COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 1, No. 3, pp. 241-254, September 2001

A COMBINATORIAL APPROACH TO INFORMATION INEQUALITIES*

TERENCE H. CHAN†

Abstract. In this paper, we establish a one-to-one correspondence between an inequality involving the entropies of discrete random variables, called an information inequality, and an inequality involving the cardinalities of the projections of what we call a quasi-uniform box assignment. We first show that a representative class of entropy functions can be characterized by quasi-uniform box assignments. Based on this result, we show a one-to-one correspondence between an information inequality and a combinatorial inequality. To demonstrate the importance of our results, we give a combinatorial proof for the nonnegativity of conditional mutual information. This shows that all Shannon-type-inequalities can be proved by methods in combinatorics. On the other hand, via a non-Shannon-type information inequality recently discovered by Zhang and Yeung, we obtain a new inequality in combinatorics whose meaning is yet to be understood.

Key Words. information inequalities, combinatorics, entropy functions.

1. Introduction.

The quest for information inequalities has been driven by the need to solve various communication problems. These inequalities play a crucial role in the proofs of almost all converse coding theorems in source and channel coding problems. In essence, they govern the impossibilities in information theory. Conventionally, we prove an information inequality by invoking certain elementary identities and inequalities in the intermediate steps of a proof. In this approach, inequalities corresponding to the nonnegativity of Shannon's information measures, called the *basic inequalities* [7], are invoked whenever we establish an inequality in an intermediate step. Proving an information inequality using this conventional approach can be quite tricky, because it may not be easy to see which identity or elementary inequality should be invoked at each step.

The framework developed in [7] renders a geometric interpretation of information inequalities. With this interpretation, all the information inequalities which are implied by the basic inequalities, called Shannon-type inequalities, have a unified description. These include all information inequalities which were known before the recent discovery of the so-called non-Shannon-type inequalities reported in [8] and [9]. Machine-proving of all Shannon-type inequalities is now possible [10].

It is evident that there are many non-Shannon-type inequalities yet to be discovered. Unfortunately, due to lack of tools, these inequalities are extremely difficult to discover and to prove. In [5][6] a combinatorial interpretation for a certain type of linear inequalities for Kolmogorov complexity (which are basically the same as linear information inequalities) is obtained. This new interpretation is important in finding

*Received on August 6, 2001, accepted for publication on September 21, 2001.

†Department of Information Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China. Email: hlchan@ie.cuhk.edu.hk

new information inequalities.

In this paper, we introduce a new concept called quasi-uniformity. By means of this concept, a combinatorial interpretation is found for all linear information inequalities. Hence, the problem of proving an information inequality can be translated to a combinatorial problem. It then opens the door to discovering and proving new information inequalities by means of tools in combinatorics.

In general, to prove/disprove an information inequality, we need to verify whether the inequality is true for all possible entropy functions. The main idea in this paper is that for a linear information inequality, it is not necessary to verify it for all possible entropy functions. It has been proved by Chan and Yeung in [1] that it is sufficient to check those entropy functions that are group-characterizable. Then they developed a group-theoretic approach to proving information inequalities. In this paper, a similar technique is used. Specificially, by showing that a linear information inequality can be proved/disproved by checking those entropy functions that are combinatorially characterizable, we develop a combinatorial approach to proving information inequalities.

2. A framework for information inequalities. Let $\mathcal{N} = \{1, \cdots, n\}$ and $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_n$ be n nonempty sets. Let Ω be the collection of all nonempty subsets of $\mathcal{N}$. For any $\alpha \in \Omega$, we define $\mathcal{X}_\alpha = \prod_{i \in \alpha} \mathcal{X}_i$ to be the Cartesian product of $\mathcal{X}_i$ for $i \in \alpha$. Unless otherwise specified, all small Greek letters ($\alpha, \beta, \gamma, \ldots, etc.$) are assumed to be elements in Ω and an element in $\mathcal{X}_\alpha$ will be denoted by $x_\alpha = (x_i : i \in \alpha)$. Sometimes, other small letters will also be used instead of x.

Let $X_1, X_2, \cdots, X_n$ be a set of n jointly distributed discrete random variables defined on $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_n$ respectively. For any $\alpha \in \Omega$, X_α denotes the jointly distributed random variable $(X_i : i \in \alpha)$. For example, $X_{(1,2)}$, or $X_{1,2}$ for simplicity, is the jointly distributed random variables of X_1 and X_2. The joint entropy of X_α is denoted by $H(X_\alpha)$.

Let $\mathcal{H}_n$ be the set of all real functions defined on Ω. In other words, $\mathcal{H}_n$ is the set of all real functions defined on the collection of nonempty subsets of $\mathcal{N}$ and hence, is a $(2^n - 1)$-dimensional Euclidean space. For simplicity, for any function $\mathbf{g} \in \mathcal{H}_n$, the function value $\mathbf{g}(\alpha)$ is denoted by g_α for all $\alpha \in \Omega$.

DEFINITION 2.1. *Let $\mathbf{g} \in \mathcal{H}_n$. Then $\mathbf{g}$ is called entropic if there exists a set of random variables $X_1, X_2, \cdots, X_n$ such that $g_\alpha = H(X_\alpha)$ for all $\alpha \in \Omega$.*

EXAMPLE 2.1. *Let X_1, X_2 be the outcomes of two fair coins respectively where head is denoted by 0 and tail is denoted by 1. Let $X_3 = X_1 + X_2$ mod 2. Then it can be checked easily that*

$$H(X_1) = H(X_2) = H(X_3) = 1 \tag{2.1}$$

$$H(X_{1,2}) = H(X_{1,3}) = H(X_{2,3}) = H(X_{1,2,3}) = 2. \tag{2.2}$$

Hence, if we define $g_\alpha = H(X_\alpha)$ for all nonempty subset α of $\{1, 2, 3\}$, then $\mathbf{g}$ is entropic.

