Hypercontractivity and Information Theory

Adway Girish
Information Theory Laboratory

EPFL

EPFL Information Processing Group

June 9, 2023

Outline

Background papers

- Ahlswede-Gács '76

  Rudolf Ahlswede and Peter Gacs. "Spreading of Sets in Product Spaces and Hypercontraction of the Markov Operator". In: *The Annals of Probability* (1976)

- Nair '14

  Chandra Nair. "Equivalent formulations of hypercontractivity using information measures". In: *Proc. International Zurich Seminar on Communications*. 2014

- Anantharam-Gohari-Kamath-Nair '13a

  Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair. "On hypercontractivity and the mutual information between Boolean functions". In: *Proc. Allerton Conference on Communication, Control, and Computing*. 2013

Hölder's inequality for probability measures

$$\mathbb{E}[|f(X)g(Y)|] \leq \|f(X)\|_{p'}\|g(Y)\|_p,\ \ p \geq 1$$

- $\dfrac{1}{p} + \dfrac{1}{p'} = 1 \longleftarrow$ Hölder conjugates
- $\|Z\|_p \triangleq \mathbb{E}[|Z|^p]^{\frac{1}{p}} \longleftarrow$ p-norm of Z

An equivalent formulation

$$\|\mathbb{E}[g(Y) \mid X]\|_p \leq \|g(Y)\|_p, \;\; p \geq 1$$

- $\dfrac{1}{p} + \dfrac{1}{p'} = 1 \longleftarrow$ Hölder conjugates
- $\|Z\|_p \triangleq \mathbb{E}[|Z|^p]^{\frac{1}{p}} \longleftarrow$ p-norm of Z

More generally?

$$\|\mathbb{E}[g(Y) \mid X]\|_p \leq \|g(Y)\|_q?$$

- $\boxed{q \geq p}$ $\longrightarrow$ always true
- $\boxed{q < p}$ $\longrightarrow$ depends on μ_{XY}

Extreme cases: $X \perp\!\!\!\perp Y \implies q \geq 1;$
$X = Y \implies q \geq p.$

Hypercontractivity parameters

Definition

For $1 \leq q \leq p < \infty$, (X, Y) is (p, q)-*hypercontractive* if

$$\|\mathbb{E}[g(Y) \mid X]\|_p \leq \|g(Y)\|_q.$$

Also define, for a given $p \geq 1$,

$$q_p(X; Y) \triangleq \inf\{q : (X, Y) \text{ is } (p, q)\text{-hypercontractive}\},$$
$$r_p(X; Y) \triangleq \frac{q_p(X; Y)}{p},$$
$$r^*(X; Y) \triangleq \inf_{p \geq 1} r_p(X; Y) = \lim_{p \to \infty} r_p(X; Y).$$

Extreme cases: $X \perp\!\!\!\perp Y \implies q_p = 1,\ r_p = \frac{1}{p},\ r^* = 0;$

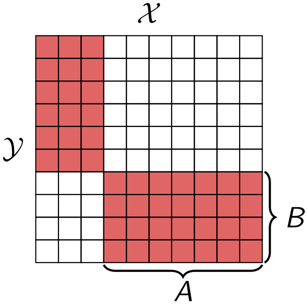
$$X = Y \implies q_p = p,\ r_p = 1,\ r^* = 1.$$

Indecomposable r.v.s

Definition

There do NOT exist nontrivial sets A and B such that

$$\mathbb{P}(X \in A \iff Y \in B) = 1.$$



$\longrightarrow$ decomposable

Probability of decoding sets

Theorem (Ahlswede-Gács '76)

For $\{(X_i, Y_i)\}_{i=1}^n$ i.i.d. and indecomposable, for any sets $A_n \subseteq \mathcal{X}^n$, $B_n \subseteq \mathcal{Y}^n$, there exist positive numbers $r < 1$ and p such that

$$\mathbb{P}(Y^n \in B_n) \geq \mathbb{P}(Y^n \in B_n \mid X^n \in A_n)^p \, \mathbb{P}(X^n \in A_n)^r.$$

- If $\mathbb{P}(Y^n \in B_n \mid X^n \in A_n) \geq \lambda$,

$$-\frac{1}{n} \log \mathbb{P}(Y^n \in B_n) \lesssim -\frac{r}{n} \log \mathbb{P}(X^n \in A_n).$$

Characterizations of r^*...

