Noise Stability: Old and New

Lei Yu

Nankai University

Presented at Institute for Advanced Study in Mathematics,
Harbin Institute of Technology
July 13, 2022

Outline

Outline

Definitions

- Let P_{XY} be a joint probability measure on $\mathcal{X} \times \mathcal{Y}$
- Let $P_{XY}^{\otimes n}$ be the n-product of P_{XY} (which is a joint probability measure on $\mathcal{X}^n \times \mathcal{Y}^n$)
  - In other words, $(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^{\otimes n}$ consists of n i.i.d. copies of $(X, Y) \sim P_{XY}$

Definitions

- Let P_{XY} be a joint probability measure on $\mathcal{X} \times \mathcal{Y}$
- Let $P_{XY}^{\otimes n}$ be the n-product of P_{XY} (which is a joint probability measure on $\mathcal{X}^n \times \mathcal{Y}^n$)
  - In other words, $(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^{\otimes n}$ consists of n i.i.d. copies of $(X, Y) \sim P_{XY}$
- The noise stability of a pair of measurable sets (A, B) with $A \subseteq \mathcal{X}^n, B \subseteq \mathcal{Y}^n$ is defined as the joint probability $P_{XY}^{\otimes n}(A \times B)$.
  - Noise stability of a pair of sets is a measure of the resistance of this pair of sets to noise corruption

Definitions

- Let P_{XY} be a joint probability measure on $\mathcal{X} \times \mathcal{Y}$
- Let $P_{XY}^{\otimes n}$ be the n-product of P_{XY} (which is a joint probability measure on $\mathcal{X}^n \times \mathcal{Y}^n$)
  - In other words, $(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^{\otimes n}$ consists of n i.i.d. copies of $(X, Y) \sim P_{XY}$
- The noise stability of a pair of measurable sets (A, B) with $A \subseteq \mathcal{X}^n, B \subseteq \mathcal{Y}^n$ is defined as the joint probability $P_{XY}^{\otimes n}(A \times B)$.
  - Noise stability of a pair of sets is a measure of the resistance of this pair of sets to noise corruption
- The noise stability problem: Estimate $P_{XY}^{\otimes n}(A \times B)$ given $P_X^{\otimes n}(A)$ and $P_Y^{\otimes n}(B)$
  - Geometrically, estimate the “area” given the “length” and “width”

Doubly Symmetric Binary Distribution

Throughout this talk, we mainly consider the doubly symmetric binary distribution (unless otherwise specified)

$$P_{XY} = \begin{array}{ccc} X\backslash Y & 0 & 1 \\ 0 & \frac{1+\rho}{4} & \frac{1-\rho}{4} \\ 1 & \frac{1-\rho}{4} & \frac{1+\rho}{4} \end{array}$$

with correlation $\rho \in (0, 1)$

Maximal/minimal noise stability

- Formally, for $a, b \in [0,1]$, define the maximal noise stability as

$$\overline{\Gamma}^{(n)}(a,b) := \max_{\substack{A,B\subseteq\{0,1\}^n : P_X^{\otimes n}(A)\le a,\\ P_Y^{\otimes n}(B)\le b}} P_{XY}^{\otimes n}(A\times B)$$

Maximal/minimal noise stability

- Formally, for $a, b \in [0,1]$, define the maximal noise stability as

$$\overline{\Gamma}^{(n)}(a,b) := \max_{\substack{A,B \subseteq \{0,1\}^n : P_X^{\otimes n}(A) \le a, \\ P_Y^{\otimes n}(B) \le b}} P_{XY}^{\otimes n}(A \times B)$$

- Define the minimal noise stability as

$$\underline{\Gamma}^{(n)}(a,b) := \min_{\substack{A,B \subseteq \{0,1\}^n : P_X^{\otimes n}(A) \ge a, \\ P_Y^{\otimes n}(B) \ge b}} P_{XY}^{\otimes n}(A \times B)$$

Maximal/minimal noise stability

- Formally, for $a, b \in [0, 1]$, define the maximal noise stability as

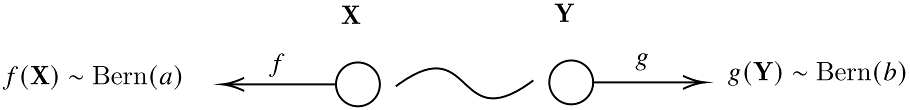
$$\overline{\Gamma}^{(n)}(a, b) := \max_{\substack{A,B\subseteq\{0,1\}^n: P_X^{\otimes n}(A)\leq a,\\ P_Y^{\otimes n}(B)\leq b}} P_{XY}^{\otimes n}(A \times B)$$

- Define the minimal noise stability as

$$\underline{\Gamma}^{(n)}(a, b) := \min_{\substack{A,B\subseteq\{0,1\}^n: P_X^{\otimes n}(A)\geq a,\\ P_Y^{\otimes n}(B)\geq b}} P_{XY}^{\otimes n}(A \times B)$$

- For dyadic rationals $a = \frac{M}{2^n}, b = \frac{N}{2^n}$ (with integers M, N), the "inequalities" in the constraints can be replaced by "equalities"
  - Moreover, for this case, $\overline{\Gamma}^{(n)}(1 - a, b) = b - \underline{\Gamma}^{(n)}(a, b)$

Interpretation: Noninteractive Correlation Distillation



$\max \mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}))$ or equivalently, $\max \mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}) = 1)$

Background

- The notion of “noise stability” was first introduced explicitly by Benjamini, Kalai, and Schramm in 1999
- However, such a quantity was in fact first studied by Witsenhausen in his classic work in 1975, which played a key role in proving a converse result for Gács–Körner common information problem.
- Noise stability was also used by Kahn, Kalai, and Linial in 1988 to prove the famous KKL theorem
- Now, noise stability is one of central topics in analysis of Boolean functions

Existing Results

- Trivial cases: a or b is 0 or 1
- Known nontrivial cases: $\overline{\Gamma}^{(n)}\left(\frac{1}{2},\frac{1}{2}\right), \overline{\Gamma}^{(n)}\left(\frac{1}{4},\frac{1}{4}\right)$ and $\underline{\Gamma}^{(n)}\left(\frac{1}{2},\frac{1}{2}\right), \underline{\Gamma}^{(n)}\left(\frac{1}{4},\frac{3}{4}\right)$
- $\overline{\Gamma}^{(n)}(a,b)$ and $\underline{\Gamma}^{(n)}(a,b)$ for other (a,b)?—unknown and difficult!

Limiting Cases

- Central Limit (CL) regime: a, b are fixed
  - $\overline{\Gamma}^{(\infty)}(a,b), \underline{\Gamma}^{(\infty)}(a,b)$ denote the limits of $\overline{\Gamma}^{(n)}(a,b), \underline{\Gamma}^{(n)}(a,b)$ as $n \to \infty$.

Limiting Cases

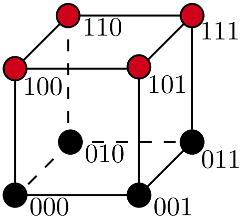
- Central Limit (CL) regime: a, b are fixed
  - $\overline{\Gamma}^{(\infty)}(a,b), \underline{\Gamma}^{(\infty)}(a,b)$ denote the limits of $\overline{\Gamma}^{(n)}(a,b), \underline{\Gamma}^{(n)}(a,b)$ as $n \to \infty$.
- Large Deviations (LD) regime: For $a = 2^{-n\alpha}, b = 2^{-n\beta}$ (with fixed $\alpha, \beta > 0$), denote

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) := -\frac{1}{n}\log\overline{\Gamma}^{(n)}\left(e^{-n\alpha}, e^{-n\beta}\right),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) := -\frac{1}{n}\log\underline{\Gamma}^{(n)}\left(e^{-n\alpha}, e^{-n\beta}\right),$$

and their limits as $\underline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta), \overline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta)$

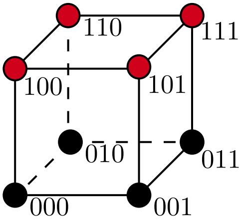
Outline

Hamming Subcubes



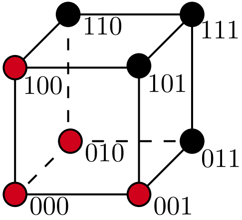
- An $(n-k)$-subcube C_{n-k} is a set of $\mathbf{x}$ with k components fixed (e.g., $\{\mathbf{1}_k\} \times \{0,1\}^{n-k}$)

