Two-Moment Inequalities for Rényi Entropy and Mutual Information

Galen Reeves

Department of ECE and Department of Statistical Science
Duke University

ISIT, June 2017

Table of Contents

How can we show mutual information is small?

1. Use $I(X;Y) = H(X) - H(X \,|\, Y)$

2. Jensen's inequality and Rényi divergence:

   $$I(X;Y) = D_1(P_{X,Y} \| P_X P_Y) \leq D_\alpha(P_{X,Y} \| P_X P_Y), \qquad \alpha > 1$$

3. A 'half Jensen' inequality and a 'two-moment' inequality
   Today's talk

Motivation: Conditional CLT for random projections

Consider $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \sqrt{t}\boldsymbol{N}$ where $\boldsymbol{A}$ is IID Gaussian random matrix and $\boldsymbol{N}$ is Gaussian perturbation. Let $G_{\boldsymbol{Y}}$ be Gaussian distribution with same mean and covariance as $\boldsymbol{Y}$.

$$P_{\boldsymbol{Y}} \approx G_{\boldsymbol{Y}} \qquad \text{CLT}$$

$$P_{\boldsymbol{Y}|\boldsymbol{A}}(\cdot \,|\, \boldsymbol{A}) \approx G_{\boldsymbol{Y}} \qquad \text{Conditional CLT}$$

To prove entropic bounds (see Friday's talk), we use

$$\{I(\boldsymbol{A};\boldsymbol{Y}) \approx 0\} \quad \text{and} \quad \text{CLT} \quad \Longleftrightarrow \quad \text{Conditional CLT}$$

Challenges:

- tight bounds on $h(\boldsymbol{Y})$ and $h(\boldsymbol{Y} \,|\, \boldsymbol{A})$ are difficult
- $p(\boldsymbol{Y} \,|\, \boldsymbol{A})/p(\boldsymbol{Y})$ can increase without bound as $t \to 0$

Table of Contents

Jensen and Rényi divergence revisited

For random variables $(X, Y) \sim p(x, y)$,

$$I(X; Y) = \mathbb{E}[\log Z], \qquad Z = \frac{p(X, Y)}{p(X)p(Y)}$$

Jensen's inequality

$$I(X; Y) = \mathbb{E}[\log Z] = \frac{1}{t}\mathbb{E}\left[\log Z^t\right] \leq \frac{1}{t}\log \mathbb{E}\left[Z^t\right]$$

This is the Rényi divergence of order $\alpha = 1 + t$.

This approach is problematic if Z has heavy tails...

A 'half Jensen' inequality [R. 2017]

Start with 'half Jensen' inequality

$$I(X;Y) = \mathbb{E}[\log Z] \leq \mathbb{E}[\log(\mathbb{E}[Z \,|\, Y])]$$

Conditional expectation of Z can be expressed in terms of the variance of conditional density:

$$\mathbb{E}[Z \,|\, Y = y] = \int \left[\frac{p(y \,|\, x)}{p(y)} \right]^2 p(x) \, \mathrm{d}x = 1 + \frac{\mathsf{Var}(p(y \,|\, X))}{p(y)}$$

For every $0 < t < 1$, the inequality $\log(1+u) \leq \frac{1}{t} u^t$ yields,

$$I(X;Y) \leq \frac{1}{t} \int [p(y)]^{1-2t} [\mathsf{Var}(p(y \,|\, X))]^t \, \mathrm{d}y$$

Special cases of the 'half Jensen' inequality

$$I(X;Y) \leq \frac{1}{t} \int [p(y)]^{1-2t} [\mathsf{Var}(p(y \,|\, X))]^t \, \mathrm{d}y$$

The case $t = 1$ gives bound in terms of chi-square divergence

The case $t = 1/2$ gives bound in terms of standard deviation of the conditional density

$$I(X;Y) \leq 2 \int \sqrt{\mathsf{Var}(p(y \,|\, X)} \, \mathrm{d}y$$

The case $0 < t < 1/2$ combined with Hölder's inequality gives bound in terms of variance of the conditional density and Rényi entropy of order $r = (1-2t)/(1-t)$

$$I(X;Y) \leq \frac{1}{t} \left[\exp(h_r(Y)) \int \mathsf{Var}(p(y \,|\, X) \, \mathrm{d}y \right]^t$$

These bounds depend on integrals of fractional powers.

A 'two-moment' inequality [R. 2017]

Proposition: For any numbers $0 < r < 1$ and $p < 1/r - 1 < q$ and non-negative function f defined on $[0, \infty)$, we have

$$\underbrace{\left(\int |f(x)|^r \, \mathrm{d}x \right)^{\frac{1}{r}}}_{\|f\|_r} \leq C \left(\underbrace{\int \|x\|^p f(x) \, \mathrm{d}x}_{p\text{-th moment}} \right)^{\lambda} \left(\underbrace{\int \|x\|^q f(x) \, \mathrm{d}x}_{q\text{-th moment}} \right)^{1-\lambda}$$

where $\lambda = (q + 1 - 1/r)/(q - p)$.

