Measure of Randomness for Stochastic Processes

Chapter 4
In this chapter, we will explore the new concept in details.

Another operational meaning, which is called resolvability, has the minimum lossless data compression ratio, the sup-entropy rate actually has another operational work of Verdú and Han in 1993. They found that, other than optimality of the designed block code.

Hence, to find an optimal block code becomes a well-defined mission since for any source with well-formulated statistical model, the sup-entropy rate can be computed and such quantity can be used as a criterion to evaluate the optimality of the designed block code.

In the previous chapter, it is shown that the sup-entropy rate is indeed the minimum lossless data compression ratio achievable for block codes.
Motivation for Resolvability

In simulations of statistical communication systems, generation of random variables by computer algorithms is very essential. The computer usually has an access to an equally likely random experiment that the most efficient algorithm requires in order to generate a random variable, which is defined as 

\[ \text{rand}() \]

throughout (0, 1).

The computer usually has an access to a basic random experiment (through a pre-defined Application Programming Interface), which generates equally likely values, such as \( \text{rand}() \), such that they are real numbers uniform random values. Hence, the computer algorithm is very essential.

In simulations of statistical communication systems, generation of random variables by computer algorithms is very essential.

Preliminary solution: One possible way to define such “complexity” measurement is:

Definition 4.1. The complexity of generating a random variable is defined as the number of random bits that the most efficient algorithm requires in order to generate the random variable with a computer that has an access to equally likely random experiments.

Conceptually, random variables with complex models are more difficult to generate than random variables with simple models. Conceptually, random variables with complex models are more difficult to generate than random variables with simple models.
Example 4.2

Consider the generation of the random variable with probability masses $P_X(1) = 1/4$, $P_X(0) = 1/2$, and $P_X(-1) = 1/4$. An algorithm is written as:

\[
\begin{align*}
\text{Flip-a-fair-coin; } & \text{ one random bit} \\
\text{If "Head", then output } & -1; \\
\text{else output } & 1; \\
\text{Flip-a-fair-coin; } & \text{ one random bit} \\
\end{align*}
\]

\text{average-case: the above algorithm requires 1.5 coin flips;}

\text{\hspace{1cm} worst-case: 2 coin flips are necessary.}

Therefore, the complexity measure can take two fundamental forms: worst-case or average-case over the range of outcomes of the random variables.

Example 4.2

Consider the generation of the random variable with probability masses $P_X(1) = 1/4$, $P_X(0) = 1/2$, and $P_X(-1) = 1/4$. An algorithm is written as:

\[
\begin{align*}
\text{Flip-a-fair-coin; } & \text{ one random bit} \\
\text{If "Head", then output } & -1; \\
\text{else output } & 1; \\
\text{Flip-a-fair-coin; } & \text{ one random bit} \\
\end{align*}
\]
• Note that we did not show in the above example that the algorithm is the most efficient one in the sense of using minimum number of random bits; however, it is indeed an optimal algorithm because it achieves the lower bound of the average minimum number of random bits. Later, we will show that such bound for generating the random variables is the entropy, which is exactly $1.5$ bits in the above example.

• As for the worse-case bound, a new terminology, resolution, will be introduced. As a result, the above algorithm also achieves the lower bound of the worst-case complexity, which is the resolution; an entropy, which is exactly $1.5$ bits in the above example.
Notations and definitions regarding resolvability

II: 4-5

Definition 4.3 (M-type)
For any positive integer \( M \), a probability distribution \( P \) is said to be \( M \)-type if \( P(\omega) \in \{0, \frac{M}{2}, \frac{M}{2^2}, \ldots, \frac{M}{2^n}, \frac{M}{2^n+1}, \ldots, 1\} \) for all \( \omega \in \Omega \).

Definition 4.4 (resolution of a random variable)
The resolution \( R(X) \) of a random variable \( X \) is the minimum \( M \) such that \( P_X \) is \( M \)-type. If \( P_X \) is not \( M \)-type for any integer \( M \), then \( R(X) = \infty \).

If the base of the logarithmic operation is 2, the resolution is measured in bits; however, if natural logarithm is taken, nats becomes the basic measurement.

• If the base of the logarithmic operation is 2, then

\[
\begin{align*}
\text{resolution of a random variable, } \log_2 \left( \frac{M}{2^n} \right) & \text{ is the minimum integer } n,
\end{align*}
\]

\( n \)-type for any integer \( M \), then \( R(X) = \infty \).

\[ R(X) = \text{resolution of a random variable} \]

\[ \text{for all } \omega \in \Omega, \quad \left\{ \frac{M}{2^n}, \frac{M}{2^n+1}, \ldots, 1 \right\} \in (\rho) \]

\( d \) is said to be \( M \)-type if \( P_X \) for any positive integer \( M \), a probability distribution \( \frac{M}{2^n} \).
As revealed previously, a random source needs to be resolved (meaning, can be generated by a computer algorithm with access to equal-probable random experiments).

As anticipated, a random variable with finite resolution is resolvable by computer algorithms.

Yet, it is possible that the resolution of a random variable is infinity.

A quick example is the random variable $X$ with distribution $P_X(0) = \frac{1}{\pi}$ and $P_X(1) = 1 - \frac{1}{\pi}$.

$X$ is 3-type, which is computer-resolvable.

With distribution $\tilde{X}$, then

$\exists \epsilon > 0$ and $\delta > 0$

such that

$||X - \tilde{X}|| < \epsilon$.

In such case, one can alternatively choose another computer-resolvable random variable, which resembles the true source within some acceptable range, to simulate the original one.

One criterion that can be used as a measure of resemblance of two random variables is the variational distance.

