Operational Interpretation of Rényi Information Measures via Composite Hypothesis Testing

Adway Girish
Information Theory Laboratory

June 21, 2023

Outline

Outline

Outline

Outline

Outline

Outline

Simple hypothesis testing

Simple hypothesis testing

X^n

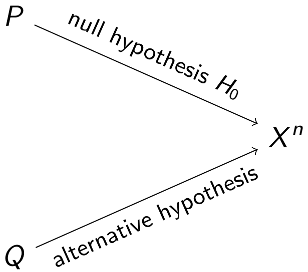
Simple hypothesis testing

P

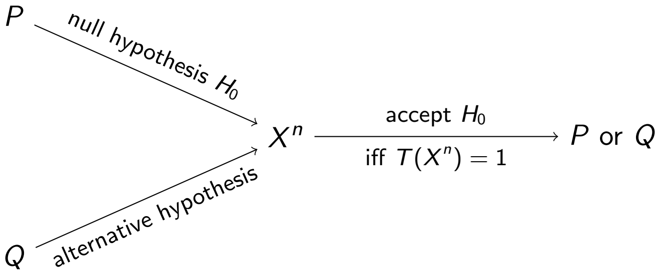
X^n

Q

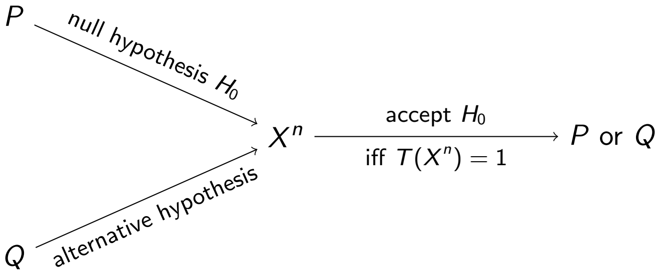
Simple hypothesis testing



Simple hypothesis testing

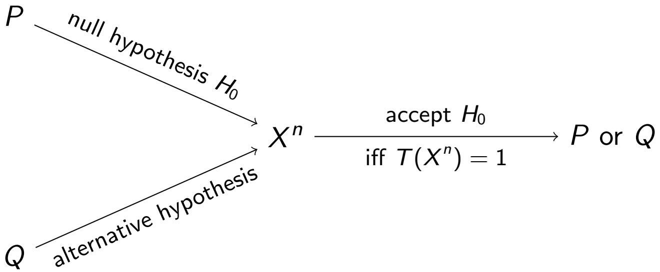


Simple hypothesis testing



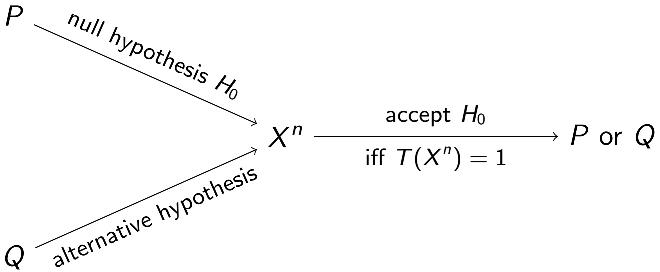
- Reject when true:
- Accept when false:

Simple hypothesis testing



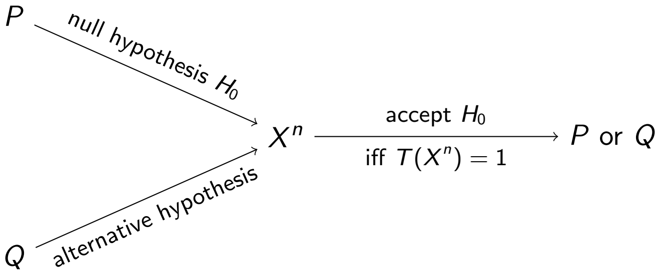
- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\}$
- Accept when false:

Simple hypothesis testing



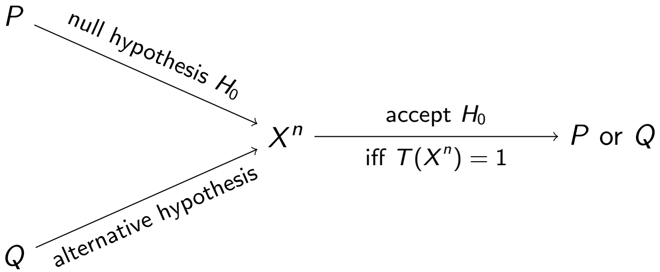
- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false:

Simple hypothesis testing



- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false: $\mathrm{q}_n = Q^n\{T(X^n) = 1\}$

Simple hypothesis testing



- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false: $\mathrm{q}_n = Q^n\{T(X^n) = 1\} \longrightarrow$ Type-II error

Information measures as error exponents

- If $\mathrm{p}_n \leq \epsilon$,

Information measures as error exponents

- If $\mathrm{p}_n \le \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P \,\|\, Q) + o(n)\big)$.[†]

[†] Herman Chernoff. "Large-sample theory: Parametric case". In: *The Annals of Mathematical Statistics* 1 (1956)

Information measures as error exponents

- If $\mathrm{p}_n \leq \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P \,||\, Q) + o(n)\big)$.
- Mutual information $I(X;Y) = D_{\mathsf{KL}}(P_{XY} \,||\, P_X P_Y)$.

Information measures as error exponents

- If $\mathrm{p}_n \leq \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P \,||\, Q) + o(n)\big)$.
- Mutual information $I(X;Y) = D_{\mathsf{KL}}(P_{XY} \,||\, P_X P_Y)$.
- Consider the following HT setup,

$$\begin{aligned} \text{null hypothesis} &: \\ \text{alternative hypothesis} &: \end{aligned}$$

Information measures as error exponents

- If $\mathrm{p}_n \leq \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P \,||\, Q) + o(n)\big)$.
- Mutual information $I(X;Y) = D_{\mathsf{KL}}(P_{XY} \,||\, P_X P_Y)$.
- Consider the following HT setup,

$$\begin{aligned} \text{null hypothesis} &: (X^n, Y^n) \sim P_{XY}^n, \\ \text{alternative hypothesis} &: \end{aligned}$$

Information measures as error exponents

- If $\mathrm{p}_n \leq \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P \,||\, Q) + o(n)\big)$.
- Mutual information $I(X;Y) = D_{\mathsf{KL}}(P_{XY} \,||\, P_X P_Y)$.
- Consider the following HT setup,

$$\text{null hypothesis} : (X^n, Y^n) \sim P_{XY}^n,$$
$$\text{alternative hypothesis} : (X^n, Y^n) \sim P_X^n P_Y^n.$$

Information measures as error exponents

- If $\mathrm{p}_n \leq \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P\,||\,Q) + o(n)\big)$.
- Mutual information $I(X;Y) = D_{\mathsf{KL}}(P_{XY}\,||\,P_X P_Y)$.
- Consider the following HT setup,

$$\text{null hypothesis} : (X^n, Y^n) \sim P^n_{XY},$$
$$\text{alternative hypothesis} : (X^n, Y^n) \sim P^n_X P^n_Y.$$

- If $\mathrm{p}_n \leq \epsilon$,

Information measures as error exponents

- If $\mathrm{p}_n \leq \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P\,||\,Q) + o(n)\big)$.
- Mutual information $I(X;Y) = D_{\mathsf{KL}}(P_{XY}\,||\,P_X P_Y)$.
- Consider the following HT setup,

$$\begin{aligned} \text{null hypothesis} : (X^n, Y^n) &\sim P_{XY}^n, \\ \text{alternative hypothesis} : (X^n, Y^n) &\sim P_X^n P_Y^n. \end{aligned}$$

- If $\mathrm{p}_n \leq \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nI(X;Y) + o(n)\big)$.

