Shannon Inequalities in Distributed Storage - Part I

Birenjith Sasidharan and P. Vijay Kumar
(joint work with Myna Vajha and Kaushik Senthoor)

Department of Electrical Communication Engineering,
Indian Institute of Science, Bangalore

Workshop on Advanced Information Theory
Commemorating the 100th Birthday of Claude Shannon

Organized by the IISc-IEEE ComSoc Student Chapter
(in association with IEEE Bangalore Section & ECE Department, IISc)

Indian Institute of Science, April 30, 2016

Outline

References

- Raymond Yeung, "Facets of Entropy' http://www.inc.cuhk.edu.hk/EII2013/entropy.pdf

Outline

Entropy of a Random Variable

Let X be a discrete random variable taking on values[1] from an alphabet $\mathcal{X}$. Then

$$\begin{aligned} H(X) &:= -\sum_{x \in \mathcal{X}} p(x) \log p(x) \\ &= -\sum_{x} p(x) \log p(x) \text{ for simplicity.} \end{aligned}$$

Clearly

$$H(X) \geq 0.$$

[1] All random variables encountered here will take on values from a finite alphabet.

Joint and Conditional Entropy of 2 Random Variables

- If X, Y are a pair of discrete random variables we define their joint entropy via:

$$H(X,Y) \;:=\; -\sum_{x,y} p(x,y)\log p(x,y).$$

- We define the conditional entropy of X given Y via:

$$\begin{aligned} H(X/Y) \;&:=\; \sum_{y} p(y)\left\{-\sum_{x} p(x/y)\log p(x/y)\right\} \\ &=\; -\sum_{x,y} p(x,y)\log p(x/y). \end{aligned}$$

- Clearly

$$\begin{aligned} H(X,Y) \;&\geq\; 0 \\ H(X/Y) \;&\geq\; 0. \end{aligned}$$

Joint and Conditional Entropy of 2 RV (continued)

Morover

$$\begin{aligned} H(X,Y) &:= -\sum_{x,y} p(x,y)\log p(x,y) \\ &= -\sum_{x,y} p(x,y)\log p(y) \; - \; \sum_{x,y} p(x,y)\log p(x/y) \\ &= H(Y) + H(X/Y). \end{aligned}$$

Mutual Information

The mutual information between X and Y is given by:

$$\begin{aligned} I(X;Y) &:= H(X) - H(X/Y) \\ &= -\sum_{x,y} p(x,y) \log \frac{p(x)}{p(x/y)} \\ &= -\sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x,y)} \\ &= H(X) + H(Y) - H(X,Y). \end{aligned}$$

Mutual Information is Non-Negative

Since the function $-\log(\cdot)$ is convex, we have that :

$$\begin{aligned} I(X;Y) &:= H(X) - H(X/Y) \\ &= \sum_{x,y} p(x,y)\{-\log \frac{p(x)p(y)}{p(x,y)}\} \\ &\geq -\log\left(\sum_{x,y} p(x,y)\frac{p(x)p(y)}{p(x,y)}\right) \\ &= -\log\left(\sum_{x,y} p(x)p(y)\right) \\ &= 0. \end{aligned}$$

Conditional Mutual Information

The conditional mutual information between X and Y, conditioned on Z, is given by:

$$\begin{aligned} I(X;Y/Z) &:= H(X/Z) - H(X/Y,Z) \\ &= -\sum_z p(z) \left\{ \sum_{x,y} p(x,y/z) \log \frac{p(x/z)}{p(x/y,z)} \right\} \\ &= -\sum_z p(z) \left\{ \sum_{x,y} p(x,y/z) \log \frac{p(x/z)p(y/z)}{p(x,y/z)} \right\} \\ &= -\sum_{x,y} p(x,y,z) \log \frac{p(x/z)p(y/z)}{p(x,y/z)}. \end{aligned}$$