Let Γ_n^* [7] be the set of all entropy functions. Γ_n^* plays an important role in information theory, especially in proving information inequalities. Γ_n^* is a subset of $\mathcal{H}_n$ and it has a very complex structure. For $n \geq 3$, Γ_n^* is not even closed [8]. It was also proved in [8] that $\overline{\Gamma}_n^*$, the closure of Γ_n^*, is a closed convex cone. Thus, $\overline{\Gamma}_n^*$ is much more manageable than Γ_n^*, and for certain applications (see Section 5), it is sufficient to consider $\overline{\Gamma}_n^*$. Hence, characterizations of Γ_n^* and $\overline{\Gamma}_n^*$ are of fundamental interest.

Notice that every linear information inequality

$$\sum_{\alpha\in\Omega} b_\alpha H(X_\alpha) \geq 0 \tag{2.3}$$

corresponds to a linear inequality $\mathbf{b}^\top\mathbf{h} \geq 0$ in $\mathcal{H}_n$, where $\mathbf{b}$ is a column vector whose components are indexed by $\alpha \in \Omega$. Hence, for simplicity, an information inequality will usually be written in the form $\mathbf{b}^\top\mathbf{h} \geq 0$.

Example 2.2. *Let $\mathcal{N} = \{1,2\}$ and $\mathbf{b} \in \mathcal{H}_2$ where $b_1 = b_2 = 1$ and $b_{1,2} = -1$. Then the information inequality $H(X_1) + H(X_2) - H(X_{1,2}) \geq 0$ corresponds to the inequality $\mathbf{b}^\top\mathbf{h} \geq 0$ in $\mathcal{H}_2$.*

Theorem 2.1. [7] *An information inequality $\sum_{\alpha\in\Omega} b_\alpha H(X_\alpha) \geq 0$ is valid if and only if for all $\mathbf{h} \in \Gamma_n^*$, $\mathbf{b}^\top\mathbf{h} \geq 0$. Since $\{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^\top\mathbf{h} \geq 0\}$ is closed and convex, an information inequality $\sum_{\alpha\in\Omega} b_\alpha H(X_\alpha) \geq 0$ is valid if and only if for all $\mathbf{h} \in \overline{\Gamma}_n^*$, $\mathbf{b}^\top\mathbf{h} \geq 0$.*

Theorem 2.1 has an important implication. It says that the validity of an information inequality $E \geq 0$ depends only on Γ_n^* (or $\overline{\Gamma}_n^*$) and $\{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^\top\mathbf{h} \geq 0\}$. Thus, if Γ_n^* is characterized explicitly, then the information inequality can be proved or disproved by comparing the two regions. Hence, the study of Γ_n^* and its underlying structure is fundamental in information theory. Although it is proved that $\overline{\Gamma}_n^*$ is a closed covex cone, $\overline{\Gamma}_n^*$ is not fully characterized for $n > 3$ yet [9] [11].

3. Box Assignment.

Definition 3.1. *Let $\mathcal{X}_1, \cdots, \mathcal{X}_n$ be n non-empty finite sets. A* **box assignment** *$\mathcal{A}$ of $\mathcal{X}_\mathcal{N}$ is a non-empty subset of $\mathcal{X}_\mathcal{N}$.*

Since $\mathcal{X}_\mathcal{N}$ is the Cartesian product of $\mathcal{X}_1, \cdots, \mathcal{X}_n$, it can be regarded as an "n-dimensional box". Then a box assignment $\mathcal{A}$ can be visualized as follows. Each element $x_\mathcal{N}$ in $\mathcal{X}_\mathcal{N}$ corresponds to a cell in the box, also denoted by $\mathcal{X}_\mathcal{N}$. If a cell $x_\mathcal{N}$ is in $\mathcal{A}$, then it recieves one ball, otherwise it recieves no ball. For simplicity, for any element $x_\mathcal{N} = (x_1, \ldots, x_n)$ in $\mathcal{X}_\mathcal{N}$ and $\alpha \in \Omega$, x_α will denote the subset of coordinates $(x_i,\ i \in \alpha)$.

Example 3.1. *Let $\mathcal{X}_1 = \{0,1,2\}$, $\mathcal{X}_2 = \{0,1,2\}$ and $\mathcal{A} = \{(1,0),(1,2),(2,1)\}$. Then $\mathcal{A}$ is a box assignment of $\mathcal{X}_1 \times \mathcal{X}_2$. See figure 3.1.*

Definition 3.2. *Let $\mathcal{A}$ be a box assignment. The α-projection of $\mathcal{A}$ is defined as follows:*

$$\mathcal{A}_\alpha = \{a_\alpha \in \mathcal{X}_\alpha : \exists\, x_\mathcal{N} \in \mathcal{A} \text{ such that } \forall i \in \alpha,\ x_i = a_i\}. \tag{3.1}$$

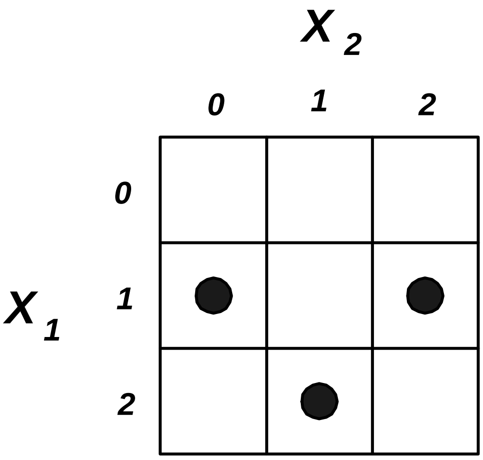

FIG. 3.1. *A Box Assignment $\mathcal{A}$ of $\mathcal{X}_1 \times \mathcal{X}_2$.*

Roughly speaking, $\mathcal{A}_\alpha$ is the "projection of the balls onto the plane $\mathcal{X}_\alpha$". In other words, $a_\alpha \in \mathcal{A}_\alpha$ if and only if there exists a ball in $\mathcal{A}$ whose i^{th} coordinate is equal to a_i for all $i \in \alpha$.

