A conjecture regarding optimality of the dictator function under Hellinger distance

Chandra Nair

The Chinese University of Hong Kong

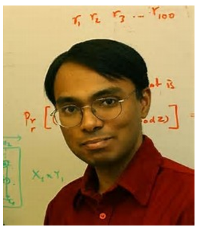
2017 ITA Workshop, UCSD

February 13, 2017

Collaborators (co-authors)

Venkat Anantharam
U.C. Berkeley

Andrej Bogdanov
CUHK



Amit Chakrabarti
Dartmouth

Thathachar Jayram
IBM Alamaden

Thanks: Simon's institute

Introduction

Starting point: the following conjecture[1] by Kumar ('12)

$\mathbf{X}$: uniform on $\{-1, +1\}^n$
$\mathbf{Y}$: obtained from $\mathbf{X}$ via the standard *noise-operator*, i.e.

- flip each bit (independently) with probability $\frac{1-\rho}{2}$.

Conjecture-MI

The dictator function $f_d(\mathbf{X}) = X_1$ maximizes the mutual information $I(f(\mathbf{X}); \mathbf{Y})$ among all boolean functions $f(\mathbf{X})$.

[1] Thomas A Courtade and Gowtham R Kumar. "Which Boolean functions maximize mutual information on noisy inputs?" In: *IEEE Transactions on Information Theory* 60.8 (2014), 4515–4525.

Introduction

Starting point: the following conjecture[1] by Kumar ('12)

$\mathbf{X}$: uniform on $\{-1,+1\}^n$
$\mathbf{Y}$: obtained from $\mathbf{X}$ via the standard *noise-operator*, i.e.

- flip each bit (independently) with probability $\frac{1-\rho}{2}$.

Conjecture-MI

The dictator function $f_d(\mathbf{X}) = X_1$ maximizes the mutual information $I(f(\mathbf{X});\mathbf{Y})$ among all boolean functions $f(\mathbf{X})$.

Alternate view
Let $\Phi_{JS}(x) := 1 - H_b(x) = JS[(x, 1-x), (1-x, x)]$.
Given $f(\mathbf{X}) : \{-1,+1\}^n \mapsto \{-1,+1\}$, let $Z_f(\mathbf{Y}) := \frac{1-(T_\rho f)(\mathbf{Y})}{2}$ where $(T_\rho f)(\mathbf{Y}) = \mathrm{E}(f(\mathbf{X})|\mathbf{Y})$.

Conjecture-MI (restatement)

The dictator function $f_d(\mathbf{X}) = X_1$ maximizes the Φ_{JS}-entropy, $\mathrm{E}(\Phi_{JS}(Z_f(\mathbf{Y}))) - \Phi_{JS}(\mathrm{E}(Z_f(\mathbf{Y})))$, among all boolean functions $f(\mathbf{X})$.

[1] Courtade and Kumar, "Which Boolean functions maximize mutual information on noisy inputs?"

Main Conjecture

Idea: Replace $\Phi_{JS}(x)$ by other convex functions.

Consider squared Hellinger distance between $(x, 1-x)$ and $(1-x, x)$

$$\Phi_{\mathcal{H}^2}(x) := 1 - 2\sqrt{x(1-x)}.$$

As before, given $f(\mathbf{X}) : \{-1,+1\}^n \mapsto \{-1,+1\}$, let $Z_f(\mathbf{Y}) := \frac{1-(T_\rho f)(\mathbf{Y})}{2}$.

Conjecture-SH

The dictator function $f_d(\mathbf{X}) = X_1$ maximizes the $\Phi_{\mathcal{H}^2}$-entropy, $\mathrm{E}(\Phi_{\mathcal{H}^2}(Z_f(\mathbf{Y}))) - \Phi_{\mathcal{H}^2}(\mathrm{E}(Z_f(\mathbf{Y})))$, among all boolean functions $f(\mathbf{X})$.

Equivalently

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \leq 1 - \sqrt{1-\rho^2}.$$

Main Conjecture

Idea: Replace $\Phi_{JS}(x)$ by other convex functions.