Clearly

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

Outline

Sets of Random Variables

Let

$$\begin{aligned} [n] &= \{1, 2, \cdots, n\} \\ X &= \{X_1, X_2, \cdots, X_n\} \end{aligned}$$

Then if

$$\begin{aligned} A = \{i_1, i_2, \cdots, i_\ell\} &\subseteq [n] \\ \text{we set } X_A = \{x_{i_1}, x_{i_2}, \cdots, x_{i_\ell}\} &\subseteq X. \end{aligned}$$

Basic Inequalities

Can show just as we have shown above that

$$\begin{aligned} H(X_A) &\geq 0 \\ H(X_A/X_B) &\geq 0 \\ I(X_A/X_B) &\geq 0 \\ I(X_A; X_B/X_c) &\geq 0. \end{aligned}$$

Note that

$$\begin{aligned} H(X_A/X_B) &= H(X_{A\cap B}, X_{A\setminus B}/X_B) \\ &= H(X_{A\setminus B}/X_B), \end{aligned}$$

etc.

Polymatroidal Axioms

These state the following:

$$\begin{aligned} H(X_A) &\geq 0 \\ H(X_A) &\leq H(X_B) \quad \text{if } A \subseteq B \\ H(X_A) + H(X_B) &\geq H(X_{A\cup B}) + H(X_{A\cap B}). \end{aligned}$$

The first statement, we have already established.

Proof of Monotonicity

To see that

$$H(X_A) \leq H(X_B) \quad \text{if } A \subseteq B \ ,$$

note that

$$\begin{aligned} H(X_B) - H(X_A) &= H(X_{B\setminus A}, X_A) - H(X_A) \\ &= H(X_{B\setminus A}/X_A) \\ &\geq 0. \end{aligned}$$

Proof of Submodularity

To see that

$$H(X_A) + H(X_B) \quad \geq \quad H(X_{A\cup B}) + H(X_{A\cap B}),$$

note that

$$\begin{aligned} H(X_A) + H(X_B) - H(X_{A\cup B}) - H(X_{A\cap B}) &\geq 0 \\ \Leftrightarrow \quad (H(X_A) - H(X_{A\cap B})) - (H(X_{A\cup B}) - H(X_B)) &\geq 0 \\ \Leftrightarrow \quad H(X_{A\setminus B}/X_{A\cap B}) - H(X_{A\setminus B}/X_B) &\geq 0 \\ \Leftrightarrow \quad I(X_{A\setminus B}; X_{B\setminus A}/X_{A\cap B}) &\geq \quad 0 \end{aligned}$$

which we know to be true.
In turns out that the basic inequalities and the polymatroidal axioms can be derived from each other.

Outline

Entropic Vectors

Let $n = 3$. Then

$$X \quad = \quad \{X_1, X_2, X_3\}$$

The vector

$$h(X) :=$$

$$(H(X_1),\ H(X_2), H(X_3), H(X_1, X_2), H(X_1, X_3), H(X_2, X_3), H(X_1, X_2, X_3))$$

is called the entropy vector associated to X. Note that in general

$$h(\cdot) : X \to \Re^{2^n - 1}.$$

The Entropic Vector Region

- A vector is called entropic if it is the entropy vector associated to some set of RV X
- The entropic vector region Γ_n^* is the region containing precisely the set of all entropic vectors $h(X)$ for all possible X.
- Both the basic inequalities (or equivalently, the polymatroidal axioms) can be re-expressed in terms of joint entropies
- Thus every entropic vector must satisfy the basic inequalities

- Does an entropic vector necessarily have to satisfy any other inequalities ?

