Chapter 1

Generalized Information Measures for Arbitrary Systems with Memory

Po-Ning Chen

Institute of Communications Engineering

National Chiao-Tung University

Hsin Chu, Taiwan 30010

Shannon's entropy

- Entropy of a discrete random variable X:

$$H(X) := -\sum_{x\in\mathcal{X}} P_X(x)\log P_X(x) = E_X\left[-\log P_X(X)\right] \text{ nats}$$

  is a measure of the average amount of uncertainty in X.

- Entropy rate for a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ is

$$\lim_{n\to\infty}\frac{1}{n}H(X^n) = \lim_{n\to\infty}\frac{1}{n}E\left[-\log P_{X^n}(X^n)\right],$$

  assuming the limit exists.

- Operation meaning: Shannon's source coding theorems for stationary and ergodic systems.

- Question: Does these measures have the same operational significance for systems with time-varying and non-stationary statistics. Answer: No.

- Solution: Require new entropy measure which can appropriately characterize the operational limits of *arbitrary* stochastic systems.

Arbitrary systems with memory

- In general, there are two indices for random variables or observations: a time index and a space index.
- When a sequence of random variables is denoted by

$$X_1, X_2, \ldots, X_n, \ldots,$$

  the subscript i of X_i can be treated as either a time index or a space index, but not both.
- Hence, when a sequence of random variables is a function of both time and space, the notation of $X_1, X_2, \ldots, X_n, \ldots$, is by no means sufficient; and therefore, a new model for a general time-varying source, such as

$$X_1^{(n)}, X_2^{(n)}, \ldots, X_t^{(n)}, \ldots,$$

  where t is the time index and n is the space or position index (or vice versa), becomes significant.

Arbitrary systems with memory

- When block-wise (fixed-length) compression of such source (with blocklength n) is considered, the same question as to the compression of i.i.d. source arises:

  > *what is the minimum compression rate (say in bits per source sample) for which the probability of error probability can be made arbitrarily small as the blocklength goes to infinity?*

- To answer the question, information theorists have to find a sequence of data compression codes for each blocklength n and investigate if the decompression error goes to zero as n approaches infinity.

Arbitrary systems with memory

- However, unlike those simple source models such as discrete memorylessness, the source being arbitrary may exhibit distinct statistics for each blocklength n; e.g., for

$$\begin{aligned} n=1 &: X_1^{(1)} \\ n=2 &: X_1^{(2)}, X_2^{(2)} \\ n=3 &: X_1^{(3)}, X_2^{(3)}, X_3^{(3)} \\ n=4 &: X_1^{(4)}, X_2^{(4)}, X_3^{(4)}, X_4^{(4)} \end{aligned} \tag{1.0.2}$$

  the statistics of $X_1^{(4)}$ could be different from $X_1^{(1)}$, $X_1^{(2)}$ and $X_1^{(3)}$ (i.e., the source statistics are not necessarily consistent).

- Since the model in question (1.0.2) is general, and the system statistics can be *arbitrarily* defined, it is therefore named an *arbitrary system with memory.*

Arbitrary systems with memory

- The triangular array of random variables in (1.0.2) is denoted by a boldface letter as

$$\boldsymbol{X} := \{X^n\}_{n=1}^{\infty},$$

where

$$X^n := \left(X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}\right);$$

for convenience, we also write

$$\boldsymbol{X} := \left\{X^n = \left(X_1^{(n)}, X_2^{(n)}, \ldots, X_n^{(n)}\right)\right\}_{n=1}^{\infty}.$$

Spectrum and Quantile

Definition 1.1 (Inf/sup-spectrum) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of random variables, then its *inf-spectrum* $\underline{u}(\cdot)$ and its *sup-spectrum* $\bar{u}(\cdot)$ are defined by

$$\underline{u}(\theta) := \liminf_{n\to\infty} \Pr\{A_n \le \theta\} \quad \text{and} \quad \bar{u}(\theta) := \limsup_{n\to\infty} \Pr\{A_n \le \theta\},$$

respectively, where $\theta \in \mathbb{R}$.

- $\underline{u}(\cdot)$ and $\bar{u}(\cdot)$ are respectively the liminf and the limsup of the cumulative distribution function (CDF) of A_n.

Definition 1.2 (Quantile of inf/sup-spectrum) For any $0 \le \delta \le 1$, For any $0 \le \delta \le 1$, the *quantile* $\underline{U}_\delta$ of the sup-spectrum $\bar{u}(\cdot)$ and the *quantile* $\bar{U}_\delta$ of the inf-spectrum $\underline{u}(\cdot)$ are defined by

$$\underline{U}_\delta := \sup\{\theta : \bar{u}(\theta) \le \delta\} \quad \text{and} \quad \bar{U}_\delta := \sup\{\theta : \underline{u}(\theta) \le \delta\},$$

respectively. It follows from the above definitions that $\underline{U}_\delta$ and $\bar{U}_\delta$ are right-continuous and non-decreasing in δ. Note that the supremum of an empty set is defined to be $-\infty$.

- If $\bar{u}(\cdot)$ is strictly increasing, then the quantile is exactly its inverse: $\underline{U}_\delta = \bar{u}^{-1}(\delta)$.