Consider squared Hellinger distance between $(x, 1-x)$ and $(1-x, x)$

$$\Phi_{\mathcal{H}^2}(x) := 1 - 2\sqrt{x(1-x)}.$$

As before, given $f(\mathbf{X}) : \{-1,+1\}^n \mapsto \{-1,+1\}$, let $Z_f(\mathbf{Y}) := \frac{1-(T_\rho f)(\mathbf{Y})}{2}$.

Conjecture-SH

The dictator function $f_d(\mathbf{X}) = X_1$ maximizes the $\Phi_{\mathcal{H}^2}$-entropy, $\mathrm{E}(\Phi_{\mathcal{H}^2}(Z_f(\mathbf{Y}))) - \Phi_{\mathcal{H}^2}(\mathrm{E}(Z_f(\mathbf{Y})))$, among all boolean functions $f(\mathbf{X})$.

Equivalently

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \leq 1 - \sqrt{1-\rho^2}.$$

1. Why is this interesting? (or, why should one care about this Φ?)
2. Evidence to the veracity of the conjecture
3. Weaker forms

About the Hellinger conjecture

In short, two lemmas:

1. Conjecture-SH implies Conjecture-MI
2. Conjecture-SH is extremal

About the Hellinger conjecture

In short, two lemmas:

1. Conjecture-SH implies Conjecture-MI
2. Conjecture-SH is extremal

Proposition

Conjecture-SH implies Conjecture-MI

Lemma

The function $H_b\left(\frac{1-\sqrt{1-x^2}}{2}\right)$ is non-negative, increasing, and convex in x for $x \in [0,1]$.

Let $\Psi(x) = \frac{1-\sqrt{1-x^2}}{2}$. Observe that

$$H\left(\frac{1-x}{2}\right) = H\left(\Psi\left(\sqrt{1-x^2}\right)\right).$$

Conjecture-MI can be expressed as

$$H\left(\Psi\left(\sqrt{1-\mathrm{E}(f)^2}\right)\right) - \mathrm{E}\left(H\left(\Psi\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right)\right)\right) \leq H(\Psi(1)) - H\left(\Psi\left(\sqrt{1-\rho^2}\right)\right).$$

Idea of proof

By the convexity of $H(\Psi(x))$ (lemma) suffices to show

$$H\left(\Psi\left(\sqrt{1-\mathrm{E}(f)^2}\right)\right) - H\left(\Psi\left(\mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right)\right)\right) \leq H(\Psi(1)) - H\left(\Psi\left(\sqrt{1-\rho^2}\right)\right).$$

However Conjecture-SH implies

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \leq 1 - \sqrt{1-\rho^2}.$$

Apply weak-majorization inequality: in particular use convexity, non-negativeness, and increasing property of $H(\Psi(x))$.

Extremality of Hellinger Conjecture

Conjecture-SH states

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \leq 1 - \sqrt{1-\rho^2}.$$

Take $\lim \rho \to 1$ (clean channel).
If Conjecture-SH is true then (for balanced Boolean functions)

$$\mathrm{E}\left(\sqrt{Sen_f(\mathbf{X})}\right) \geq 1.$$

$Sen_f(\mathbf{x})$: sensitivity at $\mathbf{x}$, number of neighbors with opposite value of $f(\mathbf{x})$.

Extremality of Hellinger Conjecture

Conjecture-SH states

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \leq 1-\sqrt{1-\rho^2}.$$

Take $\lim \rho \to 1$ (clean channel).
If Conjecture-SH is true then (for balanced Boolean functions)

$$\mathrm{E}\left(\sqrt{Sen_f(\mathbf{X})}\right) \geq 1.$$

$Sen_f(\mathbf{x})$: sensitivity at $\mathbf{x}$, number of neighbors with opposite value of $f(\mathbf{x})$.