Aside: Inequalities that "Always Hold"

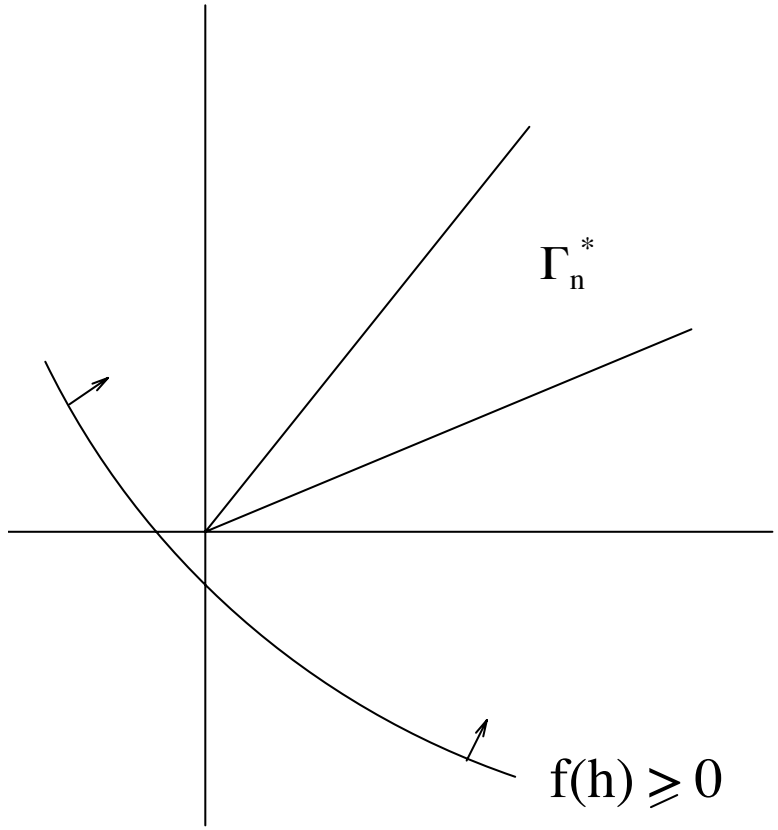

Figure taken from:

- Raymond Yeung, "Facets of Entropy' http://www.inc.cuhk.edu.hk/EII2013/entropy.pdf

Aside: Inequalities that do not "Always Hold"

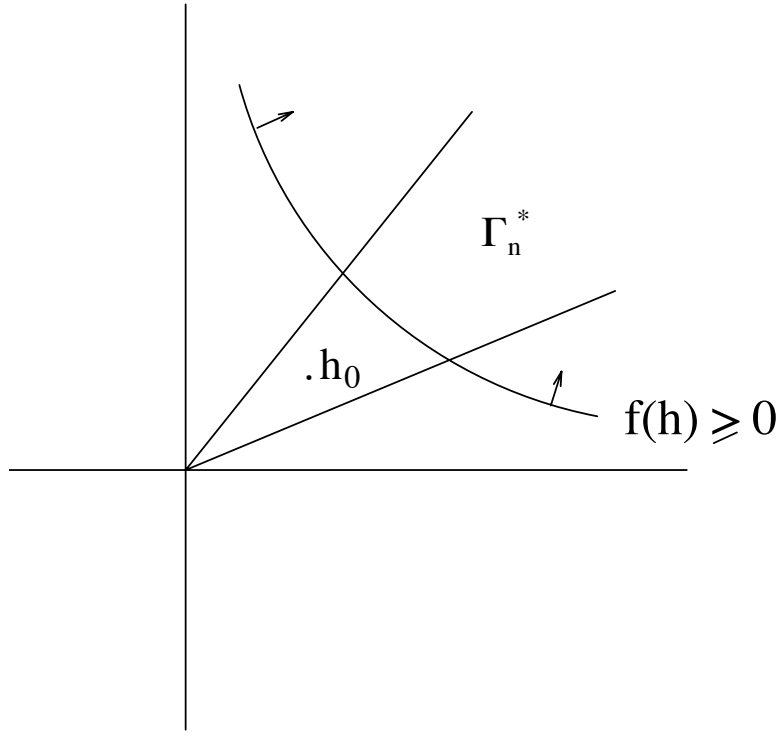

Figure taken from:

- Raymond Yeung, "Facets of Entropy' http://www.inc.cuhk.edu.hk/EII2013/entropy.pdf

