Constraint information inequalities: geometric, algorithmic and combinatorial views

Andrei Romashchenko (LIRMM)

a joint work with
T. Kaced (Hong Kong) and N. Vereshchagin (Moscow)

Journées Calculabilités, Montpellier

April 29 2014

Outline

1. Shannon entropy: basic definitions
2. "Standard" information inequalities
3. "Conditional" information inequalities
4. Conditional inequalities: geometric view
5. Information inequalities for Kolmogorov complexity
6. Towards combinatorial interpretation

Outline

1. Shannon entropy: basic definitions
2. "Standard" information inequalities
3. "Conditional" information inequalities
4. Conditional inequalities: geometric view
5. Information inequalities for Kolmogorov complexity
6. Towards combinatorial interpretation

Shannon's entropy, the basic definition

Shannon's entropy, the basic definition

random variable α :

Shannon's entropy, the basic definition

random variable α : $\mathrm{Prob}[\alpha = s_i] = p_i$

Shannon's entropy, the basic definition

random variable α : $\mathrm{Prob}[\alpha = s_i] = p_i$

distribution: $\begin{array}{c|c|c|c} s_1 & s_2 & \dots & s_k \\ \hline p_1 & p_2 & \dots & p_k \end{array}$, $p_i \geq 0$, $\sum p_i = 1$

Shannon's entropy, the basic definition

random variable α : $\mathrm{Prob}[\alpha = s_i] = p_i$

distribution: $\begin{array}{c|c|c|c} s_1 & s_2 & \dots & s_k \\ \hline p_1 & p_2 & \dots & p_k \end{array}$, $p_i \geq 0$, $\sum p_i = 1$

Definition: $H(\alpha) := \sum\limits_i p_i \log \frac{1}{p_i}$

Shannon's entropy, the basic definition

random variable α : $\mathrm{Prob}[\alpha = s_i] = p_i$

distribution: $\begin{array}{c|c|c|c} s_1 & s_2 & \dots & s_k \\ \hline p_1 & p_2 & \dots & p_k \end{array}$, $p_i \geq 0$, $\sum p_i = 1$

Definition: $H(\alpha) := \sum\limits_i p_i \log \frac{1}{p_i}$

Measure of *uncertainty* in α : $0 \leq H(\alpha) \leq \log k$

Shannon's entropy, more notation

Shannon's entropy, more notation

random variables α, β

Shannon's entropy, more notation

random variables α, β

distribution: $\mathrm{Prob}[\alpha = s_i \ \& \ \beta = t_j] = p_{ij}$

Shannon's entropy, more notation

random variables α, β

distribution: $\mathrm{Prob}[\alpha = s_i \;\&\; \beta = t_j] = p_{ij}$

Principle entropy quantities: $H(\alpha)$, $H(\beta)$, $H(\alpha, \beta)$.

Shannon's entropy, more notation

random variables α, β

distribution: $\mathrm{Prob}[\alpha = s_i \,\&\, \beta = t_j] = p_{ij}$

Principle entropy quantities: $H(\alpha)$, $H(\beta)$, $H(\alpha, \beta)$.

Conditions entropies:

$$H(\alpha \mid \beta) = H(\alpha, \beta) - H(\beta),$$
$$H(\beta \mid \alpha) = H(\alpha, \beta) - H(\alpha).$$

Shannon's entropy, more notation

random variables α, β

distribution: $\mathrm{Prob}[\alpha = s_i \ \& \ \beta = t_j] = p_{ij}$

Principle entropy quantities: $H(\alpha)$, $H(\beta)$, $H(\alpha, \beta)$.

Conditions entropies:

$$H(\alpha \mid \beta) = H(\alpha, \beta) - H(\beta),$$
$$H(\beta \mid \alpha) = H(\alpha, \beta) - H(\alpha).$$

Mutual information:

$$I(\alpha : \beta) \ = \ H(\alpha) + H(\beta) - H(\alpha, \beta)$$

Shannon's entropy, more notation

random variables α, β

distribution: $\mathrm{Prob}[\alpha = s_i \,\&\, \beta = t_j] = p_{ij}$

Principle entropy quantities: $H(\alpha)$, $H(\beta)$, $H(\alpha, \beta)$.

Conditions entropies:

$$H(\alpha \mid \beta) = H(\alpha, \beta) - H(\beta),$$
$$H(\beta \mid \alpha) = H(\alpha, \beta) - H(\alpha).$$

Mutual information:

$$\begin{aligned} I(\alpha : \beta) &= H(\alpha) + H(\beta) - H(\alpha, \beta) \\ &= H(\alpha) - H(\alpha \mid \beta) \end{aligned}$$

Shannon's entropy, more notation

random variables α, β

distribution: $\mathrm{Prob}[\alpha = s_i \,\&\, \beta = t_j] = p_{ij}$

Principle entropy quantities: $H(\alpha)$, $H(\beta)$, $H(\alpha, \beta)$.

