Facets of Entropy

Raymond W. Yeung
The Chinese University of Hong Kong

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.
- Entropy (Shannon Entropy)

$$H(X) = -\sum_x p(x) \log p(x)$$

Preliminaries

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

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

- Joint Entropy

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

Preliminaries

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

Preliminaries

Preliminaries

- For n random variables, there are $2^n - 1$ joint entropies.

Preliminaries

- 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,

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

Laws of Information Theory

Laws of Information Theory

- Constraints on entropies govern the "impossibilities" in Information Theory. Sometimes called the "Laws of Information Theory."

Laws of Information Theory

- Constraints on entropies govern the "impossibilities" in Information Theory. Sometimes called the "Laws of Information Theory."
- Pippenger's Question (1986): Are there any constraints on entropies other than the basic inequalities?

The Region Γ_n^* (Y97)

The Region Γ_n^* (Y97)

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

The Region Γ_n^* (Y97)

- 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^* (Y97)

- 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.

The Region Γ_n^* (Y97)

- 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

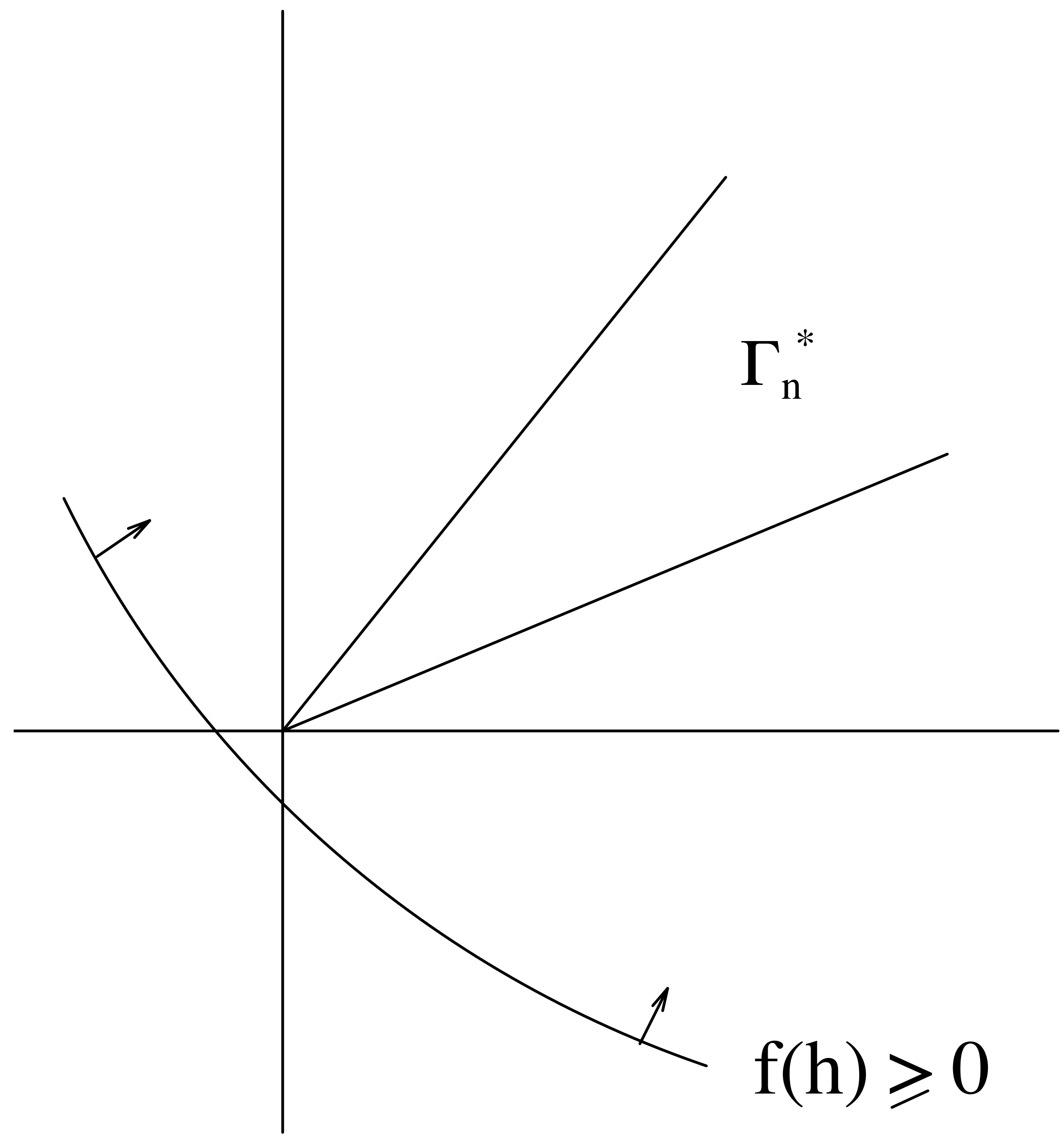
$$\Gamma_n^* \subset \{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}.$$