Yet, it is possible that the resolution of a random variable is infinity.

As anticipated, a random variable with finite resolution is resolvable by computer algorithms.

As revealed previously, a random source needs to be resolved (meaning, can be generated by a computer algorithm with access to equal-probable random experiments).
A program that generates the 3-type $\tilde{X}$ is as follows (in C language):

```c
{ 
    even = False;
    while (1) {
        Flip a fair coin;
        one random bit
        if (Head) {
            if (even == True) { output 0; even = False; }
            else { output 1; even = True; }
        }
        else { (if (Head) Print even;)
            output { even == True; output 0; even = False; }
        }
    }
}
```

A program that generates the 3-type $X$ is as follows (in C language):

```c
{ 
    if (even == True) { output 0; even = False; }
    else { output 1; even = True; }
}
```

A program that generates $W$-type random variable for any $W$ satisfying $W > 0$ is straightforward.

Physical meaning

II: 4.7
Definition 4.5 (variational distance)
The variational distance (or $\ell_1$ distance) between two distributions $P$ and $Q$ defined on common measurable space $(\Omega, F)$ is $\|P - Q\|_\Delta = \sum_{\omega \in \Omega} |P(\omega) - Q(\omega)|$.

(Note that an alternative way to formulate the variational distance is:
$$\|P - Q\| = 2 \cdot \sup E \in F |P(E) - Q(E)| = 2 \sum \{ x \in X : P(x) \geq Q(x) \} [P(x) - Q(x)]$$)

This two definitions are actually equivalent.

Definition 4.6 ($\epsilon$-achievable resolution)
Fix $\epsilon \geq 0$. $R$ is an $\epsilon$-achievable resolution for input $X$ if for all $\gamma > 0$, there exists $\tilde{X}$ satisfies $\tilde{X} \leq X$ and $R(\tilde{X}) < R + \gamma$ and $\|X - \tilde{X}\| < \epsilon$.

$\epsilon$-achievable resolution reveals the possibility that one can choose another computer-resolvable random variable whose variational distance to the true source is within an acceptable range, $\epsilon$.

$\epsilon > \|X - X\| \quad \text{and} \quad \nu + H > (X)H$

Resolution for input $X$, if for all $\gamma > 0$, there exists satisfies $\tilde{X}$ and $\epsilon < 0$. $R$ is an $\epsilon$-achievable resolution for input $X$.

The variational distance (or $\ell_1$ distance) between two distributions $\mathcal{D}$ and $\mathcal{D}$ defined on common measurable space $\mathcal{D}$ and $\mathcal{D}$ is $\|\mathcal{D} - \mathcal{D}\|$.
Next we define the \( \varepsilon \)-achievable resolution rate for a sequence of random variables, which is an extension of \( \varepsilon \)-achievable resolution defined for single random variable. Such extension is analogous to extending entropy for a single source to entropy rate for a random variable.

\[
\exists \varepsilon > 0 \text{ such that } \forall n, \exists \tilde{X} \text{ satisfying } \|X_n - \tilde{X}_n\| < \varepsilon \text{ and } R > \frac{H(X)}{n}
\]

Definition 4.7 (\( \varepsilon \)-achievable resolution rate)

Fix \( \varepsilon \geq 0 \) and input \( X \). \( R \) is an \( \varepsilon \)-achievable resolution rate for input \( X \) if for every \( \gamma > 0 \), there exists \( \tilde{X} \) satisfies for all sufficiently large \( n \).

Next we define the \( \varepsilon \)-achievable resolution rate for a sequence of random variables, which is an extension of \( \varepsilon \)-achievable resolution defined for single random variable.
Similar convention will be applied throughout the rest of this chapter.

\[
\{ \exists > \|u\tilde{X} - uX\| \text{ and } \exists + \gamma > (u\tilde{X})_U^u \}
\]

\[
(N < uA)(N \text{ and } X \in) (0 < \gamma A) : \gamma \} \in (X, \delta)_{\exists} S
\]

minimum operation, i.e.,

Here, we define \( (X, \delta)_{\exists} S \) using the "minimum" instead of a more general "infinite" minimum operation is simply because \( (X, \delta)_{\exists} S \) indeed belongs to the range of the same input, i.e.,

\[
\{ \exists > \|u\tilde{X} - uX\| \text{ and } \exists + \gamma > (u\tilde{X})_U^u \}
\]

\[
(N < uA)(N \text{ and } X \in) (0 < \gamma A) : \gamma \} \in (X, \delta)_{\exists} S
\]

Definition 4.8 \( \varepsilon \)-resolvability for \( X \)

Fix \( \varepsilon > 0 \). The \( \varepsilon \)-resolvability for \( X \) is denoted by \( (X, \delta)_{\exists} S \).
Definition 4.9 (resolvability for $X$)

The resolvability for input $X$, denoted $S(X)$, is

$$S(X) = \lim_{\varepsilon \to 0} S_{\varepsilon}(X)$$

From the definition of $\varepsilon$-resolvability, it is obvious non-increasing in $\varepsilon$. Hence,

$$S(X) = \sup_{\varepsilon > 0} S_{\varepsilon}(X)$$

The resolvability for input $X$, denoted $S(X)$, is $\varepsilon$-resolvability. 

$\frac{(X)}{(X)^{\varepsilon} S_{\varepsilon}} \vDash (X) S$

$\lim_{\varepsilon \to 0} (X)^{\varepsilon} S_{\varepsilon} \vDash (X) S$

By $S(X)$, $S(X)$ is $\varepsilon$-resolvability. 

