A Relationship Between Information Inequalities and Group Theory

Desmond S. Lun

23 October 2002

1 Outline

This report presents the result recently published in [1] that establishes a one-to-one correspondence between information inequalities and group inequalities. Our aim in this report is to present this result in as concise a manner as possible whilst not excluding any steps in the derivation, thus making it suitable for brief perusal. In Sections 2–4, we introduce the notions of entropy functions, group-characterizable functions, information inequalities, and group inequalities. We confine most of the technical details to Section 5 — a section that may we skipped if one is interested only in the general idea of the method in which the result is demonstrated. We complete the demonstration in Section 6 and conclude in Section 7.

2 Entropy functions

Let $X_1, X_2, \ldots, X_n$ be n discrete random variables whose sample spaces are $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ respectively. Let $\mathcal{N} = \{1, \ldots, n\}$ and Ω be the collection of all non-empty subsets of $\mathcal{N}$, so $|\Omega| = 2^n - 1$. For $\alpha \in \Omega$, we define X_α to be the joint random variable $(X_i)_{i \in \alpha}$ with sample space $\mathcal{X}_\alpha = \prod_{i \in \alpha} \mathcal{X}_i$ (i.e. the Cartesian product of $\mathcal{X}_i$ for $i \in \alpha$). For example, suppose $\alpha = \{1, 2\}$; then $X_{\{1,2\}}$ is the random pair (X_1, X_2).

For a given set of random variables $X_1, X_2, \ldots, X_n$, we can define a real function over Ω. Since $|\Omega|$ is finite, such a function is completely described by the set of values it takes on Ω, i.e. by a point in $\mathbb{R}^{|\Omega|}$. An entropy function, therefore, is defined a follows.

Definition 1 *Let* $\mathbf{g}$ *be a vector in* $\mathbb{R}^{|\Omega|}$ *with components* g_α *indexed by* $\alpha \in \Omega$. *Then* $\mathbf{g}$ *is an* entropy function *if there exists a set of random variables* $X_1, X_2, \ldots, X_n$ *such that* $g_\alpha = H(X_\alpha)$ *for all* α.

Let Γ_n^* be the set of all entropy functions associated with n random variables; that is

$$\Gamma_n^* = \{\mathbf{g} \in \mathbb{R}^{|\Omega|} \mid \mathbf{g} \text{ is an entropy function}\}. \tag{1}$$

3 Group-characterizable functions

We now define functions over Ω that can be characterized by a finite group G and subgroups $G_1, G_2, \dots, G_n$ in a particular way. Our ultimate aim will be to relate such group-characterizable functions with entropy functions. For a given $\alpha \in \Omega$, we denote the intersection of the subgroups G_i for $i \in \alpha$ by G_α, which is also a subgroup a G.

Definition 2 *Let* $\mathbf{h}$ *be a vector in* $\mathbb{R}^{|\Omega|}$ *with components* h_α *indexed by* $\alpha \in \Omega$*. Then* $\mathbf{h}$ *is called* group-characterizable *if there exist subgroups* $G_1, G_2, \dots, G_n$ *of a group* G *such that* $h_\alpha = \log(|G|/|G_\alpha|)$ *for all* α*.*

Let Υ_n be the set of all group-characterizable functions associated with n groups; that is

$$\Upsilon_n = \{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{h} \text{ is group-characterizable}\}. \tag{2}$$

4 Information inequalities and group inequalities

It was noted in [3] that an information inequality

$$\sum_{\alpha \in \Omega} b_\alpha H(X_\alpha) \geq 0 \tag{3}$$

is valid if and only if

$$\Gamma_n^* \subset \{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{b}^\top \mathbf{h} \geq 0\}, \tag{4}$$

where $\mathbf{b}$ is the column vector with components b_α indexed by $\alpha \in \Omega$. This observation, which follows immediately from the definition of Γ_n^*, highlights the importance of characterizing Γ_n^* to the study of information inequalities, or equivalently its closure $\overline{\Gamma}_n^*$, since because $\{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{b}^\top \mathbf{h} \geq 0\}$ is closed, the condition (4) is equivalent to

$$\overline{\Gamma}_n^* \subset \{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{b}^\top \mathbf{h} \geq 0\}. \tag{5}$$

The characterization Γ_n^* or its closure appears to be a highly non-trivial problem in general [6].