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

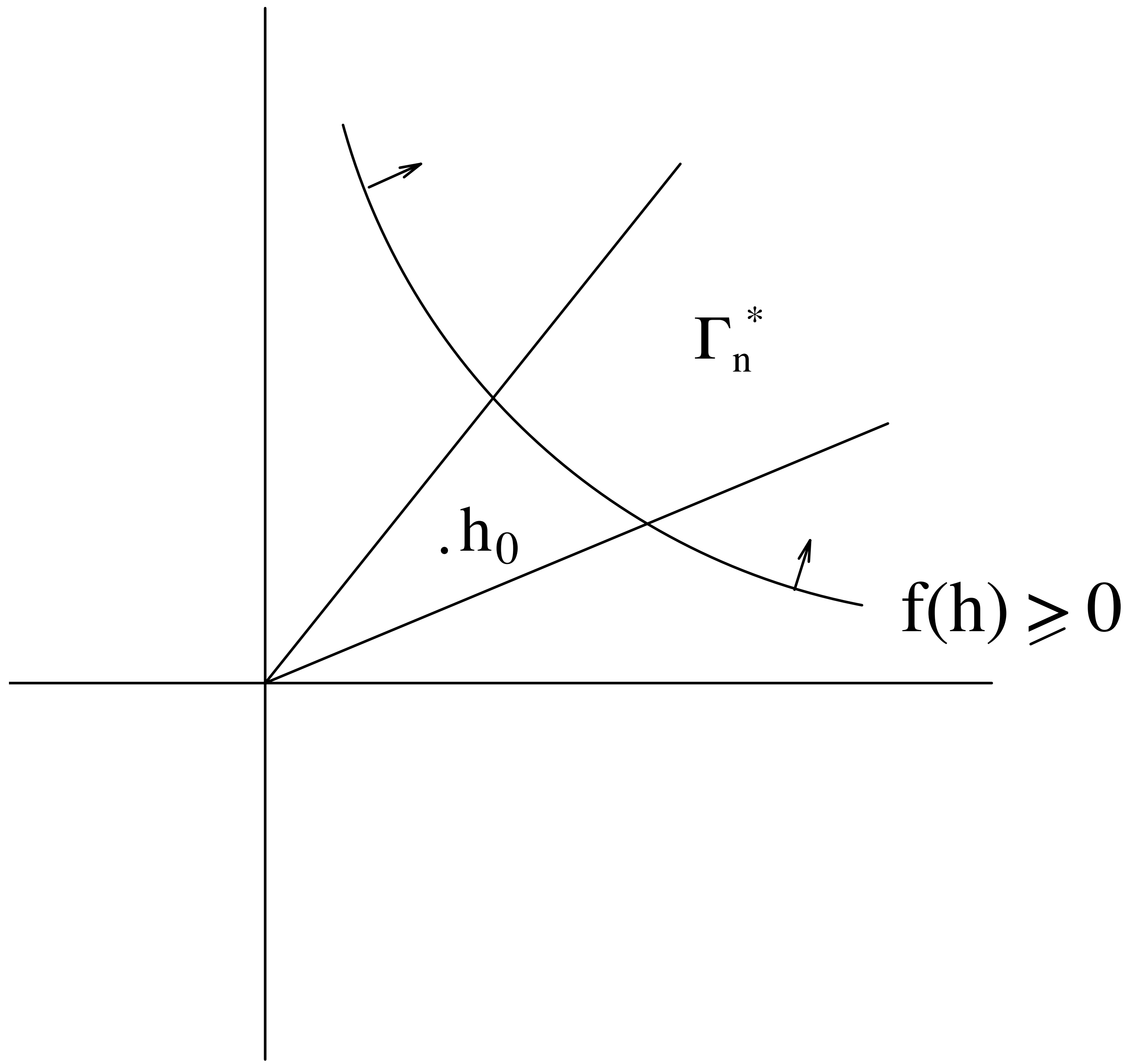
$$\overline{\Gamma}_n^* \subset \{\mathbf{h} \in \mathcal{H}_n : f(\mathbf{h}) \geq 0\}$$

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.

Pippenger's Problem Rephrased

Pippenger's Problem Rephrased

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

Pippenger's Problem Rephrased

- Is $\overline{\Gamma}_n^* = \Gamma_n$?
- 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.

Pippenger's Problem Rephrased

- Is $\overline{\Gamma}_n^* = \Gamma_n$?
- 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 $\overline{\Gamma}_3^* = \Gamma_3$

Pippenger's Problem Rephrased

- Is $\overline{\Gamma}_n^* = \Gamma_n$?
- 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.

Pippenger's Problem Rephrased

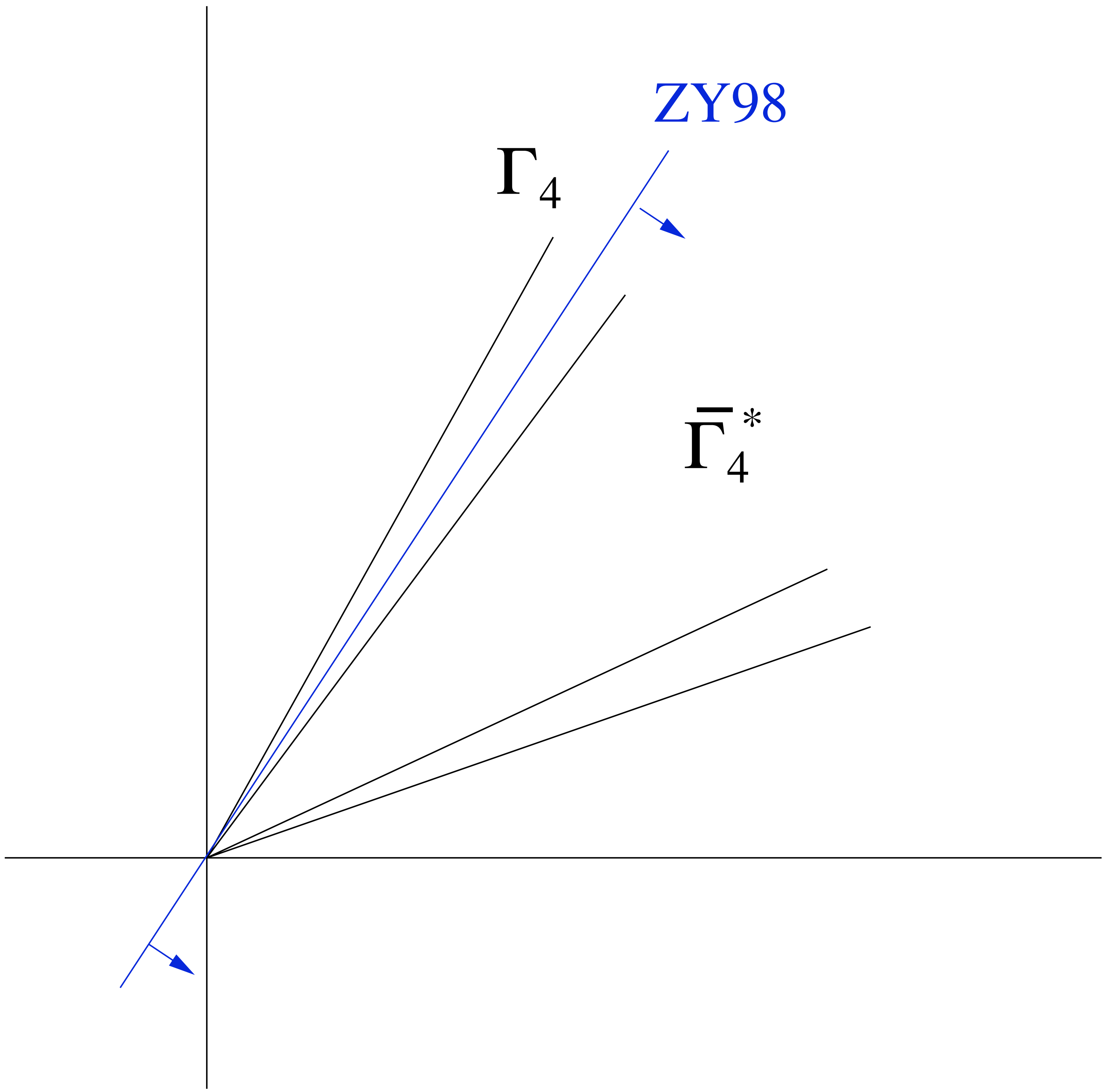
- Is $\overline{\Gamma}^*_n = \Gamma_n$?
- The answer is NO iff there exists an entropy inequality $g(\mathbf{h}) \geq 0$ which cuts between Γ_n and $\overline{\Gamma}^*_n$. Such an inequality is called a non-Shannon-type inequality.
- It is known that
  1. $\Gamma^*_2 = \Gamma_2$
  2. $\Gamma^*_3 \neq \Gamma_3$, 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 Non-Shanon-Type Inequality

