Conditional Independence Relations

Yunshu Liu

ASPITRG Research Group

2013-02-20

Motivation

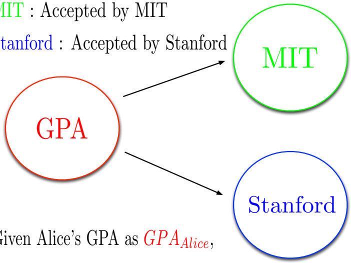
What is Conditional Independence Relations?

A is conditionally independent of B given C

Example: Suppose MIT and Stanford accepted undergraduate students only based on GPA

MIT : Accepted by MIT

Stanford : Accepted by Stanford



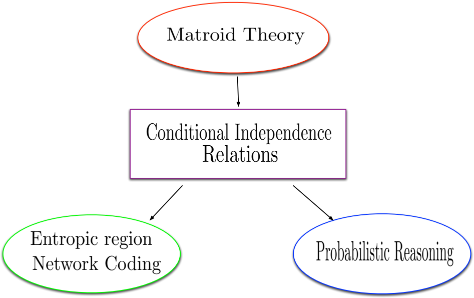
Given Alice's GPA as GPA_{Alice},

$\mathbb{P}(MIT|Stanford, GPA_{Alice}) = \mathbb{P}(MIT|GPA_{Alice})$

We say MIT is conditionally independent of Stanford given GPA_{Alice}

Sometimes use symbol $(MIT \perp Stanford | GPA_{Alice})$

Motivation



References:

[1]. F. Matúš and M. Studený, Conditional Independences among Four Random Variables I, Combinatorics, Probability and Computing, 1995, page 269-278.

[2]. F. Matúš, Infinitely Many Information Inequalities, IEEE Int. Symp. Information Theory (ISIT), 2007, page 41-44

Outline

Preliminaries on Matroid theory

Matroid theory and Conditional Independence Relations

Infinitely many Information Inequalities

Outline

- Preliminaries on Matroid theory
- Matroid theory and Conditional Independence Relations
- Infinitely many Information Inequalities

Definitions of Matroid

What is Matroid?

Matroid is an independence structure that captures and generalizes the notion of linear independence in vector spaces.

Independent Sets based definition of Matroid

We can represent a finite matroid by a pair $(E, \mathcal{I})$, where E is a finite set called ground set and $\mathcal{I}$ is a family of subset of E called independent sets obeying the following properties:

- $\emptyset \in \mathcal{I}$
- $I_1 \in \mathcal{I}$ implies that $I_2 \in \mathcal{I}$ for every subset $I_2 \subseteq I_1$
- $I_1, I_2 \in \mathcal{I}$ with $|I_1| < |I_2|$ implies there is an $e \in I_2 \backslash I_1$ such that $I_1 \cup e \in \mathcal{I}$

Definitions of Matroid

Rank Function based definition of Matroid

A rank function is a function from subsets of ground set E to integers satisfying the following conditions:

- If $X \subseteq E$ then $0 \leqslant r(X) \leqslant |X|$
- If $X \subseteq Y \subseteq E$, then $r(X) \leqslant r(Y)$
- If $X, Y \subset E$, then $r(X) + r(Y) \geqslant r(X \cup Y) + r(X \cap Y)$

NOTE: The value of the rank function is always a non-negative integer

Examples of $|E| = 4$

For $\mathbf{r}$ = $[r_1\ r_2\ r_{12}\ r_3\ r_{13}\ r_{23}\ r_{123}\ r_4\ r_{14}\ r_{24}\ r_{124}\ r_{34}\ r_{134}\ r_{234}\ r_{1234}]$

- $\mathbf{r}_1$ = [0 0 0 1 1 1 1 1 1 1 1 1 1 1 1] is a matroid
- $\mathbf{r}_2$ = [2 1 2 1 2 2 2 1 2 2 2 2 2 2 2] is not a matroid
- $\mathbf{r}_3$ = [2 2 3 2 3 3 4 2 3 3 4 4 4 4 4] is not a matroid

Polymatroids and Matroids

Polymatroidal axioms

Let f map subsets of ground set E to nonnegative real numbers, the following conditions are called Polymatroidal axioms:

- $f(\emptyset) = 0$
- If $X \subseteq Y \subseteq E$, then $r(X) \leqslant r(Y)$
- If $X, Y \subset E$, then $r(X) + r(Y) \geqslant r(X \cup Y) + r(X \cap Y)$

Examples: $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ all correspond to polymatroids.

