How did we get into this mess?

Raymond W. Yeung
Institute of Network Coding
&
Department of Information Engineering
CUHK

1988

R. W. Yeung, “Some result on multiterminal source coding,” PhD thesis, Cornell University, 1988

1988

R. W. Yeung, “Some result on multiterminal source coding,” PhD thesis, Cornell University, 1988

DISSERTATION

SOME RESULTS ON MULTITERMINAL SOURCE CODING

RAYMOND WAI HO YEUNG

1988

1991

R. W. Yeung, "A new outlook on Shannon's information measures," IEEE-IT, 1991

Shannon’s Information Measures

Shannon's Information Measures

- Entropy (Shannon Entropy)

$$H(X) = -\sum_{x} p(x) \log p(x)$$

Shannon's Information Measures

- Entropy (Shannon Entropy)

$$H(X) = -\sum_{x} p(x) \log p(x)$$

- Joint Entropy

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

Shannon's Information Measures

- Entropy (Shannon Entropy)

$$H(X) = -\sum_{x} p(x) \log p(x)$$

- Joint Entropy

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

- In information theory, entropy is the measure of the uncertainty contained in a discrete random variable, justified by fundamental coding theorems.

The Basic Inequalities

- In addition to Entropy, we also have:

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

$$H(X|Y) = H(X,Y) - H(Y)$$

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

  $$H(X|Y) = H(X,Y) - H(Y)$$

  Mutual Information

  $$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

$$H(X|Y) = H(X,Y) - H(Y)$$

  Mutual Information

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

  Conditional Mutual Information

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z)$$

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

  $$H(X|Y) = H(X,Y) - H(Y)$$

  Mutual Information

  $$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

  Conditional Mutual Information

  $$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z)$$

- These are called Shannon's information measures.

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

  $$H(X|Y) = H(X,Y) - H(Y)$$

  Mutual Information

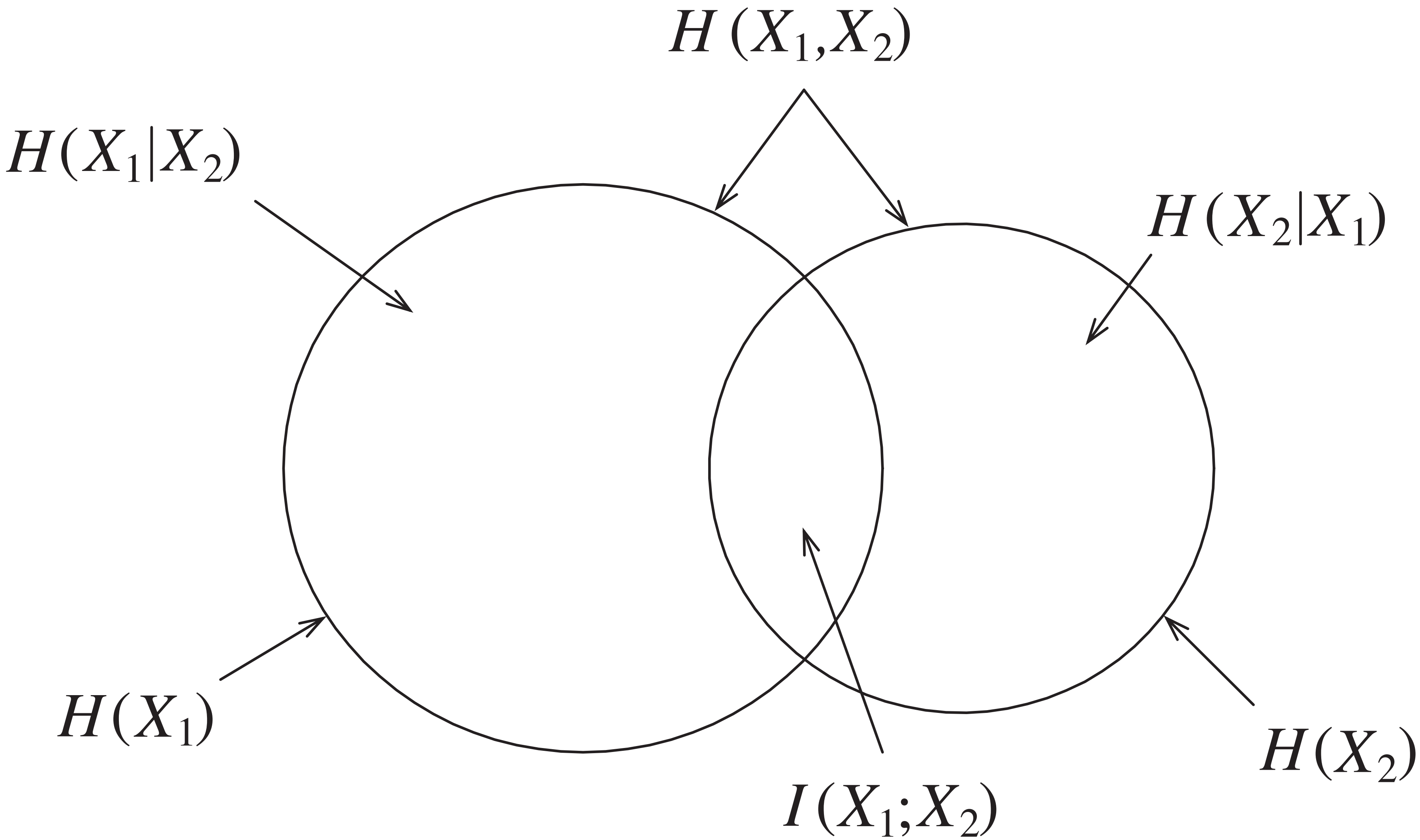
  $$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

  Conditional Mutual Information

  $$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z)$$

- These are called Shannon's information measures.

- Any Shannon's information measure is a linear combination of joint entropies.


$H(X_1,X_2)$
$H(X_1|X_2)$
$H(X_2|X_1)$
$H(X_1)$
$I(X_1;X_2)$
$H(X_2)$

- Hu Guoding (1962) showed that for every identity of Shannon's information measures, there is a corresponding set-theoretic identity.

- Hu Guoding (1962) showed that for every identity of Shannon's information measures, there is a corresponding set-theoretic identity.

- For example,

$$H(X_1, X_2) = H(X_1) + H(X_2) - I(X_1; X_2)$$

in information theory corresponds to

$$\mu(\tilde{X}_1 \cup \tilde{X}_2) = \mu(\tilde{X}_1) + \mu(\tilde{X}_2) - \mu(\tilde{X}_1 \cap \tilde{X}_2)$$

the Inclusion-Exclusion formulation in set theory, where

$$\begin{array}{ccc} H/I & \leftrightarrow & \mu \\ , & \leftrightarrow & \cup \\ ; & \leftrightarrow & \cap \\ | & \leftrightarrow & - \end{array}$$

μ is any set-additive function, and $\tilde{X}$ is a set variable corresponding to random variable X.

Some questions:

Some questions:

1. For every set-theoretic identity, does there exist an information-theoretic identity? If so, we can employ tools in set theory to study information theory.

Some questions:

1. For every set-theoretic identity, does there exist an information-theoretic identity? If so, we can employ tools in set theory to study information theory.

2. What would
$$\mu(\Omega) = \mu(A) + \mu(A^c)$$
correspond to?

The I-Measure μ^*

The I-Measure μ*

- To tackle

$$\mu(\Omega) = \mu(A) + \mu(A^c),$$

need to fix n, the number of random variables, and let $\Omega = \bigcup_{i=1}^{n} \tilde{X}_i$.

The I-Measure μ*

- To tackle

$$\mu(\Omega) = \mu(A) + \mu(A^c),$$

need to fix n, the number of random variables, and let $\Omega = \bigcup_{i=1}^n \tilde{X}_i$.

- Then the set of joint entropies of $X_1, X_2, \cdots, X_n$ defines a unique signed measure μ^*, called the *I*-Measure, which is consistent with Shannon's information measures.

The I-Measure μ*

- To tackle
$$\mu(\Omega) = \mu(A) + \mu(A^c),$$
need to fix n, the number of random variables, and let $\Omega = \bigcup_{i=1}^{n} \tilde{X}_i$.

- Then the set of joint entropies of $X_1, X_2, \cdots, X_n$ defines a unique signed measure μ^*, called the I-Measure, which is consistent with Shannon's information measures.

- This establishes the set-theoretic structure of Shannon's information measures.

The I-Measure μ*

- To tackle

$$\mu(\Omega) = \mu(A) + \mu(A^c),$$

need to fix n, the number of random variables, and let $\Omega = \bigcup_{i=1}^{n} \tilde{X}_i$.

- Then the set of joint entropies of $X_1, X_2, \cdots, X_n$ defines a unique signed measure μ^*, called the I-Measure, which is consistent with Shannon's information measures.

- This establishes the set-theoretic structure of Shannon's information measures.

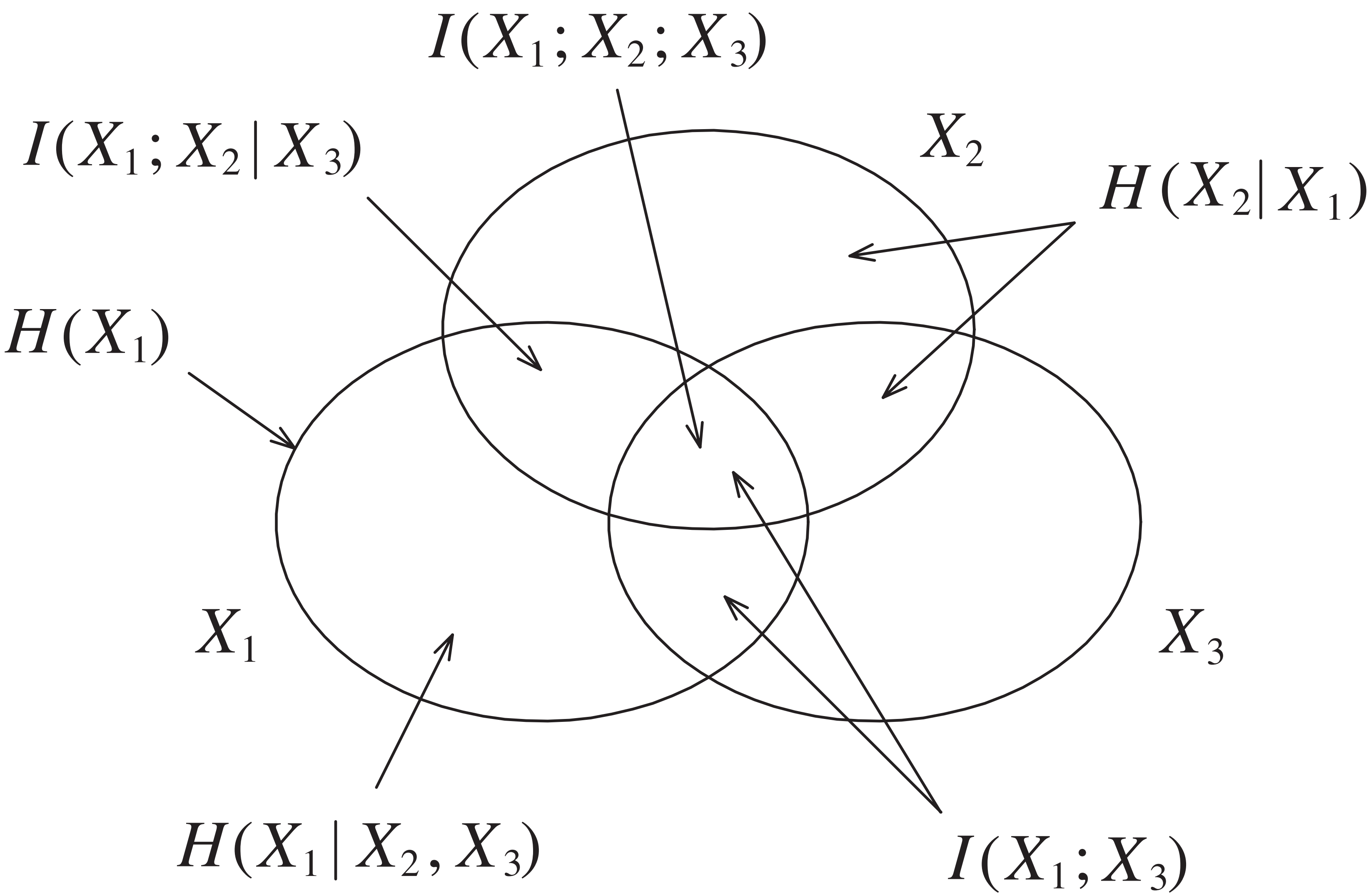
- As an example, for $n = 2$ and $A = \tilde{X}_1$,

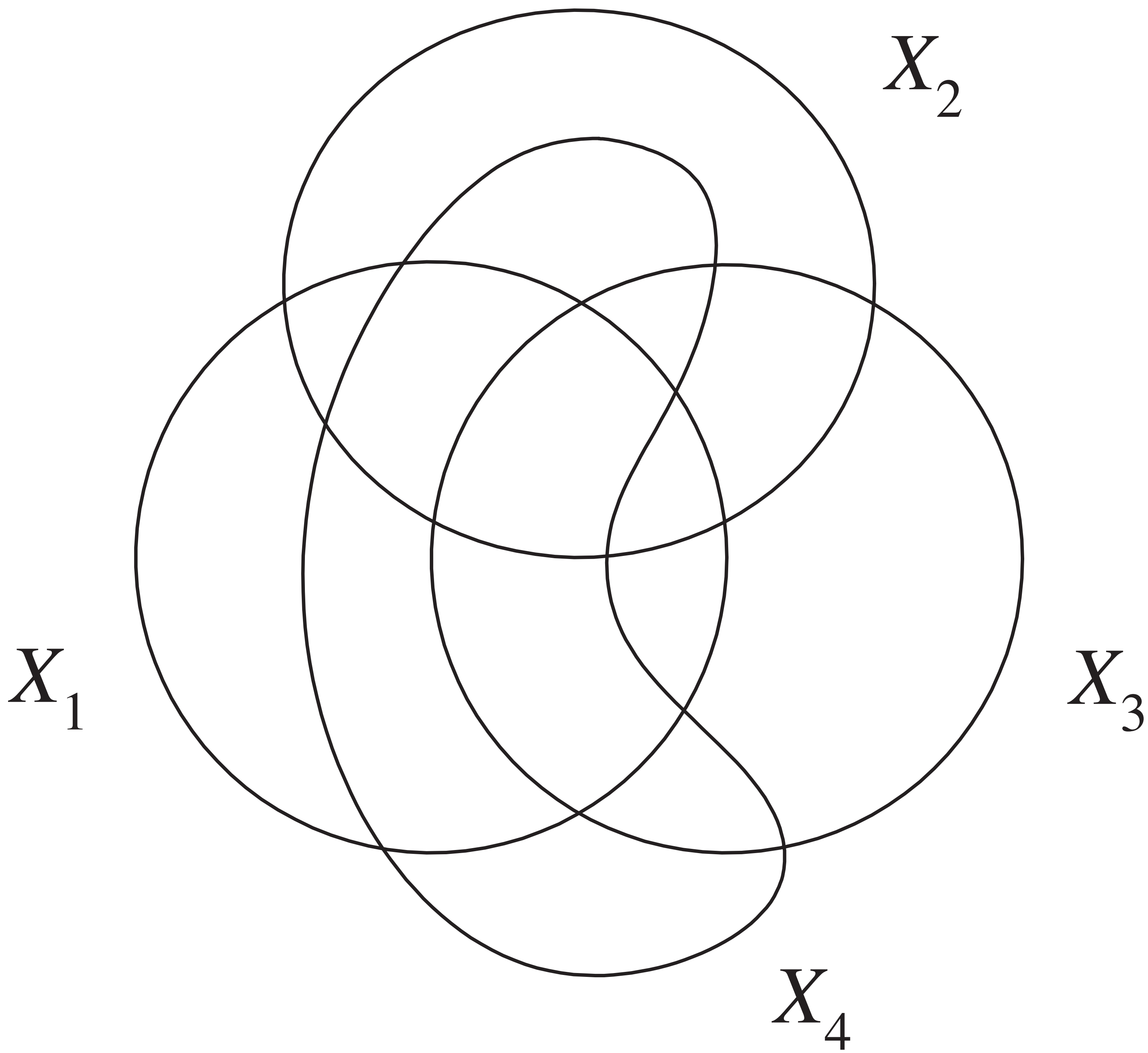
$$\mu(\Omega) = \mu(A) + \mu(A^c)$$

corresponds to

$$H(X_1, X_2) = H(X_1) + H(X_2|X_1).$$

Information Diagrams





X_2
X_1
X_3
X_4

μ^* can be Negative

μ* can be Negative

- μ^* is nonnegative for $n = 2$.

μ* can be Negative

- μ^* is nonnegative for $n = 2$.
- For $n = 3$, $\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) = I(X_1; X_2; X_3)$ can be negative.