Definition 4.9 (resolvability for $X$) 

The resolvability for input $X$, denoted $S(X)$, is $\varepsilon$-resolvability.
The resolvability is pertinent to the worse-case complexity measure for random variables (cf. Example 4.2, and the discussion following it).

With the entropy function, the information theorists also define the \( \varepsilon \)-mean-resolvability and mean-resolvability for input \( X \), which characterize the average-case complexity of random variables. With the entropy function, the information theorists also define the \( \varepsilon \)-mean-achievability rate for input \( X \) which characterize the average-case complexity of random variables (cf. Example 4.2, and the discussion following it).

**Definition 4.10** (\( \varepsilon \)-mean-achievable resolution rate)

Fix \( \varepsilon \geq 0 \).

\( R \) is an \( \varepsilon \)-mean-achievable resolution rate for input \( X \) if for all \( \gamma > 0 \), there exists \( \tilde{X} \) satisfies

\[
\left\{ \begin{array}{l}
\varepsilon > \|uX - u\tilde{X}\| \quad \text{and} \quad \gamma + H(u\tilde{X}) - I(u) \\
(N < u A)(N \text{ and } \tilde{X} \in (0 < \lambda A) : \gamma )\end{array} \right\} \min_\gamma \equiv (X)^2 S
\]

for all sufficiently large \( n \).

**Definition 4.11** (\( \varepsilon \)-mean-resolvability for \( X \))

Fix \( \varepsilon > 0 \). The \( \varepsilon \)-mean-resolvability for input \( X \), denoted by \( S(X) \), is the minimum \( \varepsilon \)-mean-achievable resolution rate for the same input, i.e.,

\[
\bar{S}_\varepsilon(X) \triangleq \min \left\{ R : (\forall \gamma > 0) \exists \tilde{X} \text{ and } \bar{N} \forall n > N \right\}
\]
The mean-resolvability for input $X$, denoted by $\bar{S}(X)$, is

$$\bar{S}(X) = \lim_{\varepsilon \to 0} \bar{S}_\varepsilon(X) = \sup_{\varepsilon > 0} \bar{S}_\varepsilon(X).$$

The only difference between resolvability and mean-resolvability is that the former employs resolution function, while the latter replaces it by entropy function.

Since entropy is the minimum average codeword length for uniquely decodable codes, an explanation for mean-resolvability is that the new random variable $\tilde{X}$ can be resolvable through realizing the optimal variable-length code for it.

You can think of the probability mass of each outcome of each $j \in \mathbb{Z} - \{0\}$, where $j$ is

$$\nu_j \left( \tilde{X} \right)$$

You can think of the probability mass of each outcome of each $j \in \mathbb{Z}$ where $j$ is

The only difference between resolvability and mean-resolvability is that the

$$(X)_{\varepsilon}^{\sup} = (X)_{\varepsilon}^{\inf} \Rightarrow (X)_{\varepsilon}$$

put $X$, denoted by $\mathcal{S}(X)$, is

\text{Definition 4.12 (mean-resolvability for input $X$)}

\text{mean-resolvability}
The operational meanings for the resolution and entropy (a new operational meaning for entropy other than the one from source coding theorem) follow the next theorem.

**Theorem 4.13**

For a single random variable $X$, the worse-case complexity is lower-bounded by its resolution $R(X)$ [Han and Verdú 1993]; and is upper-bounded by entropy $H(X)$ plus 2 bits [Knuth and Yao 1976].

Next, we reveal the operational meanings for resolvability and mean-resolvability in source coding.

**Lemma 4.14 (bound on variational distance)**

For every $\mu > 0$, $\| P - Q \| \leq \frac{2}{\mu} + 2 \cdot \sum_{x \in X} \log \frac{P(x)}{Q(x)} > \mu$.

Characterizing the resolvability.

Verdù [1993]: The worse-case complexity is lower-bounded by its resolution $R(X)$ for a single random variable $X$.
\[
\begin{align*}
\left( [(x)\partial - (x)d] \right) & \quad \left( \sum_{\{0 \leq [(x)\partial/(x)d]^{[x]} : x \in X\}} \right) + \\
& \left( [(x)(\partial - (x)d) \right) \quad \left( \sum_{\{\mu < [(x)\partial/(x)d]^{[x]} : x \in X\}} \right) \quad \exists \quad \therefore =
\end{align*}
\]

\[
\begin{align*}
& \left( [(x)\partial - (x)d] \right) \quad \left( \sum_{\{0 \leq [(x)\partial/(x)d]^{[x]} : x \in X\}} \right) \\
& \left( [(x)\partial - (x)d] \right) \quad \left( \sum_{\{\mu < [(x)\partial/(x)d]^{[x]} : x \in X\}} \right) \quad \exists \quad \therefore =
\end{align*}
\]

\[
\begin{align*}
& \left( [(x)\partial - (x)d] \right) \quad \left( \sum_{\{\mu \geq \log[(x)\partial/(x)d]^{[x]} : x \in X\}} \right) \\
& \left( [(x)\partial - (x)d] \right) \quad \left( \sum_{\{\mu \geq \log[(x)\partial/(x)d]^{[x]} : x \in X\}} \right) \quad \exists \quad \therefore = \left\| \partial - d \right\|
\end{align*}
\]

**Proof:**

The meaning of resolvability as mean-resolvability.
\[
\left( \sum_{x \in X : \log \frac{P(x)}{Q(x)} > \mu} P(x) \right) + \sum_{x \in X : \mu \geq \log \frac{P(x)}{Q(x)} \geq 0} P(x) \leq 2 \left( \sum_{x \in X : \log \frac{P(x)}{Q(x)} > \mu} P(x) \right) \]