Hamming Subcubes



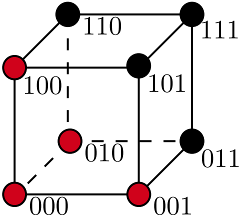
- An $(n-k)$-subcube C_{n-k} is a set of $\mathbf{x}$ with k components fixed (e.g., $\{\mathbf{1}_k\} \times \{0,1\}^{n-k}$)
  - Special case C_{n-1}: e.g., $\{1\} \times \{0,1\}^{n-1}$ (Indicator $\mathbf{x} \mapsto x_1$ called a dictator function)
- Case of $a = b = 2^{-k}$:
  - $A = B = C_{n-k}$ (identical) $\Longrightarrow P_{XY}^{\otimes n}(A \times B) = P_{XY}(1,1)^k = \left(\frac{1+\rho}{4}\right)^k$
  - $A = \mathbf{1} - B = C_{n-k}$ (anti-symmetric) $\Longrightarrow P_{XY}^{\otimes n}(A \times B) = P_{XY}(1,0)^k = \left(\frac{1-\rho}{4}\right)^k$

Hamming Balls



- Hamming Ball: For $r \in [0, n]$, $\mathbb{B}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) \leq r\} = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i \leq r\right\}$

Hamming Balls



- Hamming Ball: For $r \in [0, n]$, $\mathbb{B}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) \le r\} = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i \le r\right\}$
- CL regime: Choose $A = \mathbb{B}_{r_n}(\mathbf{0}), B = \mathbb{B}_{s_n}(\mathbf{0})$ with $r_n = \frac{n}{2} + \frac{\lambda\sqrt{n}}{2}, s_n = \frac{n}{2} + \frac{\mu\sqrt{n}}{2}$ where $\lambda, \mu \in \mathbb{R}$

Hamming Balls

- By the univariate and multivariate CL theorems,

$$P_X^{\otimes n}(A) \to \Phi(\lambda), \qquad P_Y^{\otimes n}(B) \to \Phi(\mu), \qquad P_{XY}^{\otimes n}(A \times B) \to \Phi_\rho(\lambda, \mu)$$

where Φ is the CDF of the standard Gaussian, and $\Phi_\rho(\cdot, \cdot)$ is the CDF of the zero-mean bivariate Gaussian with covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

Hamming Balls

- By the univariate and multivariate CL theorems,

$$P_X^{\otimes n}(A) \to \Phi(\lambda), \qquad P_Y^{\otimes n}(B) \to \Phi(\mu), \qquad P_{XY}^{\otimes n}(A \times B) \to \Phi_\rho(\lambda, \mu)$$

where Φ is the CDF of the standard Gaussian, and $\Phi_\rho(\cdot,\cdot)$ is the CDF of the zero-mean bivariate Gaussian with covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

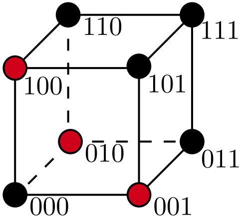
- Achievable CL probabilities:

$$\overline{\Gamma}^{(\infty)}(a,b) \ge \Lambda_\rho(a,b) \ \text{(by concentric balls } \mathbb{B}_{r_n}(\mathbf{0}), \mathbb{B}_{s_n}(\mathbf{0})\text{)}$$
$$\underline{\Gamma}^{(\infty)}(a,b) \le \Lambda_{-\rho}(a,b) \ \text{(by anti-concentric balls } \mathbb{B}_{r_n}(\mathbf{0}), \mathbb{B}_{s_n}(\mathbf{1})\text{)}$$

where bivariate normal copula (or Gaussian quadrant probability function):

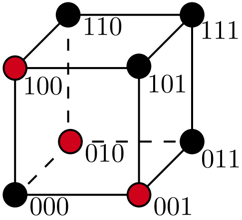
$$\Lambda_\rho(a,b) := \Phi_\rho\left(\Phi^{-1}(a), \Phi^{-1}(b)\right)$$

Hamming Spheres



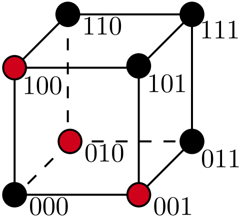
- Hamming Sphere: For $r \in [0:n]$, $\mathbb{S}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) = r\} = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i = r\right\}$

Hamming Spheres



- Hamming Sphere: For $r \in [0:n]$, $\mathbb{S}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) = r\} = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i = r\right\}$
  - a type class with type $(\bar{\lambda}, \lambda)$ in Hamming space, where $\lambda := \frac{r}{n}$ and $\bar{\lambda} := 1 - \lambda$

Hamming Spheres



- Hamming Sphere: For $r \in [0:n]$, $\mathbb{S}_r(\mathbf{0}) := \{\mathbf{x} : d_{\mathrm{H}}(\mathbf{x}, \mathbf{0}) = r\} = \left\{\mathbf{x} : \sum_{i=1}^{n} x_i = r\right\}$
  - a type class with type $(\bar{\lambda}, \lambda)$ in Hamming space, where $\lambda := \frac{r}{n}$ and $\bar{\lambda} := 1 - \lambda$
- LD regime: Choose $A = \mathbb{S}_{r_n}(\mathbf{0})$, $B = \mathbb{S}_{s_n}(\mathbf{0})$ with $r_n = \lambda n, s_n = \mu n$ where $\lambda, \mu \in [0, 1]$

Hamming Spheres

- By LD theory (or Sanov's theorem),

$$-\frac{1}{n}\log P_X^{\otimes n}(A) \to D\left(\left(\bar{\lambda},\lambda\right)\|P_X\right) = 1 - H_2(\lambda)$$
$$-\frac{1}{n}\log P_Y^{\otimes n}(B) \to D\left((\bar{\mu},\mu)\|P_Y\right) = 1 - H_2(\mu)$$
$$-\frac{1}{n}\log P_{XY}^{\otimes n}(A\times B) \to \mathbb{D}\left(\left(\bar{\lambda},\lambda\right),(\bar{\mu},\mu)\|P_{XY}\right)$$

Hamming Spheres

- By LD theory (or Sanov's theorem),

$$-\frac{1}{n}\log P_X^{\otimes n}(A) \to D\left(\left(\bar{\lambda},\lambda\right)\|P_X\right) = 1 - H_2(\lambda)$$
$$-\frac{1}{n}\log P_Y^{\otimes n}(B) \to D\left(\left(\bar{\mu},\mu\right)\|P_Y\right) = 1 - H_2(\mu)$$
$$-\frac{1}{n}\log P_{XY}^{\otimes n}(A\times B) \to \mathbb{D}\left(\left(\bar{\lambda},\lambda\right),\left(\bar{\mu},\mu\right)\|P_{XY}\right)$$

  - relative entropy: $D(Q\|P) := \sum_x Q(x)\log\frac{Q(x)}{P(x)}$
  - binary entropy function: $H_2 : t \in [0,1] \mapsto -t\log_2 t - (1-t)\log_2(1-t)$
  - minimum-relative-entropy over couplings of (Q_X, Q_Y):

$$\mathbb{D}(Q_X, Q_Y\|P_{XY}) := \min_{Q_{XY}\in C(Q_X,Q_Y)} D(Q_{XY}\|P_{XY})$$

with $C(Q_X, Q_Y) := \{Q_{XY} \text{ with marginals } Q_X, Q_Y\}$ denoting the coupling set

Hamming Spheres

- Optimizing $\mathbb{D}(Q_X, Q_Y \| P_{XY})$ over feasible $Q_X := (\lambda, \bar{\lambda}), Q_Y := (\mu, \bar{\mu}) \Longrightarrow$

$$\underline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) \le \underline{\Theta}_{\mathrm{LD}}(\alpha,\beta) := \min_{\substack{Q_X, Q_Y : D(Q_X\|P_X)\ge\alpha,\\ D(Q_Y\|P_Y)\ge\beta}} \mathbb{D}(Q_X, Q_Y \| P_{XY}),$$

$$\overline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) \ge \overline{\Theta}_{\mathrm{LD}}(\alpha,\beta) := \max_{\substack{Q_X, Q_Y : D(Q_X\|P_X)\le\alpha,\\ D(Q_Y\|P_Y)\le\beta}} \mathbb{D}(Q_X, Q_Y \| P_{XY}).$$

  - attained by concentric and anti-concentric spheres
  - or respectively attained by concentric and anti-concentric balls.