DEFINITION 3.3. *Let $\mathcal{A}$ be a box assignment and $a_\alpha \in \mathcal{X}_\alpha$. Then the β-projection of the a_α section of $\mathcal{A}$ is defined as*

$$\begin{aligned}\mathcal{A}_{\beta|\alpha}(a_\alpha) = \{b_\beta \in \mathcal{X}_\beta : \exists\, x_\mathcal{N} \in \mathcal{A} \textit{ such that } \forall i \in \alpha,\\ x_i = a_i \textit{ and } \forall j \in \beta,\ x_j = b_j\}.\end{aligned} \tag{3.2}$$

Roughly speaking, $\mathcal{A}_{\beta|\alpha}(a_\alpha)$ is the "β-projection" of the balls in the a_α section of $\mathcal{A}$. In other words, $b_\beta \in \mathcal{A}_{\beta|\alpha}(a_\alpha)$ if and only if there exists a ball in the a_α section of $\mathcal{A}$ (i.e. its i^{th} coordinate equal to a_i for all $i \in \alpha$) such that its projection on plane $\mathcal{X}_\beta$ is b_β (i.e. its j^{th} coordinate equal to b_j for all $j \in \beta$). It is easy to prove that $\mathcal{A}_{\beta|\alpha}(a_\alpha)$ is non-empty if and only if $a_\alpha \in \mathcal{A}_\alpha$.

EXAMPLE 3.2. *Let $\mathcal{A}$ be a box assignment defined in Example 3.1. Then*

$$\mathcal{A}_1 = \{1,2\},\ \mathcal{A}_2 = \{0,1,2\} \tag{3.3}$$

$$\mathcal{A}_{2|1}(0) = \{\},\ \mathcal{A}_{2|1}(1) = \{0,2\},\ \mathcal{A}_{2|1}(2) = \{1\} \tag{3.4}$$

$$\mathcal{A}_{1|2}(0) = \{1\},\ \mathcal{A}_{1|2}(1) = \{2\},\ \mathcal{A}_{1|2}(2) = \{1\}. \tag{3.5}$$

3.1. Quasi-uniform Box Assignment.

DEFINITION 3.4 (Quasi-uniform box assignment). *Let $\mathcal{A}$ be a box assignment of $\mathcal{X}_\mathcal{N}$. It is quasi-uniform if for any fixed $\alpha \in \Omega$, the cardinality of $\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)$ is constant for all $a_\alpha \in \mathcal{A}_\alpha$. In other words, the number of balls on the a_α section of $\mathcal{A}$ is the same for all $a_\alpha \in \mathcal{A}_\alpha$. The constant is denoted by $|\mathcal{A}_{\mathcal{N}|\alpha}|$ for simplicity.*

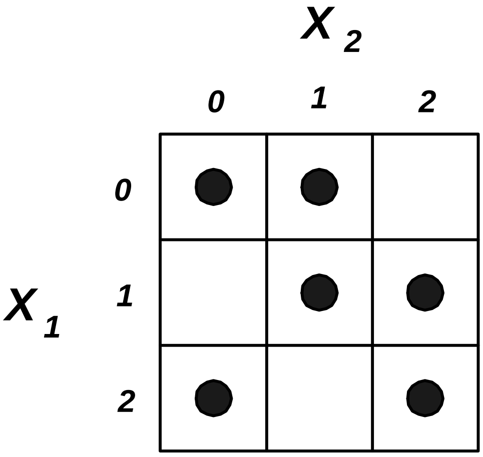

FIG. 3.2. *Quasi-uniform Box Assignment $\mathcal{A}$ of $\mathcal{X}_1 \times \mathcal{X}_2$.*

EXAMPLE 3.3. *Let $\mathcal{X}_1 = \{0,1,2\}$, $\mathcal{X}_2 = \{0,1,2\}$ and $\mathcal{A} = \{(0,0),(0,1),(1,1),(1,2),(2,0),(2,2)\}$. See figure 3.2. Here,*

$$\mathcal{A}_1 = \{0,1,2\},\ \mathcal{A}_2 = \{0,1,2\} \tag{3.6}$$

$$\mathcal{A}_{1,2|1}(0) = \{(0,0),(0,1)\},\ \mathcal{A}_{1,2|1}(1) = \{(1,1),(1,2)\},\ \mathcal{A}_{1,2|1}(2) = \{(2,0),(2,2)\} \tag{3.7}$$

$$\mathcal{A}_{1,2|2}(0) = \{(0,0),(2,0)\},\ \mathcal{A}_{1,2|2}(1) = \{(0,1),(1,1)\},\ \mathcal{A}_{1,2|2}(2) = \{(1,2),(2,2)\}. \tag{3.8}$$

Also, it is trivial that $\mathcal{A}_{1,2|1,2}(x_1,x_2) = \{(x_1,x_2)\}$ if $(x_1,x_2) \in \mathcal{A}$. Hence, for all $\alpha \in \Omega$, $|\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)|$ is constant for all $a_\alpha \in \mathcal{A}_\alpha$. Therefore, $\mathcal{A}$ is a quasi-uniform box assignment of $\mathcal{X}_1 \times \mathcal{X}_2$.