Conditions entropies:

$$H(\alpha \mid \beta) = H(\alpha, \beta) - H(\beta),$$
$$H(\beta \mid \alpha) = H(\alpha, \beta) - H(\alpha).$$

Mutual information:

$$\begin{aligned} I(\alpha : \beta) &= H(\alpha) + H(\beta) - H(\alpha, \beta) \\ &= H(\alpha) - H(\alpha \mid \beta) \\ &= H(\beta) - H(\beta \mid \alpha) \end{aligned}$$

Outline

Linear information inequalities

Linear information inequalities

Basic inequalities:

$H(a, b) \leq H(a) + H(b)$ $\quad [I(a : b) \geq 0]$

Linear information inequalities

Basic inequalities:

$$H(a,b) \leq H(a) + H(b) \qquad [I(a:b) \geq 0]$$

$$H(a,b,c) + H(c) \leq H(a,c) + H(b,c) \qquad [I(a:b\,|\,c) \geq 0]$$

Linear information inequalities

Basic inequalities:

$$H(a,b) \leq H(a) + H(b) \qquad [I(a:b) \geq 0]$$

$$H(a,b,c) + H(c) \leq H(a,c) + H(b,c) \quad [I(a:b\,|\,c) \geq 0]$$

[**Shannon type** ineq] == [combinations of basic ineq]:

example 1: $H(a) \leq H(a\,|\,b) + H(a\,|\,c) + I(b:c)$

example 2: $2H(a,b,c) \leq H(a,b) + H(a,c) + H(b,c)$

Linear information inequalities

General form: A linear information inequality is a combination of reals $\{\lambda_{i_1,\dots,i_k}\}$ such that

$$\sum \lambda_{i_1,\dots,i_k} H(a_{i_1},\dots,a_{i_k}) \geq 0$$

for all $(a_1,\dots,a_n)$.

Linear information inequalities

General form: A linear information inequality is a combination of reals $\{\lambda_{i_1,\dots,i_k}\}$ such that

$$\sum \lambda_{i_1,\dots,i_k} H(a_{i_1},\dots,a_{i_k}) \geq 0$$

for all $(a_1,\dots,a_n)$.

Applications:

- multi-source network coding
- secret sharing
- **combinatorial interpretations**
- group theoretical interpretation
- **Kolmogorov complexity**
- ...

Once again, Shannon type information inequalities :

- subadditivity,
  $H(A \cup B) \leq H(A) + H(B)$
  [in other notation $I(A : B) \geq 0$]
- submodularity,
  $H(A \cup B \cup C) + H(C) \leq H(A \cup C) + H(B \cup C)$
  [in other notation $I(A : B|C) \geq 0$]
- linear combinations of basic inequalities

Once again, Shannon type information inequalities :

- subadditivity,
  $H(A \cup B) \leq H(A) + H(B)$
  [in other notation $I(A : B) \geq 0$]
- submodularity,
  $H(A \cup B \cup C) + H(C) \leq H(A \cup C) + H(B \cup C)$
  [in other notation $I(A : B|C) \geq 0$]
- linear combinations of basic inequalities

Th [Z. Zhang, R.W. Yeung 1998] There exists a non-Shannon type information inequality:

$$\begin{aligned} I(c : d) \ \leq \ & 2I(c : d \,|\, a) + I(c : d \,|\, b) + I(a : b) \\ & + I(a : c \,|\, d) + I(a : d \,|\, c) \end{aligned}$$

Theorem [Z. Zhang, R.W. Yeung 1997] There exists a *conditional* non Shannon type inequality:

$$I(x:y) = I(x:y\,|\,a) = 0$$

$$\Downarrow$$

$$I(a:b) \le I(a:b\,|\,x) + I(a:b\,|\,y)$$

Outline

Conditional information inequalities

(a) **Trivial, Shannon-type**:

if $I(x : y) = 0$ then $H(a) \leq H(a \,|\, x) + H(a \,|\, y)$

Conditional information inequalities

(a) **Trivial, Shannon-type**:

if $I(x:y)=0$ then $H(a) \le H(a\,|\,x) + H(a\,|\,y)$

this is true since $H(a) \le H(a\,|\,x) + H(a\,|\,y) + I(x:y)$
[Shannon-type unconditional inequalitiy]