Relationship between Matroids and Polymatroids

Matroids are Polymatroids with integer rank function and cardinality constrain.

Conditional Independence Relations

Phrase $(i, j|K)$

Let $N = \{1, 2, \cdots n\}$ and $\mathcal{S}$ be the family of all couples $(i, j|K)$, where $K \subset N$ and ij is the union of two singletons i and j of $N - K$.

Example

For $N = \{1, 2, 3\}$, there are 18 such couples $(i, j|K)$, including the case when $i = j$. Listed below:
$(1, 1|\emptyset)$, $(1, 1|2)$, $(1, 1|3)$, $(1, 1|23)$,
$(2, 2|\emptyset)$, $(2, 2|1)$, $(2, 2|3)$, $(2, 2|13)$,
$(3, 3|\emptyset)$, $(3, 3|1)$, $(3, 3|2)$, $(3, 3|12)$,
$(1, 2|\emptyset)$, $(1, 2|3)$, $(1, 3|\emptyset)$, $(1, 3|2)$, $(2, 3|\emptyset)$, $(2, 3|1)$

p-representation

Probabilistically(p-) representation

A relation $\mathcal{L} \subset \mathcal{S}$ is called probabilistically representable if there exists a system of n random variables $\xi = \{\xi_i\}_{i \in N}$ such that:

- $\mathcal{L} = |[\xi]|$ = $\{(i, j|K) \in \mathcal{S}(N)|\ \xi_i$ is conditionally independent of ξ_j given ξ_K i.e. $I(\xi_i; \xi_j|\xi_K) = 0\ \}$.

We use $P(N)$ to denote the set of all p-representable relations on N.

Examples of $|E| = 4$, consider couple $(2, 3|1)$

For **r** = $[r_1\ r_2\ r_{12}\ r_3\ r_{13}\ r_{23}\ r_{123}\ r_4\ r_{14}\ r_{24}\ r_{124}\ r_{34}\ r_{134}\ r_{234}\ r_{1234}]$

- $\mathbf{r}_1$ = [0 0 0 1 1 1 1 1 1 1 1 1 1 1 1] is p-representable
- $\mathbf{r}_2$ = [2 1 2 1 2 2 2 1 2 2 2 2 2 2 2] is p-representable
- $\mathbf{r}_3$ = [2 2 3 2 3 3 4 2 3 3 4 4 4 4 4] is not p-representable

Semimatroid

Definition of Semimatroid

For $r \in \mathbf{R}^{\mathcal{P}(N)}$ we define $|[r]|$ as

- $|[r]|$ = $\{(i,j|K) \in \mathcal{S}(N)|\ r(iK) + r(jK) - r(ijK) - r(K) = 0\ \}$.

A relation $\mathcal{L} \subset \mathcal{S}(N)$ is called semimatroid if and only if $\mathcal{L} = |[r]|$ for some $r \in \Gamma_N$.

- Here Γ_N is the shannon outer bound for N random variables, which is also the region contain all polymatroids.
- We use *Semi*(N) to denote the set of all semimatroids on N.
- We say semimatroid $\mathcal{L}$ arising from r.

Examples

p-representable semimatroids are semimatroids arising from entropic vectors h
matroidal semimatroids are semimatroids arising from rank functions of matroids

Outline

- Preliminaries on Matroid theory
- Matroid theory and Conditional Independence Relations
- Infinitely many Information Inequalities

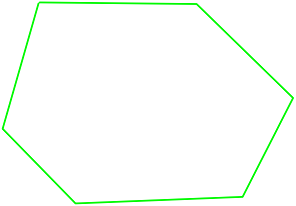
Relationship between p-representable and semimatroid

Every p-representable relation are semimatroid: $P(N) \subseteq Semi(N)$

- $|[\xi]|$ = $\{(i,j|K) \in \mathcal{S}(N) |\ I(\xi_i;\xi_j|\xi_K) = 0\ \}. \in P(N)$
- $|[r]|$ = $\{(i,j|K) \in \mathcal{S}(N) |\ r(iK) + r(jK) - r(ijK) - r(K) = 0\ \}.$ $\in Semi(N)$

