On the Non-robustness of Essentially Conditional Information Inequalities

Tarik Kaced and Andrei Romashchenko
LIRMM, CNRS & Univ. Montpellier II

Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier
LIRMM

3-7 September 2012 — IEEE Information Theory Workshop (ITW) 2012 — Lausanne, Switzerland

Introduction

What is it all about? It's a story of *conditional linear information inequalities*. i.e., linear inequalities for Shannon entropy that hold for distributions whose entropies meet some linear constraints.

We prove that some conditional information inequalities cannot be extended to any unconditional one. Some of these conditional inequalities hold for *almost entropic points*, while others do not.

Why should you care?

Answer 1: If you are working in theory, you probably want to know "the most universal laws" of information theory.

Answer 2: If you are working in applications, you should keep in mind that some information inequalities are *non robust*.

Linear information Inequalities (A)

Basic information inequalities [Shannon, 1940-s]:

- $H(a,b) \geq H(b)$, a.k.a. $H(a|b) \geq 0$,
- $H(a) + H(b) \geq H(a,b)$, a.k.a. $I(a;b) \geq 0$,
- $H(a,c) + H(b,c) \geq H(a,b,c) + H(c)$, a.k.a. $I(a;b|c) \geq 0$,

where a, b, c are (tuples of) jointly distributed discrete random variables.

Linear information Inequalities (B)

Shannon-type information inequalities :

all positive linear combinations of *basic* inequalities.

Linear information Inequalities (C)

non-Shannon-type information inequalities:

linear inequalities that hold for all distributions but **cannot** be represented as a positive combination of *basic* inequalities.

- Z. Zhang, R.W. Yeung [1998] :
$$I(c;d) \leq 2I(c;d|a) + I(c;d|b) + I(a;b) + +I(a;c|d) + I(a;d|c)$$
- R. Dougherty, C. Freiling, and K. Zeger [2006]: six other inequalities,
- several other examples,
- F. Matúš'07: there exist *infinitely many* independent linear information inequalities (with 4 random variables)

Applications of inf. inequalities

- fundamental limits in information theory:
  - in coding for noisy channels,
  - in compression,
  - in secrecy,
  - in streaming algorithms, etc.
- conditional independence relations,
- Kolmogorov complexity,
- combinatorics,
- group theory,
- etc., etc., etc.

A prominent example: **non-Shannon-type inequalities** result in progress in *secret sharing* [see A. Beimel, N. Livne, and C. Padró; A. Beimel and I. Orlov; J.R. Metcalf-Burton].

Conditional information inequalities

General form of a conditional inequality:

If [some linear constraints for entropies] then [a linear inequality for entropies].

Trivial examples of conditional information inequalities

Conditional inequalities that follow directly from unconditional ones:

- If $I(a;b) = 0$, then $H(a) + H(b) \leq H(a,b)$.
  Why? Because $H(a) + H(b) = H(a,b) + I(a;b)$.
- If $I(a;b) = 0$, then $H(a) + H(b) + H(c) \leq H(a,c) + H(b,c)$.
  Why? It follows from an **unconditional Shannon-type** inequality $H(a) + H(b) + H(c) \leq H(a,c) + H(b,c) + I(a;b)$.
- If $I(e;c|d) = I(e;d|c) = I(c;d|e) = 0$,
  then $I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b)$.
  Why? It follows from an **unconditional non-Shannon-type** inequality
  $I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b) + I(e;c|d) + I(e;d|c) + I(c;d|e)$

Essentially conditional information inequalities

($\mathcal{I}1$) [Zhang–Yeung'97]
If $I(a;b) = I(a;b|c) = 0$,
then $I(c;d) \leq I(c;d|a) + I(c;d|b)$.

($\mathcal{I}2$) [F. Matúš'99]
If $I(a;b|c) = I(b;d|c) = 0$,
then $I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b)$.

($\mathcal{I}3$) [K.-R.'11]
If $I(a;b|c) = H(c|a,b) = 0$,
then $I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b)$.

F. Matúš [2007] proved (implicitly):

($\mathcal{I}4$) If $I(a;d|c) = I(a;c|d) = 0$,
then $I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b)$.

($\mathcal{I}5$) If $I(b;c|d) = I(c;d|b) = 0$,
then $I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b)$.

+ three other conditional inequalities with 5 random variables.

Theorem 1. Inequalities ($\mathcal{I}1$–$\mathcal{I}5$) are essentially conditional.

E.g., inequality
$$I(a;b) = I(a;b|c) = 0 \Rightarrow I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b)$$
is **essentially conditional**: for all λ_1, λ_2 there exists a distribution (a,b,c,d) such that
$$I(c;d) \not\leq I(c;d|a) + I(c;d|b) + I(a;b) + \lambda_1 I(a;b) + \lambda_2 I(a;b|c)$$

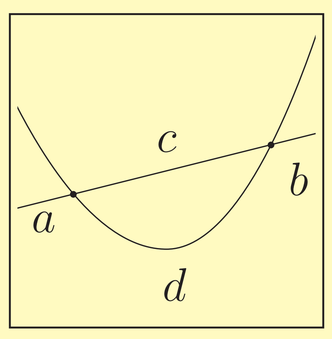
Remark. Yes, all these conditional inequalities are about one and the same **Ingleton inequality** $I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b)$. No, the Ingleton inequality is not valid for Shannon entropy *without* constraints.