Conditional information inequalities

(b) **Trivial, non Shannon-type**:

if $I(a:b\,|\,z) = I(a:z\,|\,b) = I(b:z\,|\,a) = 0$ then

$$I(a:b) \le I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$

Conditional information inequalities

(b) **Trivial, non Shannon-type**:

if $I(a:b\,|\,z) = I(a:z\,|\,b) = I(b:z\,|\,a) = 0$ then

$$I(a:b) \le I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$

this is true since

$$\begin{aligned} I(a:b) \ \le\ & I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y) \\ & + I(a:b\,|\,z) + I(a:z\,|\,b) + I(b:z\,|\,a) \end{aligned}$$

[non Shannon-type unconditional inequalitiy]

Conditional information inequalities

(c) **Non trivial, non Shannon-type**:

- **Zhang,Yeung 97:** if $I(x : y) = I(x : y \mid a) = 0$ then
$$I(a : b) \leq I(a : b \mid x) + I(a : b \mid y) + I(x : y)$$

Conditional information inequalities

(c) **Non trivial, non Shannon-type**:

- **Zhang,Yeung 97:** if $I(x:y) = I(x:y\,|\,a) = 0$ then
$$I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$
- **F. Matúš 99:** if $I(x:a|b) = I(x:b|a) = 0$ then
$$I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$

Conditional information inequalities

(c) **Non trivial, non Shannon-type**:

- **Zhang,Yeung 97:** if $I(x:y) = I(x:y\,|\,a) = 0$ then
$$I(a:b) \le I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$
- **F. Matúš 99:** if $I(x:a|b) = I(x:b|a) = 0$ then
$$I(a:b) \le I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$
- **Tarik Kaced and A.R. 2011:** if $H(a\,|\,x,y) = I(x:y\,|\,a) = 0$ then
$$I(a:b) \le I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$

$$\underbrace{I(x:y) = I(x:y|a) = 0}_{\text{[Zhang–Yeung'97]}} \quad \underbrace{I(x:a|b) = I(x:b|a) = 0}_{\text{[Matúš'99]}} \quad \underbrace{H(a|x,y) = I(x:y|a) = 0}_{\text{[T.Kaced and A.R.'11]}}$$

$$\searrow \qquad \downarrow \qquad \swarrow$$

$$I(a:b) \le I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$

$$\underbrace{I(x:y) = I(x:y|a) = 0}_{\text{[Zhang–Yeung'97]}} \quad \underbrace{I(x:a|b) = I(x:b|a) = 0}_{\text{[Matúš'99]}} \quad \underbrace{H(a|x,y) = I(x:y|a) = 0}_{\text{[T.Kaced and A.R.'11]}}$$

$$\searrow \qquad \downarrow \qquad \swarrow$$

$$I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$

Theorem. These three statements are *essentially* conditional inequalities.

Theorem The inequality

$$H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \le I(a:b|x) + I(a:b|y) + I(x:y)$$

is *essentially* conditional.

Theorem The inequality

$$H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \le I(a:b|x) + I(a:b|y) + I(x:y)$$

is *essentially* conditional.

We cannot reduce it to an unconditional inequality!

Theorem The inequality

$$H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \leq I(a:b|x) + I(a:b|y) + I(x:y)$$

is *essentially* conditional.

We cannot reduce it to an unconditional inequality!

That is, for all λ_1, λ_2 the inequality

$$I(a:b) \leq I(a:b|x) + I(a:b|y) + I(x:y) + \lambda_1 H(a|x,y) + \lambda_2 I(x:y\,|\,a)$$

does not hold.

Theorem The inequality

$$H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \leq I(a:b|x) + I(a:b|y) + I(x:y)$$

is *essentially* conditional.

We cannot reduce it to an unconditional inequality!

More precisely, for all λ_1, λ_2 there exist (a, b, c, d) such that

$$I(a:b) \not\leq I(a:b|x) + I(a:b|y) + I(x:y) + \lambda_1 H(a|x,y) + \lambda_2 I(x:y\,|\,a)$$

Claim: For any λ_1, λ_2 there exist (a, b, c, d) such that

$$I(a:b) \not\leq I(a:b|x) + I(a:b|y) + I(x:y) + \lambda_1 H(a|x,y) + \lambda_2 I(x:y\mid a)$$

Claim: For any λ_1, λ_2 there exist (a, b, c, d) such that

$$I(a:b) \not\leq I(a:b|x) + I(a:b|y) + I(x:y) + \lambda_1 H(a|x,y) + \lambda_2 I(x:y \mid a)$$

Proof : a family of counter-examples

Claim: For any λ_1, λ_2 there exist (a, b, c, d) such that

$$I(a:b) \not\leq I(a:b|x) + I(a:b|y) + I(x:y) + \lambda_1 H(a|x,y) + \lambda_2 I(x:y\,|\,a)$$

