

Article

## Recent Progresses in Characterising Information Inequalities

Terence Chan

Institute for Telecommunications Research, University of South Australia, Australia;

E-Mail: [terence.chan@unisa.edu.au](mailto:terence.chan@unisa.edu.au); Tel.: +61-8-8302-3875.

Received: 01 December 2010; in revised form: 19 January 2011 / Accepted: 28 January 2011 /

Published: 31 January 2011

---

**Abstract:** In this paper, we present a revision on some of the recent progresses made in characterising and understanding information inequalities, which are the fundamental physical laws in communications and compression. We will begin with the introduction of a geometric framework for information inequalities, followed by the first non-Shannon inequality proved by Zhang *et al.* in 1998 [1]. The discovery of this non-Shannon inequality is a breakthrough in the area and has led to the subsequent discovery of many more non-Shannon inequalities. We will also review the close relations between information inequalities and other research areas such as Kolmogorov complexity, determinantal inequalities, and group-theoretic inequalities. These relations have led to non-traditional techniques in proving information inequalities and at the same time made impacts back on those related areas by the introduction of information-theoretic tools.

**Keywords:** determinantal inequalities; Greene's Theorem; Kolmogorov complexity; quasi-uniformity; Shannon entropies; subspace rank inequalities

---

### 1. Introduction

Information inequalities are the “physical laws” that characterise the fundamental limits in communications and compression. Probably the most well-known information inequalities are the nonnegativity of entropy and mutual information, extending back to Shannon [2]. They are indispensable in proving converse coding theorems and play a critical role in information theory.

To illustrate the idea about how inequalities are invoked to prove a converse, consider the following classical scenario: Alice aims to send a source message  $M$  to Bob in a hostile environment where the

transmitted message may be eavesdropped by a malicious adversary Eve. In order to ensure that Eve will learn no knowledge about the source message  $M$ , Alice will encrypt it into a transmitted message  $X$  using a private key  $K$  which is known only by Bob and herself. It is well-known that in order to have perfect secrecy, the entropy of the key  $K$  is at least as large as the entropy of the message  $M$ . Such a result can be proved by invoking a few information inequalities as follows:

$$H(M) \stackrel{(a)}{=} H(M|X) \quad (1)$$

$$\stackrel{(b)}{=} I(M; K|X) \quad (2)$$

$$\stackrel{(c)}{\leq} H(K|X) \quad (3)$$

$$\stackrel{(d)}{\leq} H(K) \quad (4)$$

where (a) is due to perfect secrecy (*i.e.*,  $M$  and  $X$  are independent), (b) follows from that  $M$  can be reconstructed from the key  $K$  and the encrypted message  $X$ , (c) follows from the nonnegativity of conditional entropy  $H(K|M, X)$  and (d) is due to the nonnegativity of mutual information  $I(X; K)$ .

Besides their role in proving converse coding theorems, information inequalities are also shown to have close relations with inequalities for Kolmogorov complexities [3], group-theoretic inequalities [4], subspace rank inequalities [5], determinantal inequalities [6] and combinatorial inequalities [7]. Therefore, any new technique in characterising information inequalities will also have direct impact on these areas.

Despite its great importance, characterising information inequalities is not an easy task. It has been open for years whether there exists other information inequalities besides the nonnegativity of entropies and mutual information. No further information inequalities were found for fifty years, until [1] reported the first “non-Shannon” information inequality. The significance of that result lay not only in the inequality itself, but also in its construction. This particular approach for construction has been the main ingredient in every non-Shannon inequality that has been subsequently discovered. Using this approach, new inequalities can be found mechanically [8] and there are in fact infinitely many such independent inequalities even when there are only four random variables involved [9]. Despite this progress, a complete characterisation is still missing however.

In this survey paper, we will review some of the major progresses in the areas of information inequalities. The organisation of the paper is as follows. In Section 2, we will first outline a geometric framework for information inequalities, based on which we will explain how a Shannon inequality can be proved mechanically. Then we will outline the proof of a non-Shannon inequality which was first proved in [1]. A geometric perspective for the proof will also be given. Next, Matúš’ series of information inequality (and its relaxation) will be discussed.

In Section 3, we will consider several “equivalent frameworks” for information inequalities. First and the most natural one is for the scenario when random variables are continuous. We will prove that information inequalities for discrete and continuous random variables are “essentially the same”. Then we will change our focus to the one-to-one relation between information inequalities, inequalities for Kolmogorov complexity, group-theoretic inequalities and inequalities for box assignments. In Section 4, we will consider two constrained classes of information inequalities, subject to the constraint respectively that random variables are induced by vector subspaces and are Gaussian. These constrained

classes of information inequalities are equivalent to subspace rank inequalities and determinantal inequalities respectively.

## 2. Notations

Let  $\mathcal{N}_n = \{1, \dots, n\}$  be a finite set and  $2^{\mathcal{N}_n}$  be its power set. If  $n$  is understood implicitly, we will simply denote  $\mathcal{N}_n$  by  $\mathcal{N}$ . We define  $\mathcal{H}[\mathcal{N}]$  as the set of all real functions defined on  $2^{\mathcal{N}}$ . Hence,  $\mathcal{H}[\mathcal{N}]$  is a  $2^{|\mathcal{N}|}$ -dimensional Euclidean space. Elements in  $\mathcal{H}[\mathcal{N}]$  are called *rank functions* over  $\mathcal{N}$ . Let  $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$  be nonempty sets and  $\{X_1, X_2, \dots, X_n\}$  be  $n$  jointly distributed discrete random variables defined on  $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$  respectively. For any  $\alpha \subseteq \mathcal{N}$ ,  $X_\alpha$  denotes the joint random variable  $(X_i : i \in \alpha)$  defined over  $\mathcal{X}_\alpha$  (the Cartesian product of  $\mathcal{X}_i$  for  $i \in \alpha$ ). As an example,  $X_{\{1,2\}}$  is the random variable  $(X_1, X_2)$ . For simplicity, the parentheses in the subscript are usually omitted, *i.e.*,  $X_{\{1,2\}}$  is written as  $X_{1,2}$  (or even simply  $X_{12}$ ).

For a discrete random variable  $X$ ,  $\lambda(X)$  denotes the support of the probability distribution function of  $X$ . In other words,

$$\lambda(X) \triangleq \{x \in \mathcal{X} : \Pr(X = x) > 0\}$$

The (discrete) entropy of  $X$ , denoted by  $H(X)$ , is defined as

$$- \sum_{x \in \lambda(X)} p(x) \log p(x)$$

where  $p$  is the probability distribution of  $X$ . We will also use the following conventions. Singletons and sets with one element are not distinguished. For any set  $\{Y_i, i \in \mathcal{N}\}$  and subset  $\alpha \subseteq \mathcal{N}$ ,  $Y_\alpha$  denotes the subset  $\{Y_i, i \in \alpha\}$ .

## 3. A Framework for Information Inequalities

Let  $\{X_i, i \in \mathcal{N}\}$  be a set of discrete random variables. It induces a rank function  $h$  which is defined as follows: For any  $\alpha \subseteq \mathcal{N}$ ,

$$h(\alpha) \triangleq H(X_\alpha). \tag{5}$$

We call  $h$  the entropy function induced by  $\{X_1, \dots, X_n\}$ . For any function  $h$  in  $\mathcal{H}[\mathcal{N}]$ , we define

$$h(\alpha|\beta) \triangleq h(\alpha \cup \beta) - h(\beta), \tag{6}$$

$$I_h(\alpha; \beta) \triangleq h(\alpha) + h(\beta) - h(\alpha \cup \beta) - h(\alpha \cap \beta). \tag{7}$$

If  $h$  is the entropy function induced by random variables  $\{X_i, i \in \mathcal{N}\}$ , then  $h(\alpha|\beta)$  is the conditional entropy  $H(X_\alpha|X_\beta)$  and  $I_h(\alpha; \beta)$  is the mutual information  $I(X_\alpha; X_\beta|X_{\alpha \cap \beta})$ .

All entropy functions must satisfy the following polymatroidal axioms.

$$r(\emptyset) = 0 \tag{R1}$$

$$\alpha \subseteq \beta \implies h(\alpha) \leq h(\beta) \tag{R2}$$

$$h(\alpha \cup \beta) + h(\alpha \cap \beta) \leq h(\alpha) + h(\beta). \tag{R3}$$

The second axiom (R2) corresponds to that conditional entropy is nonnegative and the third axiom (R3) corresponds to that the conditional mutual information between  $X_\alpha$  and  $X_\beta$  given  $X_{\alpha \cap \beta}$  is nonnegative.

