A conjecture regarding optimality of the dictator function under Hellinger distance

Venkat Anantharam, Andrej Bogdanov, Amit Chakrabarti, T.S. Jayram, Chandra Nair

Abstract

We conjecture that the dictator function maximizes, among all Boolean functions on the hypercube, the Phi-entropy associated with Hellinger distance. This conjecture is shown to have interesting consequences and present also other related conjectures. We also give some evidence to the veracity of this conjecture.

I. Introduction

Let $\mathbf{X}$ be uniformly distributed on the hypercube $\{+1,-1\}^n$ and $\mathbf{Y}$ be a correlated random vector defined coordinatewise by $Y_i = X_i Z_i$, where $Z_1, \ldots, Z_n$ is an i.i.d. random vector distributed as $\mathrm{P}(Z_i = 1) = \frac{1+\rho}{2}$ and $\mathrm{P}(Z_i = 1) = \frac{1-\rho}{2}$. The following conjecture proposed by Kumar and appearing in [5], has attracted attention during the past few years.

Conjecture 1. *For any Boolean function, $f(\mathbf{X})$, defined on the Boolean hypercube*

$$I(f(\mathbf{X}); \mathbf{Y}) \leq I(X_1; Y_1) = 1 - H_b\left(\frac{1-\rho}{2}\right),$$

where $\mathbf{X}$ and $\mathbf{Y}$ are related as above. In the above relation $H_b(x) := -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy function and $I(\cdot;\cdot)$ denotes the mutual information, with the logarithm also being to base 2.

Remark 1. An alternate way of stating the conjecture is to say that a Boolean-valued $f(\mathbf{X})$ that maximizes the mutual information between $f(\mathbf{X})$ and $\mathbf{Y}$ is any dictator function, i.e. $f(\mathbf{X}) = X_i$ for some i.

The Jensen-Shannon divergence between two distributions $p(\mathbf{x})$ and $q(\mathbf{x})$ is defined as

$$JS(p,q) := \frac{1}{2} D\left(p \| m\right) + \frac{1}{2} D\left(q \| m\right),$$

where $m := \frac{p+q}{2}$ and $D(p\|q)$ is the Kullback-Liebler divergence (or relative entropy) between the two distributions, with the logarithm also being to base 2.

Let $\Phi_{JS}(x)$ denote the Jenson-Shannon divergence between the two-point distributions with probabilities $(x, 1-x)$ and $(1-x, x)$. Observe that $\Phi_{JS}(x) = 1 - H_b(x)$. Then the conjecture stated above can be reformulated as

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

In above $(T_\rho f)(\mathbf{Y}) := \mathrm{E}(f(\mathbf{X})|\mathbf{Y})$.

For a convex function Φ, the Φ-entropy is defined by

$$H_\Phi(Z) := \mathrm{E}(\Phi(Z)) - \Phi(\mathrm{E}(Z)).$$

For a given boolean function $f(\mathbf{X})$ defined on the hypercube, we can define an induced non-negative random variable $Z_f(\mathbf{Y}) = \frac{1-(T_\rho f)(\mathbf{Y})}{2}$, where $\mathbf{Y}$ is obtained from $\mathbf{X}$, via the noise operator T_ρ as stated above. Conjecture 1 can be phrased equivalently as $f(\mathbf{X}) = X_i$ (for some i) maximizes the Φ-entropy when $\Phi(x) = \Phi_{JS}(x)$.

Thus we can try to seek other convex functions Φ for which dictator is the maximizer. In this paper we propose the following convex function: $\Phi_{\mathcal{H}^2}(x) = 1 - 2\sqrt{x(1-x)}$, squared Hellinger distance between the probability distributions $(x, 1-x)$ and $(1-x, x)$. We make the following conjecture:

Conjecture 2. *For any Boolean function that takes values in $\{-1,+1\}$*

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

There are two reasons why this conjecture may be interesting:

- Conjecture 2 is a strengthening of the original conjecture (see Proposition 1)
- Conjecture 2 is extremal, in a sense made precise by Lemma 2, for the optimality of the dictator function.

A. Relation between the two conjectures

In this section we show that Conjecture 2 implies Conjecture 1. At the heart of this implication is the following convexity result, established in [2].

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

Proof. The second derivative yields

$$\frac{d^2 H\left(\frac{1-\sqrt{1-x^2}}{2}\right)}{dx^2} = \frac{1}{(1-x^2)^{3/2}\ln 2}\left(\ln\left(\frac{1+\sqrt{1-x^2}}{1-\sqrt{1-x^2}}\right) - 2\sqrt{1-x^2}\right) \geq 0$$

since $\ln\left(\frac{1+a}{1-a}\right) \geq 2a, a \in [0,1)$ establishing convexity. That the function is non-negative and increasing in the range is immediate. □

We use the above lemma to establish the following proposition.