EXAMPLE 3.4. *Let $\mathcal{X}_1 = \{0,1\}$, $\mathcal{X}_2 = \{0,1\}$ and $\mathcal{X}_3 = \{0,1\}$. Let $\mathcal{A} = \{(0,0,0),(1,1,0),\ (0,1,1),(1,0,1)\}$. See figure 3.3. Then $\mathcal{A}$ is a quasi-uniform box assignment of $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$.*

PROPOSITION 3.1. *Let $\mathcal{A}$ be a quasi-uniform box assignment of $\mathcal{X}_\mathcal{N}$ and $\alpha \in \Omega$. Then*

$$|\mathcal{A}_{\mathcal{N}|\alpha}| = \frac{|\mathcal{A}|}{|\mathcal{A}_\alpha|}. \tag{3.9}$$

Proof. It can be checked easily that $\mathcal{A}$ is equal to the disjoint union of $\mathcal{A}_{\mathcal{N}|\alpha}(x_\alpha)$ for $x_\alpha \in \mathcal{A}_\alpha$. Hence,

$$|\mathcal{A}| = \sum_{x_\alpha \in \mathcal{A}_\alpha} |\mathcal{A}_{\mathcal{N}|\alpha}(x_\alpha)| \tag{3.10}$$

$$= \sum_{x_\alpha \in \mathcal{A}_\alpha} |\mathcal{A}_{\mathcal{N}|\alpha}| \tag{3.11}$$

$$= |\mathcal{A}_\alpha||\mathcal{A}_{\mathcal{N}|\alpha}|. \tag{3.12}$$

The result then follows. □

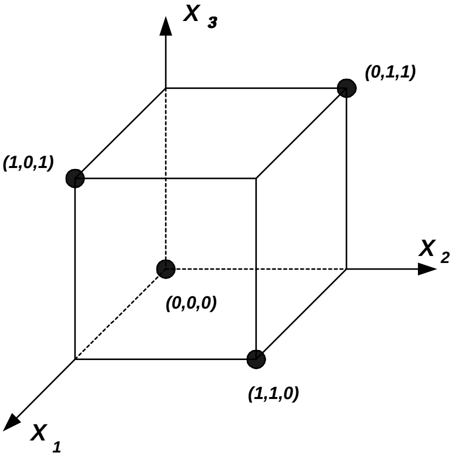

FIG. 3.3. *Quasi-uniform Box Assignment $\mathcal{A}$ of $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$.*

PROPOSITION 3.2. *Let $\mathcal{A}$ be a quasi-uniform box assignment of $\mathcal{X}_\mathcal{N}$ and $\alpha \in \Omega$. For any $a_\alpha \in \mathcal{A}_\alpha$, $|\mathcal{A}_{\beta|\alpha}(a_\alpha)| = \frac{|\mathcal{A}_{\beta\cup\alpha}|}{|\mathcal{A}_\alpha|}$.*
Proof. Let $\gamma = \alpha \cup \beta$.

$$\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha) = \{x_\mathcal{N} \in \mathcal{A}_\mathcal{N} : \forall i \in \alpha,\ x_i = a_i\} \tag{3.13}$$

$$= \bigcup_{y_\gamma \in \mathcal{A}_\gamma} \{x_\mathcal{N} \in \mathcal{A}_\mathcal{N} : \forall i \in \alpha,\ x_i = a_i \text{ and } \forall j \in \gamma,\ x_j = y_j\} \tag{3.14}$$

$$= \bigcup_{y_\gamma \in \mathcal{A}_{\gamma|\alpha}(a_\alpha)} \{x_\mathcal{N} \in \mathcal{A}_\mathcal{N} : \forall j \in \gamma,\ x_j = y_j\} \tag{3.15}$$

$$= \bigcup_{y_\gamma \in \mathcal{A}_{\gamma|\alpha}(a_\alpha)} \mathcal{A}_{\mathcal{N}|\gamma}(y_\gamma). \tag{3.16}$$

Therefore,

$$\frac{|\mathcal{A}|}{|\mathcal{A}_\alpha|} = |\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)| \tag{3.17}$$

$$= \sum_{y_\gamma \in \mathcal{A}_{\gamma|\alpha}(a_\alpha)} \frac{|\mathcal{A}|}{|\mathcal{A}_\gamma|} \tag{3.18}$$

$$= \frac{|\mathcal{A}|}{|\mathcal{A}_\gamma|} |\mathcal{A}_{\gamma|\alpha}(a_\alpha)|. \tag{3.19}$$

Hence, $|\mathcal{A}_{\gamma|\alpha}(a_\alpha)| = \frac{|\mathcal{A}_\gamma|}{|\mathcal{A}_\alpha|}$. Since $\gamma = \alpha \cup \beta$, it can be checked easily that $|\mathcal{A}_{\beta|\alpha}(a_\alpha)| = |\mathcal{A}_{\gamma|\alpha}(a_\alpha)|$. Hence $|\mathcal{A}_{\beta|\alpha}(a_\alpha)|$ is equal to $\frac{|\mathcal{A}_{\alpha\cup\beta}|}{|\mathcal{A}_\alpha|}$ for all $a_\alpha \in \mathcal{A}_\alpha$ and the constant is denoted by $|\mathcal{A}_{\beta|\alpha}|$ for simplicity. □

3.2. Constructing quasi-uniform box assignment using subgroups. Group is one of the simplest and most basic algebraic structures. Some familiar examples are: the integers under addition, the rationals excluding zero under multiplication, and the set of real-valued 2 × 2 matrices under addition, where addition and multiplication refer to the usual addition and multiplication for real numbers and matrices. The above are all examples of infinite groups. In this paper, however, we are concerned with finite groups. One example is the group of integers modulo m.