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

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

An Illustration of ZY98



Other Non-Shanon-Type Inequalities

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

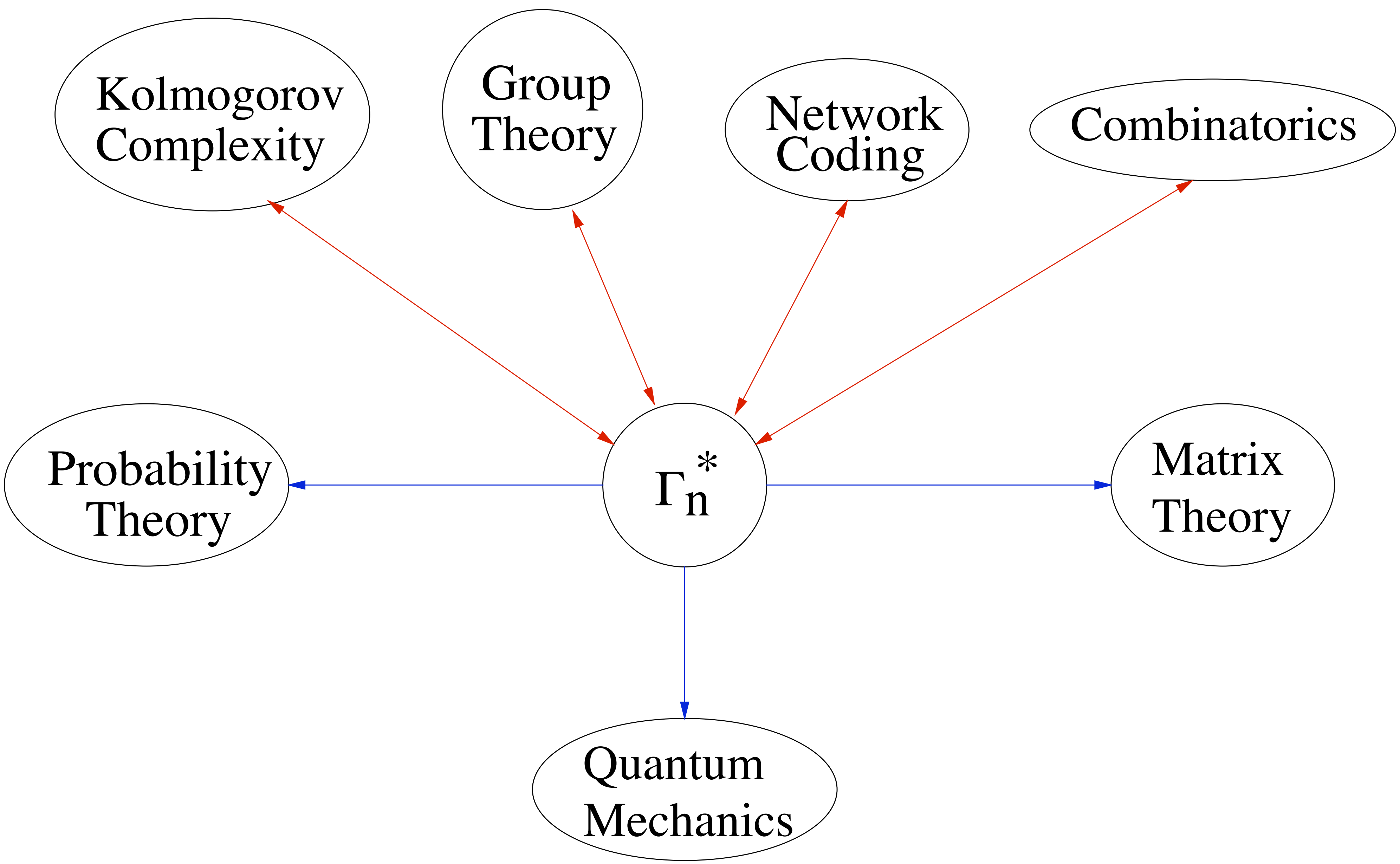
Other Non-Shanon-Type Inequalities

- 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 polytope, and hence there exist an infinitely number of linear non-Shannon-type inequalities!

Other Non-Shanon-Type Inequalities

- 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 polytope, 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 the supercomputer at UCSD.

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 $p(x)$.