Hamming Spheres

- Optimizing $\mathbb{D}\left(Q_X, Q_Y \| P_{XY}\right)$ over feasible $Q_X := \left(\lambda, \bar{\lambda}\right), Q_Y := (\mu, \bar{\mu}) \Longrightarrow$

$$\underline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) \le \underline{\Theta}_{\mathrm{LD}}(\alpha,\beta) := \min_{\substack{Q_X, Q_Y : D(Q_X \| P_X) \ge \alpha, \\ D(Q_Y \| P_Y) \ge \beta}} \mathbb{D}\left(Q_X, Q_Y \| P_{XY}\right),$$

$$\overline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) \ge \overline{\Theta}_{\mathrm{LD}}(\alpha,\beta) := \max_{\substack{Q_X, Q_Y : D(Q_X \| P_X) \le \alpha, \\ D(Q_Y \| P_Y) \le \beta}} \mathbb{D}\left(Q_X, Q_Y \| P_{XY}\right).$$

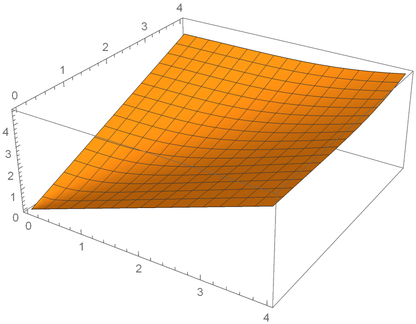
  - attained by concentric and anti-concentric spheres
  - or respectively attained by concentric and anti-concentric balls.

Conjecture (Ordentlich–Polyanskiy–Shayevitz (OPS, 2019))

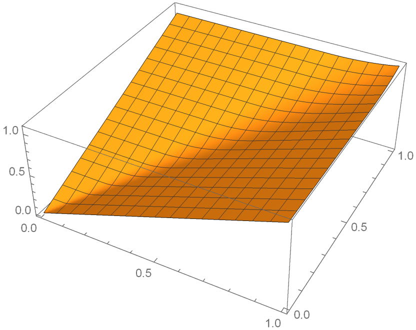
For $\alpha, \beta \in (0,1)$,

$$\underline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) = \underline{\Theta}_{\mathrm{LD}}(\alpha,\beta), \qquad \overline{\Theta}_{\mathrm{LD}}^{(\infty)}(\alpha,\beta) = \overline{\Theta}_{\mathrm{LD}}(\alpha,\beta).$$

Exponents induced by balls/spheres for $\rho = 0.9$

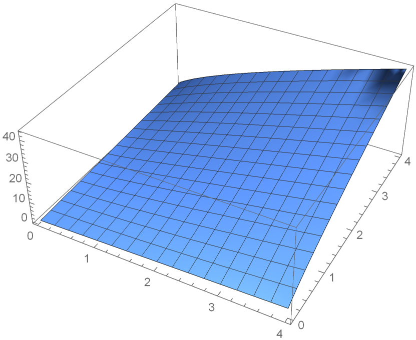


$\underline{\Theta}_{\mathrm{CL}}(\alpha, \beta) := -\log \Lambda_\rho\left(e^{-\alpha}, e^{-\beta}\right)$

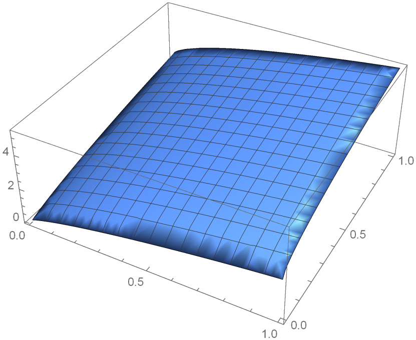


$\underline{\Theta}_{\mathrm{LD}}(\alpha, \beta)$

Exponents induced by balls/spheres for $\rho = 0.9$



$\overline{\Theta}_{\mathrm{CL}}(\alpha,\beta) := -\log \Lambda_{-\rho}\left(e^{-\alpha}, e^{-\beta}\right)$



$\overline{\Theta}_{\mathrm{LD}}(\alpha,\beta)$

Comparison: Subcubes vs. Balls/Spheres

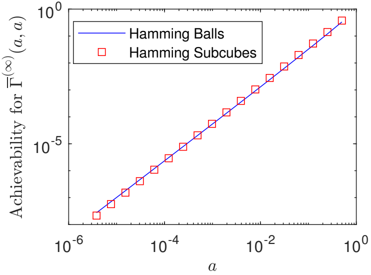
Regime	Central Limit		Large Deviations
a, b	fixed and large a, b	fixed but small a, b	exp. vanishing a, b
Subcubes	Better		
Balls		Better	Better

- For large a, b, subcubes are better; for small a, b, balls are better.

Comparison: Subcubes vs. Balls/Spheres

Regime	Central Limit		Large Deviations
a, b	fixed and large a, b	fixed but small a, b	exp. vanishing a, b
Subcubes	Better		
Balls		Better	Better

- For large a, b, subcubes are better; for small a, b, balls are better.



Natural Questions on Optimality

Questions: Hamming subcubes optimal for large a, b? Hamming balls optimal for small a, b?

Natural Questions on Optimality

Questions: Hamming subcubes optimal for large a, b? Hamming balls optimal for small a, b?

- CL regime: a, b are fixed
  - Are subcubes optimal for $a = b \in \left\{\frac{1}{2}, \frac{1}{4}\right\}$? —— including Mossel's mean-1/4 stability problem

Natural Questions on Optimality

Questions: Hamming subcubes optimal for large a, b? Hamming balls optimal for small a, b?

- CL regime: a, b are fixed
  - Are subcubes optimal for $a = b \in \left\{\frac{1}{2}, \frac{1}{4}\right\}$? —— including Mossel's mean-1/4 stability problem
- LD regime: a, b are exponentially small
  - Are balls/spheres optimal? —— Ordentlich–Polyanskiy–Shayevitz's conjecture

Natural Questions on Optimality

Questions: Hamming subcubes optimal for large a, b? Hamming balls optimal for small a, b?

- CL regime: a, b are fixed
  - Are subcubes optimal for $a = b \in \{\frac{1}{2}, \frac{1}{4}\}$? —— including Mossel's mean-1/4 stability problem
- LD regime: a, b are exponentially small
  - Are balls/spheres optimal? —— Ordentlich–Polyanskiy–Shayevitz's conjecture

- We answer these questions in the following.

Outline

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)),$$

  - $\rho(U;V) := \frac{\mathrm{cov}(U,V)}{\sqrt{\mathrm{var}(U)\mathrm{var}(V)}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)),$$

  - $\rho(U;V) := \frac{\mathrm{cov}(U,V)}{\sqrt{\mathrm{var}(U)\mathrm{var}(V)}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances
- Tensorization:

$$\rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) = \rho_{\mathrm{m}}(X;Y).$$

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X); g(Y)),$$

  - $\rho(U;V) := \frac{\mathrm{cov}(U,V)}{\sqrt{\mathrm{var}(U)\mathrm{var}(V)}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances
- Tensorization:

$$\rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) = \rho_{\mathrm{m}}(X;Y).$$

- Data Processing Inequality (DPI): For a Markov chain $U - X - Y - V$,

$$\rho_{\mathrm{m}}(U;V) \le \rho_{\mathrm{m}}(X;Y).$$

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

- Confirmed positively by Witsenhausen (1975) using maximal correlation
- The (Hirschfeld–Gebelein–Rényi) maximal correlation

$$\rho_{\mathrm{m}}(X;Y) := \sup_{f,g} \rho(f(X);g(Y)),$$

  - $\rho(U;V) := \frac{\mathrm{cov}(U,V)}{\sqrt{\mathrm{var}(U)\mathrm{var}(V)}}$ is the Pearson correlation coefficient
  - the supremum is taken over all real-valued functions with finite variances
- Tensorization:

$$\rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) = \rho_{\mathrm{m}}(X;Y).$$

- Data Processing Inequality (DPI): For a Markov chain $U - X - Y - V$,

$$\rho_{\mathrm{m}}(U;V) \le \rho_{\mathrm{m}}(X;Y).$$

- For binary X, Y, $\rho_{\mathrm{m}}(X;Y) = |\rho(X;Y)|$.

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

Theorem ([Witsenhausen, 1975])

For any A, B *with* $P_X^{\otimes n}(A) = a, P_Y^{\otimes n}(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \leq P_{XY}^{\otimes n}(A \times B) \leq ab + \rho\sqrt{a\bar{a}b\bar{b}}, \qquad \text{where } \bar{x} = 1 - x.$$

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

Theorem ([Witsenhausen, 1975])

For any A, B with $P_X^{\otimes n}(A) = a, P_Y^{\otimes n}(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \leq P_{XY}^{\otimes n}(A \times B) \leq ab + \rho\sqrt{a\bar{a}b\bar{b}}, \qquad where\ \bar{x} = 1 - x.$$

- Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y}) \implies U - \mathbf{X} - \mathbf{Y} - V \implies$

$$\frac{\left|P_{XY}^{\otimes n}(A \times B) - ab\right|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} = |\rho(U;V)| = \rho_{\mathrm{m}}(U;V) \overset{\mathrm{DPI}}{\leq} \rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) \overset{\mathrm{Tensorization}}{=} \rho_{\mathrm{m}}(X;Y)$$

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

Theorem ([Witsenhausen, 1975])

For any A, B *with* $P_X^{\otimes n}(A) = a, P_Y^{\otimes n}(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^{\otimes n}(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}, \qquad \text{where } \bar{x} = 1 - x.$$

- Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y}) \Longrightarrow U - \mathbf{X} - \mathbf{Y} - V \Longrightarrow$

$$\frac{\left|P_{XY}^{\otimes n}(A \times B) - ab\right|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} = |\rho(U;V)| = \rho_{\mathrm{m}}(U;V) \overset{\mathrm{DPI}}{\le} \rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) \overset{\mathrm{Tensorization}}{=} \rho_{\mathrm{m}}(X;Y)$$

- Important Consequence:
  - For $a = b = 1/2$, $\frac{1-\rho}{4} \le P_{XY}^{\otimes n}(A \times B) \le \frac{1+\rho}{4}$.

