On Parallels Between Shannon's and Kolmogorov's Information Theories (where the parallelism fails and why)

Andrei Romashchenko, LIRMM (Montpellier)

2 February 2016, IHP

Outline

1. Parallelism in definitions
2. Perfect parallelism: information inequalities
3. The first threat to the parallelism: conditional inequalities
   - A conditional inequality: why it is so special
   - Parallelism re-established
   - Unexpected profit from the parallelism
4. A more serious threat: conditional inequalities once again
5. Parallels for conditional descriptions: one victory and one defeat

Outline

1. Parallelism in definitions
2. Perfect parallelism: information inequalities
3. The first threat to the parallelism: conditional inequalities
   - A conditional inequality: why it is so special
   - Parallelism re-established
   - Unexpected profit from the parallelism
4. A more serious threat: conditional inequalities once again
5. Parallels for conditional descriptions: one victory and one defeat

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic (Kolmogorov complexity)

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic (Kolmogorov complexity)
- combinatorial

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic (Kolmogorov complexity)
- combinatorial (Hartley's information ?)

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic (Kolmogorov complexity)
- combinatorial (Hartley's information ?)

Why should you care of parallelism?

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic (Kolmogorov complexity)
- combinatorial (Hartley's information ?)

Why should you care of parallelism?

- to anticipate new results

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic (Kolmogorov complexity)
- combinatorial (Hartley's information ?)

Why should you care of parallelism?

- to anticipate new results
- to find the *right* form of your theorems

The sources of inspiration

Andrei Kolmogorov:
Three approaches to the quantitative definition of information,
Prob. Inform. Trans., 1965.

Three approaches:

- probabilistic (Shannon's entropy)
- algorithmic (Kolmogorov complexity)
- combinatorial (Hartley's information ?)

Why should you care of parallelism?

- to anticipate new results
- to find the *right* form of your theorems
- to see the same phenomenon from different points of view
- to not miss the cases when the parallelism fails

The definitions (1)

How to measure the amount of uncertainty in an object?

The definitions (1)

How to measure the amount of uncertainty in an object?
And in what kind of object?

- Shannon: entropy, $H(\alpha) := \sum p_i \log \frac{1}{p_i}$, for a random variable α

The definitions (1)

How to measure the amount of uncertainty in an object?
And in what kind of object?

- Shannon: entropy, $H(\alpha) := \sum p_i \log \frac{1}{p_i}$, for a random variable α
- Kolmogorov: complexity, $C(a) := \min\{|p| \,:\, p \text{ prints } a\}$, for a binary string a

The definitions (1)

How to measure the amount of uncertainty in an object?
And in what kind of object?

- Shannon: entropy, $H(\alpha) := \sum p_i \log \frac{1}{p_i}$, for a random variable α
- Kolmogorov: complexity, $C(a) := \min\{|p| \ : \ p \text{ prints } a\}$, for a binary string a
- Hartley:

The definitions (1)

How to measure the amount of uncertainty in an object?
And in what kind of object?

- Shannon: entropy, $H(\alpha) := \sum p_i \log \frac{1}{p_i}$, for a random variable α
- Kolmogorov: complexity, $C(a) := \min\{|p| \ : \ p \text{ prints } a\}$, for a binary string a
- Hartley: "combinatorial entropy", $\chi(A) := \log |A|$ for a finite set A

The definitions (2)

How to measure the amount of conditional uncertanity in an object?

The definitions (2)

How to measure the amount of conditional uncertanity in an object?

- Shannon: $H(\alpha|\beta) := \sum \mathrm{prob}[\alpha = B_i] \cdot H(\alpha|\beta = B_i)$

The definitions (2)

How to measure the amount of conditional uncertanity in an object?

- Shannon: $H(\alpha|\beta) := \sum \mathrm{prob}[\alpha = B_i] \cdot H(\alpha|\beta = B_i) = H(\alpha, \beta) - H(\beta)$

The definitions (2)

How to measure the amount of conditional uncertanity in an object?

- Shannon: $H(\alpha|\beta) := \sum \mathrm{prob}[\alpha = B_i] \cdot H(\alpha|\beta = B_i) = H(\alpha, \beta) - H(\beta)$
- Kolmogorov: $C(a|b) := \min\{|p| \ : \ p \text{ prints } a \text{ given } b\}$

The definitions (2)

How to measure the amount of conditional uncertanity in an object?

- Shannon: $H(\alpha|\beta) := \sum \mathrm{prob}[\alpha = B_i] \cdot H(\alpha|\beta = B_i) = H(\alpha, \beta) - H(\beta)$
- Kolmogorov: $C(a|b) := \min\{|p| \ : \ p \text{ prints } a \text{ given } b\}$
$$= C(a, b) - C(b)$$

The definitions (2)

How to measure the amount of conditional uncertanity in an object?

- Shannon: $H(\alpha|\beta) := \sum \mathrm{prob}[\alpha = B_i] \cdot H(\alpha|\beta = B_i) = H(\alpha,\beta) - H(\beta)$
- Kolmogorov: $C(a|b) := \min\{|p| \ : \ p \text{ prints } a \text{ given } b\}$
  $$= C(a,b) - C(b) + O(\log C(a,b))$$

The definitions (2)

How to measure the amount of conditional uncertanity in an object?

- Shannon: $H(\alpha|\beta) := \sum \mathrm{prob}[\alpha = B_i] \cdot H(\alpha|\beta = B_i) = H(\alpha, \beta) - H(\beta)$
- Kolmogorov: $C(a|b) := \min\{|p| \ : \ p \text{ prints } a \text{ given } b\}$
  $$= C(a, b) - C(b) + O(\log C(a, b))$$
- Hartley: $\chi(A_{1|2}) := \max_{y \in \pi_2(A)} \log |\{x \ : \ (x, y) \in A\}|$

The definitions (3)

How to measure the amount of useful information in an object?

The definitions (3)

How to measure the amount of useful information in an object?

- Shannon: $I_S(\alpha : \beta) := H(\beta) - H(\beta|\alpha)$

The definitions (3)

How to measure the amount of useful information in an object?

- Shannon: $I_S(\alpha : \beta) := H(\beta) - H(\beta|\alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$

The definitions (3)

How to measure the amount of useful information in an object?

- Shannon: $I_S(\alpha : \beta) := H(\beta) - H(\beta|\alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$
- Kolmogorov: $I_K(a : b) := C(b) - C(b|a)$

The definitions (3)

How to measure the amount of useful information in an object?

- Shannon: $I_S(\alpha : \beta) := H(\beta) - H(\beta|\alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$
- Kolmogorov: $I_K(a : b) := C(b) - C(b|a)$
  $= C(a) + C(b) - C(a, b)$

The definitions (3)

How to measure the amount of useful information in an object?

- Shannon: $I_S(\alpha : \beta) := H(\beta) - H(\beta|\alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$
- Kolmogorov: $I_K(a : b) := C(b) - C(b|a)$
  $= C(a) + C(b) - C(a, b) + O(\log C(a, b))$

The definitions (3)

How to measure the amount of useful information in an object?

- Shannon: $I_S(\alpha : \beta) := H(\beta) - H(\beta|\alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$
- Kolmogorov: $I_K(a : b) := C(b) - C(b|a)$
  $$= C(a) + C(b) - C(a, b) + O(\log C(a, b))$$
- Hartley:

The definitions (3)

How to measure the amount of useful information in an object?

- Shannon: $I_S(\alpha : \beta) := H(\beta) - H(\beta|\alpha) = H(\alpha) + H(\beta) - H(\alpha, \beta)$
- Kolmogorov: $I_K(a : b) := C(b) - C(b|a)$
  $= C(a) + C(b) - C(a, b) + O(\log C(a, b))$
- Hartley: well, it's getting boring...