Proposition 1. *Suppose a Boolean function $f(\mathbf{X})$ taking values $\{-1,+1\}$ satisfies*

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

then the function $f(\mathbf{X})$ also satisfies

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

Proof. Let $\Psi(x) = \frac{1-\sqrt{1-x^2}}{2}$. Lemma 1 implies that $H(\Psi(x))$ is non-negative, increasing, and convex in x. Observe that

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

The inequality in (2) can be re-expressed as

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

By the convexity of $H(\Psi(x))$ the above inequality will be implied if the following inequality holds:

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

However this inequality is an immediate consequence of (1) and using the fact that $H(\Psi(x))$ is increasing, and convex in x. □

B. Implications of the conjecture

We look at the limiting regimes of the conjecture ($\rho \to 1$ and $\rho \to 0$).

1) $\rho \to 1$ regime.: The conjecture is trivially true at $\rho = 1$ and for unbalanced functions, i.e. $\mathrm{E}(f) \neq 0$, it is also trivially true in a neighborhood (depending on $\mathrm{E}(f)$) around $\rho = 1$. Thus we focus our attention on the implication of this conjecture for balanced functions, $\mathrm{E}(f) = 0$, at ρ close to 1. Note that in this case Conjecture 2 implies

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

Elementary calculus shows that this is true as $\rho \to 1$ *if and only if*

$$\mathrm{E}\left(\sqrt{Sen_f(\mathbf{Y})}\right) \geq 1, \tag{3}$$

where $Sen_f(\mathbf{y})$ denotes the sensitivity (number of neighbors of $\mathbf{y}$ where the functions flips the value) of the function.

Remark 2. The best upper bound for $\mathrm{E}\left(\sqrt{Sen_f(\mathbf{Y})}\right)$ for a balanced Boolean function is $\sqrt{\frac{2}{\pi}}$ due to Bobkov, [4]; hence even in this limiting regime Conjecture 2, if true, would yield a non-trivial result. Note that in the same regime, Conjecture 1 reduces to $\mathrm{E}\left(Sen_f(\mathbf{Y})\right) \geq 1$. This is known as Poincare's inequality, but can also be deduced from Harper's isoperimetric inequality or by just looking at its Fourier expansion and using Parseval's identity.

The next lemma shows an extremal nature of Conjecture 2 for the optimality of dictator functions. Note that for any $\mathbf{y}$ on the hypercube, $Sen_{dict}(\mathbf{y}) = 1$.

Lemma 2. *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*

$$\mathrm{E}\left(Sen_{maj}^{\alpha}(\mathbf{Y})\right) < 1.$$

Proof. Let $n = 2k+1$. Note that

$$Sen_{maj}(\mathbf{y}) = \begin{cases} k+1 & \sum_{i=1}^{2k+1} y_i \in \{-1,1\}, \\ 0 & o.w. \end{cases}.$$

Thus

$$\mathrm{E}\left(Sen_{maj}^{\alpha}(\mathbf{Y})\right) = \frac{1}{2^{2k+1}} 2. \binom{2k+1}{k} (k+1)^{\alpha}$$

By Stirling's approximation we know that $\frac{\binom{2k+1}{k}\sqrt{k+1}}{2^{2k+1}} = \Theta(1)$; hence the result follows. □

Remark 3. When $\alpha = \frac{1}{2}$ it is not hard to show that

$$\mathrm{E}\left(Sen_{maj}^{1/2}(\mathbf{Y})\right) \geq 1, \ \forall n.$$

Thus Conjecture 2 yields the optimal exponent for the Sensitivity ($\rho \to 1$ limit) where dictator performs better than majority.

2) $\rho \to 0$ regime.: In this regime, expanding the left and side and right hand side of Conjecture 1 in powers of ρ reduces to

$$\frac{\rho^2}{2\sqrt{1-\hat{f}_\phi^2}} \sum_{i=1}^{n} \hat{f}_i^2 \leq \frac{1}{2}\rho^2 + O(\rho^3),$$

which trivially holds by Parseval's theorem. In the above, $\hat{f}_S$ denotes the Fourier co-efficients of the Boolean function.

II. Evidence for the veracity of the Conjecture

A. *Proof for certain parameter regimes*

In this section we prove the Hellinger conjecture for all functions such that the following inequality holds:

$$1+(1-\hat{f}_\phi^2)(1-\rho^2) \geq \sqrt{1-\hat{f}_\phi^2}+\sqrt{1-\rho^2}.$$

This is a slight improvement over the trivial regime given by

$$1-\sqrt{1-\rho^2} \geq \sqrt{1-\hat{f}_\phi^2}.$$

Define the function

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

where $g(\mathbf{y})=(T_\rho f)(\mathbf{y})$. The goal is to show that $G(1)\geq G(0)$. Observe that

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

Lemma 3. *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* $u\in[0,1]$.

