# Correlation Detection and an Operational Interpretation of the Rényi Mutual Information

Masahito Hayashi[1], Marco Tomamichel[2]

[1]Graduate School of Mathematics, Nagoya University, and
Centre for Quantum Technologies, National University of Singapore
[2]School of Physics, The University of Sydney


ISIT 2015, Hong Kong
(arXiv: 1408.6894)

# Outline and Motivation

- Rényi Entropy and divergence (Rényi'61) have found various applications in information theory:
  - e.g. error exponents for hypothesis testing and channel coding, cryptography, the "Honey Do" problem, etc.
- Conditional Rényi entropy and Rényi mutual information are less understood.
- Mathematical properties of different proposed definitions have recently been investigated
  - see, e.g., Fehr–Berens (TIT'14) or Verdú (ITA'15), and many works in quantum
- We want to find an operational interpretation of the measures.

# Mutual Information

- Two discrete random variables $(X, Y) \sim P_{XY}$.
- Many expression for the mutual information are available:

$$
\begin{aligned}
I(X:Y) &= H(X) + H(Y) - H(XY) && (1)\\
&= H(X) - H(X|Y) && (2)\\
&= D(P_{XY} \| P_X \times P_Y) && (3)\\
&= \min_{Q_Y} D(P_{XY} \| P_X \times Q_Y) && (4)\\
&= \min_{Q_X, Q_Y} D(P_{XY} \| Q_X \times Q_Y)\,. && (5)
\end{aligned}
$$

- Which one to generalize?

# Rényi Mutual Information

- Two discrete random variables $(X, Y) \sim P_{XY}$.
- Many expression for the mutual information are available:

$$^{1}I_\alpha(X:Y) = H_\alpha(X) + H_\alpha(Y) - H_\alpha(XY) \tag{1}$$

$$^{2}I_\alpha(X:Y) = H_\alpha(X) - H_\alpha(X|Y) \tag{2}$$

$$^{3}I_\alpha(X:Y) = D_\alpha(P_{XY} \| P_X \times P_Y) \tag{3}$$

$$^{4}I_\alpha(X:Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y) \tag{4}$$

$$^{5}I_\alpha(X:Y) = \min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X \times Q_Y)\,. \tag{5}$$

- We want the mutual information to be non-negative!
- We want it to be non-increasing under local processing!

# Rényi Mutual Information

- Two discrete random variables $(X, Y) \sim P_{XY}$.
- Many expression for the mutual information are available:

$$^{1}\cancel{I_\alpha(X:Y) = H_\alpha(X) + H_\alpha(Y) - H_\alpha(XY)} \tag{1}$$

$$^{2}I_\alpha(X:Y) = H_\alpha(X) - H_\alpha(X|Y) \tag{2}$$

$$^{3}I_\alpha(X:Y) = D_\alpha(P_{XY} \| P_X \times P_Y) \tag{3}$$

$$^{4}I_\alpha(X:Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y) \tag{4}$$

$$^{5}I_\alpha(X:Y) = \min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X \times Q_Y)\,. \tag{5}$$

- We want the mutual information to be non-negative! ✓
- We want it to be non-increasing under local processing!

# Rényi Mutual Information

- Two discrete random variables $(X, Y) \sim P_{XY}$.
- Many expression for the mutual information are available:

$$^{1}\cancel{I_\alpha(X:Y) = H_\alpha(X) + H_\alpha(Y) - H_\alpha(XY)} \tag{1}$$

$$^{2}\cancel{I_\alpha(X:Y) = H_\alpha(X) - H_\alpha(X|Y)} \tag{2}$$

$$^{3}I_\alpha(X:Y) = D_\alpha(P_{XY} \| P_X \times P_Y) \tag{3}$$

$$^{4}I_\alpha(X:Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y) \tag{4}$$

$$^{5}I_\alpha(X:Y) = \min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X \times Q_Y)\,. \tag{5}$$

- We want the mutual information to be non-negative! ✓
- We want it to be non-increasing under local processing! ✓

# Rényi Mutual Information

- Two discrete random variables $(X, Y) \sim P_{XY}$.
- Many expression for the mutual information are available:

$$^{1}\cancel{I_\alpha(X:Y) = H_\alpha(X) + H_\alpha(Y) - H_\alpha(XY)} \tag{1}$$

$$^{2}\cancel{I_\alpha(X:Y) = H_\alpha(X) - H_\alpha(X|Y)} \tag{2}$$

$$^{3}I_\alpha(X:Y) = D_\alpha(P_{XY} \| P_X \times P_Y) \tag{3}$$

$$\boxed{I_\alpha(X:Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y)} \tag{4}$$

$$^{5}I_\alpha(X:Y) = \min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X \times Q_Y)\,. \tag{5}$$

- We want the mutual information to be non-negative! ✓
- We want it to be non-increasing under local processing! ✓
- This is Sibson's proposal.