The definitions (conclusion)

Naive philosophy:

The definitions (conclusion)

Naive philosophy:

- Shannon: average complexity
- Kolmogorov: algorithmic complexity
- Hartley: worst case complexity

Outline

basic properties

- Shannon: $H(\alpha, \beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)

basic properties

- Shannon: $H(\alpha, \beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)

basic properties

- Shannon: $H(\alpha, \beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)
- Kolmogorov: $C(a, b) = C(a) + C(b|a)$

basic properties

- Shannon: $H(\alpha, \beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)
- Kolmogorov: $C(a, b) = C(a) + C(b|a) + O(\log C(a, b))$

basic properties

- Shannon: $H(\alpha, \beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)
- Kolmogorov: $C(a, b) = C(a) + C(b|a) + O(\log C(a, b))$(not *that* simple)

basic properties

- Shannon: $H(\alpha,\beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)
- Kolmogorov: $C(a,b) = C(a) + C(b|a) + O(\log C(a,b))$(not *that* simple)
  $\leq C(a) + C(b) + O(\log C(a,b))$ (trivial)

basic properties

- Shannon: $H(\alpha, \beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)
- Kolmogorov: $C(a, b) = C(a) + C(b|a) + O(\log C(a, b))$(not *that* simple)
  $\leq C(a) + C(b) + O(\log C(a, b))$ (trivial)
- Hartley: $\chi(A) \leq \chi_1(A) + \chi_{2|1}(A)$

basic properties

- Shannon: $H(\alpha, \beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)
- Kolmogorov: $C(a, b) = C(a) + C(b|a) + O(\log C(a, b))$(not *that* simple)
  $\leq C(a) + C(b) + O(\log C(a, b))$ (trivial)
- Hartley: $\chi(A) \leq \chi_1(A) + \chi_{2|1}(A)$
  $\leq \chi_1(A) + \chi_2(A)$

basic properties

- Shannon: $H(\alpha,\beta) = H(\alpha) + H(\beta|\alpha)$ (trivial)
  $\leq H(\alpha) + H(\beta)$ (Jensen's inequality)
- Kolmogorov: $C(a,b) = C(a) + C(b|a) + O(\log C(a,b))$(not *that* simple)
  $\leq C(a) + C(b) + O(\log C(a,b))$ (trivial)
- Hartley: $\chi(A) \leq \chi_1(A) + \chi_{2|1}(A)$
  $\leq \chi_1(A) + \chi_2(A)$ (just trivial)

less standard properties

- Shannon:
  $2H(\alpha,\beta,\gamma) \leq H(\alpha,\beta) + H(\beta,\gamma) + H(\alpha,\gamma)$

less standard properties

- Shannon:
  $2H(\alpha,\beta,\gamma) \le H(\alpha,\beta) + H(\beta,\gamma) + H(\alpha,\gamma)$
- Kolmogorov:
  $2C(a,b,c) \le C(a,b) + C(b,c) + C(a,c) + O(\log C(a,b,c))$

less standard properties

- Shannon:
  $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- Kolmogorov:
  $2C(a, b, c) \leq C(a, b) + C(b, c) + C(a, c) + O(\log C(a, b, c))$
- Hartley:
  $2\chi(A) \leq \chi(A_{12}) + \chi(A_{23}) + \chi(A_{13})$
  equivalent version:
  $|A|^2 \leq |\pi_{12}(A)| \cdot |\pi_{23}(A)| \cdot |\pi_{13}(A)|$

less standard properties

- Shannon:
  $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- Kolmogorov:
  $2C(a, b, c) \leq C(a, b) + C(b, c) + C(a, c) + O(\log C(a, b, c))$
- Hartley:
  $2\chi(A) \leq \chi(A_{12}) + \chi(A_{23}) + \chi(A_{13})$
  equivalent version:
  $|A|^2 \leq |\pi_{12}(A)| \cdot |\pi_{23}(A)| \cdot |\pi_{13}(A)|$ (not trivial!)

less standard properties

- Shannon:
  $2H(\alpha,\beta,\gamma) \leq H(\alpha,\beta) + H(\beta,\gamma) + H(\alpha,\gamma)$
- Kolmogorov:
  $2C(a,b,c) \leq C(a,b) + C(b,c) + C(a,c) + O(\log C(a,b,c))$
- Hartley:
  $2\chi(A) \leq \chi(A_{12}) + \chi(A_{23}) + \chi(A_{13})$
  equivalent version:
  $|A|^2 \leq |\pi_{12}(A)| \cdot |\pi_{23}(A)| \cdot |\pi_{13}(A)|$ (not trivial!)

Hammer and Shen:
Kolmogorov's or Shannon's inequality implies Hartley's inequality
(*A strange application of Kolmogorov complexity, 1998*)

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \leq H(\alpha, \beta)$

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \leq H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$

inequalities: parallelism is getting evident

Shannon:

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

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \leq H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$
- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$
- $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \le H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$
- $H(\alpha, \beta) \le H(\alpha) + H(\beta)$
- $2H(\alpha, \beta, \gamma) \le H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- $H(\alpha, \beta, \gamma) + H(\alpha) \le H(\alpha, \beta) + H(\alpha, \gamma)$

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \leq H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$
- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$
- $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- $H(\alpha, \beta, \gamma) + H(\alpha) \leq H(\alpha, \beta) + H(\alpha, \gamma)$

Kolmogorov:

- $C(a) \leq C(a, b) + O(1)$

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \leq H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$
- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$
- $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- $H(\alpha, \beta, \gamma) + H(\alpha) \leq H(\alpha, \beta) + H(\alpha, \gamma)$

Kolmogorov:

- $C(a) \leq C(a, b) + O(1)$
- $I_K(a : b) = I_K(b : a) + O(\log C(a, b))$

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \leq H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$
- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$
- $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- $H(\alpha, \beta, \gamma) + H(\alpha) \leq H(\alpha, \beta) + H(\alpha, \gamma)$

Kolmogorov:

- $C(a) \leq C(a, b) + O(1)$
- $I_K(a : b) = I_K(b : a) + O(\log C(a, b))$
- $C(a, b) \leq C(a) + H(b) + O(\log \ldots)$

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \le H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$
- $H(\alpha, \beta) \le H(\alpha) + H(\beta)$
- $2H(\alpha, \beta, \gamma) \le H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- $H(\alpha, \beta, \gamma) + H(\alpha) \le H(\alpha, \beta) + H(\alpha, \gamma)$

Kolmogorov:

- $C(a) \le C(a, b) + O(1)$
- $I_K(a : b) = I_K(b : a) + O(\log C(a, b))$
- $C(a, b) \le C(a) + H(b) + O(\log \ldots)$
- $2C(a, b, c) \le C(a, b) + C(b, c) + C(a, c) + O(\log \ldots)$

inequalities: parallelism is getting evident

Shannon:

- $H(\alpha) \leq H(\alpha, \beta)$
- $I_S(\alpha : \beta) = I_S(\beta : \alpha)$
- $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$
- $2H(\alpha, \beta, \gamma) \leq H(\alpha, \beta) + H(\beta, \gamma) + H(\alpha, \gamma)$
- $H(\alpha, \beta, \gamma) + H(\alpha) \leq H(\alpha, \beta) + H(\alpha, \gamma)$

Kolmogorov:

- $C(a) \leq C(a, b) + O(1)$
- $I_K(a : b) = I_K(b : a) + O(\log C(a, b))$
- $C(a, b) \leq C(a) + H(b) + O(\log \ldots)$
- $2C(a, b, c) \leq C(a, b) + C(b, c) + C(a, c) + O(\log \ldots)$
- $C(a, b, c) + C(a) \leq C(a, b) + C(a, c) + O(\log \ldots)$

general inequalities

Equivalence Theorem [Hammer, R., Shen, Vereshchagin]
Exactly the same linear inequality are valid for Shannon's entropy and for Kolmogorov complexity.