$$\mathrm{P(N)} = \mathrm{Semi(N)} \text{ for } N \leq 3$$

$$\text{Recall } \Gamma_N = \bar{\Gamma}^*_N \text{ for } N \leq 3$$

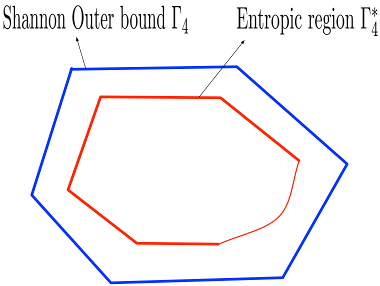


Every p-representable relations are semimatroid: $P(N) \subseteq Semi(N)$

- $|[\xi]|$ = $\{(i,j|K) \in \mathcal{S}(N) |\ I(\xi_i;\xi_j|\xi_K) = 0\ \} \in P(N)$
- $|[r]|$ = $\{(i,j|K) \in \mathcal{S}(N) |\ r(iK) + r(jK) - r(ijK) - r(K) = 0\ \}$ $\in Semi(N)$

$\mathrm{P(N)} \subsetneq \mathrm{Semi(N)}$ for $N = 4$

Recall $\bar{\Gamma}_4^* \subsetneq \Gamma_4$

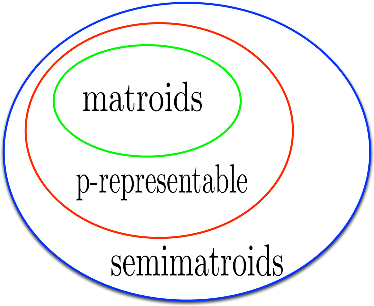


Examples:

- $|[\mathbf{r}_1]|$ and $|[\mathbf{r}_2]|$ are p-representable
- $|[\mathbf{r}_3]|$ is a semimatroid but not p-representable

Matroid, p-representable semimatroid and semimatroid for N = 4

Since every matroid which is linearly representable over a finite field is also p-representable, all matroids for $|N| \leq 4$ are p-representable.



$\mathrm{N} = 4$

Characterization of the extreme rays of Γ_4

Matroids that are extreme rays of Γ_4

For $N = \{1,2,3,4\}$, with $I \subseteq N$ and $0 \leq t \leq |N \backslash I|$, define

$$r_t^I(J) = \min\{t, |J \backslash I|\} \textit{ with } J \subseteq N$$

Then r_t^I is the matroid of rank t with loops I

Γ_4 has 27 matroid extreme rays, all of which can be expressed as r_t^I for some t and I

Characterization of the extreme rays of Γ_4

non-matroid extreme rays of Γ_4

For $N = \{1, 2, 3, 4\}$, define the following functions:

$$g_i^{(2)}(J) = \begin{cases} 2 & \text{if } J = i \\ min\{2, |J|\} & \text{if } J \neq i \end{cases}$$

$$g_i^{(3)}(J) = \begin{cases} |J| & \text{if } i \notin J \\ min\{3, |J| + 1\} & \text{if } i \in J \end{cases}$$

$$f_{ij}(K) = \begin{cases} 3 & \text{if } K \in \{ik, jk, il, jl, kl\} \\ min\{4, 2|K|\} & \text{otherwise} \end{cases}$$

Characterization of the extreme rays of Γ_4

extreme rays of Γ_4

Γ_4 has 41 extreme rays, including
27 matroid of the form r_t^l for some t and l,
4 extreme rays of the form $g_i^{(2)}$ for $i = 1, 2, 3, 4$,
4 extreme rays of the form $g_i^{(3)}$ for $i = 1, 2, 3, 4$,
6 extreme rays of the form f_{ij} for $i, j \in N$ and $i \neq j$.

Examples of $|E| = 4$

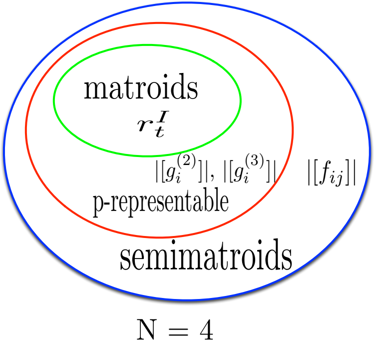
For $\mathbf{r}$ = $[r_1\ r_2\ r_{12}\ r_3\ r_{13}\ r_{23}\ r_{123}\ r_4\ r_{14}\ r_{24}\ r_{124}\ r_{34}\ r_{134}\ r_{234}\ r_{1234}]$

- $\mathbf{r}_1$ = r_1^{12}= [0 0 0 1 1 1 1 1 1 1 1 1 1 1 1]
- $\mathbf{r}_2$ = $g_1^{(2)}$ = [2 1 2 1 2 2 2 1 2 2 2 2 2 2 2]
- $\mathbf{r}_3$ = f_{34} = [2 2 3 2 3 3 4 2 3 3 4 4 4 4 4]

Characterization of the extreme rays of Γ_4

extreme rays of Γ_4

$|[g_i^{(2)}]|$, $|[g_i^{(3)}]|$ and $|[f_{ij}]|$ are all semimatroids, among them $|[g_i^{(2)}]|$ and $|[g_i^{(3)}]|$ are p-representable, $|[f_{ij}]|$ are not p-representable.