Proof : a family of counter-examples

- fix a finite field $\mathbb{F}$
- a is a random *line* (a polynomial of degree 1 over $\mathbb{F}$)

Claim: For any λ_1, λ_2 there exist (a, b, c, d) such that

$$I(a:b) \not\leq I(a:b|x) + I(a:b|y) + I(x:y) + \lambda_1 H(a|x,y) + \lambda_2 I(x:y \mid a)$$

Proof : a family of counter-examples

- fix a finite field $\mathbb{F}$
- a is a random *line* (a polynomial of degree 1 over $\mathbb{F}$)
- x and y are two different *points* in this line

Claim: For any λ_1, λ_2 there exist (a, b, c, d) such that

$$I(a : b) \not\leq I(a : b|x) + I(a : b|y) + I(x : y) + \lambda_1 H(a|x, y) + \lambda_2 I(x : y \mid a)$$

Proof : a family of counter-examples

- fix a finite field $\mathbb{F}$
- a is a random *line* (a polynomial of degree 1 over $\mathbb{F}$)
- x and y are two different *points* in this line
- b is a *parabola* (a polynomial of degree 2) that intersects a at x and y

Claim: For any λ_1, λ_2 there exist (a, b, c, d) such that

$$I(a:b) \not\leq I(a:b|x) + I(a:b|y) + I(x:y) + \lambda_1 H(a|x,y) + \lambda_2 I(x:y\,|\,a)$$

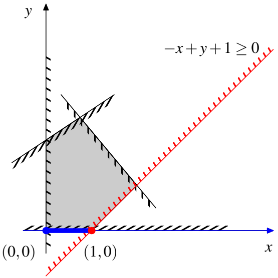
Proof : a family of counter-examples

- fix a finite field $\mathbb{F}$
- a is a random *line* (a polynomial of degree 1 over $\mathbb{F}$)
- x and y are two different *points* in this line
- b is a *parabola* (a polynomial of degree 2) that intersects a at x and y

$$\begin{array}{ccccccccccc} I(a:b) & \not\leq & I(a:b\,|\,x) & + & I(a:b\,|\,y) & + & I(x:y) & + & \lambda_1 H(a|x,y) & + & \lambda_2 I(x:y\,|\,a) \\ \| & & \| & & \| & & \| & & \| & & \| \\ 1+o(1) & \not\leq & o(1) & + & o(1) & + & o(1) & + & 0 & + & o(1) \end{array}$$

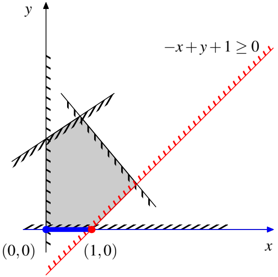
Outline

A geometric view on conditional inequalities:



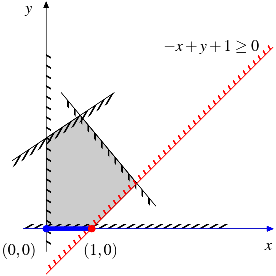
if $y = 0$ then $x \leq 1$

A geometric view on conditional inequalities:



if $y = 0$ then $x \leq 1 \quad \Longleftarrow x \leq 1 + y$

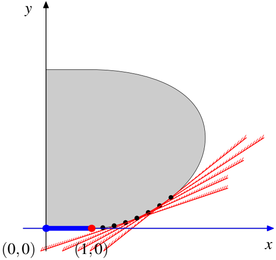
A geometric view on conditional inequalities:



if $y=0$ then $x\leq 1$ $\Longleftarrow$ $x\leq 1+y$

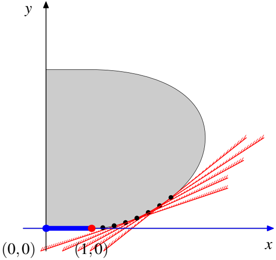
NOT essentially conditional

A geometric view on conditional inequalities:



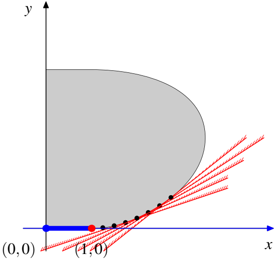
if $y = 0$ then $x \leq 1$

A geometric view on conditional inequalities:



if $y = 0$ then $x \leq 1$ $\Longleftarrow$ from an *infinite* family of linear inequalities

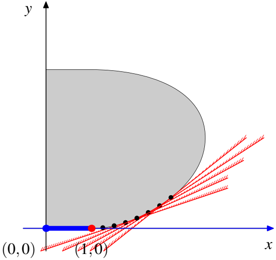
A geometric view on conditional inequalities:



if $y = 0$ then $x \leq 1 \Longleftarrow$ from an *infinite* family of linear inequalities