μ* can be Negative

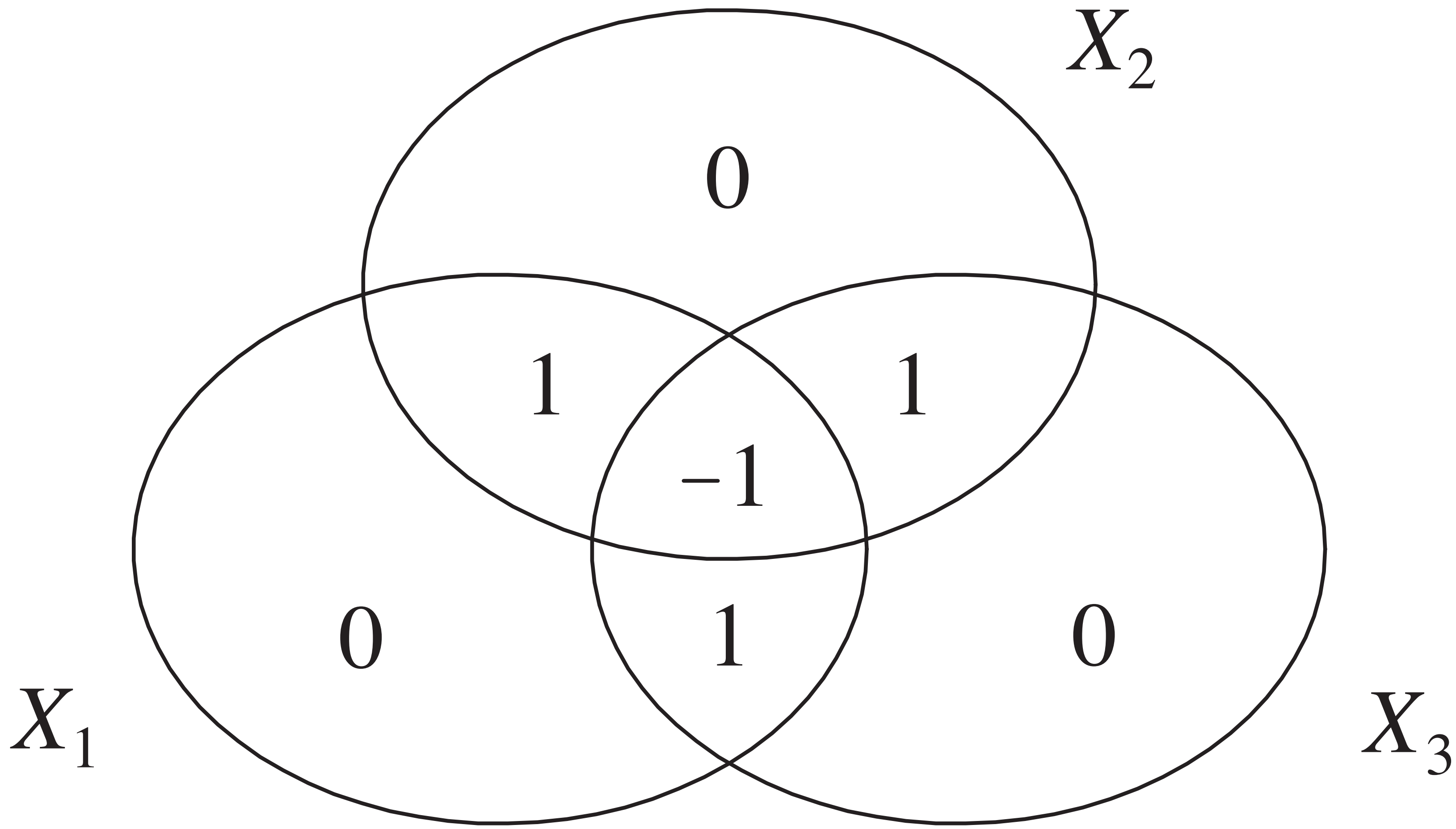
- μ^* is nonnegative for $n = 2$.
- For $n = 3$, $\mu^*(\tilde{X}_1 \cap \tilde{X}_2 \cap \tilde{X}_3) = I(X_1; X_2; X_3)$ can be negative.

Example 3.10

- X_1, X_2 – i.i.d. binary r.v.'s uniform on $\{0, 1\}$
- $X_3 = X_1 + X_2 \bmod 2$
- Easy to check:
  - $H(X_i) = 1$, for all i
  - X_1, X_2, X_3 are pairwise independent, so that

$$H(X_i, X_j) = 2 \text{ and } I(X_i; X_j) = 0, \text{ for all } i \neq j$$

  - Under these constraints, $I(X_1; X_2; X_3) = -1$.



The information diagram for Example 3.10

Example 3.15 (Imperfect Secrecy Theorem) Let X be the plain text, Y be the cipher text, and Z be the key in a secret key cryptosystem. Since X can be recovered from Y and Z, we have

$$H(X|Y,Z) = 0.$$

Show that this constraint implies

$$I(X;Y) \geq H(X) - H(Z).$$

Remark Do not need to make these assumptions about the scheme:

- $H(Y|X,Z) = 0$
- $I(X;Z) = 0$

VII. Future Work

We now address some issues for further investigations.

a) We have constructed a real measure μ^* on $\mathcal{F}$, which we call the I-Measure, from the joint distribution of the random variables involved. It should be pointed out that not every real measure μ on $\mathcal{F}$ is an I-Measure. For μ to be an I-Measure, it is necessary that the value of μ on the elements of $\mathcal{F}$, which correspond to Shannon's information measures are nonnegative. However, given such a measure, it is not clear whether we can always find a joint distribution for the random variables such that Shannon's information measures on these random variables agree with the value of μ on the corresponding elements of $\mathcal{F}$. This is a very fundamental question to be answered.

b) The value of μ^* on the elements of $\mathcal{F}$ that correspond to Shannon's information measures are always nonnegative. A question of interest is: What are the elements of $\mathcal{F}$ on which the value of μ^* are always nonnegative? The more general question of what linear combinations of entropies are always nonnegative was raised by Te Sun Han [13].

Markov Chains

Markov Chains

- For a Markov chain of any length n, the information diagram can be displayed in two dimensions because certain atoms can be suppressed.

Markov Chains

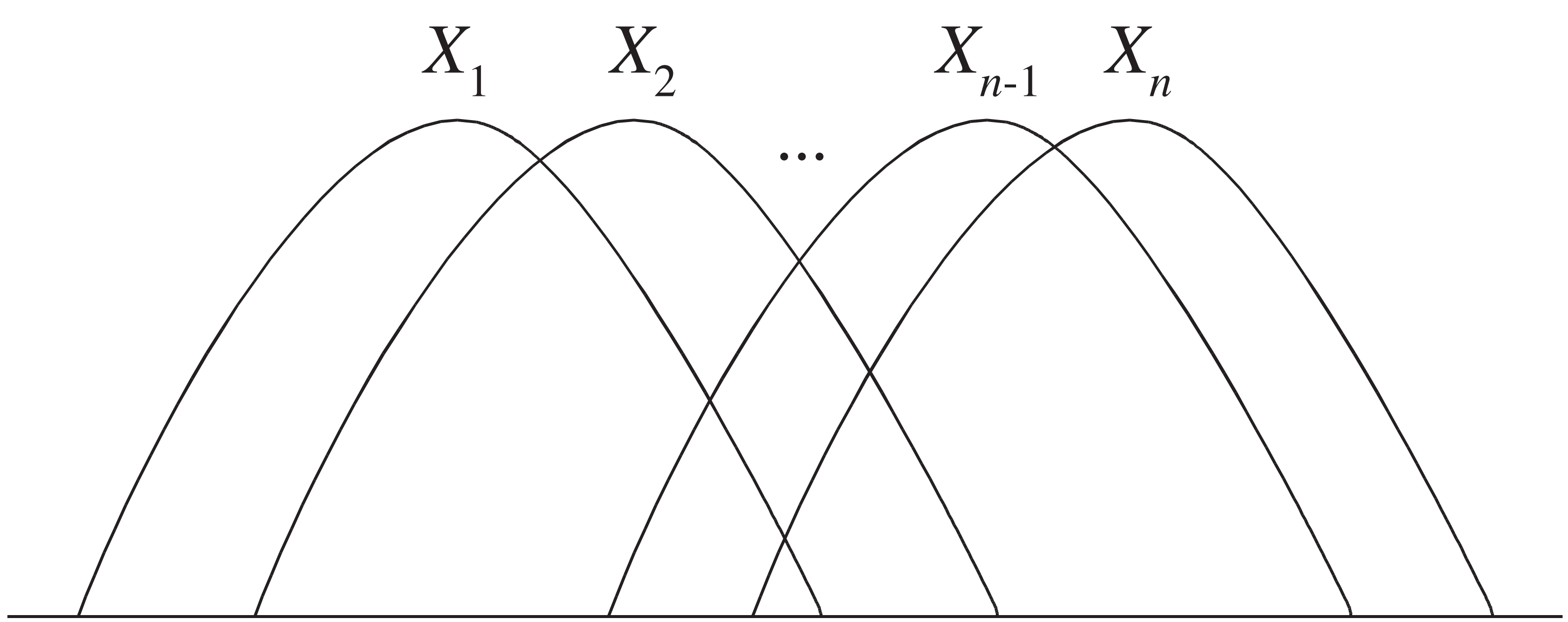
- For a Markov chain of any length n, the information diagram can be displayed in two dimensions because certain atoms can be suppressed.

- The values of μ^* on the remaining atoms correspond to Shannon's information measures and hence are nonnegative. Thus, μ^* is a measure. (Kawabata and Y, 1992).

Markov Chains

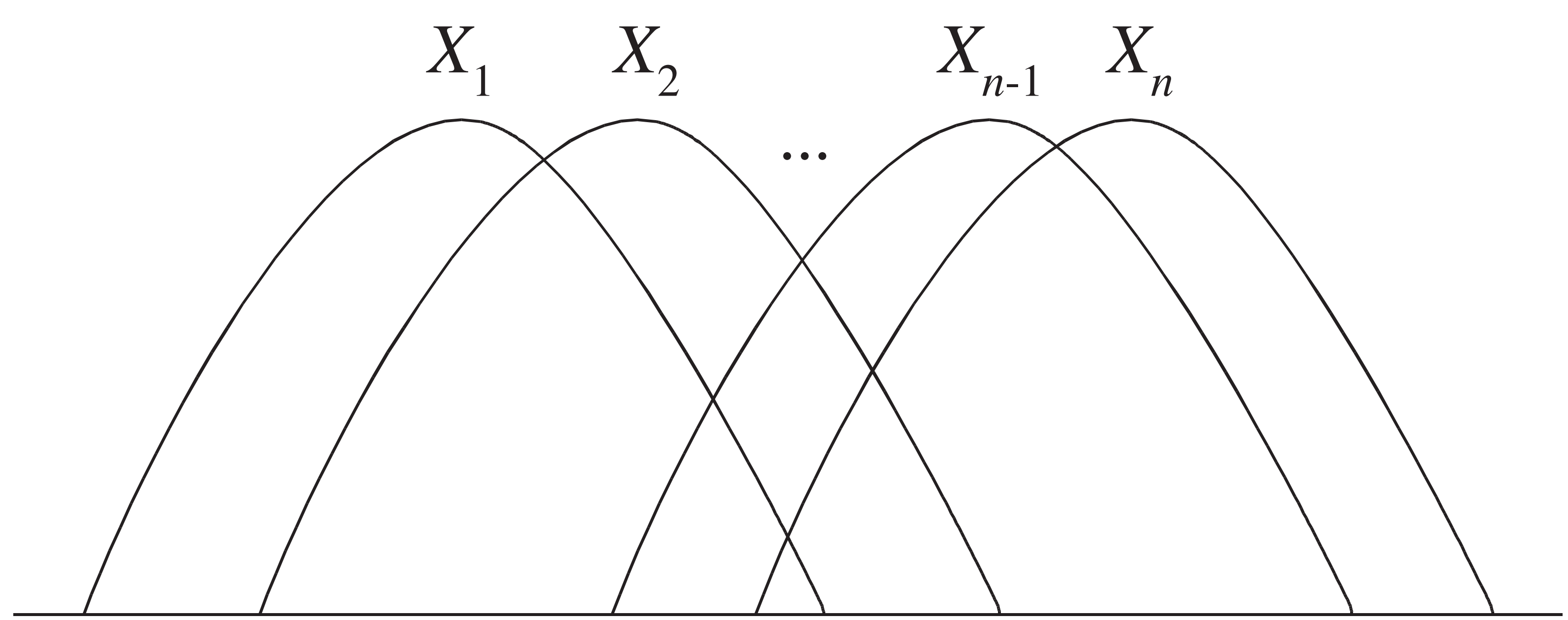
- For a Markov chain of any length n, the information diagram can be displayed in two dimensions because certain atoms can be suppressed.

- The values of μ^* on the remaining atoms correspond to Shannon's information measures and hence are nonnegative. Thus, μ^* is a measure. (Kawabata and Y, 1992).



Markov Chains

- For a Markov chain of any length n, the information diagram can be displayed in two dimensions because certain atoms can be suppressed.
- The values of μ^* on the remaining atoms correspond to Shannon's information measures and hence are nonnegative. Thus, μ^* is a measure. (Kawabata and Y, 1992).
- This theme can be extended to Markov random field. (Y, Ye and Lee, 2002).



Gaussian Multiterminal Source Coding

（正規多重情報源符号化問題の研究）

February, 1980

Supervisor: Prof. Nobunori Oshima

Tsutomu Kawabata

Course of Mathematical Engineering
and Instrumentation Physics,
Division of Engineering,
Graduate School,
University of Tokyo

In his book [7], Abramson uses interesting figures to help their readers to memorize various relationship among amounts of information. For example the relationships among H(X), H(Y), H(Y|X), H(Y|X), H(X,Y) and I(X;Y) are precisely expressed in the quantitative relationships of areas [Fig. A.1]. The consistencies on additivities that each form of amount of information has an aspect as a measure. However the situation is not so simple when we consider three random variables X,Y and Z. Pictures in [Fig. A.2] which also from [7] exactly express the relationships such as

$$I(X;Y|Z)=H(X|Z)-H(X|Y,Z) \quad . \qquad (A.1)$$

However there appears in this cace a quantity $I(X;Y;Z)\overset{d}{=}H(X,Y,Z)-H(X,Y)-H(Y,Z)-H(Z,X)+H(X)+H(Y)+H(Z)$, whose meaning is not easy to understand. In fact I(X;Y;Z) is not always positive.

Our major interest is not in the generally correlated random

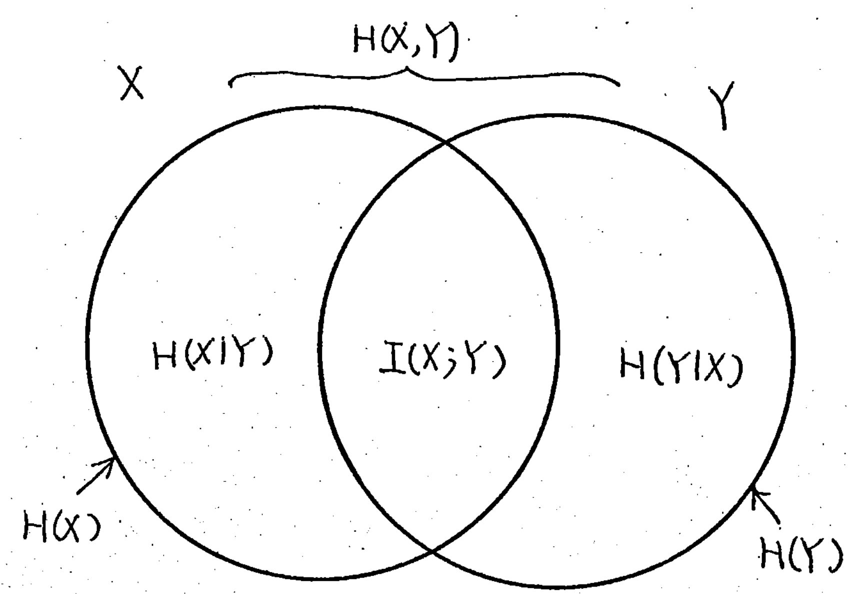


Fig. A.1

Relationships among information (Abramson [7])

We want to define a measure μ on the Boolean algebra generated by $\tilde{X}_1, \cdots \tilde{X}_m$, which we denote by $\mathcal{B}$, such that the equalities

$$H(X_A) = \mu(\tilde{X}_A) \tag{A.3}$$

$$H(X_A|X_B) = \mu(\tilde{X}_A - \tilde{X}_B) \tag{A.4}$$

$$I(X_A;X_B) = \mu(\tilde{X}_A \cap \tilde{X}_B) \tag{A.5}$$

$$I(X_A;X_B|X_C) = \mu(\tilde{X}_A \cap \tilde{X}_B - \tilde{X}_C) \tag{A.6}$$

are satisfied.

Next we define concepts and prove some lemmas.

Def. A.1.

Non-empty $A \in \mathcal{B}$ is called p-atom if it is not $\mathbb{R}^2 - \bigcup_{i=1}^{n} \tilde{X}_i$ and for any $B \in \mathcal{B}$, $B \supset A$ or $B \cap A = \phi$.

An p-atom is illustrated in [Fig. A.8]. It is noted that a p-atom is not a usual atom in the Boolean algebraic sence. It is really the stuff (not empty) set.

Lemma A.1.