Ingleton inequality

Ingleton inequality

$$
\begin{aligned}
&Ingleton_{12} \\
&= I(X_1;X_2|X_3) + I(X_1;X_2|X_4) + I(X_3;X_4|\emptyset) - I(X_1;X_2|\emptyset) \\
&= h_{12} + h_{13} + h_{23} + h_{14} + h_{24} - h_1 - h_2 - h_{34} - h_{123} - h_{124} \geqslant 0
\end{aligned}
$$

R_4: generated by four variable shannon-type inequalities and 6 Ingleton inequalities

R_4 has 35 extreme rays, including
27 matroid of the form r_t^I for some t and I,
8 extreme rays of the form $g_i^{(2)}$ and $g_i^{(3)}$ for $i = 1, 2, 3, 4$.
Note: $f_{ij} \notin R_4$ for $i, j \in N$ and $i \neq j$

Ingleton inequality

Ingleton semimatroid

A relation $\mathcal{L} \subset \mathcal{S}(N)$ is called Ingleton semimatroid if and only if $\mathcal{L} = |[r]|$ for some $r \in R_N$.

For $|N| = 4$ every Ingleton semimatroid is p-representable

Reason:
Define the convex cone of all Ingleton semimatroid as InSemi(N), all the extreme rays of InSemi(N) are p-representable.

G_4^{ij}: The gap between R_4 and Γ_4

$G_4^{ij} = \{h \in \Gamma_4 | Ingleton_{ij} \leq 0\}$

G_4^{ij} is the convex hull of 15 extreme rays, the V-representation are generated by the 15 linearly independent functions f_{ij}, r_1^{ijk}, r_1^{ijl}, r_1^{ikl}, r_1^{jkl}, $r_1^{\emptyset}$, $r_3^{\emptyset}$, r_1^{i}, r_1^{j}, r_1^{ik}, r_1^{jk}, r_1^{il}, r_1^{jl}, r_2^{k}, r_2^{l}.

The H-representation of G_4^{ij} are 14 shannon-type inequality together with $-Ingleton_{ij} \leq 0$

List of semimatroids on four discrete variables

Ingleton semimatroid

There are 120 irreducible p-representable semimatroids of sixteen types(means remove all permutations) over four-element set N. Among which there are 36 ingleton semimatroids of 11 types:
$|[0]|$, $|[r_1^{N-i}]|$ for $i \in N$, $|[r_1^{ij}]|$ for $i, j \in N$ distinct, $|[r_1^{i}]|$ for $i \in N$, $|[r_1]|$, $|[r_2^{i}]|$ for $i \in N$, $|[r_2^{i||j}]|$ for $i, j \in N$ distinct, $|[r_2]|$, $|[r_3]|$, $|[g_i^{(2)}]|$ for $i \in N$, $|[g_i^{(3)}]|$ for $i \in N$.

List of semimatroids on four discrete variables

non-Ingleton semimatroid

There are 120 irreducible p-representable semimatroids of sixteen types(means remove all permutations) over four-element set *N*. Among which there are 84 non-Ingleton semimatroids of 5 types:

$$\begin{aligned}
\mathcal{L}_{ij}^{kl|\emptyset} &= \{(kl|i),(kl|j)(ij|\emptyset)\} \cup \{(kl|ij)\} \cup \{(k|ij),(l|ij),(i|jkl),(j|ikl),(k|ijl),(l|ijk)\} \\
\mathcal{L}_{ij}^{(ij|kl)} &= \{(ij|k),(ij|l),(kl|ij)\} \cup \{(kl|i),(kl|j)\} \\
\mathcal{L}_{ij}^{(ik|jl)} &= \{(kl|ij),(ij|k),(ik|l)\} \cup \{(kl|j)\} \cup \{(l|ij),(l|ijk)\} \\
\mathcal{L}_{ij}^{ik|j} &= \{(ij|k),(ik|l),(kl|j)\} \cup \{(i|jkl),(j|ikl),(k|ijl),(l|ijk)\} \\
\mathcal{L}_{ij}^{jl|\emptyset} &= \{(kl|i),(jl|k),(ij|\emptyset)\} \cup \{(kl|ij)\} \cup \{(k|ij),(l|ij),(i|jkl),(j|ikl),(k|ijl),(l|ijk)\}
\end{aligned}$$