Proof. Observe that

$$\begin{aligned}
f'(u)&=\frac{1}{\sqrt{1-\lambda\rho^2-(1-\lambda)u}}+\frac{(1-\lambda)(u-\rho^2)}{2(1-\lambda\rho^2-(1-\lambda)u)^{3/2}}\\
&=\frac{2-(1+\lambda)\rho^2-(1-\lambda)u}{2(1-\lambda\rho^2-(1-\lambda)u)^{3/2}}.\\
f''(u)&=\frac{-(1-\lambda)}{2(1-\lambda\rho^2-(1-\lambda)u)^{3/2}}+\frac{3(1-\lambda)}{2}\frac{(2-(1+\lambda)\rho^2-(1-\lambda)u)}{2(1-\lambda\rho^2-(1-\lambda)u)^{5/2}}\\
&=\frac{1-\lambda}{4(1-\lambda\rho^2-(1-\lambda)u)^{5/2}}(-2(1-\lambda\rho^2-(1-\lambda)u)+3(2-(1+\lambda)\rho^2-(1-\lambda)u))\\
&=\frac{1-\lambda}{4(1-\lambda\rho^2-(1-\lambda)u)^{5/2}}(4-(3+\lambda)\rho^2-(1-\lambda)u).
\end{aligned}$$

□

An immediate corollary of Lemma 3 is that if U is a random variable that takes values in $[0,1]$ then

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

where $f(u)$ is the function defined earlier.

Denoting $\alpha=\mathrm{E}(g^2(Y))$ and observing that $g^2(Y)\in[0,1]$ we use the above observation to obtain

$$G'(\lambda)\geq\frac{\hat{f}_\phi^2}{2\sqrt{1-(1-\lambda)\hat{f}_\phi^2}}+(1-\alpha)\frac{\rho^2}{2\sqrt{1-\lambda\rho^2}}-\alpha\frac{\sqrt{1-\rho^2}}{2\sqrt{\lambda}}. \tag{4}$$

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

$$\begin{aligned}
\int_0^1 G'(\lambda)d\lambda&\geq 1-\sqrt{1-\hat{f}_\phi^2}+(1-\alpha)(1-\sqrt{1-\rho^2})-\alpha\sqrt{1-\rho^2}\\
&=2-\alpha-\sqrt{1-\hat{f}_\phi^2}-\sqrt{1-\rho^2}.
\end{aligned}$$

Since $\alpha \leq \hat{f}_\phi^2 + \rho^2(1-\hat{f}_\phi^2)$ (by Parseval) we have, as long as

$$1 + (1-\hat{f}_\phi^2)(1-\rho^2) \geq \sqrt{1-\hat{f}_\phi^2} + \sqrt{1-\rho^2}$$

we will have

$$\int_0^1 G'(\lambda) d\lambda \geq 0$$

and we are done.

B. *Numerical verification until n=8*

We performed numerical verification until n=8. The key to performing this verification (since the number of Boolean functions grow doubly exponentially) are the following lemmas that allow us to restrict our attention to *pairwise monotone* functions. We will indeed show that such a result holds for the maximization over Boolean functions of $H_\Phi(Z_f(\mathbf{Y}))$ for any convex function Φ.

1) The family of $\mathcal{L}$-operators on Boolean functions: In this section, we define an $\mathcal{L}$-operator on a Boolean function and study its properties. In the background, we have a lexicographic ordering of the vertices of the hypercube where the vector of all -1 is the lowest element of the order and the vector of all $+1$ is the highest element. Further we assume that the coordinates are from the most significant bit to the least significant bit.

Definition 1. For any $k \in \{1,\ldots,n\}$ and set $\mathcal{S} \subseteq [1:n], |S| = k$ the $\mathcal{L}_\mathcal{S}$ operator transforms a boolean function $f(\mathbf{x})$ into another function $f^\uparrow(\mathbf{x})$ as follows: for any $\mathbf{x}' \in \{-1,+1\}^{n-k}$, let $\mathcal{V} = \{\mathbf{x}_1, ..., \mathbf{x}_{2^k}\}$ be the (increasing) lexicographic ordering of the set of co-ordinates whose restriction to $[1:n] \setminus S$ agrees with $\mathbf{x}'$. Then $\big(f^\uparrow(\mathbf{x}_1), ..., f^\uparrow(\mathbf{x}_{2^k})\big)$ is defined to be the ascending sort of the vector $(f(\mathbf{x}_1), ..., f(\mathbf{x}_{2^k}))$.

We also define 2^k functions on $\{-1,+1\}^{n-k}$ according to

$$f_j(\mathbf{x}') = f(\mathbf{x}_j), \quad j = 1,..,2^k.$$

Using this definition, we can view $\big(f^\uparrow(\mathbf{x}_1), ..., f^\uparrow(\mathbf{x}_{2^k})\big)$ as the ascending sort of the vector $(f_1(\mathbf{x}'), ..., f_{2^k}(\mathbf{x}'))$.