### 3.1. Geometric Framework

Characterisation of entropic functions is one of the most important and challenging problems in information theory. In the following, we will review the geometric framework proposed in [10] which has greatly simplified our understanding about information inequalities.

A function  $h \in \mathcal{H}[\mathcal{N}]$  is called **weakly entropic** if there exists  $\delta > 0$  such that  $\delta \cdot h$  is entropic, and is called **almost entropic** if it is the limit of a sequence of weakly entropic functions. Let  $\Gamma^*(\mathcal{N})$  be the set of all entropic functions and  $\bar{\Gamma}^*(\mathcal{N})$  be its closure. Then  $\bar{\Gamma}^*(\mathcal{N})$  is a closed and convex cone, and in fact is the set of all almost entropic functions. Compared to  $\Gamma^*(\mathcal{N})$ , its closure  $\bar{\Gamma}^*(\mathcal{N})$  is more manageable. In fact, for many application, it is sufficient to consider  $\bar{\Gamma}^*(\mathcal{N})$ . The following proves that characterising all linear information inequalities is equivalent to characterising the set  $\bar{\Gamma}^*(\mathcal{N})$ .

**Theorem 1 (Yeung [10])** *An information inequality  $\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0$  is valid (i.e., holds for all discrete random variables) if and only if*

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha h(\alpha) \geq 0, \quad \forall h \in \bar{\Gamma}^*(\mathcal{N}).$$

Unfortunately,  $\bar{\Gamma}^*(\mathcal{N})$  is still extremely difficult to characterise explicitly for  $n \geq 4$ . As we shall see, the cone is not polyhedral and hence cannot be defined by a finite number of linear inequalities. Theorem 1 offers a geometric perspective in understanding information inequalities. Based on the theorem, Yan *et al.* [11] wrote the software called Information-Theoretic Inequality Prover (ITIP) which can mechanically verify all Shannon inequalities.

The idea behind ITIP is very simple: Suppose we have a cone  $\Upsilon$  of  $\mathcal{H}[\mathcal{N}]$  such that  $\Gamma^*(\mathcal{N}) \subseteq \Upsilon$ . Consider an information inequality

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0. \tag{8}$$

Suppose one can verify that

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha h(\alpha) \geq 0, \quad \forall h \in \Upsilon.$$

Then by Theorem 1, the information inequality (8) will be valid. In other words, if the minimum of the following optimisation problem is nonnegative,

$$\text{Minimise } \sum_{\alpha \subseteq \mathcal{N}} c_\alpha h(\alpha)$$

subject to

$$h \in \Upsilon,$$

then the information inequality (8) is valid.

As  $\Upsilon$  is a cone (hence,  $\delta h \in \Upsilon$  for all  $\delta \geq 0$  and  $h \in \Upsilon$ ), it is only required to test if the origin  $\mathbf{0}$  is a global minimum or not in the above optimisation problem. Furthermore, as the optimisation problem is convex, the optimality of  $\mathbf{0}$  can be verified by checking the Karush–Kuhn–Tucker (KKT) condition.

In ITIP,  $\Upsilon$  is chosen as the cone  $\Gamma(\mathcal{N})$  whose elements are all rank functions  $h$  that satisfies the polymatroidal axioms (R1)-(R3). By picking such a cone, the ITIP can prove all inequalities that are implied by the three axioms (or equivalently, all Shannon inequalities).

### 3.2. Non-Shannon Inequalities

It has been an open question for many years whether there exist information inequalities that are not implied by Shannon’s information inequalities. This question was finally answered in [1] where non-Shannon type inequalities were constructed explicitly. The proof was based on the use of auxiliary random variables. This turns out to be a very powerful technique. In fact, all subsequently discovered non-Shannon type information inequalities are essentially proved by the same technique.

**Theorem 2 (Non-Shannon’s inequality [1])** *Let  $\{X_1, X_2, X_3, X_4\}$  be random variables. Then*

$$2I(X_3; X_4) \leq I(X_1; X_2) + I(X_1; X_3, X_4) + 3I(X_3; X_4|X_1) + I(X_3; X_4|X_2). \tag{9}$$

*Or equivalently, if  $h$  is entropic, then*

$$2I_h(3; 4) \leq I_h(1; 2) + I_h(1; 3, 4) + 3I_h(3; 4|1) + I_h(3; 4|2). \tag{10}$$

The information inequality in Theorem 2 is a non-Shannon’s inequality because one can construct a rank function  $h \in \mathcal{H}(\mathcal{N}_4)$  such that (1)  $h$  satisfies all the polymatroidal axioms (R1)-(R3) and (2)  $h$  violates the inequality (10)

To illustrate the technique in proving new inequalities, we will sketch the proof for Theorem 2. Further details can be found in [1,12].

**Sketch of proof of Theorem 2:**

Let  $h$  be the entropy function induced by a set of discrete random variables  $\{X_1, X_2, X_3, X_4\}$  whose underlying distribution is  $p$ . Construct two auxiliary random variables  $X'_1$  and  $X'_2$  such that

$$\Pr(x_1, x_2, x_3, x_4, x'_1, x'_2) = \begin{cases} p(x_3, x_4)p(x_1, x_2|x_3, x_4)p(x'_1, x'_2|x_3, x_4) & \text{if } p(x_3, x_4) > 0 \\ 0 & \text{if } p(x_3, x_4) = 0. \end{cases} \tag{11}$$

It is easy to see that the marginals of  $\{X_1, X_2, X_3, X_4\}$  and  $\{X'_1, X'_2, X_3, X_4\}$  are the same. By invoking the basic Shannon inequalities (involving six random variables), we can prove that

$$\begin{aligned} I(X_3; X_4) - I(X_3; X_4|X_1) - I(X_3; X_4|X_2) \\ = I(X_1; X'_2) - I(X_1; X'_2|X_4) - I(X_1; X'_2|X_3) - I(X_3; X_4|X_1, X'_2). \end{aligned} \tag{12}$$

Hence,

$$I(X_3; X_4) - I(X_3; X_4|X_1) - I(X_3; X_4|X_2) \leq I(X_1; X'_2). \tag{13}$$

Similarly, we can also prove that

$$I(X_3; X_4) - 2I(X_3; X_4|X_1) \leq I(X_1; X'_1), \tag{14}$$

and consequently,

$$2I(X_3; X_4) - 3I(X_3; X_4|X_1) - I(X_3; X_4|X_2) \leq I(X_1; X'_1) + I(X_1; X'_2). \tag{15}$$

Again, by invoking only Shannon’s inequalities, it can be proved that

$$I(X_1; X'_1) + I(X_1; X'_2) \leq I(X_1; X_3, X_4) + I(X'_1; X'_2) \tag{16}$$

$$= I(X_1; X_3, X_4) + I(X_1; X_2) \tag{17}$$

Combining (15) and (17), the theorem is proved.□

**Remark:** In the above proof of Theorem 2, the non-Shannon inequality is proved by invoking only a sequence of Shannon inequalities. This seems impossible at the first glance, as by definition, non-Shannon inequalities are all inequalities that are not implied by Shannon inequalities. The trick however is to apply Shannon inequalities over a larger set of random variables.

Using the geometric framework obtained earlier, we will describe in the following a “geometric interpretation” for the proof of the non-Shannon’s inequality.

Consider a set  $\mathcal{M}$  such that  $\mathcal{N} \subseteq \mathcal{M}$ . Let  $h \in \mathcal{H}[\mathcal{M}]$ . We define  $\text{proj}_{\mathcal{N}}[h]$  as a function  $g \in \mathcal{H}[\mathcal{N}]$  such that

$$g(\alpha) = h(\alpha)$$

for all  $\alpha \subseteq \mathcal{N}$ . Similarly, for any subset  $\mathcal{A}$  of  $\mathcal{H}[\mathcal{M}]$ ,  $\text{proj}_{\mathcal{N}}[\mathcal{A}]$  is the following subset

$$\text{proj}_{\mathcal{N}}[\mathcal{A}] \triangleq \{\text{proj}_{\mathcal{N}}[h] : h \in \mathcal{A}\}.$$

Now, suppose that one can construct two cones  $\Upsilon$  and  $\mathcal{C}$  such that

1.  $\Gamma^*(\mathcal{M}) \subseteq \Upsilon$ ;
2. For any  $g \in \Gamma^*(\mathcal{N})$ , there exists a  $h \in \Gamma^*(\mathcal{M}) \cap \mathcal{C}$  such that  $g = \text{proj}_{\mathcal{N}}[h]$ . Or equivalently,  $\Gamma^*(\mathcal{N}) \subseteq \text{proj}_{\mathcal{N}}[\Gamma^*(\mathcal{M}) \cap \mathcal{C}]$ .