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

Theorem ([Witsenhausen, 1975])

For any A, B *with* $P_X^{\otimes n}(A) = a, P_Y^{\otimes n}(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^{\otimes n}(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}, \qquad \textit{where } \bar{x} = 1 - x.$$

- Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y}) \Longrightarrow U - \mathbf{X} - \mathbf{Y} - V \Longrightarrow$

$$\frac{\left|P_{XY}^{\otimes n}(A \times B) - ab\right|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} = |\rho(U;V)| = \rho_{\mathrm{m}}(U;V) \overset{\text{DPI}}{\le} \rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) \overset{\text{Tensorization}}{=} \rho_{\mathrm{m}}(X;Y)$$

- Important Consequence:
  - For $a = b = 1/2$, $\frac{1-\rho}{4} \le P_{XY}^{\otimes n}(A \times B) \le \frac{1+\rho}{4}$.
  - Dictators (subcubes) are optimal for $a = b = 1/2$, i.e., $\overline{\Gamma}^{(n)}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1+\rho}{4}$ and $\underline{\Gamma}^{(n)}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1-\rho}{4}$

Optimality for $a = b = \frac{1}{2}$: subcubes (dictators) optimal?

Theorem ([Witsenhausen, 1975])

For any A, B *with* $P_X^{\otimes n}(A) = a, P_Y^{\otimes n}(B) = b$,

$$ab - \rho\sqrt{a\bar{a}b\bar{b}} \le P_{XY}^{\otimes n}(A \times B) \le ab + \rho\sqrt{a\bar{a}b\bar{b}}, \qquad \textit{where } \bar{x} = 1 - x.$$

- Proof: Setting $U = 1_A(\mathbf{X}), V = 1_B(\mathbf{Y}) \Longrightarrow U - \mathbf{X} - \mathbf{Y} - V \Longrightarrow$

$$\frac{\left|P_{XY}^{\otimes n}(A \times B) - ab\right|}{\sqrt{a\bar{a}}\sqrt{b\bar{b}}} = |\rho(U;V)| = \rho_{\mathrm{m}}(U;V) \overset{\text{DPI}}{\le} \rho_{\mathrm{m}}(\mathbf{X};\mathbf{Y}) \overset{\text{Tensorization}}{=} \rho_{\mathrm{m}}(X;Y)$$

- Important Consequence:
  - For $a = b = 1/2$, $\frac{1-\rho}{4} \le P_{XY}^{\otimes n}(A \times B) \le \frac{1+\rho}{4}$.
  - Dictators (subcubes) are optimal for $a = b = 1/2$, i.e., $\overline{\Gamma}^{(n)}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1+\rho}{4}$ and $\underline{\Gamma}^{(n)}\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1-\rho}{4}$
  - Note: This point also can be proven by hypercontractivity and Fourier analysis

Optimality for $a = b = \frac{1}{4}$——Mossel's mean-1/4 stability problem

- Confirmed positively by Yu and Tan (2021) using Fourier analysis

Optimality for $a = b = \frac{1}{4}$——Mossel's mean-1/4 stability problem

- Confirmed positively by Yu and Tan (2021) using Fourier analysis
- Fourier coefficients of $f : \{0,1\}^n \to \{0,1\}$ are

$$\hat{f}(\mathbf{y}) := \frac{1}{2^n} \sum_{\mathbf{x}} f(\mathbf{x}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

Optimality for $a = b = \frac{1}{4}$——Mossel's mean-1/4 stability problem

- Confirmed positively by Yu and Tan (2021) using Fourier analysis
- Fourier coefficients of $f : \{0,1\}^n \to \{0,1\}$ are

$$\hat{f}(\mathbf{y}) := \frac{1}{2^n} \sum_{\mathbf{x}} f(\mathbf{x}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

- Fourier expansion of f is

$$f(\mathbf{x}) = \sum_{\mathbf{y}} \hat{f}(\mathbf{y}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

Optimality for $a = b = \frac{1}{4}$—Mossel's mean-1/4 stability problem

- Confirmed positively by Yu and Tan (2021) using Fourier analysis
- Fourier coefficients of $f : \{0,1\}^n \to \{0,1\}$ are

$$\hat{f}(\mathbf{y}) := \frac{1}{2^n} \sum_{\mathbf{x}} f(\mathbf{x}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

- Fourier expansion of f is

$$f(\mathbf{x}) = \sum_{\mathbf{y}} \hat{f}(\mathbf{y}) (-1)^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

- Define the k-degree Fourier weight as

$$\mathbf{W}_k(f) := \sum_{|\mathbf{y}|=k} \hat{f}(\mathbf{y})^2$$

where $|\mathbf{y}|$ denotes the Hamming weight of $\mathbf{y}$.

Optimality for $a = b = \frac{1}{4}$: subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0(f) = a^2, \qquad \sum_{k=0}^{n} \mathbf{W}_k(f) = a, \qquad \mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1) = \sum_{k=0}^{n} \mathbf{W}_k(f)\,\rho^k.$$

Optimality for $a = b = \frac{1}{4}$: subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0(f) = a^2, \qquad \sum_{k=0}^{n} \mathbf{W}_k(f) = a, \qquad \mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1) = \sum_{k=0}^{n} \mathbf{W}_k(f)\,\rho^k.$$

- LP bound on $\mathbf{W}_1$ [Fu et al., 2001, Yu and Tan, 2019]:

$$\mathbf{W}_1(f) \le \varphi(a) := \begin{cases} 2a\left(\sqrt{a} - a\right) & 0 \le a \le \frac{1}{4} \\ \frac{a}{2} & \frac{1}{4} < a \le \frac{1}{2} \end{cases}$$

Optimality for $a = b = \frac{1}{4}$: subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0(f) = a^2, \qquad \sum_{k=0}^{n} \mathbf{W}_k(f) = a, \qquad \mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1) = \sum_{k=0}^{n} \mathbf{W}_k(f)\,\rho^k.$$

- LP bound on $\mathbf{W}_1$ [Fu et al., 2001, Yu and Tan, 2019]:

$$\mathbf{W}_1(f) \le \varphi(a) := \begin{cases} 2a\left(\sqrt{a} - a\right) & 0 \le a \le \frac{1}{4} \\ \frac{a}{2} & \frac{1}{4} < a \le \frac{1}{2} \end{cases}$$

- Fact:

$$\mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}) = 1) \le \max\{\mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1), \mathbb{P}(g(\mathbf{X}) = g(\mathbf{Y}) = 1)\}$$

Optimality for $a = b = \frac{1}{4}$: subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0(f) = a^2, \qquad \sum_{k=0}^{n} \mathbf{W}_k(f) = a, \qquad \mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1) = \sum_{k=0}^{n} \mathbf{W}_k(f)\rho^k.$$

- LP bound on $\mathbf{W}_1$ [Fu et al., 2001, Yu and Tan, 2019]:

$$\mathbf{W}_1(f) \le \varphi(a) := \begin{cases} 2a\left(\sqrt{a} - a\right) & 0 \le a \le \frac{1}{4} \\ \frac{a}{2} & \frac{1}{4} < a \le \frac{1}{2} \end{cases}$$

- Fact:

$$\mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}) = 1) \le \max\{\mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1), \mathbb{P}(g(\mathbf{X}) = g(\mathbf{Y}) = 1)\}$$

Theorem ([Yu and Tan, 2019, Yu and Tan, 2021])

$$\overline{\Gamma}^{(n)}(a,a) \le a^2 + \rho\varphi(a) + \rho^2\left(a - a^2 - \varphi(a)\right).$$

Optimality for $a = b = \frac{1}{4}$: subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0(f) = a^2, \qquad \sum_{k=0}^{n} \mathbf{W}_k(f) = a, \qquad \mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1) = \sum_{k=0}^{n} \mathbf{W}_k(f)\,\rho^k.$$

- LP bound on $\mathbf{W}_1$ [Fu et al., 2001, Yu and Tan, 2019]:

$$\mathbf{W}_1(f) \le \varphi(a) := \begin{cases} 2a\left(\sqrt{a} - a\right) & 0 \le a \le \frac{1}{4} \\ \frac{a}{2} & \frac{1}{4} < a \le \frac{1}{2} \end{cases}$$

- Fact:
$$\mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}) = 1) \le \max\{\mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1), \mathbb{P}(g(\mathbf{X}) = g(\mathbf{Y}) = 1)\}$$

Theorem ([Yu and Tan, 2019, Yu and Tan, 2021])

$$\overline{\Gamma}^{(n)}(a, a) \le a^2 + \rho\varphi(a) + \rho^2\left(a - a^2 - \varphi(a)\right).$$

- Consequence: For $a = 1/4$, the upper bound reduces to $\left(\frac{1+\rho}{4}\right)^2 \implies$ $\overline{\Gamma}^{(n)}\left(\frac{1}{4}, \frac{1}{4}\right) = \left(\frac{1+\rho}{4}\right)^2$ for $n \ge 2$, attained by subcubes

Optimality for $a = b = \frac{1}{4}$: subcubes optimal?