Liminf in probability and limsup in probability

- *liminf in probability* $\underline{U}$ of $\{A_n\}_{n=1}^{\infty}$ is the largest extended real number such that for all $\xi > 0$,

$$\lim_{n\to\infty} \Pr[A_n \le \underline{U} - \xi] = 0.$$

- *limsup in probability* $\bar{U}$ of $\{A_n\}_{n=1}^{\infty}$ is the smallest extended real number such that for all $\xi > 0$,

$$\lim_{n\to\infty} \Pr[A_n \ge \bar{U} + \xi] = 0.$$

-

$$\underline{U} = \lim_{\delta\downarrow 0} \underline{U}_\delta = \underline{U}_0$$

and

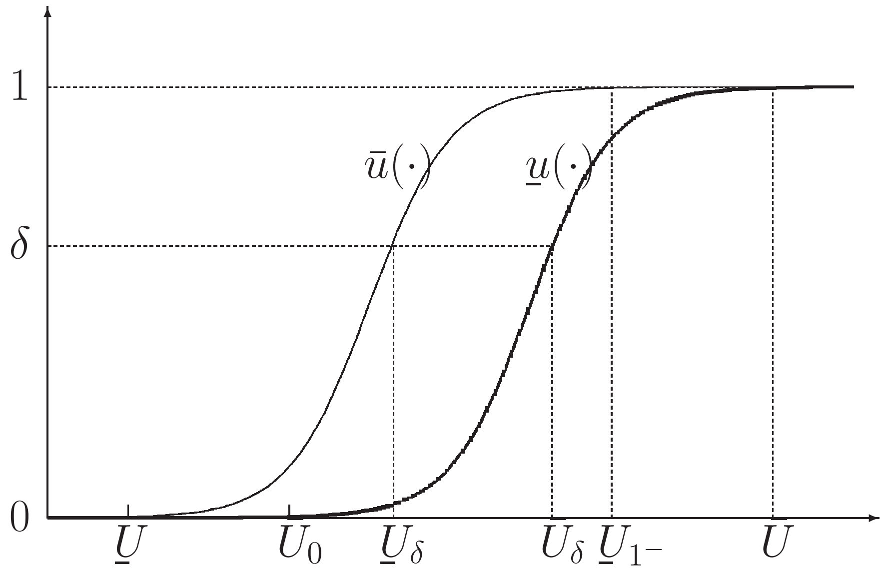
$$\bar{U} = \lim_{\delta\uparrow 1} \bar{U}_\delta = \sup\{\theta : \underline{u}(\theta) < 1\}.$$

- It readily follows from the above definitions that

$$\underline{U} \le \underline{U}_\delta \le \bar{U}_\delta \le \bar{U} \quad \text{for } \delta \in [0, 1).$$

- $\bar{U}_1 = \underline{U}_1 = \infty.$

Liminf in probability and limsup in probability



The asymptotic CDFs (spectrums) of $\{A_n\}_{n=1}^{\infty}$ and their quantiles:

$\bar{u}(\cdot)$ = sup-spectrum of $\{A_n\}$, $\underline{u}(\cdot)$ = inf-spectrum of $\{A_n\}$,

$\underline{U}_\delta$ = quantile of $\bar{u}(\cdot)$, $\bar{U}_\delta$ = quantile of $\underline{u}(\cdot)$,

$\underline{U} = \lim_{\delta\downarrow 0} \underline{U}_\delta = \underline{U}_0$, $\underline{U}_{1^-} = \lim_{\xi\uparrow 1} \underline{U}_\xi$, $\bar{U} = \lim_{\delta\uparrow 1} \bar{U}_\delta$.

Properties of quantile

Lemma 1.4 Assume:

- Two random sequences: $\{A_n\}_{n=1}^{\infty}$ and $\{B_n\}_{n=1}^{\infty}$;
- $\bar{u}(\cdot)$ = sup-spectrum of $\{A_n\}_{n=1}^{\infty}$; $\underline{U}_\delta$ = quantile of $\bar{u}(\cdot)$;
- $\underline{u}(\cdot)$ = inf-spectrum of $\{A_n\}_{n=1}^{\infty}$; $\bar{U}_\delta$ = quantile of $\underline{u}(\cdot)$;
- $\bar{v}(\cdot)$ = sup-spectrum of $\{B_n\}_{n=1}^{\infty}$; $\underline{V}_\delta$ = quantile of $\bar{u}(\cdot)$;
- $\underline{v}(\cdot)$ = inf-spectrum of $\{B_n\}_{n=1}^{\infty}$; $\bar{V}_\delta$ = quantile of $\underline{u}(\cdot)$;
- $(\overline{u+v})(\cdot)$ = sup-spectrum of $\{A_n + B_n\}_{n=1}^{\infty}$, i.e.,

$$(\overline{u+v})(\theta) := \limsup_{n\to\infty} \Pr\{A_n + B_n \le \theta\};$$

  $(\underline{U+V})_\delta$ = quantile of $(\overline{u+v})(\cdot)$;

- $(\underline{u+v})(\cdot)$ = inf-spectrum of $\{A_n + B_n\}_{n=1}^{\infty}$, i.e.,

$$(\underline{u+v})(\theta) := \liminf_{n\to\infty} \Pr\{A_n + B_n \le \theta\};$$

  $(\overline{U+V})_\delta$ = quantile of $(\underline{u+v})(\cdot)$.

Then the following statements hold.