From the conditions 1 and 2, we have

$$\Gamma^*(\mathcal{N}) \subseteq \text{proj}_{\mathcal{N}}[\Upsilon \cap \mathcal{C}].$$

Again, using Theorem 1, we can prove that an information inequality

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0 \tag{18}$$

is valid if

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha g(\alpha) \geq 0, \quad \forall g \in \text{proj}_{\mathcal{N}}(\mathcal{C} \cap \Upsilon).$$

Equivalently, the inequality (18) is valid if the minimum of the following linear program is zero.

$$\begin{aligned} &\text{Minimise } \sum_{\alpha \subseteq \mathcal{N}} c_\alpha h(\alpha) \\ &\text{subject to} \\ &h \in \mathcal{C} \cap \Upsilon. \end{aligned} \tag{19}$$

**Remark:** Instead of verifying if an information inequality is valid or not, we can also use the Fourier-Motzkin elimination method to find all linear inequalities that defines the cone  $\text{proj}_{\mathcal{N}}(\mathcal{C} \cap \Upsilon)$ . Clearly, each such inequality corresponds to a valid information inequality over  $\{X_i, i \in \mathcal{N}\}$ .

Now, we will revisit the non-Shannon inequality in Theorem 2. Let  $\mathcal{N} = \{1, 2, 3, 4\}$  and  $\mathcal{M} = \{1, 2, 3, 4, 1', 2'\}$ . Given any random variables  $\{X_1, \dots, X_4\}$ , construct two random variables  $X'_1$  and  $X'_2$  such that the probability distribution of  $\{X_1, X_2, X_3, X_4, X'_1, X'_2\}$  is given by (11). Let  $g$  be the entropy function of  $\{X_1, \dots, X_4\}$  and  $h$  be the entropy function of  $\{X_1, X_2, X_3, X_4, X'_1, X'_2\}$ . Then it is easy to see that for all  $i \in \{1, 2\}$  and  $\beta \subseteq \{3, 4\}$ ,

$$I_h(1, 2; 1', 2'|3, 4) = 0, h(1, 2, \beta) = h(1', 2', \beta), h(i, \beta) = h(i', \beta)$$

and  $\text{proj}_{\mathcal{N}}[h] = g$ .

Let

$$\mathcal{C} \triangleq \left\{ h \in \mathcal{H}[\mathcal{M}] : \begin{aligned} &I_h(1, 2; 1', 2'|3, 4) = 0, h(1, 2, \beta) = h(1', 2', \beta), \\ &h(i, \beta) = h(i', \beta), \forall i \in \{1, 2\} \text{ and } \beta \subseteq \{3, 4\} \end{aligned} \right\}$$

and  $\Upsilon = \Gamma(\mathcal{M})$  (which is the set of all functions  $h$  that satisfies the polymatroidal axioms). Then clearly  $\Gamma^*(\mathcal{M}) \subseteq \Gamma(\mathcal{M})$  and  $\Gamma^*(\mathcal{N}) \subseteq \text{proj}_{\mathcal{N}}[\Upsilon \cap \mathcal{C}]$ . It can be numerically verified that the minimum of the linear program in (19) is zero when the information inequality is the non-Shannon inequality (9). Consequently, the non-Shannon inequality is indeed proved.

### 3.3. Non-Polyhedral Property

In the pervious subsection, we have discussed a promising technique in proving (or even discovering) new information inequalities. Using the same technique proposed in [1], more and more linear information inequalities have been discovered [8,13–15]. Later in [9], Matúš obtained a countable infinite set of linear information inequalities for a set of four random variables. Using the same set of inequalities, Matúš further proved that  $\bar{\Gamma}^*(\mathcal{N}_4)$  is not a polyhedral. In the following, we will review Matúš’ inequalities and its relaxation.

**Remark:** The non-polyhedral property of  $\bar{\Gamma}^*(\mathcal{N}_4)$  was later used by [16] to show that the set of achievable tuples of a network is in general also non-polyhedral. As a result, this proved that the Linear Programming bounds is not tight in general.

**Theorem 3 (Matúš)** Let  $s \in \mathbb{Z}^+$  and  $g \in \Gamma^*(\mathcal{N}_4)$ . Then

$$s (\square_{12,34} g + \triangle_{34|2} g + \triangle_{24|3} g) + \triangle_{23|4} g + \frac{s(s-1)}{2} (\triangle_{24|3} g + \triangle_{34|2} g) \geq 0 \tag{20}$$

where for any distinct elements  $i, j, k \in \mathcal{N}_4$ ,

$$\begin{aligned} \triangle_{ij|k} g &\triangleq I_g(X_i; X_j|X_k), \\ \square_{12,34} g &\triangleq g(13) + g(23) + g(14) + g(24) + g(34) \\ &\quad - g(12) - g(3) - g(4) - g(134) - g(234). \end{aligned}$$

While Matúš proved a series of linear information inequalities, it is sometimes difficult to use these infinitely number of inequalities at the same time. In [17], the series of Matúš’ inequalities is relaxed to a single non-linear inequality.

**Remark:** Using one single nonlinear inequality, it can be proved that the set of all almost entropic functions is not polyhedral.

**Theorem 4 (Quadratic information inequality [17])** Let  $g \in \bar{\Gamma}^*(\mathcal{N})$ ,

$$\begin{aligned} a(g) &\triangleq \frac{1}{2} (\Delta_{24|3} g + \Delta_{34|2} g) \\ b(g) &\triangleq \square_{12,34} g + \Delta_{34|2} g + \Delta_{24|3} g \\ c(g) &\triangleq \Delta_{23|4} g \\ w(g) &\triangleq \begin{cases} -\frac{b(g)-a(g)}{2a(g)} & \text{if } a(g) > 0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{21}$$

If  $b(g) \leq 2a(g)$ , then

$$(b(g) - a(g))^2 - 4a(g)c(g) \leq \min(4a(g)^2(w(g) - \lfloor w(g) \rfloor)^2, 4a(g)^2(\lceil w(g) \rceil - w(g))^2) \tag{22}$$

and consequently,

$$\left( \square_{12,34} g + \frac{\Delta_{24|3} g + \Delta_{34|2} g}{2} \right)^2 - 2(\Delta_{24|3} g + \Delta_{34|2} g)\Delta_{32|4} g \leq \frac{(\Delta_{24|3} g + \Delta_{34|2} g)^2}{4}. \tag{23}$$

**Remark:** Subject to the constraint that  $b(g) > 2a(g)$ , then the series of linear inequalities (20) is implied by the Shannon inequalities. Therefore, the constraint (i.e.,  $b(g) > 2a(g)$ ) we imposed on Theorem 4 is not critical.

**Conjecture 1** (20) holds for all  $s \geq 0$ . Consequently, if  $b(g) \leq 2a(g)$ , then

$$(b(g) - a(g))^2 - 4a(g)c(g) \leq 0.$$

#### 4. Equivalent Frameworks

In the previous section, we have described a framework for information inequalities for discrete random variables. We have also demonstrated the common proving technique. In this section, we will construct several different frameworks which are “equivalent” or “almost equivalent” to the earlier one. These equivalence relations among different frameworks will turn out to be very useful in deriving new information theoretic tools.

##### 4.1. Differential Entropy

The previous framework for information inequalities assumes that all random variables are discrete. A very natural extension of the framework is thus to relax the restriction by allowing random variables to be continuous. To achieve this goal, we will first need an analogous definition of discrete entropy in the domain of continuous random variables.