- Properties: For a Boolean f with mean a,

$$\mathbf{W}_0(f) = a^2, \qquad \sum_{k=0}^{n} \mathbf{W}_k(f) = a, \qquad \mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1) = \sum_{k=0}^{n} \mathbf{W}_k(f)\rho^k.$$

- LP bound on $\mathbf{W}_1$ [Fu et al., 2001, Yu and Tan, 2019]:

$$\mathbf{W}_1(f) \le \varphi(a) := \begin{cases} 2a\left(\sqrt{a} - a\right) & 0 \le a \le \frac{1}{4} \\ \frac{a}{2} & \frac{1}{4} < a \le \frac{1}{2} \end{cases}$$

- Fact:
$$\mathbb{P}(f(\mathbf{X}) = g(\mathbf{Y}) = 1) \le \max\{\mathbb{P}(f(\mathbf{X}) = f(\mathbf{Y}) = 1), \mathbb{P}(g(\mathbf{X}) = g(\mathbf{Y}) = 1)\}$$

Theorem ([Yu and Tan, 2019, Yu and Tan, 2021])

$$\overline{\Gamma}^{(n)}(a, a) \le a^2 + \rho\varphi(a) + \rho^2\left(a - a^2 - \varphi(a)\right).$$

- Consequence: For $a = 1/4$, the upper bound reduces to $\left(\frac{1+\rho}{4}\right)^2 \implies$
$\overline{\Gamma}^{(n)}\left(\frac{1}{4}, \frac{1}{4}\right) = \left(\frac{1+\rho}{4}\right)^2$ for $n \ge 2$, attained by subcubes
- However, $\underline{\Gamma}^{(n)}\left(\frac{1}{4}, \frac{1}{4}\right)$ is still open!

Another bound better for small a, b

Theorem (Small-Set Expansion [Ahlswede and Gács, 1976, Kahn et al., 1988, Mossel et al., 2006, O'Donnell, 2014])

For any $n \geq 1$ and $\alpha, \beta > 0$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \geq \underline{\theta}(\alpha,\beta),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \leq \overline{\theta}(\alpha,\beta),$$

where

$$\underline{\theta}(\alpha,\beta) = \begin{cases} \frac{\alpha+\beta-2\rho\sqrt{\alpha\beta}}{1-\rho^2}, & \rho^2\alpha \leq \beta \leq \frac{\alpha}{\rho^2}, \\ \alpha, & \beta < \rho^2\alpha, \\ \beta, & \alpha < \rho^2\beta \end{cases},$$

$$\overline{\theta}(\alpha,\beta) = \frac{\alpha+\beta+2\rho\sqrt{\alpha\beta}}{1-\rho^2}.$$

Proof

- Proof of SSE Theorem: Based on the classic hypercontractivity inequalities [Bonami, 1968, Gross, 1975, Borell, 1982]:

Proof

- Proof of SSE Theorem: Based on the classic hypercontractivity inequalities [Bonami, 1968, Gross, 1975, Borell, 1982]:
  - Forward part: For $p, q \geq 1$, s.t. $(p-1)(q-1) \geq \rho^2$,

$$\langle f, g \rangle \leq \|f\|_p \|g\|_q, \qquad \forall f : \mathcal{X} \to \mathbb{R}_{\geq 0}, g : \mathcal{Y} \to \mathbb{R}_{\geq 0}$$

    where $\langle f, g \rangle := \mathbb{E}[f(\mathbf{X}) g(\mathbf{Y})]$, $\|f\|_p := (\mathbb{E}[f^p(\mathbf{X})])^{1/p}$, and $\|g\|_q := (\mathbb{E}[g^q(\mathbf{Y})])^{1/q}$.

Proof

- Proof of SSE Theorem: Based on the classic hypercontractivity inequalities [Bonami, 1968, Gross, 1975, Borell, 1982]:
  - Forward part: For $p, q \geq 1$, s.t. $(p-1)(q-1) \geq \rho^2$,

$$\langle f, g \rangle \leq \|f\|_p \|g\|_q , \qquad \forall f : \mathcal{X} \to \mathbb{R}_{\geq 0}, g : \mathcal{Y} \to \mathbb{R}_{\geq 0}$$

    where $\langle f, g \rangle := \mathbb{E}\left[f(\mathbf{X})\, g(\mathbf{Y})\right]$, $\|f\|_p := \left(\mathbb{E}\left[f^p(\mathbf{X})\right]\right)^{1/p}$, and $\|g\|_q := \left(\mathbb{E}\left[g^q(\mathbf{Y})\right]\right)^{1/q}$.
  - Substituting $f \leftarrow 1_A, g \leftarrow 1_B$ and optimizing p, q, SSE theorem follows.

An even better one

Theorem (Strong SSE [Yu et al., 2021, Yu, 2021c, Yu, 2021a])

For any $n \geq 1$ *and* $\alpha, \beta \in (0, 1]$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \geq \underline{\Theta}_{\mathrm{LD}}(\alpha, \beta),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \leq \overline{\Theta}_{\mathrm{LD}}(\alpha, \beta).$$

An even better one

Theorem (Strong SSE [Yu et al., 2021, Yu, 2021c, Yu, 2021a])

For any $n \geq 1$ and $\alpha, \beta \in (0, 1]$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \geq \underline{\Theta}_{\mathrm{LD}}(\alpha, \beta),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \leq \overline{\Theta}_{\mathrm{LD}}(\alpha, \beta).$$

- stronger than SSE theorem (can recover the latter by letting $\alpha, \beta \to 0$)

An even better one

Theorem (Strong SSE [Yu et al., 2021, Yu, 2021c, Yu, 2021a])

For any $n \geq 1$ *and* $\alpha, \beta \in (0, 1]$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \geq \underline{\Theta}_{\mathrm{LD}}(\alpha,\beta),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \leq \overline{\Theta}_{\mathrm{LD}}(\alpha,\beta).$$

- stronger than SSE theorem (can recover the latter by letting $\alpha, \beta \to 0$)
- The bounds in the strong SSE theorem are asymptotically tight as $n \to \infty$.

OPS's conjecture is true: Balls/spheres are optimal in LD regime!

An even better one

Theorem (Strong SSE [Yu et al., 2021, Yu, 2021c, Yu, 2021a])

For any $n \geq 1$ and $\alpha, \beta \in (0, 1]$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \geq \underline{\Theta}_{\mathrm{LD}}(\alpha,\beta),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \leq \overline{\Theta}_{\mathrm{LD}}(\alpha,\beta).$$

- stronger than SSE theorem (can recover the latter by letting $\alpha, \beta \to 0$)
- The bounds in the strong SSE theorem are asymptotically tight as $n \to \infty$.

OPS's conjecture is true: Balls/spheres are optimal in LD regime!