for all $(\alpha_1, \ldots, \alpha_n)$

$$\begin{aligned}\lambda_1 H(\alpha_1) + \lambda_2 H(\alpha_2) + \lambda_3 H(\alpha_3) + \ldots \\ +\lambda_{12} H(\alpha_1, \alpha_2) + \lambda_{13} H(\alpha_1, \alpha_3) + \ldots \\ + \lambda_{123} H(\alpha_1, \alpha_2, \alpha_3) + \ldots \ \ldots \ \ldots \geq 0\end{aligned}$$

if and only if

there exists a $C > 0$ *such that for all* $(a_1, \ldots, a_n)$

$$\begin{aligned}\lambda_1 C(a_1) + \lambda_2 C(a_2) + \lambda_3 C(a_3) + \ldots \\ +\lambda_{12} C(a_1, a_2) + \lambda_{13} C(a_1, a_3) + \ldots \\ + \lambda_{123} C(a_1, a_2, a_3) + \ldots \ \ldots \ \ldots \\ + C \log(|a_1| + \ldots + |a_n|) \geq 0\end{aligned}$$

What about combinatorial inequalities?

What about combinatorial inequalities?

Information inequalities have some combinatorial interpretation. . .

What about combinatorial inequalities?

Information inequalities have some combinatorial interpretation...

...but we do not discuss them in this talk.

What about combinatorial inequalities?

Information inequalities have some combinatorial interpretation...

...but we do not discuss them in this talk.

See

- Chan-Yeung

What about combinatorial inequalities?

Information inequalities have some combinatorial interpretation...

...but we do not discuss them in this talk.

See

- Chan-Yeung
- R.-Shen-Vereshchegin

What about combinatorial inequalities?

Information inequalities have some combinatorial interpretation. . .

. . . but we do not discuss them in this talk.

See

- Chan-Yeung
- R.-Shen-Vereshchegin
- Alon-Newman-Shen-Tardos-Vereshchagin

the intuition behind the information inequalities

We know from Shannon:

the intuition behind the information inequalities

We know from Shannon:

- monotonicity: $H(\alpha) \le H(\alpha, \beta)$

the intuition behind the information inequalities

We know from Shannon:

- monotonicity: $H(\alpha) \leq H(\alpha, \beta)$ (a.k.a. $H(\beta|\alpha) \geq 0$)

the intuition behind the information inequalities

We know from Shannon:

- monotonicity: $H(\alpha) \leq H(\alpha, \beta)$ (a.k.a. $H(\beta|\alpha) \geq 0$)
- subadditivity: $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$

the intuition behind the information inequalities

We know from Shannon:

- monotonicity: $H(\alpha) \le H(\alpha, \beta)$ (a.k.a. $H(\beta|\alpha) \ge 0$)
- subadditivity: $H(\alpha, \beta) \le H(\alpha) + H(\beta)$ (a.k.a. $I_S(\alpha : \beta) \ge 0$)

the intuition behind the information inequalities

We know from Shannon:

- monotonicity: $H(\alpha) \leq H(\alpha, \beta)$ (a.k.a. $H(\beta|\alpha) \geq 0$)
- subadditivity: $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$ (a.k.a. $I_S(\alpha : \beta) \geq 0$)
- submodularity: $H(\alpha, \beta, \gamma) + H(\alpha) \leq H(\alpha, \beta) + H(\alpha, \gamma)$

the intuition behind the information inequalities

We know from Shannon:

- monotonicity: $H(\alpha) \leq H(\alpha, \beta)$ (a.k.a. $H(\beta|\alpha) \geq 0$)
- subadditivity: $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$ (a.k.a. $I_S(\alpha : \beta) \geq 0$)
- submodularity: $H(\alpha, \beta, \gamma) + H(\alpha) \leq H(\alpha, \beta) + H(\alpha, \gamma)$ (a.k.a. $I_S(\beta : \gamma|\alpha) \geq 0$)

the intuition behind the information inequalities

We know from Shannon:

- monotonicity: $H(\alpha) \le H(\alpha, \beta)$ (a.k.a. $H(\beta|\alpha) \ge 0$)
- subadditivity: $H(\alpha, \beta) \le H(\alpha) + H(\beta)$ (a.k.a. $I_S(\alpha : \beta) \ge 0$)
- submodularity: $H(\alpha, \beta, \gamma) + H(\alpha) \le H(\alpha, \beta) + H(\alpha, \gamma)$ (a.k.a. $I_S(\beta : \gamma|\alpha) \ge 0$)

plus all substitutions

the intuition behind the information inequalities

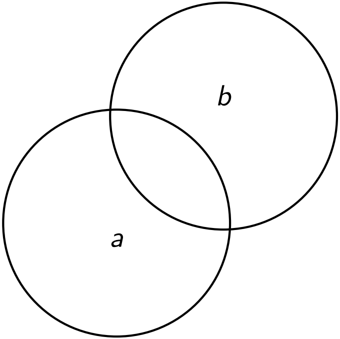
We know from Shannon:

- monotonicity: $H(\alpha) \leq H(\alpha, \beta)$ (a.k.a. $H(\beta|\alpha) \geq 0$)
- subadditivity: $H(\alpha, \beta) \leq H(\alpha) + H(\beta)$ (a.k.a. $I_S(\alpha : \beta) \geq 0$)
- submodularity: $H(\alpha, \beta, \gamma) + H(\alpha) \leq H(\alpha, \beta) + H(\alpha, \gamma)$ (a.k.a. $I_S(\beta : \gamma|\alpha) \geq 0$)