# Rényi Entropy and Divergence

- For two pmf's $P_X \ll Q_X$, the Rényi divergence is defined as

$$D_\alpha(P_X\|Q_X) = \frac{1}{\alpha-1}\log\left(\sum_x P_X(x)^\alpha Q_X(x)^{1-\alpha}\right).$$

  for any $\alpha \in (0,1)\cup(1,\infty)$ and as a limit for $\alpha \in \{0,1,\infty\}$.
- Monotonicity: for $\alpha \geq \beta$, we have

$$D_\alpha(P_X\|Q_X) \geq D_\beta(P_X\|Q_X).$$

- Kullback-Leibler divergence:

$$\lim_{\alpha\to 1} D_\alpha(P_X\|Q_X) = D(P_X\|Q_X) = \sum_x P_X(x)\log\frac{P_X(x)}{Q_X(x)}.$$

- Data-processing inequality (DPI): for any channel $W$, we have

$$D_\alpha(P_X\|Q_X) \geq D_\alpha(P_XW\|Q_XW).$$

# Rényi Mutual Information

Recall: $I_\alpha(X:Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y)$

- Inherits monotonicity and DPI from divergence.
- We have $\lim_{\alpha \to 1} I_\alpha(X:Y) = I(X:Y)$.
- Sibson's identity (Sibson'69): minimizer satisfies

$$Q_Y(y)^\alpha \propto \sum_x P_X(x) P_{Y|X}(y|x)^\alpha \,,$$

$$I_\alpha(X:Y) = \frac{1}{\alpha - 1} \log \left( \sum_y \left( \sum_x P_X(x) P_{Y|X}(y|x)^\alpha \right)^{\frac{1}{\alpha}} \right) .$$

- Additivity: $(X_1, X_2, Y_1, Y_2) \leftarrow P_{X_1Y_1} \times P_{X_2Y_2}$ independent:

$$I_\alpha(X_1X_2 : Y_1Y_2) = I_\alpha(X_1 : Y_1) + I_\alpha(X_2 : Y_2) \,.$$

# Correlation Detection and One-Shot Converse

- Correlation Detection: given a pmf $P_{XY}$, consider

  Null Hypothesis: $(X, Y) \sim P_{XY}$
  Alternative Hypothesis: $X \sim P_X$ independent of $Y$

- For a test $T_{Z|XY}$ with $Z \in \{0, 1\}$ define errors

$$\alpha(T) = \Pr[Z = 1], \qquad (X, Y, Z) \sim P_{XY} \times T_{Z|XY}$$
$$\beta(T) = \max_{Q_Y} \Pr[Z = 0], \quad (X, Y, Z) \sim P_X \times Q_Y \times T_{Z|XY}$$

- The one-shot (meta-) converse can be stated in terms of this composite hypothesis testing problem (Polyanskiy'13).
- Any code on $W_{Y|X}$ with input distribution $P_X$ using $M$ codewords and average error $\varepsilon$ satisfies ($P_{XY} = P_X \times W_{Y|X}$):

$$M \leq \frac{1}{\hat{\beta}(\varepsilon)}, \quad \hat{\beta}(\varepsilon) = \min \left\{ \beta(T) \,\middle|\, T_{Z|XY} \text{ s.t. } \alpha(T) \leq \varepsilon \right\}.$$

# Asymptotic Correlation Detection

- Consider the asymptotics $n \to \infty$ for the sequence of problems

  Null Hypothesis: $(X^n, Y^n) \sim P_{XY}^{\times n}$
  Alternative Hypothesis: $X^n \sim P_X^{\times n}$ independent of $Y^n$

- For a test $T^n_{Z|X^nY^n}$ with $Z \in \{0,1\}$ define errors

$$\alpha(T^n) = \Pr[Z=1], \qquad (X,Y,Z) \sim P_{XY}^{\times n} \times T^n_{Z|X^nY^n}$$
$$\beta(T^n) = \max_{Q_{Y^n}} \Pr[Z=0], \quad (X,Y,Z) \sim P_X^{\times n} \times Q_{Y^n} \times T^n_{Z|X^nY^n}$$