The Region Γ_n

Let Γ_n denote the region in $\Re^{2^n-1}$ of vectors that satisfy the basic inequalities. The big question, is:

$$\Gamma_n^* = \Gamma_n?$$

Turns out that:

$$\begin{aligned} \Gamma_2^* &= \Gamma_2 \\ \Gamma_3^* &\neq \Gamma_3 \\ \text{but } \bar{\Gamma}_3^* &= \Gamma_3 \end{aligned}$$

In general,

$$\Gamma_n^* \neq \Gamma_n$$

The Zhang-Yeung Counter Example when $n = 4$

Turns out [2]

$$2I(X_3; X_4) \leq$$
$$I(X_1; X_2) + I(X_1; X_3, X_4) + 3I(X_3; X_4/X_1) + I(X_3; X_4/X_2),$$

is an example of an inequality that is satisfied by all vectors in Γ_4^*, but that this inequality is not derivable from the inequalities defining Γ_4.

[2] Zhang and Yeung, "On characterization of entropy function via information inequalities," T-IT, 1998.

The Consequent Picture for $n = 4$

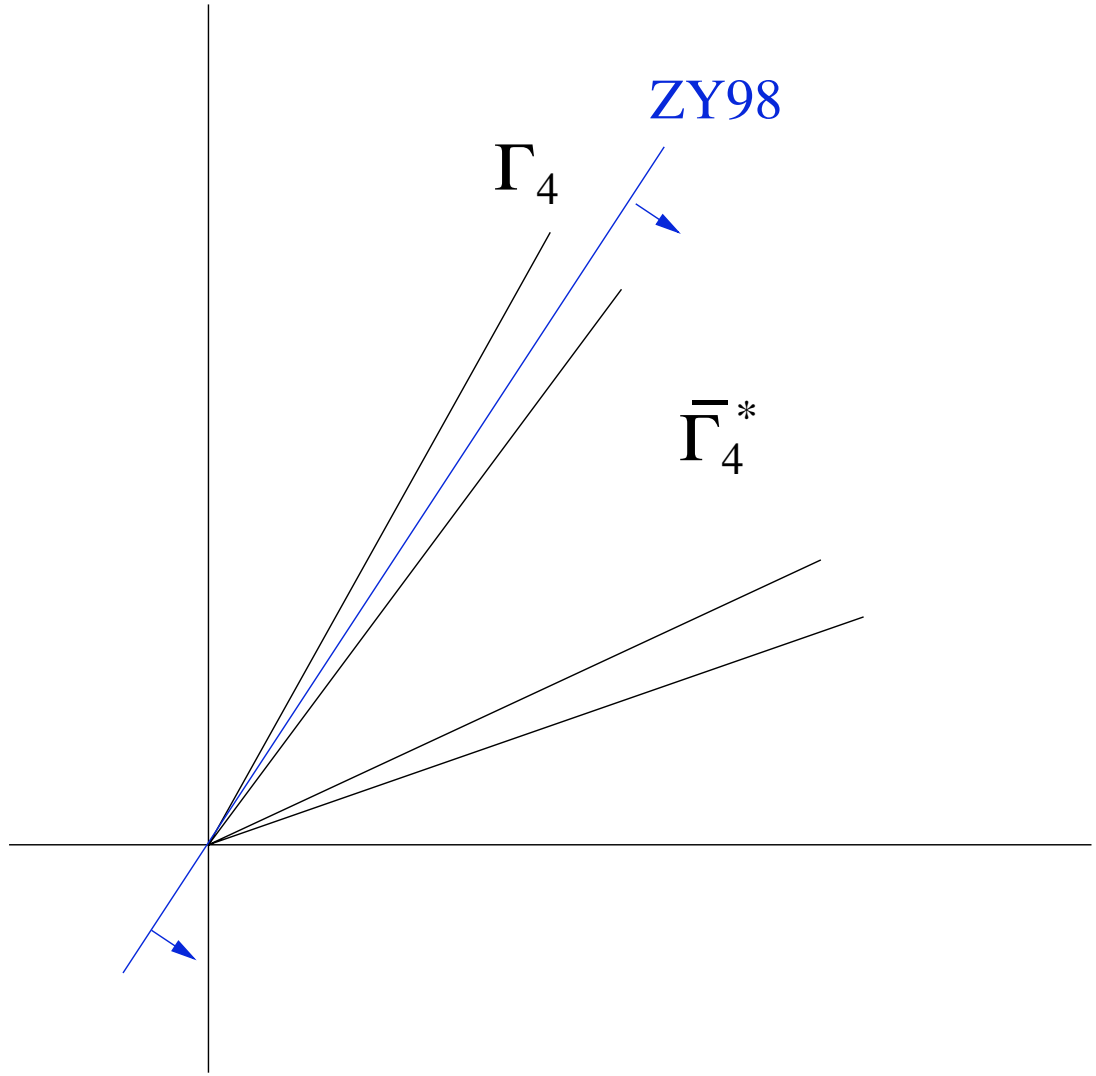

Figure taken from:

- Raymond Yeung, "Facets of Entropy' http://www.inc.cuhk.edu.hk/EII2013/entropy.pdf

Pipenger (1986)

What Are the Laws of Information Theory?

Nicholas Pippenger
IBM Almaden Research Laboratory K51-801
650 Harry Road
San Jose, California 95120-6099

Shannon defined the *entropy* $H(X)$ of a random variable X assuming values in a finite set $\mathcal{X}$ to be $-\sum_{x\in\mathcal{X}} \Pr(X=x)\log\Pr(X=x)$. The entropy $H(X,Y,Z)$ of a finite set $\{X,Y,Z\}$ of random variables is defined by regarding the tuple (X,Y,Z) as a single random variable. In information theory, one also deals with *conditional entropies*, like $H(X \mid Y) = H(X,Y) - H(Y)$; *mutual informations*, like $I(X;Y) = H(X) + H(Y) - H(X,Y)$; and *conditional