If $B \in \mathcal{B}$ is not empty and not $\mathbb{R}^2 - \bigcup_{i=1}^{n} \tilde{X}_i$, it is expressed uniquely as a finite direct sum of p-atom.

[proof]

This is a Boolean algebraic result. By the condition (A.2) many atoms vanishes. [Fig. A.8] show the situation.

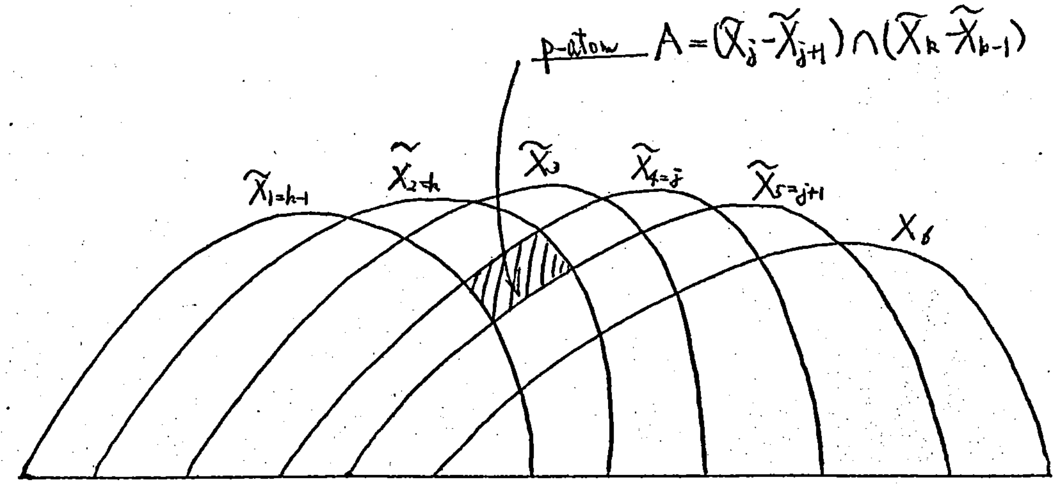


Fig. A.8

An p-atom

The Structure of the *I*-Measure of a Markov Chain

Tsutomu Kawabata and Raymond W. Yeung

***Abstract*—The underlying mathematical structure of Shannon's information measures was studied in a recent paper by Yeung, and the *I*-Measure μ^*, which is a signed measure defined on a proper σ-field $\mathcal{F}$, was introduced. The *I*-Measure is a natural extension of Shannon's information measures and is uniquely defined by them. They also introduced as a consequence the *I*-Diagram as a geometric tool to visualize the relationship among the information measures. In general, an *I*-Diagram for n random variables must be constructed in n-1 dimensions. It is shown that for any finite collection of random variables forming a Markov chain, μ^* assumes a very simple structure which can be illustrated by an *I*-Diagram in two dimensions, and μ^* is a nonnegative measure.**

***Index Terms*—Shannon's information measures, *I*-Measure, *I*-Diagram, Markov Chain.**

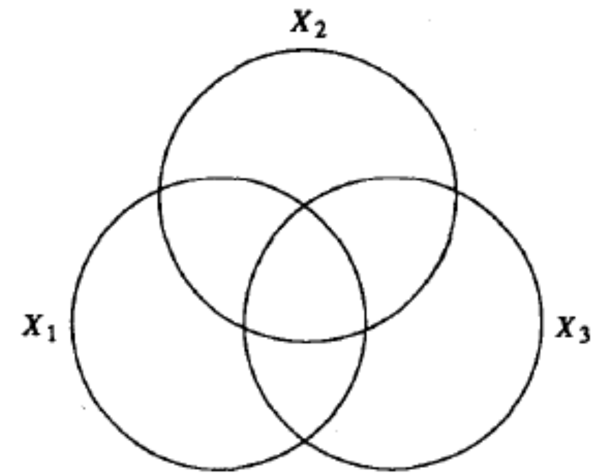


Fig. 1.

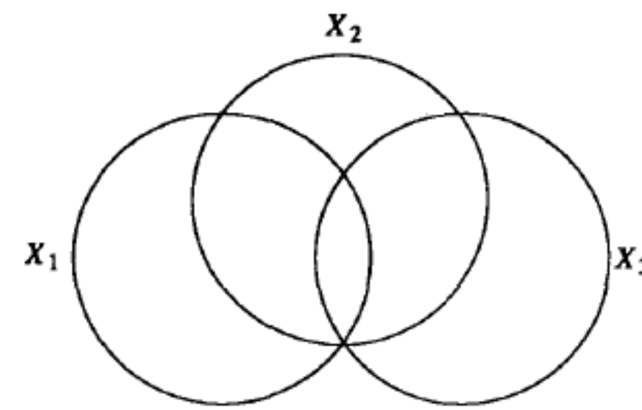


Fig. 2.

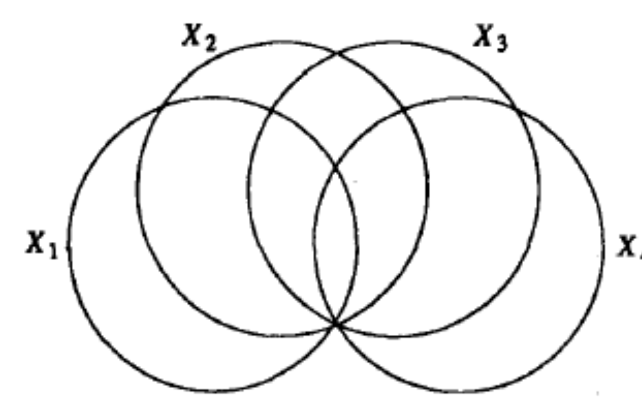


Fig. 3.

2) $I(X_2; X_3) = I(X_1; X_4) + I(X_1; X_3 \mid X_4) + I(X_2; X_4 \mid X_1) + I(X_2; X_3 \mid X_1, X_4)$,
3) $H(X_2, X_4 \mid X_1, X_3) = H(X_2 \mid X_1, X_3) + H(X_4 \mid X_3)$,
4) $H(X_2 \mid X_1, X_4) = H(X_2 \mid X_1, X_3) + I(X_2; X_3 \mid X_1, X_4)$.

Theorem 1: Let $X_i, i = 1, \cdots, n, n \geq 2$, be collection of random variables and $-(X_i)-$. Then $I(X_1; \cdots; X_n) = I(X_1; X_n)$.

Proof: We shall prove the theorem by induction. It is clear that it is true for $n = 2$. Assume it is true for some $n \geq 2$. Then

$$I(X_1; X_2; \cdots; X_{n+1}) = I(X_1; X_3; \cdots; X_{n+1}) - I(X_1; X_3; \cdots; X_{n+1} | X_2).$$

By the induction hypothesis,

$$I(X_1; X_3; \cdots; X_{n+1}) = I(X_1; X_{n+1}).$$

Now by Theorem 5 in [1],

$$I(X_1; X_3; \cdots; X_{n+1} | X_2) = \sum_{x_2} P[X_2 = x_2] I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2).$$

Conditioning on $X_2 = x_2$, the random variables $X_1, X_3, \cdots, X_{n+1}$ still form a Markov chain. By the induction hypothesis, we conclude that

$$I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2) = I(X_1; X_{n+1} | X_2 = x_2).$$

Thus,

$$\begin{aligned} I(X_1; X_3; \cdots; X_{n+1} | X_2) &= \sum_{x_2} P[X_2 = x_2] I(X_1; X_{n+1} | X_2 = x_2) \\ &= I(X_1; X_{n+1} | X_2) = 0, \end{aligned}$$

proving the theorem.

Theorem 1: Let $X_i, i = 1, \cdots, n, n \geq 2$, be collection of random variables and $-(X_i)-$. Then $I(X_1; \cdots; X_n) = I(X_1; X_n)$.

Proof: We shall prove the theorem by induction. It is clear that it is true for $n = 2$. Assume it is true for some $n \geq 2$. Then

$$I(X_1; X_2; \cdots; X_{n+1}) = I(X_1; X_3; \cdots; X_{n+1}) - I(X_1; X_3; \cdots; X_{n+1} | X_2).$$

By the induction hypothesis,

$$I(X_1; X_3; \cdots; X_{n+1}) = I(X_1; X_{n+1}).$$

Now by Theorem 5 in [1],

$$I(X_1; X_3; \cdots; X_{n+1} | X_2) = \sum_{x_2} P[X_2 = x_2] I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2).$$

Conditioning on $X_2 = x_2$, the random variables $X_1, X_3, \cdots, X_{n+1}$ still form a Markov chain. By the induction hypothesis, we conclude that

$$I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2) = I(X_1; X_{n+1} | X_2 = x_2).$$

Thus,

$$\begin{aligned} I(X_1; X_3; \cdots; X_{n+1} | X_2) &= \sum_{x_2} P[X_2 = x_2] I(X_1; X_{n+1} | X_2 = x_2) \\ &= I(X_1; X_{n+1} | X_2) = 0, \end{aligned}$$

proving the theorem.

Theorem 1: Let $X_i, i = 1, \cdots, n, n \geq 2$, be collection of random variables and $-(X_i)-$. Then $I(X_1; \cdots; X_n) = I(X_1; X_n)$.

Proof: We shall prove the theorem by induction. It is clear that it is true for $n = 2$. Assume it is true for some $n \geq 2$. Then

$$I(X_1; X_2; \cdots; X_{n+1}) = I(X_1; X_3; \cdots; X_{n+1}) - I(X_1; X_3; \cdots; X_{n+1} | X_2).$$

By the induction hypothesis,

$$I(X_1; X_3; \cdots; X_{n+1}) = I(X_1; X_{n+1}).$$

Now by Theorem 5 in [1],

$$I(X_1; X_3; \cdots; X_{n+1} | X_2) = \sum_{x_2} P[X_2 = x_2] I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2).$$

Conditioning on $X_2 = x_2$, the random variables $X_1, X_3, \cdots, X_{n+1}$ still form a Markov chain. By the induction hypothesis, we conclude that

$$I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2) = I(X_1; X_{n+1} | X_2 = x_2).$$

Thus,

$$I(X_1; X_3; \cdots; X_{n+1} | X_2) = \sum_{x_2} P[X_2 = x_2] I(X_1; X_{n+1} | X_2 = x_2) = I(X_1; X_{n+1} | X_2) = 0,$$

proving the theorem.

Theorem 1: Let $X_i, i = 1, \cdots, n, n \geq 2$, be collection of random variables and $-(X_i)-$. Then $I(X_1; \cdots; X_n) = I(X_1; X_n)$.

Proof: We shall prove the theorem by induction. It is clear that it is true for $n = 2$. Assume it is true for some $n \geq 2$. Then

$$I(X_1; X_2; \cdots; X_{n+1}) = I(X_1; X_3; \cdots; X_{n+1}) - I(X_1; X_3; \cdots; X_{n+1} | X_2).$$

By the induction hypothesis,

$$I(X_1; X_3; \cdots; X_{n+1}) = I(X_1; X_{n+1}).$$

Now by Theorem 5 in [1],

$$I(X_1; X_3; \cdots; X_{n+1} | X_2) = \sum_{x_2} P[X_2 = x_2] I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2).$$

Conditioning on $X_2 = x_2$, the random variables $X_1, X_3, \cdots, X_{n+1}$ still form a Markov chain. By the induction hypothesis, we conclude that

$$I(X_1; X_3; \cdots; X_{n+1} | X_2 = x_2) = I(X_1; X_{n+1} | X_2 = x_2).$$

Thus,

$$I(X_1; X_3; \cdots; X_{n+1} | X_2) = \sum_{x_2} P[X_2 = x_2] I(X_1; X_{n+1} | X_2 = x_2) = I(X_1; X_{n+1} | X_2) = 0,$$

proving the theorem.

Acknowledgment

The author thanks Prof. I. Czsisar for his creative suggestions that significantly simplify the proof of the results, and Prof. T. Cover and Prof. L. Campbell for pointing out the references.

1995

Zhen Zhang visited CUHK in Fall 1995

R.W. Yeung, "A framework for linear information inequalities," IEEE-IT, 1997

Z. Zhang and R.W. Yeung, " A non-Shannon-type conditional inequality of information quantities," IEEE-IT, 1997

Z. Zhang and R.W. Yeung, " On characterization of entropy function via information inequalities," IEEE-IT, 1997

Preliminaries

Preliminaries

- $\mathcal{N}_n = \{1, \ldots, n\}$

Preliminaries

- $\mathcal{N}_n = \{1, \ldots, n\}$
- $\Omega = \{X_i, i \in \mathcal{N}_n\}$; X_i is a discrete random variable.

Preliminaries

- $\mathcal{N}_n = \{1, \ldots, n\}$
- $\Omega = \{X_i, i \in \mathcal{N}_n\}$; X_i is a discrete random variable.
- For n random variables, there are $2^n - 1$ joint entropies.

Preliminaries

- $\mathcal{N}_n = \{1, \ldots, n\}$
- $\Omega = \{X_i, i \in \mathcal{N}_n\}$; X_i is a discrete random variable.
- For n random variables, there are $2^n - 1$ joint entropies.
- E.g., $n = 3$, the $2^3 - 1 = 7$ joint entropies are

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

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

Preliminaries

Preliminaries

- Let $X_\alpha = (X_i, i \in \alpha)$. E.g., $X_{\{1,2,3\}} = (X_1, X_2, X_3)$.

Preliminaries

- Let $X_\alpha = (X_i, i \in \alpha)$. E.g., $X_{\{1,2,3\}} = (X_1, X_2, X_3)$.
- Define $H_\Omega(\alpha) = H(X_\alpha)$. E.g., $H_\Omega(\{1,2,3\}) = H(X_1, X_2, X_3)$.

Preliminaries

- Let $X_\alpha = (X_i, i \in \alpha)$. E.g., $X_{\{1,2,3\}} = (X_1, X_2, X_3)$.
- Define $H_\Omega(\alpha) = H(X_\alpha)$. E.g., $H_\Omega(\{1,2,3\}) = H(X_1, X_2, X_3)$.
- $H_\Omega : 2^{\mathcal{N}_n} \to \mathbb{R}$ is set function with $H_\Omega(\phi) = 0$.

Preliminaries

- Let $X_\alpha = (X_i, i \in \alpha)$. E.g., $X_{\{1,2,3\}} = (X_1, X_2, X_3)$.
- Define $H_\Omega(\alpha) = H(X_\alpha)$. E.g., $H_\Omega(\{1,2,3\}) = H(X_1, X_2, X_3)$.
- $H_\Omega : 2^{\mathcal{N}_n} \rightarrow \mathbb{R}$ is set function with $H_\Omega(\phi) = 0$.
- H_Ω is called the entropy function of Ω.

The Entropy Function as a Polymatroid

- It is well-known that for any Ω, H_Ω satisfies the following *polymatroidal axioms*. For any $\alpha, \beta \subset \mathcal{N}_n$,

  (P1) $H_\Omega(\phi) = 0$;

  (P2) $H_\Omega(\alpha) \leq H_\Omega(\beta)$ if $\alpha \subset \beta$;

  (P3) $H_\Omega(\alpha) + H_\Omega(\beta) \geq H_\Omega(\alpha \cap \beta) + H_\Omega(\alpha \cup \beta)$.

The Basic Inequalities

- In addition to Entropy, we also have:

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

$$H(X|Y) = H(X,Y) - H(Y)$$

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

$$H(X|Y) = H(X,Y) - H(Y)$$

  Mutual Information

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

$$H(X|Y) = H(X,Y) - H(Y)$$

  Mutual Information

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

  Conditional Mutual Information

$$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z)$$

The Basic Inequalities

- In addition to Entropy, we also have:

  Conditional Entropy

  $$H(X|Y) = H(X,Y) - H(Y)$$

  Mutual Information

  $$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

  Conditional Mutual Information

  $$I(X;Y|Z) = H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z)$$

- These are called Shannon's information measures.

The Basic Inequalities

- The polymatriodal axioms are equivalent to the nonnegativity of Shannon's information measures, called the basic inequalities.

The Basic Inequalities

- The polymatriodal axioms are equivalent to the nonnegativity of Shannon's information measures, called the basic inequalities.
- That is,

$$\begin{aligned} \text{entropy} &\geq 0 \\ \text{mutual info} &\geq 0 \\ \text{conditional entropy} &\geq 0 \\ \text{conditional mutual info} &\geq 0 \end{aligned}$$

The Region Γ_n^*

The Region Γ_n^*

- Fix n. For each Ω, H_Ω defines a vector in $\mathcal{H}_n = \mathbb{R}^{2^n - 1}$.

The Region Γ_n^*

- Fix n. For each Ω, H_Ω defines a vector in $\mathcal{H}_n = \mathbb{R}^{2^n-1}$.
- $\mathcal{H}_n$ is called the *entropy space* for n r.v.'s

The Region Γ_n^*

- Fix n. For each Ω, H_Ω defines a vector in $\mathcal{H}_n = \mathbb{R}^{2^n-1}$.
- $\mathcal{H}_n$ is called the *entropy space* for n r.v.'s
- A vector

$$\mathbf{h} = (h_\alpha : \alpha \in 2^{\mathcal{N}_n \backslash \emptyset})$$

  in $\mathcal{H}_n$ is called entropic if it corresponds to the entropy function H_Ω for some Ω.