NO unconditional inequality $x \leq 1 + \lambda y$

A geometric view on conditional inequalities:

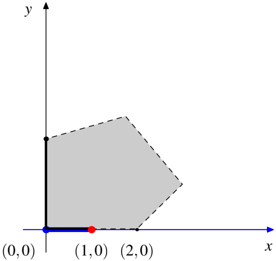


if $y = 0$ then $x \leq 1$ $\Longleftarrow$ from an *infinite* family of linear inequalities

NO unconditional inequality $x \leq 1 + \lambda y$

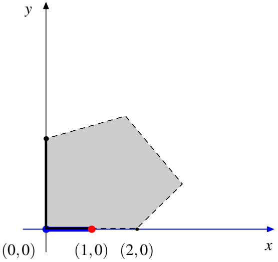
this inequality is essentially conditional

A geometric view on conditional inequalities:



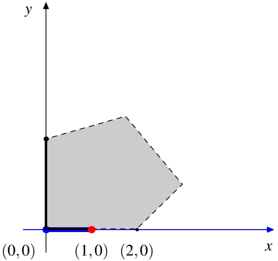
if $y = 0$ then $x \leq 1$

A geometric view on conditional inequalities:



if $y = 0$ then $x \leq 1$ $\Longleftarrow$ from a complex structure of the borderline

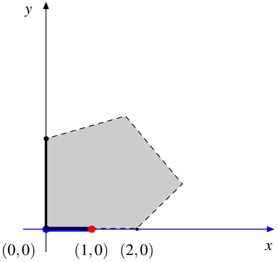
A geometric view on conditional inequalities:



if $y = 0$ then $x \leq 1$ $\Longleftarrow$ from a complex structure of the borderline

NO unconditional inequality $x \leq 1 + \lambda y$

A geometric view on conditional inequalities:



if $y = 0$ then $x \leq 1$ $\Longleftarrow$ from a complex structure of the borderline

NO unconditional inequality $x \leq 1 + \lambda y$

this inequality is also essentially conditional

Outline

1. Shannon entropy: basic definitions
2. "Standard" information inequalities
3. "Conditional" information inequalities
4. Conditional inequalities: geometric view
5. Information inequalities for Kolmogorov complexity
6. Towards combinatorial interpretation

Inequalities for Kolmogorov complexity:

- Exactly the same classes of unconditional linear inequalities hold for Shannon's entropy and for Kolmogorov complexity.

Inequalities for Kolmogorov complexity:

- Exactly the same classes of unconditional linear inequalities hold for Shannon's entropy and for Kolmogorov complexity.

e.g. $H(a_1, a_2) \leq H(a_1) + H(a_2)$ vs $C(a_1, a_2) \leq C(a_1) + C(a_2) + O(\log N)$

Inequalities for Kolmogorov complexity:

- Exactly the same classes of unconditional linear inequalities hold for Shannon's entropy and for Kolmogorov complexity.

e.g. $H(a_1, a_2) \leq H(a_1) + H(a_2)$ vs $C(a_1, a_2) \leq C(a_1) + C(a_2) + O(\log N)$

the general scheme: $\lambda_1 H(a_1) + \lambda_2 H(a_2) + \ldots + \lambda_{12} H(a_1, a_2) + \ldots \geq 0$

is equivalent to

$$\lambda_1 C(a_1) + \lambda_2 C(a_2) + \ldots + \lambda_{12} C(a_1, a_2) + \ldots + O(\log N) \geq 0$$

Inequalities for Kolmogorov complexity:

- Exactly the same classes of unconditional linear inequalities hold for Shannon's entropy and for Kolmogorov complexity.

  e.g. $H(a_1, a_2) \leq H(a_1) + H(a_2)$ vs $C(a_1, a_2) \leq C(a_1) + C(a_2) + O(\log N)$

  the general scheme: $\lambda_1 H(a_1) + \lambda_2 H(a_2) + \ldots + \lambda_{12} H(a_1, a_2) + \ldots \geq 0$

  is equivalent to

  $$\lambda_1 C(a_1) + \lambda_2 C(a_2) + \ldots + \lambda_{12} C(a_1, a_2) + \ldots + O(\log N) \geq 0$$

- Essentially conditional inequality [Matúš'99] is valid for Kolmogorov complexity (in *some* sense).