1. $\underline{U}_\delta$ and $\bar{U}_\delta$ are both non-decreasing and right-continuous functions of δ for $\delta \in [0, 1]$.
2. $\lim_{\delta \downarrow 0} \underline{U}_\delta = \underline{U}_0$ and $\lim_{\delta \downarrow 0} \bar{U}_\delta = \bar{U}_0$.
3. For $\delta \geq 0$, $\gamma \geq 0$, and $\delta + \gamma \leq 1$,

$$(\underline{U + V})_{\delta+\gamma} \geq \underline{U}_\delta + \underline{V}_\gamma, \tag{1.2.1}$$

   and

$$(\overline{U + V})_{\delta+\gamma} \geq \underline{U}_\delta + \bar{V}_\gamma. \tag{1.2.2}$$

4. For $\delta \geq 0$, $\gamma \geq 0$, and $\delta + \gamma \leq 1$,

$$(\underline{U + V})_{\delta} \leq \underline{U}_{\delta+\gamma} + \bar{V}_{(1-\gamma)}, \tag{1.2.3}$$

   and

$$(\overline{U + V})_{\delta} \leq \bar{U}_{\delta+\gamma} + \bar{V}_{(1-\gamma)}. \tag{1.2.4}$$

Generalized information measures

In Definitions 1.1 and 1.2,

- A_n = *normalized entropy density*, i.e.,

$$\frac{1}{n}h_{X^n}(X^n) := -\frac{1}{n}\log P_{X^n}(X^n),$$

$$\delta\text{-}\textit{inf-entropy rate}\, \underline{H}_\delta(\boldsymbol{X}) = \text{quantile of the sup-spectrum of } \frac{1}{n}h_{X^n}(X^n)$$

$$\delta\text{-}\textit{sup-entropy rate}\ \bar{H}_\delta(\boldsymbol{X}) = \text{quantile of the inf-spectrum of } \frac{1}{n}h_{X^n}(X^n).$$

- A_n = *normalized information density*, i.e.,

$$\frac{1}{n}i_{X^nW^n}(X^n;Y^n) = \frac{1}{n}i_{X^n,Y^n}(X^n;Y^n) := \frac{1}{n}\log\frac{P_{X^n,Y^n}(X^n,Y^n)}{P_{X^n}(X^n)P_{Y^n}(Y^n)},$$

$$\delta\text{-}\textit{inf-information-rate}\ \underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) = \text{quantile of the sup-spectrum of } \frac{1}{n}i_{X^nW^n}(X^n;Y^n)$$

$$\delta\text{-}\textit{sup-information-rate}\ \bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) = \text{quantile of the inf-spectrum of } \frac{1}{n}i_{X^nW^n}(X^n;Y^n).$$

Generalized information measures

- A_n = *normalized log-likelihood ratio*, i.e.,

$$\frac{1}{n} d_{X^n}(X^n \| \hat{X}^n) := \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\hat{X}^n}(X^n)}$$

δ-inf-divergence rate $\underline{D}_\delta(\boldsymbol{X} \| \hat{\boldsymbol{X}})$ = quantile of the sup-spectrum of $\frac{1}{n} d_{X^n}(X^n \| \hat{X}^n)$

δ-sup-divergence rate $\bar{D}_\delta(\boldsymbol{X} \| \hat{\boldsymbol{X}})$ = quantile of the inf-spectrum of $\frac{1}{n} d_{X^n}(X^n \| \hat{X}^n)$.

- Notes:
  - The *inf-entropy-rate* $\underline{H}(\boldsymbol{X})$ and the *sup-entropy-rate* $\bar{H}(\boldsymbol{X})$ are special cases of the δ-inf/sup-entropy rate measures:

$$\underline{H}(\boldsymbol{X}) = \underline{H}_0(\boldsymbol{X}), \quad \text{and} \quad \bar{H}(\boldsymbol{X}) = \lim_{\delta \uparrow 1} \bar{H}_\delta(\boldsymbol{X}).$$

  - Concept: If the random variable $(1/n)h(X^n)$ exhibits a limiting distribution, and suppose the limiting distribution of $(1/n)h_{X^n}(X^n)$ is positive over $(-2, 2)$; and zero, otherwise. Then $\bar{H}(\boldsymbol{X}) = 2$ and $\underline{H}(\boldsymbol{X}) = -2$.

Generalized information measures

Entropy Measures	
system	arbitrary source $\boldsymbol{X}$
norm. entropy density	$\frac{1}{n}h_{X^n}(X^n) := -\frac{1}{n}\log P_{X^n}(X^n)$
entropy sup-spectrum	$\bar{h}(\theta) := \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \theta\right\}$
entropy inf-spectrum	$\underline{h}(\theta) := \liminf_{n\to\infty} \Pr\left\{\frac{1}{n}h_{X^n}(X^n) \le \theta\right\}$
δ-inf-entropy rate	$\underline{H}_\delta(\boldsymbol{X}) := \sup\{\theta : \bar{h}(\theta) \le \delta\}$
δ-sup-entropy rate	$\bar{H}_\delta(\boldsymbol{X}) := \sup\{\theta : \underline{h}(\theta) \le \delta\}$
sup-entropy rate	$\bar{H}(\boldsymbol{X}) := \lim_{\delta\uparrow 1} \bar{H}_\delta(\boldsymbol{X})$
inf-entropy rate	$\underline{H}(\boldsymbol{X}) := \underline{H}_0(\boldsymbol{X})$

Generalized entropy measures where $\delta \in [0, 1]$.