Let G be a group. A subset S of G is a subgroup of G if S is also a group and each subgroup S of G partitions G into left cosets of S. In this section, we will construct quasi-uniform box assigments using subgroups of a finite group. We first state without proof the following basic facts in group theory. The proof of these facts is straightforward (see for example [2][11]).

THEOREM 3.1. *Let $G_1, \ldots, G_n$ be n subgroups of a finite group G. For any $\alpha \in \Omega$, let $G_\alpha = \bigcap_{i\in\alpha} G_i$. Then*

1. *G_α is a subgroup of G.*
2. **(Lagrange's Theorem)** *There are $\frac{|G|}{|G_\alpha|}$ distinct left cosets of G_α in G. Thus, there are $\frac{|G|}{|G_\mathcal{N}|}$ distinct left cosets of $G_\mathcal{N}$ in G.*
3. *Let $K_1, \ldots, K_n$ be left cosets of $G_1, \ldots, G_n$ respectively. Then $\bigcap_{i\in\alpha} K_i$ is either a left coset of G_α or is empty.*
4. *Let K be a left coset of $G_\mathcal{N}$. Then there exists unique left cosets K_1 of G_1, $\ldots$, K_n of G_n such that $K = \bigcap_{i\in\mathcal{N}} K_i$.*
5. *Let $K_1, \ldots, K_n$ be left cosets of $G_1, \ldots, G_n$ respectively. If $\bigcap_{i\in\alpha} K_i$ is non-empty, then there are $\frac{|G_\alpha|}{|G_\mathcal{N}|}$ left cosets of $G_\mathcal{N}$ in $\bigcap_{i\in\alpha} K_i$.*

THEOREM 3.2. *Let G be a finite group, $G_1, \ldots, G_n$ be n subgroups of G and $\mathcal{X}_i$ be the index of the set of left cosets of G_i (i.e. the left cosets of G_i is denoted by K_{i,x_i} for $x_i \in \mathcal{X}_i$). Define*

$$\mathcal{A} = \left\{ x_\mathcal{N} \in \mathcal{X}_\mathcal{N} : \bigcap_{i=1}^{n} K_{i,x_i} \text{ is a left coset of } G_\mathcal{N} \right\}. \tag{3.20}$$

Then $\mathcal{A}$ is a quasi-uniform box assignment of $\mathcal{X}_\mathcal{N}$. In addition, for any $\alpha \in \Omega$, $|\mathcal{A}_\alpha| = \frac{|G|}{|G_\alpha|}$.

Proof. Let $a_\alpha \in \mathcal{A}_\alpha$. It can be checked easily that $x_\mathcal{N} \in \mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)$ if and only if $\bigcap_{i=1}^n K_{i,x_i}$ is a left coset of $G_\mathcal{N}$ in $\bigcap_{i\in\alpha} K_{i,a_i}$. By Theorem 3.1, there are $\frac{|G_\alpha|}{|G_\mathcal{N}|}$'s left cosets of $G_\mathcal{N}$ in $\bigcap_{i\in\alpha} K_{i,a_i}$ and hence, $\frac{|G_\alpha|}{|G_\mathcal{N}|}$'s balls on the a_α section. Thus, $|\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)|$ is constant for all $a_\alpha \in \mathcal{A}_\alpha$, i.e., $\mathcal{A}$ is a quasi-uniform box assignment. In addition, since there are $\frac{|G|}{|G_\mathcal{N}|}$ distinct left cosets of $G_\mathcal{N}$ in G, $|\mathcal{A}| = \frac{|G|}{|G_\mathcal{N}|}$. By Proposition 3.2,

$$|\mathcal{A}_\alpha| = \frac{|\mathcal{A}|}{|\mathcal{A}_{\mathcal{N}|\alpha}|} \tag{3.21}$$

$$= \frac{\frac{|G|}{|G_\mathcal{N}|}}{\frac{|G_\alpha|}{|G_\mathcal{N}|}} \tag{3.22}$$

$$= \frac{|G|}{|G_\alpha|}. \tag{3.23}$$

The result then follows. □

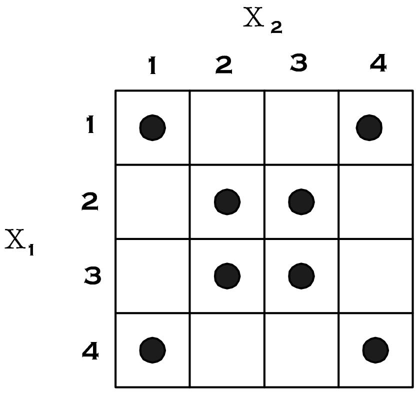

FIG. 3.4. *Quasi-uniform Box Assignment $\mathcal{A}$ of $\mathcal{X}_1 \times \mathcal{X}_2$.*

EXAMPLE 3.5. *Let $G = \{(i,j,k) : i,j,k = 0$ or $1\}$, where the group operation is the componentwise modulo 2 addition. Let $G_1 = \{(0,0,0),\ (1,0,0)\}$ and $G_2 = \{(0,0,0),\ (1,1,1)\}$. There are four left cosets of G_1 in G:*

$$\begin{array}{ll} K_{1,1} = \{(0,0,0),(1,0,0)\} & K_{1,2} = \{(0,0,1),(1,0,1)\} \\ K_{1,3} = \{(0,1,0),(1,1,0)\} & K_{1,4} = \{(0,1,1),(1,1,1)\}. \end{array} \tag{3.24}$$

Similarly, there are four left cosets of G_2 in G:

$$\begin{array}{ll} K_{2,1} = \{(0,0,0),(1,1,1)\} & K_{2,2} = \{(0,0,1),(1,1,0)\} \\ K_{2,3} = \{(0,1,0),(1,0,1)\} & K_{2,4} = \{(0,1,1),(1,0,0)\}. \end{array} \tag{3.25}$$

Then a quasi-uniform box can be constructed by putting a ball in the cell (i,j) if $K_{1,i} \bigcap K_{2,j}$ is non-empty, as depicted in Figure 3.4.