**Definition 1 (Differential entropies)** Let  $\{X_i, i \in \mathcal{N}\}$  be a set of continuous random variables such that  $X_i$  are real numbers. For any  $\alpha \subseteq \mathcal{N}$ , let  $f_\alpha(x_i, i \in \alpha)$  be the density functions for  $(X_i, i \in \alpha)$ . Then the **differential entropy** of  $(X_i, i \in \alpha)$  is denoted by

$$H(X_\alpha) \triangleq - \int f(x_\alpha) \log f(x_\alpha) d_{x_\alpha}.$$

**Remark:** For notation simplicity, we abuse our notations by using  $H(X)$  to denote both discrete and differential entropies. However, its exact meaning should be clear from the context.

Discrete and differential entropies shared similar and dissimilar properties. The main difference is that differential entropy can be negative, unlike discrete entropy. However, mutual information and its conditional counterpart (by defined analogously as in (7)) remain nonnegative. In fact, as we shall see, the sets of information inequalities for discrete and continuous random variables are almost the same.

**Definition 2 (Balanced inequalities)** An information inequality  $\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0$  (for either discrete or continuous random variables) is called **balanced** if for all  $n \in \mathcal{N}$ ,  $\sum_{\alpha \subseteq \mathcal{N}: n \in \alpha} c_\alpha = 0$ .

For any information inequality  $\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0$  or expression  $\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha)$ , its  $n^{th}$  residual weight  $r_n$  is defined as

$$r_n \triangleq \sum_{\alpha \subseteq \mathcal{N}: n \in \alpha} c_\alpha. \tag{24}$$

Clearly, an information inequality is balanced if and only if  $r_n = 0$  for all  $n \in \mathcal{N}$ .

**Example 1** The residual weights  $r_1, r_2$  of the information inequality  $H(X_1) + H(X_2) \geq 0$  are both equal to one. Hence, the inequality is not balanced.

For any information inequality  $\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0$ , its balanced counterpart is the following inequality

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) - \sum_{n \in \mathcal{N}} r_n H(X_n | X_i, i \neq n) \geq 0, \tag{25}$$

which is balanced (as its name suggests).

**Proposition 1 (Necessary and sufficiency of balanced inequalities [6])** For any valid information inequality  $\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0$ , it is a valid discrete information inequality if and only if

1. its residual weights  $r_n \geq 0$  for all  $n$ , and
2. its balanced counterpart is also valid.

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) - \sum_{n \in \mathcal{N}} r_n H(X_n | X_i, i \neq n) \geq 0.$$

Consequently, all valid discrete information inequalities are implied by the set of all valid balanced inequalities and the nonnegativity of (conditional) entropies.

It turns out that this set of balanced information inequalities also play the same significant role for inequalities involving continuous random variables.

**Theorem 5 (Equivalence [6])** All information inequalities for continuous random variables are balanced. Furthermore, a balanced information inequality

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) - \sum_{n \in \mathcal{N}} r_n H(X_n | X_i, i \neq n) \geq 0$$

is valid for continuous random variable if and only if it is also valid for discrete random variables.

By Theorem 5, to characterise information inequalities, it is sufficient to consider only balanced information inequalities which are the same for either discrete or continuous random variables.



Since  $U$  is uniformly distributed over  $G$ , we can easily prove that  $X_K$  is uniformly distributed over  $\Omega_K$  and that

$$H(X_K) = \log |G|/|K|.$$

The above construction of a random variable from a subgroup can be extended naturally to multiple subgroups.

**Theorem 7 (Group characterisable random variables [4])** *Let  $G$  be a finite group and  $\{G_i, i \in \mathcal{N}\}$  be a set of subgroups of  $G$ . For each  $i \in \mathcal{N}$ , let  $X_i$  be the random variable induced by the subgroup  $G_i$  as defined above. Then for any  $\alpha \subseteq \mathcal{N}$ ,*

1.  $H(X_i, i \in \alpha) = \log |G|/|\cap_{i \in \alpha} G_i|$ ,
2.  $|\lambda(X_i, i \in \alpha)| = |\cap_{i \in \alpha} G_i|$ ,
3.  $(X_i, i \in \alpha)$  is uniformly distributed over its support. In other word, the value of the probability distribution function of  $(X_i, i \in \alpha)$  is either zero or is a constant.

**Definition 4** *A function  $h \in \mathcal{H}[\mathcal{N}]$  is called **group characterisable** if it is the entropy function of a set of random variables  $\{X_1, \dots, X_n\}$  induced by a finite group  $G$  and its subgroups  $\{G_1, \dots, G_n\}$ . Furthermore,  $h$  is*

1. **representable** if  $\{G, G_1, \dots, G_n\}$  are all vector space, and
2. **abelian** if  $G$  is abelian.

Clearly, random variables induced by a set of subgroups must satisfy all valid information inequalities. Therefore, we have the following theorem.

**Theorem 8 (Group-theoretic inequalities [4])** *Let*

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0 \tag{30}$$

*be a valid information inequality. Then for any finite group  $G$  and its subgroups  $\{G_i, i \in \mathcal{N}\}$ , we have*

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha \log \frac{|G|}{|\cap_{i \in \alpha} G_i|} \geq 0, \tag{31}$$

*or equivalently,*

$$|G|^{\sum_{\alpha \subseteq \mathcal{N}} c_\alpha} \geq \prod_{\alpha \subseteq \mathcal{N}} |\cap_{i \in \alpha} G_i|^{c_\alpha}. \tag{32}$$

Theorem 8 proved that we can directly “translate” any information inequality into a group-theoretic inequality. A very surprising result proved in [4] was that the the converse also holds.

**Theorem 9 (Converse [4])** *The information inequality (30) is valid if it is satisfied by all random variables induced by groups, or equivalently, the group-theoretic inequality (32) is valid.*

Theorems 8 and 9 suggested that to prove an information inequality, it is necessary and sufficient to verify if the inequality is satisfied by all random variables induced by groups. Later, we will further illustrate how to use the two theorems to derive a group-theoretic proof for information inequalities.

In the following, we will further prove that many statistical properties of random variables induced by groups will have analogous algebraic interpretations.

**Lemma 1 (Properties of group induced random variables)** *Suppose that  $\{X_i, i \in \mathcal{N}\}$  is a set of random variables induced by a finite group  $G$  and its subgroups  $\{G_i, i \in \mathcal{N}\}$ . Then*

1. **(Functional dependency)**  $H(X_l|X_i, i \in \alpha) = 0$  (i.e.,  $X_l$  is a function of  $X_\alpha$ ) if and only if  $\bigcap_{i \in \alpha} G_i \subseteq G_l$ . Hence, functional dependency is equivalent to subset relation;
2. **(Independency)**  $I(X_i; X_j|X_l) = 0$  if and only if

$$|G_i \cap G_l| |G_j \cap G_l| = |G_l| |G_i \cap G_j \cap G_l|; \tag{33}$$

3. **(Conditioning preserves group characterisation)** for any fixed any  $\alpha \subseteq \mathcal{N}$ , the group  $K \triangleq \bigcap_{i \in \alpha} K_i$  and its subgroups  $K_i \triangleq K \cap G_i$  for  $i \in \mathcal{N}$  induce a set of random variables  $\{Y_i, i \in \mathcal{N}\}$  such that

$$H(Y_i, i \in \beta) = H(X_i, i \in \beta | X_j, j \in \alpha)$$

for all  $\beta \subseteq \mathcal{N}$ . In other words, for any group characterisable  $h \in \mathcal{H}[\mathcal{N}]$ , let  $g \in \mathcal{H}[\mathcal{N}]$  such that

$$g(\beta) = h(\beta|\alpha)$$

for all  $\beta \subseteq \mathcal{N}$ . Then  $g$  is also group characterisable.

**Proposition 2 (Duality [19])** *Let  $\{V_1, \dots, V_n\}$  be a set of vector subspaces of  $V \triangleq \mathbb{F}^m$  over the finite field  $\mathbb{F}$ . Define the following subspace  $W_i$  for  $i \in \mathcal{N}$ :*

$$W_i = \{w \in V : v^\top w = 0\}. \tag{34}$$

Then, for any  $\alpha \subseteq \mathcal{N}$ ,

$$\dim \langle V_i, i \in \alpha \rangle = \dim V - \dim \bigcap_{i \in \alpha} W_i = \log \frac{|V|}{|\bigcap_{i \in \alpha} W_i|}.$$

Hence, if  $h \in \mathcal{H}[\mathcal{N}]$  such that  $h(\alpha) \triangleq \dim \langle V_i, i \in \alpha \rangle$  for all  $\alpha \subseteq \mathcal{N}$ , then  $h$  is weakly representable.