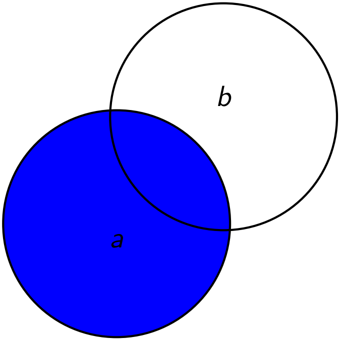
plus all substitutions

plus all (positive) linear combinations

information diagrams

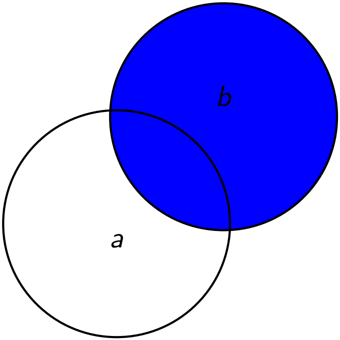


information diagrams



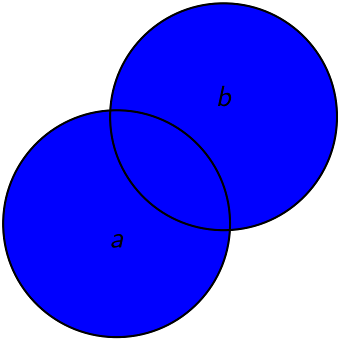
$H(a)$

information diagrams



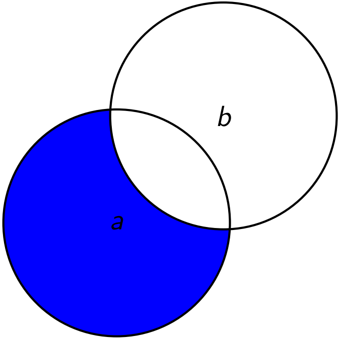
$H(b)$

information diagrams



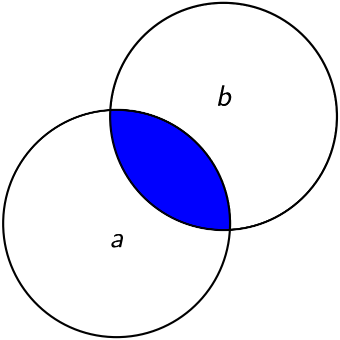
$H(a, b)$

information diagrams



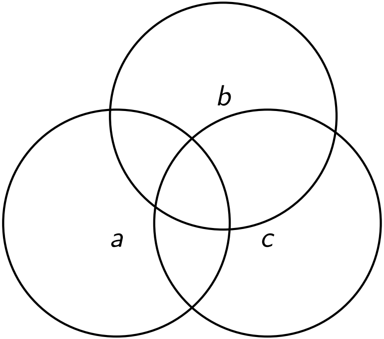
$H(a|b) = H(a, b) - H(b)$

information diagrams

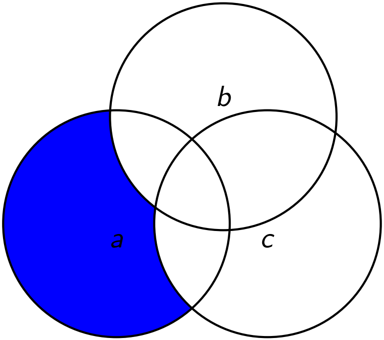


$I_K(a : b) = H(a) + H(b) - H(a, b)$

information diagrams

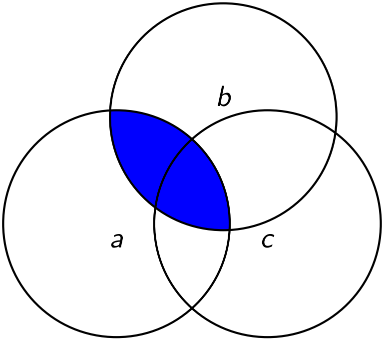


information diagrams



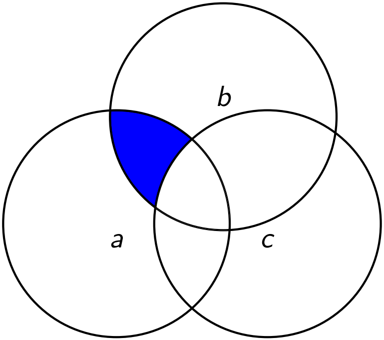
$H(a|b,c) = H(a,b,c) - H(b,c)$

information diagrams



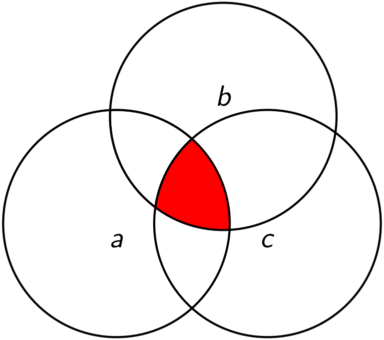
$I_S(a:b) = H(a) + H(b) - H(a,b)$

information diagrams



$I_S(a : b|c) = H(a,c) + H(b,c) - H(a,b,c) - H(c)$

information diagrams



$$I_S(a:b:c) = H(a)+H(b)+H(c)-H(a,b)-H(a,c)-H(b,c)+H(a,b,c)$$

the intuition behind the information inequalities

What are all these linear "information inequalities" ?

We know from Shannon:

- monotonicity: $H(\alpha) \le H(\alpha, \beta)$
- subadditivity: $H(\alpha, \beta) \le H(\alpha) + H(\beta)$
  (a.k.a. $I_S(\alpha : \beta) \ge 0$)
- submodularity: $H(\alpha, \beta, \gamma) + H(\alpha) \le H(\alpha, \beta) + H(\alpha, \gamma)$
  (a.k.a. $I_S(\beta : \gamma|\alpha) \ge 0$)

plus all substitutions

plus all (positive) linear combinations

Anything else?

non Shannon type inequalities

Yes!

non Shannon type inequalities

Yes!

Z. Zhang and R. W. Yeung, [1998]:
the first non Shannon type information inequality

non Shannon type inequalities

Yes!

Z. Zhang and R. W. Yeung, [1998]:
the first non Shannon type information inequality
$I_S(a:b) \leq 2I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + I_S(a:x|b) + I_S(b:x|a)$

non Shannon type inequalities

Yes!

Z. Zhang and R. W. Yeung, [1998]:
the first non Shannon type information inequality
$I_S(a:b) \leq 2I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + I_S(a:x|b) + I_S(b:x|a)$

Since 1998
we got *other* examples of non Shannon type information inequalities

non Shannon type inequalities

Yes!

Z. Zhang and R. W. Yeung, [1998]:
the first non Shannon type information inequality
$I_S(a:b) \leq 2I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + I_S(a:x|b) + I_S(b:x|a)$

Since 1998
we got *other* examples of non Shannon type information inequalities
we got *many* other examples

non Shannon type inequalities

Yes!

Z. Zhang and R. W. Yeung, [1998]:
the first non Shannon type information inequality
$I_S(a:b) \leq 2I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + I_S(a:x|b) + I_S(b:x|a)$

Since 1998
we got *other* examples of non Shannon type information inequalities
we got *many* other examples
we got *infinitely many* other examples...

non Shannon type inequalities

Yes!

Z. Zhang and R. W. Yeung, [1998]:
the first non Shannon type information inequality
$I_S(a:b) \le 2I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + I_S(a:x|b) + I_S(b:x|a)$

Since 1998
we got *other* examples of non Shannon type information inequalities
we got *many* other examples
we got *infinitely many* other examples...

F. Matúš [1999]:
For $n > 3$ random variables the cone of all linear information inequalities is not polyhedral !

non Shannon type inequalities

Yes!

Z. Zhang and R. W. Yeung, [1998]:
the first non Shannon type information inequality
$I_S(a:b) \leq 2I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + I_S(a:x|b) + I_S(b:x|a)$

Since 1998
we got *other* examples of non Shannon type information inequalities
we got *many* other examples
we got *infinitely many* other examples...

F. Matúš [1999]:
For $n > 3$ random variables the cone of all linear information inequalities is not polyhedral !

Applications: lower bounds for secret sharing schemes, for admissible rates in communication networks, etc.

Outline

Outline

conditional inequalities: parallelism under a threat

Another kind of information inequalities for Shannon's entropy:

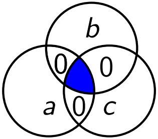
if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

conditional inequalities: parallelism under a threat

Another kind of information inequalities for Shannon's entropy:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

Sketch of the proof:
If $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $\exists w$ such that $H(w|a) = H(w|b) = 0$ and $H(w) = I_S(a:b)$

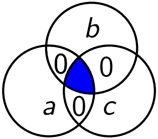


conditional inequalities: parallelism under a threat

Another kind of information inequalities for Shannon's entropy:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

Sketch of the proof:
If $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $\exists w$ such that $H(w|a) = H(w|b) = 0$ and $H(w) = I_S(a:b)$



Then we apply a Shannon type inequality

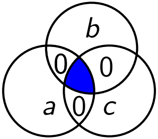
$$H(w) \quad \le \quad H(w|x) \quad + \quad H(w|y) \quad + \quad I_S(x:y)$$

conditional inequalities: parallelism under a threat

Another kind of information inequalities for Shannon's entropy:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

Sketch of the proof:
If $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $\exists w$ such that $H(w|a) = H(w|b) = 0$ and $H(w) = I_S(a:b)$



Then we apply a Shannon type inequality

$$\begin{array}{ccccccc} H(w) & \leq & H(w|x) & + & H(w|y) & + & I_S(x:y) \\ || & & || & & || & & \\ I_S(a:b) & & I_S(a:b|x) & & I_S(a:b|y) & & \end{array}$$

conditional inequalities: parallelism under a threat

How to re-formulate it for Kolmogorov complexity?