(by fundamental inequality)
The lemma then holds:

\[(uX)^{\mathcal{H}} \supseteq (uX)^{uX} d \supseteq \frac{u}{1} \]

For all \( uX \in uX \). Hence, for all \( uX \in uX \).

\[
\{ (uX)^{\mathcal{H} -} \} \exists x \in \supseteq (uX)^{uX} d
\]

\( (uX)^{\mathcal{H}} \)

\[\text{Proof: By definition of } \mathcal{H}\]

For every \( n \).

\[
I = \left\{ (uX)^{\mathcal{H}} \supseteq (uX)^{uX} d \supseteq \frac{u}{1} : uX \in uX \right\}^{uX} d
\]

Lemma 4.15

\[
\left( n + \left[ n < \frac{(x)O \supseteq \mathcal{O}}{(x)d \supseteq \mathcal{O}} : x \in x \right] d \right) \exists =
\left( \left\{ 0 < \frac{(x)O \supseteq \mathcal{O}}{(x)d \supseteq \mathcal{O}} \subseteq n : x \in x \right\} d \right) \exists =
\]

Olpe. meanings of resolvability & mean-resolvability
Theorem 4.16

The resolvability for input $X$ is equal to its sup-entropy rate, i.e.,

$$S(X) = \bar{H}(X).$$

Proof:

1. By definition of $\bar{H}(X)$,

$$\limsup_{n \to \infty} P_X^n(D_0) > 0.$$

2. Let $\varnothing + (X)S < (u^x)u^x d \log \frac{u}{1 - u^x} \ni u^x \in u^x \} \equiv 0^d$

3. Suppose $(X)H > (X)S$.

Then there exists $\varnothing < \varnothing$ such that $(X)H > (X)S$.

It suffices to show that contradicts to Lemma 4.15.

\[ (X)H \geq (X)S. \]

\[ (X)H = (X)S. \]

The resolvability for input $X$ is equal to its sup-entropy rate, i.e., the resolvability.
\[
\{ (u_x)^u X d \ : \ \exists \ \gamma \geq |(u_x)^u X d - (u_x)^u X d| \text{ and } \\
0 < (u_x)^u X d : u_x \ni u_x \} \equiv A
\]

For sufficiently large \( n \), define

\[
\exists > \| u_x \sim - u_x \| \quad \text{and} \quad \frac{\gamma}{\varphi} + (x)^s > (u_x)^H u
\]

Select \( \varepsilon > 0 \) such that

\[
\{ (x)^s \geq (x)^2 s \text{ and observe that } \{ 1, 2, \cdots \} \text{ is infinitely often in } n. \}
\]

Resolvability
\[(0_x)^{uX}d (\exists \wedge - 1) \geq (0_x)^{uX}d\]

which holds infinitely often in \(n\) and every 0\(\mathcal{A}\) \(\cup 1\mathcal{A}\) in \(0_x\) satisfies

\[
\{0 < \exists \wedge - \alpha \geq (1/\sqrt{\epsilon})\} \leq (0\mathcal{A} \cup 1\mathcal{A})^{uX}d
\]

Consider that

\[
\exists \wedge = \frac{\exists \wedge}{\exists} >
\]

\[
\{(u_x)^{uX}d - (u_x)^{uX}d \}^{\exists \wedge > (u_x)^{uX}d : uX \in uX}\} =
\]

\[
\{(u_x)^{uX}d \cdot \exists \wedge < \{(u_x)^{uX}d - (u_x)^{uX}d \} : uX \in uX\}^{uX}d =
\]

\[
\{(u_x)^{uX}d \cdot \exists \wedge < \{(u_x)^{uX}d - (u_x)^{uX}d \} : uX \in uX\}^{uX}d +
\]

\[
\{0 = (u_x)^{uX}d : uX \in uX\}^{uX}d \geq
\]

\[
\{(u_x)^{uX}d \cdot \exists \wedge < \{(u_x)^{uX}d - (u_x)^{uX}d \} : 0\}^{uX}d = (1\mathcal{A})^{uX}d
\]

Then

Resolvability
which contradicts the result of Lemma 4.15.

Therefore, for those \( n \) that (4.3.1) holds, \( P_{\tilde{X}_n} \{ x_n \in X_n : -\frac{1}{n} \log P_{\tilde{X}_n}(x_n) > R(\tilde{X}_n) \} \geq P_{\tilde{X}_n} \{ x_n \in X_n : -\frac{1}{n} \log P_{\tilde{X}_n}(x_n) > S(X) + \delta \} \geq 0 \), which contradicts to the result of Lemma 4.15.
It suffices to show the existence of arbitrary $\mathbf{X}$ such that

$$(\mathbf{X})_H \succeq (\mathbf{X})_S.$$
Resolvability

For each $G$ chosen, we obtain from Lemma 4.14 that

$$\|X_n - \tilde{X}_n\| \leq 2\mu + 2 \cdot P_{\tilde{X}_n}(x_n \in X_n) \log P_X(x_n) > \mu$$

(since $P_{\tilde{X}_n}(G^c) = 0$)

$$= 2\mu + P_{\tilde{X}_n}\{x_n \in G : \frac{1}{n} \log P_X(x_n) > \bar{H}(X) + \gamma + \mu\}$$

Since $G$ is chosen randomly, we can take the expectation values (w.r.t. the random $G$) of the above inequality to obtain:

$$\mathbb{E}_{G} [\|X_n - \tilde{X}_n\|] \leq 2\mu + \mathbb{E}_{G} \left\{ \frac{W}{I} + n\zeta \right\}$$