Similar limit for Conjecture-MI would be equivalent to (for balanced Boolean functions)

$$\mathrm{E}\left(Sen_f(\mathbf{X})\right) \geq 1.$$

This is known to be true (Poincare's inequality, Pareseval's theorem, Harper's isoperimetric inequality)

Extremality of Hellinger Conjecture

Conjecture-SH states

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \le 1 - \sqrt{1-\rho^2}.$$

Take $\lim \rho \to 1$ (clean channel).
If Conjecture-SH is true then (for balanced Boolean functions)

$$\mathrm{E}\left(\sqrt{Sen_f(\mathbf{X})}\right) \ge 1.$$

$Sen_f(\mathbf{x})$: sensitivity at $\mathbf{x}$, number of neighbors with opposite value of $f(\mathbf{x})$.

On the other hand, best lower bound for balanced functions

$$\mathrm{E}\left(\sqrt{Sen_f(\mathbf{X})}\right) \ge \sqrt{\frac{2}{\pi}} \text{ (Bobkov '98).}$$

Therefore even in this limit, the conjecture would imply something new.

Extremality of Hellinger Conjecture

Conjecture-SH states

$$\sqrt{1 - \mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1 - (T_\rho f)^2(\mathbf{Y})}\right) \leq 1 - \sqrt{1 - \rho^2}.$$

Take $\lim \rho \to 1$ (clean channel).
If Conjecture-SH is true then (for balanced Boolean functions)

$$\mathrm{E}\left(\sqrt{Sen_f(\mathbf{X})}\right) \geq 1.$$

$Sen_f(\mathbf{x})$: sensitivity at $\mathbf{x}$, number of neighbors with opposite value of $f(\mathbf{x})$.

Lemma

For any $\alpha < \frac{1}{2}$, let $maj(\mathbf{Y})$ denote the majority function (assume that n is odd).
Then there exists large enough n such that

$$\mathrm{E}\left(Sen^{\alpha}_{maj}(\mathbf{Y})\right) < \mathrm{E}\left(Sen^{\alpha}_{dic}(\mathbf{Y})\right) = 1.$$

Evidence to the veracity of Conjecture-SH

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \leq 1 - \sqrt{1-\rho^2}.$$

1. verified numerically until $n = 8$
2. Conjecture-SH is true if

$$\sqrt{1-\rho^2} + \sqrt{1-\mathrm{E}(f)^2} \leq 1$$

Evidence to the veracity of Conjecture-SH

$$\sqrt{1-\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-(T_\rho f)^2(\mathbf{Y})}\right) \leq 1 - \sqrt{1-\rho^2}.$$

1. verified numerically until $n = 8$
2. Conjecture-SH is true if

$$\sqrt{1-\rho^2} + \sqrt{1-\mathrm{E}(f)^2} \leq 1 + (1-\rho^2)(1-\mathrm{E}(f)^2)$$

On numerical verification

- Issue: Number of Boolean functions is 2^{2^n}
- **Lemma**: For any convex Φ, there is a *doubly monotone* boolean function that maximizes the Φ-entropy, $\mathrm{E}(\Phi(Z_f)) - \Phi(\mathrm{E}(Z_f))$, where maximization is over all boolean functions.
- While the number of *doubly monotone* boolean functions also grows doubly exponentially, still amenable till $n = 8$ (or a bit more).

On lemma

Doubly-monotone: A boolean function is said to be doubly-monotone if it is monotone, and for any $1 \leq i < j \leq n$,

$$f(S \cup \{i, j\}) \geq f(S \cup \{i\}) \geq f(S \cup \{j\}) \geq f(S), \quad \forall S \subseteq [1:n].$$

Proof: Follows from majorization and Karamata's inequality.
Similar argument also present in (Courtade-Kumar '14)

2nd evidence: Proof in the parameter regime

$$G(\lambda) := \sqrt{1-(1-\lambda)\,\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-\lambda\rho^2-(1-\lambda)g^2(\mathbf{y})}\right),$$

where $g(\mathbf{y}) = (T_\rho f)(\mathbf{y})$. Want to show $G(1) \geq G(0)$.