- Define minimal error for fixed rate $R > 0$:

$$\hat{\alpha}(R;n) = \min\left\{\alpha(T^n) \,\middle|\, T^n_{Z|X^nY^n} \text{ s.t. } \beta(T^n) \leq \exp(-nR)\right\}.$$

# Error Exponents (Hoeffding)

Recall: $I_s(X:Y) = \min_{Q_Y} D_s(P_{XY} \| P_X \times Q_Y)$

$$\hat{\alpha}(R;n) = \min\left\{\alpha(T^n) \,\middle|\, T^n_{Z|X^nY^n} \text{ s.t. } \beta(T^n) \leq \exp(-nR)\right\}$$

**Result (Error Exponent)**

*For any $R > 0$, we have*

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\log\hat{\alpha}(R;n)\right\} = \sup_{s\in(0,1)}\left\{\frac{1-s}{s}\big(I_s(X:Y) - R\big)\right\}.$$

- If $R \geq I(X:Y)$ it evaluates to $0$, else it is positive.
  - $I(X:Y)$ is the critical rate (cf. Stein's Lemma).
- If $R < I_0(X:Y)$ it diverges to $+\infty$.
  - This is the zero-error regime.

# Strong Converse Exponents (Han–Kobayashi)

Recall: $I_s(X:Y) = \min_{Q_Y} D_s(P_{XY} \| P_X \times Q_Y)$

$$\hat{\alpha}(R;n) = \min\left\{\alpha(T^n) \,\middle|\, T^n_{Z|X^nY^n} \text{ s.t. } \beta(T^n) \leq \exp(-nR)\right\}$$

**Result (Strong Converse Exponent)**

*For any* $0 < R < I_\infty(X:Y)$, *we have*

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\log\left(1-\hat{\alpha}(R;n)\right)\right\} = \sup_{s>1}\left\{\frac{s-1}{s}\left(R - I_s(X:Y)\right)\right\}.$$

- If $R \leq I(X:Y)$ it evaluates to $0$, otherwise it is positive.
  - This implies the strong converse to Stein's Lemma.
- What if $R = I(X:Y)$ ?

# Second Order Expansion

- For small deviations $r$ from the rate $R$, define

$$\hat{\alpha}(R,r;n) = \min\left\{\alpha(T^n)\,\middle|\, T^n_{Z|X^nY^n} \text{ s.t. } \beta(T^n) \leq \exp(-nR-\sqrt{n}r)\right\}.$$

**Result (Second Order Expansion)**

*For any $r \in \mathbb{R}$, we have*

$$\lim_{n\to\infty} \hat{\alpha}\big(I(X:Y),r;n\big) = \Phi\left(\frac{r}{\sqrt{V(X:Y)}}\right).$$

- $\Phi$ is cumulative (normal) Gaussian distribution function.
- $V(X:Y) = V(P_{XY}\|P_X\times P_Y)$ where $V(\cdot\|\cdot)$ is the divergence variance

$$\frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=1} I_s(X:Y) = \frac{1}{2}V(X:Y).$$

# Universal Distribution

- For every $n$, consider the universal pmf (Hayashi'09)

$$T^n_{Y^n}(y^n) = \sum_{\lambda \in \mathcal{P}_n(Y)} \frac{1}{|\mathcal{P}_n(Y)|} U_\lambda(y^n)\,,$$

  where $U_\lambda$ is the uniform distribution over the type class $\lambda$.
- Every $S_n$-invariant pmf $Q_{Y^n}$ satisfies

$$Q_{Y^n}(y^n) \leq |\mathcal{P}_n(Y)| \cdot T^n_{Y^n}(y^n) \qquad \forall y^n\,.$$

- **Main idea:** test $P_{XY}^{\times n}$ vs. $P_X^{\times n} \times T^n_{Y^n}$.

**Lemma**

*For any joint pmf $P_{XY}$, the universal pmf satisfies*

$$D_\alpha\left(P_{XY}^{\times n} \,\middle|\, P_X^{\times n} \times T^n_{Y^n}\right) = nI_\alpha(X:Y) + O(\log n)\,.$$

# Error Exponent: Achievability (1)

- Fix $s \in (0,1)$. Fix sequence $\{\lambda_n\}_n$ to be chosen later.
- We use Neyman-Pearson tests for $P_{XY}^{\times n}$ vs. $P_X^{\times n} \times T_{Y^n}^n$:

$$Z(x^n, y^n) = 1\left\{ \log \frac{P_{XY}^{\times n}(x^n, y^n)}{P_X^{\times n}(x^n) T_{Y^n}^n(y^n)} \geq \lambda_n \right\}.$$