**Remark:** While  $W^\perp$  and  $W$  are both subspaces of  $V$  and  $\dim W + \dim W^\perp = \dim V$ ,  $\langle W, W^\perp \rangle \neq V$  in general. If  $\mathbb{F} = \mathbb{R}$ , then  $W_i$  (defined as in (34)) is the orthogonal complement of  $V_i$ .

Theorems 8 and 9 suggested that proving an information inequality (30) is equivalent to proving a group-theoretic inequality (32). In the following, we will illustrate the idea by providing a group-theoretic proof for nonnegativity of mutual information

$$H(X_1) + H(X_2) \geq H(X_1, X_2). \tag{35}$$

**Example 2 (Group-theoretic Proof)** Let  $G$  be a finite group and  $G_1$  and  $G_2$  be its subgroups. Let

$$S = \{a \circ b : a \in G_1, b \in G_2\}$$

where  $\circ$  is the binary group operator. As  $S$  is a subset of  $|G|$ ,  $|S| \leq |G|$ . With a simple counting argument (by removing duplications), it can be proved easily that

$$|S| = \frac{|G_1||G_2|}{|G_1 \cap G_2|}.$$

Therefore,

$$|G||G_1 \cap G_2| \geq |G_1||G_2|.$$

Finally, according to Theorems 8 and 9, the inequality (35) follows.

It is worth mentioning that Theorems 8 and 9 also suggested an information-theoretic proof for group-theoretic inequalities. For example, the following information inequality

$$H(X_1) + H(X_2) + 2H(X_1, X_2) + 4H(X_3) + 4H(X_4) \tag{36}$$

$$+ 5H(X_1, X_3, X_4) + 5H(X_2, X_3, X_4) \tag{37}$$

$$\leq 6H(X_3, X_4) + 4H(X_1, X_3) + 4H(X_1, X_4) \tag{38}$$

$$+ 4H(X_2, X_3) + 4H(X_2, X_4), \tag{39}$$

implies the following group-theoretic inequality

$$|G_{34}|^6 |G_{13}|^4 |G_{14}|^4 |G_{23}|^4 |G_{24}|^4 \leq |G_1| |G_2| |G_3|^4 |G_4|^4 |G_{12}|^2 |G_{134}|^5 |G_{234}|^5 \tag{40}$$

The meaning of this inequality and its implications in group theory are yet to be understood.

#### 4.4. Combinatorial Perspective

Random variables that are induced by groups have many interesting properties. One interesting property is that they are *quasi-uniform* in nature.

**Definition 5 (Quasi-uniform random variables)** A set of random variables  $\{X_1, \dots, X_n\}$  is called quasi-uniform if for all  $\alpha \subseteq \mathcal{N}$ ,  $X_\alpha \triangleq (X_i, i \in \alpha)$  is uniformly distributed over its support  $\lambda(X_\alpha)$ . In other words,

$$\Pr(X_\alpha = x_\alpha) = \begin{cases} 1/|\lambda(X_\alpha)| & \text{if } x_\alpha \in \lambda(X_\alpha) \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

Since  $X_\alpha$  is uniformly distributed for all  $\alpha \subseteq \mathcal{N}$ , the entropy  $H(X_\alpha)$  is thus equal to  $\log |\lambda(X_\alpha)|$ .

According to the Asymptotic Equipartition Property (AEP) [12], for a sufficiently long sequence of independent and identically distributed random variables, the set of typical sequences has a total probability close to one and the probability of each typical sequence is approximately the same. In certain sense, quasi-uniform random variables possess the *non-asymptotic* equipartition property that the probabilities are completely concentrated and uniformly distributed over their supports. As a result,

quasi-uniform random variables can be fully characterised by their supports (because the probability distributions are uniform over the supports). This offers a combinatorial interpretation for quasi-uniform random variables. And it turns out that this interpretation offers a combinatorial approach to proving information inequalities.

**Definition 6 (Box assignment)** Let  $\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$  be nonempty finite sets and  $\mathcal{X}$  be their Cartesian product  $\prod_{i=1}^n \mathcal{X}_i$ . A box assignment  $\mathcal{A}$  in  $\mathcal{X}$  is a nonempty subset of  $\mathcal{X}_{\mathcal{N}}$ .

For any  $\alpha \subseteq \mathcal{N}$  and  $a_\alpha \triangleq (a_i, i \in \alpha) \in \prod_{i \in \alpha} \mathcal{X}_i$ , we define

$$\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha) \triangleq \{(x_j, j \in \mathcal{N}) \in \mathcal{A} : x_i = a_i, i \in \alpha\}, \tag{42}$$

$$\mathcal{A}_\alpha \triangleq \{(a_i, i \in \alpha) : |\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)| \geq 1\}. \tag{43}$$

Roughly speaking,  $\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)$  is the set of elements in  $\mathcal{A}$  such that its “ $i^{\text{th}}$ -coordinate” is  $a_i$  for  $i \in \alpha$ . The set  $\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)$  will be called the  $a_\alpha$ -layer of  $\mathcal{A}$ . And hence,  $\mathcal{A}_\alpha$  contains all  $a_\alpha$  such that the  $a_\alpha$ -layer of  $\mathcal{A}$  is nonempty. And we will call  $\mathcal{A}_\alpha$  the  $\alpha$ -projection of  $\mathcal{A}$ .

**Definition 7 (Quasi-uniform box assignment)** A box assignment  $\mathcal{A}$  is called **quasi-uniform** if for any  $\alpha \subseteq \mathcal{N}$ , the cardinality of  $\mathcal{A}_{\mathcal{N}|\alpha}(a_\alpha)$  is constant for all  $a_\alpha \in \mathcal{A}_\alpha$ . And we will denote the constant by  $|\mathcal{A}_{\mathcal{N}|\alpha}|$  for simplicity.

The following proposition proves that quasi-uniform box assignment and quasi-uniform random variables are in fact equivalent.

**Proposition 3 (Equivalence [7])** Let  $\{X_1, \dots, X_n\}$  be a set of quasi-uniform random variables and  $\mathcal{A}$  be its probability distribution’s support. Then  $\mathcal{A}$  is a quasi-uniform box assignment in  $\prod_{i \in \mathcal{N}} \mathcal{X}_i$ . Furthermore, for all  $\alpha \subseteq \mathcal{N}$ ,

$$H(X_\alpha) = \log |\mathcal{A}_\alpha|. \tag{44}$$

Conversely, for any quasi-uniform box assignment  $\mathcal{A}$ , there exists a set of quasi-uniform random variables  $\{X_1, \dots, X_n\}$  whose probability distribution’s support is indeed  $\mathcal{A}$ .

As random variables induced by groups are quasi-uniform, by Theorems 8 and 9, we have the following combinatorial interpretation for information inequalities.

**Theorem 10 (Combinatorial interpretation [7])** An information inequality

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0 \tag{45}$$

is valid if and only if the following box assignment inequality is valid

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha \log |\mathcal{A}_\alpha| \geq 0, \tag{46}$$

or equivalently,

$$\prod_{\alpha \subseteq \mathcal{N}} |\mathcal{A}_\alpha|^{c_\alpha} \geq 1 \tag{47}$$

for all quasi-uniform box assignments  $\mathcal{A}$ .

Again, in the following example, we will illustrate how to use the combinatorial interpretation to derive a “combinatorial proof” for information inequality.

**Example 3 (Combinatorial proof)** Let  $\mathcal{A}$  be a quasi-uniform box assignment in  $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ . Suppose  $(a_1, a_2) \in \mathcal{A}$ . Then it is obvious that  $a_1 \in \mathcal{A}_1$  and  $a_2 \in \mathcal{A}_2$ . In other words,  $\mathcal{A} \subseteq \mathcal{A}_1 \times \mathcal{A}_2$  and consequently,

$$|\mathcal{A}_1| \times |\mathcal{A}_2| \geq |\mathcal{A}_{1,2}|.$$

By Theorem 10, we prove that  $H(X_1) + H(X_2) \geq H(X_1, X_2)$ .

#### 4.5. Coding Perspective

We can also view a box assignment  $\mathcal{A}$  as an error correcting code such that  $\mathcal{A}$  is the set of all codewords. For each codeword  $(a_1, \dots, a_n)$ ,  $a_i$  is the  $i^{th}$  symbol to be transmitted across a channel. Taking this coding perspective, in the following, a box assignment will simply be called a *code*. Also, a code  $\mathcal{C}$  is called a *quasi-uniform code* if  $\mathcal{C}$  is a quasi-uniform box assignment. Again, each quasi-uniform code  $\mathcal{C}$  will induce a set of quasi-uniform random variables  $\{X_1, \dots, X_n\}$ .

