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

\[ H(\alpha) := \sum_i p_i \log \frac{1}{p_i} \]

Measure of uncertainty in \( \alpha \): 0 ≤ \( H(\alpha) \) ≤ \( \log k \)

Andrei Romashchenko
Constraint Information Inequalities
April 29 2014 4 / 31
Shannon’s entropy, the basic definition

random variable $\alpha$:

\[
\begin{align*}
\text{Prob} \left[ \alpha = s_i \right] &= p_i \\
\text{distribution:} &
s_1 s_2 \ldots s_k \\
p_1 &\geq 0, \sum p_i = 1
\end{align*}
\]

Definition:
\[
H(\alpha) := \sum_i p_i \log \frac{1}{p_i}
\]

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

Andrei Romashchenko
Constraint Information Inequalities
April 29 2014
Shannon’s entropy, the basic definition

random variable $\alpha : \text{Prob}[\alpha = s_i] = p_i$
Shannon’s entropy, the basic definition

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

distribution:

<table>
<thead>
<tr>
<th>$s_1$</th>
<th>$s_2$</th>
<th>\ldots</th>
<th>$s_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_1$</td>
<td>$p_2$</td>
<td>\ldots</td>
<td>$p_k$</td>
</tr>
</tbody>
</table>

$p_i \geq 0$, $\sum p_i = 1$
Shannon’s entropy, the basic definition

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

distribution: $\begin{array}{ccc} s_1 & s_2 & \cdots & s_k \\ p_1 & p_2 & \cdots & p_k \end{array}$, $p_i \geq 0$, $\sum p_i = 1$

Definition: $H(\alpha) := \sum_i p_i \log \frac{1}{p_i}$
Shannon’s entropy, the basic definition

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

distribution: $\begin{array}{cccc}
s_1 & s_2 & \ldots & s_k \\
p_1 & p_2 & \ldots & p_k \\
\end{array}$, $p_i \geq 0$, $\sum p_i = 1$

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

Measure of uncertainty in $\alpha$ : $0 \leq H(\alpha) \leq \log k$
Shannon’s entropy, more notation

\[ \text{Pr} \left[ \alpha = s_i \land \beta = t_j \right] = p_{ij} \]

Principle entropy quantities:

\[ H(\alpha), H(\beta), H(\alpha, \beta) \]

Conditions entropies:

\[ H(\alpha | \beta) = H(\alpha, \beta) - H(\beta) \]
\[ H(\beta | \alpha) = H(\alpha, \beta) - H(\alpha) \]

Mutual information:

\[ I(\alpha : \beta) = H(\alpha) + H(\beta) - H(\alpha, \beta) \]
\[ = H(\alpha) - H(\alpha | \beta) \]
\[ = H(\beta) - H(\beta | \alpha) \]
Shannon’s entropy, more notation
random variables $\alpha, \beta$

Distribution: $\text{Prob}[\alpha = s, \beta = t] = p_{ij}$

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

Conditions entropies:
- $H(\alpha | \beta) = H(\alpha, \beta) - H(\beta)$
- $H(\beta | \alpha) = H(\alpha, \beta) - H(\alpha)$

Mutual information:
- $I(\alpha : \beta) = H(\alpha) + H(\beta) - H(\alpha, \beta)$
Shannon’s entropy, more notation

random variables $\alpha, \beta$

distribution: $\text{Prob}[\alpha = s_i \& \beta = t_j] = p_{ij}$
Shannon’s entropy, more notation
random variables $\alpha, \beta$
distribution: $\text{Prob}[\alpha = s_i \& \beta = t_j] = p_{ij}$
Principle entropy quantities: $H(\alpha), H(\beta), H(\alpha, \beta)$. 
Shannon’s entropy, more notation
random variables $\alpha, \beta$
distribution: $\text{Prob}[\alpha = s_i \& \beta = t_j] = p_{ij}$

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 $\alpha, \beta$

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

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

Conditions entropies:

\[
H(\alpha | \beta) = H(\alpha, \beta) - H(\beta),
\]
\[
H(\beta | \alpha) = H(\alpha, \beta) - H(\alpha).
\]

Mutual information:

\[
I(\alpha : \beta) = H(\alpha) + H(\beta) - H(\alpha, \beta)
\]
Shannon’s entropy, more notation
random variables $\alpha, \beta$
distribution: $\text{Prob}[\alpha = s_i \& \beta = t_j] = p_{ij}$

