) + H(V|U) - H(V|Z,U) \\ &= I(Z;U|V) + I(V;U|Z) + I(V;Z|U). \end{aligned}$$

From the above expression, we may measure this effectiveness by

$$W(Z,U,V) \triangleq I(Z;U|V) + I(U;V|Z) + I(Z;V|U)$$

without requiring $H(V|Z,U) = 0$. (This modification in the definition of this effectiveness measure is justified in Theorem proved below.)

The effectiveness of the optimal approximate representation is given by

$$W(Z,U) = \inf_V W(Z,U,V).$$

THEOREM 5.

Let (Z^n, U^n) be n independently and identically distributed copies of (Z, U) .

1. Let V be a random variable jointly distributed with Z, U . Then there exists $V^{(n)} = f(Z^n, U^n)$, such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(V^{(n)}|Z^n) = I(V;U|Z)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(V^{(n)}|U^n) = I(V;Z|U)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(Z^n;U^n|V^{(n)}) = I(Z;U|V).$$

2. $W(Z^n, U^n) = nW(Z, U)$.

Proof. Let V, U, Z take their values from finite sets $\mathcal{V}, \mathcal{U}, \mathcal{Z}$. We prove the following Lemma:

LEMMA 1. Let (Z^n, U^n) be n i.i.d. copies of (Z, U) . Then for any $\epsilon > 0$, when n is sufficiently large, there exists

$$2^{n(I(V;Z,U)+\epsilon)}$$

elements (this set is denoted by \mathcal{V}_0) in \mathcal{V}^n and a function

$$f_n : \mathcal{Z}^n \times \mathcal{U}^n \rightarrow \mathcal{V}_0$$

such that the random variable $V^{(n)} = f_n(Z^n, U^n)$ has the following properties:

1.

$$I(Z^n; U^n | V^{(n)}) \leq n(I(Z; U | V) + \epsilon)$$

2.

$$H(V^{(n)} | Z^n) \leq n(I(V; U | Z) + \epsilon)$$

3.

$$H(V^{(n)} | U^n) \leq n(I(V; Z | U) + \epsilon).$$

If this lemma is proved, then the first conclusion of the theorem is true. We prove this lemma by random coding argument. Let $2^{n(I(V; Z, U) + 2\epsilon)}$ elements are selected from the ϵ -typical set $A_\epsilon^{(n)}(V)$ independently with uniform distribution. Then by standard argument it is easy to verify that

$$P_e \triangleq P\{(z^n, u^n) : \text{there doesn't exist } v^n \in \mathcal{V}_0, \\ \text{s.t. } (z^n, u^n, v^n) \in A_\epsilon^{(n)}(Z, U, V)\} \xrightarrow{n \rightarrow \infty} 0$$

where $A_\epsilon^{(n)}(Z, U, V)$ is the set of ϵ -typical sequences for random variables Z, U, V . Define $f(z^n, u^n)$ as randomly selected element v^n of \mathcal{V}_0 with respect to uniform distribution over all elements that satisfy $(z^n, u^n, v^n) \in A_\epsilon^{(n)}(Z, U, V)$ if there exists one, otherwise define $f(z^n, u^n)$ as a fixed element v_0^n of \mathcal{V}_0 . In the following, the expected value operator is taken with respect to the joint distribution of the random function f_n and other random variables involved. We have

$$\begin{aligned} \mathbf{E}H(V^{(n)} | Z^n) &\leq \log \mathbf{E}[|\{v^n \in \mathcal{V}_0 : (v^n, Z^n) \text{ are jointly } \epsilon\text{-typical}\}| + 1] \\ &\leq \log(2^{n(I(V; Z, U) + 2\epsilon)} \cdot 2^{-nH(V) + O(\ln n)} \cdot 2^{n(H(V|Z) + \epsilon)}) \\ &\leq n(I(V; U | Z) + 4\epsilon). \end{aligned}$$

Similarly, we can prove

$$H(V^{(n)} | U^n) \leq n(I(V; Z | U) + 4\epsilon).$$

We also have,

$$\begin{aligned} \mathbf{E}I(Z^n; U^n | V^{(n)}) &= \mathbf{E}[H(v^{(n)} | Z^n) + H(v^{(n)} | U^n) \\ &\quad + H(Z^n) + H(U^n) - H(V^{(n)}) - H(Z^n, U^n, V^{(n)})] \\ &\leq n(I(V; U | Z) + 4\epsilon + I(V; Z | U) + 4\epsilon) \\ &\quad - n(I(V; Z, U) - 2\epsilon) + n(H(Z) + H(U) - H(Z, U)) \\ &= n(I(Z; U) | V) + 10\epsilon. \end{aligned}$$