conditional inequalities: parallelism under a threat

How to re-formulate it for Kolmogorov complexity?

if

$$\begin{cases} I_K(a:b|c) & = & O(\log\ldots), \\ I_K(a:c|b) & = & O(\log\ldots), \\ I_K(b:c|a) & = & O(\log\ldots), \end{cases}$$

then $I_K(a:b) \le I_K(a:b|x) + I_K(a:b|y) + I_K(x:y) + O(\log\ldots)$

How to prove it?

conditional inequalities: soft version

the conjectured inequality for Kolmogorov complexity:

if

$$\left\{ \begin{array}{rcl} I_K(a:b|c) &=& O(?), \\ I_K(a:c|b) &=& O(?), \\ I_K(b:c|a) &=& O(?), \end{array} \right.$$

then $I_K(a:b) \le I_K(a:b|x) + I_K(a:b|y) + I_K(x:y) + O(?)$

conditional inequalities: soft version

the conjectured inequality for Kolmogorov complexity:

if

$$\begin{cases} I_K(a:b|c) &= O(?), \\ I_K(a:c|b) &= O(?), \\ I_K(b:c|a) &= O(?), \end{cases}$$

then $I_K(a:b) \le I_K(a:b|x) + I_K(a:b|y) + I_K(x:y) + O(?)$

Shannon's version with "soft" constraints:

if

$$\begin{cases} I_S(a:b|c) &= \text{[sth small]}, \\ I_S(a:c|b) &= \text{[sth small]}, \\ I_S(b:c|a) &= \text{[sth small]}, \end{cases}$$

then $I_S(a:b) \le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \text{[sth small]}$

Outline

conditional inequalities: coping with the threat

Revisit the Shannon's version of the inequality:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

conditional inequalities: coping with the threat

Revisit the Shannon's version of the inequality:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

Alternative proof:

Step 1: prove an unconditional non Shannon type inequality:

$$I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \\ + I_S(a:b|c) + I_S(a:c|b) + I_S(b:c|a)$$

conditional inequalities: coping with the threat

Revisit the Shannon's version of the inequality:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

Alternative proof:

Step 1: prove an unconditional non Shannon type inequality:

$$I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \\ + I_S(a:b|c) + I_S(a:c|b) + I_S(b:c|a)$$

Step 2: deduce the conditional inequality

conditional inequalities: coping with the threat

Revisit the Shannon's version of the inequality:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

Alternative proof:

Step 1: prove an unconditional non Shannon type inequality:

$$I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \\ + I_S(a:b|c) + I_S(a:c|b) + I_S(b:c|a)$$

Step 2: deduce the conditional inequality (even with "soft" constraints!)

conditional inequalities: coping with the threat

Revisit the Shannon's version of the inequality:

if $I_S(a:b|c) = I_S(a:c|b) = I_S(b:c|a) = 0$,
then $I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

Alternative proof:

Step 1: prove an unconditional non Shannon type inequality:

$$I_S(a:b) \leq I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \\ + I_S(a:b|c) + I_S(a:c|b) + I_S(b:c|a)$$

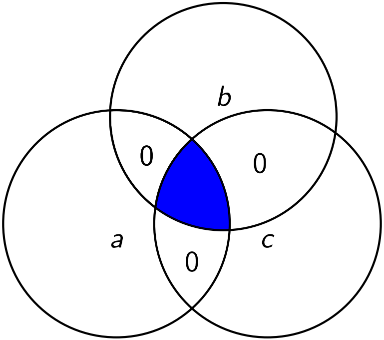
Step 2: deduce the conditional inequality (even with "soft" constraints!)

Towards Kolmogorov's version: the unconditional inequality rewrites to
Step 1': $I_K(a:b) \leq I_K(a:b|x) + I_K(a:b|y) + I_K(x:y) +$
$+ I_K(a:b|c) + I_K(a:c|b) + I_K(b:c|a) + O(\log\ldots)$

Step 2': deduce the conditional version for Kolmogorov complexity

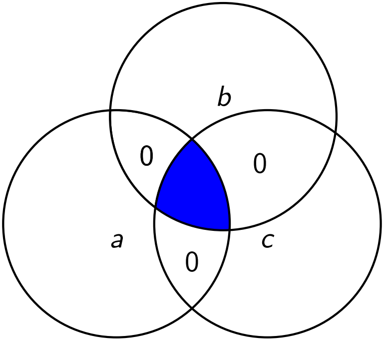
Outline

Ahlswede-Körner and extracting the mutual information



We can *materialize* the mutual information $I_S(a : b : c)$:

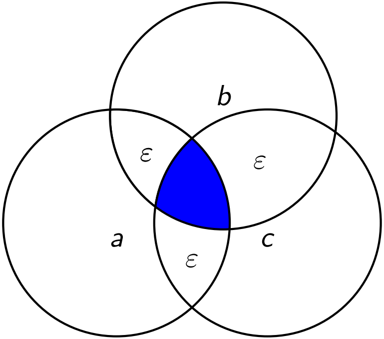
Ahlswede-Körner and extracting the mutual information



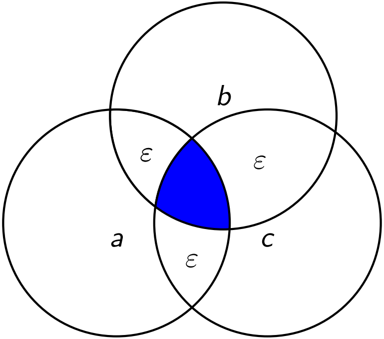
We can *materialize* the mutual information $I_S(a : b : c)$: $\exists w$ s.t.

- $H(w) = I_S(a : b : c)$,
- $H(w|a) = 0$,
- $H(w|b) = 0$,
- $H(w|c) = 0$.

Ahlswede-Körner and extracting the mutual information

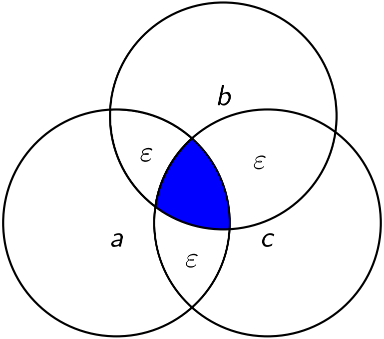


Ahlswede-Körner and extracting the mutual information



Can we now *materialize* the mutual information $I_S(a : b : c)$?

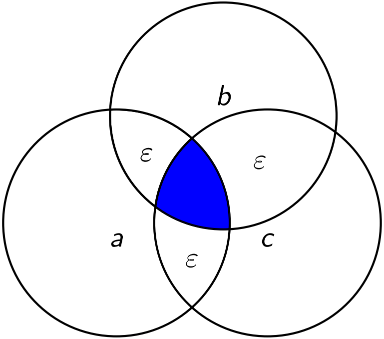
Ahlswede-Körner and extracting the mutual information



Can we now *materialize* the mutual information $I_S(a : b : c)$?

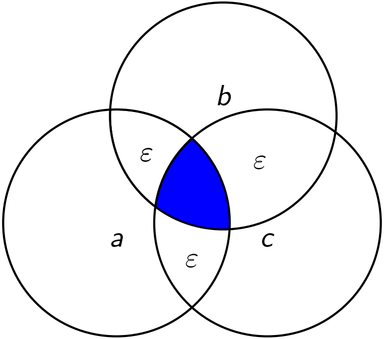
Yes, in some sense...

Ahlswede-Körner and extracting the mutual information



Can we now *materialize* the mutual information $I_S(a:b:c)$?
Yes, in some sense...
Let $(a_1, b_1, c_1), \ldots, (a_n, b_n, c_n)$ be i.i.d., distributed as (a, b, c).