Observe that each $U_j$ is either in or not in $G$, and will contribute weight $1/M$ when it is in $G$. From the i.i.d. assumption of $\mathbb{I}_{\{\Omega\}_I}$, we can then

$$\cdot |G \cup B| \frac{W}{I} + n\zeta \geq \left[ \|uX - uX\| \right] \mathbb{E}$$

Since $G$ is chosen randomly, we can take the expectation values (w.r.t. the random $G$) of the above inequality to obtain:

$$\mathbb{E}_{G} [\|X_n - \tilde{X}_n\|] \leq 2\mu + \mathbb{E}_{G} \left\{ \frac{W}{I} + n\zeta \right\}$$

where

$$(0 = (G \cup B)^{uX}d + n\zeta \geq$$

$$\left\{ \frac{u}{n} + \zeta + (X)H < (uX)^{uX}d \quad \forall \left\{ \frac{u}{n} \right\} : B \ni u, \; (uX)^{uX}d \geq \bar{H}(X) + \gamma + \mu \right\}$$

$$(= (G \cup B)^{uX}d)$$

since

$$n \leq \frac{(uX)^{uX}d \quad \forall \left\{ \frac{u}{n} \right\} : B \ni u, \; (uX)^{uX}d \geq \bar{H}(X) + \gamma + \mu \right\}$$

$$\geq \frac{uXd \cdot \zeta + n\zeta}{\|uX - uX\|}$$

For each chosen, we obtain from Lemma 4.14 that

$$\mathbb{E}_{G} [\|X_n - \tilde{X}_n\|] \leq 2\mu + \mathbb{E}_{G} \left\{ \frac{W}{I} + n\zeta \right\}$$
tence of the desired \( X \). Since can be chosen arbitrarily small, therefore guarantees the exis-

\[
0 = \left[ \| uX - uX \| \right] \delta A \limsup_{\| uX - uX \|} = \left[ \| uX - uX \| \right] \delta A \limsup_{\| uX - uX \|} \leq \text{by Equation (3.2)}
\]

which implies \( n \zeta = [\mathcal{A}]_{uX}^d \limsup_{\| uX - uX \|} n \zeta \geq \left[ \| uX - uX \| \right] \delta A \limsup_{\| uX - uX \|} \leq \text{by Equation (3.2)}
\]

\[\text{Hence,} \quad [\mathcal{A}]_{uX}^d = \left( [\mathcal{A}]_{uX}^d I \right) \frac{W}{I} = \left( [\mathcal{A}]_{uX}^d I - [\mathcal{A}]_{uX}^d I - W \right) \frac{W}{I} = [\mathcal{A}]_{uX}^d [\mathcal{A}]_{uX}^d \frac{W}{I} + \cdots + [\mathcal{A}]_{uX}^d [\mathcal{A}]_{uX}^d \frac{W}{I} + [\mathcal{A}]_{uX}^d \frac{W}{I} \]

evaluate \( \| \mathcal{A} \cup \delta \| \delta A \frac{W}{I} \) by \( \| \mathcal{A} \cup \delta \| \delta A (W / I) \)
Theorem 4.17

For any $\mathbf{X}$, $\overline{S}(\mathbf{X}) = \limsup_{n \to \infty} \frac{1}{n} H(X^n)$.

Proof:

1. For sufficiently large $u$, there exists such that $0 < \zeta$ for any $\varepsilon > 0$. Then for any $u$, observe that for any $(\mathbf{X})^\varepsilon S \preceq (\mathbf{X})S$ we have

\[ (uX)H(u/1) \preceq (\mathbf{X})S. \]

2. Let

\[ 0 = \|uX - u\mathbf{X}\| \quad \text{and} \quad \zeta + (uX)H(u/1) \sup_{I} \limsup_{n \to \infty} \frac{1}{n} H(X^n) > (\mathbf{X})H(u/1) \]

for sufficiently many $u$.

This can be trivially achieved by letting $\mathbf{X} = \mathbf{X}$, since $\overline{S}(\mathbf{X}) \geq \overline{S}(\mathbf{X})^\varepsilon$ for any $0 < \varepsilon < 1/2$. Then for any $\gamma > 0$ and all sufficiently large $n$, there exists $\tilde{X}$ such that $\tilde{X} - X^n < \varepsilon$, for every $\varepsilon > 0$. This is equivalent to show that for all such that $0 < \zeta$ for every $\varepsilon \preceq (\mathbf{X})H(u/1) \sup_{I} \limsup_{n \to \infty} \|X^n - \tilde{X}\| \geq (\mathbf{X})S$

\[ (uX)H(u/1) \sup_{I} \limsup_{n \to \infty} \frac{1}{n} H(X^n) = (\mathbf{X})S. \]

Theorem 4.17 For any $\mathbf{X}$

Mean-resolvability
\[ (uX)^{H - 1}_I \sup_{\|X\| \leq \varepsilon} \leq (X)_{\varepsilon} \]

Since \( \varepsilon \) and can be taken arbitrarily small, we have

\[ \varepsilon + (X)_{\varepsilon} > |X| \sup_{\|X\| \leq \varepsilon} \|X\| - (uX)^{H - 1}_I \sup_{\|X\| \leq \varepsilon} \]

which implies that

\[ \varepsilon + (X)_{\varepsilon} > (uX)^{H - 1}_I \sup_{\|X\| \leq \varepsilon} \|X\| - (uX)^{H - 1}_I \]

and (4.3.3) we obtain

\[ \frac{\varepsilon}{u|X|} \sup_{\|X\| \leq \varepsilon} \geq |(uX)^{H - (uX)|} \]

Using the fact [1], pp. 333 that \( \varepsilon \geq \|uX - uX\| \)

\[ \varepsilon > \|uX - uX\| \]

and

\[ \overline{\text{Mean-resolvability}} \]
In the previous chapter, we have proved that the lossless data compression rate for block codes is lower bounded by \( H(X) \).