Generalized information measures

Mutual Information Measures	
system	arbitrary channel $P_{\boldsymbol{W}} = P_{\boldsymbol{Y}\vert\boldsymbol{X}}$ with input $\boldsymbol{X}$ and output $\boldsymbol{Y}$
norm. information density	$\frac{1}{n} i_{X^n W^n}(X^n; Y^n) := \frac{1}{n} \log \frac{P_{X^n,Y^n}(X^n, Y^n)}{P_{X^n}(X^n) \times P_{Y^n}(Y^n)}$
information sup-spectrum	$\bar{i}(\theta) := \limsup_{n\to\infty} \Pr\left\{\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \le \theta\right\}$
information inf-Spectrum	$\underline{i}(\theta) := \liminf_{n\to\infty} \Pr\left\{\frac{1}{n} i_{X^n W^n}(X^n; Y^n) \le \theta\right\}$
δ-inf-information rate	$\underline{I}_\delta(\boldsymbol{X}; \boldsymbol{Y}) := \sup\{\theta : \bar{i}(\theta) \le \delta\}$
δ-Sup-Information Rate	$\bar{I}_\delta(\boldsymbol{X}; \boldsymbol{Y}) := \sup\{\theta : \underline{i}(\theta) \le \delta\}$
sup-information rate	$\bar{I}(\boldsymbol{X}; \boldsymbol{Y}) := \lim_{\delta\uparrow 1} \bar{I}_\delta(\boldsymbol{X}; \boldsymbol{Y})$
inf-information rate	$\underline{I}(\boldsymbol{X}; \boldsymbol{Y}) := \underline{I}_0(\boldsymbol{X}; \boldsymbol{Y})$

Generalized mutual information measures where $\delta \in [0, 1]$.

Generalized information measures

Divergence Measures	
system	arbitrary sources $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$
norm. log-likelihood ratio	$\frac{1}{n} d_{X^n}(X^n \\| \hat{X}^n) := \frac{1}{n} \log \frac{P_{X^n}(X^n)}{P_{\hat{X}^n}(X^n)}$
divergence sup-spectrum	$\bar{d}(\theta) := \limsup_{n\to\infty} \Pr\left\{ \frac{1}{n} d_{X^n}(X^n \\| \hat{X}^n) \le \theta \right\}$
divergence inf-spectrum	$\underline{d}(\theta) := \liminf_{n\to\infty} \Pr\left\{ \frac{1}{n} d_{X^n}(X^n \\| \hat{X}^n) \le \theta \right\}$
δ-inf-divergence rate	$\underline{D}_\delta(\boldsymbol{X} \\| \hat{\boldsymbol{X}}) := \sup\{\theta : \bar{d}(\theta) \le \delta\}$
δ-sup-divergence rate	$\bar{D}_\delta(\boldsymbol{X} \\| \hat{\boldsymbol{X}}) := \sup\{\theta : \underline{d}(\theta) \le \delta\}$
sup-divergence rate	$\bar{D}(\boldsymbol{X} \\| \hat{\boldsymbol{X}}) := \lim_{\delta \uparrow 1} \bar{D}_\delta(\boldsymbol{X} \\| \hat{\boldsymbol{X}})$
inf-divergence rate	$\underline{D}(\boldsymbol{X} \\| \hat{\boldsymbol{X}}) := \underline{D}_0(\boldsymbol{X} \\| \hat{\boldsymbol{X}})$

Generalized divergence measures where $\delta \in [0, 1]$.

Properties of generalized information measures

- An example of basic properties for Shannon's entropy:
  $I(X;Y) = H(Y) - H(Y|X)$.

  – By taking $\delta = 0$ and letting $\gamma \downarrow 0$ in

  $$\underline{(U+V)}_{\delta+\gamma} \geq \underline{U}_\delta + \underline{V}_\gamma \text{ for } \delta \geq 0, \gamma \geq 0, \text{ and } \delta + \gamma \leq 1$$

  and

  $$\underline{(U+V)}_{\delta} \leq \underline{U}_{\delta+\gamma} + \bar{V}_{(1-\gamma)} \text{ for } \delta \geq 0, \gamma \geq 0, \text{ and } \delta + \gamma \leq 1,$$

  we obtain

  $$\underline{(U+V)} \geq \underline{U}_0 + \lim_{\gamma \downarrow 0} \underline{V}_\gamma \geq \underline{U} + \underline{V}$$

  and

  $$\underline{(U+V)} \leq \lim_{\gamma \downarrow 0} \underline{U}_\gamma + \lim_{\gamma \downarrow 0} \bar{V}_{(1-\gamma)} = \underline{U} + \bar{V}.$$

  – Meaning: The liminf in probability of a sequence of random variables $A_n + B_n$ is upper bounded by the liminf in probability of A_n plus the limsup in probability of B_n; and is lower bounded by the sum of the liminfs in probability of A_n and B_n.

Properties of generalized information measures

– This fact can be used to show that

$$\underline{I}(\boldsymbol{X};\boldsymbol{Y}) + \underline{H}(\boldsymbol{Y}|\boldsymbol{X}) \le \underline{H}(\boldsymbol{Y}) \le \underline{I}(\boldsymbol{X};\boldsymbol{Y}) + \bar{H}(\boldsymbol{Y}|\boldsymbol{X}),$$

or equivalently,

$$\underline{H}(\boldsymbol{Y}) - \bar{H}(\boldsymbol{Y}|\boldsymbol{X}) \le \underline{I}(\boldsymbol{X};\boldsymbol{Y}) \le \underline{H}(\boldsymbol{Y}) - \underline{H}(\boldsymbol{Y}|\boldsymbol{X}).$$

Lemma 1.5 For a source $\boldsymbol{X}$ with finite alphabet $\mathcal{X}$ and arbitrary sources $\boldsymbol{Y}$ and $\boldsymbol{Z}$, the following properties hold.