Re-scaling the parameter ϵ , the lemma is proved. In the proof, we used the fact that $\mathbf{E}H(V^{(n)}) \geq n(I(V; Z, U) - 2\epsilon)$ which is proved as follows. We have

$$H(Z^n, U^n | V^{(n)}) \leq n(H(Z, U | V) + \epsilon) + nP_\epsilon H(Z, U).$$

For large n , this is at most $n(H(Z, U | V) + 2\epsilon)$. This is because Z^n, U^n and $V^{(n)}$ are jointly ϵ -typical unless $V^{(n)}$ is the fixed element v_0^n . Therefore

$$\begin{aligned} H(V^{(n)}) &\leq nH(Z, U) - n(H(Z, U | V) + 2\epsilon) \\ &= n(I(V; Z, U) - 2\epsilon). \end{aligned}$$

To prove the second conclusion, we observe that if $V^{(n)}$ is a random variable satisfying $W(Z^n, U^n, V^{(n)}) \leq W(Z^n, U^n) + \epsilon$, then

$$\begin{aligned} &I(Z^n; U^n | V^{(n)}) + I(Z^n; V^{(n)} | U^n) + I(U^n; V^{(n)} | Z^n) \\ &= H(Z^n | V^{(n)}) + H(U^n | V^{(n)}) - H(Z^n, U^n | V^{(n)}) + \\ &\quad H(Z^n | U^n) - H(Z^n | U^n, V^{(n)}) + H(U^n | Z^n) - H(U^n | Z^n, V^{(n)}) \\ &= \sum_{i=1}^n (H(Z_i | Z^{i-1}, V^{(n)}) + H(U_i | U^{i-1}, V^{(n)}) - H(Z_i, U_i | Z^{i-1}, U^{i-1}, V^{(n)})) + \\ &\quad H(Z_i | U_i) - H(Z_i | Z^{i-1}, U^n, V^{(n)}) + H(U_i | Z_i) - H(U_i | U^{i-1}, Z^n, V^{(n)}) \\ &\geq \sum_{i=1}^n [H(Z_i | V_i) + H(U_i | V_i) - H(Z_i, U_i | V_i)] + \\ &\quad H(Z_i | U_i) - H(Z_i | V_i, U_i) + H(U_i | Z_i) - H(U_i | Z_i, V_i) \\ &= \sum_{i=1}^n W(Z_i, U_i, V_i) \\ &\geq nW(Z, U) \end{aligned}$$

where $V_i = (Z^{i-1}, U^{i-1}, V^{(n)})$. Since ϵ is arbitrary, this implies

$$W(Z^n, U^n) \geq nW(Z, U).$$

Apparently,

$$W(Z^n, U^n) \leq nW(Z, U).$$

The theorem is proved. \square

This theorem shows that $W(Z, U)$ is single-letterizable. Although we have abandoned the requirements that V is a function of Z and U in the definition of $W(Z, U, V)$, when n i.i.d. copies are considered, we proved that $V^{(n)}$ is a function of Z^n, U^n .

THEOREM 6.

$$W(Z, U) \leq \min\{H(Z|U), H(U|Z), I(Z; U)\}.$$

Proof. Taking V as a random variable independent of Z, U gives $W(Z, U, V) = I(Z; U)$.

Taking $V = Z$ gives $W(Z, U, V) = H(Z|U)$

and taking $V = U$ gives $W(Z, U, V) = H(U|Z)$. These special values of $W(Z, U, V)$ prove the theorem. \square

Since $W(Z, U) \leq I(U; Z)$, we can define

$$\text{coef}_{ar}(Z, U) \triangleq \begin{cases} \frac{W(Z, U)}{I(Z; U)} & I(Z; U) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

Dividing by $I(U; Z)$ eliminates the impact of the quantity $I(U; Z)$ on the measurement of effectiveness for the approximate representation. We call coef_{ar} the *coefficient of optimal approximate representation* of $I(Z; U)$. We have

$$0 \leq \text{coef}_{ar} \leq 1$$

DEFINITION 3. *If $\text{coef}_{ar}(Z, U) = 1$, we say that the mutual information of Z and U is not approximately representable.*

