Which Boolean Functions are Most Informative?

Gowtham R Kumar, Thomas Courtade

Electrical Engineering, Stanford University

Introduction

Let X^n be a sequence of i.i.d. Bernoulli $(1/2)$ random variables, Z^n be a sequence of i.i.d. Bernoulli (α) random variables independent of X^n, and $Y^n = X^n \oplus Z^n$, where "$\oplus$" denotes coordinate-wise XOR.

A *boolean function* is a function $b : \{0,1\}^n \to \{0,1\}$. We pose the following conjecture:

Conjecture 1: For any boolean function $b : \{0,1\}^n \to \{0,1\}$,

$$I(b(X^n); Y^n) \leq 1 - H(\alpha),$$

where $H(\cdot)$ is the entropy function.

Remark 1: The conjectured upper bound is attained by $b(X^n) = X_1$.

Despite its simple formulation, Conjecture 1 remains unproven. Traditional tools from information theory appear to be ineffective here due to the fact that the range of b is fixed to be one bit, and does not grow asymptotically large as it would in a typical information-theoretic formulation.

Motivation

Boolean functions are the fundamental building blocks of all data processing. The fact that (the apparently simple) Conjecture 1 cannot be solved indicates that we do not understand such functions at even the most basic information-theoretic level. Other applications include computational biology and communication with feedback.

The Lexicographic Ordering on $\{0,1\}^n$

The lexicographic ordering on $\{0,1\}^n$ is induced by the usual ordering on the integer representations of binary vectors. For example, the lexicographic ordering on $\{0,1\}^3$ is given by

$$000 \prec 001 \prec 010 \prec 011 \prec 100 \prec 101 \prec 110 \prec 111.$$

An initial segment of size k of the lexicographic ordering on $\{0,1\}^n$ consists of the first k elements of $\{0,1\}^n$ with respect to the lexicographic ordering. For instance, the initial segment of $\{0,1\}^3$ of length 4 is given by the set $\{000, 001, 010, 011\}$.

Initial segments of the lexicographic order play an important role in discrete isoperimetric inequalities. To this end, the *hypercube* is defined to be the graph with vertex set $\{0,1\}^n$, where two vertices are connected by an edge iff their binary representations have Hamming distance 1. The classical edge-isoperimetric inequality for the hypercube (cf. [Harper, 1964]) states the following:

Theorem 1: For a subset $A \subset \{0,1\}^n$ with fixed cardinality $|A| = k$, the number of edges in the hypercube connecting A to A^c is minimized when A is an initial segment of the lexicographic order.

Refined Conjectures

Attempts to establish Conjecture 1 directly yielded few results, therefore we pose the following refinements.

Conjecture 2: For a fixed $\Pr(b(X^n) = 0)$, the conditional entropy $H(b(X^n)|Y^n)$ is minimized when $b^{-1}(0)$ is an initial segment of the lexicographic order on $\{0,1\}^n$.

- Conjecture 2 is an information-theoretic analog of Theorem 1. Roughly speaking, it states that for a fixed preimage size $|b^{-1}(0)| = k$, the uncertainty of $b(X^n)$ evaluated on noisy inputs is minimized when b is an indicator function for an initial segment of the lexicographic order.

Conjecture 3: If $b^{-1}(0)$ is an initial segment of the lexicographic order on $\{0,1\}^n$, then

$$H(b(X^n)|Y^n) \geq H(b(X^n)) \cdot H(\alpha).$$

- Referring to Conjecture 3 as a "conjecture" is perhaps too modest. Indeed, we give a computer aided proof of the conjecture for α ranging from 0 to $1/2$ in increments of 0.001.
- Conjecture 3 can be interpreted as a strong data-processing inequality. Indeed, by rearranging terms, an equivalent formulation of Conjecture 3 is that