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:

$$I(\alpha : \beta) = H(\alpha) + H(\beta) - H(\alpha, \beta)$$
$$= H(\alpha) - H(\alpha \mid \beta)$$
Shannon’s entropy, more notation
random variables $\alpha, \beta$
distribution: $\text{Prob}[\alpha = s_i \& \beta = t_j] = p_{ij}$

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:

\[ I(\alpha : \beta) = H(\alpha) + H(\beta) - H(\alpha, \beta) \]
\[ = H(\alpha) - H(\alpha \mid \beta) \]
\[ = H(\beta) - H(\beta \mid \alpha) \]
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
Linear information inequalities

Basic inequalities:

\[ H(a, b) \leq H(a) + H(b) \]

\[ I(a:b) \geq 0 \]

\[ H(a, b, c) + H(c) \leq H(a, c) + H(b, c) \]

\[ I(a:b|c) \geq 0 \]

Shannon type inequations:

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

**Basic inequalities:**

\[ H(a, b) \leq H(a) + H(b) \]

\[ I(a : b) \geq 0 \]
Linear information inequalities

Basic inequalities:

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

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

**Basic inequalities:**

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

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

[**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 \( \{\lambda_{i_1, \ldots, i_k}\} \) such that

\[
\sum \lambda_{i_1, \ldots, i_k} H(a_{i_1}, \ldots, a_{i_k}) \geq 0
\]

for all \( (a_1, \ldots, a_n) \).
Linear information inequalities

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

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

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

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:

\[ I(c : d) \leq 2I(c : d | a) + I(c : d | b) + I(a : b) + I(a : c | d) + I(a : d | c) \]
Theorem [Z. Zhang, R.W. Yeung 1997] There exists a conditional non Shannon type inequality:

\[ I(x : y) = I(x : y \mid a) = 0 \]

\[ \Downarrow \]

\[ I(a : b) \leq I(a : b \mid x) + I(a : b \mid y) \]
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
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 $H(a) \leq H(a \mid x) + H(a \mid y)$

this is true since $H(a) \leq H(a \mid x) + H(a \mid y) + I(x : y)$

[Shannon-type unconditional inequality]
Conditional information inequalities

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

If \( I(a : b | z) = I(a : z | b) = I(b : z | a) = 0 \) then

\[
I(a : b) \leq I(a : b | x) + I(a : b | y) + I(x : y)
\]
Conditional information inequalities

(b) Trivial, non Shannon-type:

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

\[
I(a : b) \leq I(a : b | x) + I(a : b | y) + I(x : y)
\]

this is true since

\[
I(a : b) \leq I(a : b | x) + I(a : b | y) + I(x : y) + I(a : b | z) + I(a : z | b) + I(b : z | a)
\]

[non Shannon-type unconditional inequality]
Conditional information inequalities

(c) Non trivial, non Shannon-type:

- **Zhang, Yeung 97**: if \( I(x : y) = I(x : y | a) = 0 \) then
  \[
  I(a : b) \leq I(a : b | x) + I(a : b | y) + I(x : y)
  \]
Conditional information inequalities

(c) Non trivial, non Shannon-type:

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

- **F. Matúš 99**: if \( I(x : a \mid b) = I(x : b \mid a) = 0 \) 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 
  \[
  I(a : b) \leq I(a : b | x) + I(a : b | y) + I(x : y)
  \]

- **Tarik Kaced and A.R. 2011**: if 
  \[
  H(a | x, y) = I(x : y | a) = 0
  \]
  \[
  I(a : b) \leq I(a : b | x) + I(a : b | y) + I(x : y)
  \]
\[
\begin{align*}
I(x : y) = I(x : y|a) &= 0 \quad \blacktriangleright \quad [\text{Zhang–Yeung'97}] \\
I(x : a|b) = I(x : b|a) &= 0 \quad \blacktriangleright \quad [\text{Matúš'99}] \\
H(a|x, y) = I(x : y|a) &= 0 \quad \blacktriangleright \quad [\text{T.Kaced and A.R.'11}]
\end{align*}
\]

\[
I(a : b) \leq I(a : b|x) + I(a : b|y) + I(x : y)
\]
\[ I(x : y) = I(x : y \mid a) = 0 \]
\[ I(x : a \mid b) = I(x : b \mid a) = 0 \]
\[ H(a \mid x, y) = I(x : y \mid a) = 0 \]

\[ I(a : b) \leq I(a : b \mid x) + I(a : b \mid y) + I(x : y) \]