$$G'(\lambda) = \frac{\mathrm{E}(f)^2}{2\sqrt{1-(1-\lambda)\,\mathrm{E}(f)^2}} - \mathrm{E}\left(\frac{g^2(\mathbf{y})-\rho^2}{2\sqrt{1-\lambda\rho^2-(1-\lambda)g^2(\mathbf{y})}}\right)$$

2nd evidence: Proof in the parameter regime

$$G(\lambda) := \sqrt{1-(1-\lambda)\,\mathrm{E}(f)^2} - \mathrm{E}\left(\sqrt{1-\lambda\rho^2-(1-\lambda)g^2(\mathbf{y})}\right),$$

where $g(\mathbf{y}) = (T_\rho f)(\mathbf{y})$. Want to show $G(1) \geq G(0)$.

$$G'(\lambda) = \frac{\mathrm{E}(f)^2}{2\sqrt{1-(1-\lambda)\,\mathrm{E}(f)^2}} - \mathrm{E}\left(\frac{g^2(\mathbf{y})-\rho^2}{2\sqrt{1-\lambda\rho^2-(1-\lambda)g^2(\mathbf{y})}}\right)$$

Lemma

For any $0 \leq \rho^2, \lambda \leq 1$ *the function*

$$f(u) := \frac{u-\rho^2}{\sqrt{1-\lambda\rho^2-(1-\lambda)u}}$$

is convex and increasing in u *when* $u \in [0,1]$.

If U is a random variable that takes values in $[0,1]$ then

$$E(f(U)) \leq (1-E(U))f(0) + E(U)f(1).$$

Calculus of variations..

Denoting $\alpha = \mathrm{E}(g^2(Y))$ we obtain

$$G'(\lambda) \geq \frac{\mathrm{E}(f)^2}{2\sqrt{1-(1-\lambda)\,\mathrm{E}(f)^2}} + (1-\alpha)\frac{\rho^2}{2\sqrt{1-\lambda\rho^2}} - \alpha\frac{\sqrt{1-\rho^2}}{2\sqrt{\lambda}}. \tag{1}$$

Thus, integrating both sides with respect to λ from 0 to 1 we obtain

$$\int_0^1 G'(\lambda)d\lambda \geq 2 - \alpha - \sqrt{1-\mathrm{E}(f)^2} - \sqrt{1-\rho^2}.$$

Since $\alpha \leq \mathrm{E}(f)^2 + \rho^2(1-\mathrm{E}(f)^2)$ (by Parseval) we are done if

$$1 + (1-\mathrm{E}(f)^2)(1-\rho^2) \geq \sqrt{1-\mathrm{E}(f)^2} + \sqrt{1-\rho^2}.$$

Weaker form

Conjecture[†]-SH-W

For any pairs of boolean functions $f(\mathbf{X}), g(\mathbf{Y})$ on the hypercube taking values in $\{-1, +1\}$,

$$\sqrt{1 - \mathrm{E}(f(\mathbf{X}))^2} - \mathrm{E}\left(\sqrt{1 - \mathrm{E}(f(\mathbf{X})|g(\mathbf{Y}))^2}\right) \leq 1 - \sqrt{1 - \rho^2}.$$

Weaker form

Conjecture†-SH-W

For any pairs of boolean functions $f(\mathbf{X}), g(\mathbf{Y})$ on the hypercube taking values in $\{-1, +1\}$,

$$\sqrt{1 - \mathrm{E}(f(\mathbf{X}))^2} - \mathrm{E}\left(\sqrt{1 - \mathrm{E}(f(\mathbf{X})|g(\mathbf{Y}))^2}\right) \leq 1 - \sqrt{1 - \rho^2}.$$

If this is true, then it would imply the following:

Proposition (Pichler et.al. '16)

For any pairs of boolean functions $f(\mathbf{X}), g(\mathbf{Y})$ on the hypercube,

$$I(f(\mathbf{X}); g(\mathbf{Y})) \leq 1 - H_b\left(\frac{1-\rho}{2}\right)$$

However the techniques used here is insufficient to prove Conjecture†-SH-W.

