On the Non-robustness of Essentially Conditional Information Inequalities

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

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]:

\[ H(a,b) \geq H(b), \]

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

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

\[ I(a,b,c) = 0, \]

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.

\[ I(c;d) \leq I(c;d|a) + I(c;d|b) + (I(a) + I(b) + I(a|c) + I(b|c)) \]

\[ 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 1 random variable)

Applications of inf. inequalities

\[ \text{Definition 1. A point } h \in \mathbb{R}^d \text{ is called entropic if there exists a distribution } (x_1, \ldots, x_n) \text{ such that } h = (H(x_1), \ldots, H(x_n)). \]

\[ \text{Definition 2. A point } h \in \mathbb{R}^d \text{ is called almost entropic if it is a limit of a sequence of entropic points.} \]

Trivial examples of conditional information inequalities

If \( I(a,b) = 0 \), then \( H(a) + H(b) \leq H(a,b) \).

Why? Because \( H(a) + H(b) = H(a) + H(b) + H(c) \).

If \( I(a,b) = 0 \), then \( H(a) + H(b) = H(a) + H(b) + H(c) \).

Why? It follows from an unconditional Shannon-type inequality \( H(a) + H(b) + H(c) \leq H(a,c) + H(b,c) \).

Theorem 2. \((1-\Omega)\) hold for a.e. points.
Proof: Just inspect the proof of \((1-\Omega)\).

Theorem 3. \((1-\Omega)\) do not hold for a.e. points.
Proof: The construction from Th 1 + Spielman-Wolf.

Entropic and almost entropic points

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

Sketch of the proof for \((1)\):

Assume that for some \( \lambda_1, \lambda_2 \):

\[ I(c;d) \leq I(c;d|a) + I(c;d|b) + I(a) + \lambda_1 I(a|b) + \lambda_2 I(b|a). \]

Consider an affine plane over a finite field \( F_q \). Define a distribution \((a, b, c, d)\) as follows:

\[ \text{Let } c \text{ be a random non-vertical line } \}

\[ \text{pick independently and uniformly two points } a \text{ and } b \text{ on } \]

\[ \text{pick a random parabola \( d \) that intersects } \]

\[ \text{at points } a \text{ and } b \text{.} \]

\[ (c,d) \equiv 1, \] since independently chosen line and parabola on the plane intersect almost half of the time. Also we have \( I(c;d|a) = I(c;d|b) = I(a) = O(1/q) \) and \( (a,b) \equiv O(1/q) \).

With a more accurate calculation, \( (*) \) results in a contradiction (for large enough \( q \)).

Open Problems

Does inequality \((1)\) 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 \((1), (\Omega)\)?