- Then, with $(X^n, Y^n) \sim P_{XY}^{\times n}$, we have

$$\begin{aligned}
\Pr[Z=1] &= \sum_{x^n, y^n} P_{XY}^{\times n}(x^n, y^n)\, 1\left\{ \log \frac{P_{XY}^{\times n}(x^n, y^n)}{P_X^{\times n}(x^n) T_{Y^n}^n(y^n)} \geq \lambda_n \right\} \\
&\leq \exp\big((1-s)\lambda_n\big) \sum_{x^n, y^n} \big(P_{XY}^{\times n}(x^n, y^n)\big)^s \big(P_X^{\times n}(x^n) T_{Y^n}^n(y^n)\big)^{1-s} \\
&= \exp\Big((1-s)\big(\lambda_n - D_s(P_{XY}^{\times n} \| P_X^{\times n} \times T_{Y^n}^n)\big)\Big).
\end{aligned}$$

# Error Exponent: Achievability (2)

- And, with $(X^n, Y^n) \sim P_X^{\times n} \times Q_{Y^n}$, we have

$$\Pr[Z=0] = \sum_{x^n,y^n} P_{X^n}(x^n) Q_{Y^n}(y^n) 1\left\{ \log \frac{P_{XY}^{\times n}(x^n,y^n)}{P_X^{\times n}(x^n) T_{Y^n}^n(y^n)} < \lambda_n \right\}$$

$$= \sum_{x^n,y^n} P_{X^n}(x^n) \widetilde{Q}_{Y^n}(y^n) 1\left\{ \log \frac{P_{XY}^{\times n}(x^n,y^n)}{P_X^{\times n}(x^n) T_{Y^n}^n(y^n)} < \lambda_n \right\}.$$

  where $\widetilde{Q}_{Y^n}(y^n) = \sum_{\pi \in S_n} \frac{1}{|S_n|} Q_{Y^n}(P(\pi) y^n)$ is $S_n$-invariant.

- Now we can bring in the universal pmf again:

$$\Pr[Z=0] \le |\mathcal{P}_n(Y)| \sum_{x^n,y^n} P_{X^n}(x^n) T_{Y^n}^n(y^n) 1\left\{ \log \frac{P_{XY}^{\times n}(x^n,y^n)}{P_X^{\times n}(x^n) T_{Y^n}^n(y^n)} < \lambda_n \right\}$$

$$\le |\mathcal{P}_n(Y)| \exp\Big( -s\lambda_n - (1-s) D_s(P_{XY}^{\times n} \| P_X^{\times n} \times T_{Y^n}^n) \Big).$$

- Choose $\{\lambda_n\}$ such that $\Pr[Z=0] \le \exp(-nR)$.

# Second Order: Achievability

- There exits $\{\lambda_n\}_n$ such that

$$\Pr[Z=0] \le \exp\left(-nI(X:Y) - \sqrt{n}r\right) \quad (X^n, Y^n) \sim P_X^{\times n} \times Q_{Y^n}$$
$$\Pr[Z=1] = \Pr[F_n(X_n, Y_n) < r] \quad (X^n, Y^n) \sim P_{XY}^{\times n}\,.$$

  with a new sequence of random variables

$$F_n(X_n, Y_n) = \frac{1}{\sqrt{n}}\left(\log \frac{P_{XY}^{\times n}(X^n, Y^n)}{P_X^{\times n}(X^n) T_{Y^n}^n(Y^n)} - nR - \log|\mathcal{P}_n(Y)|\right).$$

- Asymptotic cumulant generating function:

$$\begin{aligned}\Lambda_F(t) &= \lim_{n\to\infty} \log \mathbb{E}[\exp(tF_n)] \\ &= \lim_{n\to\infty} \frac{t}{\sqrt{n}}\left(D_{1+\frac{t}{\sqrt{n}}}(P_{XY}^{\times n} \| P_X^{\times n} \times T_{Y^n}^n) - nI(X:Y)\right) \\ &= \frac{t^2}{2} V(P_{XY} \| P_X \times P_Y)\,.\end{aligned}$$

- $F_n$ converges in distribution to a Gaussian $F$ with variance $V$ (by a variation of Lévi's continuity theorem).