Information measures as error exponents

- If $\mathrm{p}_n \le \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nD_{\mathsf{KL}}(P \,||\, Q) + o(n)\big)$.
- Mutual information $I(X;Y) = D_{\mathsf{KL}}(P_{XY} \,||\, P_X P_Y)$.
- Consider the following HT setup,

$$\text{null hypothesis} : (X^n, Y^n) \sim P_{XY}^n,$$
$$\text{alternative hypothesis} : (X^n, Y^n) \sim P_X^n P_Y^n.$$

- If $\mathrm{p}_n \le \epsilon$, then the optimal $\mathrm{q}_n = \exp\big(-nI(X;Y) + o(n)\big)$.
- Operational interpretation of $I(X;Y)$.

Rényi divergences as error exponents

- Rényi divergence of order α, for $\alpha > 0$, $\alpha \neq 1$

Rényi divergences as error exponents

- Rényi divergence of order α, for $\alpha > 0$, $\alpha \neq 1$

$$D_\alpha(P \,||\, Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_P \left[\left(\frac{P}{Q} \right)^{\alpha - 1} \right].$$

Rényi divergences as error exponents

- Rényi divergence of order α, for $\alpha > 0$, $\alpha \neq 1$

$$D_\alpha(P \,||\, Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_P \left[\left(\frac{P}{Q} \right)^{\alpha - 1} \right].$$

- If $\mathrm{q}_n \leq \exp(-nR)$ for some $0 < R < D_{\mathsf{KL}}(P \,||\, Q)$,

Rényi divergences as error exponents

- Rényi divergence of order α, for $\alpha > 0$, $\alpha \neq 1$

$$D_\alpha(P \,||\, Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_P \left[\left(\frac{P}{Q} \right)^{\alpha - 1} \right].$$

- If $\mathrm{q}_n \leq \exp(-nR)$ for some $0 < R < D_{\mathrm{KL}}(P \,||\, Q)$, then the optimal

$$\mathrm{p}_n = \exp\left(-n \sup_{0<\alpha<1} \frac{1-\alpha}{\alpha} (D_\alpha(P \,||\, Q) - R) + o(n) \right).$$

Rényi information measures

Rényi information measures

$I(X;Y)$	$I_\alpha(X;Y)$

Rényi information measures

$I(X;Y)$	$I_\alpha(X;Y)$
$D_{\mathsf{KL}}(P_{XY} \,\|\, P_X P_Y)$	
$\min_{Q_Y} D_{\mathsf{KL}}(P_{XY} \,\|\, P_X Q_Y)$	
$\min_{Q_X,\, Q_Y} D_{\mathsf{KL}}(P_{XY} \,\|\, Q_X Q_Y)$	

Rényi information measures

$I(X;Y)$	$I_\alpha(X;Y)$
$D_{\mathrm{KL}}(P_{XY} \,\|\, P_X P_Y)$	$D_\alpha(P_{XY} \,\|\, P_X P_Y)$?
$\min_{Q_Y} D_{\mathrm{KL}}(P_{XY} \,\|\, P_X Q_Y)$	$\min_{Q_Y} D_\alpha(P_{XY} \,\|\, P_X Q_Y)$?
$\min_{Q_X, Q_Y} D_{\mathrm{KL}}(P_{XY} \,\|\, Q_X Q_Y)$	$\min_{Q_X, Q_Y} D_\alpha(P_{XY} \,\|\, Q_X Q_Y)$?

Rényi information measures

$I(X;Y)$	$I_\alpha(X;Y)$
$D_{\mathrm{KL}}(P_{XY} \,\|\, P_X P_Y)$	$D_\alpha(P_{XY} \,\|\, P_X P_Y)$
$\min_{Q_Y} D_{\mathrm{KL}}(P_{XY} \,\|\, P_X Q_Y)$	$\boxed{\min_{Q_Y} D_\alpha(P_{XY} \,\|\, P_X Q_Y)}$
$\min_{Q_X, Q_Y} D_{\mathrm{KL}}(P_{XY} \,\|\, Q_X Q_Y)$	$\min_{Q_X, Q_Y} D_\alpha(P_{XY} \,\|\, Q_X Q_Y)$