Ahlswede-Körner and extracting the mutual information



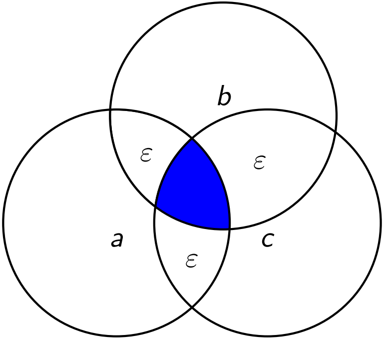
Can we now *materialize* the mutual information $I_S(a : b : c)$?

Yes, in some sense...

Let $(a_1, b_1, c_1), \ldots, (a_n, b_n, c_n)$ be i.i.d., distributed as (a, b, c). $\exists W$ s.t.

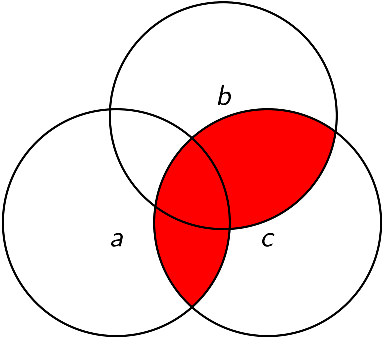
- $H(W) = n \cdot I_S(a : b : c) + O(\varepsilon) + o(n)$,
- $H(W|a_1 \ldots a_n) = O(\varepsilon) + o(n)$,
- $H(W|b_1 \ldots b_n) = O(\varepsilon) + o(n)$,
- $H(W|c_1 \ldots c_n) = O(\varepsilon) + o(n)$.

Ahlswede-Körner and extracting the mutual information



How to *materialize* the mutual information $I_S(a : b : c)$?

Ahlswede-Körner and extracting the mutual information



Let $(a_1, b_1, c_1), \dots, (a_n, b_n, c_n)$ be i.i.d., distributed as (a, b, c).

Ahlswede and Körner [1975]: there exists a W such that

- $H(W) = n \cdot I_S(a, b : c) + o(n)$,
- $H(a_1 \dots a_n | W) = n \cdot H(a|c) + o(n)$,
- $H(b_1 \dots b_n | W) = n \cdot H(b|c) + o(n)$,
- $H(a_1 \dots a_n, b_1 \dots b_n | W) = n \cdot H(a, b|c) + o(n)$.

from Ahlswede-Körner to Zhang-Yeung

Step 0. take n i.i.d. copies of (a, b, c, x, y)

from Ahlswede-Körner to Zhang-Yeung

Step 0. take n i.i.d. copies of (a, b, c, x, y)
$\bar{a} := a_1 \ldots a_n$, $\bar{b} := b_1 \ldots b_n$, $\bar{c} := c_1 \ldots c_n$, $\bar{x} := x_1 \ldots x_n$, $\bar{y} := y_1 \ldots y_n$

from Ahlswede-Körner to Zhang-Yeung

Step 0. take n i.i.d. copies of (a, b, c, x, y)
$\bar{a} := a_1 \dots a_n$, $\bar{b} := b_1 \dots b_n$, $\bar{c} := c_1 \dots c_n$, $\bar{x} := x_1 \dots x_n$, $\bar{y} := y_1 \dots y_n$

Step 1. take W from Ahslwede-Körner (for the triple a, b, c)

from Ahlswede-Körner to Zhang-Yeung

Step 0. take n i.i.d. copies of (a, b, c, x, y)
$\bar{a} := a_1 \dots a_n$, $\bar{b} := b_1 \dots b_n$, $\bar{c} := c_1 \dots c_n$, $\bar{x} := x_1 \dots x_n$, $\bar{y} := y_1 \dots y_n$

Step 1. take W from Ahslwede-Körner (for the triple a, b, c)

Step 2. Apply a Shannon type inequality

$$H(W) \quad \leq \quad H(W|\bar{x}) \quad + \quad H(W|\bar{y}) \quad + \quad I_S(\bar{x} : \bar{y})$$

from Ahlswede-Körner to Zhang-Yeung

Step 0. take n i.i.d. copies of (a, b, c, x, y)
$\bar{a} := a_1 \dots a_n$, $\bar{b} := b_1 \dots b_n$, $\bar{c} := c_1 \dots c_n$, $\bar{x} := x_1 \dots x_n$, $\bar{y} := y_1 \dots y_n$

Step 1. take W from Ahslwede-Körner (for the triple a, b, c)

Step 2. Apply a Shannon type inequality

$$\begin{array}{ccccccc} H(W) & \leq & H(W|\bar{x}) & + & H(W|\bar{y}) & + & I_S(\bar{x} : \bar{y}) \\ \wr\wr & & \wr\wr & & \wr\wr & & \\ I_S(\bar{a} : \bar{b}) & & I_S(\bar{a} : \bar{b}|\bar{x}) & & I_S(\bar{a} : \bar{b}|\bar{y}) & & \end{array}$$

from Ahlswede-Körner to Zhang-Yeung

Step 0. take n i.i.d. copies of (a, b, c, x, y)
$\bar{a} := a_1 \dots a_n$, $\bar{b} := b_1 \dots b_n$, $\bar{c} := c_1 \dots c_n$, $\bar{x} := x_1 \dots x_n$, $\bar{y} := y_1 \dots y_n$

Step 1. take W from Ahslwede-Körner (for the triple a, b, c)

Step 2. Apply a Shannon type inequality

$$\begin{array}{ccccccc} H(W) & \leq & H(W|\bar{x}) & + & H(W|\bar{y}) & + & I_S(\bar{x} : \bar{y}) \\ \wr & & \wr & & \wr & & \\ I_S(\bar{a} : \bar{b}) & & I_S(\bar{a} : \bar{b}|\bar{x}) & & I_S(\bar{a} : \bar{b}|\bar{y}) & & \end{array}$$

Conclusion: if $I_S(a : b|c) \approx I_S(a : c|b) \approx I_S(b : c|a) \approx 0$, then
$I_S(a : b) \leq I_S(a : b|x) + I_S(a : b|x) + I_S(x : y) +$ [small residue term]

from Ahlswede-Körner to Zhang-Yeung

Step 0. take n i.i.d. copies of (a, b, c, x, y)
$\bar{a} := a_1 \dots a_n$, $\bar{b} := b_1 \dots b_n$, $\bar{c} := c_1 \dots c_n$, $\bar{x} := x_1 \dots x_n$, $\bar{y} := y_1 \dots y_n$

Step 1. take W from Ahslwede-Körner (for the triple a, b, c)

Step 2. Apply a Shannon type inequality

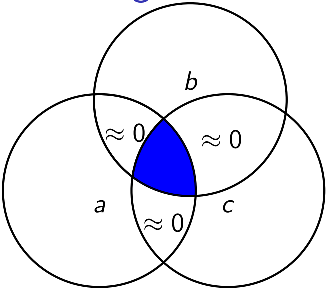
$$\begin{array}{ccccccc} H(W) & \leq & H(W|\bar{x}) & + & H(W|\bar{y}) & + & I_S(\bar{x} : \bar{y}) \\ \wr\wr & & \wr\wr & & \wr\wr & & \\ I_S(\bar{a} : \bar{b}) & & I_S(\bar{a} : \bar{b}|\bar{x}) & & I_S(\bar{a} : \bar{b}|\bar{y}) & & \end{array}$$

Conclusion: if $I_S(a : b|c) \approx I_S(a : c|b) \approx I_S(b : c|a) \approx 0$, then
$I_S(a : b) \leq I_S(a : b|x) + I_S(a : b|x) + I_S(x : y) +$ [small residue term]

More precisely:

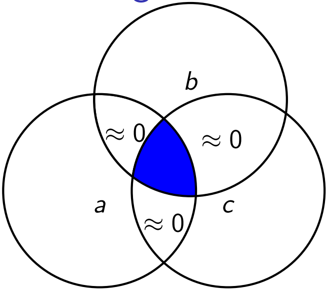
$$I_S(a : b) \leq I_S(a : b|x) + I_S(a : b|y) + I_S(x : y) + \\ + I_S(a : b|c) + I_S(a : c|b) + I_S(b : c|a)$$

extracting mutual information in Kolmogorov's framework



$$\left\{ \begin{array}{lcl} I_K(a:c|b) & \approx & 0, \\ I_K(a:b|c) & \approx & 0, \\ I_K(b:c|a) & \approx & 0. \end{array} \right. \quad \Longrightarrow \quad \text{the mutual information can be "materialized"}$$

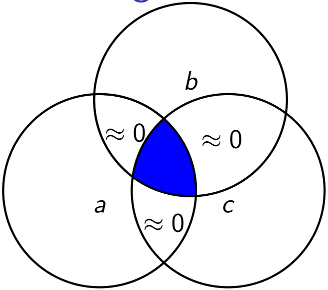
extracting mutual information in Kolmogorov's framework



$$\begin{cases} I_K(a:c|b) & \approx & 0, \\ I_K(a:b|c) & \approx & 0, \\ I_K(b:c|a) & \approx & 0. \end{cases} \implies \text{the mutual information can be "materialized"}$$

No Ahslwede-Körner Lemma for Kolmogorov compl.

extracting mutual information in Kolmogorov's framework



$$\left\{ \begin{array}{lcl} I_K(a:c|b) & \approx & 0, \\ I_K(a:b|c) & \approx & 0, \\ I_K(b:c|a) & \approx & 0. \end{array} \right. \implies \text{the mutual information can be "materialized"}$$

No Ahslwede-Körner Lemma for Kolmogorov compl. However, we can prove $\exists w$:

- $C(w) \approx I_K(a:b:c)$,
- $C(w|a) \approx 0$,
- $C(w|b) \approx 0$,
- $C(w|c) \approx 0$.

Outline

conditional inequalities: another threat

Another conditional information inequality for Shannon's entropy, F. Matúš:

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

conditional inequalities: another threat

Another conditional information inequality for Shannon's entropy, F. Matúš:

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

Sketch of the proof:
Step 1. if $I_S(a:x|b) = I_S(b:x|a) = 0$,
then $\exists w$ such that $H(w|x) = H(w|y) = 0$ and $I_S(a:w) = I_S(a:xy)$.

conditional inequalities: another threat

Another conditional information inequality for Shannon's entropy, F. Matúš:

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

Sketch of the proof:
Step 1. if $I_S(a : x|b) = I_S(b : x|a) = 0$,
then $\exists w$ such that $H(w|x) = H(w|y) = 0$ and $I_S(a : w) = I_S(a : xy)$.

Step 2. apply a suitable Shannon type inequality, etc.

conditional inequalities: another threat

Another conditional information inequality for Shannon's entropy, F. Matúš:

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

Sketch of the proof:
Step 1. if $I_S(a : x|b) = I_S(b : x|a) = 0$,
then $\exists w$ such that $H(w|x) = H(w|y) = 0$ and $I_S(a : w) = I_S(a : xy)$.

Step 2. apply a suitable Shannon type inequality, etc.

A version with "soft constraints"?

conditional inequalities: another threat

Another conditional information inequality for Shannon's entropy, F. Matúš:

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

Sketch of the proof:
Step 1. if $I_S(a:x|b) = I_S(b:x|a) = 0$,
then $\exists w$ such that $H(w|x) = H(w|y) = 0$ and $I_S(a:w) = I_S(a:xy)$.

Step 2. apply a suitable Shannon type inequality, etc.

A version with "soft constraints"?

A version for Kolmogorov complexity?

conditional inequalities: parallelism undermined

hard-constraints Shannon's entropy:

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

conditional inequalities: parallelism undermined

hard-constraints Shannon's entropy:

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

No soft-constraints linear version for Shannon's entropy:
whatever are coefficients λ_1, λ_2, for some distribution (a, b, x, y)
$I_S(a:b) \not\le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \lambda_1 I_S(a:x|b) + \lambda_2 I_S(b:x|a)$

conditional inequalities: parallelism undermined

hard-constraints Shannon's entropy:

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

No soft-constraints linear version for Shannon's entropy:
whatever are coefficients λ_1, λ_2, for some distribution (a, b, x, y)
$I_S(a:b) \not\le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \lambda_1 I_S(a:x|b) + \lambda_2 I_S(b:x|a)$

However, a *kind of* Kolmogorov's version:
if $I_K(a:x|b) \le \sqrt{n}$ and $I_K(b:x|a) \le \sqrt{n}$, then
$I_K(a:b) \le I_K(a:b|x) + I_K(a:b|y) + I_K(x:y) + O(n^{3/4})$

conditional inequalities: parallelism fails

hard-constraints Shannon's inequality [Kaced, R.]:

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

conditional inequalities: parallelism fails

hard-constraints Shannon's inequality [Kaced, R.]:

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

No soft-constraints version Shannon's:
whatever are coefficients λ_1, λ_2, for some distributions
$I_S(a : b) \not\leq I_S(a : b|x) + I_S(a : b|y) + I_S(x : y) + \lambda_1 I_S(x : y|a) + \lambda_2 H(a|x, y)$

conditional inequalities: parallelism fails

hard-constraints Shannon's inequality [Kaced, R.]:

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

No soft-constraints version Shannon's:
whatever are coefficients λ_1, λ_2, for some distributions
$I_S(a : b) \not\leq I_S(a : b|x) + I_S(a : b|y) + I_S(x : y) + \lambda_1 I_S(x : y|a) + \lambda_2 H(a|x, y)$

No Kolmogorov's version: for some strings a, b, x, y
$I_K(x : y|a) = O(\log n)$ and $C(a|x, y) = O(\log n)$
but $I_K(a : b) \gg I_K(a : b|x) + I_K(a : b|y) + I_K(x : y)$

conditional inequalities: parallelism fails

hard-constraints Shannon's inequality [Kaced, R.]:

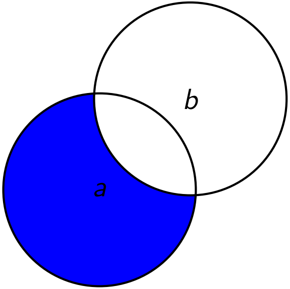
if $I_S(x:y|a) = H(a|x,y) = 0$,
then $I_S(a:b) \le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y)$

No soft-constraints version Shannon's:
whatever are coefficients λ_1, λ_2, for some distributions
$I_S(a:b) \not\le I_S(a:b|x) + I_S(a:b|y) + I_S(x:y) + \lambda_1 I_S(x:y|a) + \lambda_2 H(a|x,y)$

No Kolmogorov's version: for some strings a, b, x, y
$I_K(x:y|a) = O(\log n)$ and $C(a|x,y) = O(\log n)$
but $I_K(a:b) \gg I_K(a:b|x) + I_K(a:b|y) + I_K(x:y)$ (the gap $= \Omega(n)$)

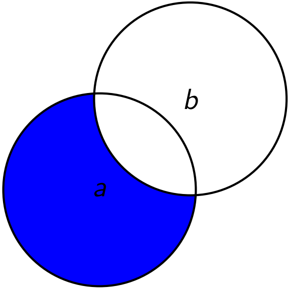
Outline

conditional encoding: slepian-wolf vs muchnik



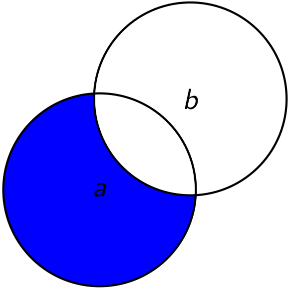
$H(a|b) = H(a,b) - H(b)$.

conditional encoding: slepian-wolf vs muchnik



$H(a|b) = H(a,b) - H(b)$. Can we *materialize* the part $a \setminus b$?

conditional encoding: slepian-wolf vs muchnik



$H(a|b) = H(a,b) - H(b)$. Can we *materialize* the part $a \setminus b$?