4. Combinatorial characterizations. In this section, we will construct entropy functions based on quasi-uniform box assignments. We then show that in order to prove/disprove an information inequality, it is sufficient to check those entropy functions that can be constructed from quasi-uniform box assignments.

THEOREM 4.1. *Let $\mathcal{A}$ be a quasi-uniform box assignment of $\mathcal{X}_{\mathcal{N}}$. Then $\mathbf{h} \in \mathcal{H}_n$ defined by*

$$h_\alpha = \log|\mathcal{A}_\alpha| \tag{4.1}$$

for all $\alpha \in \Omega$ is entropic, i.e., $\mathbf{h} \in \Gamma_n^*$.

Proof. It suffices to show that there exists a collection of random variables $X_1, \ldots, X_n$ such that for all $\alpha \in \Omega$, the entropy $H(X_\alpha)$ is equal to h_α. Let the joint probability mass function of $X_1, \ldots, X_n$ be

$$P(x_1, \ldots, x_n) = \begin{cases} \frac{1}{|\mathcal{A}|} & \text{if } (x_1, \cdots, x_n) \in \mathcal{A} \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

It can be checked easily that $P(x_i : i \in \alpha) = \frac{1}{|\mathcal{A}|}|\mathcal{A}_{\mathcal{N}|\alpha}(x_\alpha)|$. Therefore,

$$P(x_i : i \in \alpha) = \begin{cases} \frac{1}{|\mathcal{A}|}|\mathcal{A}_{\mathcal{N}|\alpha}| & \text{if } x_\alpha \in \mathcal{A}_\alpha \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Hence, the random variable X_α is uniformly distributed over $\mathcal{A}_\alpha$ and $H(X_\alpha) = \log|\mathcal{A}_\alpha|$. □

DEFINITION 4.1. *Let $\mathcal{A}$ be a quasi-uniform box assignment of $\mathcal{X}_\mathcal{N}$ and $\mathbf{h} \in \mathcal{H}_n$ such that $h_\alpha = \log|\mathcal{A}_\alpha|$ for all $\alpha \in \Omega$. Then $\mathcal{A}$ is a combinatorial characterization of* $\mathbf{h}$.

Theorem 4.1 asserts that certain entropy functions in Γ_n^* have a combinatorial characterization. These are called combinatorially characterizable entropy functions, which will be used in the next section to obtain a combinatorial characterization of the region $\overline{\Gamma}_n^*$.

EXAMPLE 4.1. *Let $G_1, \ldots, G_n$ be n subgroups of a finite group G and $\mathbf{h} \in \mathcal{H}_n$ be defined by*

$$h_\alpha = \log\frac{|G|}{|G_\alpha|} \tag{4.4}$$

for all $\alpha \in \Omega$. In section 3.2, we have constructed a quasi-uniform box assignment $\mathcal{A}$ such that $|\mathcal{A}_\alpha| = \frac{|G|}{|G_\alpha|}$. Therefore, $\mathbf{h}$ is combinatorially characterizable.

We have introduced the class of entropy functions in Γ_n^* which have a combinatorial characterization. However, an entropy function $\mathbf{h} \in \Gamma_n^*$ may not have a combinatorial characterization due to the following observation. Suppose $\mathbf{h} \in \Gamma_n^*$. Then there exists a collection of random variables $X_1, X_2, \cdots, X_n$ such that $h_\alpha = H(X_\alpha)$ for all $\alpha \in \Omega$. If $\mathcal{A}$ is a combinatorial characterization of $\mathbf{h}$, then $H(X_\alpha) = \log|\mathcal{A}_\alpha|$ for all $\alpha \in \Omega$. Since $|\mathcal{A}_\alpha|$ is an integer, $H(X_\alpha)$ must be the logarithm of an integer. However, the joint entropy of a set of random variables in general is not necessarily the logarithm of an integer. Therefore, it is possible to construct an entropy function $\mathbf{h} \in \Gamma_n^*$ which has no combinatorial characterization.

Although $\mathbf{h} \in \Gamma_n^*$ does not imply $\mathbf{h}$ has a combinatorial characterization, it turns out that the set of all $\mathbf{h} \in \Gamma_n^*$ which have a combinatorial characterization is almost good enough to characterize the region Γ_n^*, as we will see next.

DEFINITION 4.2. *Define the following region in $\mathcal{H}_n$:*

$$\Lambda_n = \{\mathbf{h} \in \mathcal{H}_n : \mathbf{h} \text{ has a combinatorial characterization } \}. \tag{4.5}$$

By Theorem 4.1, if $\mathbf{h} \in \mathcal{H}_n$ has a combinatorial characterization, then $\mathbf{h} \in \Gamma_n^*$. Therefore, $\Lambda_n \subseteq \Gamma_n^*$. We will prove in the next theorem that $\overline{con}(\Lambda_n)$, the convex closure of Λ_n, is in fact equal to $\overline{\Gamma}_n^*$, the closure of Γ_n^*.

THEOREM 4.2. $\overline{con}(\Lambda_n) = \overline{\Gamma}_n^*$.