For any code  $\mathcal{C}$  (which is just a box assignment) and two codewords  $c, c' \in \mathcal{C}$ , the *Hamming distance* between codewords  $c \triangleq (c_1, \dots, c_n)$  and  $c' \triangleq (c'_1, \dots, c'_n)$  is defined as

$$D(c, c') \triangleq |\{i \in \mathcal{N} : c_i \neq c'_i\}|.$$

In addition, the minimum Hamming distance of the code  $\mathcal{C}$  is defined as

$$D(\mathcal{C}) \triangleq \min_{c \neq c', c, c' \in \mathcal{C}} D(c, c')$$

The minimum Hamming distance of a code characterises how strong the error correcting capability of the code is. Specifically, a code  $\mathcal{C}$  with a minimum Hamming distance  $d$  can correct up to  $\lfloor \frac{d-1}{2} \rfloor$ 's symbol errors.

**Example 4** Let  $\mathcal{C}$  be a length-3 code containing only two codewords  $(0, 0, 0)$  and  $(1, 1, 1)$ . The minimum Hamming distance of this code is 3 and hence can correct any single symbol error. For instance, suppose the codeword  $(0, 0, 0)$  is transmitted. If a symbol error occurs, the receiver will receive either  $(1, 0, 0)$ ,  $(0, 1, 0)$  or  $(0, 0, 1)$ . In any case, the receiver can always determine which symbol is erroneous (by using a bounded-distance decoder) and hence can correct it.

In addition to the minimum Hamming distance, in many cases, a code's *distance profile* is also of great importance: Let  $\mathcal{C}$  be a code and  $c$  be a codeword in  $\mathcal{C}$ . The distance profile of  $\mathcal{C}$  centered at  $c$  is a set of integers  $A(\mathcal{C}, c) \triangleq \{A_r(c) : r = 1, \dots, n\}$  where

$$A_r(c) \triangleq |\{c' \in \mathcal{C} : D(c, c') = r\}|.$$

In other words,  $A_r(c)$  is the number of codewords in  $\mathcal{C}$  such that their Hamming distances to the centering codeword  $c$  is  $r$ .

The profile  $A(\mathcal{C}, c)$  contains information about how likely a decoding error (*i.e.*, the receiver decodes a wrong codeword) occurs if the transmitted codeword is  $c$ . In general, the distance profile  $A(\mathcal{C}, c)$  depends

on the choice of  $c$ . A code is called *distance-invariant* if its distance profile  $A(\mathcal{C}, c)$  is independent of  $c$ . Roughly speaking, a distance-invariant code is one where the probability of decoding error is the same for all transmitted codewords  $c \in \mathcal{C}$ .

**Theorem 11 (Distance invariance [20])** *Quasi-uniform codes are distance-invariant.*

**Example 5 (Linear codes)** *Let  $P$  be a  $n - k \times n$  parity check matrix (over a finite field  $\mathbb{F}$ ) and the code  $\mathcal{C}$  is defined by*

$$\mathcal{C} = \{c \in \mathbb{F}^n : Pc = \mathbf{0}\}.$$

*Then  $\mathcal{C}$  is called a linear code. Note that, for a linear code, if  $c_1, c_2 \in \mathcal{C}$ , then  $c_1 + c_2$  is also contained in  $\mathcal{C}$ . Linear codes are quasi-uniform codes and hence are also distance invariant.*

In the following, we will consider only quasi-uniform codes. For simplicity, we will assume without loss of generality that there is a zero-codeword  $\mathbf{0} \in \mathcal{C}$  (by renaming). Also, for any  $c \in \mathcal{C}$ , we define the Hamming weight of the codeword  $c$  (denoted by  $D(c)$ ) as  $D(c, \mathbf{0})$ .

**Definition 8 (Weight enumerator)** *The weight enumerator of a quasi-uniform code  $\mathcal{C}$  with length  $n$  is*

$$W_{\mathcal{C}}(x, y) = \sum_{r=1}^n A_r x^{n-r} y^r$$

*where  $x$  and  $y$  are indeterminates, and  $A_r \triangleq A_r(\mathbf{0})$ . Using simple counting, it is easy to prove that*

$$W_{\mathcal{C}}(x, y) \triangleq \sum_{c \in \mathcal{C}} x^{n-D(c)} y^{D(c)}. \tag{48}$$

In many cases, it is more convenient to work with weight enumerator than distance profile. However, conceptually, they are equivalent (*i.e.*, they can be uniquely obtained from each other). Clearly, the weight enumerator is uniquely determined from the code  $\mathcal{C}$ . However, what “structural property” of the code  $\mathcal{C}$  determines the weight enumerator? For example, suppose that we construct a new code from  $\mathcal{C}$  by exchanging the first and the second codeword symbols. It is obvious that this modification will not affect the weight enumerator. In other words, ordering of the codeword symbols has no effects on the weight enumerator. The question therefore is: What property of a code has direct effects on the weight enumerator?

To answer the question, let us use the old perspective that a quasi-uniform code is merely a quasi-uniform box assignment (and also its associated set of quasi-uniform random variables). These random variables  $\{X_1, \dots, X_n\}$  have a simple interpretation here: Suppose a codeword  $C = (C_1, \dots, C_n)$  is randomly and uniformly selected from  $\mathcal{C}$ . Then  $X_i$  is the  $i^{th}$  symbol in the random codeword  $C$ , *i.e.*,  $X_i = C_i$ . Our answer to the above question is given in the following theorem.

**Theorem 12 (Generalised Greene’s Theorem [20])** *Let  $\mathcal{C}$  be a quasi-uniform code and  $\{X_1, \dots, X_n\}$  be its induced quasi-uniform random variables. Suppose that  $\rho$  is the entropy function of  $\{X_1, \dots, X_n\}$ . In other words,  $\rho(\alpha) = H(X_i, i \in \alpha)$ . Then*

$$W_{\mathcal{C}}(x, y) = \sum_{\alpha \subseteq \mathcal{N}} 2^{\rho(\mathcal{N}) - \rho(\alpha)} (x - y)^{|\alpha|} y^{n - |\alpha|}. \tag{49}$$

**Remark:** The Greene’s Theorem is a special case of Theorem 12 when the code  $\mathcal{C}$  is a linear code.

By Theorem 12, the weight enumerator (and also the error-correcting capability) of a quasi-uniform code depends only on the entropy function induced by the codeword symbol random variables. By exploiting the relation between the entropy function of a set of quasi-uniform random variables and the weight enumerator of the induced code, we open a new door on how to harness coding theory results to derive new information theory results.

**Example 6 (Code-theoretic proof)** Consider a set of quasi-uniform random variables  $\{X_1, X_2\}$  which induces a length-2 quasi-uniform code  $C$ . The length of the code is 2. By the Generalised Greene’s Theorem, the number of codewords which have Hamming weights 1 is given by

$$A_1 = (2^{H(X_1, X_2) - H(X_1)} + 2^{H(X_1, X_2) - H(X_2)} - 2) . \tag{50}$$

As  $A_1$  is nonnegative, (50) implies that

$$\min(H(X_1), H(X_2)) \leq H(X_1, X_2). \tag{51}$$

Finally, by Theorem 10 (a variation of which to be precise), an information inequality holds if and only if it also holds for all quasi-uniform random variables. Consequently, we prove that (51) holds for all random variables.

### 5. Constrained Information Inequalities

In pervious sections, we considered general information inequalities where we do not impose any constraint on the choice of random variables. In the following, we will focus on two constrained classes of information inequalities: *subspace rank inequalities* and *determinantal inequalities*.