We also show that \( H(X) \) is also the resolvability for source \( X \).

Thus, if we can find such a typical set, the Shannon’s source coding theorem is actually the existence of a typical set for block codes.

The key to the Shannon’s source coding theorem is actually the existence of a data compression rate for block codes.

We can therefore conclude that resolvability is equal to the minimum lossless data compression rate for block codes.

We also show that \( H(X) \) is also the resolvability for source \( X \).

For block codes is lower bounded by \( H(X) \).

In the previous chapter, we have proved that the lossless data compression rate

Resolvability and source coding

Furthermore, extension of the theorems to codes of some specific types becomes feasible.

For block codes can actually be generalized to more general sources, such as non-stationary sources.
\[ R = \limsup_{n \to \infty} \frac{1}{n} \log |A_n| \leq R \]

where \( R \) is the average codeword length of all such length codes if there exists a sequence of error-free prefix codes such that \( H \) is an achievable source compression rate for variable-length codes.

**Definition 4.18** (minimum \( \varepsilon \)-source compression rate for fixed-length codes)

\[ T_{\varepsilon}(X) = \min \{ R \} \]

\[ T_{\varepsilon}(X) = \lim_{\varepsilon \to 0} T_{\varepsilon}(X) \]

\[ T_{\varepsilon}(X) = \min \{ R \} \]

Note that the definition of \( T_{\varepsilon}(X) \) is equivalent to the one in Definition 3.2.

**Definition 4.19** (minimum source compression rate for fixed-length codes)

\[ T(X) = \lim_{\varepsilon \to 0} T_{\varepsilon}(X) \]

\[ T(X) = \min \{ R \} \]

where \( T(X) \) represents the minimum source compression rate for fixed-length codes, which is defined as:

**Definition 4.20** (minimum source compression rate for variable-length codes)

\[ R \] is an achievable source compression rate for variable-length codes if there exists a sequence of error-free prefix codes such that

\[ H \geq \limsup_{n \to \infty} \frac{1}{n} \log |C_n| \]

where \( \bar{C}_n \) is the average codeword length of all such codes.

**Resolvability and source coding**
Recall that for a single source, the measure of its uncertainty is entropy. Although the entropy can also be used to characterize the overall uncertainty of a random sequence $X$, the source coding however concerns more on the "average" entropy of it.

So far, we have seen four expressions of "average" entropy:

$$\begin{align*}
\bar{H}(X) &= \inf_{\beta \in \mathbb{R}} \{ \beta : \lim \sup_{n \to \infty} \frac{1}{n} \log P_X(X^n) > \beta \} \\
H(X) &= \sup_{\alpha \in \mathbb{R}} \{ \alpha : \lim \sup_{n \to \infty} \frac{1}{n} \log P_X(X^n) < \alpha \}
\end{align*}$$

If

$$\lim_{n \to \infty} \frac{1}{n} H(X^n) = \lim \sup_{n \to \infty} \frac{1}{n} H(X^n) = \lim \inf_{n \to \infty} \frac{1}{n} H(X^n),$$

then $\lim_{n \to \infty} \frac{1}{n} H(X^n)$ is named the entropy rate of the source.
If we can show that, for any $\epsilon > 0$, 
\[
(X)_S \geq (X)_L
\]
then the proof is completed. This claim is proved as follows. 

Next we will prove that 
\[
(X)_S = (X)_L = \overline{H}(X)
\]
and 
\[
(X)_S = (X)_L = \overline{H}(X)
\]
where a source for a source \( (uX)H(u/L)^{\infty-u} \) is introduced into the operational characteristics of \( \lim_{N \to \infty} \) which were already introduced. The operational characteristics of \( \lim_{N \to \infty} \) which were already introduced, 
\[
\overline{H}(X)
\]
and \( (X)_H \) are called the sup-entropy rate and inf-entropy rate, which are introduced in Chapter 6.
By definition of $S^2_\varepsilon(X)$, we know that for any $\gamma > 0$, there exists $\tilde{X}$ and $N$ such that for $n > N$, $1/nR(\tilde{X}^n) < S^2_\varepsilon(X) + \gamma$ and $\|X^n - \tilde{X}^n\| < 2\varepsilon$.

Let $A_n \triangleq \{x^n : P^n(X^n(x^n)) > 0\}$. Since $(1/n)R(\tilde{X}^n) < S^2_\varepsilon(X) + \gamma$ and $\|X^n - \tilde{X}^n\| < 2\varepsilon$, $|A_n| \leq \exp\{R(\tilde{X}^n)\} < \exp\{n(S^2_\varepsilon(X) + \gamma)\}$.

Therefore, $\limsup_{n \to \infty} \frac{1}{n} \log |A_n| \leq S^2_\varepsilon(X) + \gamma$.

Also, $\limsup_{n \to \infty} P^n(X^n(A_n^c)) \leq \varepsilon$. Hence, $\limsup_{n \to \infty} P^n(X^n(A_n^c)) \leq \varepsilon$.

Since $S^2_\varepsilon(X) + \gamma$ is just one of the rates that satisfy the conditions of the minimum $\varepsilon$-source compression rate, and $T_\varepsilon(X)$ is the smallest one of such rates, $T_\varepsilon(X) \leq S^2_\varepsilon(X) + \gamma$ for any $\gamma > 0$. 