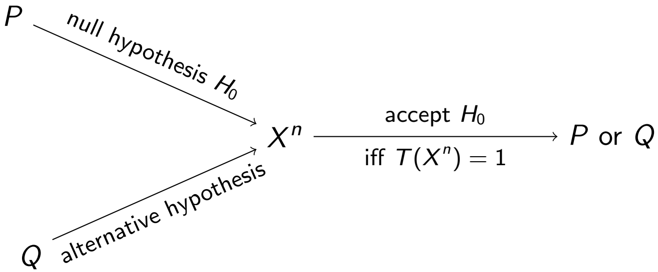
Rényi information measures

$I(X;Y)$	$I_\alpha(X;Y)$
$D_{\mathrm{KL}}(P_{XY} \,\|\, P_X P_Y)$	$D_\alpha(P_{XY} \,\|\, P_X P_Y)$
$\min_{Q_Y} D_{\mathrm{KL}}(P_{XY} \,\|\, P_X Q_Y)$	$\boxed{\min_{Q_Y} D_\alpha(P_{XY} \,\|\, P_X Q_Y)}$
$\min_{Q_X, Q_Y} D_{\mathrm{KL}}(P_{XY} \,\|\, Q_X Q_Y)$	$\min_{Q_X, Q_Y} D_\alpha(P_{XY} \,\|\, Q_X Q_Y)$

Sibson's mutual information[†]

[†] Robin Sibson. "Information radius". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 2 (1969)

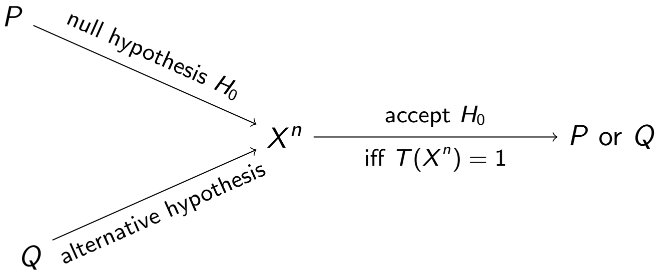
Simple hypothesis testing again



- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false: $\mathrm{q}_n = Q^n\{T(X^n) = 1\} \longrightarrow$ Type-II error

Composite hypothesis testing

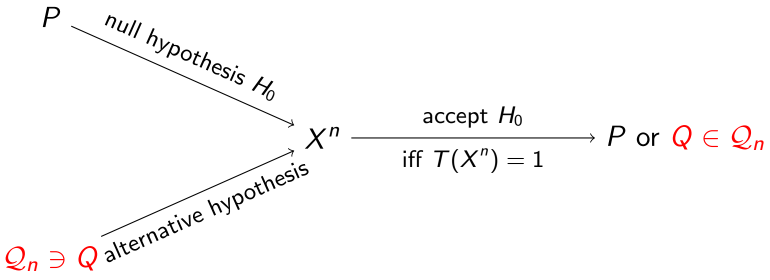
- For a fixed set of distributions $\mathcal{Q}_n$,



- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false: $\mathrm{q}_n = Q^n\{T(X^n) = 1\} \longrightarrow$ Type-II error

Composite hypothesis testing

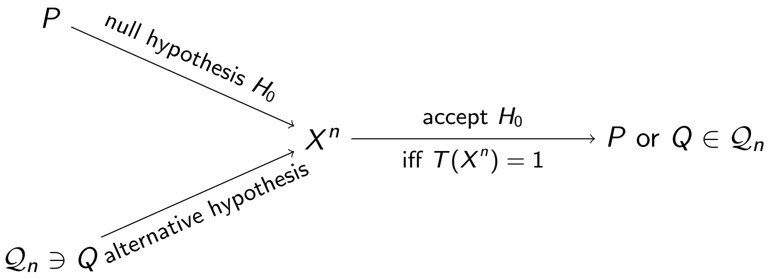
- For a fixed set of distributions $\mathcal{Q}_n$,



- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false: $\mathrm{q}_n = Q^n\{T(X^n) = 1\} \longrightarrow$ Type-II error

Composite hypothesis testing

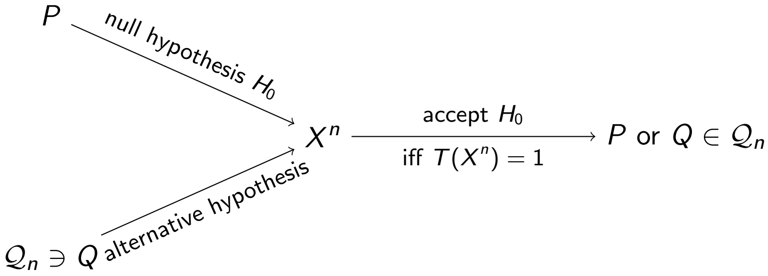
- For a fixed set of distributions $\mathcal{Q}_n$,



- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false: $\mathrm{q}_n = \max_{Q \in \mathcal{Q}_n} Q^n\{T(X^n) = 1\} \longrightarrow$ Type-II error

Composite hypothesis testing

- For a fixed set of distributions $\mathcal{Q}_n$,



- Reject when true: $\mathrm{p}_n = P^n\{T(X^n) = 0\} \longrightarrow$ Type-I error
- Accept when false: $\mathrm{q}_n = \max_{Q \in \mathcal{Q}_n} Q^n\{T(X^n) = 1\} \longrightarrow$ Type-II error

Error exponents for composite hypothesis testing

Error exponents for composite hypothesis testing

- $D_\alpha(P \,||\, \mathcal{Q}_n) \triangleq \min_{Q \in \mathcal{Q}_n} D_\alpha(P \,||\, Q)$.

Error exponents for composite hypothesis testing

- $D_\alpha(P \,||\, \mathcal{Q}_n) \triangleq \min_{Q \in \mathcal{Q}_n} D_\alpha(P \,||\, Q)$.
- [†]If $\mathrm{q}_n \leq \exp(-nR)$ for some $0 < R < D_{\mathsf{KL}}(P \,||\, \mathcal{Q}_n)$,

[†]Marco Tomamichel and Masahito Hayashi. "Operational Interpretation of Rényi Information Measures via Composite Hypothesis Testing Against Product and Markov Distributions". In: *IEEE Transactions on Information Theory* 2 (2018)

Error exponents for composite hypothesis testing

- $D_\alpha(P \,||\, \mathcal{Q}_n) \triangleq \min_{Q \in \mathcal{Q}_n} D_\alpha(P \,||\, Q)$.
- [†]If $\mathrm{q}_n \leq \exp(-nR)$ for some $0 < R < D_{\mathsf{KL}}(P \,||\, \mathcal{Q}_n)$, then the optimal

$$\mathrm{p}_n = \exp\left(-n \sup_{0<\alpha<1} \frac{1-\alpha}{\alpha}(D_\alpha(P \,||\, \mathcal{Q}_n) - R) + o(n)\right),$$

[†]Marco Tomamichel and Masahito Hayashi. "Operational Interpretation of Rényi Information Measures via Composite Hypothesis Testing Against Product and Markov Distributions". In: *IEEE Transactions on Information Theory* 2 (2018)

Error exponents for composite hypothesis testing

- $D_\alpha(P \,||\, \mathcal{Q}_n) \triangleq \min_{Q \in \mathcal{Q}_n} D_\alpha(P \,||\, Q)$.

- [†]If $\mathrm{q}_n \leq \exp(-nR)$ for some $0 < R < D_{\mathsf{KL}}(P \,||\, \mathcal{Q}_n)$, then the optimal

$$\mathrm{p}_n = \exp\left(-n \sup_{0<\alpha<1} \frac{1-\alpha}{\alpha}(D_\alpha(P \,||\, \mathcal{Q}_n) - R) + o(n)\right),$$

  but *only* under some conditions on $\mathcal{Q}_n$.