Likewise, a group inequality of the form

$$\sum_{\alpha \in \Omega} b_\alpha \log \frac{|G|}{|G_\alpha|} \geq 0 \tag{6}$$

is valid if and only if

$$\Upsilon_n \subset \{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{b}^\top \mathbf{h} \geq 0\}. \tag{7}$$

Thus, we are interested in the relationship between Υ_n and Γ_n^*. We observe that all the function values of group-characterizable functions are in one-to-one correspondence with rational numbers, so there are at most countably many group-characterizable functions. There are, however, uncountably many entropy functions. Hence, there are far fewer group-characterizable functions than there are entropy functions. It turns out, however, that Υ_n is almost good enough to characterize Γ_n^*, and in fact $\overline{\text{conv}}(\Upsilon_n)$, the convex closure of Υ_n, is equal to $\overline{\Gamma}_n^*$, the closure of Γ_n^*, as we shall establish in the next section.

5 The equivalence of $\overline{\text{conv}}(\Upsilon_n)$ and $\overline{\Gamma}_n^*$

Theorem 1 $\overline{\text{conv}}(\Upsilon_n) = \overline{\Gamma}_n^*$ *for all natural numbers n.*

To establish this theorem, we require a number of lemmas.

Lemma 1 *[1, Theorem 3.1] If $\mathbf{h}$ is group-characterizable, then it is an entropy function, i.e. $\mathbf{h} \in \Gamma_n^*$.*

Proof. Let Λ be a uniformly-distributed discrete random variable with sample space G. For any $i \in \mathcal{N}$, let random variable X_i be a function of Λ such that X_i is the left coset aG_i if $\Lambda = a$. Let α be an element of Ω. Then

$$P((X_i = a_i G_i)_{i\in\alpha}) = P(\Lambda \in \cap_{i\in\alpha} a_i G_i) = \frac{|\cap_{i\in\alpha} a_i G_i|}{|G|}. \tag{8}$$

Consider $\cap_{i\in\alpha} a_i G_i$. If it is non-empty, then let b be an element of $\cap_{i\in\alpha} a_i G_i$, and we have

$$\begin{aligned} \cap_{i\in\alpha} a_i G_i &= \cap_{i\in\alpha} b G_i \\ &= b \cap_{i\in\alpha} G_i = bG_\alpha. \end{aligned} \tag{9}$$

Therefore, the intersection $\cap_{i\in\alpha} a_i G_i$ is either empty or of size $|G_\alpha|$. Moreover, by Lagrange's theorem, there are $|G|/|G_\alpha|$ distinct such non-empty intersections. Hence, we have

$$P((X_i = a_i G_i)_{i\in\alpha}) = \begin{cases} \frac{|G_\alpha|}{|G|} & \text{if } \cap_{i\in\alpha} a_i G_i \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

And it is evident that the entropy of $(X_i)_{i\in\alpha}$ is $\log(|G|/|G_\alpha|) = h_\alpha$. Therefore, $\mathbf{h}$ is an entropy function. □

Lemma 2 *[5, Theorem 1] $\overline{\Gamma}_n^*$ is a convex cone.*

Proof. We shall first demonstrate that $\overline{\Gamma}_n^*$ is convex. Let $\mathbf{u}, \mathbf{v}$ be the entropy functions of two sets of random variables $Y_1, Y_2, \ldots, Y_n$ and $Z_1, Z_2, \ldots, Z_n$ respectively. Thus $\mathbf{u}, \mathbf{v} \in \Gamma_n^*$. It suffices to show that for any $0 < b < 1$,

$b\mathbf{u} + (1-b)\mathbf{v} \in \overline{\Gamma}_n^*$ (for it is then straightforward to extend the result to any convex combination of points in the closure of Γ_n^*).