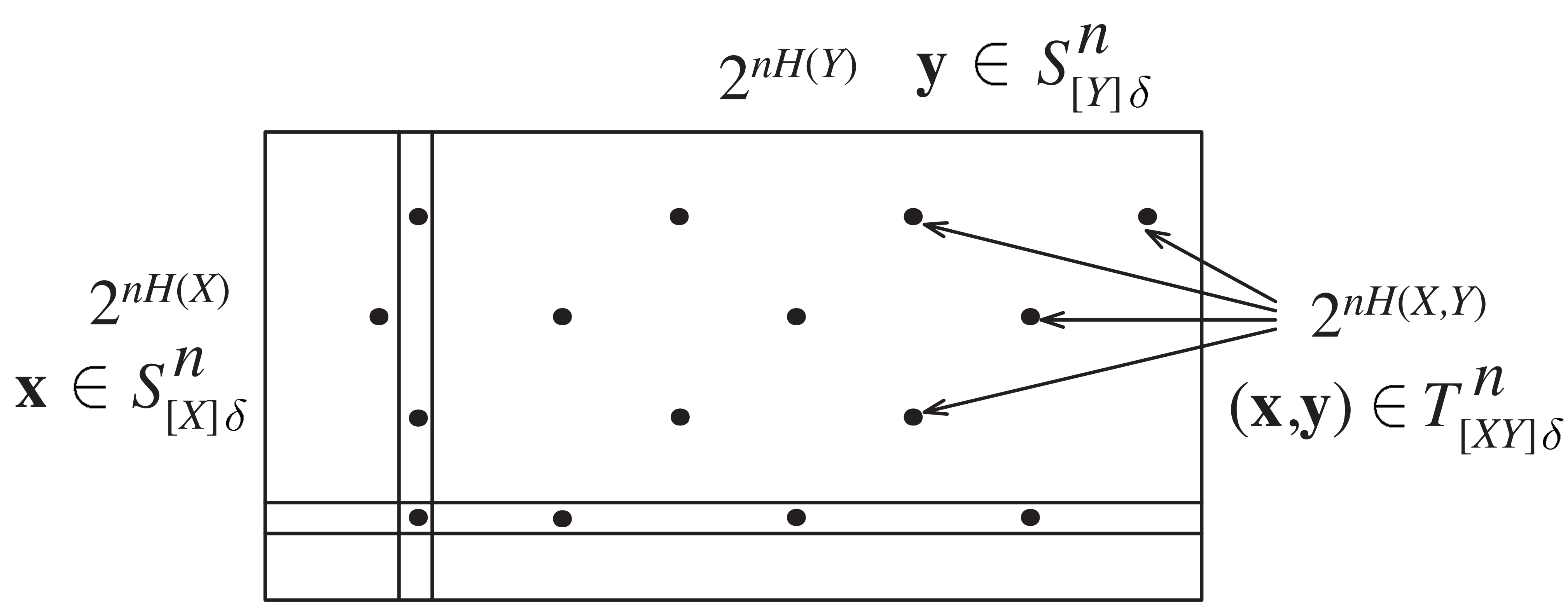
2-D Quasi-Uniform Array

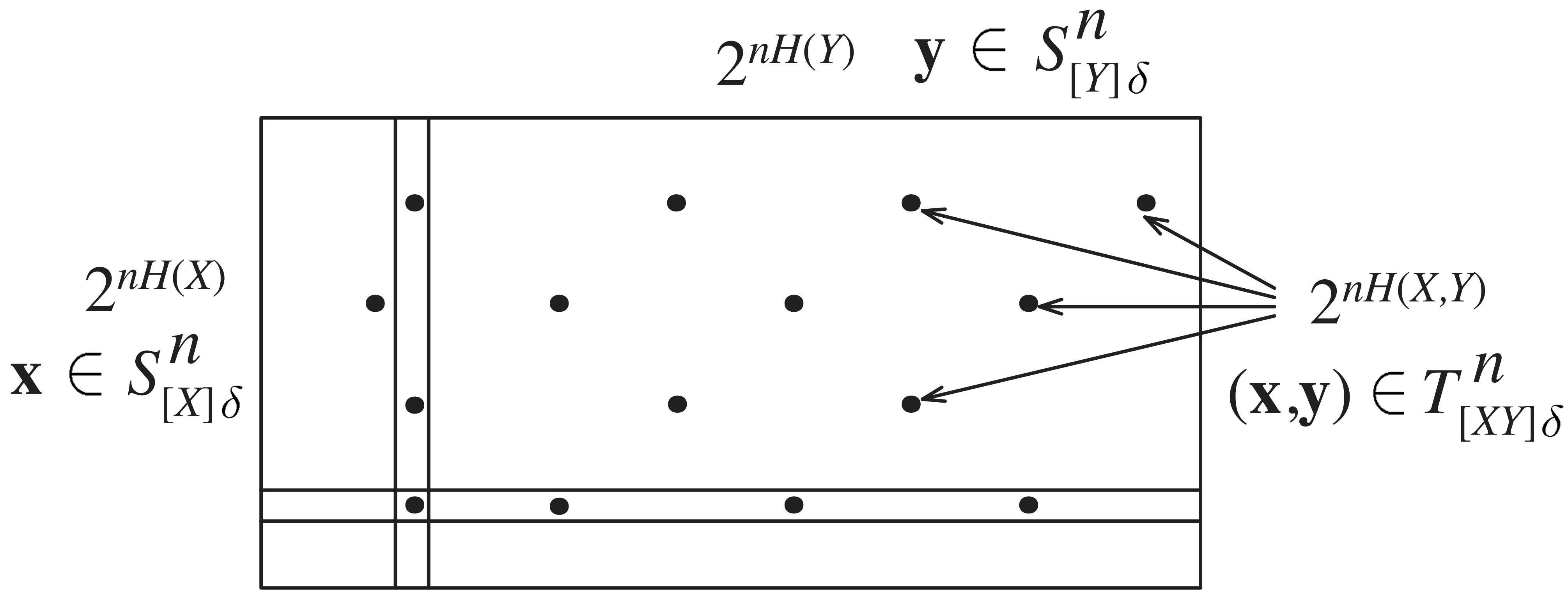
- For a distribution $p(x)$, a sequence $\mathbf{x}$ of length n is strongly typical if the empirical distribution of $\mathbf{x}$ is approximately equal to $p(x)$.