The Region Γ_n^*

- Fix n. For each Ω, H_Ω defines a vector in $\mathcal{H}_n = \mathbb{R}^{2^n-1}$.
- $\mathcal{H}_n$ is called the *entropy space* for n r.v.'s
- A vector
$$\mathbf{h} = (h_\alpha : \alpha \in 2^{\mathcal{N}_n \setminus \emptyset})$$
in $\mathcal{H}_n$ is called entropic if it corresponds to the entropy function H_Ω for some Ω.
- Define the region in $\mathcal{H}_n$:
$$\Gamma_n^* = \{\mathbf{h} \in \mathcal{H}_n : \mathbf{h} \text{ is entropic}\}$$

Entropy Inequalities:
A Geometric View

Entropy Inequalities: A Geometric View

- An entropy inequality has the form $f(\mathbf{h}) \geq 0$.

Entropy Inequalities: A Geometric View

- An entropy inequality has the form $f(\mathbf{h}) \geq 0$.
- $f(\mathbf{h}) \geq 0$ always holds if and only if

$$\Gamma_n^* \subset \{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}.$$

Entropy Inequalities: A Geometric View

- An entropy inequality has the form $f(\mathbf{h}) \geq 0$.
- $f(\mathbf{h}) \geq 0$ always holds if and only if

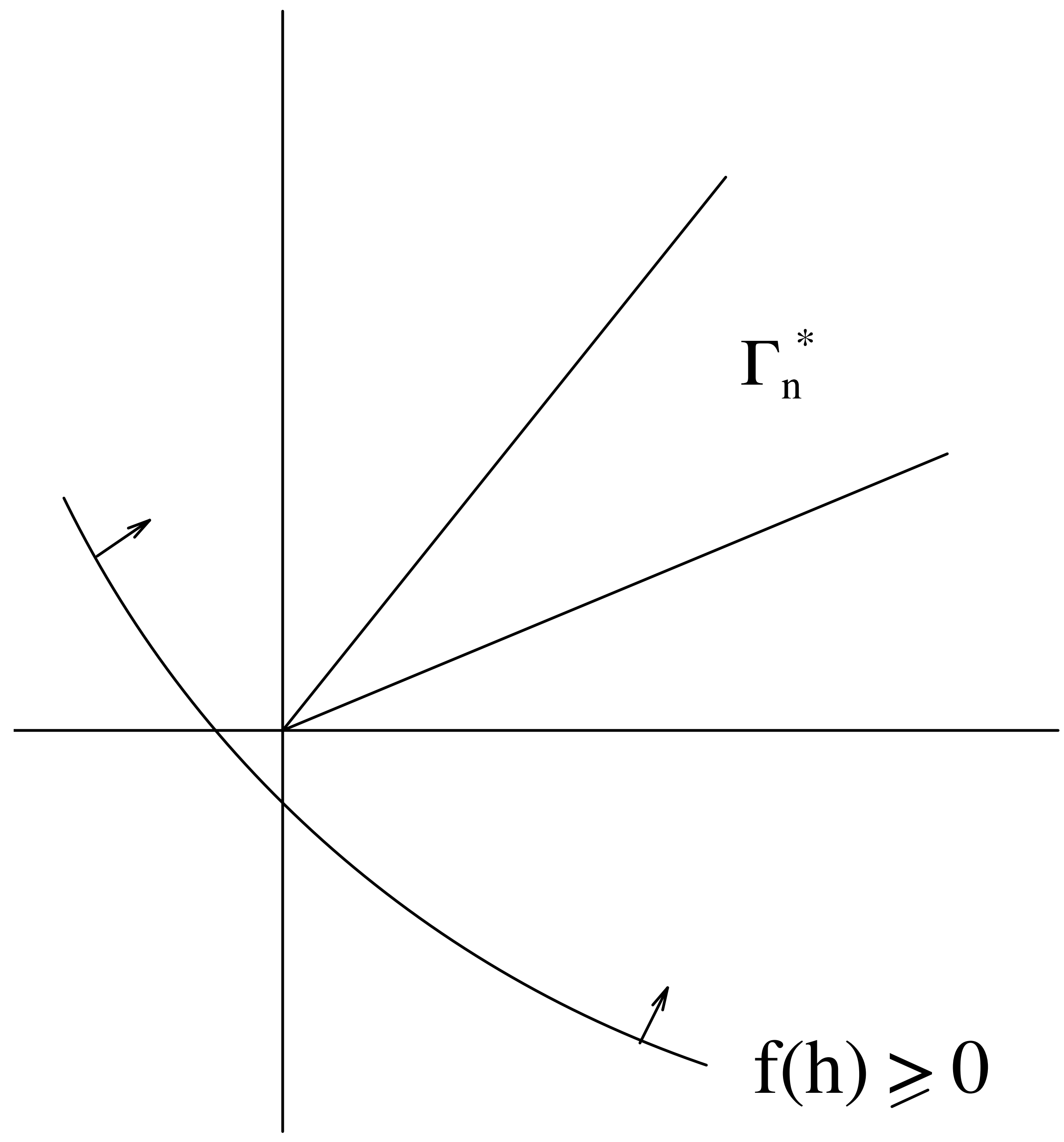
$$\Gamma_n^* \subset \{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}.$$

- In fact, $f(\mathbf{h}) \geq 0$ always holds if and only if

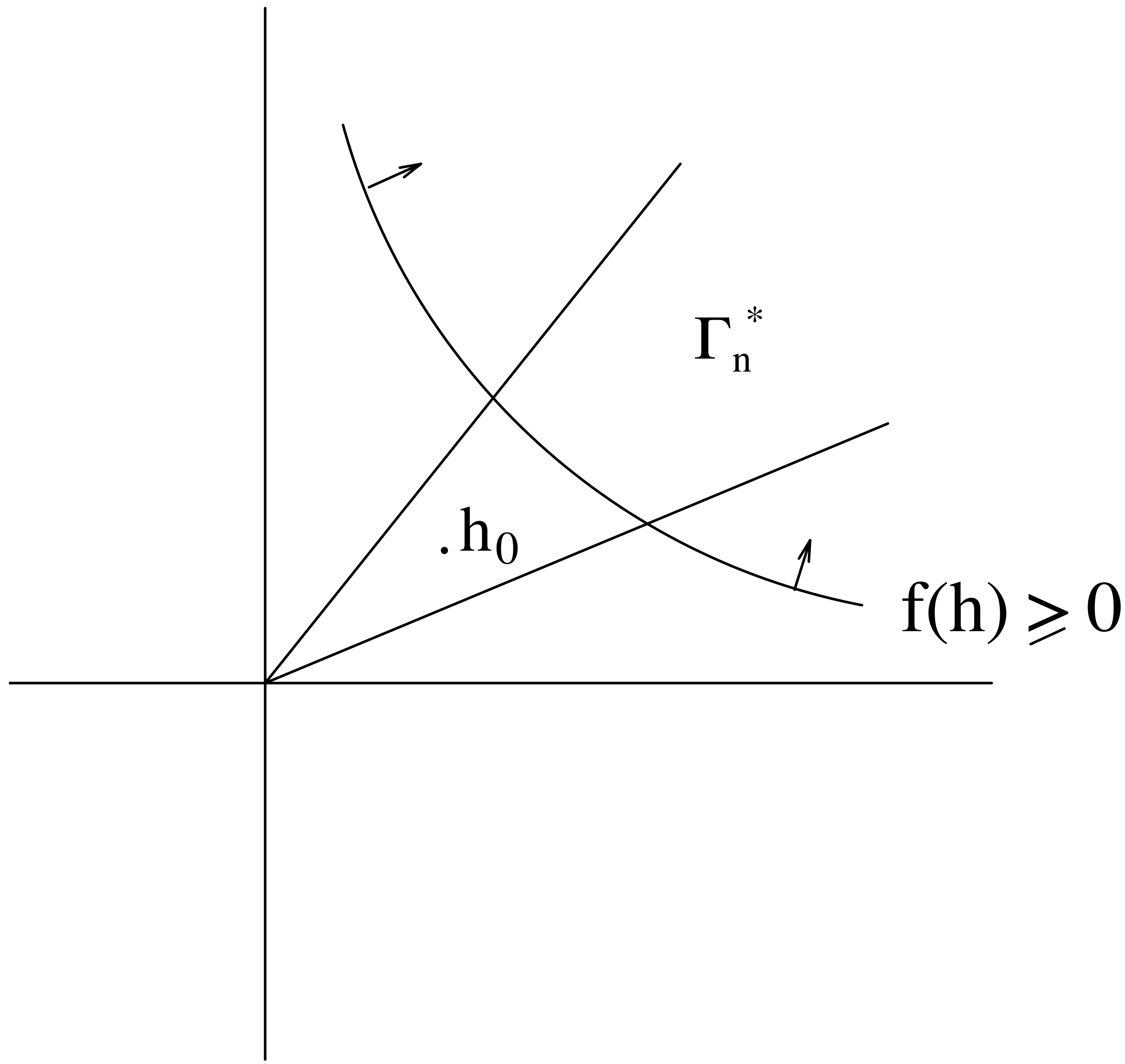
$$\overline{\Gamma}_n^* \subset \{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}$$

because $\{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}$ is closed.

$f(\mathbf{h}) \geq 0$ Always holds



$f(\mathbf{h}) \geq 0$ Does Not Always holds



The Region Γ_n

The Region Γ_n

- Define the following region in $\mathcal{H}_n$:

$$\Gamma_n = \{\mathbf{h} \in \mathcal{H}_n : \mathbf{h} \text{ satisfies the basic inequalities}\}$$

The Region Γ_n

- Define the following region in $\mathcal{H}_n$:

$$\Gamma_n = \{\mathbf{h} \in \mathcal{H}_n : \mathbf{h} \text{ satisfies the basic inequalities}\}$$

- $\Gamma_n^* \subset \Gamma_n$ since the basic inequalities are satisfied by any $X_1, \ldots, X_n$.

The Region Γ_n

- Define the following region in $\mathcal{H}_n$:

$$\Gamma_n = \{\mathbf{h} \in \mathcal{H}_n : \mathbf{h} \text{ satisfies the basic inequalities}\}$$

- $\Gamma_n^* \subset \Gamma_n$ since the basic inequalities are satisfied by any $X_1, \ldots, X_n$.

- An entropy inequality $f(\mathbf{h}) \geq 0$ is called a Shannon-type inequality if it is implied by the basic inequalities, or

$$\Gamma_n \subset \{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}.$$

Machine-Proving of Entropy Inequalities

The geometric view of entropy inequalities enables machine-proving of entropy inequalities. The following applications have been developed:

Machine-Proving of Entropy Inequalities

The geometric view of entropy inequalities enables machine-proving of entropy inequalities. The following applications have been developed:

1. ITIP (Information-Theoretic Inequality Prover) at CUHK

Machine-Proving of Entropy Inequalities

The geometric view of entropy inequalities enables machine-proving of entropy inequalities. The following applications have been developed:

1. ITIP (Information-Theoretic Inequality Prover) at CUHK (Y.-O.Yan and Y, 1996)

2. Xitip at EPFL (Pulikkoonattu, Perron, Diggavi, 2007)

Machine-Proving of Entropy Inequalities

The geometric view of entropy inequalities enables machine-proving of entropy inequalities. The following applications have been developed:

1. ITIP (Information-Theoretic Inequality Prover) at CUHK (Y.-O.Yan and Y, 1996)
2. Xitip at EPFL (Pulikkoonattu, Perron, Diggavi, 2007)
3. ITTP (Information-Theoretic Theorem Prover) at KAIST (S.-Y. Chung, 2009)

Machine-Proving of Entropy Inequalities

The geometric view of entropy inequalities enables machine-proving of entropy inequalities. The following applications have been developed:

1. ITIP (Information-Theoretic Inequality Prover) at CUHK (Y.-O.Yan and Y, 1996)

2. Xitip at EPFL (Pulikkoonattu, Perron, Diggavi, 2007)

3. ITTP (Information-Theoretic Theorem Prover) at KAIST (S.-Y. Chung, 2009)

ITIP and Xitip are linear programming based, while ITTP is axiom based.

Is $\bar{\Gamma}_n^* = \Gamma_n$?

Is $\bar{\Gamma}_n^* = \Gamma_n$?

- The answer is NO iff there exists an entropy inequality $g(\mathbf{h}) \geq 0$ which cuts between Γ_n and $\overline{\Gamma}_n^*$. Such an inequality is called a non-Shannon-type inequality.

Is $\bar{\Gamma}_n^* = \Gamma_n$?

- The answer is NO iff there exists an entropy inequality $g(\mathbf{h}) \geq 0$ which cuts between Γ_n and $\overline{\Gamma}_n^*$. Such an inequality is called a non-Shannon-type inequality.

- It is known that

  1. $\Gamma_2^* = \Gamma_2$
  2. $\Gamma_3^* \neq \Gamma_3$, but $\overline{\Gamma}_3^* = \Gamma_3$

Is $\bar{\Gamma}^*_n = \Gamma_n$?

- The answer is NO iff there exists an entropy inequality $g(\mathbf{h}) \geq 0$ which cuts between Γ_n and $\overline{\Gamma}^*_n$. Such an inequality is called a non-Shannon-type inequality.

- It is known that

  1. $\Gamma^*_2 = \Gamma_2$
  2. $\Gamma^*_3 \neq \Gamma_3$, but $\overline{\Gamma}^*_3 = \Gamma_3$

- Therefore, unconstrained non-Shannon-type inequalities can exist only for 4 or more random variables.

Is $\bar{\Gamma}^*_n = \Gamma_n$?

- The answer is NO iff there exists an entropy inequality $g(\mathbf{h}) \geq 0$ which cuts between Γ_n and $\overline{\Gamma}^*_n$. Such an inequality is called a non-Shannon-type inequality.

- It is known that

  1. $\Gamma^*_2 = \Gamma_2$
  2. $\Gamma^*_3 \neq \Gamma_3$, but $\overline{\Gamma}^*_3 = \Gamma_3$

- Therefore, unconstrained non-Shannon-type inequalities can exist only for 4 or more random variables.

- In general,

  – Γ^*_n is neither closed nor convex, but $\overline{\Gamma}^*_n$ is a convex cone.

A Constrained Non-Shannon-Type Inequality (ZY97)

A Constrained Non-Shannon-Type Inequality (ZY97)

- The following constrained inequality was discovered for any 4 random variables (Zhang and Y, 1997):

  If $X_1 \perp X_2 | X_3$ and $X_1 \perp X_2$, then

$$I(X_3; X_4) \leq I(X_3; X_4|X_1) + I(X_3; X_4|X_2)$$

A Constrained Non-Shannon-Type Inequality (ZY97)

- The following constrained inequality was discovered for any 4 random variables (Zhang and Y, 1997):

  If $X_1 \perp X_2 | X_3$ and $X_1 \perp X_2$, then

  $$I(X_3; X_4) \leq I(X_3; X_4|X_1) + I(X_3; X_4|X_2)$$

- Under the constraint $X_1 \perp X_2$, this is equivalent to the Ingleton inequality in matroid theory:

  $$\begin{aligned} & H(X_1, X_3) + H(X_1, X_4) + H(X_2, X_3) + H(X_2, X_4) + H(X_3, X_4) \\ \geq \quad & H(X_3) + H(X_4) + H(X_1, X_2) + H(X_1, X_3, X_4) + H(X_2, X_3, X_4) \end{aligned}$$

A Constrained Non-Shannon-Type Inequality (ZY97)

- The following constrained inequality was discovered for any 4 random variables (Zhang and Y, 1997):

  If $X_1 \perp X_2 | X_3$ and $X_1 \perp X_2$, then

$$I(X_3; X_4) \le I(X_3; X_4|X_1) + I(X_3; X_4|X_2)$$

- Under the constraint $X_1 \perp X_2$, this is equivalent to the Ingleton inequality in matroid theory:

$$\begin{aligned} & H(X_1, X_3) + H(X_1, X_4) + H(X_2, X_3) + H(X_2, X_4) + H(X_3, X_4) \\ \ge \quad & H(X_3) + H(X_4) + H(X_1, X_2) + H(X_1, X_3, X_4) + H(X_2, X_3, X_4) \end{aligned}$$

- ZY97 says that a certain region on the boundary of Γ_n is not entropic.

A Constrained Non-Shannon-Type Inequality (ZY97)

- The following constrained inequality was discovered for any 4 random variables (Zhang and Y, 1997):

  If $X_1 \perp X_2 | X_3$ and $X_1 \perp X_2$, then

$$I(X_3; X_4) \leq I(X_3; X_4|X_1) + I(X_3; X_4|X_2)$$

- Under the constraint $X_1 \perp X_2$, this is equivalent to the Ingleton inequality in matroid theory:

$$\begin{aligned} H(X_1, X_3) + H(X_1, X_4) + H(X_2, X_3) + H(X_2, X_4) + H(X_3, X_4) \\ \geq \quad H(X_3) + H(X_4) + H(X_1, X_2) + H(X_1, X_3, X_4) + H(X_2, X_3, X_4) \end{aligned}$$

- ZY97 says that a certain region on the boundary of Γ_n is not entropic.

- However, it is not strong enough to imply that $\overline{\Gamma}_4^* \neq \Gamma_4$.