Let $(Y_1^{(i)}, Y_2^{(i)}, \ldots, Y_n^{(i)})$ for $1 \leq i \leq k$ be k independent vectors each distributed identically to $(Y_1, Y_2, \ldots, Y_n)$. Likewise, let $(Z_1^{(i)}, Z_2^{(i)}, \ldots, Z_n^{(i)})$ for $1 \leq i \leq k$ be k independent vectors each distributed identically to $(Z_1, Z_2, \ldots, Z_n)$. Let U be a random variable independent of all other random variables having the distribution

$$p_U(u) = \begin{cases} 1-\delta-\mu & \text{if } u = 0, \\ \delta & \text{if } u = 1, \\ \mu & \text{if } u = 2. \end{cases} \tag{11}$$

Observe that $H(U) \to 0$ as $\delta, \mu \to 0$.

We now construct the random variables $X_1, X_2, \ldots, X_n$ by

$$X_i = \begin{cases} 0 & \text{if } U = 0, \\ (Y_i^{(1)}, Y_i^{(2)}, \ldots, Y_i^{(k)}) & \text{if } U = 1, \\ (Z_i^{(1)}, Z_i^{(2)}, \ldots, Z_i^{(k)}) & \text{if } U = 2. \end{cases} \tag{12}$$

So for any $\alpha \in \Omega$, we have

$$\begin{aligned} H(X_\alpha) \leq H(X_\alpha, U) &= H(U) + H(X_\alpha | U) \\ &= H(U) + \delta k H(Y_\alpha) + \mu k H(X_\alpha) \end{aligned} \tag{13}$$

and

$$H(X_\alpha) \geq H(X_\alpha | U) = H(U) + \delta k H(Y_\alpha) + \mu k H(X_\alpha). \tag{14}$$

Combining the two previous equations, we have

$$0 \leq H(X_\alpha) - (\delta k H(Y_\alpha) + \mu k H(Z_\alpha)) \leq H(U). \tag{15}$$

And by taking $\delta = b/k$ and $\mu = (1-b)/k$, we obtain

$$0 \leq H(X_\alpha) - (b H(Y_\alpha) + (1-b) H(Z_\alpha)) \leq H(U). \tag{16}$$

By taking k sufficiently large, the upper bound can be made arbitrarily small. Hence we have a sequence of points in Γ_n^* whose limit point is $b\mathbf{u} + (1-b)\mathbf{v}$. So $b\mathbf{u} + (1-b)\mathbf{v} \in \overline{\Gamma}_n^*$, concluding the demonstration of the convexity of $\overline{\Gamma}_n^*$.

We next observe that if $\mathbf{v} \in \overline{\Gamma}_n^*$ then for any positive integer k, $k\mathbf{v} \in \overline{\Gamma}_n^*$ (for any point in Γ_n^*, we simply consider k independent copies of its associated random variables, then straightforwardly extend the result to the closure). In addition, by letting $X_1, X_2, \ldots, X_n$ be random variables taking constant values with probability 1, we see that Γ_n^* contains the origin.

Now consider a non-negative combination of points $\{\mathbf{v}_i\}$ in $\overline{\Gamma}_n^*$, $\sum_i \alpha_i \mathbf{v}_i$, where $\alpha_i \geq 0$ for all i. Let $\alpha = \sum_i \alpha_i$. By the convexity of $\overline{\Gamma}_n^*$,

$$\sum_i \frac{\alpha_i}{\lceil \alpha \rceil} \mathbf{v}_i + \left(1 - \frac{\alpha}{\lceil \alpha \rceil}\right) \mathbf{0} = \sum_i \frac{\alpha_i}{\lceil \alpha \rceil} \mathbf{v}_i$$

is an element of $\overline{\Gamma}^*_n$. Multiplying this vector by $\lceil \alpha \rceil$, we see that $\sum_i \alpha_i \mathbf{v}_i \in \overline{\Gamma}^*_n$, thus completing the proof. □

Lemma 3 *[1, Theorem 4.1] For any* $\mathbf{h} \in \Gamma^*_n$*, there exists a sequence* $\{\mathbf{f}^{(r)}\}$ *in* Υ_n *such that* $\lim_{r\to\infty}(\mathbf{f}^{(r)}/r) = \mathbf{h}$.