#### 5.1. Rank Inequalities

Let  $\{V_1, \dots, V_n\}$  be a set of vector subspaces over a field  $\mathbb{F}$ . A subspace rank inequality is an inequality about the rank or dimension of subspaces in the following form:

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha \dim \langle V_i, i \in \alpha \rangle \geq 0. \tag{52}$$

For example, it is straightforward to prove that

$$\dim \langle V_1 \rangle + \dim \langle V_2 \rangle \geq \dim \langle V_1, V_2 \rangle, \tag{53}$$

which is a direct consequence of the following identity

$$\dim \langle V_1 \rangle + \dim \langle V_2 \rangle = \dim \langle V_1, V_2 \rangle + \dim V_1 \cap V_2. \tag{54}$$

Subspace rank inequalities are in fact constrained information inequalities subject to the criteria that random variables are induced by vector subspaces over a field. Clearly, all valid information inequalities (including all Shannon inequalities) are subspace rank inequalities. For example, the subspace rank inequality (53) is indeed equivalent to the nonnegativity of mutual information. Besides all these known

unconstrained information inequalities, one of the most well-known subspace rank inequalities is the Ingleton inequalities [21]. A recent work [22] proved that Ingleton inequalities also include Shannon inequalities as special cases and determined the unique minimal set of Ingleton inequalities that imply all the others.

**Theorem 13 (Ingleton inequality)** *Suppose  $r$  is a representable polymatroid over  $\mathcal{X}$ . Then for every choice of subsets  $X_1, X_2, X_3, X_4 \subseteq \mathcal{X}$*

$$0 \leq r(\mathcal{X}_1 \cup \mathcal{X}_2) + r(\mathcal{X}_1 \cup \mathcal{X}_3) + r(\mathcal{X}_1 \cup \mathcal{X}_4) + r(\mathcal{X}_2 \cup \mathcal{X}_3) + r(\mathcal{X}_2 \cup \mathcal{X}_4) - r(\mathcal{X}_1) - r(\mathcal{X}_2) - r(\mathcal{X}_3 \cup \mathcal{X}_4) - r(\mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_3) - r(\mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_4). \tag{55}$$

It has been open for years whether there exists subspace rank inequalities that are not implied by Ingleton inequalities and Shannon inequalities. It was until recently that the question was finally answered. In [5], insufficiency of Ingleton inequality to characterise all subspace rank inequalities was proved. And in [23,24], new subspace rank inequalities not implied by Ingleton inequalities were explicitly constructed. In fact, the set of subspace rank inequalities for up to five variables have all been determined. However, the complete characterisation involving more than five variables is still missing. In the following, we will review some of the important results along this line of work.

**Theorem 14 (Kinser [23])** *Suppose  $\mathcal{X} = \{X_1, \dots, X_n\}$  and  $h$  is representable over  $\mathcal{X}$ . Then*

$$h(X_{1,2}) + h(X_{1,3,n}) + h(X_3) + \sum_{i=4}^n (h(X_i) + h(X_{2,i-1,i})) \leq h(X_{1,3}) + h(X_{1,n}) + h(X_{2,3}) + \sum_{i=4}^n (h(X_{2,i}) + h(X_{i-1,i})). \tag{56}$$

Or equivalently,

$$I_h(X_2; X_3) \leq I_h(X_1; X_2) + I_h(X_3; X_n|X_1) + \sum_{i=4}^n I_h(X_2; X_{i-1}|X_i). \tag{57}$$

**Theorem 15 (Dougherty et al. [24])** *Suppose  $\mathcal{X} = \{A, B, C_1, \dots, C_n\}$  and  $h$  is representable over  $\mathcal{X}$ . Then*

$$(n - 1)I_h(A; B) \leq \sum_{i=1}^n I_h(A; B|C_i) + \sum_{i=1}^n h(C_i) - h(C_1, \dots, C_n). \tag{58}$$

**Remark:** In addition to the inequalities obtained in Theorem 15, the work [24] found all subspace rank inequalities in five variables (called *DFZ inequalities*) and many more other new inequalities in six variables.

**Definition 9 ( $\epsilon$ -truncation)** *Let  $h$  be a polymatroid over  $\mathcal{Y}$  and  $0 \leq \epsilon \leq h(\mathcal{Y})$ . Define  $g$  as follows where*

$$g(\alpha) \triangleq \min(h(\alpha), h(\mathcal{Y}) - \epsilon), \forall \alpha \subseteq \mathcal{Y}. \tag{59}$$

Then  $g$  is called the  $\epsilon$ -truncation of  $h$ .

**Definition 10 (Truncation-preserving inequalities)** Let  $\mathcal{Y} = \{Y_i, i \in \mathcal{N}_m\}$ . A set of rank inequalities

$$\left\{ \sum_{\alpha \subseteq \mathcal{N}_m} c_\alpha^\ell H(Y_i, i \in \alpha) \geq 0, \ell \in \Delta \right\} \tag{60}$$

is said to preserve truncation (or is truncation-preserving) if for any  $h$  satisfying all the inequalities in (60), its truncation also satisfies all the inequalities.

**Proposition 4 (Chan et al. [5])** DFZ inequalities are truncation preserving.

**Theorem 16 (Insufficiency of truncation preserving inequalities [5])** Let  $\Delta_n$  be the set of all subspace rank inequalities involving  $n$  variables (or subspaces). Then for sufficiently large  $n$ ,  $\Delta_n$  is not truncation-preserving.

### 5.2. Determinantal Inequalities

Information inequalities for Gaussian random variables are another interesting class of information inequalities. As we shall see, they are equivalent to determinantal inequalities.

**Definition 11 (Gaussian polymatroid)** Let  $h$  be a polymatroid over  $\mathcal{N}$ . It is called **Gaussian** if there exists a set of jointly Gaussian random variables  $\{Y_j, j \in \mathcal{K}\}$  with a  $|\mathcal{K}| \times |\mathcal{K}|$  covariance matrix and a partition of  $\mathcal{K}$  into  $n$  disjoint nonempty subsets  $\beta_1, \dots, \beta_n$  such that for any  $\alpha \subseteq \mathcal{N}$ ,

$$h(\alpha) = H(X_i, i \in \alpha). \tag{61}$$

where  $X_i = (Y_j, j \in \beta_i)$  for all  $i \in \mathcal{N}$ . Furthermore,  $h$  is called **weakly Gaussian** if there exists  $\delta > 0$  such that  $\delta h$  is Gaussian, and **almost Gaussian** if  $h$  is the limit of a sequence of weakly Gaussian functions.

It is straightforward to prove that the weakly Gaussian property is closed under addition. In other words, if  $h$  and  $g$  are weakly Gaussian, then their sum  $h + g$  is also weakly Gaussian. Furthermore, like information inequality for any continuous random variables, if an inequality

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha H(X_\alpha) \geq 0 \tag{62}$$

holds for all Gaussian random variables  $\{X_i, i \in \mathcal{N}\}$  [25], then it must be balanced. Therefore, in the following, we will only consider balanced information inequalities.

Let  $\{Y_j, j \in \mathcal{K}\}$  be a set of jointly Gaussian random variables with covariance matrix  $K$  which is a  $|\mathcal{K}| \times |\mathcal{K}|$  positive definite matrix. Suppose  $\mathcal{K}$  is partitioned into  $n$  disjoint nonempty subsets  $\beta_1, \dots, \beta_n$ . A very compelling property of a set of Gaussian random variable is that its entropy and the determinant of its covariance matrix is related by the following relation:

$$H(Y_i, i \in \beta) = \frac{1}{2} \log [(2\pi e)^{|\beta|} \det(K_\beta)] = |\beta| \frac{\log(2\pi e)}{2} + \frac{\log \det(K_\beta)}{2}. \tag{63}$$

where  $K_\beta$  be the principal submatrix of  $K$  by deleting rows and columns that are not indexed by  $\beta$ . Substitute (63) back into (62), the inequality (62) is satisfied by all Gaussian random variables  $\{X_i, i \in \mathcal{N}\}$  if and only if

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha \left( |\beta_\alpha| \frac{\log(2\pi e)}{2} + \frac{\log \det(K_{\beta_\alpha})}{2} \right) \geq 0 \tag{64}$$

where  $\beta_\alpha = \bigcup_{i \in \alpha} \beta_i$ . Since the inequality (62) is balanced,

$$\sum_{\alpha \subseteq \mathcal{N}: j \in \alpha} c_\alpha = 0 \tag{65}$$

for all  $j \in \mathcal{N}$ . On the other hand,

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha |\beta_\alpha| = \sum_{\alpha \subseteq \mathcal{N}} c_\alpha \sum_{j \in \alpha} |\beta_j| \tag{66}$$

$$= \sum_{j \in \mathcal{N}} \sum_{\alpha \subseteq \mathcal{N}: j \in \alpha} c_\alpha |\beta_j| \tag{67}$$

$$= \sum_{j \in \mathcal{N}} |\beta_j| \sum_{\alpha \subseteq \mathcal{N}: j \in \alpha} c_\alpha \tag{68}$$

$$= 0. \tag{69}$$

Therefore, the inequality (62) holds for all Gaussian random variables if and only if the following determinantal inequality holds for all positive definite matrix  $K$

$$\sum_{\alpha \subseteq \mathcal{N}} c_\alpha \log \det(K_{\beta_\alpha}) \geq 0 \tag{70}$$

or equivalently,

$$\prod_{\alpha \subseteq \mathcal{N}} \det(K_{\beta_\alpha})^{c_\alpha} \geq 1. \tag{71}$$

As a direct consequence, for any valid information inequality, we can use the above relation to derive a corresponding determinantal inequality. For example, the following well-known determinantal inequalities can all be proved using this “information-theoretical method”.