Define

$$\mathcal{R}_{ar} \triangleq \{(r_1, r_2) : 1 > r_1 > 0, \max\{r_1, 1 - r_1\} \leq r_2 \leq 1, \exists(Z, U) \text{ s.t.}$$

$$\frac{H(Z)}{H(Z) + H(U)} = r_1, \frac{H(Z, U)}{H(Z) + H(U)} = r_2,$$

and $I(U; Z)$ is not approximately representable.}

The determination of \mathcal{R}_{ar} is a fundamental problem for the concept of approximate representation of the mutual information between two random variables. We list some basic properties of this region as follows:

1. $1 \geq r_3 > r_2$, and $(r_1, r_2) \in \mathcal{R}_{ar} \Rightarrow (r_1, r_3) \in \mathcal{R}_{ar}$,
2. \mathcal{R}_{ar} is convex.

4.2. Approximately Independent Coverings of Mutual Information. As we know, the concept of independent perfect coverings of the mutual information is a very restrictive concept. For most pairs of random variables (Z, U) , independent perfect coverings do not exist. To relax the requirements of the concept, we first allow approximate coverings. Let X, Y be two approximate coverings of $I(Z; U)$. Then the sum $I(Z; U|X) + I(Z; U|Y)$ can be used to measure the covering effectiveness. Next we allow that X and Y be dependent and $I(X; Y)$ is used to measure the dependence of X and Y . Then

$$T(Z, U; X, Y) \triangleq I(Z; U|X) + I(Z; U|Y) + I(X; Y)$$

can be used to measure the effectiveness of the pair (X, Y) as *approximately independent coverings* (AIC) of $I(Z; U)$. Define

$$T(Z, U) \triangleq \inf_{X, Y} T(Z, U; X, Y)$$

This information quantity measures how efficient the optimal AIC of $I(Z; U)$ can be.

THEOREM 7.

$$0 \leq T(Z, U) \leq I(Z; U).$$

Proof. Taking $X = Z$ and $Y = U$ gives $T(Z, U; X, Y) = I(Z; U)$, which implies $T(Z, U) \leq I(Z; U)$. Meanwhile, $T(Z, U) \geq 0$ is obvious.

Currently, we do not know whether $T(Z, U)$ is single-letterizable. That is, we do not know whether it is true that, for n i.i.d copies (Z^n, U^n) of (Z, U) , $T(Z^n, U^n) = nT(Z, U)$. Obviously, we have

$$T(Z^n, U^n) \leq nT(Z, U).$$

So we define

$$\tilde{T}(Z, U) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} T(Z^n, U^n).$$

Since $\tilde{T}(Z, U) \leq I(Z; U)$, we define the *coefficient of the optimal approximately independent coverings* of $I(Z; U)$ as

$$0 \leq \text{coef}_{aic}(Z, U) \triangleq \frac{\tilde{T}(Z, U)}{I(Z; U)} \leq 1.$$

5. An Information Theoretical Explanation of the New Inequality. The new inequality can be recast in the form

$$I(Z; U) \leq I(Z; U|X) + I(Z; U|Y) + I(X; Y) + I(Z; V|U) + I(Z; U|V) + I(U; V|Z)$$

which is also correct for n i.i.d copies (Z^n, U^n) of (Z, U) , i.e.,

$$\begin{aligned} I(Z^n; U^n) &\leq I(Z^n; U^n|X^{(n)}) + I(Z^n; U^n|Y^{(n)}) + I(X^{(n)}; Y^{(n)}) + \\ &\quad I(Z^n; V^{(n)}|U^n) + I(Z^n; U^n|V^{(n)}) + I(U^n; V^{(n)}|Z^n). \end{aligned}$$

Since this is true for any triplet $X^{(n)}, Y^{(n)}, V^{(n)}$, we obtain

$$I(Z^n; U^n) \leq n\tilde{T}(Z, U) + nW(Z, U).$$

Dividing by $nI(Z; U)$, we obtain the following theorem

THEOREM 8.

$$\text{coef}_{ar}(Z, U) + \text{coef}_{aic}(Z, U) \geq 1.$$

This theorem shows that the approximate representation (AR) problem and the approximately independent coverings (AIC) problem are contradictory in the sense that one of the two problems has a good solution (the corresponding coefficient is close to zero) implies that the solution for the other problem is poor (the corresponding coefficient is close to 1). Finally, we have the following relationship between \mathcal{R}_{ar} and \mathcal{R}_{ipc} :