Outline

- Preliminaries on Matroid theory
- Matroid theory and Conditional Independence Relations
- Infinitely many Information Inequalities

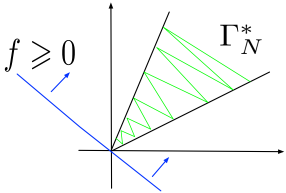
Information Inequalities

Information Inequalities: Definition

If we consider the function of all $(2^n - 1)$ joint entropies associated with **X**:

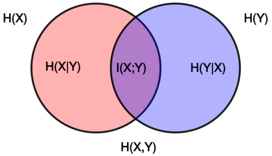
$$f = \sum_i \lambda_{\mathcal{A}_i} h_{\mathcal{A}_i} \geqslant 0$$

where $\lambda_{\mathcal{A}_i}$ are real coefficients. If a inequality of this form is true for any random variables $\{X_1, \cdots X_n\}$, we call it an Information inequality.



Information Inequalities: examples

For jointly related discrete random variables A, B and C, we can define the conditional entropy of A given B by $H(A|B)$, the mutual information between A and B by $I(A;B)$, the conditional mutual information is given by $I(A;B|C)$.



$$H(A|B) = H(A,B) - H(B) \geqslant 0$$
$$I(A;B) = H(A) + H(B) - H(A,B) \geqslant 0$$
$$I(A;B|C) = H(A,C) + H(B,C) - H(A,B,C) - H(C) \geqslant 0$$

Information Inequalities

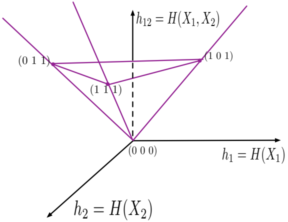
Shannon-Type Information inequality

$$\Gamma_N = \left\{ \boldsymbol{h} \in \mathbb{R}^{2^N-1} \;\middle|\; \begin{array}{c} h_{\mathcal{A}} + h_{\mathcal{B}} \geq h_{\mathcal{A}\cap\mathcal{B}} + h_{\mathcal{A}\cup\mathcal{B}} \; \forall \mathcal{A}, \mathcal{B} \subseteq \mathcal{N} \\ h_{\mathcal{P}} \geq h_{\mathcal{Q}} \geq 0 \quad \forall \mathcal{Q} \subseteq \mathcal{P} \subseteq \mathcal{N} \end{array} \right\}$$

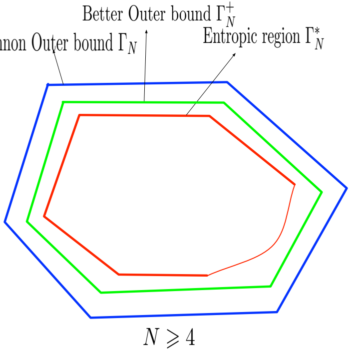
Relationship between Γ_n^* and Γ_n

$\Gamma_2^* = \Gamma_2$ and $\bar{\Gamma}_3^* = \Gamma_3$:
All the Information inequalities on
$N \leqslant 3$ variables are Shannon-Type.

For N = 2, Use $h_1 + h_2 \geqslant h_{12}$,
$h_{12} \geqslant h_1$ and $h_{12} \geqslant h_2$, we get



Information Inequalities



For $N \geq 4$ we have $\Gamma_N \neq \bar{\Gamma}_N^*$

Non-Shannon-Type Information inequalities exist for $N \geqslant 4$

The first discovered Non-Shannon-Type Information inequality:Zhang-Yeung Inequality

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

Zhang-Yeung Inequality and DFZ Inequality

Recall:

$$\begin{aligned}&Ingleton_{12}\\&= I(X_1;X_2|X_3) + I(X_1;X_2|X_4) + I(X_3;X_4|\emptyset) - I(X_1;X_2|\emptyset)\end{aligned}$$

Zhang-Yeung Inequality

$$Ingleton_{12} + I(X_1;X_2|X_3) + I(X_1;X_3|X_2) + I(X_2;X_3|X_1) \geqslant 0$$

DFZ Inequality

$$2Ingleton_{12} + 3I(X_1;X_2|X_3) + 3I(X_1;X_3|X_2) + I(X_2;X_3|X_1) \geqslant 0$$