Proof. It is trivial to prove that $\overline{con}(\Lambda_n) \subseteq \overline{\Gamma}_n^*$. Recall that in Example 4.1, we have constructed combinatorially characterizable functions using subgroups of finite groups. Let Υ_n be the set of all combinatorially characterizable functions that can be constructed as in Example 4.1 using subgroups of finite groups. It has been proved in [1] that $\overline{con}(\Upsilon_n) = \overline{\Gamma}^*_n$. Therefore, $\overline{\Gamma}^*_n \subseteq \overline{con}(\Lambda_n)$ and the result follows. □

5. Information inequalities. As we have stated in Section 2 that an information inequality[1]

$$\mathbf{b}^\top \mathbf{h} \geq 0 \tag{5.1}$$

always holds if and only if

$$\overline{\Gamma}^*_n \subseteq \{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^\top \mathbf{h} \geq 0\}. \tag{5.2}$$

In other words, all unconditional information inequalities are fully characterized by $\overline{\Gamma}^*_n$. We also have proved at the end of the last section that $\overline{con}(\Lambda_n) = \overline{\Gamma}^*_n$. Since $\Lambda_n \subseteq \Gamma^*_n \subseteq \overline{\Gamma}^*_n$, if (5.2) holds, then

$$\Lambda_n \subseteq \{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^\top \mathbf{h} \geq 0\}. \tag{5.3}$$

On the other hand, since $\{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^\top \mathbf{h} \geq 0\}$ is closed and convex, by taking convex closure in (5.3), we obtain

$$\overline{\Gamma}^*_n = \overline{con}(\Lambda_n) \subseteq \{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}^\top \mathbf{h} \geq 0\}. \tag{5.4}$$

Therefore, (5.2) and (5.3) are equivalent.

For each $\mathbf{h} \in \Lambda_n$, $h_\alpha = \log |\mathcal{A}_\alpha|$ for all $\alpha \in \Omega$ for some quasi-uniform box assignment $\mathcal{A}$ of $\mathcal{X}_\mathcal{N}$. Hence, the information inequality $\sum_{\alpha \in \mathcal{H}_n} b_\alpha H(X_\alpha) \geq 0$ holds for all random variables $X_1, X_2, \cdots, X_n$ if and only if the corresponding combinatorial inequality $\sum_{\alpha \in \mathcal{H}_n} b_\alpha \log |\mathcal{A}_\alpha| \geq 0$ holds for all quasi-uniform box assigments of $\mathcal{X}_\mathcal{N}$. In other words, for every unconditional information inequality, there is a corresponding combinatorial inequality, and vice versa. Therefore, inequalities in information theory can be proved by methods in combinatorics and vice versa.

In the rest of the section, we explore this one-to-one correspondence between information theory and combinatorics. We first give a combinatorial proof of the basic inequalities in information theory. At the end of the section, we will give an information-theoretic proof for a new combinatorial inequality.

THEOREM 5.1. *Let $\mathcal{A}$ be any quasi-uniform box assignment of $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$. Then*

$$|\mathcal{A}_{1,2,3}||\mathcal{A}_3| \leq |\mathcal{A}_{1,3}||\mathcal{A}_{2,3}|. \tag{5.5}$$

Proof.

[1] the results still hold for general information inequality $\mathbf{b}(\mathbf{h}) \geq 0$ as long as $\{\mathbf{h} \in \mathcal{H}_n : \mathbf{b}(\mathbf{h}) \geq 0\}$ is a closed and convex cone.

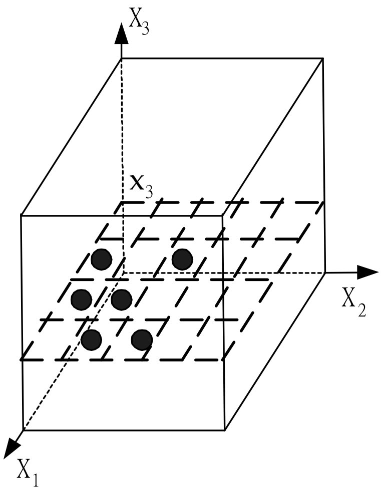

FIG. 5.1. *A 2-dimensional box assignment induced by $\mathcal{A}$*

Let $\mathcal{A}$ be any quasi-uniform box assignment of $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$. Fix $x_3 \in \mathcal{A}_3$. Then let $\mathcal{B} = \mathcal{A}_{1,2|3}(x_3)$. It can be checked easily that $\mathcal{B}_1 = \mathcal{A}_{1|3}(x_3)$ and $\mathcal{B}_2 = \mathcal{A}_{2|3}(x_3)$. See Figure 5.1. Also, it is trivial to prove that $\mathcal{B}$ is a subset of $\mathcal{B}_1 \times \mathcal{B}_2$. Hence,

$$|\mathcal{B}| \leq |\mathcal{B}_1||\mathcal{B}_2|. \tag{5.6}$$

By Proposition 3.2,

$$|\mathcal{B}| = \frac{|\mathcal{A}_{1,2,3}|}{|\mathcal{A}_3|}; \; |\mathcal{B}_1| = \frac{|\mathcal{A}_{1,3}|}{|\mathcal{A}_3|}; \; |\mathcal{B}_2| = \frac{|\mathcal{A}_{2,3}|}{|\mathcal{A}_3|}. \tag{5.7}$$

Hence,

$$\frac{|\mathcal{A}_{1,2,3}|}{|\mathcal{A}_3|} \leq \frac{|\mathcal{A}_{1,3}|}{|\mathcal{A}_3|}\frac{|\mathcal{A}_{2,3}|}{|\mathcal{A}_3|} \tag{5.8}$$

which is equivalent to $|\mathcal{A}_{1,2,3}||\mathcal{A}_3| \leq |\mathcal{A}_{1,3}||\mathcal{A}_{2,3}|$. □