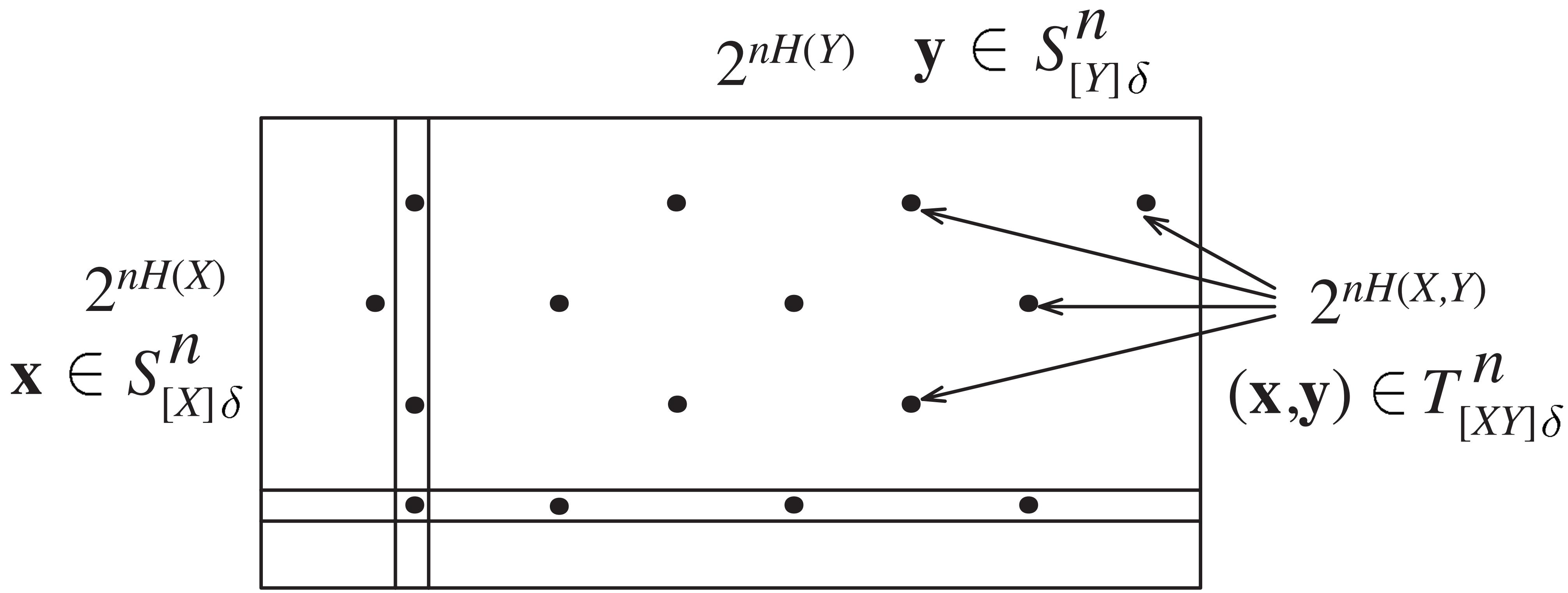
- Let $p(x, y)$ be a joint distribution. The strongly typical sequences w.r.t. $p(x, y)$, $p(x)$, and $p(y)$ can be illustrated by a 2-D quasi-uniform array.

2-D Quasi-Uniform Array

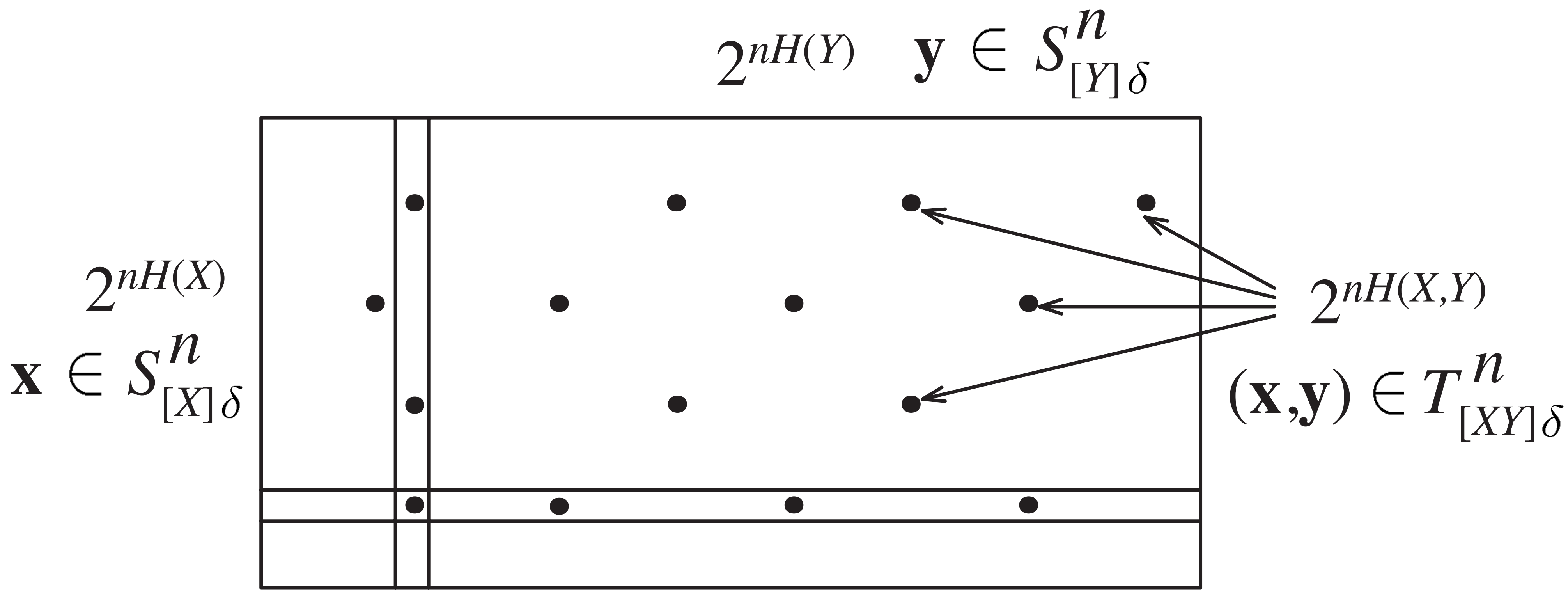
- For a distribution $p(x)$, a sequence $\mathbf{x}$ of length n is strongly typical if the empirical distribution of $\mathbf{x}$ is approximately equal to $p(x)$.
- Let $p(x,y)$ be a joint distribution. The strongly typical sequences w.r.t. $p(x,y)$, $p(x)$, and $p(y)$ can be illustrated by a 2-D quasi-uniform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