Information Inequality

A sequence of Information Inequalities on four variables

$$s\,Ingleton_{12} + I(X_2; X_3|X_1) + \frac{s(s+1)}{2}[I(X_1; X_2|X_3) + I(X_1; X_3|X_2)] \geqslant 0$$

Derived by F. Matúš using adhesivity of polymatroid in [2];
Later by R. Dougherty and C. Freiling and K. Zeger using D-copy in their preprint paper http://arxiv.org/abs/1104.3602

The closure of Γ_N^* is not polyhedral for $N \geq 4$

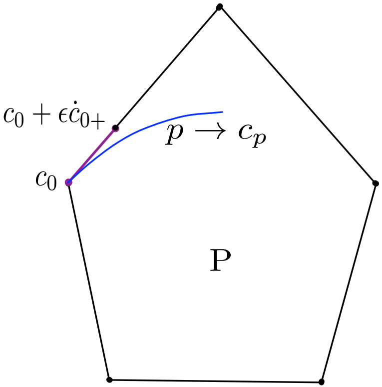
Ideas and procedure of the proof:

- Generate a sequence of Information Inequalities on four variables.
- Prove a Lemma such that if a cone is a polyhedral set, a curve in this set must have certain property .
- Construct a curve in the closure of Γ_4^*.
- Prove the property of this curve contradict the results of the Lemma, thus the closure of Γ_4^* is not polyhedral.

The closure of Γ_4^* is not polyhedral

Geometrical Lemma

If P is a polyhedral and a curve $c : [0,1] \to P$ has a right tangent $\dot{c}_{0+} = \lim_{p \to 0+} \frac{1}{p}[c_p - c_0]$, then P contains the segment with two endpoints c_0 and $c_0 + \epsilon \dot{c}_{0+}$ for some $\epsilon > 0$.



The closure of Γ_4^* is not polyhedral

Construction of the curve: 4 atoms

Consider four binary random variables: $\mathbf{X} = [x_1\ x_2\ x_3\ x_4]$, where x_1 and x_2 are independent, i.e. $I(x_1; x_2) = 0$, and the marginal distributions of x_1 and x_2 are
$p(x_1 = 0) = 2p$, $p(x_1 = 1) = 1 - 2p$ (where $0 \leq p \leq \frac{1}{2}$);
$p(x_2 = 0) = \frac{1}{2}$, $p(x_2 = 1) = \frac{1}{2}$.
Furthermore,
$x_3 = x_1 \cdot x_2$
$x_4 = (1 - x_1) \cdot (1 - x_2)$
4 atoms: For all the outcomes of **X**, only four or less have non-zero probability.

h_x^p: the entropy function of X

The closure of Γ_4^* is not polyhedral

Construction of the curve: h_x^p and r_t^l

$$\ln 2 \cdot c(p) = h_x^p + H(p) r_1^{14} + [\ln 2 + 2p \ln 2 - \frac{1}{2} H(2p)][r_1^{23} + r_2^4]$$

where $0 \leq p \leq \frac{1}{2}$, $\quad H(p) = -p \ln p - (1-p) \ln(1-p)$,
r_1^{14}, r_1^{23} and r_2^4 are linear matroids, also notice $h_x^0 = r_1^{14}$,
$h_x^{\frac{1}{2}} = r_1^{13}$.

After some calculation:

$$c(0) = r_1^{14} + r_1^{23} + r_2^4$$
$$\dot{c}_{0+} = f_{12} + r_1^{24} + r_2^3$$

The closure of Γ_4^* is not polyhedral

Contradiction

- $h(\epsilon) = c_0 + \epsilon \dot{c}_{0+} = [r_1^{14} + r_1^{23} + r_2^{4}] + \epsilon[f_{12} + r_1^{24} + r_2^{3}]$
- By the geometrical lemma, if the closure of Γ_4^* is polyhedral, then $h(\epsilon)$ should be entropic for some ϵ.
- Plug $h(\epsilon)$ in the sequence of Information Inequalities:

$$s\ Ingleton_{12} + \frac{s(s+1)}{2}[I(X_1;X_2|X_3) + I(X_1;X_3|X_2)]$$
$$+I(X_2;X_3|X_1) = 1 - \epsilon s$$

- For s large enough, $1 - \epsilon s \leq 0$, contradiction!

Thanks!

Questions!