1. $\bar{H}_\delta(\boldsymbol{X}) \geq 0$ for $\delta \in [0,1]$. (This property also applies to $\underline{H}_\delta(\boldsymbol{X})$, $\bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y})$, $\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y})$, $\bar{D}_\delta(\boldsymbol{X}\|\hat{\boldsymbol{X}})$, and $\underline{D}_\delta(\boldsymbol{X}\|\hat{\boldsymbol{X}})$.)

2. $\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) = \underline{I}_\delta(\boldsymbol{Y};\boldsymbol{X})$ and $\bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) = \bar{I}_\delta(\boldsymbol{Y};\boldsymbol{X})$ for $\delta \in [0,1]$.

3. For $0 \leq \delta < 1$, $0 \leq \gamma < 1$ and $\delta + \gamma \leq 1$,
$$\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \leq \underline{H}_{\delta+\gamma}(\boldsymbol{Y}) - \underline{H}_\gamma(\boldsymbol{Y}|\boldsymbol{X}), \tag{1.4.1}$$
$$\underline{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \leq \bar{H}_{\delta+\gamma}(\boldsymbol{Y}) - \bar{H}_\gamma(\boldsymbol{Y}|\boldsymbol{X}), \tag{1.4.2}$$
$$\bar{I}_\delta(\boldsymbol{X};\boldsymbol{Y}) \leq \bar{H}_{\delta+\gamma}(\boldsymbol{Y}) - \underline{H}_\gamma(\boldsymbol{Y}|\boldsymbol{X}), \tag{1.4.3}$$
$$\underline{I}_{\delta+\gamma}(\boldsymbol{X};\boldsymbol{Y}) \geq \underline{H}_\delta(\boldsymbol{Y}) - \bar{H}_{(1-\gamma)}(\boldsymbol{Y}|\boldsymbol{X}), \tag{1.4.4}$$
and
$$\bar{I}_{\delta+\gamma}(\boldsymbol{X};\boldsymbol{Y}) \geq \bar{H}_\delta(\boldsymbol{Y}) - \bar{H}_{(1-\gamma)}(\boldsymbol{Y}|\boldsymbol{X}). \tag{1.4.5}$$
(Note that the case of $(\delta,\gamma) = (1,0)$ holds for (1.4.1) and (1.4.2), and the case of $(\delta,\gamma) = (0,1)$ holds for (1.4.3), (1.4.4) and (1.4.5).)

4. $0 \leq \underline{H}_\delta(\boldsymbol{X}) \leq \bar{H}_\delta(\boldsymbol{X}) \leq \log|\mathcal{X}|$ for $\delta \in [0,1)$, where each $X_i^{(n)}$ takes values in $\mathcal{X}$ for $i = 1,\ldots,n$ and $n = 1,2,\ldots$.

5. $\underline{I}_\delta(\boldsymbol{X},\boldsymbol{Y};\boldsymbol{Z}) \geq \underline{I}_\delta(\boldsymbol{X};\boldsymbol{Z})$ for $\delta \in [0,1]$.

Property 1:

$$
\begin{aligned}
\Pr\left\{-\frac{1}{n}\log P_{X^n}(X^n) < 0\right\} &= 0, \\
\Pr\left\{\frac{1}{n}\log\frac{P_{X^n}(X^n)}{P_{\hat{X}^n}(X^n)} < -\nu\right\} &= P_{X^n}\left\{x^n \in \mathcal{X}^n \ : \ \frac{1}{n}\log\frac{P_{X^n}(x^n)}{P_{\hat{X}^n}(x^n)} < -\nu\right\} \\
&= \sum_{\left\{x^n\in\mathcal{X}^n \ : \ P_{X^n}(x^n)<P_{\hat{X}^n}(x^n)e^{-n\nu}\right\}} P_{X^n}(x^n) \\
&\leq \sum_{\left\{x^n\in\mathcal{X}^n \ : \ P_{X^n}(x^n)<P_{\hat{X}^n}(x^n)e^{n\nu}\right\}} P_{\hat{X}^n}(x^n)e^{-n\nu} \\
&\leq e^{-n\nu}\cdot \sum_{\left\{x^n\in\mathcal{X}^n \ : \ P_{X^n}(x^n)<P_{\hat{X}^n}(x^n)e^{n\nu}\right\}} P_{\hat{X}^n}(x^n) \\
&\leq e^{-\nu n}, \qquad (1.4.6)
\end{aligned}
$$

and, by following the same procedure as (1.4.6),

$$
\Pr\left\{\frac{1}{n}\log\frac{P_{X^n,Y^n}(X^n,Y^n)}{P_{X^n}(X^n)P_{Y^n}(Y^n)} < -\nu\right\} \leq e^{-\nu n}.
$$

Properties of generalized information measures

Property 2: An immediate consequence of the definition.