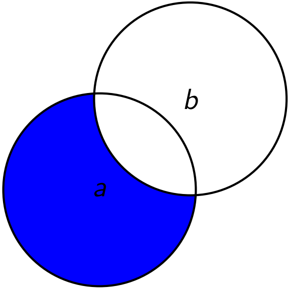
A formal question: can we find a w such that

- $H(w) = H(a|b)$,
- $H(w|a) = 0$,
- $H(a|b,w) = 0$

?

In general, **no**!

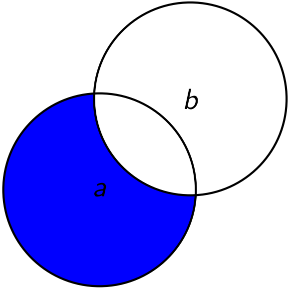
conditional encoding: slepian-wolf vs muchnik



$H(a|b) = H(a,b) - H(b)$. Can we *materialize* the part $a \setminus b$?

Let $(a_1, b_1), \ldots, (a_n, b_n)$ be i.i.d., all with the same distribution as (a, b)

conditional encoding: slepian-wolf vs muchnik



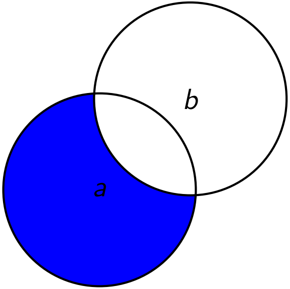
$H(a|b) = H(a, b) - H(b)$. Can we *materialize* the part $a \setminus b$?

Let $(a_1, b_1), \ldots, (a_n, b_n)$ be i.i.d., all with the same distribution as (a, b)

Selpian and Wolf [1973]: Then there exists a W such that

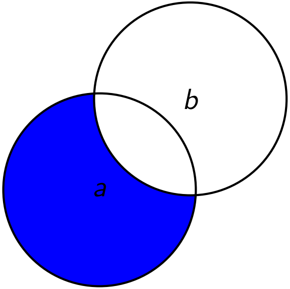
- $H(W) = n \cdot H(a|b) + o(n)$,
- $H(W|a_1 \ldots a_n) = 0$,
- $H(a_1 \ldots a_n | b_1 \ldots b_n, W) = 0$

conditional encoding: slepian-wolf vs muchnik



Kolmogorov's framework:
Can we *materialize* the part $a \setminus b$?

conditional encoding: slepian-wolf vs muchnik

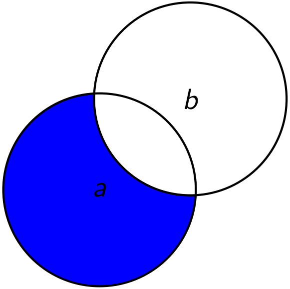


Kolmogorov's framework:
Can we *materialize* the part $a \setminus b$?

Yes, there exists a w such that

- $C(w) = C(a|b) + O(\log \ldots)$,
- $C(w|a) = O(\log \ldots)$,
- $H(a|b, w) = O(\log \ldots)$.

conditional encoding: slepian-wolf vs muchnik



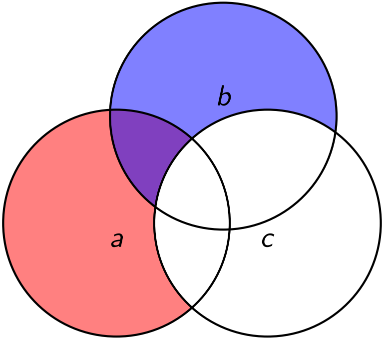
Kolmogorov's framework:
Can we *materialize* the part $a \setminus b$?

Yes, there exists a w such that

- $C(w) = C(a|b) + O(\log \dots)$,
- $C(w|a) = O(\log \dots)$,
- $H(a|b, w) = O(\log \dots)$.

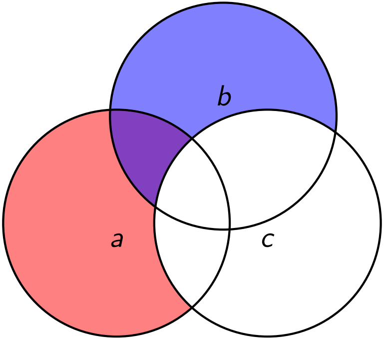
[Bennett-Gács-Li-Vitanyi-Zureck, Fortnow-Laplante, An. Muchnik]

two conditional descriptions: when the parallelism fails



Let $(a_1, b_1, c_1), \ldots, (a_n, b_n, c_n)$ be i.i.d., all distributed as (a, b, c).

two conditional descriptions: when the parallelism fails

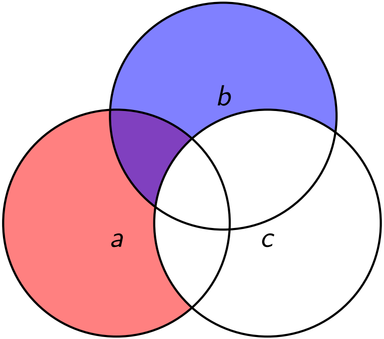


Let $(a_1, b_1, c_1), \ldots, (a_n, b_n, c_n)$ be i.i.d., all distributed as (a, b, c).

Then there exist V and W such that

- $H(V) = n \cdot H(a|c) + o(n)$ and $H(a_1 \ldots a_n | c_1 \ldots c_n, V) = 0$,
- $H(W) = n \cdot H(b|c) + o(n)$ and $H(b_1 \ldots b_n | c_1 \ldots c_n, W) = 0$,
- $H(V, W) = n \cdot H(a, b|c) + o(n)$.

two conditional descriptions: when the parallelism fails

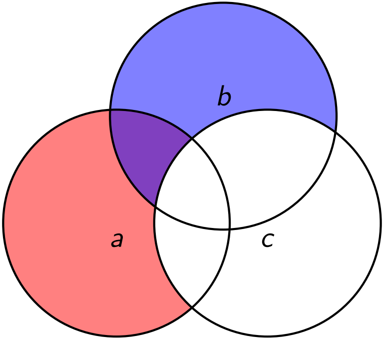


Let $(a_1, b_1, c_1), \ldots, (a_n, b_n, c_n)$ be i.i.d., all distributed as (a, b, c).

Then there exist V and W such that

- $H(V) = n \cdot H(a|c) + o(n)$ and $H(a_1 \ldots a_n | c_1 \ldots c_n, V) = 0$,
- $H(W) = n \cdot H(b|c) + o(n)$ and $H(b_1 \ldots b_n | c_1 \ldots c_n, W) = 0$,
- $H(V, W) = n \cdot H(a, b|c) + o(n)$. (i.e., $I_S(V, W : c_1 \ldots c_n) = o(n)$)

two conditional descriptions: when the parallelism fails



Let $(a_1, b_1, c_1), \ldots, (a_n, b_n, c_n)$ be i.i.d., all distributed as (a, b, c).

Then there exist V and W such that

- $H(V) = n \cdot H(a|c) + o(n)$ and $H(a_1 \ldots a_n | c_1 \ldots c_n, V) = 0$,
- $H(W) = n \cdot H(b|c) + o(n)$ and $H(b_1 \ldots b_n | c_1 \ldots c_n, W) = 0$,
- $H(V, W) = n \cdot H(a, b|c) + o(n)$. (i.e., $I_S(V, W : c_1 \ldots c_n) = o(n)$)

The counterpart of this statement for Kolmogorov complexity is wrong!