CHAPTER 12
Rate-Distortion Theory

12.1. Introduction

The rate-distortion problem begins with a source alphabet $X$ and a reproduction alphabet $Y$. For $n = 1, 2, \ldots$, we consider mappings $q: X^n \rightarrow Y^n$ with the property that $q$ takes $N < |X^n|$ distinct values in $Y^n$. The idea is that $q(x)$ is an “approximation” or “quantization” of $x$. To measure the error or distortion, we use a function $d_n: X^n \times Y^n \rightarrow [0, \infty)$. If $X = (X_1, \ldots, X_n)$ is an $n$-tuple of $X$-valued random variables, we can consider the mean distortion,

$$E[d_n(X, q(X))].$$

**Example 12.1** (Probability of Error). Suppose $X = Y$ and $d_n(x, y) = 1$ if $x \neq y$ and $d_n(x, y) = 0$ if $x = y$. In other words, $d_n$ is the trivial metric on $X^n$. Then

$$E[d_n(X, q(X))] = P(q(X) \neq X))$$

is just the probability error in a block source code.

**Example 12.2** (Hamming Distance). The Hamming distance between $x$ and $y$ is the number of entries in which they differ. Since $I_{\{x\}}(y) = 1$ if $y \neq x$ and is zero otherwise, the formula

$$\sum_{k=1}^{n} I_{\{x_k\}}(y_k)$$

counts the number of positions in which $x$ and $y$ differ. The normalized Hamming distance distortion measure is

$$d_n(x, y) := \frac{1}{n} \sum_{k=1}^{n} I_{\{x_k\}}(y_k).$$
Example 12.3 (Squared Error). If $X = Y = \mathbb{IR}$, then the Euclidean distance between $x$ and $y$ is
\[ \|x - y\|^2 = \sum_{k=1}^{n} |x_k - y_k|^2. \]
We call $\|x - y\|^2$ the **squared error**. The **normalized squared error** distortion measure is
\[ d_n(x, y) = \frac{1}{n} \|x - y\|^2 = \frac{1}{n} \sum_{k=1}^{n} |x_k - y_k|^2. \]
Thus,
\[ E[d_n(X, q(X))] = \frac{1}{n} E[\|X - q(X)\|^2] \]
differs from the mean-squared error formula in Chapter ?? only by the factor of $1/n$.

The last two examples illustrate the concept of **single-letter distortion measures**, that is, distortion measures of the form
\[ d_n(x, y) = \frac{1}{n} \sum_{k=1}^{n} d(x_k, y_k), \]
where $d: X \times Y \rightarrow [0, \infty)$. The normalized Hamming distance uses $d(x, y) = I_{\{x\}}(y)$, and the normalized squared-error uses $d(x, y) = |x - y|$.

### 12.2. Achievable Rate-Distortion Pairs

**Definition 12.4.** We say that a pair of nonnegative real numbers $(R, D)$ is an **achievable rate-distortion pair** for a source $X_1, X_2, \ldots$, if $\forall \Delta D > 0$, $\forall \Delta R > 0$, $\exists n_0$ such that $\forall n \geq n_0$, $\exists q: X^n \rightarrow Y^n$ taking at most $N$ distinct values with
\[ \frac{\log N}{n} < R + \Delta R \quad \text{and} \quad E[d_n(X, q(X))] < D + \Delta D, \]
where $X = (X_1, \ldots, X_n)$. The **rate-distortion region** is the set of achievable rate-distortion pairs.

For fixed $R \geq 0$, if $(R, D)$ is achievable, then it is easy to see that for any $D' > D$, the pair $(R, D')$ is also achievable. Similarly, for fixed $D \geq 0$, if $(R, D)$ is achievable, then for any $R' > R$, the pair $(R', D)$ is also achievable. Hence, to determine the rate-distortion region, it suffices to characterize either the **rate-distortion function**
\[ R(D) := \inf\{R \geq 0 : (R, D) \text{ is achievable}\}, \]
or the **distortion-rate function**

\[ D(R) := \inf \{ D \geq 0 : (R, D) \text{ is achievable} \} . \]

Of course, \( R(D) \) and \( D(R) \) depend on the distribution of the source \( X_1, X_2, \ldots, \) although we suppress this dependence in the notation.

Let \( X \) and \( Y \) be finite sets, and let \( p \) be a pmf on \( X \). For \( D \geq 0 \), we put

\[ \tilde{R}(D) := \inf_{W : E[d(X,Y)] \leq D} I(X \wedge Y) , \]

where in computing the mutual information and the expectation, we use \( P_{XY}(x,y) = p(x)W(y|x) \). In other words, to compute \( \tilde{R}(D) \), we minimize \( I(p \times W) \) as a function of the transition probability \( W \), subject to the inequality constraint on \( W \) that

\[ \sum_x \sum_y d(x,y)p(x)W(y|x) \leq D . \]

For a discrete memoryless source and a single-letter distortion measure, we will show that for fixed \( D \geq 0 \), the pair \( (R, D) \) is achievable if and only if \( R \geq \tilde{R}(D) \). Thus, \( R(D) = \tilde{R}(D) \) for a discrete memoryless source.