Why is the weak form true[†]?

Conjecture[†]

For any pair of binary random variables (U, V) with V taking values in $\{-1, 1\}$ the following holds:

$$\sqrt{1 - \mathrm{E}(V)^2} - \mathrm{E}\left(\sqrt{1 - \mathrm{E}(V|U)^2}\right) + \sqrt{1 - s_\dagger(U;V)} \leq 1,$$

where $s_\dagger(U;V) = \lim_{p\to 0} s_p(U;V)$.

$s_p(U;V)$: reverse-hypercontractivity parameter.

Why is the weak form true[†]?

Conjecture[†]

For any pair of binary random variables (U, V) with V taking values in $\{-1, 1\}$ the following holds:

$$\sqrt{1 - \mathrm{E}(V)^2} - \mathrm{E}\left(\sqrt{1 - \mathrm{E}(V|U)^2}\right) + \sqrt{1 - s_\dagger(U; V)} \leq 1,$$

where $s_\dagger(U; V) = \lim_{p \to 0} s_p(U; V)$.

$s_p(U; V)$: reverse-hypercontractivity parameter.

Reverse-hypercontractivity

A pair of random variables (U, V) is said to be (p, q)-*reverse-hypercontractive* for $1 > q \geq p$ if

$$E(f(U)g(V)) \geq \|f(U)\|_{p'} \|g(V)\|_q.$$

For an fixed $p < 1$ define

$$s_p(U; V) := \sup \left\{ \frac{q-1}{p-1} : (U, V) \text{ is } (p, q)\text{-}\textit{reverse-hypercontractive} \right\}.$$

An inequality†

For any $(s, c, d) \in [0, 1]$, the inequality

$$\sqrt{1-(s(\bar{d}-d)+\bar{s}(c-\bar{c}))^2} - s\sqrt{1-(\bar{d}-d)^2} - \bar{s}\sqrt{1-(\bar{c}-c)^2} + \sqrt{\frac{D(s\bar{c}+\bar{s}d\|s\bar{d}+\bar{s}c)}{sD(c\|d)+\bar{s}D(d\|c)}} \leq 1.$$

- Seems to be true (numerical simulations)
- If true, will imply Conjecture† (hence Conjecture†-SH-W)
- Can formally establish it for certain parameters, including a neighborhood of equality achieving points*.

An inequality†

For any $(s, c, d) \in [0, 1]$, the inequality

$$\sqrt{1 - (s(\bar{d} - d) + \bar{s}(c - \bar{c}))^2} - s\sqrt{1 - (\bar{d} - d)^2} - \bar{s}\sqrt{1 - (\bar{c} - c)^2} + \sqrt{\frac{D(s\bar{c} + \bar{s}d \| s\bar{d} + \bar{s}c)}{sD(c\|d) + \bar{s}D(d\|c)}} \leq 1.$$

- Seems to be true (numerical simulations)
- If true, will imply Conjecture† (hence Conjecture†-SH-W)
- Can formally establish it for certain parameters, including a neighborhood of equality achieving points*.

Remarks:

- Conjecture† may also be of independent interest
- Can possibly obtain a computer assisted formal proof of the inequality above

Conclusion

Proposed $\Phi_{\mathcal{H}^2}(x)$ to be *an extremal* function for which dictator mamizes the Φ-entropy.

- Gave a proof in some (limited) parameter regimes

Proposed an explicit three variable inequality that establishes the weaker form

- This is done via another inequality involving hypercontracitivity

Conclusion

Proposed $\Phi_{\mathcal{H}^2}(x)$ to be *an extremal* function for which dictator mamizes the Φ-entropy.

- Gave a proof in some (limited) parameter regimes

Proposed an explicit three variable inequality that establishes the weaker form

- This is done via another inequality involving hypercontracitivity

Lots of open conjectures

Connections to deeper stuff

- Talagrand's inequality (via Bobkov)

Thank You