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



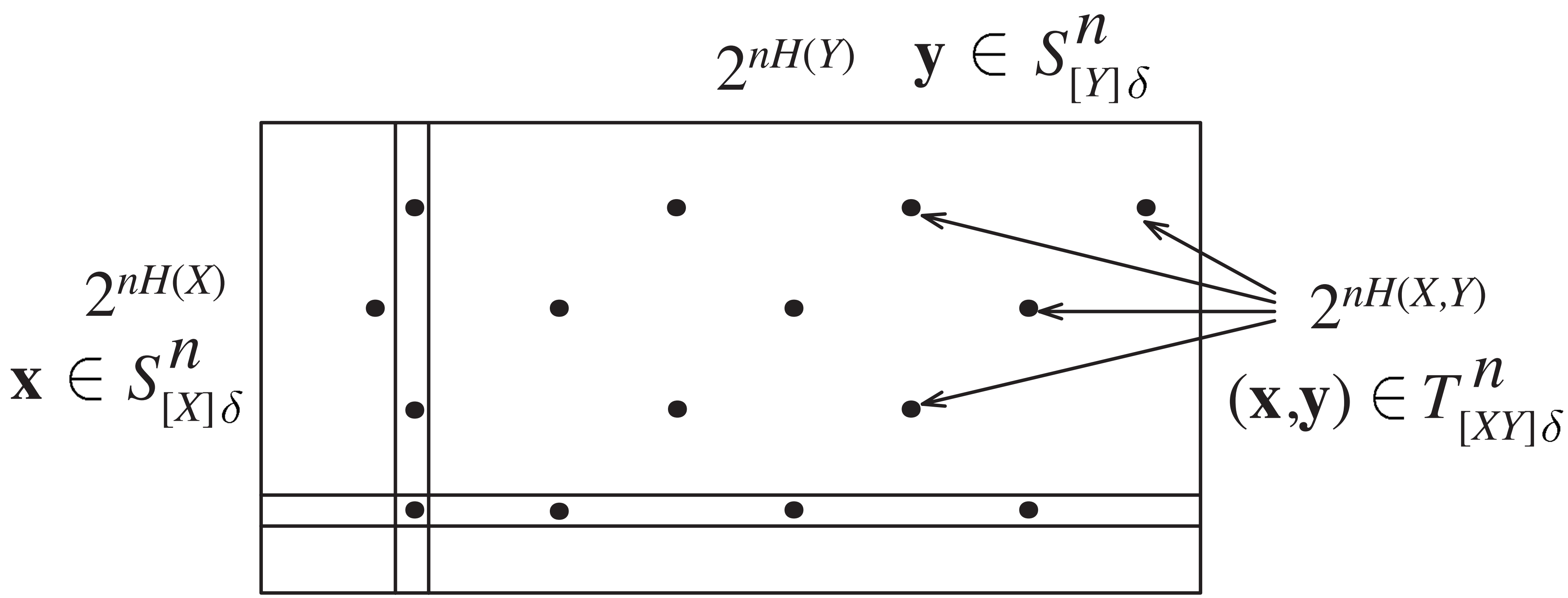
- Each row has approximately the same number of dots ($\sim 2^{nH(Y|X)}$) and each column has approximately the same number of dots ($\sim 2^{nH(X|Y)}$).



- 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

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



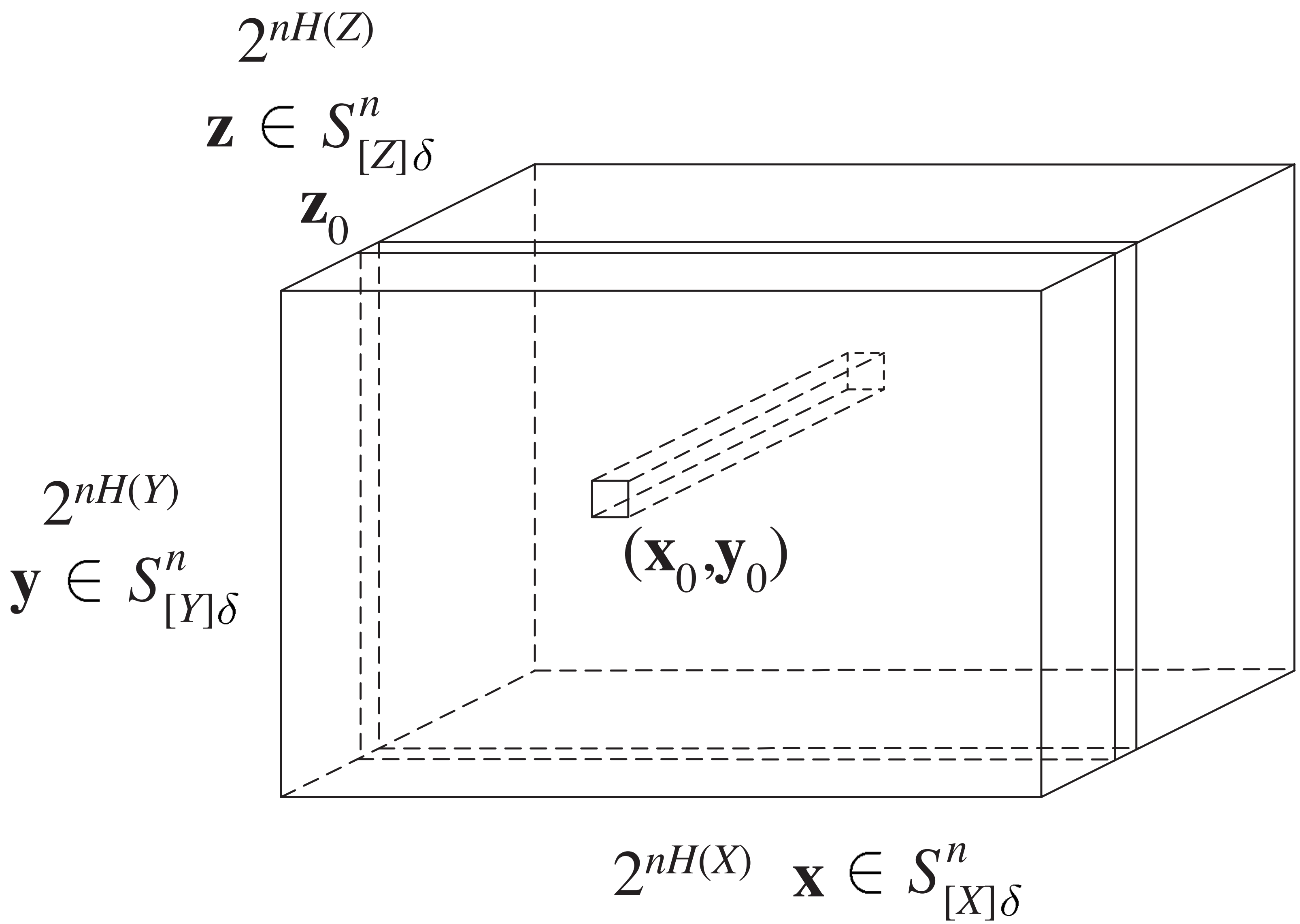
- Each row has approximately the same number of dots ($\sim 2^{nH(Y|X)}$) and each column has approximately the same number of dots ($\sim 2^{nH(X|Y)}$).