mutual informations*, like $I(X;Y \mid Z) = H(X,Y) + H(X,Z) - H(X,Y,Z) - H(Z)$. All identities and inequalities concerning these quantities, however, can be reduced to ones involving only "plain" entropies, like $H(X,Y,Z)$, by invoking these definitions. The identities are known (see [H] and [R]). The problem posed here is to determine the inequalities.

If $\{X_t\}_{t\in T}$ is a family of random variables, and if $S \subseteq T$, let H_S denote the entropy of the subfamily $\{X_s\}_{s\in S}$. The resulting map $H : 2^T \to \mathbf{R}$ satisfies the following conditions (known as the *polymatroid axioms*).

(1) $H_S \geq 0$ and $H_\emptyset = 0$.
(2) $H_R \leq H_S$ if $R \subseteq S$.
(3) $H_{R\cup S} + H_{R\cap S} \leq H_R + H_S$.

Raymond Yeung

Zhen Zhang and Others

Zhen Zhang
University of Southern California

Terence Chan
University of South Australia

Imre Csiszár
Hungarian Academy of Sciences

Outline

Polyhedra

- a closed half space:

$$H^+ \quad = \quad \{x \in V \mid \lambda(x) + a_0 \geq 0\}.$$

- a subset $P \subseteq V$ is called a polyhedron if it is the intersection of finitely many closed half spaces.
- a polytope is a bounded polyhedron
- a subset $P \subseteq V$ is called a cone if it is the intersection of finitely many linear closed half spaces (i.e., half spaces where $a_0 = 0$)

Other Results

- $h_Z = h_X + h_Y \in \Gamma_n^*$; proof use $Z_i = (X_i, Y_i)$, X, Y independent
- Hence for any integer k, $kh \in \Gamma_n^*$
- $0 \in \Gamma_n^*$ (use the constant random variable)
- Γ_n^* is convex by appropriate construction of random variables