An Unconstrained Non-Shannon-Type Inequality (ZY98)

An Unconstrained Non-Shannon-Type Inequality (ZY98)

- The following unconstrained non-Shannon-type inequality was discovered for any 4 random variables (Zhang and Y, 1998):

$$
\begin{aligned}
& I(Z;U) - I(Z;U|X) - I(Z;U|Y) \\
& \quad \leq \quad \frac{1}{2}I(X;Y) + \frac{1}{4}[I(X;Z,U) + I(Y;Z,U)]
\end{aligned}
$$

An Unconstrained Non-Shannon-Type Inequality (ZY98)

- The following unconstrained non-Shannon-type inequality was discovered for any 4 random variables (Zhang and Y, 1998):

$$
\begin{aligned}
& I(Z;U) - I(Z;U|X) - I(Z;U|Y) \\
\leq \quad & \frac{1}{2}I(X;Y) + \frac{1}{4}[I(X;Z,U) + I(Y;Z,U)]
\end{aligned}
$$

- This implies $\overline{\Gamma}_4^* \neq \Gamma_4$!

Proof Idea of ZY98

Proof Idea of ZY98

- For any X_1, X_2, X_3, X_4, construct two auxiliary random variables $\tilde{X}_1, \tilde{X}_2$ by

$$(X_1, X_2) \xleftarrow{p(x_1,x_2|x_3,x_4)} (X_3, X_4) \xrightarrow{p(x_1,x_2|x_3,x_4)} (\tilde{X}_1, \tilde{X}_2)$$

Proof Idea of ZY98

- For any X_1, X_2, X_3, X_4, construct two auxiliary random variables $\tilde{X}_1, \tilde{X}_2$ by

$$(X_1, X_2) \xleftarrow{p(x_1,x_2|x_3,x_4)} (X_3, X_4) \xrightarrow{p(x_1,x_2|x_3,x_4)} (\tilde{X}_1, \tilde{X}_2)$$

- Since $(X_1, X_2, X_3, X_4) \sim (\tilde{X}_1, \tilde{X}_2, X_3, X_4)$, we have

$$\begin{aligned} H(X_1) &= H(\tilde{X}_1) \\ H(X_1, X_2) &= H(\tilde{X}_1, \tilde{X}_2) \\ H(X_1, X_2, X_3) &= H(\tilde{X}_1, \tilde{X}_2, X_3) \\ &\cdots \end{aligned}$$

Proof Idea of ZY98

- For any X_1, X_2, X_3, X_4, construct two auxiliary random variables $\tilde{X}_1, \tilde{X}_2$ by

$$(X_1, X_2) \overset{p(x_1,x_2|x_3,x_4)}{\longleftarrow} (X_3, X_4) \overset{p(x_1,x_2|x_3,x_4)}{\longrightarrow} (\tilde{X}_1, \tilde{X}_2)$$

- Since $(X_1, X_2, X_3, X_4) \sim (\tilde{X}_1, \tilde{X}_2, X_3, X_4)$, we have

$$\begin{aligned} H(X_1) &= H(\tilde{X}_1) \\ H(X_1, X_2) &= H(\tilde{X}_1, \tilde{X}_2) \\ H(X_1, X_2, X_3) &= H(\tilde{X}_1, \tilde{X}_2, X_3) \\ &\cdots \end{aligned}$$

- Using all these constraints and invoking the basic inequalities for 6 random variables to obtain ZY98, which does not involve $\tilde{X}_1, \tilde{X}_2$!

Proof Idea of ZY98

- For any X_1, X_2, X_3, X_4, construct two auxiliary random variables $\tilde{X}_1, \tilde{X}_2$ by

$$(X_1, X_2) \overset{p(x_1,x_2|x_3,x_4)}{\longleftarrow} (X_3, X_4) \overset{p(x_1,x_2|x_3,x_4)}{\longrightarrow} (\tilde{X}_1, \tilde{X}_2)$$

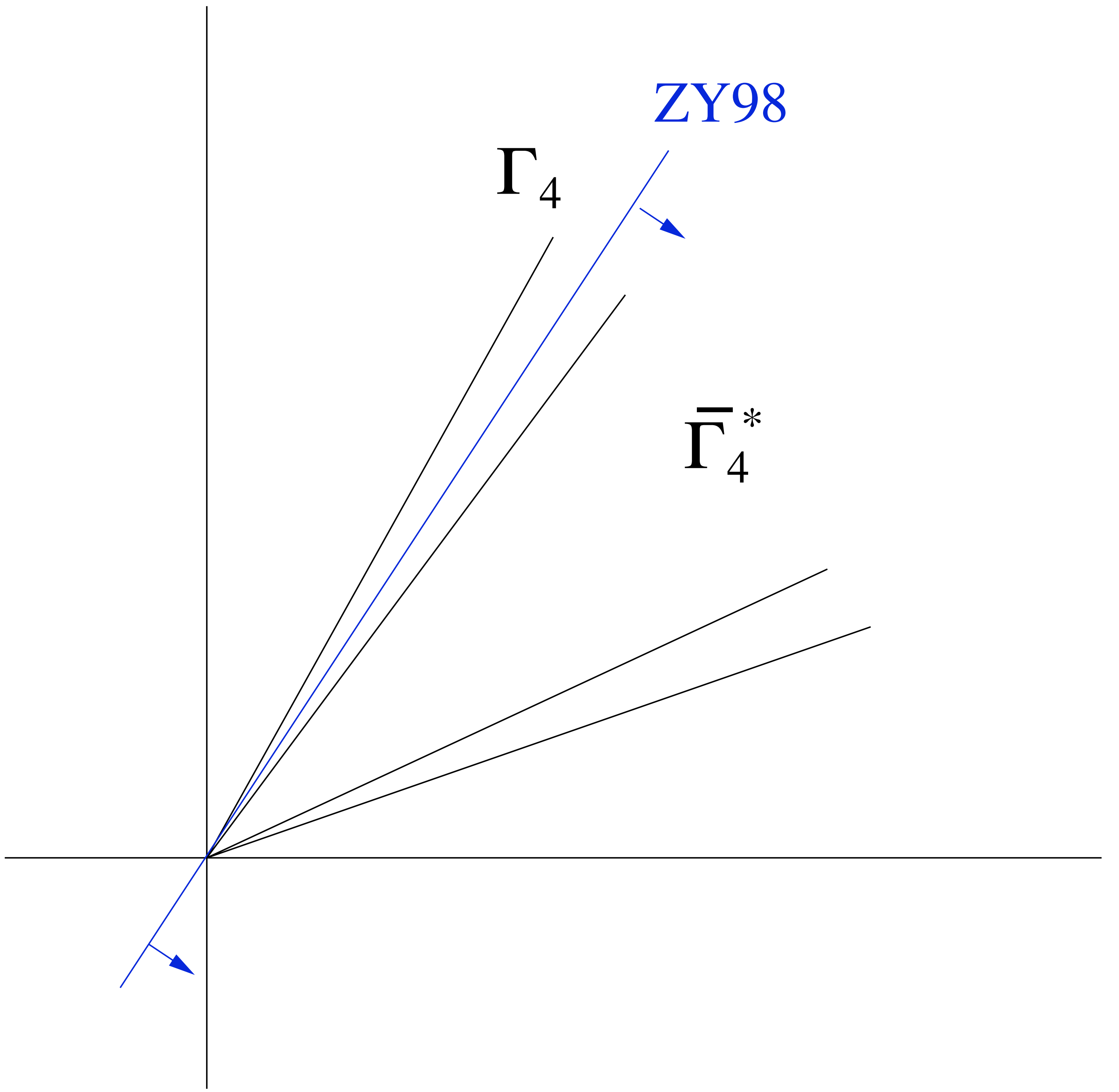
- Since $(X_1, X_2, X_3, X_4) \sim (\tilde{X}_1, \tilde{X}_2, X_3, X_4)$, we have

$$\begin{aligned} H(X_1) &= H(\tilde{X}_1) \\ H(X_1, X_2) &= H(\tilde{X}_1, \tilde{X}_2) \\ H(X_1, X_2, X_3) &= H(\tilde{X}_1, \tilde{X}_2, X_3) \\ &\cdots \end{aligned}$$

- Using all these constraints and invoking the basic inequalities for 6 random variables to obtain ZY98, which does not involve $\tilde{X}_1, \tilde{X}_2$!

- Can be proved by using ITIP.

An Illustration of ZY98



ITIP

ITIP

1.
```
>> ITIP('H(XYZ) <= H(X) + H(Y) + H(Z)')
True
```

ITIP

1. ```
   >> ITIP('H(XYZ) <= H(X) + H(Y) + H(Z)')
   True
   ```

2. ```
   >> ITIP('I(X;Z) = 0','I(X;Z|Y) = 0','I(X;Y) = 0')
   True
   ```

ITIP

1. `>> ITIP(’H(XYZ) <= H(X) + H(Y) + H(Z)’)`
   `True`

2. `>> ITIP(’I(X;Z) = 0’,’I(X;Z|Y) = 0’,’I(X;Y) = 0’)`
   `True`

3. `>> ITIP(’I(Z;U) - I(Z;U|X) - I(Z;U|Y) <=`
   ` 0.5 I(X;Y) + 0.25 I(X;ZU) + 0.25 I(Y;ZU)’)`
   `Not provable by ITIP`

ITIP

1.
```
>> ITIP('H(XYZ) <= H(X) + H(Y) + H(Z)')
True
```

2.
```
>> ITIP('I(X;Z) = 0','I(X;Z|Y) = 0','I(X;Y) = 0')
True
```

3.
```
>> ITIP('I(Z;U) - I(Z;U|X) - I(Z;U|Y) <=
 0.5 I(X;Y) + 0.25 I(X;ZU) + 0.25 I(Y;ZU)')
Not provable by ITIP
```

ZY98

1986

From nicholas@cs.ubc.ca Tue Jul 7 23:04:11 1998
X400-Received: by /PRMD=ca/ADMD=telecom.canada/C=ca/; Relayed; Tue, 7 Jul
1998
8:03:59 UTC-0700
Date: Tue, 7 Jul 1998 8:03:59 UTC-0700
X400-Originator: nicholas@cs.ubc.ca
X400-Recipients: non-disclosure:;
X400-Content-Type: P2-1984 (2)
X400-MTS-Identifier: [/PRMD=ca/ADMD=telecom.canada/C=ca/;980707080359]
Content-Identifier: 4429
X-UIDL: 900031892.048
From: Nicholas Pippenger <nicholas@cs.ubc.ca>
To: zzhang@milly.usc.edu, whyeung@ie.cuhk.edu.hk
MIME-Version: 1.0 (Generated by Ean X.400 to MIME gateway)

I have just seen your paper "On the Characterization of Entropy Function via Information Inequalities" in the IEEE Transactions on Information Theory. Please allow me to congratulate you on a most beautiful result! I worked on the problem of whether \overbar{\Gamma}^*_n = \Gamma_n during the 80s, without any success. I presented it as an open problem at the SPOC (Specific Problems on Communication and Computation) Conference in 1986--I believe there were proceedings published by Springer, but they seem to be out of print now.

It was wonderful to see your paper.

 - Nick Pippenger

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) $H_{R\cup S} + H_{R\cap S} \leq H_R + H_S$.

These conditions are immediate consequences of the fact that the logarithm vanishes at unity, is increasing and is concave. Are there any other conditions? If so, what are they? If not, show that any function satisfying (1), (2) and (3) can be approximated arbitrarily closely by the entropies of some family of random variables. (I say "approximated arbitrarily closely" to avoid the question of what happens on the boundary of the polytope defined by (1), (2) and (3).)

[H] K. T. Hu, "On the Amount of Information", *Theory of Prob. and Appl.*, 7 (1962) 439–447.

[R] F. M. Reza, *An Introduction to Information Theory*, McGraw-Hill, New York, 1961.

Other Non-Shannon-Type Inequalities

Other Non-Shannon-Type Inequalities

- Matúš (1999) proved the Ingleton inequality under the constraints $X_1 \perp X_2 | X_3$ and $X_2 \perp X_4 | X_3$ ($X_1 \perp X_2$ for ZY98).

Other Non-Shannon-Type Inequalities

- Matúš (1999) proved the Ingleton inequality under the constraints $X_1 \perp X_2|X_3$ and $X_2 \perp X_4|X_3$ ($X_1 \perp X_2$ for ZY98).

- ZY98 have been further generalized by Makarychev et al. (2002), Zhang (2003), and Matúš (2007).

Other Non-Shannon-Type Inequalities

- Matúš (1999) proved the Ingleton inequality under the constraints $X_1 \perp X_2 | X_3$ and $X_2 \perp X_4 | X_3$ ($X_1 \perp X_2$ for ZY98).

- ZY98 have been further generalized by Makarychev et al. (2002), Zhang (2003), and Matúš (2007).

- In particular, Matúš showed that $\overline{\Gamma}^*_n$ is not a polyhedral cone, and hence there exist an infinitely number of linear non-Shannon-type inequalities!

Other Non-Shannon-Type Inequalities

- Matúš (1999) proved the Ingleton inequality under the constraints $X_1 \perp X_2 | X_3$ and $X_2 \perp X_4 | X_3$ ($X_1 \perp X_2$ for ZY98).

- ZY98 have been further generalized by Makarychev et al. (2002), Zhang (2003), and Matúš (2007).

- In particular, Matúš showed that $\overline{\Gamma}^*_n$ is not a polyhedral cone, and hence there exist an infinitely number of linear non-Shannon-type inequalities!

- Dougherty, Freiling and Zeger (2006) have discovered several tens of non-Shannon-type inequalities by a search on a supercomputer.

Other Non-Shannon-Type Inequalities

- Matúš (1999) proved the Ingleton inequality under the constraints $X_1 \perp X_2 | X_3$ and $X_2 \perp X_4 | X_3$ ($X_1 \perp X_2$ for ZY98).
- ZY98 have been further generalized by Makarychev et al. (2002), Zhang (2003), and Matúš (2007).
- In particular, Matúš showed that $\overline{\Gamma}^*_n$ is not a polyhedral cone, and hence there exist an infinitely number of linear non-Shannon-type inequalities!
- Dougherty, Freiling and Zeger (2006) have discovered several tens of non-Shannon-type inequalities by a search on a supercomputer.

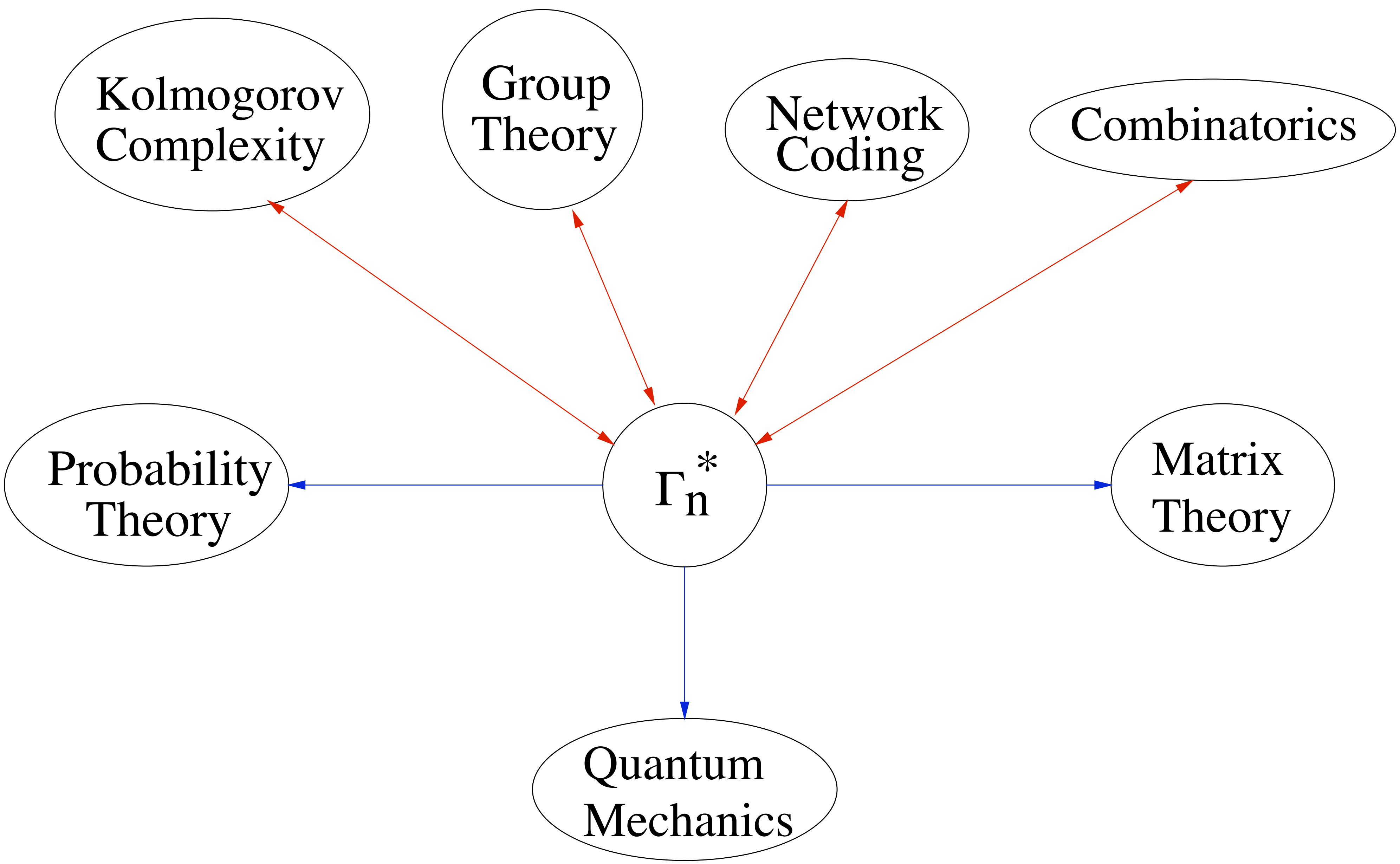
Remark

- Matúš (2007) also showed a fundamental property of Γ^*_n:

$$int(cl(\Gamma^*_n)) \subset \Gamma^*_n$$

i.e., Γ^*_n has a solid core.