Resolvability and source coding
It can then be easily verified that $uX$ satisfies the next four properties:

1. $uX^0 = uX \not\supset (uX)^\gamma \sum - uW \right\} = (uX)^\gamma$
2. $\forall u \not\supset uX \not\supset \left\{(uX)^uX \not\supset uW \right\}$

where

3. $\forall u \not\supset \{0\} \not\supset uX \not\supset \left\{(uX)^uX \not\supset 0 \right\}$
4. $\forall u \not\supset \{0\} \not\supset uX \not\supset \left\{0 \right\}$

$\forall u \not\supset \{0\} \not\supset uX \not\supset \left\{0 \right\}$

This claim can be proved as follows. Fix $0 < \alpha < 1$. By definition of $uW$, there exists a sequence of sets $\forall u \not\supset \{0\}$ such that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By the definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$. By definition of $uW$, we know that for any $\forall u \not\supset \{0\}$, there exists $\forall u \not\supset \{0\}$ such that $uX \not\supset \{0\}$.
Resolvability and source coding

\[ \langle X \rangle L \geq (uX)_I \]

Consequently

\[ I = (0 uX)_d + (uA) uX d \]

\[ \frac{uW}{I} \geq (uX)_d - \left[ \frac{uW}{(uX)_d d W} \right] = \left| (uX)_d - (uX)_d d \right| \]

\[ \forall \ uX \not \in uX \text{ for all } (c) \]

\[ \forall \ uX \not \in 0 uX \text{ since } \forall + \exists > (uA) uX d \geq (0 uX)_d \]

\[ \forall \text{ is } u \text{-type } (a) \]

\[ \forall \text{ is } u \text{-type } (a) \]
\begin{align*}
&\mathcal{L}/(\mathcal{V} + \varepsilon) \subseteq \mathcal{H} \quad \Rightarrow \\
&(\mathcal{V} + \varepsilon) + (\mathcal{V} + \varepsilon) + \mathcal{V} \cup \varepsilon \quad \Rightarrow \\
&(u^V)uX + (\mathcal{V} + \varepsilon) + \frac{((\mathcal{L} + (X)^L)u)}{((\mathcal{L} + (X)^L)u)dx} \quad \Rightarrow \\
&(u^V)uX + (\mathcal{V} + \varepsilon) + \frac{uW}{|u^V|} = \\
&(u^X)uX \quad \subseteq + \quad \frac{uW}{|u^V|} \quad \subseteq \\
&\left[(\mathcal{V} + \varepsilon)uX + (\mathcal{V} + \varepsilon)uX\right] + \left[(u^X)uX - (u^X)uX\right] \quad \subseteq \\
&(u^X)uX - (u^X)uX \quad \subseteq + \\
&\left|(u^X)uX - (u^X)uX\right| \quad \subseteq = \|uX - uX\|
\end{align*}
Resolvability and source coding

Resolvability, which in turn is equal to the sup-entropy rate.

This theorem tells us that the minimum source compression ratio for fixed-length

The proof is completed by noting that $\alpha$ can be made arbitrarily small.

\[(X)^{\beta} L \geq (X)^{(\alpha + \varepsilon)\beta} S\]

resolvability, and $S$ is the smallest one of such quantities.

Resolvability and source coding

Since $\alpha + \varepsilon > 0$.
Theorem 4.22 (equality of mean-resolvability and minimum source coding rate for variable-length codes)

\[\bar{T}(X) = \bar{S}(X) = \limsup_{n \to \infty} \frac{1}{n} H(X^n).\]

Proof: Equality of \( \bar{S}(X) \) and \( \limsup_{n \to \infty} \frac{1}{n} H(X^n) \) is already given in Theorem 4.17.

Definition 4.20 states that there exists, for all \( \gamma > 0 \) and all sufficiently large \( n \), an error-free variable-length code whose average codeword length satisfies

\[\forall \gamma > 0 \exists n \text{ such that } \mathbb{E}(\ell_n) < \frac{1}{n} H(X^n) + \gamma.\]

Moreover, the fundamental source coding lower bound for a uniquely decodable code (cf. Theorem 4.18 of Volume I of the lecture notes) is

\[H(X^n) \leq \mathbb{E}(\ell_n).\]

Thus, by letting \( X = \tilde{X} \) we obtain, for \( \gamma > 0 \),

\[\forall \gamma > 0 \exists n \text{ such that } \mathbb{E}(\ell_n) < \frac{1}{n} H(X^n) + \gamma.\]

Therefore, by letting \( \tilde{X} = X \), we obtain \( \mathbb{E}(\|X^n - \tilde{X}^n\|) = 0 \) and

\[\frac{1}{n} H(\tilde{X}^n) < \frac{1}{n} H(X^n) < \bar{T}(X) + \gamma,
\]

which concludes the proof.
By the fact [1, pp. 33] that $\bar{\mathcal{S}}(\tilde{X})$ is a rate-achievable mean-resolution rate of $X$ for any $\varepsilon > 0$, i.e.,

$\bar{\mathcal{S}}(X) = \lim_{\varepsilon \to 0} \bar{\mathcal{S}}_{\varepsilon}(X) \leq \bar{T}(X)$.

2. $\bar{T}(X) \leq \bar{\mathcal{S}}(X)$.