Property 3: Follow from the facts that

$$\frac{1}{n}h_{Y^n}(Y^n) = \frac{1}{n}i_{X^n,Y^n}(X^n;Y^n) + \frac{1}{n}h_{X^n,Y^n}(Y^n|X^n),$$

where

$$\frac{1}{n}h_{X^n,Y^n}(Y^n|X^n) := -\frac{1}{n}\log P_{Y^n|X^n}(Y^n|X^n).$$

Property 4: $\bar{H}_\delta(\cdot)$ is non-decreasing in δ, $\bar{H}_\delta(\boldsymbol{X}) \leq \bar{H}(\boldsymbol{X})$, and $\bar{H}(\boldsymbol{X}) \leq \log|\mathcal{X}|$. The last inequality can be proved as follows:

$$\begin{aligned}
&\Pr\left\{\frac{1}{n}h_{X^n}(X^n) \leq \log|\mathcal{X}| + \nu\right\} | \\
&= 1 - P_{X^n}\left\{x^n \in \mathcal{X}^n \;:\; \frac{1}{n}\log\frac{P_{X^n}(X^n)}{1/|\mathcal{X}|^n} < -\nu\right\} \\
&\geq 1 - e^{-n\nu},
\end{aligned}$$

where the last step can be obtained by using the same procedure as (1.4.6). Therefore, $\underline{h}(\log|\mathcal{X}| + \nu) = 1$ for any $\nu > 0$, which indicates that $\bar{H}(\boldsymbol{X}) \leq \log|\mathcal{X}|$.

Properties of generalized information measures

Property 5:

$$\frac{1}{n} i_{X^n,Y^n,Z^n}(X^n, Y^n; Z^n) = \frac{1}{n} i_{X^n,Z^n}(X^n; Z^n) + \frac{1}{n} i_{X^n,Y^n,Z^n}(Y^n; Z^n | X^n).$$

□

Properties of generalized information measures

Lemma 1.6 (Data processing lemma) Fix $\delta \in [0,1]$. Consider a channel with finite input and output alphabets, $\mathcal{X}$ and $\mathcal{Y}$, respectively, and with distribution $P_{W^n}(y^n|x^n) = P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n P_{Y_i|X_i}(y_i|x_i)$ for all n, $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$. Then

$$\underline{I}_\delta(\boldsymbol{X}_1; \boldsymbol{X}_3) \le \underline{I}_\delta(\boldsymbol{X}_1; \boldsymbol{X}_2).$$

Proof: By Property 5 of Lemma 1.5, we get

$$\underline{I}_\delta(\boldsymbol{X}_1; \boldsymbol{X}_3) \le \underline{I}_\delta(\boldsymbol{X}_1; \boldsymbol{X}_2, \boldsymbol{X}_3) = \underline{I}_\delta(\boldsymbol{X}_1; \boldsymbol{X}_2),$$

where the equality holds because

$$\frac{1}{n} \log \frac{P_{X_1^n, X_2^n, X_3^n}(x_1^n, x_2^n, x_3^n)}{P_{X_1^n}(x_1^n) P_{X_2^n, X_3^n}(x_2^n, x_3^n)} = \frac{1}{n} \log \frac{P_{X_1^n, X_2^n}(x_1^n, x_2^n)}{P_{X_1^n}(x_1^n) P_{X_2^n}(x_2^n)}.$$

□

Lemma 1.7 (Optimality of independent inputs) Fix $\delta \in [0,1)$. Consider a finite-alphabet channel with $P_{W^n}(y^n|x^n) = P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n P_{Y_i|X_i}(y_i|x_i)$ for all n. For any input $\boldsymbol{X}$ and its corresponding output $\boldsymbol{Y}$,

$$\underline{I}_\delta(\boldsymbol{X}; \boldsymbol{Y}) \le \underline{I}_\delta(\bar{\boldsymbol{X}}; \bar{\boldsymbol{Y}}) = \underline{I}(\bar{\boldsymbol{X}}; \bar{\boldsymbol{Y}}),$$

where $\bar{\boldsymbol{Y}}$ is the output due to $\bar{\boldsymbol{X}}$, which is an independent process with the same first order statistics as $\boldsymbol{X}$, i.e., $P_{\overline{X}^n}(x^n) = \prod_{i=1}^n P_{X_i}(x_i)$.

Computation of general information measures

System setting:

- Let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and
$$Y_i^{(n)} = X_i^{(n)} \oplus Z_i^{(n)}$$
where the arbitrary noise $\boldsymbol{Z}$ is independent of the channel input $\boldsymbol{X}$.
- Assume that $\boldsymbol{X}$ is a Bernoulli uniform input, i.e., an i.i.d. random process with uniform marginal distribution.
- Then the resultant channel output $\boldsymbol{Y}$ is also Bernoulli uniform no matter what distribution $\boldsymbol{Z}$ has.

Computation of general information measures

Derivations:

$$
\begin{aligned}
\bar{i}(\theta) &:= \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log\frac{P_{Y^n|X^n}(Y^n|X^n)}{P_{Y^n}(Y^n)} \le \theta\right\} \\
&= \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log P_{Z^n}(Z^n) - \frac{1}{n}\log P_{Y^n}(Y^n) \le \theta\right\} \\
&= \limsup_{n\to\infty} \Pr\left\{\frac{1}{n}\log P_{Z^n}(Z^n) \le \theta - \log(2)\right\} \\
&= \limsup_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) \ge \log(2) - \theta\right\} \\
&= 1 - \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) < \log(2) - \theta\right\}.
\end{aligned}
$$