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

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

is essentially conditional.
Theorem The inequality

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

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 \( \lambda_1, \lambda_2 \) 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 \( \lambda_1, \lambda_2 \) 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 $\lambda_1, \lambda_2$ 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 $\lambda_1, \lambda_2$ 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
Claim: For any $\lambda_1, \lambda_2$ 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 \( \lambda_1, \lambda_2 \) 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
Claim: For any $\lambda_1, \lambda_2$ there exist $(a, b, c, d)$ such that

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

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 $y$
**Claim:** For any $\lambda_1, \lambda_2$ 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 $y$

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

$$\parallel \quad \parallel \quad \parallel \quad \parallel \quad \parallel$$

$$1 + o(1) \not\leq o(1) + o(1) + o(1) + 0 + o(1)$$
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
A geometric view on conditional inequalities:

\[-x + y + 1 \geq 0\]

if \(y = 0\) then \(x \leq 1\)
A geometric view on conditional inequalities:

\[-x + y + 1 \geq 0\]

if \(y = 0\) then \(x \leq 1\) \(\iff\) \(x \leq 1 + y\)
A geometric view on conditional inequalities:

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

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 \( x \leq 1 \) \( \iff \) from an *infinite* family of linear inequalities
A geometric view on conditional inequalities:

if \( y = 0 \) then \( x \leq 1 \) \iff \text{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 \( x \leq 1 \) \iff \text{from an } infinite \text{ 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 \( x \leq 1 \) \( \iff \) from a complex structure of the borderline
A geometric view on conditional inequalities:

if \( y = 0 \) then \( x \leq 1 \) \iff \text{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 \(x \leq 1\) \iff \text{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.

\[ H(a_1, a_2) \leq H(a_1) + H(a_2) \leq C(a_1, a_2) \leq C(a_1) + C(a_2) + O(\log N) \]
Inequalities for Kolmogorov complexity:

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

  \[ H(a_1, a_2) \leq H(a_1) + H(a_2) \text{ vs } C(a_1, a_2) \leq C(a_1) + C(a_2) + O(\log N) \]
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. $H(a_1, a_2) \leq H(a_1) + H(a_2)$ 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).

  $I(x : a | b) \leq \sqrt{N}$ & $I(x : b | a) \leq \sqrt{N} \Rightarrow I(a : b) \leq I(a : b | x) + I(a : b | y) + I(x : y) + O(N^{3/4})$
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 [Matůš'99] is valid for Kolmogorov complexity (in some sense).

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

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
Once again,

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

- $G = \text{bi-partite graph}$
The same statement in terms of graphs:

- $G =$ bi-partite graph
- each edge has a color
The same statement in terms of graphs:

- $G = \text{bi-partite graph}$
- 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:

- $G = \text{bi-partite graph}$
- each edge has a color

From a graph to a distribution:

- take a random edge
- $x = \text{the left end of the edge}$
- $y = \text{the right end of the edge}$
- $a = \text{the color of the edge}$
The same statement in terms of graphs:

- $G = \text{bi-partite graph}$
- each edge has a color

From a graph to a distribution:

- take a random edge
- $x = \text{the left end of the edge}$
- $y = \text{the right end of the edge}$
- $a = \text{the color of the edge}$

Andrei Romashchenko
Constraint Information Inequalities
April 29 2014 27 / 31
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:

- $G = \text{bi-partite graph}$
- each edge has a color

From a graph to a distribution:

- take a random edge
- $x = \text{the left end of the edge}$
- $y = \text{the right end of the edge}$
- $a = \text{the color of the edge}$

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) ⇒ $H(a|x, b) + H(a|y, b) ≤ H(a|b)$**
• $x =$ 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) $\Rightarrow H(a| x, b) + H(a| y, b) \leq H(a| b)$**

**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 $(a, b)$
\[ x = \text{the left end of the edge} \]
\[ y = \text{the right end of the edge} \]
\[ a = \text{the color of the edge} \]
\[ \text{edges of each color make a clique} \]

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

**Proof:**

- keep the distribution of \((a, b)\), take \(x\) and \(y\) independently given \((a, b)\)
- trivial: \(H(a, b, x', y') = H(a, b) + H(x | a, b) + H(y | a, b)\)
\[ x = \text{the left end of the edge} \]
\[ y = \text{the right end of the edge} \]
\[ a = \text{the color of the edge} \]
\[ \text{edges of each color make a clique} \]