Observe that $\bar{\mathcal{S}}_{\varepsilon}(X) \leq \bar{\mathcal{S}}(X)$ for $0 < \varepsilon < 1/2$. Hence, by taking $\gamma$ satisfying

$2 \varepsilon \log |X| > \gamma > \varepsilon \log |X|$ and for all sufficiently large $n$,

there exists such $\tilde{X}_n$ that $\|X_n - \tilde{X}_n\| = \varepsilon$. Moreover, for $\varepsilon > 0$ we have

$(\mathbf{x})_{\mathcal{S}} \geq (\mathbf{x})_{\mathcal{L}}$.

and

$(\mathbf{x})_{\mathcal{L}} \geq (\mathbf{x})_{\mathcal{S}} 0^{-\varepsilon}_{\text{lim}} = (\mathbf{x})_{\mathcal{S}}$

which concludes that $(\mathbf{x})_{\mathcal{L}}$ is an $\varepsilon$-achievable mean-resolution rate of $X$.
Mean-resolvability and source coding

Given, the above theorem tells us that the minimum source compression ratio for

\[(\sigma) \mathcal{S} \supset (\sigma) \mathcal{L} \]

Hence, \((\sigma) \mathcal{S} \) is an achievable source compression rate for variable-length codes, and

\[ \frac{u}{I} + \frac{\varepsilon \log \varepsilon}{I} - \frac{|X| \varepsilon \log \varepsilon + \gamma + (\sigma) \mathcal{S}}{I} \geq \]

\[ \frac{u}{I} + \frac{\varepsilon \log \varepsilon}{I} - \frac{|X| \varepsilon \log \varepsilon + (\sigma) \mathcal{H}/u}{I} \geq \frac{u}{I} + (\sigma) \mathcal{H}/u \geq \frac{u \gamma}{I} \]

and \((\sigma) \mathcal{S} \) we obtain

\[(uX) \mathcal{H}(u/I)^{\infty-u} \]
Discussions

\[ (uX)^H \limsup_{n \to \infty} \frac{1}{n} H(X) \leq \bar{H}(X), \]

which follows straightforwardly by the fact that the mean of the random variable \(-\frac{1}{n} \log P(X)\) is no greater than its right margin of the support.

Also note that for stationary-ergodic source, all these quantities are equal, i.e.,

\[ \bar{S} = \bar{I} = \bar{H} = \bar{T} = \bar{S} = \limsup_{n \to \infty} \frac{1}{n} H(X_n). \]

Note that \(\lim sup\) is no greater than its right margin of the support.

By the fact that the mean of the random variable \(\frac{u}{1} \log (u/1)\) is no greater than the \(uX\) which follows straightforwardly.
Example 4.23

Consider a binary random source \( X_1, X_2, \ldots \) where \( \{X_i\}_{i=1}^{\infty} \) are pair-wise independent with common uniform marginal distribution.

You may imagine that the source is formed by selecting from infinitely many binary number generators. The selecting process \( Z \) is independent for each time instance.

\[
Z - 1 = (1)^i X_i \\
Z = (0)^i X_i
\]

Independent random variables with individual distribution \( \{X_i\}_{i=1}^{\infty} \) where \( \{X_i\}_{i=1}^{\infty} \) are...
Source generator: \{X_t\}_{t \in I} (I = (0,1)) is an independent random process with \(P(X_t(0)) = 1 - P(X_t(1)) = 1\) and is also independent of the selector \(Z\), where \(X_t\) is outputted if \(Z = t\). Source generator of each time instance is independent temporally.
It can be shown that such source is not stationary.

Nevertheless, by means of similar argument as AEP theorem, we can show:

\[
\log P_X(X_1) + \log P_X(X_2) + \cdots + \log P_X(X_n) \xrightarrow{n} h_b(Z) \quad \text{in probability}
\]

where

\[
h_b(a) = -a \log_2(a) - (1-a) \log_2(1-a)
\]

is the binary entropy function.

To compute the ultimate average entropy rate in terms of the random variable \( h_b(Z) \), it requires that

\[
\log P_X(X_1) + \log P_X(X_2) + \cdots + \log P_X(X_n) \xrightarrow{\text{in mean}} h_b(Z)
\]

which is a stronger result than convergence in probability.

\[\text{Example}\]

It can be shown that such source is not stationary.
With the fundamental properties for convergence, convergence-in-probability implies convergence-in-mean provided the sequence of random variables is uniformly integrable, which is true for

\[ \sum_{i=1}^{n} \log \mathbb{P}(X_i) \]

because of the i.i.d. nature of \( \{X_i\} \). Similarly, convergence-in-mean of \( \{X_i\} \)

\[ \mathbb{E} \left[ |(X^X d Y)| \right] \]

implies convergence-in-probability of \( \{X_i\} \) for convergence-in-mean of \( \{X_i\} \).
We therefore have:

\[
\left| \frac{1}{n} \log \left( \frac{u}{1} \right) - \left[ (uX)_{uX} d \log(u/1) \right] \right| \leq \left| \frac{1}{n} \log \left( \frac{u}{1} \right) - \left[ (uX)_{uX} d \log(u/1) \right] \right| \to 0.
\]

Consequently,

\[
\limsup_{n \to \infty} \frac{1}{n} H(X^n) = \mathbb{E} \left( \log \left( \frac{u}{1} \right) \right) = \int_0^1 P_Z \left( \log \left( \frac{u}{1} \right) > t \right) \, dt = 0.859912 \text{ bits}.
\]

However, it can be shown that the ultimate CDF of

\[
\left( \frac{1}{n} \log \left( \frac{u}{1} \right) \right)
\]

is

\[
P \left[ t < \left( Z \right)^{qy} d \right] = \int_0^1 \mathbb{E} \left( \log \left( \frac{u}{1} \right) \right) \to \infty.
\]

We therefore have:

\[
\text{Example}
\]
The ultimate CDF of $\log(u/1)$.

\[
\left\{ t \geq (Z)^{\eta t} \right\} \mathcal{D} : (uX)^{\log(u/1)}.
\]