Remark 4. The following remark is worth noting: for any $\mathcal{S}$ the operator $\mathcal{L}_\mathcal{S}$ performs a permutation of the function values on the points of the hypercube. Hence $\mathrm{E}(f)$ is not altered by $\mathcal{L}_\mathcal{S}$.

In this section we assume that $0 < \rho < 1$ (without loss of generality).

Proposition 2. *For any $i \in \{1,..,n\}$ and for any boolean function $f(\mathbf{X})$, define $f^\uparrow = \mathcal{L}_{\{i\}}(f)$. Then for any convex function ϕ we have*

$$\mathrm{E}(\Phi(Z_{f^\uparrow}(\mathbf{Y}))) - \Phi(\mathrm{E}(Z_{f^\uparrow}(\mathbf{Y}))) \geq \mathrm{E}(\Phi(Z_f(\mathbf{Y}))) - \Phi(\mathrm{E}(Z_f(\mathbf{Y}))).$$

Here $Z_f(\mathbf{Y}) = \frac{1-(T_\rho f)(\mathbf{Y})}{2}$.

Proof. Since the $\mathcal{L}_\mathcal{S}$ operators preserve the means, hence it suffices to show that $\mathrm{E}(\Phi(Z_{f^\uparrow}(\mathbf{Y}))) \geq E(\Phi(Z_f(\mathbf{Y})))$. Observe that

$$f(\mathbf{x}) = \frac{(1-x_i)}{2} f_1(\mathbf{x}') + \frac{(1+x_i)}{2} f_2(\mathbf{x}')$$
$$f^\uparrow(\mathbf{x}) = \frac{(1-x_i)}{2} f_1^\uparrow(\mathbf{x}') + \frac{(1+x_i)}{2} f_2^\uparrow(\mathbf{x}').$$

Note that we have

$$g(\mathbf{y}) = \frac{(1-\rho y_i)}{2} \mathrm{E}(f_1(\mathbf{x}')|\mathbf{Y} = \mathbf{y}') + \frac{(1+\rho y_i)}{2} \mathrm{E}(f_2(\mathbf{x}')|\mathbf{Y} = \mathbf{y}')$$
$$g^\uparrow(\mathbf{y}) = \frac{(1-\rho y_i)}{2} \mathrm{E}(f_1^\uparrow(\mathbf{x}')|\mathbf{Y} = \mathbf{y}') + \frac{(1+\rho y_i)}{2} \mathrm{E}(f_2^\uparrow(\mathbf{x}')|\mathbf{Y} = \mathbf{y}').$$

We will establish that for any $\mathbf{y}'$, the vector

$$\begin{bmatrix} g^\uparrow(\mathbf{y}', y_i = -1) \\ g^\uparrow(\mathbf{y}', y_i = 1) \end{bmatrix} = \begin{bmatrix} \frac{(1+\rho)}{2} \mathrm{E}(f_1^\uparrow(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') + \frac{(1-\rho)}{2} \mathrm{E}(f_2^\uparrow(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') \\ \frac{(1-\rho)}{2} \mathrm{E}(f_1^\uparrow(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') + \frac{(1+\rho)}{2} \mathrm{E}(f_2^\uparrow(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') \end{bmatrix}$$

majorizes the vector

$$\begin{bmatrix} g(\mathbf{y}', y_i = -1) \\ g(\mathbf{y}', y_i = 1) \end{bmatrix} = \begin{bmatrix} \frac{(1+\rho)}{2}\,\mathrm{E}(f_1(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') + \frac{(1-\rho)}{2}\,\mathrm{E}(f_2(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') \\ \frac{(1-\rho)}{2}\,\mathrm{E}(f_1(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') + \frac{(1+\rho)}{2}\,\mathrm{E}(f_2(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') \end{bmatrix}.$$

Towards this consider the vectors

$$\begin{bmatrix} \frac{(1+\rho)}{2} f_1^{\uparrow}(\mathbf{x}') + \frac{(1-\rho)}{2} f_2^{\uparrow}(\mathbf{x}') \\ \frac{(1-\rho)}{2} f_1^{\uparrow}(\mathbf{x}') + \frac{(1+\rho)}{2} f_2^{\uparrow}(\mathbf{x}') \end{bmatrix} \text{ and } \begin{bmatrix} \frac{(1+\rho)}{2} f_1(\mathbf{x}') + \frac{(1-\rho)}{2} f_2(\mathbf{x}') \\ \frac{(1-\rho)}{2} f_1(\mathbf{x}') + \frac{(1+\rho)}{2} f_2(\mathbf{x}'). \end{bmatrix}$$