- Notice: OPS's conjecture was proven previously for
  - limiting cases as $\rho \to 0$ or 1 in [Ordentlich et al., 2020]
  - the case $\alpha = \beta$ in [Kirshner and Samorodnitsky, 2021]

Summary of bounds

<table>
<tr><th rowspan="2">Bounds</th><th colspan="2">Central Limit</th><th>Large Deviations</th></tr>
<tr><th>fixed and large a, b</th><th>fixed but small a, b</th><th>exp. vanishing a, b</th></tr>
<tr><td>Maximal Correlation</td><td>Sharp for $a = b = 1/2$ (Subcubes)</td><td></td><td></td></tr>
<tr><td>Fourier Analysis</td><td>Sharp for $a = b = 1/2$ and $a = b = 1/4$ (Subcubes)</td><td></td><td></td></tr>
<tr><td>SSE</td><td></td><td rowspan="2">Almost sharp (Balls)</td><td></td></tr>
<tr><td>Strong SSE</td><td></td><td>Sharp (Balls/Spheres)</td></tr>
</table>

Outline

Extension to Gaussian Distributions

- The noise stability problem for Gaussian distributions? —completely solved

Extension to Gaussian Distributions

- The noise stability problem for Gaussian distributions? —completely solved

Theorem (Borell's Isoperimetric Theorem [Borell, 1985, Mossel and Neeman, 2015])

Let $(\mathbf{X}, \mathbf{Y})$ *be a sequence of Gaussian pairs with corr.* $\rho \in (0,1)$*. Then, for any* $a, b \in [0,1]$*,*

$$\overline{\Gamma}^{(n)}(a,b) = \Lambda_{\rho}(a,b)$$
$$\underline{\Gamma}^{(n)}(a,b) = \Lambda_{-\rho}(a,b).$$

Extension to Gaussian Distributions

- The noise stability problem for Gaussian distributions? —completely solved

Theorem (Borell's Isoperimetric Theorem [Borell, 1985, Mossel and Neeman, 2015])

Let $(\mathbf{X}, \mathbf{Y})$ *be a sequence of Gaussian pairs with corr.* $\rho \in (0, 1)$. *Then, for any* $a, b \in [0, 1]$,

$$\overline{\Gamma}^{(n)}(a, b) = \Lambda_{\rho}(a, b)$$
$$\underline{\Gamma}^{(n)}(a, b) = \Lambda_{-\rho}(a, b).$$

- Recall: Bivariate normal copula:

$$\Lambda_{\rho}(a, b) := \Phi_{\rho}\left(\Phi^{-1}(a), \Phi^{-1}(b)\right)$$

Extension to Gaussian Distributions

- The noise stability problem for Gaussian distributions? —completely solved

Theorem (Borell's Isoperimetric Theorem [Borell, 1985, Mossel and Neeman, 2015])

Let $(\mathbf{X},\mathbf{Y})$ *be a sequence of Gaussian pairs with corr.* $\rho \in (0,1)$*. Then, for any* $a,b \in [0,1]$*,*

$$\overline{\Gamma}^{(n)}(a,b) = \Lambda_{\rho}(a,b)$$
$$\underline{\Gamma}^{(n)}(a,b) = \Lambda_{-\rho}(a,b).$$

- Recall: Bivariate normal copula:

$$\Lambda_{\rho}(a,b) := \Phi_{\rho}\left(\Phi^{-1}(a),\Phi^{-1}(b)\right)$$

- Parallel half-spaces are optimal (e.g., $A=\{x_1\le r\}, B=\{y_1\le s\}$)

Extension to General Distributions

- Strong SSE is still true for distributions defined on Polish spaces.

Extension to General Distributions

- Strong SSE is still true for distributions defined on Polish spaces.

Theorem (Strong SSE [Yu, 2021c])

Let P_{XY} be a joint distribution defined on a Polish space. For any $n \geq 1$ and $\alpha, \beta > 0$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \geq \mathfrak{K}\left[\underline{\Theta}_{\mathrm{LD}}\right](\alpha, \beta),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \leq \mathfrak{C}\left[\overline{\varphi}_{\mathrm{LD}}\right](\alpha, \beta),$$

where $\mathfrak{K}[f], \mathfrak{C}[f]$ respectively denote the lower convex and upper concave envelopes of a function f, and

$$\overline{\varphi}(s,t) := \sup_{Q_X, Q_Y : D(Q_X \| P_X) = s, D(Q_Y \| P_Y) = t} \mathbb{D}(Q_X, Q_Y \| P_{XY}).$$

Extension to General Distributions

- Strong SSE is still true for distributions defined on Polish spaces.

Theorem (Strong SSE [Yu, 2021c])

Let P_{XY} be a joint distribution defined on a Polish space. For any $n \geq 1$ and $\alpha, \beta > 0$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \geq \mathfrak{K}\left[\underline{\Theta}_{\mathrm{LD}}\right](\alpha,\beta),$$
$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha,\beta) \leq \mathfrak{C}\left[\overline{\varphi}_{\mathrm{LD}}\right](\alpha,\beta),$$

where $\mathfrak{K}[f], \mathfrak{C}[f]$ respectively denote the lower convex and upper concave envelopes of a function f, and

$$\overline{\varphi}(s,t) := \sup_{Q_X,Q_Y: D(Q_X\|P_X)=s, D(Q_Y\|P_Y)=t} \mathbb{D}(Q_X,Q_Y\|P_{XY}).$$

- Tightness: The bounds above are asymptotically sharp as $n \to \infty$.

Extension to General Distributions

- Strong SSE is still true for distributions defined on Polish spaces.

Theorem (Strong SSE [Yu, 2021c])

Let P_{XY} be a joint distribution defined on a Polish space. For any $n \geq 1$ and $\alpha, \beta > 0$,

$$\underline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \geq \mathfrak{K}\left[\underline{\Theta}_{\mathrm{LD}}\right](\alpha, \beta),$$

$$\overline{\Theta}_{\mathrm{LD}}^{(n)}(\alpha, \beta) \leq \mathfrak{C}\left[\overline{\varphi}_{\mathrm{LD}}\right](\alpha, \beta),$$

where $\mathfrak{K}[f], \mathfrak{C}[f]$ respectively denote the lower convex and upper concave envelopes of a function f, and

$$\overline{\varphi}(s,t) := \sup_{Q_X, Q_Y : D(Q_X\|P_X)=s, D(Q_Y\|P_Y)=t} \mathbb{D}(Q_X, Q_Y \| P_{XY}).$$

- Tightness: The bounds above are asymptotically sharp as $n \to \infty$.
- For the doubly symmetric binary distribution, $\overline{\Theta}_{\mathrm{LD}} = \overline{\varphi}_{\mathrm{LD}}$.

Outline

Hypercontractivity (HC) Inequalities

- Forward HC: For any nonnegative f, g,

$$\langle f, g \rangle \leq \|f\|_p \, \|g\|_q$$

where $p, q \geq 1$ s.t. $(p-1)(q-1) \geq \rho^2$

Hypercontractivity (HC) Inequalities

- Forward HC: For any nonnegative f, g,

$$\langle f, g\rangle \leq \|f\|_p \|g\|_q$$

  where $p, q \geq 1$ s.t. $(p-1)(q-1) \geq \rho^2$

  - strengthening the Hölder inequality for which $p, q \geq 1$ s.t. $(p-1)(q-1) \geq 1$

Hypercontractivity (HC) Inequalities

- Forward HC: For any nonnegative f, g,

$$\langle f, g \rangle \leq \|f\|_p \, \|g\|_q$$

where $p, q \geq 1$ s.t. $(p-1)(q-1) \geq \rho^2$
  - strengthening the Hölder inequality for which $p, q \geq 1$ s.t. $(p-1)(q-1) \geq 1$

- Reverse HC: For any nonnegative f, g,

$$\langle f, g \rangle \geq \|f\|_p \, \|g\|_q$$

where $p, q \leq 1$ s.t. $(p-1)(q-1) \geq \rho^2$

Hypercontractivity (HC) Inequalities

- Forward HC: For any nonnegative f, g,

$$\langle f, g\rangle \leq \|f\|_p \|g\|_q$$

where $p, q \geq 1$ s.t. $(p-1)(q-1) \geq \rho^2$

  - strengthening the Hölder inequality for which $p, q \geq 1$ s.t. $(p-1)(q-1) \geq 1$

- Reverse HC: For any nonnegative f, g,

$$\langle f, g\rangle \geq \|f\|_p \|g\|_q$$

where $p, q \leq 1$ s.t. $(p-1)(q-1) \geq \rho^2$

  - strengthening the reverse Hölder inequality for which $p, q \leq 1$ s.t. $(p-1)(q-1) \geq 1$

HC $\Longrightarrow$ SSE

- HC $\Longrightarrow$ SSE: By setting $f \leftarrow 1_A, g \leftarrow 1_B$,

$$P_{XY}^{\otimes n}(A \times B) \le P_X^{\otimes n}(A)^{1/p} P_Y^{\otimes n}(B)^{1/q},$$
$$P_{XY}^{\otimes n}(A \times B) \ge P_X^{\otimes n}(A)^{1/p} P_Y^{\otimes n}(B)^{1/q}.$$

SSE $\Longrightarrow$ HC? —Yes!

SSE $\Longrightarrow$ HC? —Yes!

- The "level-partition" technique [Kirshner and Samorodnitsky, 2021]:

$$f \approx \sum_{i=1}^{M} c_i 1_{A_i}, \qquad g \approx \sum_{j=1}^{N} d_j 1_{B_j}$$

SSE $\Longrightarrow$ HC? —Yes!