THEOREM 9.

$$\mathcal{R}_{ipc} \subseteq \mathcal{R}_{ar}.$$

Proof. If $(r_1, r_2) \in \mathcal{R}_{ipc}$, there exists (Z, U) for which $coef_{aiac}(Z, U) = 0$ and

$$\frac{H(Z)}{H(Z) + H(U)} = r_1, \frac{H(Z, U)}{H(Z) + H(U)} = r_2.$$

Then we have $coef_{ar}(Z, U) \geq 1$. Since $coef_{ar}(Z, U) \leq 1$ always holds, we have

$$coef_{ar}(Z, U) = 1 \Rightarrow (r_1, r_2) \in \mathcal{R}_{ar}$$

□

The following is a new conditional information inequality.

THEOREM 10. $I(U; Z|X) + I(U; Z|Y) + I(X; Y) = 0$ implies that

$$H(Z, U) \geq \frac{3}{4}(H(Z) + H(U)).$$

Proof. $H(Z, U) < \frac{3}{4}(H(Z) + H(U))$ implies that $coef_{ar}(U, Z) < 1$ by Theorem 6. This contradicts the condition that $coef_{aic}(U, Z) = 0$. □

This result is not the best possible. There exists an absolute constant $C > 0.75$ for which the following result is true

THEOREM 11. $I(U; Z|X) + I(U; Z|Y) + I(X; Y) = 0$ implies

$$H(Z, U) \geq C(H(Z) + H(U)).$$

We believe that the best constant C is

$$C = \frac{3}{4h_2(0.25)} = 0.9248.$$

A proof has not been found.

6. Discussion. In this article, we provided a simple proof for a recently found non-Shannon-type information inequality. We formulated two optimization problems of information quantities, the AR problem and AIC problem. We defined two information quantities $coef_{ar}(Z, U)$ and $coef_{aic}(Z, U)$ for a pair of random variables (Z, U) and proved that $coef_{ar}(Z, U) + coef_{aic}(Z, U) \geq 1$. Two related regions \mathcal{R}_{ar} and \mathcal{R}_{ipc} are also defined and we proved that $\mathcal{R}_{ipc} \subseteq \mathcal{R}_{ar}$. We propose the following open problems:

1. Determine the regions \mathcal{R}_{ar} and \mathcal{R}_{ipc} (We believe that $\mathcal{R}_{ar} \neq \mathcal{R}_{ipc}$);
2. Use the new inequality to derive a better inner bound for Γ_4^* ;
3. Determine the constant C .
4. Using the conditional version of the new inequality, we can generalize all concepts introduced in the paper to their conditional versions. The key results in the paper can be easily generalized, too. A particularly interesting observation is that for a pair of random variables Z and U , there always exists a random variable W such that $coef_{aic}(Z, U|W) = 0$. For instance, $W = (Z, U)$ will work. Many questions can be raised based on this observation. We mention the following as an example. Determine

$$\inf\{H(W) : coef_{aic}(Z, U|W) = 0\}.$$

REFERENCES

- [1] Z. ZHANG AND R. W. YEUNG, *A non-Shannon type conditional inequality of information quantities*, IEEE Trans. Inform. Theory, 43(1997), pp. 1982–1985.
- [2] Z. ZHANG AND R. W. YEUNG, *On characterization of entropy function via information inequalities*, IEEE Trans. Inform. Theory, 44(1998), pp. 1440–1452.
- [3] K. MAKARYCHEV, Y. MAKARYCHEV, A. ROMASHCHEKO, AND N. VERESHCHAGIN, *New Class of Non-Shannon Type Information Inequalities for Entropies*, Preprint 2001.
- [4] R. W. YEUNG AND Z. ZHANG, *A class of non-Shannon-type information inequalities and their applications*, Communications in Information and Systems, 1(2001), pp. 87–100.
- [5] R. W. YEUNG AND Y.-O. YAN, *Information Theoretic Inequality Prover*, <http://www.ie.cuhk.edu.hk/~ITIP>.
- [6] P. GÁCS, J. KÖRNER, *Common Information is Far Less Than Mutual Information*, Probl. of Control and Inform. Theory, 2:2(1973), pp. 149–162.