Proof. For any $\mathbf{h} \in \Gamma^*_n$, there exists a collection of random variables $X_1, X_2, \ldots, X_n$ such that $h_\alpha = H(X_\alpha)$ for all $\alpha \in \Omega$. We first consider the special case where $|\mathcal{X}_i| < \infty$ for all $i \in \mathcal{N}$ and the joint distribution $X_1, X_2, \ldots, X_n$ is rational. We shall construct a sequence $\{\mathbf{f}^{(r)}\}$ in Υ_n such that $\lim_{r\to\infty}(\mathbf{f}^{(r)}/r) = \mathbf{h}$.

For any $\alpha \in \Omega$, let Q_α be the marginal distribution of X_α. We assume without loss of generality that for any $\alpha \in \Omega$ and for all $\mathbf{x} \in \mathcal{X}_\alpha$, $Q_\alpha(\mathbf{x})$ is a rational number with denominator q.

For each $r = q, 2q, 3q, \ldots$, we fix an $n \times r$ matrix A

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,r} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,r} \end{bmatrix} \tag{17}$$

such that for all $\mathbf{x} \in \mathcal{X}_\mathcal{N}$ the number of columns in A that are equal to $\mathbf{x}$ is $rQ_\mathcal{N}(\mathbf{x})$. The existence of such a matrix is guaranteed because the joint distribution of $\mathcal{X}_\mathcal{N}$ is rational with denominator q.

For any $\alpha \in \Omega$, we denote by A_α the submatrix of A obtained by extracting the rows of A indexed by α. We observe that, for all $\mathbf{x} \in \mathcal{X}_\alpha$, the number of columns of A_α that are equal to the $\mathbf{x}$ is $rQ_\alpha(\mathbf{x})$, where Q_α is the marginal distribution of the random variables X_i for $i \in \alpha$.

Let G be the group of permutation on $\{1, \ldots, r\}$. (Although the group G depends on r, this dependency is not explicitly stated for the sake of simplicity.) For any $i \in \mathcal{N}$, we define

$$G_i = \{\sigma \in G \,|\, \sigma[A_i] = A_i\}, \tag{18}$$

that is those permutations of r elements that keep the ith row fixed. It is straightforward to check that G_i is a subgroup of G.

Let $\alpha \in \Omega$. Then

$$\begin{aligned} G_\alpha &= \cap_{i\in\alpha} G_i \\ &= \cap_{i\in\alpha}\{\sigma \in G \,|\, \sigma[A_i] = A_i\} \\ &= \{\sigma \in G \,|\, \sigma[A_i] = A_i \text{ for all } i \in \alpha\} \\ &= \{\sigma \in G \,|\, \sigma[A_\alpha] = A_\alpha\}. \end{aligned} \tag{19}$$

Since for all $\mathbf{x} \in \mathcal{X}_\alpha$ the number of columns in A_α that are equal to $\mathbf{x}$ is $rQ_\alpha(\mathbf{x})$, it follows that

$$|G_\alpha| = \prod_{\mathbf{x}\in\mathcal{X}_\alpha} (rQ_\alpha(\mathbf{x}))! \tag{20}$$

and hence

$$\frac{|G|}{|G_\alpha|} = \frac{r!}{\prod_{\mathbf{x}\in\mathcal{X}_\alpha}(rQ_\alpha(\mathbf{x}))!}. \tag{21}$$

Using Theorem 12.1.3 from [2], we obtain

$$\lim_{r\to\infty}\frac{1}{r}\log\frac{|G|}{|G_\alpha|} = H(X_\alpha) = h_\alpha. \tag{22}$$