- 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 99)

Entropy and Groups
(Chan-Y 99)

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

Entropy and Groups (Chan-Y 99)

- 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 99)

- 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

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

Non-shannon-Type Group Inequalities

Non-shannon-Type Group Inequalities

- This corresponds to

$$\begin{aligned} |G_3 \cap G_4|^6 |G_1 \cap G_3|^4 & & |G_1||G_2||G_3|^4|G_4|^4 \\ |G_1 \cap G_4|^4 |G_2 \cap G_3|^4 & \leq & |G_1 \cap G_2|^2 |G_1 \cap G_3 \cap G_4|^5 \\ |G_2 \cap G_4|^4 & & |G_2 \cap G_3 \cap G_4|^5 \end{aligned}$$

Non-shannon-Type Group Inequalities

- This corresponds to

$$\begin{array}{rcl} |G_3 \cap G_4|^6 |G_1 \cap G_3|^4 & & |G_1||G_2||G_3|^4 |G_4|^4 \\ |G_1 \cap G_4|^4 |G_2 \cap G_3|^4 & \leq & |G_1 \cap G_2|^2 |G_1 \cap G_3 \cap G_4|^5 \\ |G_2 \cap G_4|^4 & & |G_2 \cap G_3 \cap G_4|^5 \end{array}$$

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

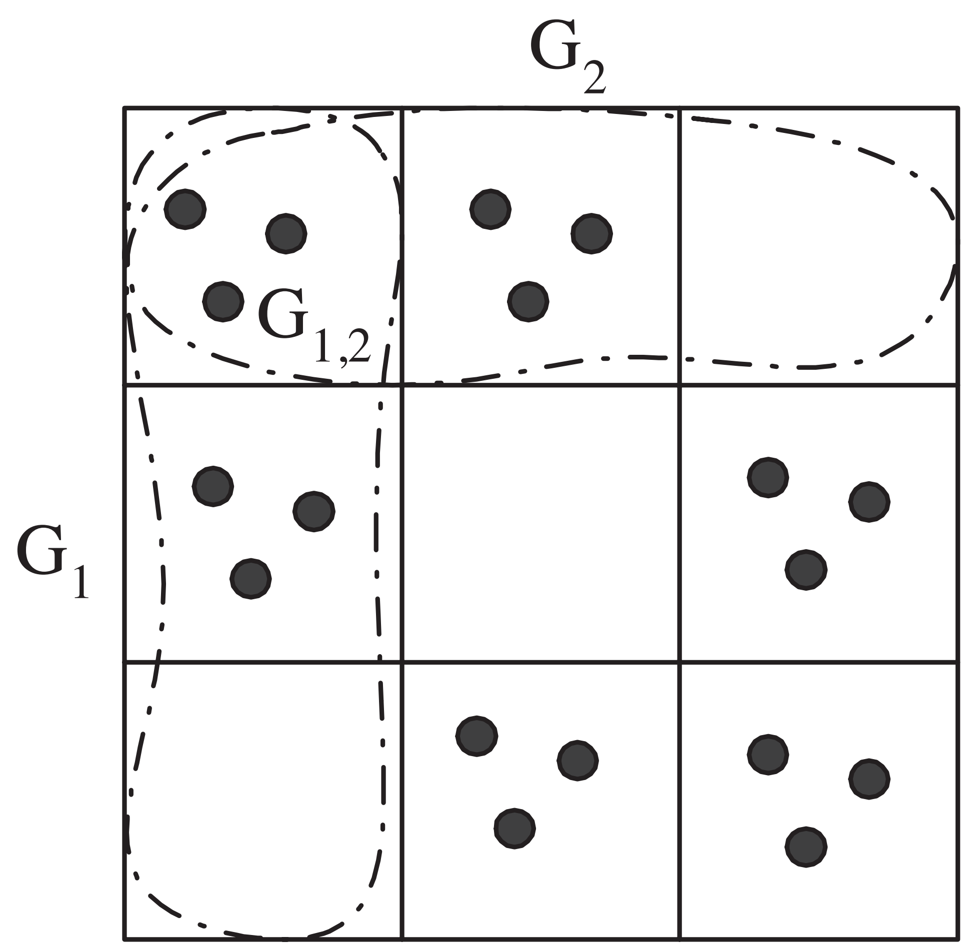
Relation between Finite Group and Quasi-Uniform Array

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.



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,

$$K(x) + K(y) \geq K(x, y)$$

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,

$$K(x) + K(y) \geq K(x, y)$$

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

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

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

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

- 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.

- Very hard for $n \geq 4$.

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.

- $n = 4$ was settled by F. Matúš (1999) with the help of the following non-Shannon-type information inequality (Ingleton inequality):

  If $X \perp Y$ or $Y \perp U|Z$, then

$$H(XYZ) + H(XU) + H(YU) + H(ZU) - H(XY) - H(U)$$

$$-H(XZU) - H(YZU) \geq 0$$

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.

- $n = 4$ was settled by F. Matúš (1999) with the help of the following non-Shannon-type information inequality (Ingleton inequality):

  If $X \perp Y$ or $Y \perp U|Z$, then

$$H(XYZ) + H(XU) + H(YU) + H(ZU) - H(XY) - H(U)$$

$$-H(XZU) - H(YZU) \geq 0$$

- Matroid Theory is a powerful tool for studying this problem.

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 (2006) 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

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

gives the Hadamard inequality

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

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

$$\begin{aligned} & |K_1||K_2||K_{1,2}|^2|K_3|^4|K_4|^4|K_{1,3,4}|^5|K_{2,3,4}|^5 \\ \leq \quad & |K_{3,4}|^6|K_{1,3}|^4|K_{1,4}|^4|K_{2,3}|^4|K_{2,4}|^4 \end{aligned}$$

for all positive definte 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

$$\begin{aligned} & |K_1||K_2||K_{1,2}|^2|K_3|^4|K_4|^4|K_{1,3,4}|^5|K_{2,3,4}|^5 \\ \leq \quad & |K_{3,4}|^6|K_{1,3}|^4|K_{1,4}|^4|K_{2,3}|^4|K_{2,4}|^4 \end{aligned}$$

for all positive definte matrices.

- Many other "non-Shannon-type" determinant inequalities can be obtained this way.