Since $\rho \in (0, 1)$ and $f_1^{\uparrow}(\mathbf{x}') \leq f_2^{\uparrow}(\mathbf{x}')$ by construction, note that

$$\frac{(1+\rho)}{2} f_1^{\uparrow}(\mathbf{x}') + \frac{(1-\rho)}{2} f_2^{\uparrow}(\mathbf{x}') \leq \frac{(1-\rho)}{2} f_1^{\uparrow}(\mathbf{x}') + \frac{(1+\rho)}{2} f_2^{\uparrow}(\mathbf{x}').$$

Further, since $f_1^{\uparrow}(\mathbf{x}') = \min\{f_1(\mathbf{x}'), f_2(\mathbf{x}')\}$ and $f_2^{\uparrow}(\mathbf{x}') = \max\{f_1(\mathbf{x}'), f_2(\mathbf{x}')\}$, and $\rho \in (0, 1)$ we have

$$\frac{(1+\rho)}{2} f_1^{\uparrow}(\mathbf{x}') + \frac{(1-\rho)}{2} f_2^{\uparrow}(\mathbf{x}') = \min\left\{\frac{(1+\rho)}{2} f_1(\mathbf{x}') + \frac{(1-\rho)}{2} f_2(\mathbf{x}'), \frac{(1-\rho)}{2} f_1(\mathbf{x}') + \frac{(1+\rho)}{2} f_2(\mathbf{x}')\right\}.$$

Taking conditional expectations at $\mathbf{y}'$ we have

$$\begin{aligned} g^{\uparrow}(\mathbf{y}', y_i = -1) &= \frac{(1+\rho)}{2}\,\mathrm{E}(f_1^{\uparrow}(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') + \frac{(1-\rho)}{2}\,\mathrm{E}(f_2^{\uparrow}(\mathbf{X}')|\mathbf{Y} = \mathbf{y}') \\ &= \mathrm{E}\left(\min\left\{\frac{(1+\rho)}{2} f_1(\mathbf{X}') + \frac{(1-\rho)}{2} f_2(\mathbf{X}'), \frac{(1-\rho)}{2} f_1(\mathbf{X}') + \frac{(1+\rho)}{2} f_2(\mathbf{X}')\right\} |\mathbf{Y}' = \mathbf{y}'\right) \\ &\leq \min\left\{E\left(\frac{(1+\rho)}{2} f_1(\mathbf{X}') + \frac{(1-\rho)}{2} f_2(\mathbf{X}')|\mathbf{Y}' = \mathbf{y}'\right),\right. \\ &\qquad\qquad \left. E\left(\frac{(1-\rho)}{2} f_1(\mathbf{X}') + \frac{(1+\rho)}{2} f_2(\mathbf{X}')|\mathbf{Y}' = \mathbf{y}'\right)\right\} \\ &= \min\{g(\mathbf{y}', y_i = -1), g(\mathbf{y}', y_i = +1)\}. \end{aligned}$$

Since $p(\mathbf{x}'|\mathbf{y}') > 0$ for every pair $(\mathbf{x}', \mathbf{y}')$, equality in the above holds *if and only if* for every $\mathbf{x}'$, $f_1(\mathbf{x}') \leq f_2(\mathbf{x}')$ or for every $\mathbf{x}'$, $f_1(\mathbf{x}') \leq f_2(\mathbf{x}')$. Since $f_1^{\uparrow}(\mathbf{x}') + f_2^{\uparrow}(\mathbf{x}') = f_1(\mathbf{x}') + f_2(\mathbf{x}')$ for every $\mathbf{x}'$ it is clear that $g^{\uparrow}(\mathbf{y}', y_i = -1) + g^{\uparrow}(\mathbf{y}', y_i = 1) = g(\mathbf{y}', y_i = -1) + g(\mathbf{y}', y_i = +1)$. Thus we have shown the majorization between the vectors. From Karamata's inequality concerning majorization and concave functions, the proposition immediately follows. □

By iterating over singleton sets (or coordinates) we can restrict our search for maximizers to monotone functions.

Proposition 3. *For any $(i, j) \in \{1, .., n\}$ and for any monotone boolean function $f(\mathbf{X})$, define $f^{\uparrow} = \mathcal{L}_{\{i,j\}}(f)$. Then for any strictly concave function ϕ we have*

$$\mathrm{E}(\phi(g^{\uparrow}(\mathbf{Y}))) \leq \mathrm{E}(\phi(g(\mathbf{Y}))),$$

where $g(\mathbf{Y}) = \mathrm{E}(f(\mathbf{X})|\mathbf{Y}) = (T_\rho f)(\mathbf{Y})$, $g^{\uparrow}(\mathbf{Y}) = \mathrm{E}(f^{\uparrow}(\mathbf{X})|\mathbf{Y}) = (T_\rho f^{\uparrow})(\mathbf{Y})$, with equality holding if and only if $f_2(\mathbf{x}') \leq f_3(\mathbf{x}')\ \forall \mathbf{x}'$ or $f_3(\mathbf{x}') \geq f_2(\mathbf{x}')\ \forall \mathbf{x}'$.