Theorem (Ahlswede-Gács '76)

$$r^*(X;Y) = \sup_{\nu_X : \substack{\nu_X \neq \mu_X, \\ \nu_X \ll \mu_X}} \frac{D_{\mathrm{KL}}(\nu_Y \,||\, \mu_Y)}{D_{\mathrm{KL}}(\nu_X \,||\, \mu_X)}.$$

Theorem (Anantharam-Gohari-Kamath-Nair '13b[†])

$$r^*(X;Y) = \sup_{\nu_{UX} : \substack{I(U;X) > 0, \\ U - X - Y}} \frac{I(U;Y)}{I(U;X)}.$$

[†] Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair. "On Maximal Correlation, Hypercontractivity, and the Data Processing Inequality studied by Erkip and Cover". In: *CoRR* (2013)

...and r_p

Theorem (Nair '14)

$$r_p(X;Y) = \sup_{\nu_{XY}:\substack{\nu_{XY}\neq\mu_{XY},\\ \nu_{XY}\ll\mu_{XY}}} \frac{D_{\mathrm{KL}}(\nu_Y \,||\, \mu_Y)}{D_{\mathrm{KL}}(\nu_X \,||\, \mu_X) + p\begin{pmatrix} D_{\mathrm{KL}}(\nu_{XY} \,||\, \mu_{XY}) \\ -\ D_{\mathrm{KL}}(\nu_X \,||\, \mu_X) \end{pmatrix}}.$$

Theorem (Nair '14)

$$r_p(X;Y) = \sup_{\nu_{UXY}:\substack{I(U;XY)>0,\\ \sum_{u\in\mathcal{U}} \nu_{UXY}(u,\cdot,\cdot)=\mu_{XY}}} \frac{I(U;Y)}{I(U;X) + p\begin{pmatrix} I(U;XY) \\ -\ I(U;X) \end{pmatrix}}.$$

A conjecture on Boolean functions

Conjecture (Courtade-Kumar '14[†])

For $\{(X_i, Y_i)\}_{i=1}^n$ *i.i.d.* $\mathrm{DSBS}(\alpha)$*, for any Boolean functions* b_1*,* b_2*,*

$$I(b_1(X^n); b_2(Y^n)) \leq 1 - h_2(\alpha).$$

[†] Thomas A. Courtade and Gowtham R. Kumar. "Which Boolean Functions Maximize Mutual Information on Noisy Inputs?" In: *IEEE Transactions on Information Theory* (2014)

A stronger conjecture on r^* of binary r.v.s

Conjecture (Anantharam-Gohari-Kamath-Nair '13a)

For any binary-valued W and Z,

$$I(W; Z) \leq 1 - h_2\left(\frac{1 - \sqrt{r^*(W; Z)}}{2}\right).$$

- $W - X^n - Y^n - Z \implies r^*(W; Z) \leq r^*(X^n; Y^n)$;
- Tensorization: $r_p(X^n; Y^n) = \max_{i=1,\ldots,n} r_p(X_i; Y_i)$;
- $(X_i, Y_i) \sim \mathrm{DSBS}(\alpha) \implies r^*(X_i; Y_i) = (1 - 2\alpha)^2$.

Takeaways

- Connections to information-theoretic quantities
- Applications interesting, but...
- ...more work to be done!

Research proposal

- Computable characterizations of r_p and r^*
- Closed form expressions in particular cases
- More connections and applications to information theory

Thank you!