Best possible constant is given by

$$C = \left[\frac{1}{(q-p)} \widetilde{\mathrm{B}} \left(\frac{r\lambda}{1-r}, \frac{r(1-\lambda)}{1-r} \right) \right]^{\frac{1-r}{r}}$$

with $\widetilde{B}(a, b) = \mathrm{B}(a, b)(a + b)^{a+b} a^{-a} b^{-b}$.

A ‘two-moment’ inequality [R. 2017]

Proposition: For any numbers $0 < r < 1$ and $p < 1/r - 1 < q$ and non-negative function f defined on $S \subseteq \mathbb{R}^n$, we have

$$\underbrace{\left(\int |f(x)|^r \, \mathrm{d}x \right)^{\frac{1}{r}}}_{\|f\|_r} \leq C \left(\underbrace{\int \|x\|^{np} f(x) \, \mathrm{d}x}_{np\text{-th moment}} \right)^{\lambda} \left(\underbrace{\int \|x\|^{nq} f(x) \, \mathrm{d}x}_{nq\text{-th moment}} \right)^{1-\lambda}$$

where $\lambda = (q + 1 - 1/r)/(q - p)$.

Best possible constant is given by

$$C = \left[\frac{\mathrm{Vol}(B^n \cap \mathrm{cone}(S))}{(q-p)} \widetilde{\mathrm{B}} \left(\frac{r\lambda}{1-r}, \frac{r(1-\lambda)}{1-r} \right) \right]^{\frac{1-r}{r}}$$

with $\widetilde{B}(a,b) = \mathrm{B}(a,b)(a+b)^{a+b} a^{-a} b^{-b}$.

Remarks on the ‘two-moment’ inequality

The proof is a straightforward consequence of Hölder’s inequality and integral representation of the Beta function.

For $r = 1/2$, Euler’s reflection formula for the Beta function leads to simplified expression

$$C = \frac{\pi \lambda^{-\lambda} (1-\lambda)^{-(1-\lambda)}}{(q-p)\sin(\pi\lambda)}$$

It is possible that variations of these inequalities exist in the literature on weighted L^p-norm inequalities. So far, I have been unable to find an explicit reference.

Table of Contents

Rényi entropy

Let X have density $p(x)$ on $S \subseteq \mathbb{R}^n$. The Rényi entropy of order $r \in (0,1) \cup (1,\infty)$ is defined according to

$$h_r(X) = \frac{1}{1-r} \log\left(\int_S |p(x)|^r \, \mathrm{d}x \right).$$

Properties:

- Decreasing in r
- Limit as $r \to 0$ depends on volume of support
- Limit as $r \to 1$ is Shannon entropy

Upper bound for Rényi entropy [R. 2017]

Proposition: For any numbers $0 < r < 1$ and $p < 1/r - 1 < q$ and density function $p(x)$ on $S \subseteq \mathbb{R}^n$, we have

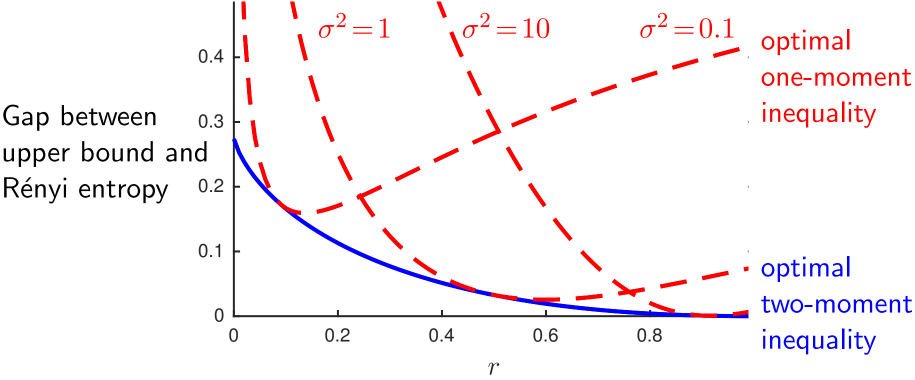
$$h_r(X) \le \log \widetilde{C} + \frac{r\lambda}{1-r} \log \mathbb{E}[\|X\|^{np}] + \frac{r(1-\lambda)}{1-r} \log \mathbb{E}[\|X\|^{nq}]$$

with $\lambda = (q + 1 - 1/r)/(q - p)$.

Relationship to existing results:

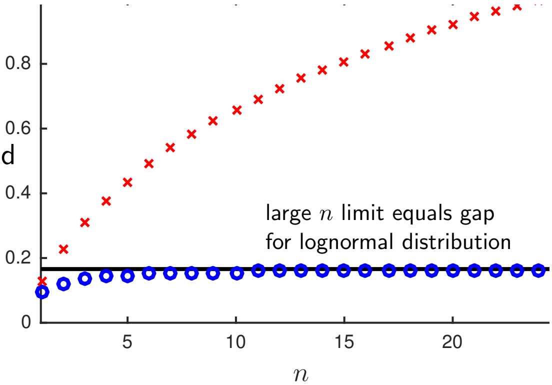
- **One-moment bound:** Evaluating with $p = 0$ recovers entropy–moment inequalities of Costa et al. 2002 and Lutwak et al. 2002. Bound is attained by a maximum entropy distribution.
- **Two-moment bound:** Evaluating with two carefully chosen moments (p, q) can lead to significant improvements.