$$I(b(X^n); Y^n) \leq I(b(X^n); X^n) \cdot (1 - H(\alpha))$$

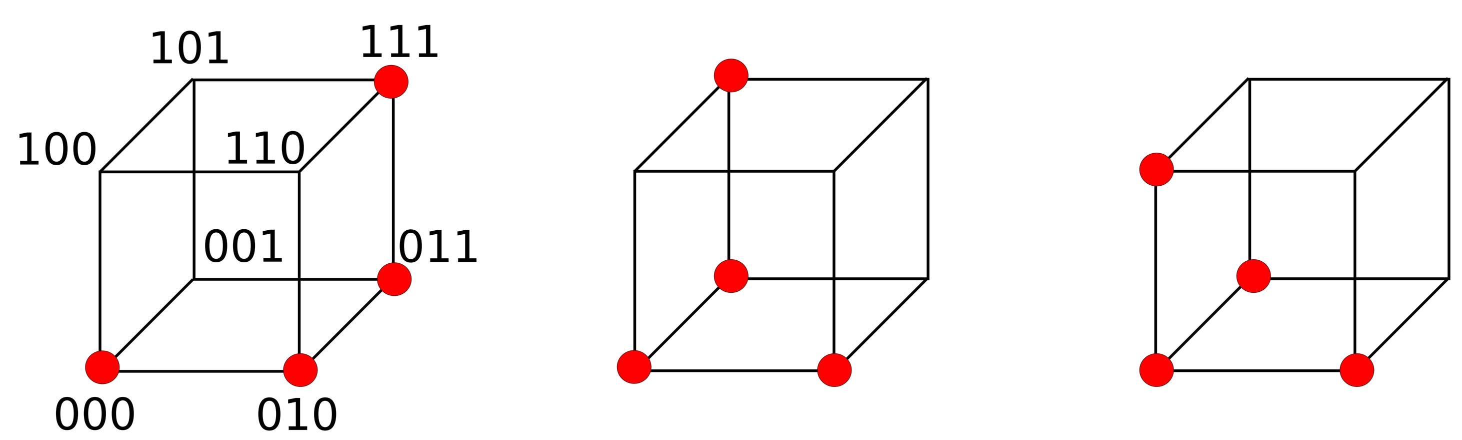
Assuming the validity of Conjectures 2 and 3, this inequality would hold for *all* boolean functions $b : \{0,1\}^n \to \{0,1\}$, and hence implies the claim of Conjecture 1.

Progress toward Conjecture 2

Although we have not yet proven Conjecture 2, we have been able to establish intermediate results. For example:

Theorem 2: For any b, there is a monotone function $\tilde{b}$ for which $\Pr(b(X^n) = 0) = \Pr(\tilde{b}(X^n) = 0)$ and $H(\tilde{b}(X^n)|Y^n) \leq H(b(X^n)|Y^n)$.

The proofs of these intermediate results are based on the notion of *compression operators* (see [Bollobás & Leader, 1991]). Compression operators modify the set $b^{-1}(0)$ in a way that preserves the cardinality, without increasing $H(b(X^n)|Y^n)$. Graphically, a sequence of 1-dimensional compressions on the set $A = \{000, 010, 011, 111\}$ proceeds as:



A $\quad$ $C_2(A)$ $\quad$ $C_3(C_2(A))$

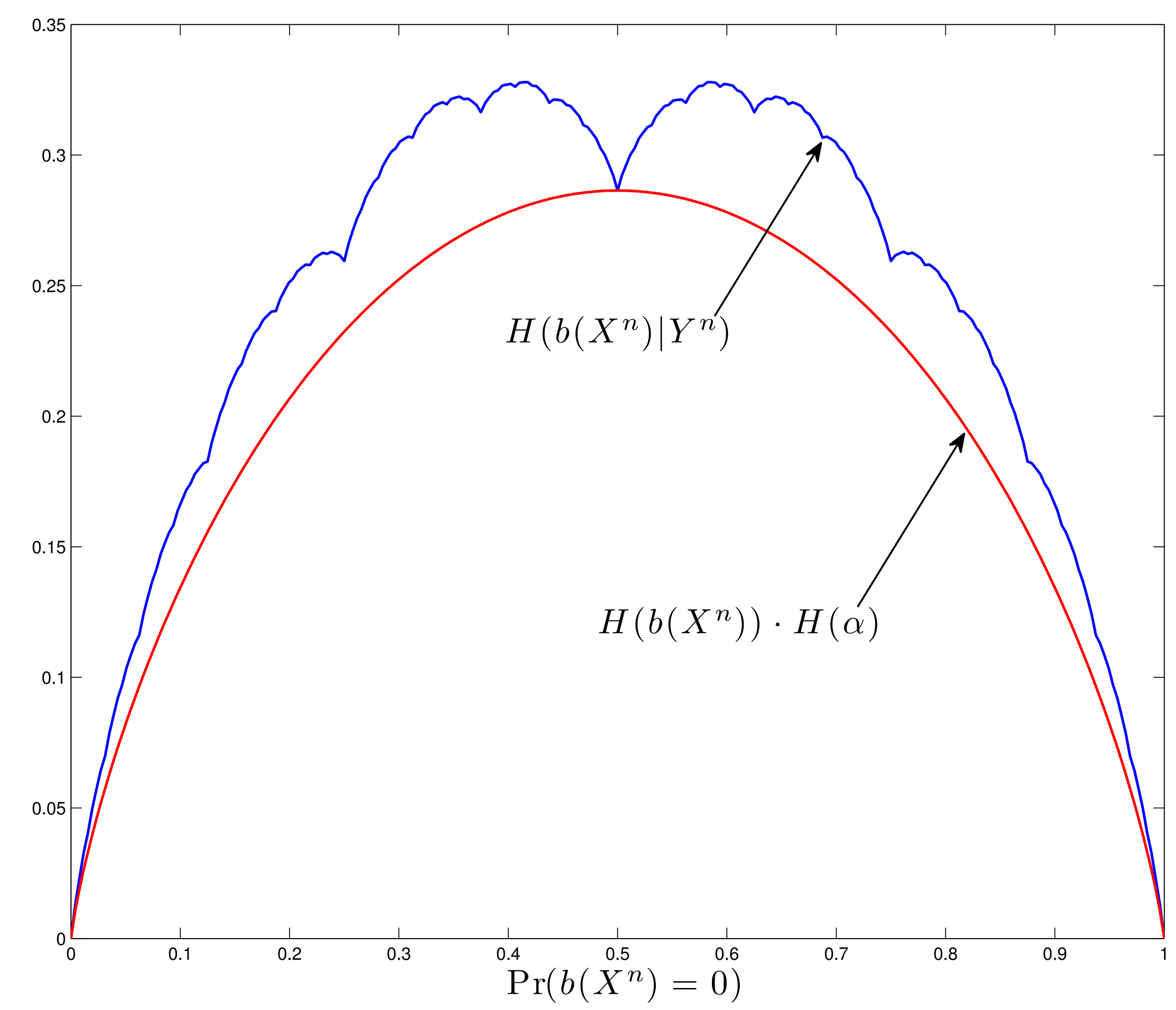
In addition to 1-dimensional compressions, 2-dimensional compression operators also no not increase $H(b(X^n)|Y^n)$. This allows the set of "candidate" functions which minimize $H(b(X^n)|Y^n)$ to be significantly reduced from the entire set of 2^{2^n} different boolean functions on $\{0,1\}^n$. Using these reductions, we have numerically validated Conjecture 2 for all $\alpha \in [0,1]$ for $n \leq 6$.

Assuming we could compute $H(b(X^n)|Y^n)$ for 10^6 different boolean functions per second, this validation would have required roughly $600,000$ years of computation without the reductions afforded by the compression operations!

Progress toward Conjecture 3

Even under the assumption that $b^{-1}(0)$ is an initial segment of the lexicographic order, the quantity $H(b(X^n)|Y^n)$ is very difficult to deal with. For instance, $H(b(X^n)|Y^n)$ is not monotone in $\Pr(b(X^n) = 0) \in [0, 1/2]$ as one might initially suspect. However, under the restriction that $b^{-1}(0)$ is an initial segment of the lexicographic order, the function $H(b(X^n)|Y^n)$ is "pseudo-concave" in $\Pr(b(X^n) = 0) \in [0, 1/2]$. Exploiting pseudo-concavity, we can give a computerized proof of Conjecture 3 for a fixed α.

Experiments illustrate a relationship between $H(b(X^n)|Y^n)$ and the Takagi function (an everywhere continuous, but nowhere differentiable function), which reinforces the connection to classical discrete isoperimetric inequalities (see [Lev, 2012]), and reveals the depth of the original conjecture.



Acknowledgement

- This work is supported in part by Air Force grant FA9550- 10-1-0124 and by the NSF Center for Science of Information under grant agreement CCF-0939370.
- We gratefully acknowledge helpful conversations with Abbas El Gamal and Chandra Nair.