\textbf{Th. (1) + (2) ⇒} \[ H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) \]

\textbf{Proof:}

\[ \text{keep the distribution of } (a, b), \text{ take } x \text{ and } y \text{ independently given } (a, b) \]
\[ \text{trivial: } H(a, b, x', y') = H(a, b) + H(x \mid a, b) + H(y \mid a, b) \]
\[ \text{evident: } H(a, b, x', y') \leq H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y') \]
\( x = \) 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) \leq 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: \(H(a, b, x', y') \leq H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y')\)
- simple: \(H(a, b, x', y') \leq H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y')\)
\[ x = \text{the left end of the edge} \]
\[ y = \text{the right end of the edge} \]
\[ a = \text{the color of the edge} \]

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:

1. keep the distribution of \((a, b)\), take \(x\) and \(y\) independently given \((a, b)\)
2. trivial: \(H(a, b, x', y') = H(a, b) + H(x \mid a, b) + H(y \mid a, b)\)
3. evident: \(H(a, b, x', y') \leq H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y')\)
4. simple: \(H(a, b, x', y') \leq H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y')\)
5. result: \(H(a, b) + H(x \mid a, b) + H(y \mid a, b) \leq H(b) + H(x \mid b) + H(y \mid b)\)
\( x = \) 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) \( \Rightarrow \) \( H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) \)

**Proof:**

- keep the distribution of \((a, b)\), take \(x\) and \(y\) independently given \((a, b)\)
- trivial: \( H(a, b, x', y') = H(a, b) + H(x \mid a, b) + H(y \mid a, b) \)
- evident: \( H(a, b, x', y') \leq H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y') \)
- simple: \( H(a, b, x', y') \leq H(b) + H(x \mid b) + H(y \mid b) + H(a \mid x', y') \)
- result: \( H(a, b) + H(x \mid a, b) + H(y \mid a, b) \leq H(b) + H(x \mid b) + H(y \mid b) \)
- which implies \( H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) \)
So what?
So what? Why do we need all these inequalities?
This is all about (bi-)clique covering!

Let $G$ be a bi-partite graph with colored edges. Assume that edges of each color can be covered by $N$ bi-cliques. Let $x$ be the left end of the edge, $y$ be the right end of the edge, $a$ be the color of the edge, and $w$ be 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 $(a,x,y)$ [e.g., a uniform distribution on edges] and add a suitable $b$ [don't ask me how to invent this]. Observe

$$H(a|x,b) + H(a|y,b) \neq H(a|b).$$

Then the bi-clique covering number $\geq 2(H(a|x,b) + H(a|y,b) - H(a|b))/2$. 

Andrei Romashchenko
Constraint Information Inequalities
April 29 2014 30 / 31
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

\[
H(a|x, b, w) + H(a|y, b, w) \leq H(a|b, w)
\]

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

\[
H(a|x, b) + H(a|y, b) \neq H(a|b)
\]

then [bi-clique covering number] $\geq 2(H(a|x, b) + H(a|y, b) - H(a|b))/2$
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 $H(a \mid x, b, w) + H(a \mid y, b, w) \leq H(a \mid b, w)$. 

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 $H(a \mid x, b, w) + H(a \mid y, b, w) \neq H(a \mid b, w)$ then [bi-clique covering number] $\geq 2 \left( H(a \mid x, b, w) + H(a \mid y, b, w) - H(a \mid b, w) \right) / 2$
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

\[ H(a \mid x, b, w) + H(a \mid y, b, w) \leq H(a \mid b, w). \]

\[ \text{hence} \ H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) + 2H(w) \]
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 $H(a \mid x, b, w) + H(a \mid y, b, w) \leq H(a \mid b, w)$.

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

it follows: $H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b) + 2 \log N$
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
This is all about (bi-)clique covering!

Let \( G \) be a 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 \((a, x, y)\)
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]
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 $b$
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 $b$]
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 $H(a \mid x, b) + H(a \mid y, b) \leq H(a \mid b)$
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 $H(a | x, b, w) + H(a | y, b, w) \leq H(a | b, w)$.

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 $(a, x, y)$ [e.g., a uniform distribution on edges]
- add a suitable $b$ [don’t ask me how to invent this $b$]
- observe $H(a | x, b) + H(a | y, b) \not\leq H(a | b)$
- then [bi-clique covering number] $\geq 2(H(a | x, b) + H(a | y, b) - H(a | b))/2$
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 than the conventional arguments (applications to communication complexity).