Subjects Related to Γ_n^*



COMBINATORICS

2-D Quasi-Uniform Array

2-D Quasi-Uniform Array

- For a distribution $p(x)$, a sequence $\mathbf{x}$ of length n is strongly typical if the empirical distribution of $\mathbf{x}$ is approximately equal to $p(x)$.

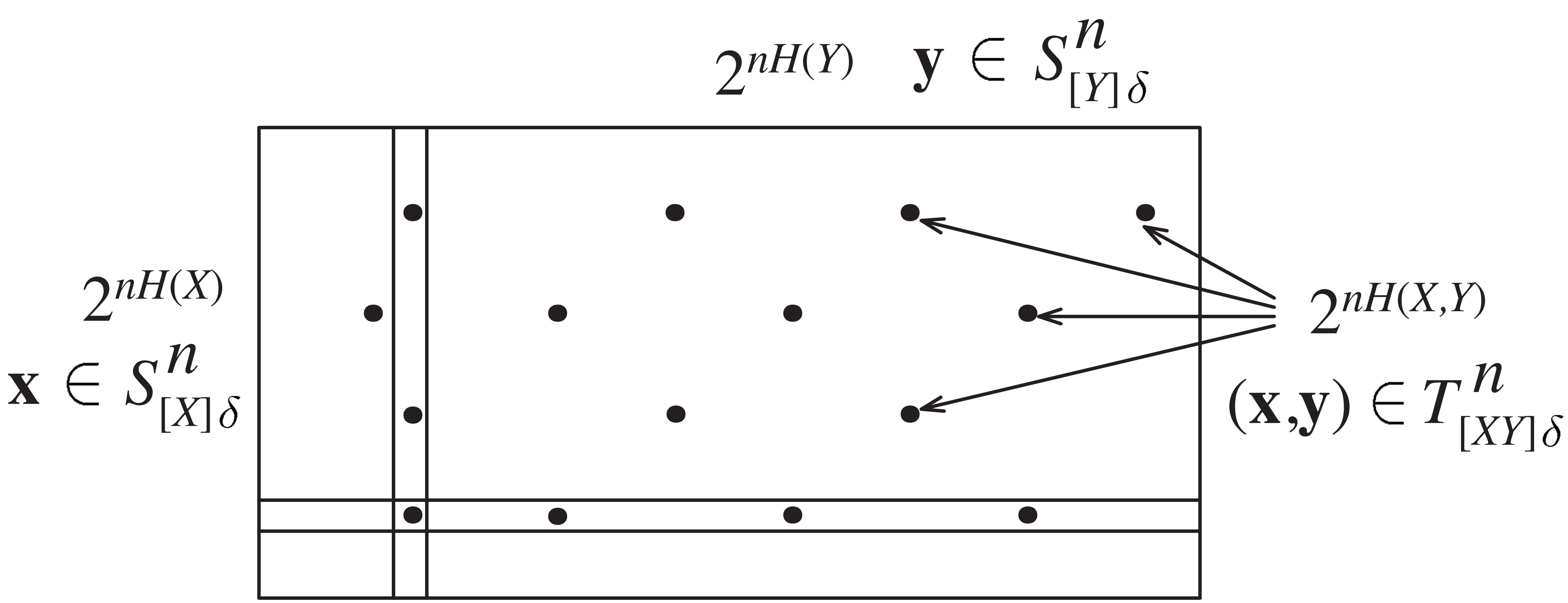
2-D Quasi-Uniform Array

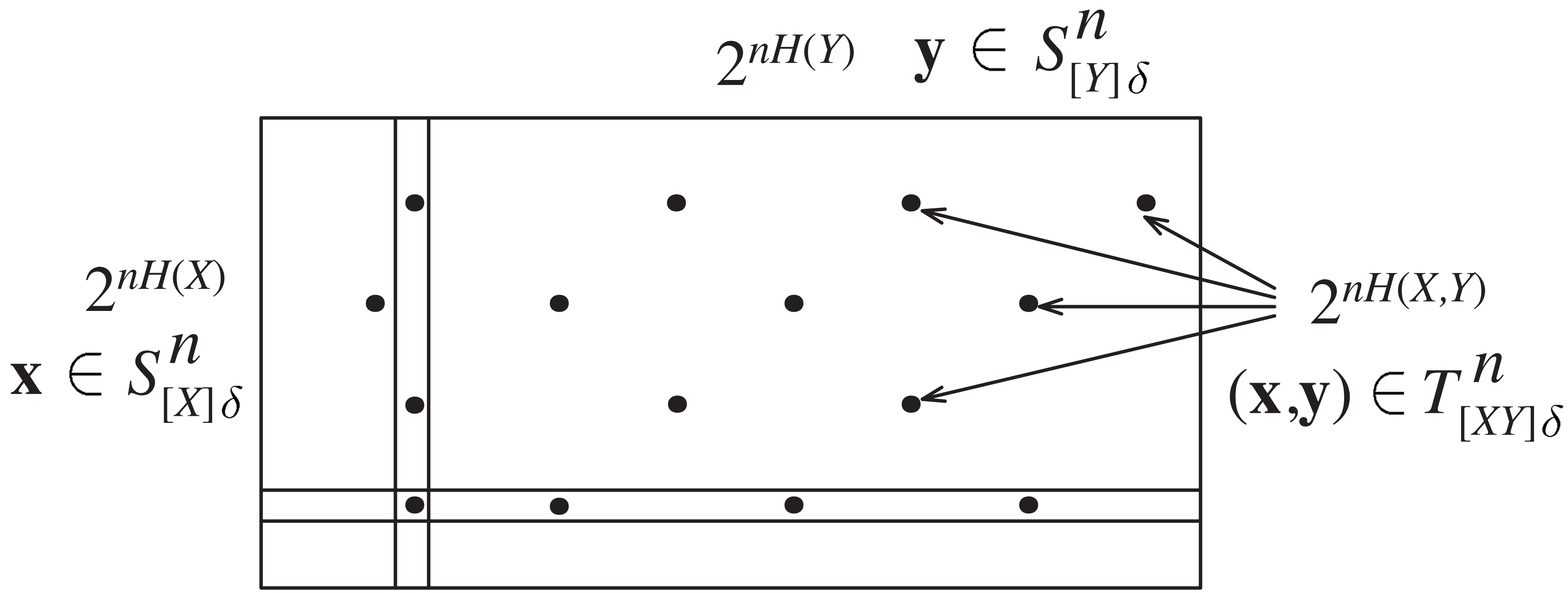
- For a distribution $p(x)$, a sequence $\mathbf{x}$ of length n is strongly typical if the empirical distribution of $\mathbf{x}$ is approximately equal to $p(x)$.

- Let $p(x, y)$ be a joint distribution. The strongly typical sequences w.r.t. $p(x, y)$, $p(x)$, and $p(y)$ can be illustrated by a 2-D quasi-uniform array.

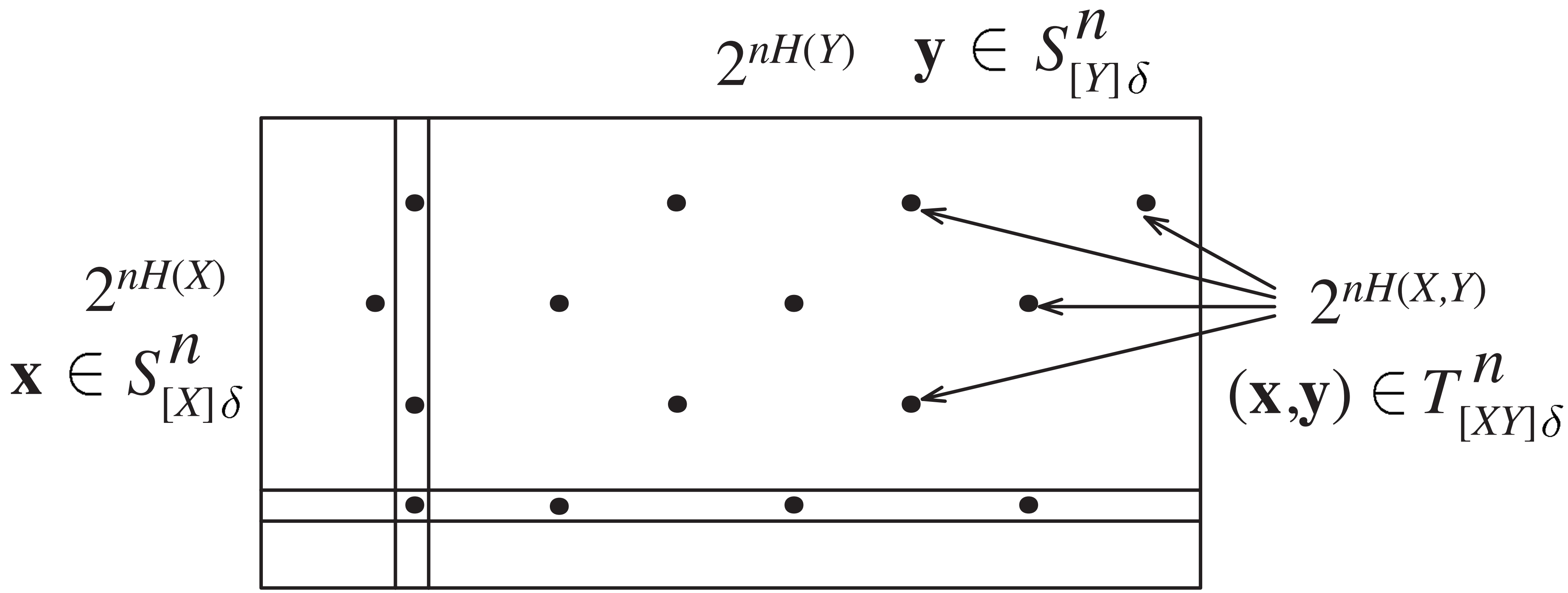
2-D Quasi-Uniform Array

- For a distribution $p(x)$, a sequence $\mathbf{x}$ of length n is strongly typical if the empirical distribution of $\mathbf{x}$ is approximately equal to $p(x)$.
- Let $p(x, y)$ be a joint distribution. The strongly typical sequences w.r.t. $p(x, y)$, $p(x)$, and $p(y)$ can be illustrated by a 2-D quasi-uniform array.

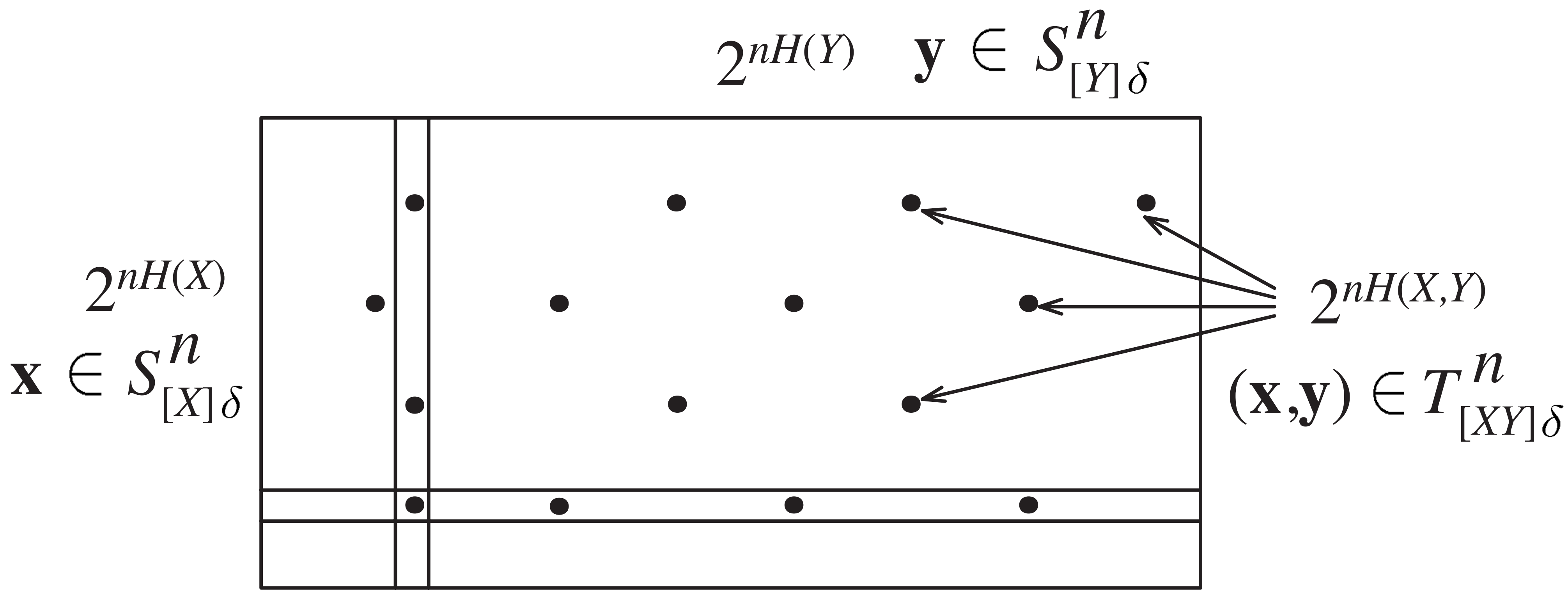




$2^{nH(Y)}$ $\mathbf{y} \in S^n_{[Y]\delta}$
$2^{nH(X)}$
$\mathbf{x} \in S^n_{[X]\delta}$
$2^{nH(X,Y)}$
$(\mathbf{x},\mathbf{y}) \in T^n_{[XY]\delta}$

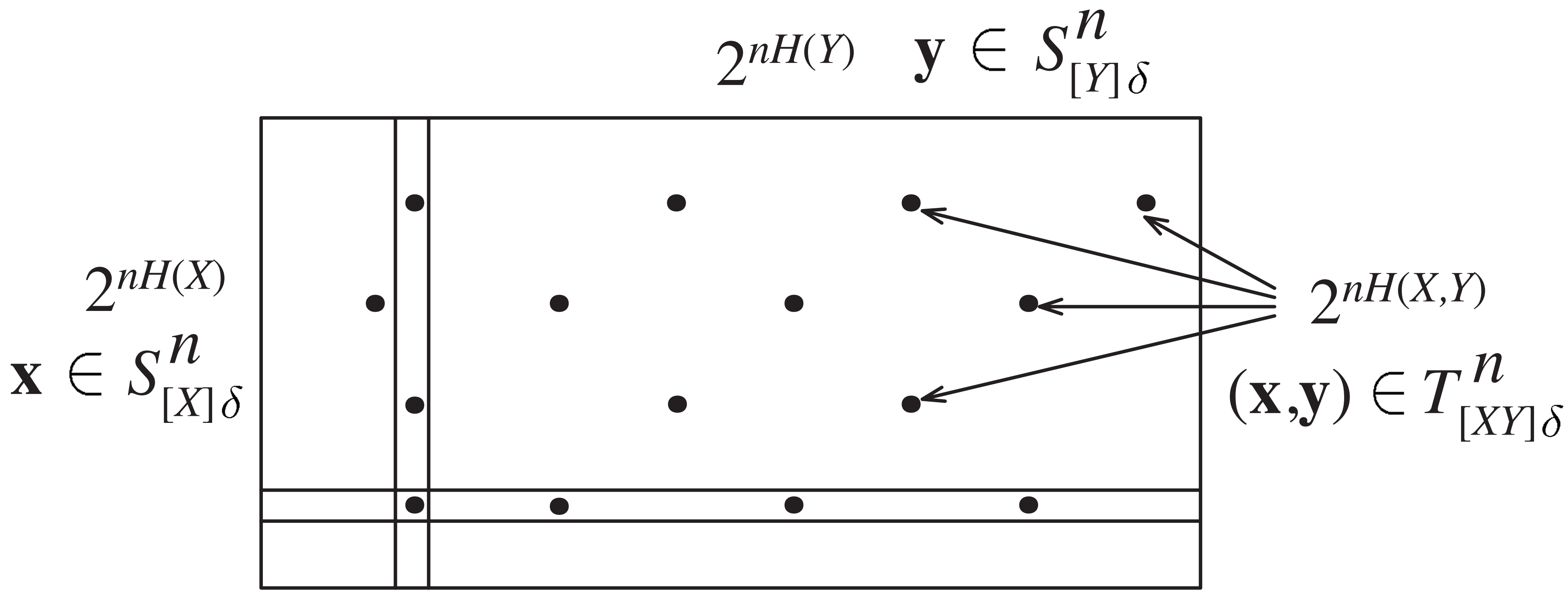


- Each row has approximately the same number of dots ($\sim 2^{nH(Y|X)}$) and each column has approximately the same number of dots ($\sim 2^{nH(X|Y)}$).