Inequalities for Kolmogorov complexity:

- Exactly the same classes of unconditional linear inequalities hold for Shannon's entropy and for Kolmogorov complexity.

e.g. $H(a_1, a_2) \le H(a_1) + H(a_2)$ vs $C(a_1, a_2) \le C(a_1) + C(a_2) + O(\log N)$

the general scheme: $\lambda_1 H(a_1) + \lambda_2 H(a_2) + \ldots + \lambda_{12} H(a_1, a_2) + \ldots \ge 0$

is equivalent to

$$\lambda_1 C(a_1) + \lambda_2 C(a_2) + \ldots + \lambda_{12} C(a_1, a_2) + \ldots + O(\log N) \ge 0$$

- Essentially conditional inequality [Matúš'99] is valid for Kolmogorov complexity (in *some* sense).

$I(x : a|b) \le \sqrt{N}$ & $I(x : b|a) \le \sqrt{N} \Rightarrow I(a : b) \le I(a : b \mid x) + I(a : b \mid y) + I(x : y) + O(N^{3/4})$

Inequalities for Kolmogorov complexity:

- Exactly the same classes of unconditional linear inequalities hold for Shannon's entropy and for Kolmogorov complexity.

e.g. $H(a_1, a_2) \leq H(a_1) + H(a_2)$ vs $C(a_1, a_2) \leq C(a_1) + C(a_2) + O(\log N)$

the general scheme: $\lambda_1 H(a_1) + \lambda_2 H(a_2) + \ldots + \lambda_{12} H(a_1, a_2) + \ldots \geq 0$

is equivalent to

$$\lambda_1 C(a_1) + \lambda_2 C(a_2) + \ldots + \lambda_{12} C(a_1, a_2) + \ldots + O(\log N) \geq 0$$

- Essentially conditional inequality [Matúš'99] is valid for Kolmogorov complexity (in *some* sense).

$I(x:a|b) \leq \sqrt{N}$ & $I(x:b|a) \leq \sqrt{N} \Rightarrow I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y) + O(N^{3/4})$

- Essentially conditional inequalities [Zhang–Yeung'97] and [Kaced–R.'11] are **not** for Kolmogorov complexity

Inequalities for Kolmogorov complexity:

- Exactly the same classes of unconditional linear inequalities hold for Shannon's entropy and for Kolmogorov complexity.

  e.g. $H(a_1, a_2) \leq H(a_1) + H(a_2)$ vs $C(a_1, a_2) \leq C(a_1) + C(a_2) + O(\log N)$

  the general scheme: $\lambda_1 H(a_1) + \lambda_2 H(a_2) + \ldots + \lambda_{12} H(a_1, a_2) + \ldots \geq 0$

  is equivalent to

  $$\lambda_1 C(a_1) + \lambda_2 C(a_2) + \ldots + \lambda_{12} C(a_1, a_2) + \ldots + O(\log N) \geq 0$$

- Essentially conditional inequality [Matúš'99] is valid for Kolmogorov complexity (in *some* sense).

  $I(x : a|b) \leq \sqrt{N}$ & $I(x : b|a) \leq \sqrt{N} \Rightarrow I(a : b) \leq I(a : b \mid x) + I(a : b \mid y) + I(x : y) + O(N^{3/4})$

- Essentially conditional inequalities [Zhang–Yeung'97] and [Kaced–R.'11] are **not** for Kolmogorov complexity (in any reasonable sense)

Outline

Once again,

Theorem: $H(a|x, y) = I(x : y|a) = 0 \Rightarrow I(a : b) \leq I(a : b \mid x) + I(a : b \mid y) + I(x : y)$

Once again,

Theorem: $H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$

Why is it valid?

Once again,

Theorem: $H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$

Why is it valid? And what does it mean?

Once again,

Theorem: $H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$

Why is it valid? And what does it mean?

We relax the constraint and make the inequality stronger:

(1) a is a function of (x, y)

(2) $\Pr[X_i\,|\,A_k] > 0$ & $\Pr[Y_j\,|\,A_k] > 0 \Rightarrow \Pr[X_i, Y_j\,|\,A_k] > 0$

Theorem:

$$(1) + (2) \implies H(a\,|\,x,b) + H(a\,|\,y,b) \leq H(a\,|\,b)$$

Once again,

Theorem: $H(a|x,y) = I(x:y|a) = 0 \Rightarrow I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$

Why is it valid? And what does it mean?