1. **(Hadamard inequality)** Let  $K$  be a positive definite matrix  $K$ . Then

$$\det K \leq \prod_{i=1}^{|\mathcal{K}|} K_{i,i} \tag{72}$$

where  $K_{i,i}$  is the  $i^{th}$  diagonal entry of  $K$ . This inequality follows from the following information inequality

$$H(Y_1, \dots, Y_k) \leq \sum_{i=1}^k H(Y_i).$$

2. (**Szasz inequality**) For any  $1 \leq l < k$ ,

$$\left( \prod_{\beta:|\beta|=l} \det(K_\beta) \right)^{1/\binom{k-1}{l-1}} \geq \left( \prod_{\beta:|\beta|=l+1} \det(K_\beta) \right)^{1/\binom{k-1}{l}}. \tag{73}$$

This determinantal inequality follows from the following information inequality

$$\frac{1}{\binom{k}{l}} \sum_{\beta:|\beta|=l} \frac{H(Y_i, i \in \beta)}{l} \geq \frac{1}{\binom{k}{l+1}} \sum_{\beta:|\beta|=l+1} \frac{H(Y_i, i \in \beta)}{l+1}. \tag{74}$$

Finally, we will conclude this section by the following open question: While Gaussian polymatroid is clearly almost entropic, is it true that an almost entropic polymatroid almost Gaussian? In other words, for any almost entropic polymatroid  $h$ , can we construct a sequence of Gaussian polymatroids  $\{g_i, i = 1, \dots\}$  such that

$$\lim_{i \rightarrow \infty} \delta_i g_i = h$$

for some  $\delta_i > 0$  for all  $i$ .

## 6. Summary and Conclusions

In this paper, we have reviewed some of the recent progresses in characterisation of information inequalities. We first began with a geometric framework for information inequalities which has simplified the understanding of information inequalities. We also reviewed how the first non-Shannon inequality was proved and highlighted the general idea behind the proof. Next, we studied the infinite series of inequalities over  $\mathcal{N}_4$  and considered a nonlinear relaxation of the series of inequalities.

We have also reviewed how information inequalities are related to Kolmogorov complexity inequalities, group-theoretic inequalities and inequalities for box assignments. Based on their relations, we demonstrated non-traditional approaches to proving information inequalities.

Finally, we investigated two constrained classes of information inequalities. The first class is when random variables are induced by vector spaces. In this case, the constrained inequalities are equivalent to subspace rank inequalities. We showed that Ingleton and DFZ inequalities are insufficient to characterise all subspace rank inequalities in general where the set of all subspace rank inequalities is not truncation-preserving. The second constrained class of inequalities is when random variables are Gaussian. We have showed that these constrained inequalities are in fact determinantal inequalities.

As a final remark, we would like to emphasise that this survey paper aims not to cover every aspect about information inequalities. In fact, there are many interesting pieces of work that we did not cover. For example, as pointed out by one of the reviewers, one very interesting area is about the relation between convex body inequalities and information inequalities [26,27]. We strongly encourage readers who are interested to further explore those relevant areas.

## Acknowledgements

This work was supported by the Australian Government under ARC grant DP1094571.

## References and Notes

1. Zhang, Z.; Yeung, R.W. On the characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory* **1998**, *44*, 1440–1452.
2. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.
3. Hammer, D.; Romashchenko, A.; Shen, A.; Vereshchagin, N. Inequalities for Shannon Entropy and Kolmogorov Complexity. *J. Computer Syst. Sci.* **2000**, *60*, 442–464.
4. Chan, T.H.; Yeung, R.W. On a relation between information inequalities and group theory. *IEEE Trans. Inform. Theory* **2002**, *48*, 1992–1995.
5. Chan, T.H.; Grant, A.; Kern, D. Novel technique in characterising representable polymatroids. *IEEE Trans. Inform. Theory* **2009**, submitted for publication.
6. Chan, T.H. Balanced information inequalities. *IEEE Trans. Inform. Theory* **2003**, *49*, 3261–3267.
7. Chan, T.H. A combinatorial approach to information inequalities. *Commun. Inform. Syst.* **2001**, *1*, 1–14.
8. Dougherty, R.; Freiling, C.; Zeger, K. Six New Non-Shannon Information Inequalities. *IEEE Int. Symp. Inform. Theory* **2006**, July, 233–236.
9. Matus, F. Infinitely Many Information Inequalities. In Proceedings of ISIT 2007, Nice, France, June 2007.
10. Yeung, R. A framework for linear information inequalities. *IEEE Trans. Inform. Theory* **1997**, *43*, 1924–1934.
11. Yeung, R.; Yan Y. Information Theoretic Inequality Prover. Available online: <http://user-www.ie.cuhk.edu.hk/ITIP/> (accessed on 27 January 2011)
12. Yeung, R. *A First Course in Information Theory*; Kluwer Academic/Plenum Publisher: New York, NY, USA, 2002.
13. Yeung, R.W.; Zhang, Z. A class of non-Shannon-type information inequalities and their applications. *Commun. Inform. Syst.* **2001**, *1*, 87–100.
14. Sason, I. Identification of new classes of non-Shannon type constrained information inequalities and their relation to finite groups. In Proceedings of 2002 IEEE International Symposium, Lausanne, Switzerland, 30 June–5 July 2002.
15. Makarychev, K.; Makarychev, Y.; Romashchenko, A.; Vereshchagin, N. A new class of non-Shannon-type inequalities for entropies. *Commun. Inform. Syst.* **2002**, *2*, 147–165.
16. Chan, T.H.; Grant, A. Dualities between Entropy Functions and network codes. *IEEE Trans. Inform. Theory* **2008**, *54*, 4470–4487.
17. Chan, T.; Grant, A. Non-linear Information Inequalities. *Entropy J.* **2008**, *10*, 765–775.
18. Strictly speaking, the Kolmogorov complexity of a string depends on the chosen “computer model”. However, the choice of the computer model will only affect the resulting Kolmogorov up to a constant difference (because different computer models can emulate each other). Asymptotically, such a difference will not cause a significant difference.
19. Chan, T.H.; Grant, A. Linear programming bounds for network coding. *IEEE Trans. Inform. Theory* **2011**, submitted to be published.

20. Chan, T.H.; Grant, A.; Britz, T. Properties of quasi-uniform codes. In Proceedings of 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, June 2010.
21. Ingleton inequalities are not valid information inequalities, as there exists almost entropic polymatroids violating the inequalities.
22. Guille, L.; Chan, T.H.; Grant, A. The minimal set of Ingleton inequalities. *IEEE Trans. Inform. Theory* **2009**, arXiv:0802.2574.
23. Kinser, R. New inequalities for subspace arrangements. *J. Combin. Theory Ser. A* **2010**, 10.1016/j.jcta.2009.10.014.
24. Dougherty, R.; Freiling, C.; Zeger, K. Linear rank inequalities on five or more variables. *Arxiv Preprint* **2009**, cs.IT/0910.0284v3
25. Each  $X_i$  can be a vector of jointly distributed Gaussian random variables as defined in (61).
26. Lutwak, E.; Yang, D.; Zhang, G. Cramer-Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information. *IEEE Trans. Info. Theory* **2005**, 51, 473–478.
27. Lutwak, E.; Yang, D.; Zhang, G. Moment-entropy inequalities for a random vector. *IEEE Trans. Info. Theory* **2007**, 53, 1603–1607.

© 2011 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license <http://creativecommons.org/licenses/by/3.0/>.