Computation of general information measures

Hence, for $\varepsilon \in (0, 1)$,

$$
\begin{aligned}
\underline{I}_{\varepsilon}(\boldsymbol{X};\boldsymbol{Y}) &= \sup\left\{\theta : \bar{i}(\theta) \le \varepsilon\right\} \\
&= \sup\left\{\theta : 1 - \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) < \log(2) - \theta\right\} \le \varepsilon\right\} \\
&= \sup\left\{\theta : \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) < \log(2) - \theta\right\} \ge 1 - \varepsilon\right\} \\
&= \sup\left\{(\log(2) - \beta) : \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) < \beta\right\} \ge 1 - \varepsilon\right\} \\
&= \log(2) + \sup\left\{-\beta : \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) < \beta\right\} \ge 1 - \varepsilon\right\} \\
&= \log(2) - \inf\left\{\beta : \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) < \beta\right\} \ge 1 - \varepsilon\right\} \\
&= \log(2) - \sup\left\{\beta : \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) < \beta\right\} < 1 - \varepsilon\right\} \\
&\le \log(2) - \sup\left\{\beta : \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) \le \beta\right\} < 1 - \varepsilon\right\} \\
&= \log(2) - \lim_{\delta\uparrow(1-\varepsilon)} \bar{H}_{\delta}(\boldsymbol{Z}).
\end{aligned}
$$

Also, for $\varepsilon \in (0,1)$,

$$\begin{aligned} \underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}) &\geq \sup\left\{\theta : \limsup_{n\to\infty} \Pr\left[\frac{1}{n}\log\frac{P_{X^n,Y^n}(X^n,Y^n)}{P_{X^n}(X^n)P_{Y^n}(Y^n)} < \theta\right] < \varepsilon\right\} \quad (1.5.3) \\ &= \log(2) - \sup\left\{\beta : \liminf_{n\to\infty} \Pr\left\{-\frac{1}{n}\log P_{Z^n}(Z^n) \leq \beta\right\} \leq 1-\varepsilon\right\} \\ &= \log(2) - \bar{H}_{(1-\varepsilon)}(\boldsymbol{Z}), \end{aligned}$$

where (1.5.3) follows from the fact described in Footnote 3. Therefore,

$$\log(2) - \bar{H}_{(1-\varepsilon)}(\boldsymbol{Z}) \leq \underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}) \leq \log(2) - \lim_{\gamma\uparrow(1-\varepsilon)} \bar{H}_\gamma(\boldsymbol{Z}) \quad \text{for } \varepsilon \in (0,1).$$

By taking $\varepsilon \downarrow 0$, we obtain

$$\underline{I}(\boldsymbol{X};\boldsymbol{Y}) = \underline{I}_0(\boldsymbol{X};\boldsymbol{Y}) = \log(2) - \bar{H}(\boldsymbol{Z}).$$

Based on this result, we can now compute $\underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y})$ for some specific examples.

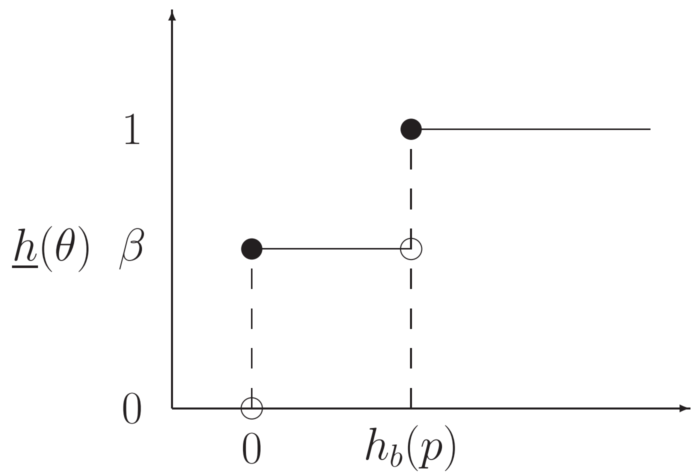
Example 1.8

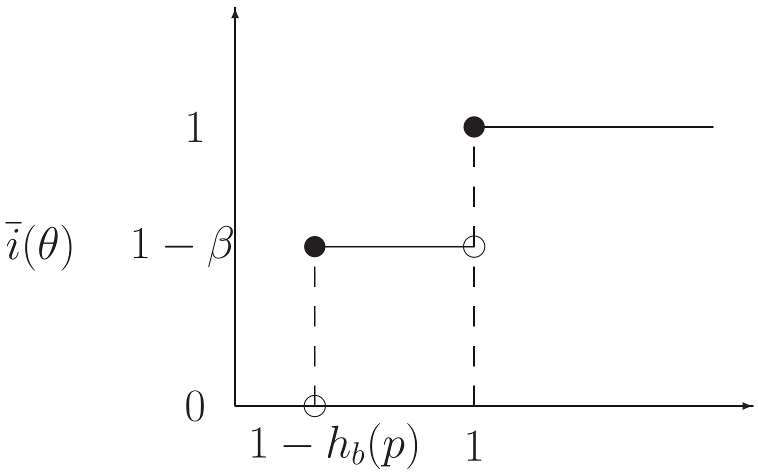
$$\boldsymbol{Z} = \begin{cases} \text{all-zero sequence with probability } \beta; \\ \text{Bernoulli (with parameter } p\text{) with probability } 1-\beta. \end{cases}$$

Then

$$\frac{1}{n} h_{Z^n}(Z^n) \to \begin{cases} 0, & \text{with probability } \beta; \\ h_b(p), & \text{with probability } 1-\beta, \end{cases}$$

where $h_b(p) := -p\log p - (1-p)\log(1-p)$.