Proof. W.l.o.g we assume $i < j$. The proof basically mimics the proof of Proposition 2, in that we will establish that for any $\mathbf{y}'$, the vector

$$\begin{bmatrix} g^{\uparrow}(\mathbf{y}', y_i = -1, y_j = -1) \\ g^{\uparrow}(\mathbf{y}', y_i = -1, y_j = 1) \\ g^{\uparrow}(\mathbf{y}', y_i = 1, y_j = -1) \\ g^{\uparrow}(\mathbf{y}', y_i = 1, y_j = 1) \end{bmatrix} = \mathrm{E}\left(\begin{bmatrix} \frac{(1+\rho)^2}{4} f_1^{\uparrow}(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_2^{\uparrow}(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_3^{\uparrow}(\mathbf{X}') + \frac{(1+\rho)^2}{4} f_4^{\uparrow}(\mathbf{X}') \\ \frac{(1-\rho^2)}{4} f_1^{\uparrow}(\mathbf{X}') + \frac{(1+\rho)^2}{4} f_2^{\uparrow}(\mathbf{X}') + \frac{(1-\rho)^2}{4} f_3^{\uparrow}(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_4^{\uparrow}(\mathbf{X}') \\ \frac{(1-\rho^2)}{4} f_1^{\uparrow}(\mathbf{X}') + \frac{(1-\rho)^2}{4} f_2^{\uparrow}(\mathbf{X}') + \frac{(1+\rho)^2}{4} f_3^{\uparrow}(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_4^{\uparrow}(\mathbf{X}') \\ \frac{(1-\rho)^2}{4} f_1^{\uparrow}(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_2^{\uparrow}(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_3^{\uparrow}(\mathbf{X}') + \frac{(1-\rho)^2}{4} f_4^{\uparrow}(\mathbf{X}') \end{bmatrix} \middle| \mathbf{Y}' = \mathbf{y}'\right)$$

majorizes the vector

$$\begin{bmatrix} g(\mathbf{y}', y_i=-1, y_j=-1) \\ g(\mathbf{y}', y_i=-1, y_j=1) \\ g(\mathbf{y}', y_i=1, y_j=-1) \\ g(\mathbf{y}', y_i=1, y_j=1) \end{bmatrix} = \mathrm{E}\left(\begin{bmatrix} \frac{(1+\rho)^2}{4} f_1(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_2(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_3(\mathbf{X}') + \frac{(1+\rho)^2}{4} f_4(\mathbf{X}') \\ \frac{(1-\rho^2)}{4} f_1(\mathbf{X}') + \frac{(1+\rho)^2}{4} f_2(\mathbf{X}') + \frac{(1-\rho)^2}{4} f_3(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_4(\mathbf{X}') \\ \frac{(1-\rho^2)}{4} f_1(\mathbf{X}') + \frac{(1-\rho)^2}{4} f_2(\mathbf{X}') + \frac{(1+\rho)^2}{4} f_3(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_4(\mathbf{X}') \\ \frac{(1-\rho)^2}{4} f_1(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_2(\mathbf{X}') + \frac{(1-\rho^2)}{4} f_3(\mathbf{X}') + \frac{(1-\rho)^2}{4} f_4(\mathbf{X}') \end{bmatrix} \middle| \mathbf{Y}' = \mathbf{y}' \right)$$

Since the function $f(\mathbf{X})$ is monotone we have $f_1(\mathbf{x}') \leq \min\{f_2(\mathbf{x}'), f_3(\mathbf{x}')\} \leq f_4(\mathbf{x}')$. Hence $f_1(\mathbf{x}') = f_1^{\uparrow}(\mathbf{x}'), f_4(\mathbf{x}') = f_4^{\uparrow}(\mathbf{x}'), f_2^{\uparrow}(\mathbf{x}') = \min\{f_2(\mathbf{x}'), f_3(\mathbf{x}')\}, f_3^{\uparrow}(\mathbf{x}') = \max\{f_2(\mathbf{x}'), f_3(\mathbf{x}')\}$. This immediately implies that $g^{\uparrow}(\mathbf{y}', y_i = -1, y_j = -1) = g(\mathbf{y}', y_i = -1, y_j = -1)$ and $g^{\uparrow}(\mathbf{y}', y_i = 1, y_j = 1) = g(\mathbf{y}', y_i = 1, y_j = 1)$. In an argument identical to that of Proposition 2, we obtain that

$$g^{\uparrow}(\mathbf{y}', y_i=-1, y_j=1) \leq \min\{g(\mathbf{y}', y_i=-1, y_j=1), g(\mathbf{y}', y_i=1, y_j=-1)\}$$

with equality holding *if and only if* $f_2(\mathbf{x}') \leq f_3(\mathbf{x}')\ \forall \mathbf{x}'$ or $f_3(\mathbf{x}') \geq f_2(\mathbf{x}')\ \forall \mathbf{x}'$. Clearly we also have $g^{\uparrow}(\mathbf{y}', y_i = -1, y_j = 1) + g^{\uparrow}(\mathbf{y}', y_i = 1, y_j = -1) = g(\mathbf{y}', y_i = -1, y_j = 1) + g(\mathbf{y}', y_i = 1, y_j = -1)$. This completes the proof and observe that the equality condition means that $f^{\uparrow}(\mathbf{x}) = f(\mathbf{x})$ or $f^{\uparrow}(\mathbf{x}) = f_{i\leftrightarrow j}(\mathbf{x})$. □

Remark 5. The above two propositions were also established in [] for $JS(x)$ which also helped them numerically verify their conjecture till $n = 7$.