Example: Lognormal distribution



Example: Multivariate Gaussian distribution

Gap between upper bound and Rényi entropy ($r = 0.2$)



optimal one-moment inequality

optimal two-moment inequality

Table of Contents

Variance of conditional density

Let (X, Y) be a random pair such that the conditional distribution of Y given X has density $p(y\,|\,x)$ on $\mathbb{R}^n$.

Variance of conditional density defined by

$$\mathsf{Var}(f(y\,|\,X)) = \mathbb{E}\Big[\big|p(y\,|\,X) - p(y)\big|^2\Big], \qquad X \sim P_X$$

Define s-th moment of the variance:

$$V_s(Y\,|\,X) = \int \|y\|^s \,\mathsf{Var}(p(y\,|\,X))\,\mathrm{d}y.$$

Note that $V_s(Y|X)$ is nonnegative and equal to zero if and only if X and Y are independent.

A mutual information bound [R. 2017]

Combining the 'half Jensen' and 'two-moment' inequalities yields:

$$I(X;Y) \leq C_\lambda \sqrt{\frac{\omega(S) V_{np}^{\lambda}(Y|X) V_{nq}^{1-\lambda}(Y|X)}{(q-p)}}, \qquad q < 1 < p$$

where $\lambda = (q-1)/(q-p)$ and $\omega(S) = \mathrm{Vol}(B^n \cap \mathrm{cone}(S))$

Useful properties of $V_s(Y \,|\, X)$

The s-th moment of the variance can be expressed as

$$V_s(Y \,|\, X) = \mathbb{E}[K_s(X, X) - K_s(X_1, X_2)],$$

where X_1 and X_2 are independent and

$$K_s(x_1, x_2) = \int \|y\|^s f(y|x_1) f(y|x_2)\, \mathrm{d}y$$

is a p.d. kernel that does not depend on the distribution of X.

If $U \to X \to Y$ forms a Markov chain then

$$V_s(Y \,|\, U) = \mathbb{E}\big[K_s(X_1', X_2') - K_s(X_1, X_2)\big]$$

where X_1' and X_2' are conditionally independent given U

Illustrative example

Consider Markov chain $U \to X \to Y$ given by

$$X \mid U \sim \mathcal{N}(0, U), \qquad Y \mid X \sim \mathcal{N}(X, 1).$$

Then

$$K_s(x_1, x_2) = 2^{-\frac{1+s}{2}} \mathbb{E}\left[\left|\mathcal{N}(0,1) + \frac{x_1 + x_2}{\sqrt{2}}\right|^s\right] \phi\left(\frac{x_1 - x_2}{\sqrt{2}}\right),$$

and

$$V_s(Y|U) = \frac{\Gamma(\frac{1+s}{2})}{2\pi} \mathbb{E}\left[(1+U)^{\frac{s-1}{2}} - \frac{(1+U_1)^{\frac{s}{2}}(1+U_2)^{\frac{s}{2}}}{(1+\frac{1}{2}(U_1+U_2))^{\frac{s+1}{2}}}\right],$$

where (U_1, U_2) are independent copies of U.

Illustrative example

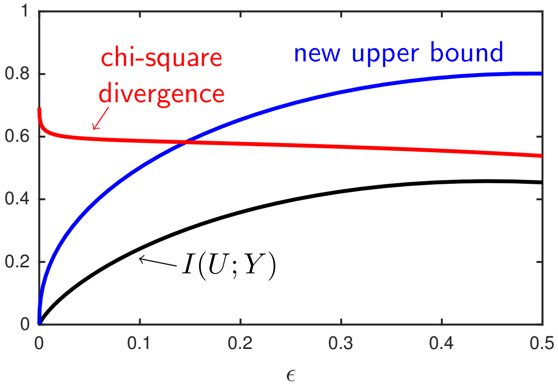


Table of Contents

Conclusion

- Bounds on mutual information using a 'half Jensen' inequality and a 'two-moment' inequality.
- Two carefully chosen moments can lead to significant improvements for Rényi entropy.
- New measure of dependence with interesting properties.
- One application is precise convergence rates for entropic conditional central limit theorem [R. 2017]

References I

- G. Reeves, "Two-moment inequailties for Rényi entropy and mutual information," 2017, [Online]. Available: https://arxiv.org/abs/1702.07302.
- ——, "Two-moment inequailties for Rényi entropy and mutual information," in *Proc. IEEE Int. Symp. Inform. Theory*, Aachen, Germany, Jun. 2017.
- ——, "Conditional central limit theorems for Gaussian projections," Dec. 2016, [Online]. Available: https://arxiv.org/abs/1612.09252.
- ——, "Conditional central limit theorems for Gaussian projections," in *Proc. IEEE Int. Symp. Inform. Theory*, Aachen, Germany, Jun. 2017.