Therefore,

$$\underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}) = \begin{cases} 1-h_b(p), & \text{if } 0 < \varepsilon < 1-\beta; \\ 1, & \text{if } 1-\beta \le \varepsilon < 1. \end{cases}$$

Example 1.9 $\boldsymbol{Z}$ =non-stationary binary independent sequence with

$$\Pr\left\{Z_i^{(n)}=0\right\}=1-\Pr\left\{Z_i^{(n)}=1\right\}=p_i,$$

then by the fact that

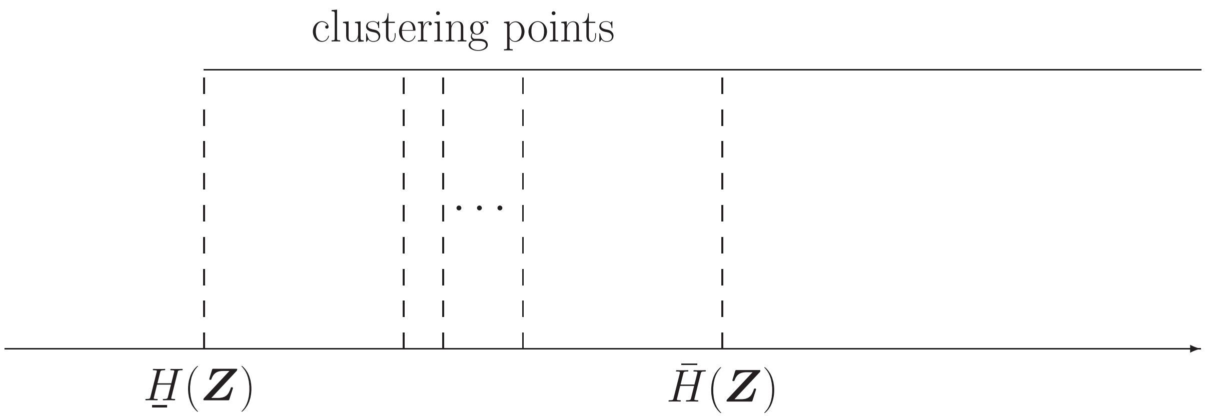
$$\begin{aligned}\mathrm{Var}\left[-\log P_{Z_i^{(n)}}(Z_i^{(n)})\right] &\le E\left[\left(\log P_{Z_i^{(n)}}(Z_i^{(n)})\right)^2\right]\\ &\le \sup_{0<p_i<1}\left[p_i(\log p_i)^2+(1-p_i)(\log(1-p_i))^2\right]\\ &< \log(2),\end{aligned}$$

we have (by Chebyshev's inequality) that as $n\to\infty$,

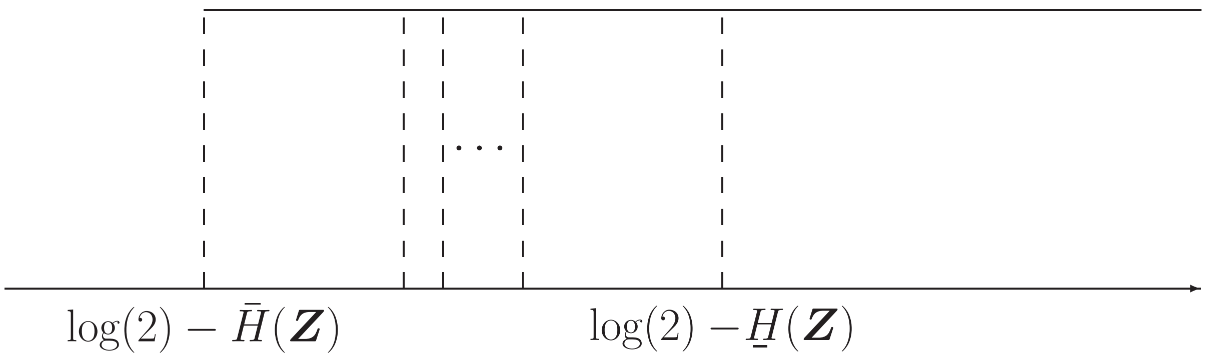
$$\Pr\left\{\left|-\frac{1}{n}\log P_{Z^n}(Z^n)-\frac{1}{n}\sum_{i=1}^{n}H\left(Z_i^{(n)}\right)\right|>\gamma\right\}\to 0,$$

for any $\gamma>0$.

Computation of general information measures



The limiting spectrum of $(1/n)h_{Z^n}(Z^n)$.



The possible limiting spectrums of $(1/n)i_{X^n,Y^n}(X^n;Y^n)$.

Computation of general information measures

Therefore, $\bar{H}_{(1-\varepsilon)}(\boldsymbol{Z})$ is equal to

$$\bar{H}_{(1-\varepsilon)}(\boldsymbol{Z}) = \begin{cases} \begin{aligned} \bar{H}(\boldsymbol{Z}) &= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} H\left(Z_i^{(n)}\right) \\ &= \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} h_b(p_i) \end{aligned}, & \text{for } \varepsilon \in (0,1]; \\ +\infty, & \text{for } \varepsilon = 0. \end{cases}$$

Consequently,

$$\underline{I}_\varepsilon(\boldsymbol{X};\boldsymbol{Y}) = \begin{cases} 1 - \bar{H}(\boldsymbol{Z}) = 1 - \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} h_b(p_i), & \text{for } \varepsilon \in [0,1), \\ \infty, & \text{for } \varepsilon = 1. \end{cases}$$