We define $\mathbf{f}^{(r)}$ by

$$f_\alpha^{(r)} = \log\frac{|G|}{|G_\alpha|} \tag{23}$$

for all $\alpha\in\Omega$. Then $\mathbf{f}^{(r)}\in\Upsilon_n$ by construction, and

$$\lim_{r\to\infty}\frac{1}{r}\mathbf{f}^{(r)} = \mathbf{h}. \tag{24}$$

Thus the theorem is proved for the special case of entropy functions of random variables with finite sample space and rational joint distribution. In general, for any $\mathbf{h}\in\Gamma_n^*$, we can construct a sequence $\{\mathbf{h}^{(k)}\}$ in Γ_n^* that converges to $\mathbf{h}$ where, for all k, $\mathbf{h}^{(k)}$ is the entropy function of a collection of random variables with finite sample space and rational joint distribution. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. We have established that $\Upsilon_n\subset\Gamma_n^*$ with Lemma 1 and that $\overline{\Gamma}_n^*$ is convex with Lemma 2, hence

$$\overline{\mathrm{conv}}(\Upsilon_n)\subset\overline{\mathrm{conv}}(\Gamma_n^*)\subset\overline{\mathrm{conv}}(\overline{\Gamma}_n^*) = \overline{\Gamma}_n^*. \tag{25}$$

By taking any finite group G and subgroups $G_1, G_2, \ldots, G_n$ such that $G = G_1 = G_2 = \cdots = G_n$, we see that Υ_n contains the origin. Therefore, it follows from Lemma 3 that $\overline{\Gamma}_n^*\subset\overline{\mathrm{conv}}(\Upsilon_n)$. We thus conclude that $\overline{\Gamma}_n^* = \overline{\mathrm{conv}}(\Upsilon_n)$. □

6 Correspondence between information inequalities and group inequalities

Recall from Section 4 that an information inequality

$$\sum_{\alpha\in\Omega} b_\alpha H(X_\alpha)\geq 0 \tag{26}$$

is valid if and only if

$$\Gamma_n^*\subset\{\mathbf{h}\in\mathbb{R}^{|\Omega|}\mid\mathbf{b}^\top\mathbf{h}\geq 0\}, \tag{27}$$

and the group inequality

$$\sum_{\alpha \in \Omega} b_\alpha \log \frac{|G|}{|G_\alpha|} \geq 0 \tag{28}$$

is valid if and only if

$$\Upsilon_n \subset \{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{b}^\top \mathbf{h} \geq 0\}. \tag{29}$$

Suppose (27) is satisfied. Then since $\Upsilon_n \subset \Gamma_n^*$, (29) is satisfied. Conversely, if (29) is satisfied, then since $\{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{b}^\top \mathbf{h} \geq 0\}$ is closed and convex, we have

$$\Gamma_n^* \subset \overline{\Gamma}_n^* = \overline{\text{conv}}(\Upsilon_n) \subset \{\mathbf{h} \in \mathbb{R}^{|\Omega|} \,|\, \mathbf{b}^\top \mathbf{h} \geq 0\}. \tag{30}$$

So (27) and (29) are equivalent, and therefore (26) and (28) are equivalent, that is, for every information inequality of the form of (26), there is a corresponding group inequality of the form of (28) and vice versa.

7 Conclusion

We have reviewed a demonstration that establishes a one-to-one correspondence between information inequalities and group inequalities. In essence, the demonstration comes down to showing that the set of points in $\mathbb{R}^{|\Omega|}$ that are group-characterizable, Υ_n, whilst sparser than the set of points that are entropy functions, Γ_n^*, can characterize $\overline{\Gamma}_n^*$ with its convex closure, which is a sufficiently strong characterization if one is interested in the space of linear non-strict inequalities.

Though it is unclear what the utility of this result is or whether or not it has any 'deeper meaning', it is nevertheless interesting, and contributes another method to proving information inequalities or group inequalities at the very least.

The reader who wishes to learn more about information inequalities and their relationship with group inequalities, or in general, is referred [4].

References

[1] Terence H. Chan and Raymond W. Yeung. On a relation between information inequalities and group theory. *IEEE Trans. Inform. Theory*, 48(7):1992–1995, July 2002.

[2] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., New York, NY, 1991.

[3] Raymond W. Yeung. A framework for linear information inequalities. *IEEE Trans. Inform. Theory*, 43(6):1924–1934, November 1997.

[4] Raymond W. Yeung. *A First Course in Information Theory.* Kluwer Academic/Plenum, 2002.

[5] Zhen Zhang and Raymond W. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Inform. Theory*, 43(6):1982–1986, November 1997.

[6] Zhen Zhang and Raymond W. Yeung. On characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory*, 44(4):1440–1452, July 1998.