Sketch of the proof for ($\mathcal{I}1$): Assume that for some λ_1, λ_2

$$(*)\quad I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a;b) + \lambda_1 I(a;b) + \lambda_2 I(a;b|c)$$

Consider an affine plane over a finite field $\mathbb{F}_q$. Define a distribution (a,b,c,d) as follows:

- Let c be a random non-vertical line c;
- pick independently and uniformly two points a and b in line c.
- pick a random parabola d that intersect c at points a and b.



$I(c;d) \approx 1$, since independently chosen line and parabola on the plane intersect almost half of the time. Also we have $I(c;d|a) = I(c;d|b) = I(a;b|c) = 0$ and $I(a;b) = O(\frac{\log q}{q})$. With a more accurate calculation, (*) results in
$$1 - \frac{1}{q} \leq \lambda_1 \frac{\log q}{q},$$
a contradiction (for large enough q).

Entropic and almost entropic points

Definition 1. A point $\mathbf{h} \in \mathbb{R}^{2^n-1}$ is called entropic if there exists a distribution $(x_1, \ldots, x_n)$ such that
$$\mathbf{h} = (H(x_1), \ldots, H(x_1, x_2), \ldots, H(x_1, \ldots, x_n)).$$

Definition 2. A point $\mathbf{h} \in \mathbb{R}^{2^n-1}$ is called almost entropic (a.e.) if it is a limit of a sequence of entropic points.

Remark 1. For each n the set of a.e. points is known to be *a closed convex cone*.

Remark 2. For each n the set of a.e. points is exactly the set of all points in $\mathbb{R}^{2^n-1}$ that satisfy all valid unconditional linear information inequalities involving n random variables.

Two kinds of essentially conditional information inequalities

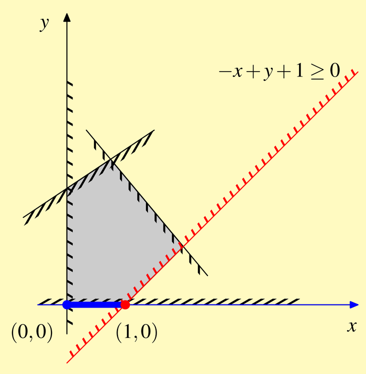
Theorem 2. ($\mathcal{I}4$–$\mathcal{I}5$) hold for a.e. points.
Proof: just inspect the proof of ($\mathcal{I}4$–$\mathcal{I}5$).

Theorem 3. ($\mathcal{I}1$–$\mathcal{I}3$) do not hold for a.e. points.
Proof: The construction from Th 1 + Slepian–Wolf.

Geometric intuition.

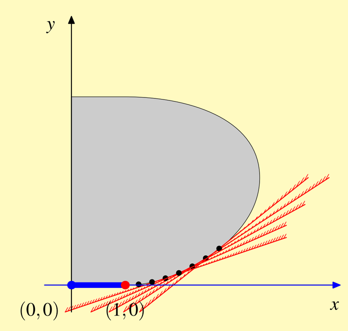
For simplicity we draw 2D polygons instead of a 15D cone:

Fig. 1. A conditional inequality is **not** essentially conditional, it is a "shade" of **one** unconditional inequality:



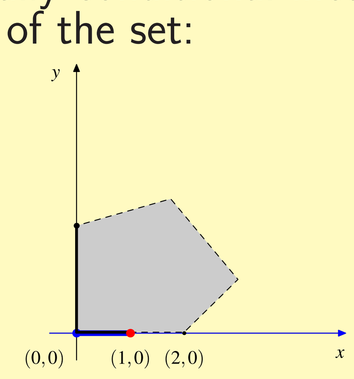
If $y = 0$ then $x \leq 1$. It follows from $-x + y + 1 \geq 0$.
cf. "Trivial" examples of conditional information inequalities.

Fig. 2. An essentially conditional inequality:



If $y = 0$ then $x \leq 1$. This conditional inequality is implied by an **infinite** family of tangent half-planes.
cf. Th. 2 on essentially conditional inequalities ($\mathcal{I}4$–$\mathcal{I}5$).

Fig. 3. An essentially conditional inequality that does not hold for the closure of the set:



If $y = 0$ then $x \leq 1$. For the closure of this set with the same constraint $y = 0$ we have only $x \leq 2$.
cf. Th. 3 on essentially conditional inequalities ($\mathcal{I}1$), ($\mathcal{I}3$).

Theorem 4 [Matúš'07]. For $n \geq 4$ the cone of all almost entropic points is not polyhedral.
Proof: Follows from ($\mathcal{I}4$), from ($\mathcal{I}5$), or any other essentially conditional inequality for a.e. points, see Fig. 2 above.

Open Problems

- Does inequality ($\mathcal{I}2$) hold for almost entropic points?
- Do there exist any essentially conditional inequalities with a constraint of co-dimension 1?
- What is the geometric, physical meaning of ($\mathcal{I}1$), ($\mathcal{I}3$)?