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 problem has been studied by Y and Zhang (1999), Song, Y, and Cai (2006). Yan, Y and Zhang (2007) 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 (2007) proved a one-to-one correspondence between 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 (2007) proved a one-to-one correspondence between entropy functions and multi-source network coding problems.

- Thus

Every constraint on the entropy function is useful in
some 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 (2007) proved a one-to-one correspondence between entropy functions and multi-source network coding problems.
- Thus

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

- The implications of non-Shannon-type inequalities in information theory is finally understood in the context of network coding.

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 lower bound in a secret sharing problem.

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 lower bound in a secret sharing problem.
- Secret sharing can be regarded as a special case of secure network coding (Cai and Y, 2002).

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

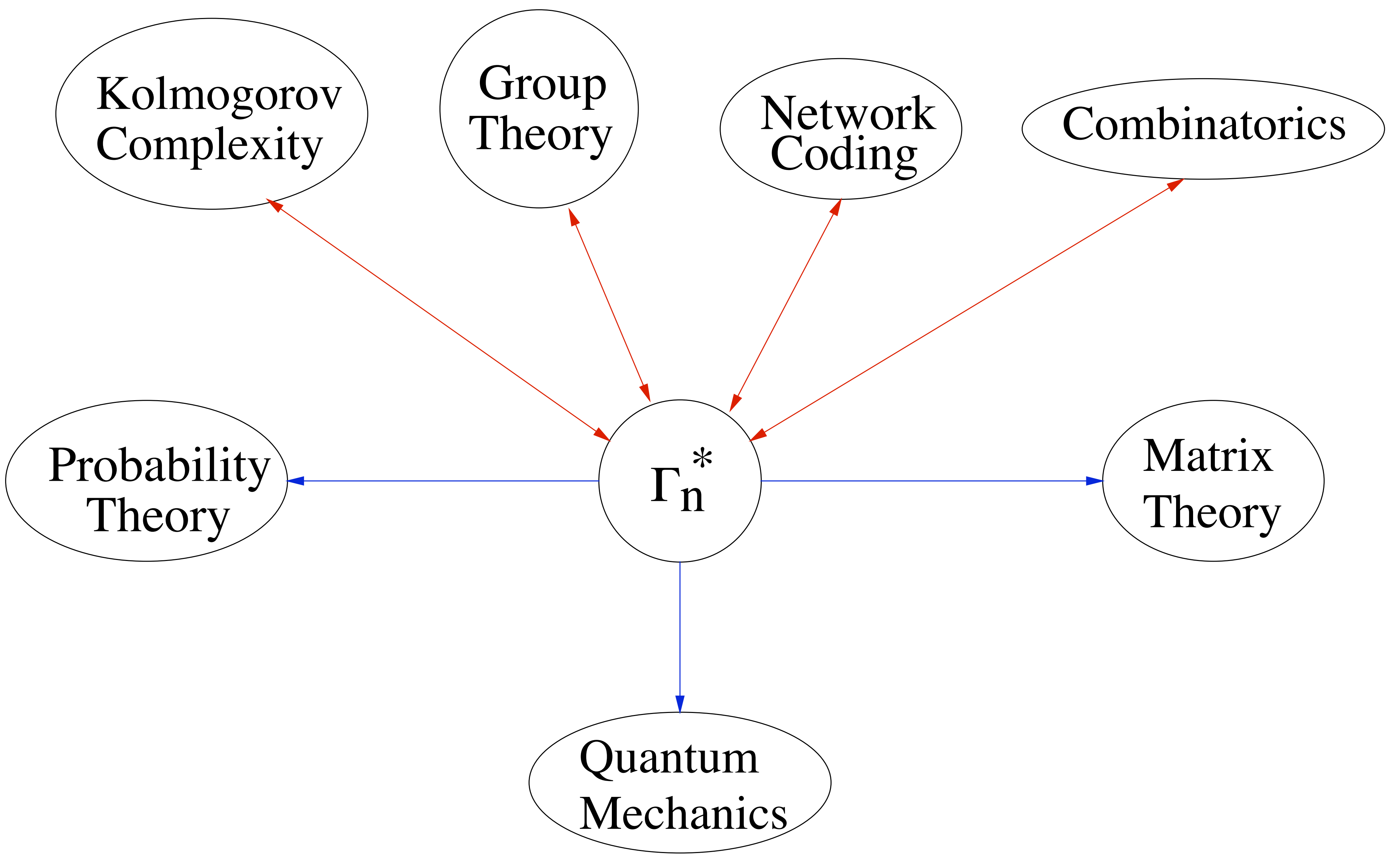
- The von Neumann entropy is an extension of the Shannon entropy to quantum mechanics.

- 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

- I. Csiszár advocates that information theory should be an integral part of mathematics.

Concluding Remarks

- I. Csiszár advocates that information theory should be an integral part of mathematics.
- It is clear that constraints on the entropy function has fundamental implications in a number of fields in information science, mathematics, and physics.

Concluding Remarks

- I. Csiszár advocates that information theory should be an integral part of mathematics.
- It is clear that constraints on the entropy function has fundamental implications in a number of fields in information science, 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

- I. Csiszár advocates that information theory should be an integral part of mathematics.
- It is clear that constraints on the entropy function has fundamental implications in a number of fields in information science, 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.
- Matroid theory plays a role here and there, in particular in the study of conditional independence and network coding.

Concluding Remarks

- I. Csiszár advocates that information theory should be an integral part of mathematics.
- It is clear that constraints on the entropy function has fundamental implications in a number of fields in information science, 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.
- Matroid theory plays a role here and there, in particular in the study of conditional independence and network coding.
- The relations between these fields need a deeper understanding. The combinatorial structure to study is the quasi-uniform array.

Concluding Remarks

- I. Csiszár advocates that information theory should be an integral part of mathematics.
- It is clear that constraints on the entropy function has fundamental implications in a number of fields in information science, 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.
- Matroid theory plays a role here and there, in particular in the study of conditional independence and network coding.
- The relations between these fields need a deeper understanding. The combinatorial structure to study is the quasi-uniform array.
- "Non-Shannon-type" inequalities in different fields need further understanding.

ARCHIVE

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.

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

ACKNOWLEDGMENTS

Zhen Zhang
University of Southern California

Zhen Zhang
University of Southern California

Terence Chan
University of South Australia

Zhen Zhang
University of Southern California

Terence Chan
University of South Australia

Imre Csiszár
Hungarian Academy of Sciences

Thank You