- Each row has approximately the same number of dots ($\sim 2^{nH(Y|X)}$) and each column has approximately the same number of dots ($\sim 2^{nH(X|Y)}$).
- Thus

$$2^{nH(X,Y)} \leq 2^{nH(X)} 2^{nH(Y)} \Rightarrow H(X,Y) \leq H(X) + H(Y)$$



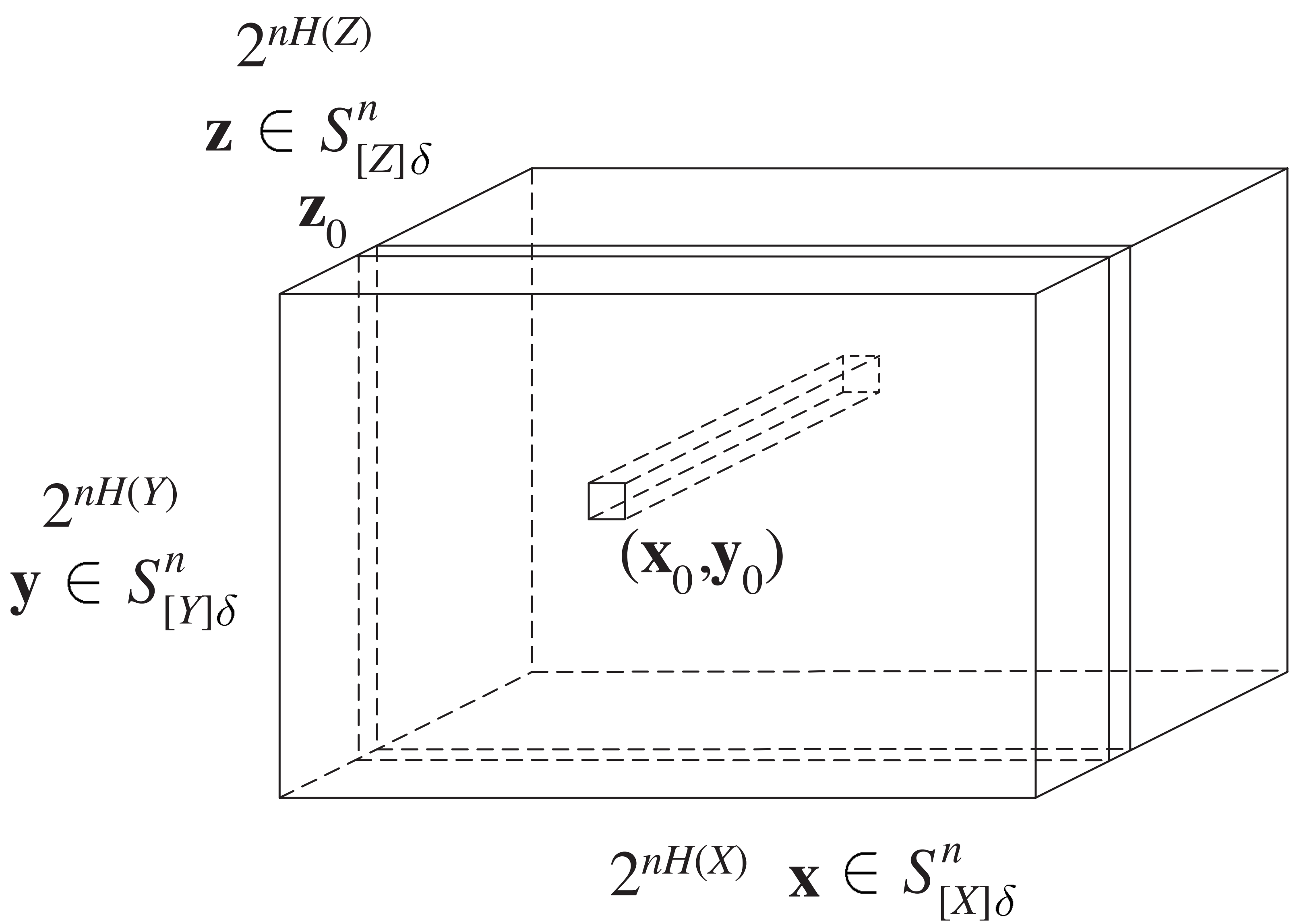
- Each row has approximately the same number of dots ($\sim 2^{nH(Y|X)}$) and each column has approximately the same number of dots ($\sim 2^{nH(X|Y)}$).

- Thus

$$2^{nH(X,Y)} \leq 2^{nH(X)} 2^{nH(Y)} \Rightarrow H(X,Y) \leq H(X) + H(Y)$$

- Then the basic inequality $I(X;Y) \geq 0$ is about the unfilled entries in the array.

3-D Quasi-Uniform Array



Quasi-Uniform Arrays and Entropy Inequalities

Quasi-Uniform Arrays and Entropy Inequalities

- For an n-dimensional quasi-uniform array, if all the “dots” are assigned equal probabilities, then the projection on every lower dimensional plane has a uniform distribution over its support.

Quasi-Uniform Arrays and Entropy Inequalities

- For an n-dimensional quasi-uniform array, if all the “dots” are assigned equal probabilities, then the projection on every lower dimensional plane has a uniform distribution over its support.

- Do quasi-uniform arrays fully capture all constraints on the entropy function?

Quasi-Uniform Arrays and Entropy Inequalities

- For an n-dimensional quasi-uniform array, if all the “dots” are assigned equal probabilities, then the projection on every lower dimensional plane has a uniform distribution over its support.

- Do quasi-uniform arrays fully capture all constraints on the entropy function?

- YES. T. Chan (2001) showed that all constraints on the entropy function can be obtained through quasi-uniform arrays, and vice versa.

GROUP THEORY

Entropy and Groups
(Chan-Y 2002)

Entropy and Groups
(Chan-Y 2002)

- Let G be a finite group and $G_1, G_2, \ldots, G_n$ be subgroups of G.

Entropy and Groups
(Chan-Y 2002)

- Let G be a finite group and $G_1, G_2, \ldots, G_n$ be subgroups of G.
- Let $G_\alpha = \cap_{i \in \alpha} G_i$, also a subgroup.

Entropy and Groups
(Chan-Y 2002)

- Let G be a finite group and $G_1, G_2, \ldots, G_n$ be subgroups of G.
- Let $G_\alpha = \cap_{i \in \alpha} G_i$, also a subgroup.
- A probability distribution for n random variables $X_1, X_2, \ldots, X_n$ can be constructed from any finite group G and subgroups $G_1, G_2, \ldots, G_n$, with

$$H(X_\alpha) = \log \frac{|G|}{|G_\alpha|}$$

  which depends only on the orders of G and $G_1, G_2, \ldots, G_n$.

Entropy and Groups

Entropy and Groups

- Substituting the joint entropies into any entropy inequality gives a group inequality.

Entropy and Groups

- Substituting the joint entropies into any entropy inequality gives a group inequality.
- For example, for any X_1, X_2,

$$H(X_1) + H(X_2) \geq H(X_1, X_2)$$

corresponds to for any finite group G and subgroups G_1, G_2,

$$\log \frac{|G|}{|G_1|} + \log \frac{|G|}{|G_2|} \geq \log \frac{|G|}{|G_1 \cap G_2|}$$

or

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

Non-Shannon-Type Group Inequalities

Non-Shannon-Type Group Inequalities

- "Non-Shannon-type" group inequalities can be obtained accordingly.

Non-Shannon-Type Group Inequalities

- "Non-Shannon-type" group inequalities can be obtained accordingly.
- For example, ZY98 can be written as

$$\left.\begin{array}{r} H(X_1)+H(X_1,X_2)+2H(X_3) \\ +2H(X_4)+4H(X_1,X_3,X_4) \\ +H(X_2,X_3,X_4) \end{array}\right\} \leq \left\{\begin{array}{l} 3H(X_1,X_3)+3H(X_1,X_4) \\ +3H(X_3,X_4)+H(X_2,X_3) \\ +H(X_2,X_4) \end{array}\right.$$

Non-Shannon-Type Group Inequalities

- "Non-Shannon-type" group inequalities can be obtained accordingly.
- For example, ZY98 can be written as

$$\left.\begin{array}{r} H(X_1)+H(X_1,X_2)+2H(X_3) \\ +2H(X_4)+4H(X_1,X_3,X_4) \\ +H(X_2,X_3,X_4) \end{array}\right\} \leq \left\{\begin{array}{l} 3H(X_1,X_3)+3H(X_1,X_4) \\ +3H(X_3,X_4)+H(X_2,X_3) \\ +H(X_2,X_4) \end{array}\right.$$

- This corresponds to

$$\left.\begin{array}{r} |G_1||G_1\cap G_2||G_3|^2 \\ \cdot\ |G_4|^2|G_1\cap G_3\cap G_4|^4 \\ \cdot\ |G_2\cap G_3\cap G_4| \end{array}\right\} \geq \left\{\begin{array}{l} |G_1\cap G_3|^3|G_1\cap G_4|^3 \\ \cdot\ |G_3\cap G_4|^3|G_2\cap G_3| \\ \cdot\ |G_2\cap G_4| \end{array}\right.$$

Non-shannon-Type Group Inequalities

- "Non-Shannon-type" group inequalities can be obtained accordingly.
- For example, ZY98 can be written as

$$\left.\begin{array}{r} H(X_1)+H(X_1,X_2)+2H(X_3) \\ +2H(X_4)+4H(X_1,X_3,X_4) \\ +H(X_2,X_3,X_4) \end{array}\right\} \le \left\{\begin{array}{l} 3H(X_1,X_3)+3H(X_1,X_4) \\ +3H(X_3,X_4)+H(X_2,X_3) \\ +H(X_2,X_4) \end{array}\right.$$

- This corresponds to

$$\left.\begin{array}{r} |G_1||G_1\cap G_2||G_3|^2 \\ \cdot\ |G_4|^2|G_1\cap G_3\cap G_4|^4 \\ \cdot\ |G_2\cap G_3\cap G_4| \end{array}\right\} \ge \left\{\begin{array}{l} |G_1\cap G_3|^3|G_1\cap G_4|^3 \\ \cdot\ |G_3\cap G_4|^3|G_2\cap G_3| \\ \cdot\ |G_2\cap G_4| \end{array}\right.$$

- It can be proved that the correspondence between entropy inequalities and group inequalities is one-to-one.

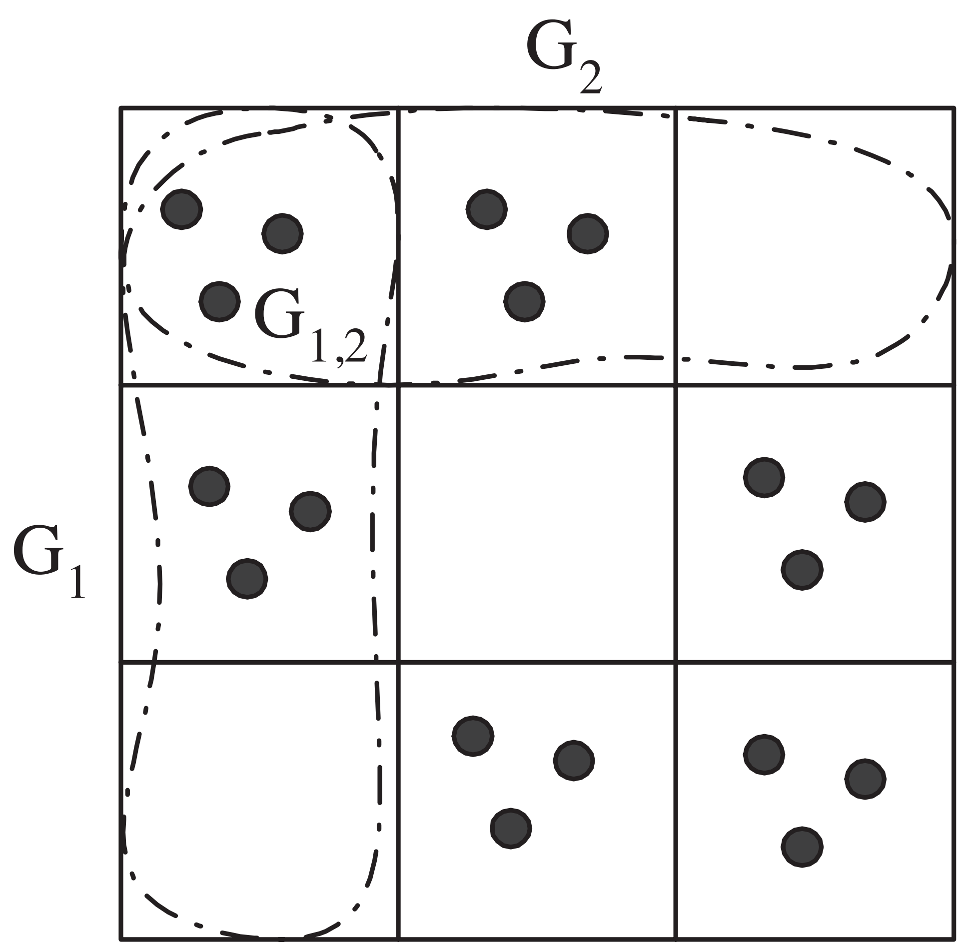
Relation between Finite Group and Quasi-Uniform Array

Relation between Finite Group and Quasi-Uniform Array

- The distribution of the elements of a finite group among its subgroups exhibits a quasi-uniform structure.

Relation between Finite Group and Quasi-Uniform Array

- The distribution of the elements of a finite group among its subgroups exhibits a quasi-uniform structure.



PROBABILITY THEORY

Compatibility of Conditional Independence

Compatibility of Conditional Independence

- The Implication Problem Is a given conditional independency implied by a given set of conditional independencies?

Compatibility of Conditional Independence

- The Implication Problem Is a given conditional independency implied by a given set of conditional independencies?

- Example

$$\left.\begin{array}{l} X \rightarrow Y \rightarrow Z \\ X \perp Y \end{array}\right\} \Rightarrow X \perp Z$$

Compatibility of Conditional Independence

- The Implication Problem Is a given conditional independency implied by a given set of conditional independencies?

- Example

$$\left.\begin{array}{l} X \to Y \to Z \\ X \perp Y \end{array}\right\} \Rightarrow X \perp Z$$

- A very basic problem in probability theory.

Compatibility of Conditional Independence

Compatibility of Conditional Independence

- This is a subproblem of characterizing Γ_n^* because a conditional independence relation is just a hyperplane in $\mathcal{H}_n$.

Compatibility of Conditional Independence

- This is a subproblem of characterizing Γ_n^* because a conditional independence relation is just a hyperplane in $\mathcal{H}_n$.
- For example, $X \perp Y | Z \Leftrightarrow I(X;Y|Z) = 0$.

Compatibility of Conditional Independence

- This is a subproblem of characterizing Γ_n^* because a conditional independence relation is just a hyperplane in $\mathcal{H}_n$.
- For example, $X \perp Y|Z \Leftrightarrow I(X;Y|Z) = 0$.
- Thus the conditional independence problem is

A discrete problem imbedded in a continuous problem.

Compatibility of Conditional Independence

- This is a subproblem of characterizing Γ_n^* because a conditional independence relation is just a hyperplane in $\mathcal{H}_n$.
- For example, $X \perp Y|Z \Leftrightarrow I(X;Y|Z) = 0$.
- Thus the conditional independence problem is

A discrete problem imbedded in a continuous problem.

- The more general p-representability problem was studied by Matúš and Studený (1995).

Compatibility of Conditional Independence

- This is a subproblem of characterizing Γ_n^* because a conditional independence relation is just a hyperplane in $\mathcal{H}_n$.
- For example, $X \perp Y|Z \Leftrightarrow I(X;Y|Z) = 0$.
- Thus the conditional independence problem is

A discrete problem imbedded in a continuous problem.

- The more general p-representability problem was studied by Matúš and Studený (1995).
- $n = 4$ was settled by F. Matúš (1999) by proving a constrained version of the Ingleton inequality (non-Shannon-type).

Compatibility of Conditional Independence

- This is a subproblem of characterizing Γ_n^* because a conditional independence relation is just a hyperplane in $\mathcal{H}_n$.

- For example, $X \perp Y|Z \Leftrightarrow I(X;Y|Z) = 0$.

- Thus the conditional independence problem is

A discrete problem imbedded in a continuous problem.

- The more general p-representability problem was studied by Matúš and Studený (1995).

- $n = 4$ was settled by F. Matúš (1999) by proving a constrained version of the Ingleton inequality (non-Shannon-type).

- Very hard for $n \geq 4$.

KOLMOGOROV COMPLEXITY

Entropy and Kolmogorov Complexity

Entropy and Kolmogorov Complexity

- Let $K(\cdot)$ denotes the Kolmogorov complexity of a collection of sequences.