III. Weaker conjecture relating pairs of Boolean functions

In this section I will document the weaker version of the conjectures that relate pairs of Boolean functions $f(\mathbf{X})$ and $g(\mathbf{Y})$.

Proposition 4 ([7])**.** *For any pair of Boolean functions $f(\mathbf{X})$ and $g(\mathbf{Y})$ that takes values $\{-1, +1\}$*

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

Conjecture 3. *For any pair of Boolean functions that takes values $\{-1, +1\}$*

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

Proposition 5. *Suppose a pair of Boolean functions $f(\mathbf{X}), g(\mathbf{Y})$ taking values $\{-1, +1\}$ satisfies*

$$\sqrt{1 - \hat{f}_{\Phi}^2} - \mathrm{E}\left(\sqrt{1 - (\mathrm{E}(f(\mathbf{X})|g(\mathbf{Y})))^2}\right) \leq 1 - \sqrt{1-\rho^2} \tag{5}$$

then the function $f(\mathbf{X})$ also satisfies

$$I(f(\mathbf{X}); g(\mathbf{Y})) \leq 1 - H\left(\frac{1-\rho}{2}\right). \tag{6}$$

Proof. The proof mimics the earlier similar proof. Let $\Psi(x) = \frac{1-\sqrt{1-x^2}}{2}$. Lemma 1 implies that $H(\Psi(x))$ is non-negative, increasing, and convex in x. Observe that

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

The inequality in (6) can be re-expressed as

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

By the convexity of $H(\Psi(x))$ the above inequality will be implied if the following inequality holds:

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

However this inequality is an immediate consequence of (5) and using the fact that $H(\Psi(x))$ is non-negative, increasing, and convex in x. □

A. Evidence in support of Conjecture 3

In this section we will present an inequality involving three variables that, if true, would imply Conjecture 3. The idea is similar to the one in [2] where a similar approach was attempted to resolve Proposition 4 (when it was still a conjecture). The inequality may be of independent interest. As is wont in problems involving Boolean functions, the proposed inequality relies on hypercontractivity parameters between pairs of binary random variables.

Preliminaries involving hypercontractivity: A pair of random variables (U,V) is said to be (p,q)-*hypercontractive* for $1 < q \leq p$ if

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

holds for every non-negative functions $f(U)$ and $g(V)$. Here $p' = \frac{p}{p-1}$ is the Hölder conjugate. For an fixed $p > 1$ define

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

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

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

holds for every non-negative functions $f(U)$ and $g(V)$. Here $p' = \frac{p}{p-1}$ is the Hölder conjugate. For an fixed $p < 1$ define

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

The hypercontractivity parameters (both forward and reverse) satisfy the tensorization and the data-processing properties: (a) if (U^n, V^n) are i.i.d. with each-component distributed as (U,V) then $s_p(U^n;V^n) = s_p(U;V)$; (b) if $U' - U - V - V'$ is a Markov chain, then $s_p(U';V') \leq s_p(U;V)$.

Coming back to the Boolean functions problem, it is known [], [] that $s_p(\mathbf{X};\mathbf{Y}) = s_p(\mathbf{Y};\mathbf{X}) = \rho^2$. Hence for any $f(\mathbf{X})$ and $g(\mathbf{Y})$ we have $s_p(f(\mathbf{X}); g(\mathbf{Y})) \leq \rho^2$. The following conjecture, though of independent interest, was proposed as a way to resolve Proposition 4 (when it was still a conjecture).

Conjecture 4 ([2]). *For any pair of binary random variables (U,V) the following inequality holds:*

$$I(U;V) + \sqrt{1 - s_*(U;V)} \leq 1,$$

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

Remark 6. This conjecture is still formally unestablished, but if true, it would immediately imply Proposition 4 since $s_*(f(\mathbf{X}); g(\mathbf{Y})) \leq \rho^2$, and $f(\mathbf{X}), g(\mathbf{Y})$ are binary-valued. Let $(U,V) \sim p_U(u)p_{V|U}(v|u)$. It has been shown in [1] that

$$s_*(U;V) = \sup_{q_U \ll p_U} \frac{D(p_{V|U} \circ q_U \| p_{V|U} \circ p_U)}{D(q_U \| p_U)},$$

where $p_{V|U} \circ q_U$ (similarly the other term) denotes the induced distribution on V by composing the conditional lay $p_{V|u}$ with the law q_U.