- The "level-partition" technique [Kirshner and Samorodnitsky, 2021]:

$$f \approx \sum_{i=1}^{M} c_i 1_{A_i}, \qquad g \approx \sum_{j=1}^{N} d_j 1_{B_j}$$

- Roughly speaking, applying SSE to each pair of $(A_i, B_j) \Longrightarrow$ HC

SSE $\Longrightarrow$ HC? —Yes!

- The "level-partition" technique [Kirshner and Samorodnitsky, 2021]:

$$f \approx \sum_{i=1}^{M} c_i 1_{A_i}, \qquad g \approx \sum_{j=1}^{N} d_j 1_{B_j}$$

- Roughly speaking, applying SSE to each pair of $(A_i, B_j) \Longrightarrow$ HC
- Strong SSE + "level-partition" technique $\Longrightarrow$ Strong HC [Kirshner and Samorodnitsky, 2021, Yu et al., 2021]
    - Strong HC: An improved version of HC inequalities with the support sizes of f, g known

SSE $\Longrightarrow$ HC? —Yes!

- The "level-partition" technique [Kirshner and Samorodnitsky, 2021]:

$$f \approx \sum_{i=1}^{M} c_i 1_{A_i}, \qquad g \approx \sum_{j=1}^{N} d_j 1_{B_j}$$

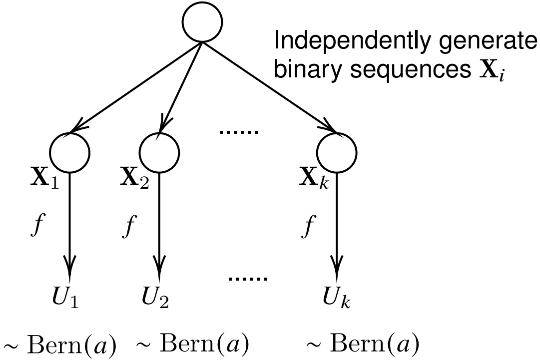
- Roughly speaking, applying SSE to each pair of $(A_i, B_j) \Longrightarrow$ HC
- Strong SSE + "level-partition" technique $\Longrightarrow$ Strong HC [Kirshner and Samorodnitsky, 2021, Yu et al., 2021]
  - Strong HC: An improved version of HC inequalities with the support sizes of f, g known

(Strong) SSE $\Longleftrightarrow$ (Strong) HC (in some sense)!

Outline

NICD with k Users

$$\mathbf{Y} \sim \mathrm{Bern}^{\otimes n}(1/2)$$



Asymmetric Version: $\max \mathbb{P}(U_1 = U_2 = \ldots = U_k = 1)$

Symmetric Version: $\max \mathbb{P}(U_1 = U_2 = \ldots = U_k)$

q-Stability

- (Asymmetric) max q-stability [Li and Médard, 2021]: For $q \in [1, \infty)$,

$$\overline{\Gamma}_q(a) := \max_{A: P_X^{\otimes n}(A)=a} \mathbb{E}_{\mathbf{Y}} \left[P_{X|Y}^{\otimes n}(A|\mathbf{Y})^q \right]$$

q-Stability

- (Asymmetric) max q-stability [Li and Médard, 2021]: For $q \in [1, \infty)$,

$$\overline{\Gamma}_q(a) := \max_{A: P_X^{\otimes n}(A)=a} \mathbb{E}_{\mathbf{Y}} \left[P_{X|Y}^{\otimes n}(A|\mathbf{Y})^q \right]$$

- Symmetric max q-stability: For $q \in [1, \infty)$,

$$\overline{\Gamma}_q^{\mathrm{sym}}(a) := \max_{A: P_X^{\otimes n}(A)=a} \mathbb{E}_{\mathbf{Y}} \left[P_{X|Y}^{\otimes n}(A|\mathbf{Y})^q + P_{X|Y}^{\otimes n}(A^c|\mathbf{Y})^q \right]$$

q-Stability

- (Asymmetric) max q-stability [Li and Médard, 2021]: For $q \in [1, \infty)$,

$$\overline{\Gamma}_q(a) := \max_{A: P_X^{\otimes n}(A)=a} \mathbb{E}_{\mathbf{Y}} \left[P_{X|Y}^{\otimes n}(A|\mathbf{Y})^q \right]$$

- Symmetric max q-stability: For $q \in [1, \infty)$,

$$\overline{\Gamma}_q^{\mathrm{sym}}(a) := \max_{A: P_X^{\otimes n}(A)=a} \mathbb{E}_{\mathbf{Y}} \left[P_{X|Y}^{\otimes n}(A|\mathbf{Y})^q + P_{X|Y}^{\otimes n}(A^c|\mathbf{Y})^q \right]$$

- We only consider $a = 1/2$.

Dictator functions optimal?

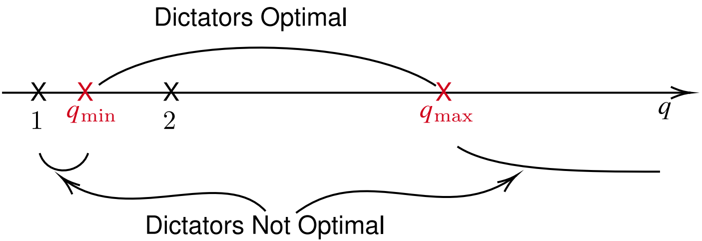
Lemma ([Barnes and Özgür, 2020])

For $a = 1/2$, there are two thresholds $1 \leq q_{\min} < 2 < q_{\max}$ *such that dictators (i.e., $(n-1)$-subcubes) are optimal if and only if $q \in [q_{\min}, q_{\max}]$.*

Dictator functions optimal?

Lemma ([Barnes and Özgür, 2020])

For $a = 1/2$, there are two thresholds $1 \le q_{\min} < 2 < q_{\max}$ such that dictators (i.e., $(n-1)$-subcubes) are optimal if and only if $q \in [q_{\min}, q_{\max}]$.



Dictator functions optimal?

- What are $q_{\min}, q_{\max}$?

Dictator functions optimal?

- What are $q_{\min}, q_{\max}$?
  - Li–Médard conjecture (2019): $q_{\min} = 1$
    - Equivalent to Courtade–Kumar conjecture (2013): Dictator functions maximize $I(f(\mathbf{X}); \mathbf{Y})$ over all balanced Boolean functions? (Derivative of q-stability at $q = 1$ is conditional entropy)

Dictator functions optimal?

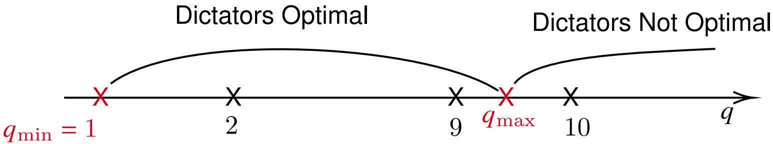
- What are $q_{\min}, q_{\max}$?
  - Li–Médard conjecture (2019): $q_{\min} = 1$
    - Equivalent to Courtade–Kumar conjecture (2013): Dictator functions maximize $I\left(f\left(\mathbf{X}\right);\mathbf{Y}\right)$ over all balanced Boolean functions? (Derivative of q-stability at $q = 1$ is conditional entropy)
  - Mossel-O'Donnell conjecture (2005): $q_{\max} \geq 9$

Dictator functions optimal?

- What are $q_{\min}, q_{\max}$?
  - Li–Médard conjecture (2019): $q_{\min} = 1$
    - Equivalent to Courtade–Kumar conjecture (2013): Dictator functions maximize $I(f(\mathbf{X});\mathbf{Y})$ over all balanced Boolean functions? (Derivative of q-stability at $q = 1$ is conditional entropy)
  - Mossel-O'Donnell conjecture (2005): $q_{\max} \geq 9$
  - It is known that $q_{\max} < 10$ [Mossel and O'Donnell, 2005]

Dictator functions optimal?