COROLLARY 5.1. *For random variables X_1, X_2 and X_3,*

$$I(X_1; X_2|X_3) \geq 0. \tag{5.9}$$

Proof. Let $\mathcal{A}$ be a quasi-uniform box assignment of $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$. Then by Theorem 5.1,

$$|\mathcal{A}_{1,2,3}||\mathcal{A}_3| \leq |\mathcal{A}_{1,3}||\mathcal{A}_{2,3}| \tag{5.10}$$

Hence,

$$\log|\mathcal{A}_{1,2,3}| + \log|\mathcal{A}_3| \leq \log|\mathcal{A}_{1,3}| + \log|\mathcal{A}_{2,3}|. \tag{5.11}$$

This combinatorial inequality corresponds to the information inequality

$$H(X_1, X_2, X_3) + H(X_3) \leq H(X_1, X_3) + H(X_2, X_3) \tag{5.12}$$

which is equivalent to

$$I(X_1;X_2|X_3) \geq 0. \tag{5.13}$$

□

EXAMPLE 5.1. *Recently, the following highly non-trivial information inequality, which cannot be deduced by invoking the basic Shannon inequalities directly, has been proved in [9].*

$$\begin{aligned} &H(X_1)+H(X_2)+2H(X_1,X_2)+4H(X_3)+4H(X_4)\\ &+5H(X_1,X_3,X_4)+5H(X_2,X_3,X_4)\\ &\leq 6H(X_3,X_4)+4H(X_1,X_3)+4H(X_1,X_4)\\ &\quad +4H(X_2,X_3)+4H(X_2,X_4), \end{aligned} \tag{5.14}$$

Such an information inequality is referred to as a non-Shannon-type information inequality. This information inequality corresponds to the following combinatorial inequality

$$\begin{aligned} &\log|\mathcal{A}_1|+\log|\mathcal{A}_2|+2\log|\mathcal{A}_{1,2}|+4\log|\mathcal{A}_3|+4\log|\mathcal{A}_4|\\ &+5\log|\mathcal{A}_{1,3,4}|+5\log|\mathcal{A}_{2,3,4}|\\ &\leq 6\log|\mathcal{A}_{3,4}|+4\log|\mathcal{A}_{1,3}|+4\log|\mathcal{A}_{1,4}|\\ &\quad +4\log|\mathcal{A}_{2,3}|+4\log|\mathcal{A}_{2,4}| \end{aligned} \tag{5.15}$$

Taking exponentiation on both sides, we obtain

$$\begin{aligned} &|\mathcal{A}_1||\mathcal{A}_2||\mathcal{A}_{1,2}|^2|\mathcal{A}_3|^4|\mathcal{A}_4|^4|\mathcal{A}_{1,3,4}|^5|\mathcal{A}_{2,3,4}|^5\\ &\leq |\mathcal{A}_{3,4}|^6|\mathcal{A}_{1,3}|^4|\mathcal{A}_{1,4}|^4|\mathcal{A}_{2,3}|^4|\mathcal{A}_{2,4}|^4. \end{aligned} \tag{5.16}$$

The meaning of this combinatorial inequality is yet to be understood.

6. Conclusion. In this paper, we have identified a class of entropy functions which have combinatorial characterizations. These functions are called combinatorially characterizable entropy functions. The discovery in this paper is particularly important for studying Γ_n^*, the set of all entropy functions. It has been shown that Γ_n^* plays a crucial role in information theory. However, it is extremely difficult to characterize this set. One possible way to characterize Γ_n^* is to find new information inequalities which give tighter outer bounds on Γ_n^*. The most important result along this line was reported in [9], in which a new non-Shannon-type information inequality was obtained. However, due to lack of tools, to find new information inequalities is an extremely difficult task. The result in this paper turns the problem of characterizing entropy functions into the problem of characterizing the sizes of the projections of a quasi-uniform box assignment. Hence, it may be possible that we can use some existing results in combinatorics to attack the corresponding problem in information theory.

Acknowledgment

The author is grateful to Professor Raymond Yeung. His detailed comments and advice have greatly improved the readability of the paper.

REFERENCES

[1] T.H. Chan and R.W. Yeung, *On a relation between information inequalities and group theory*, submitted to IEEE Transactions of Information Theory.

[2] T.H. Chan, *Aspects of information inequalities and its applications*, M. Phil thesis, 1998.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.

[4] N. Pippenger, *What are the laws of information theory?* 1986 Specific Problems on Communication and Computation Conference, Palo Alto, California, Sept. 3-5, 1986

[5] D. Hammer and A. Shen, *A strange application of Kolmogorov complexity*, Theory of Computing Systems, 31:1(1998), pp. 1–4.

[6] A. Romashchenko, A. Shen, and N. K. Vereshchagin, *Combinatorial interpretation of Kolmogorov complexity*, Electronic Colloquium on Computational Complexity, vol.7, 2000.

[7] R.W. Yeung, *A framework for linear information inequality*, IEEE Transactions of Information Theory, 43(1997), pp. 1924–1934.

[8] Z. Zhang and R. W. Yeung, *A non-Shannon-type conditional information inequality*, IEEE Transactions of Information Theory, 43(1997), pp. 1982–1986.

[9] Z. Zhang and R. W. Yeung, *On the characterization of entropy function via information inequalities*, IEEE Transactions of Information Theory, 44(1998), pp. 1440–1452.

[10] R. W. Yeung and Y.-O. Yan, ITIP, `http://www.ie.cuhk.edu.hk/~ITIP`.

[11] R. W. Yeung, *A First Course in Information Theory*, to be published by Kluwer Academic Publishers.