We relax the constraint and make the inequality stronger:

(1) a is a function of (x,y)

(2) $\Pr[X_i\,|\,A_k] > 0$ & $\Pr[Y_j\,|\,A_k] > 0 \Rightarrow \Pr[X_i, Y_j\,|\,A_k] > 0$

Theorem:

$$(1) + (2) \implies H(a\,|\,x,b) + H(a\,|\,y,b) \leq H(a\,|\,b)$$

$$\Downarrow$$

$$I(a:b) \leq I(a:b\,|\,x) + I(a:b\,|\,y) + I(x:y)$$

The same statement in terms of graphs:

- $G =$ bi-partite graph

The same statement in terms of graphs:

- G = bi-partite graph
- each edge has a color

The same statement in terms of graphs:

- G = bi-partite graph
- each edge has a color

From a graph to a distribution:

- take a random edge

The same statement in terms of graphs:

- G = bi-partite graph
- each edge has a color

From a graph to a distribution:

- take a random edge
- x = the left end of the edge
- y = the right end of the edge

The same statement in terms of graphs:

- G = bi-partite graph
- each edge has a color

From a graph to a distribution:

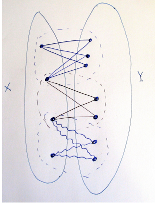
- take a random edge
- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge

The same statement in terms of graphs:

- G = bi-partite graph
- each edge has a color

From a graph to a distribution:

- take a random edge
- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge

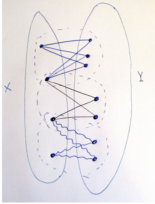


The same statement in terms of graphs:

- G = bi-partite graph
- each edge has a color

From a graph to a distribution:

- take a random edge
- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge



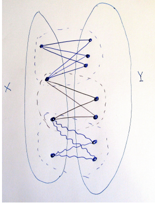
And b is whatever you want!

The same statement in terms of graphs:

- G = bi-partite graph
- each edge has a color

From a graph to a distribution:

- take a random edge
- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge



And b is whatever you want!

Our conditions:

(1) a is uniquely defined by x and y

(2) edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b)$

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b)$

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b)$

Proof:

- keep the distribution of (a, b),

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b)$

Proof:

- keep the distribution of (a, b), take x and y independently given (a, b)

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a \mid x, b) + H(a \mid y, b) \le H(a \mid b)$

Proof:

- keep the distribution of (a, b), take x and y independently given (a, b)
- trivial: $H(a, b, x', y') = H(a, b) + H(x \mid a, b) + H(y \mid a, b)$

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a \mid x, b) + H(a \mid y, b) \le H(a \mid b)$

Proof:

- keep the distribution of (a, b), take x and y independently given (a, b)
- trivial: $H(a, b, x', y') = H(a, b) + H(x \mid a, b) + H(y \mid a, b)$
- evident: $H(a, b, x', y') \le H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y')$

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a \mid x, b) + H(a \mid y, b) \le H(a \mid b)$

Proof:

- keep the distribution of (a, b), take x and y independently given (a, b)
- trivial: $H(a, b, x', y') = H(a, b) + H(x \mid a, b) + H(y \mid a, b)$
- evident: $H(a, b, x', y') \le H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y')$
- simple: $H(a, b, x', y') \le H(b) + H(x \mid b) + H(y \mid b)$ ~~$+ H(a \mid x', y')$~~

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b)$

Proof:

- keep the distribution of (a,b), take x and y independently given (a,b)
- trivial: $H(a,b,x',y') = H(a,b) + H(x\,|\,a,b) + H(y\,|\,a,b)$
- evident: $H(a,b,x',y') \le H(b) + H(x\,|\,b) + H(y\,|\,b) + H(a\,|\,x',y')$
- simple: $H(a,b,x',y') \le H(b) + H(x\,|\,b) + H(y\,|\,b)$ ~~$+H(a\,|\,x',y')$~~
- result: $H(a,b) + H(x\,|\,a,b) + H(y\,|\,a,b) \le H(b) + H(x\,|\,b) + H(y\,|\,b)$

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- edges of each color make a clique

Th. (1) + (2) $\Rightarrow H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b)$

Proof:

- keep the distribution of (a,b), take x and y independently given (a,b)
- trivial: $H(a,b,x',y') = H(a,b) + H(x\,|\,a,b) + H(y\,|\,a,b)$
- evident: $H(a,b,x',y') \le H(b) + H(x\,|\,b) + H(y\,|\,b) + H(a\,|\,x',y')$
- simple: $H(a,b,x',y') \le H(b) + H(x\,|\,b) + H(y\,|\,b) ~~+ H(a\,|\,x',y')~~$
- result: $H(a,b) + H(x\,|\,a,b) + H(y\,|\,a,b) \le H(b) + H(x\,|\,b) + H(y\,|\,b)$
- which implies $H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b)$

So what?

So what? Why do we need all these inequalities?

This is all about (bi-)clique covering!

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.