For a pair of binary-valued random variables (U,V), characterize the joint distribution in terms of $(s,c,d) \in [0,1]$ where $\mathrm{P}(U=1) = s, \mathrm{P}(V=0|U=1) = d, \mathrm{P}(V=1|U=0) = c$. As before let $\bar{x} = 1-x$ for a generic variable x. Then clearly

$$s_*(U;V) \geq \frac{D(r\bar{d} + \bar{r}c \| s\bar{d} + \bar{s}c)}{D(r\|s)} =: s_*^{lb}(U;V),$$

where $r = \frac{\bar{s}c\bar{c}}{\bar{s}c\bar{c} + sd\bar{d}}$, and $D(x\|y)$ is the relative entropy between the two distributions $[x, 1-x]$ and $[y, 1-y]$. Numerical simulations show that the explicit three-variable inequality

$$I(U;V) + \sqrt{1 - s_*^{lb}(U;V)} \leq 1, \tag{7}$$

seems to hold. Further one can formally prove the inequality in the neighborhood (by considering expansions) of the equality achieving points; however a complete formal proof for the entire regime is still missing. Note that (7) implies Conjecture 4, yielding very strong evidence in support of its validity. Note that $s_*^{lb}(U;V)$ is just a explicit lower bound for $s_*(U;V)$ that seems to work. □

Unfortunately neither $s_*(U;V)$ nor the techniques in [7] seem to work for the stronger version, Conjecture 3. However the following conjecture appears to be true and as before could be of independent interest.

Conjecture 5. *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 $s_\dagger(U;V) = \lim_{p\to 0} s_p(U;V)$.

Remark 7. Taking $V = f(\mathbf{X})$ and $U = g(\mathbf{Y})$ and by noting that $s_\dagger(g(\mathbf{Y}); f(\mathbf{X})) \leq \rho^2$, the above conjecture immediately implies Conjecture 3. Let $(U,V) \sim p_U(u)p_{V|U}(v|u)$. It has been shown in [3], [6] that

$$1 - s_\dagger(U;V) = \inf_{q_V \neq p_V} \frac{D(q_V \| p_V)}{\inf_{q_{V|U}: q_{V|U}\circ p_U = q_V} \mathrm{E}_U\, D(q_{V|U}\|p_{V|U})},$$

where the outer infimum is taken over all distributions q_V and the inner infimum is taken over all conditional laws $q_{V|U}$ such that $q_{V|U}\circ p_U$ yields q_V.

As before, for a pair of binary-valued random variables (U,V), characterize the joint distribution in terms of $(s,c,d)\in[0,1]$ where $\mathrm{P}(U=1)=s, \mathrm{P}(V=-1|U=1)=d, \mathrm{P}(V=1|U=-1)=c$. Taking $q_V(1) = s\bar{c}+\bar{s}d$, it turns out the the inner infimum (a convex minimization) is obtained by swapping c and d. This leads us to the following inequality:

$$1 - s_\dagger(U;V) \leq \frac{D(s\bar{c}+\bar{s}d\|s\bar{d}+\bar{s}c)}{sD(c\|d)+\bar{s}D(d\|c)} =: 1 - s_\dagger^{lb}(U;V).$$

Numerical simulations indicate that for binary (U,V) the inequality

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

seems to hold. Writing this more explicitly, 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 hold (numerically). We can formally establish this for certain range of parameters (including near equality points) but a complete formal proof is still lacking.

References

[1] Rudolf Ahlswede and Peter Gács, *Spreading of sets in product spaces and hypercontraction of the markov operator*, The Annals of Probability (1976), 925–939.

[2] Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair, *On hypercontractivity and the mutual information between boolean functions.*, 51st Annual Allerton conference on communication, control, and computing, 2013, pp. 13–19.

[3] S. Beigi and C. Nair, *Equivalent characterization of reverse brascamp-lieb-type inequalities using information measures*, 2016 IEEE International Symposium on Information Theory (ISIT), July 2016, pp. 1038–1042.

[4] S. G. Bobkov, *An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in gauss space*, Ann. Probab. **25** (1997), no. 1, 206–214.

[5] Thomas A Courtade and Gowtham R Kumar, *Which boolean functions maximize mutual information on noisy inputs?*, IEEE Transactions on Information Theory **60** (2014), no. 8, 4515–4525.

[6] Sudeep Kamath, *Reverse hypercontractivity using information measures*, Proceedings of the 53rd Annual Allerton Conference on Communications, Control and Computing, 2015.

[7] Georg Pichler, Pablo Piantanida, and Gerald Matz, *Two dictator functions maximize mutual information*, CoRR **abs/1604.02109** (2016).