Entropy and Kolmogorov Complexity

- Let $K(\cdot)$ denotes the Kolmogorov complexity of a collection of sequences.
- Hammer et al. (2000) showed that there exists a one-to-one correspondence between entropy inequalities and Kolmogorov complexity inequalities.

Entropy and Kolmogorov Complexity

- Let $K(\cdot)$ denotes the Kolmogorov complexity of a collection of sequences.
- Hammer et al. (2000) showed that there exists a one-to-one correspondence between entropy inequalities and Kolmogorov complexity inequalities.
- For example, for any X_1, X_2,

$$H(X_1) + H(X_2) \geq H(X_1, X_2)$$

corresponds to for any two sequences x and y,

$$K(x_1) + K(x_2) \geq K(x_1, x_2)$$

Entropy and Kolmogorov Complexity

- Let $K(\cdot)$ denotes the Kolmogorov complexity of a collection of sequences.
- Hammer et al. (2000) showed that there exists a one-to-one correspondence between entropy inequalities and Kolmogorov complexity inequalities.
- For example, for any X_1, X_2,

$$H(X_1) + H(X_2) \geq H(X_1, X_2)$$

corresponds to for any two sequences x and y,

$$K(x_1) + K(x_2) \geq K(x_1, x_2)$$

- "Non-Shannon-type" Kolmogorov complexity inequalities can be obtained accordingly.

NETWORK CODING

Entropy and Network Coding

Entropy and Network Coding

- For single-source network coding, the network capacity is completely characterized by the maximum flows in the network.

Entropy and Network Coding

- For single-source network coding, the network capacity is completely characterized by the maximum flows in the network.

- For multi-source network coding, the network capacity can be characterized by Γ_n^*.

Entropy and Network Coding

- For single-source network coding, the network capacity is completely characterized by the maximum flows in the network.
- For multi-source network coding, the network capacity can be characterized by Γ_n^*.
- The problem has been studied by Y and Zhang (1999), Song, Y, and Cai (2006). Yan, Y and Zhang (2007, 2012) finally obtained a complete characterization (implicit) of the network capacity in terms of Γ_n^*.

Entropy and Network Coding

Entropy and Network Coding

- Dougherty, Freiling and Zeger (2007) constructed the first multi-source network coding example whose characterization of the network capacity requires ZY98.

Entropy and Network Coding

- Dougherty, Freiling and Zeger (2007) constructed the first multi-source network coding example whose characterization of the network capacity requires ZY98.

- The construction is based on the Vámos matroid.

Entropy and Network Coding

- Dougherty, Freiling and Zeger (2007) constructed the first multi-source network coding example whose characterization of the network capacity requires ZY98.

- The construction is based on the Vámos matroid.

- Chan and Grant (2008) proved a duality between almost entropy functions and multi-source network coding problems.

Entropy and Network Coding

- Dougherty, Freiling and Zeger (2007) constructed the first multi-source network coding example whose characterization of the network capacity requires ZY98.

- The construction is based on the Vámos matroid.

- Chan and Grant (2008) proved a duality between almost entropy functions and multi-source network coding problems.

- Thus

Every constraint on the entropy function is useful in
some multi-source network coding problems!

Secret Sharing

Secret Sharing

- Secret sharing in cryptography was introduced independently by Blakley and Shamir (1979).

Secret Sharing

- Secret sharing in cryptography was introduced independently by Blakley and Shamir (1979).
- Recently, Beimel et al. (2008) has applied ZY98 to obtain a new performance bound in secret sharing.

Secret Sharing

- Secret sharing in cryptography was introduced independently by Blakley and Shamir (1979).
- Recently, Beimel et al. (2008) has applied ZY98 to obtain a new performance bound in secret sharing.
- Secret sharing can be regarded as a special case of secure network coding (Cai and Y, 2002).

MATRIX THEORY

Differential Entropy Inequalities

Differential Entropy Inequalities

- Differential Entropy

$$h(X) = -\int f(x) \log f(x) dx$$

Differential Entropy Inequalities

- Differential Entropy

$$h(X) = -\int f(x) \log f(x) dx$$

- Joint Differential Entropy

$$h(X, Y) = -\int f(x, y) \log f(x, y) dx dy$$

Differential Entropy Inequalities

- Differential Entropy

$$h(X) = -\int f(x) \log f(x) dx$$

- Joint Differential Entropy

$$h(X, Y) = -\int f(x, y) \log f(x, y) dx dy$$

- Chan (2003) showed that a differential entropy inequality is valid iff the coefficients of the random variables are *balanced* and its discrete counterpart is valid.

Balanced Entropy Inequalities

Balanced Entropy Inequalities

- For example,

$$h(X|Y) = h(X,Y) - h(Y) \geq 0$$

is not valid because the coefficients are not balanced.

Balanced Entropy Inequalities

- For example,
$$h(X|Y) = h(X,Y) - h(Y) \geq 0$$
is not valid because the coefficients are not balanced.

- On the other hand,
$$I(X;Y) = h(X) + h(Y) - h(X,Y) \geq 0$$
is valid.

Balanced Entropy Inequalities

- For example,
$$h(X|Y) = h(X,Y) - h(Y) \geq 0$$
is not valid because the coefficients are not balanced.

- On the other hand,
$$I(X;Y) = h(X) + h(Y) - h(X,Y) \geq 0$$
is valid.

- The coefficients in ZY98 are balanced, so it is also valid for differential entropy.

Gaussian Distribution

Gaussian Distribution

- Any (symmetric) positive definite matrix is a valid covariance matrix, so that it defines the joint pdf of a Gaussian random vector

$$\mathbf{X} = [\ X_1\ X_2\ \cdots\ X_n\ ].$$

Gaussian Distribution

- Any (symmetric) positive definite matrix is a valid covariance matrix, so that it defines the joint pdf of a Gaussian random vector

$$\mathbf{X} = [\ X_1\ X_2\ \cdots\ X_n\ ].$$

- Then

$$h(\mathbf{X}) = \frac{1}{2}\log\left[(2\pi e)^n |K|\right]$$

and for any subset α of $\{1, 2, \ldots, n\}$,

$$h(\mathbf{X}_\alpha) = \frac{1}{2}\log\left[(2\pi e)^{|\alpha|} |K_\alpha|\right]$$

where K_α is the corresponding submatrix of K.

Non-Shannon-Type Matrix Inequalities

Non-Shannon-Type Matrix Inequalities

- Substituting these joint differential entropies into the inequality

$$h(X_1, X_2, \ldots, X_n) \leq \sum_i h(X_i)$$

gives the Hadamard inequality

$$|K| \leq \prod_i |K_i| = \prod_i k_{ii}$$

Non-Shannon-Type Matrix Inequalities

- Substituting these joint differential entropies into the inequality

$$h(X_1, X_2, \ldots, X_n) \leq \sum_i h(X_i)$$

gives the Hadamard inequality

$$|K| \leq \prod_i |K_i| = \prod_i k_{ii}$$

- Substituting these joint differential entropies into ZY98 gives

$$\begin{aligned} & |K_1||K_{12}||K_3|^2|K_4|^2|K_{134}|^4|K_{234}| \\ \leq \quad & |K_{13}|^3|K_{14}|^3|K_{34}|^3|K_{23}||K_{24}| \end{aligned}$$

for all positive definite matrices.

Non-Shannon-Type Matrix Inequalities

- Substituting these joint differential entropies into the inequality

$$h(X_1, X_2, \ldots, X_n) \leq \sum_i h(X_i)$$

gives the Hadamard inequality

$$|K| \leq \prod_i |K_i| = \prod_i k_{ii}$$

- Substituting these joint differential entropies into ZY98 gives

$$\begin{aligned} & |K_1||K_{12}||K_3|^2|K_4|^2|K_{134}|^4|K_{234}| \\ \leq \quad & |K_{13}|^3|K_{14}|^3|K_{34}|^3|K_{23}||K_{24}| \end{aligned}$$

for all positive definite matrices.

- Many other “non-Shannon-type” inequalities of the principal minors of positive definite matrices can be obtained this way.

Principal Minor Inequalities

Principal Minor Inequalities

- Can all constraints on the principal minors of positive definite matrices be obtained through differential entropy inequalities and the Gaussian distribution?

Principal Minor Inequalities

- Can all constraints on the principal minors of positive definite matrices be obtained through differential entropy inequalities and the Gaussian distribution?
- Chan, Guo, and Y (2012) showed that this is the case for 3×3 positive definite matrices.

Principal Minor Inequalities

- Can all constraints on the principal minors of positive definite matrices be obtained through differential entropy inequalities and the Gaussian distribution?
- Chan, Guo, and Y (2012) showed that this is the case for 3×3 positive definite matrices.
- Previously, Hassibi and Shadbakht (2008) studied normalized Gaussian entropy functions and obtained a characterization for 3 Gaussian random variables.

QUANTUM MECHANICS

The von Neumann Entropy

The von Neumann Entropy

- The von Neumann entropy is an extension of the Shannon entropy to quantum mechanics.

The von Neumann Entropy

- The von Neumann entropy is an extension of the Shannon entropy to quantum mechanics.

- The strong subadditivity of the von Neumann entropy (analogous to the basic inequalities for the Shannon inequalities) was proved by Lieb ad Ruskai (1973).

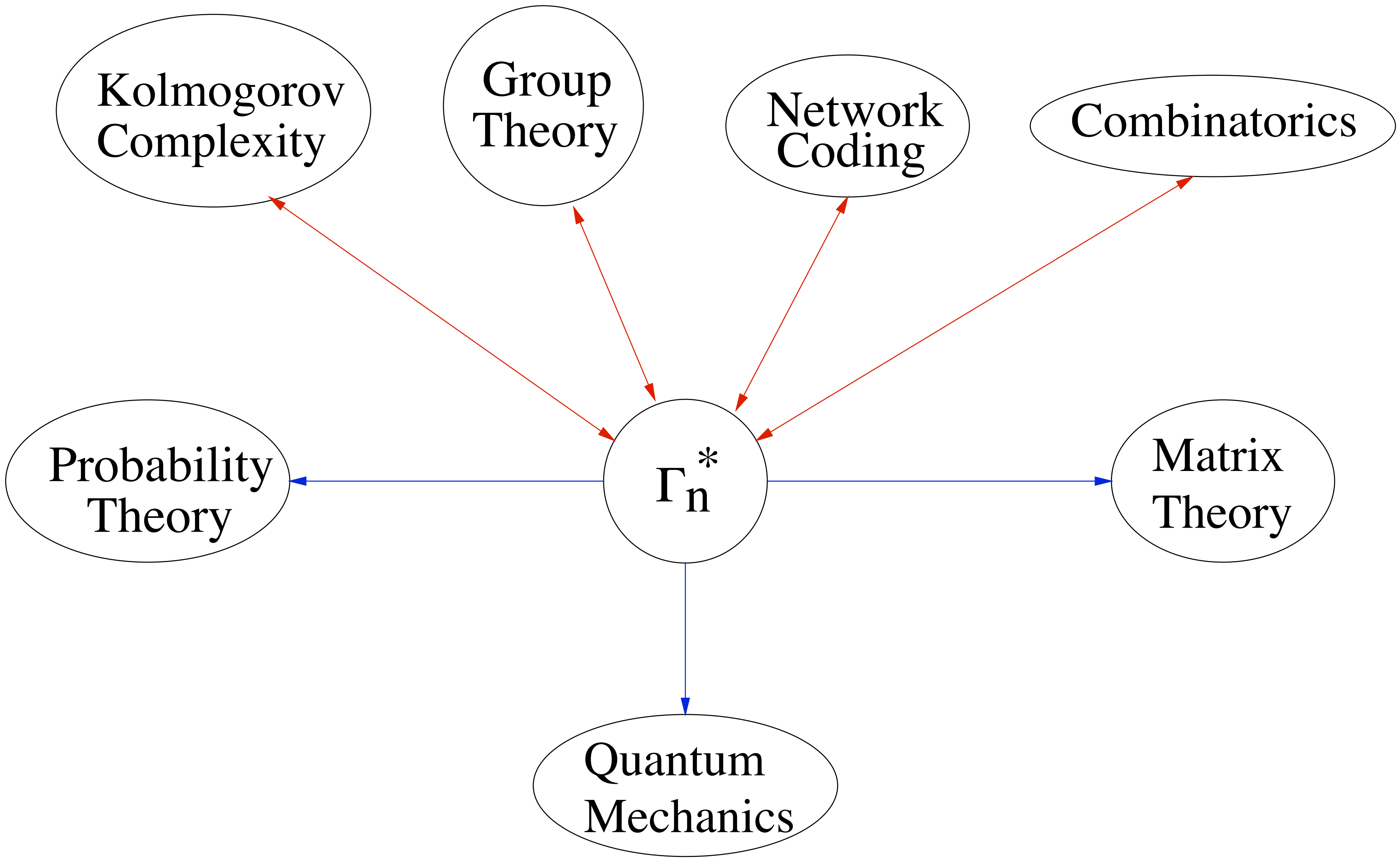
The von Neumann Entropy

- The von Neumann entropy is an extension of the Shannon entropy to quantum mechanics.
- The strong subadditivity of the von Neumann entropy (analogous to the basic inequalities for the Shannon inequalities) was proved by Lieb ad Ruskai (1973).
- Inspired by the discovery of non-Shannon-type inequalities, Pippenger (2003) proved that for a 3-party system, there exists no inequality for the von Neumann entropy beyond strong subadditivity.

The von Neumann Entropy

- The von Neumann entropy is an extension of the Shannon entropy to quantum mechanics.
- The strong subadditivity of the von Neumann entropy (analogous to the basic inequalities for the Shannon inequalities) was proved by Lieb ad Ruskai (1973).
- Inspired by the discovery of non-Shannon-type inequalities, Pippenger (2003) proved that for a 3-party system, there exists no inequality for the von Neumann entropy beyond strong subadditivity.
- Linden and Winter (2005) discovered for a 4-party system a constrained inequality for the von Neumann entropy which is independent of strong subadditivity.

Summary


Kolmogorov Complexity
Group Theory
Network Coding
Combinatorics
Probability Theory
Γ_n^*
Matrix Theory
Quantum Mechanics

Concluding Remarks

Concluding Remarks

- The entropy function connects together a number of seemingly unrelated fields in information sciences, mathematics, and physics.

Concluding Remarks

- The entropy function connects together a number of seemingly unrelated fields in information sciences, mathematics, and physics.
- There exist one-to-one correspondences among the entropy function, group theory, Kolmogorov complexity, and network coding, suggesting that they share the same underlying structure.

Concluding Remarks

- The entropy function connects together a number of seemingly unrelated fields in information sciences, mathematics, and physics.
- There exist one-to-one correspondences among the entropy function, group theory, Kolmogorov complexity, and network coding, suggesting that they share the same underlying structure.
- The relations between these fields need a deeper understanding. The quasi-uniform array plays a central role.

Concluding Remarks

- The entropy function connects together a number of seemingly unrelated fields in information sciences, mathematics, and physics.
- There exist one-to-one correspondences among the entropy function, group theory, Kolmogorov complexity, and network coding, suggesting that they share the same underlying structure.
- The relations between these fields need a deeper understanding. The quasi-uniform array plays a central role.
- "Non-Shannon-type" inequalities in different fields need further understanding.

Concluding Remarks

- The entropy function connects together a number of seemingly unrelated fields in information sciences, mathematics, and physics.
- There exist one-to-one correspondences among the entropy function, group theory, Kolmogorov complexity, and network coding, suggesting that they share the same underlying structure.
- The relations between these fields need a deeper understanding. The quasi-uniform array plays a central role.
- "Non-Shannon-type" inequalities in different fields need further understanding.
- Information theory is clearly an integral part of mathematics.

Can we get out of this mess?

Facets of Entropy

Raymond W. Yeung*

October 4, 2012

Constraints on the entropy function are of fundamental importance in information theory. For a long time, the polymatroidal axioms, or equivalently the nonnegativity of the Shannon information measures, are the only known constraints. Inequalities that are implied by nonnegativity of the Shannon information measures are categorically referred to as Shannon-type inequalities. If the number of random variables is fixed, a Shannon-type inequality can in principle be verified by a software package known as ITIP. A non-Shannon-type inequality is a constraint on the entropy function which is not implied by the nonnegativity of the Shannon information measures. In the late 1990s, the discovery of a few such inequalities revealed that Shannon-type inequalities alone do not constitute a complete set of constraints on the entropy function. In the past decade or so, connections between the entropy function and a number of subjects in information sciences, mathematics, and physics have been established. These subjects include probability theory, network coding, combinatorics, group theory, Kolmogorov complexity, matrix theory, and quantum mechanics. This expository work is an attempt to present a picture for the many facets of the entropy function.[1]

Keywords: Entropy, polymatroid, non-Shannon-type inequalities, positive definite matrix, quasi-uniform array, Kolmogorov complexity, conditional independence, network coding, quantum information theory.

---

*R. W. Yeung is with the Institute of Network Coding and Department of Information Engineering, The Chinese University of Hong Kong, N.T., Hong Kong. Email: whyeung@ie.cuhk.edu.hk His work was partially supported by a grant from the University Grants Committee (Project No. AoE/E-02/08) of the Hong Kong Special Administrative Region, China.

[1]This work is based on the author's plenary talk with the same title at the 2009 IEEE International Symposium on Information Theory, Seoul, Korea, Jun 28 - Jul 3, 2009.