- What are $q_{\min}, q_{\max}$?
  - Li–Médard conjecture (2019): $q_{\min} = 1$
    - Equivalent to Courtade–Kumar conjecture (2013): Dictator functions maximize $I(f(\mathbf{X});\mathbf{Y})$ over all balanced Boolean functions? (Derivative of q-stability at $q = 1$ is conditional entropy)
  - Mossel-O'Donnell conjecture (2005): $q_{\max} \geq 9$
  - It is known that $q_{\max} < 10$ [Mossel and O'Donnell, 2005]



Progress on Courtade–Kumar/Li–Médard Conjecture

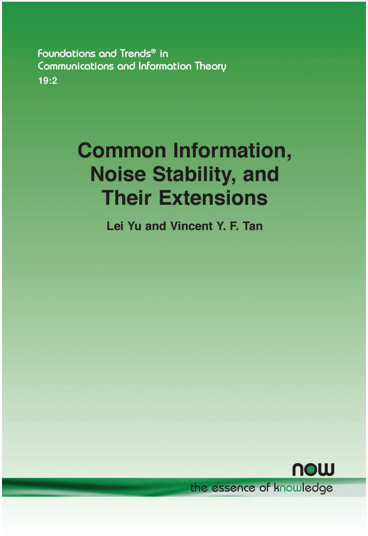
Related Works	Upper Bounds on $\max_{\text{Boolean } f:\mathbb{E}f=1/2} I(f(\mathbf{X});\mathbf{Y})$	Tools
CK/LM Conjecture	$1-H_2\left(\frac{1-\rho}{2}\right)$	
[Witsenhausen and Wyner, 1975]	ρ^2	Mrs. Gerber's lemma (or HC)
[Ordentlich et al., 2016]	$\frac{\log_2 e}{2}\rho^2+9\left(1-\frac{\log_2 e}{2}\right)\rho^4$ for $0\le\rho\le\frac{1}{\sqrt{3}}$ (asymp. tight as $\rho\to 0$)	Fourier analysis + HC
[Samorodnitsky, 2016]	tight bound for $\rho\in[0,\rho_0]$ with some $0<\rho_0<1$	Fourier analysis + Random restrictions + ...
[Yu, 2021b]	tight bound for $\rho\in[0,\rho_1]$ with $\rho_1\approx 0.46$ (explicitly given)	Fourier analysis + KKT conditions
[Pichler et al., 2018]	A weaker version: $\max_{\text{Boolean } f,g} I(f(\mathbf{X});g(\mathbf{Y}))=1-H_2\left(\frac{1-\rho}{2}\right)$	Fourier analysis + Partition technique

Progress on Mossel–O'Donnell Conjecture

Related Works	Dictators are optimal for ...	Tools
Mossel–O'Donnell Conjecture	$\mathbf{2 < q \leq 9}$ (both symmetric and asymmetric max q-stability)	
[Mossel and O'Donnell, 2005]	$2 < q \leq 3$ (symmetric)	Reducing $q = 3$ to $q = 2$
[Witsenhausen, 1975]	$q = 2$ (asymmetric)	Maximal correlation
[Yu, 2021b]	$2 < q \leq 5$ (symmetric); $2 < q \leq 3$ (asymmetric)	Fourier analysis + KKT conditions
[Mossel and O'Donnell, 2005, Li and Médard, 2021]	$\rho \to 0$ or 1 (symmetric and asymmetric)	Fourier analysis

Summary of tools

Methods		Central Limit		Moderate Deviations	Large Deviations
		fixed and large a (and b)	fixed but small a (and b)	subexp. vanishing a (and b)	exp. vanishing a (and b)
Information-Theoretic Methods	Maximal Correlation	Sharp for noise stability with $a = b = 1/2$			
	HC/SSE (stronger than MC)		Almost sharp	Sharp	
	Strong HC/SSE (stronger than HC/SSE)				Sharp
Fourier Analysis	Combined with Optimization Theory (LP or KKT)	Sharp for noise stability with $a = b = 1/2, 1/4$; Sharp for q-stability with $a = 1/2$ and certain (q, ρ)			



Common Information, Noise Stability, and Their Extensions

Lei Yu and Vincent Y. F. Tan

Common information measures the amount of matching variables in two or more information sources. It is ubiquitous in information theory and related areas such as theoretical computer science and discrete probability. However, because there are multiple notions of common information, a unified understanding of the deep interconnections between them is lacking. In this monograph the authors fill this gap by leveraging a small set of mathematical techniques that are applicable across seemingly disparate problems.

The reader is introduced in Part I to the operational tasks and properties associated with the two main measures of common information, namely Wyner's and Gács–Körner–Witsenhausen's (GKW). In the subsequent two Parts, the authors take a deeper look at each of these. In Part II they discuss extensions to Wyner's common information from the perspective of distributed source simulation, including the Rényi common information. In Part III, GKW common information comes under the spotlight. Having laid the groundwork, the authors seamlessly transition to discussing their connections to various conjectures in information theory and discrete probability.

This monograph provides students and researchers in Information Theory with a comprehensive resource for understanding common information and points the way forward to creating a unified set of techniques applicable over a wide range of problems.

This book is originally published as
Foundations and Trends® in Communications and Information Theory
Volume 19 Issue 2, ISSN: 1567-2190.

ISBN 978-1-63828-014-9

References I

Ahlswede, R. and Gács, P. (1976).
Spreading of sets in product spaces and hypercontraction of the markov operator.
The Annals of Probability, pages 925–939.

Barnes, L. P. and Özgür, A. (2020).
The Courtade-Kumar most informative Boolean function conjecture and a symmetrized Li-Médard conjecture are equivalent.
In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2205–2209. IEEE.

Bonami, A. (1968).
Ensembles $\Lambda(p)$ dans le dual de D^∞.
In *Annales de l'institut Fourier*, volume 18, pages 193–204.

Borell, C. (1982).
Positivity improving operators and hypercontractivity.
Mathematische Zeitschrift, 180(3):225–234.

Borell, C. (1985).
Geometric bounds on the Ornstein-Uhlenbeck velocity process.
Probability Theory and Related Fields, 70(1):1–13.

Fu, F.-W., Wei, V. K., and Yeung, R. W. (2001).
On the minimum average distance of binary codes: Linear programming approach.
Discrete Applied Mathematics, 111(3):263–281.

Gross, L. (1975).
Logarithmic sobolev inequalities.
American Journal of Mathematics, 97(4):1061–1083.

References II

Kahn, J., Kalai, G., and Linial, N. (1988).
The influence of variables on Boolean functions.
In *29th Annual Symposium on Foundations of Computer Science*, pages 68–80. IEEE.

Kirshner, N. and Samorodnitsky, A. (2021).
A moment ratio bound for polynomials and some extremal properties of Krawchouk polynomials and Hamming spheres.
IEEE Trans. Inf. Theory, 67(6):3509–3541.

Li, J. and Médard, M. (2021).
Boolean functions: noise stability, non-interactive correlation distillation, and mutual information.
IEEE Trans. Inf. Theory, 67(2):778–789.

Mossel, E. and Neeman, J. (2015).
Robust optimality of Gaussian noise stability.
Journal of the European Mathematical Society, 17(2):433–482.

Mossel, E. and O'Donnell, R. (2005).
Coin flipping from a cosmic source: On error correction of truly random bits.
Random Structures & Algorithms, 26(4):418–436.

Mossel, E., O'Donnell, R., Regev, O., Steif, J. E., and Sudakov, B. (2006).
Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality.
Israel Journal of Mathematics, 154(1):299–336.

References III

O'Donnell, R. (2014).
Analysis of Boolean Functions.
Cambridge University Press.

Ordentlich, O., Polyanskiy, Y., and Shayevitz, O. (2020).
A note on the probability of rectangles for correlated binary strings.
IEEE Trans. Inf. Theory.

Ordentlich, O., Shayevitz, O., and Weinstein, O. (2016).
An improved upper bound for the most informative Boolean function conjecture.
In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 500–504. IEEE.

Pichler, G., Piantanida, P., and Matz, G. (2018).
Dictator functions maximize mutual information.
The Annals of Applied Probability, 28(5):3094–3101.

Samorodnitsky, A. (2016).
On the entropy of a noisy function.
IEEE Trans. Inf. Theory, 62(10):5446–5464.

Witsenhausen, H. and Wyner, A. (1975).
A conditional entropy bound for a pair of discrete random variables.
IEEE Trans. Inf. Theory, 21(5):493–501.

Witsenhausen, H. S. (1975).
On sequences of pairs of dependent random variables.
SIAM Journal on Applied Mathematics, 28(1):100–113.

References IV

Yu, L. (2021a).
The convexity and concavity of envelopes of the minimum-relative-entropy region for the DSBS.
arXiv preprint arXiv:2106.03654.

Yu, L. (2021b).
On the Φ-stability and related conjectures.
arXiv preprint arXiv:2104.08740.

Yu, L. (2021c).
Strong Brascamp–Lieb inequalities.
arXiv preprint arXiv:2102.06935.

Yu, L., Anantharam, V., and Chen, J. (2021).
Graphs of joint types, noninteractive simulation,and stronger hypercontractivity.
arXiv preprint arXiv:2102.00668.

Yu, L. and Tan, V. Y. F. (2019).
An improved linear programming bound on the average distance of a binary code.
arXiv preprint arXiv:1910.09416.

Yu, L. and Tan, V. Y. F. (2021).
On non-interactive simulation of binary random variables.
IEEE Trans. Inf. Theory, 67(4):2528–2538.

Thank you for your attention!