[†]Marco Tomamichel and Masahito Hayashi. "Operational Interpretation of Rényi Information Measures via Composite Hypothesis Testing Against Product and Markov Distributions". In: *IEEE Transactions on Information Theory* 2 (2018)

Sibson's mutual information as an error exponent

- $I_\alpha^S(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \,||\, P_X Q_Y)$

Sibson's mutual information as an error exponent

- $I_\alpha^S(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \,||\, P_X Q_Y)$

- $D_\alpha(P \,||\, \mathcal{Q}_n) = I_\alpha^S(X;Y)$ for the following HT setup,

$$\begin{aligned} \text{null hypothesis} &: \\ \text{alternative hypothesis} &: \qquad\qquad . \end{aligned}$$

Sibson's mutual information as an error exponent

- $I_\alpha^S(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \,||\, P_X Q_Y)$

- $D_\alpha(P \,||\, \mathcal{Q}_n) = I_\alpha^S(X;Y)$ for the following HT setup,

$$\text{null hypothesis}: (X^n, Y^n) \sim P_{XY}^n,$$
$$\text{alternative hypothesis}: \qquad\qquad\qquad .$$

Sibson's mutual information as an error exponent

- $I_\alpha^S(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \,||\, P_X Q_Y)$

- $D_\alpha(P \,||\, \mathcal{Q}_n) = I_\alpha^S(X;Y)$ for the following HT setup,

$$\text{null hypothesis} : (X^n, Y^n) \sim P_{XY}^n,$$
$$\text{alternative hypothesis} : X^n \sim P_X^n, \text{ independent of } Y^n.$$

Sibson's mutual information as an error exponent

- $I_\alpha^S(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \,||\, P_X Q_Y)$

- $D_\alpha(P \,||\, \mathcal{Q}_n) = I_\alpha^S(X;Y)$ for the following HT setup,

$$\text{null hypothesis} : (X^n, Y^n) \sim P_{XY}^n,$$
$$\text{alternative hypothesis} : X^n \sim P_X^n, \ \text{independent of } Y^n.$$

- If $\mathrm{q}_n \leq \exp(-nR)$ for some $0 < R < \min_{Q_Y} D_{\mathrm{KL}}(P_{XY} \,||\, P_X Q_Y)$,

Sibson's mutual information as an error exponent

- $I_\alpha^S(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \,||\, P_X Q_Y)$

- $D_\alpha(P \,||\, \mathcal{Q}_n) = I_\alpha^S(X;Y)$ for the following HT setup,

$$\text{null hypothesis} : (X^n, Y^n) \sim P_{XY}^n,$$
$$\text{alternative hypothesis} : X^n \sim P_X^n, \text{ independent of } Y^n.$$

- If $\mathrm{q}_n \leq \exp(-nR)$ for some $0 < R < \min_{Q_Y} D_{\mathsf{KL}}(P_{XY} \,||\, P_X Q_Y)$, then the optimal

$$\mathrm{p}_n = \exp\left(-n \sup_{0<\alpha<1} \frac{1-\alpha}{\alpha}(I_\alpha^S(X;Y) - R) + o(n)\right).$$

Closing remarks

- Extension of results from simple to composite hypothesis testing

Closing remarks

- Extension of results from simple to composite hypothesis testing
- Operational interpretation of Sibson's mutual information

Closing remarks

- Extension of results from simple to composite hypothesis testing
- Operational interpretation of Sibson's mutual information
- Future work:

Closing remarks

- Extension of results from simple to composite hypothesis testing
- Operational interpretation of Sibson's mutual information
- Future work:
  - Relax conditions on $\mathcal{Q}_n$ and generalize these results

Closing remarks

- Extension of results from simple to composite hypothesis testing
- Operational interpretation of Sibson's mutual information
- Future work:
  - Relax conditions on $\mathcal{Q}_n$ and generalize these results
  - Clearer interpretations

Thank you!