Assume that edges of each color can be covered by N bi-cliques

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- $x =$ the left end of the edge
- $y =$ the right end of the edge
- $a =$ the color of the edge
- $w =$ the index of a bi-clique covering the edge

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a \mid x, b, w) + H(a \mid y, b, w) \leq H(a \mid b, w)$.

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a \mid x, b, w) + H(a \mid y, b, w) \leq H(a \mid b, w)$.
hence $H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) + 2H(w)$

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a \,|\, x, b, w) + H(a \,|\, y, b, w) \leq H(a \,|\, b, w)$.
hence $H(a \,|\, x, b) + H(a \,|\, y, b) \leq H(a \,|\, b) + 2H(w)$
it follows: $H(a \,|\, x, b) + H(a \,|\, y, b) \leq H(a \,|\, b) + 2 \log N$

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a \mid x, b, w) + H(a \mid y, b, w) \leq H(a \mid b, w)$.
hence $H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) + 2H(w)$
it follows: $H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) + 2 \log N$

Problem: Given G, we want to estimate its *bi-clique covering number*

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a \,|\, x, b, w) + H(a \,|\, y, b, w) \leq H(a \,|\, b, w)$.
hence $H(a \,|\, x, b) + H(a \,|\, y, b) \leq H(a \,|\, b) + 2H(w)$
it follows: $H(a \,|\, x, b) + H(a \,|\, y, b) \leq H(a \,|\, b) + 2 \log N$

Problem: Given G, we want to estimate its *bi-clique covering number*

Recipe:

- take a distribution (a, x, y)

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a\,|\,x,b,w) + H(a\,|\,y,b,w) \le H(a\,|\,b,w)$.
hence $H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b) + 2H(w)$
it follows: $H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b) + 2\log N$

Problem: Given G, we want to estimate its *bi-clique covering number*
Recipe:

- take a distribution (a,x,y) [e.g., a uniform distribution on edges]

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a\,|\,x, b, w) + H(a\,|\,y, b, w) \le H(a\,|\,b, w)$.
hence $H(a\,|\,x, b) + H(a\,|\,y, b) \le H(a\,|\,b) + 2H(w)$
it follows: $H(a\,|\,x, b) + H(a\,|\,y, b) \le H(a\,|\,b) + 2\log N$

Problem: Given G, we want to estimate its *bi-clique covering number*
Recipe:

- take a distribution (a, x, y) [e.g., a uniform distribution on edges]
- add a suitable b

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a\,|\,x,b,w) + H(a\,|\,y,b,w) \le H(a\,|\,b,w)$.
hence $H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b) + 2H(w)$
it follows: $H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b) + 2\log N$

Problem: Given G, we want to estimate its *bi-clique covering number*
Recipe:

- take a distribution (a,x,y) [e.g., a uniform distribution on edges]
- add a suitable b [don't ask me how to invent this b]

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a \mid x, b, w) + H(a \mid y, b, w) \leq H(a \mid b, w)$.
hence $H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) + 2H(w)$
it follows: $H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) + 2 \log N$

Problem: Given G, we want to estimate its *bi-clique covering number*
Recipe:

- take a distribution (a, x, y) [e.g., a uniform distribution on edges]
- add a suitable b [don't ask me how to invent this b]
- observe $H(a \mid x, b) + H(a \mid y, b) \not\leq H(a \mid b)$

This is all about (bi-)clique covering!

Let G be bi-partite graph with colored edges.
Assume that edges of each color can be covered by N bi-cliques

- x = the left end of the edge
- y = the right end of the edge
- a = the color of the edge
- w = the index of a bi-clique covering the edge

then $H(a\,|\,x,b,w) + H(a\,|\,y,b,w) \le H(a\,|\,b,w)$.
hence $H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b) + 2H(w)$
it follows: $H(a\,|\,x,b) + H(a\,|\,y,b) \le H(a\,|\,b) + 2\log N$

Problem: Given G, we want to estimate its *bi-clique covering number*
Recipe:

- take a distribution (a,x,y) [e.g., a uniform distribution on edges]
- add a suitable b [don't ask me how to invent this b]
- observe $H(a\,|\,x,b) + H(a\,|\,y,b) \not\le H(a\,|\,b)$
- then [bi-clique covering number] $\ge 2^{(H(a\,|\,x,b)+H(a\,|\,y,b)-H(a\,|\,b))/2}$

Open problems:

1. **[geometry]** The form of the cone of "entropic" points: infinitely many flat facets? or a curved surface?
2. **[complexity]** Another conditional inequality by F. Matúš: is it valid for Kolmogorov complexity?
3. **[combinatorics/complexity]** Lower bounds for clique covering: find examples where this technique is more effective then the conventional arguments (applications to communication complexity).