# Quantum Hypothesis Testing

- Given a bipartite quantum state $\rho_{AB}$, consider

  Null Hypothesis: state is $\rho_{AB}$
  Alternative Hypothesis: state is $\rho_A \otimes \sigma_B$ for some state $\sigma_B$

- Using the same notation:

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\log\hat{\alpha}(R;n)\right\} = \sup_{s\in(0,1)}\left\{\frac{1-s}{s}\big(\bar{I}_s(A:B)-R\big)\right\},$$

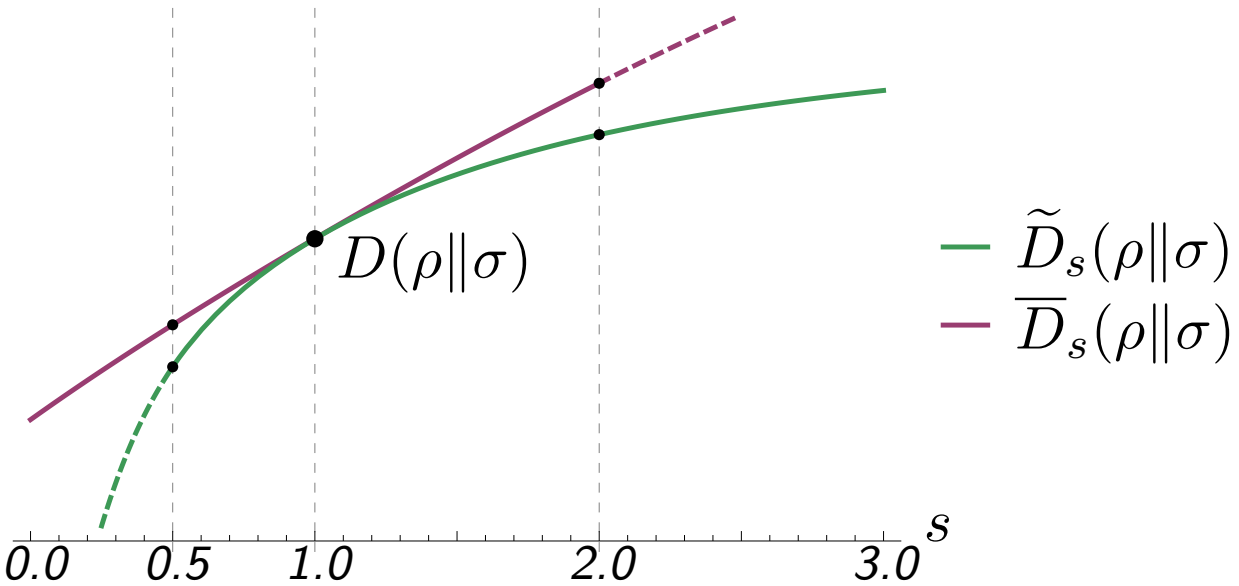
$$\lim_{n\to\infty}\left\{-\frac{1}{n}\log\big(1-\alpha(R;n)\big)\right\} = \sup_{s>1}\left\{\frac{s-1}{s}\big(R-\widetilde{I}_s(A:B)\big)\right\}.$$

- The definition are similar,

$$\left.\begin{matrix}\bar{I}_s(A:B)\\ \widetilde{I}_s(A:B)\end{matrix}\right\} = \min_{\sigma_B}\left\{\begin{matrix}\bar{D}_s(\rho_{AB}\|\rho_A\otimes\sigma_B)\\ \widetilde{D}_s(\rho_{AB}\|\rho_A\otimes\sigma_B)\end{matrix}\right..$$

- But $\bar{D}_s$ and $\widetilde{D}_s$ are different!

# Two Quantum Rényi Divergences




- They agree with the classical quantity for commuting states.

$$\overline{D}_s(\rho\|\sigma) = \frac{1}{s-1}\log \operatorname{tr}\left(\rho^s \sigma^{1-s}\right) ,$$

$$\widetilde{D}_s(\rho\|\sigma) = \frac{1}{s-1}\log \operatorname{tr}\left(\left(\sigma^{\frac{1-s}{2s}}\rho\sigma^{\frac{1-s}{2s}}\right)^s\right) .$$

# Summary and Outlook

- Correlation detection gives operational meaning to

$$I_\alpha(X:Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y)\,.$$

- Similarly Arimoto's conditional Rényi entropy

$$H_\alpha(X|Y) = \log|X| - \min_{Q_Y} D_\alpha(P_{XY} \| U_X \times Q_Y)\,.$$

has an operational interpretation:

Null Hypothesis: $(X,Y) \sim P_{XY}$
Alternative Hypothesis: $X \sim U_X$ uniform and indep. of $Y$

- Does the symmetric mutual information

$$I'_\alpha(X:Y) = \min_{Q_X, Q_Y} D_\alpha(P_{XY} \| Q_X \times Q_Y)$$

have a natural operational interpretation?