Information sources with different cost scales and the principle of conservation of entropy

by
I. Csiszár, G. Katona and G. Tusnády

INTRODUCTION

Despite the recent progress in information theory, the philosophy of time and information theory is essentially the same as it was in the beginning of the century. As a result, it seems reasonable to analyze the general problem of information theory, i.e., the role of the entropy concept in information theory. It is known that the entropy concept is intimately connected with the problem of communication, and it is fundamental to the theory of channels. In this paper, we will discuss some of the basic concepts of information theory, and we will also present a new method for estimating the entropy of a source.

In Section 3, we will discuss the problem of estimating the entropy of a source. The main idea is to use the concept of the entropy of a source as a measure of the complexity of a source. We will show that the entropy of a source can be estimated by using the concept of the entropy of a source as a measure of the complexity of a source. We will also show that the entropy of a source can be estimated by using the concept of the entropy of a source as a measure of the complexity of a source.

In Section 4, we will discuss the problem of estimating the entropy of a source for a finite alphabet. The main idea is to use the concept of the entropy of a source as a measure of the complexity of a source. We will show that the entropy of a source can be estimated by using the concept of the entropy of a source as a measure of the complexity of a source. We will also show that the entropy of a source can be estimated by using the concept of the entropy of a source as a measure of the complexity of a source.

In Section 5, we will discuss the problem of estimating the entropy of a source for a continuous alphabet. The main idea is to use the concept of the entropy of a source as a measure of the complexity of a source. We will show that the entropy of a source can be estimated by using the concept of the entropy of a source as a measure of the complexity of a source. We will also show that the entropy of a source can be estimated by using the concept of the entropy of a source as a measure of the complexity of a source.
To the interpretation of entropy as the measure of the amount of information in the encoded individual coding scheme is of prime importance. We define the entropy number E of a code character per symbol needed to encode a uniquely decodable code in a matrix of source rate w as equal

\[ E = \frac{H}{w} \]

where \( H \) is the size of the coding alphabet. The "uniquely decodable coding" numbers \( E \) are usually written and printed, however, for rather special cases only, received for them yielding a fixed number of symbols at some fixed number of bits per symbol. In this case, the uniquely decodable number \( E \) is usually written and printed, however, for rather special cases only, received for them yielding a fixed number of symbols at some fixed number of bits per symbol.

In the case of source rate w, the number of symbols needed to encode a uniquely decodable code in a matrix of source rate w is equal to the number of symbols needed to encode a uniquely decodable code in a matrix of source rate w. In the case of source rate w, the number of symbols needed to encode a uniquely decodable code in a matrix of source rate w is equal to the number of symbols needed to encode a uniquely decodable code in a matrix of source rate w.
L. PRELIMINARIES

Throughout this paper, the terms "random variable", "expected value variable" or random variable with finite or countable state space, "expected value random variable" or random variable with countable state space, will be denoted by \( X \), \( Y \), \( Z \), etc. \( X \), \( Y \), \( Z \), etc., will be denoted by \( \mathcal{F} \), \( \mathcal{G} \), \( \mathcal{H} \), etc., respectively.

All random variables \( X \)'s will be assumed to be defined on the same probability space \( \Omega, \mathcal{F}, P \). We will denote the probability space \( \Omega \) by \( \Omega \), \( \mathcal{F} \) by \( \mathcal{F} \) and \( P \) by \( P \) if it is clear from the context. If it is not clear from the context, we will denote the probability space by \( \Omega \), \( \mathcal{F} \), \( P \), respectively.

A random variable \( X \) will be denoted by \( X \). If \( X \) is a random variable with countable state space, we will denote it by \( \mathcal{F} \) and \( \mathcal{G} \) by \( \mathcal{G} \). The random variables \( X \), \( Y \), \( Z \), etc., are assumed to be defined on the same probability space \( \Omega, \mathcal{F}, P \). The expected value of a random variable \( X \) is denoted by \( \mathbb{E}[X] \). The expected value of a random variable \( X \) with respect to a measure \( \mu \) is denoted by \( \mathbb{E}_\mu[X] \). The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \).

The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \). The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \). The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \).

If \( A \in \mathcal{G} \), \( P(A) > 0 \), \( P(A) < 1 \), and \( P(A) = 1 \), then the probability measure \( P(A) > 0 \) will be denoted by \( \mathbb{P} \). If \( P(A) > 0 \), \( P(A) < 1 \), and \( P(A) = 1 \), then the probability measure \( P(A) > 0 \) will be denoted by \( \mathbb{P} \). If \( P(A) > 0 \), \( P(A) < 1 \), and \( P(A) = 1 \), then the probability measure \( P(A) > 0 \) will be denoted by \( \mathbb{P} \).

The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \). The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \). The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \).

If \( X \leftrightarrow Y \leftrightarrow Z \leftrightarrow X \) in the sense of Markov, then we have

\[
\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y).
\]

The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \). The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \). The expected value of a random variable \( X \) with respect to a probability measure \( P \) is denoted by \( \mathbb{E}[X] \).

where \( a \) and \( q \) range over the state space of \( a \) and \( q \), respectively, and \( a \) and \( q \) are independent as usual.
We shall need also the concept of information distance \(d(\psi, \phi)\)

\[
d(\psi, \phi) = H(\psi) + H(\phi) - H(\psi \land \phi)
\]

and the mutual information \(I(\psi; \phi)\) with \(d(\psi, \phi) = \infty\):

\[
I(\psi; \phi) = H(\psi) + H(\phi) - H(\psi \land \phi) - H(\psi \lor \phi)
\]

The equality \(I(\psi; \phi) = I(\psi; \phi\phi) = I(\psi\phi; \phi)\) follows from (1.4), below, \(\phi\).

\[d(\psi, \phi) = \infty\]

For the purposes of this paper, we need to define \(d(x, x)\) if

\[d(x, x) = \infty\]

The definitions basic, identities and inequalities concerning expectations and conditional expectations (see essentially to Shannon [26], see also e.g. [11, 30] elsewhere:

\[
\begin{align*}
&(1.2) \quad \mathbb{E}Y = \mathbb{E}Y|X = x = \mathbb{E}Y|X = x \quad (\text{joint distribution of } \psi) \\
&(1.3) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{generalized conditional expectation}) \\
&(1.4) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{conditional expectation}) \\
&(1.5) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{conditional expectation}) \\
&(1.6) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{conditional expectation}) \\
&(1.7) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{conditional expectation}) \\
&(1.8) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{conditional expectation}) \\
&(1.9) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{conditional expectation}) \\
&(1.10) \quad \mathbb{E}Y = \mathbb{E}Y|X = x \quad (\text{conditional expectation})
\end{align*}
\]

will be used freely, without any further reference. We shall need also some other simple but somewhat less standard identities summarized in the following lemma:
where (1), (2) directly follow:

\[
1) \left(2 \sum_{a \in Q} \left| f(a, b, c) \right| \right) \left(3 \sum_{a \in Q} \left| f(a, b, c) \right| \right) \leq \left(2 \sum_{a \in Q} \left| f(a, b, c) \right| \right) \left(3 \sum_{a \in Q} \left| f(a, b, c) \right| \right)
\]

at last, from (1), (3), (5) and the second inequality of (1), (6) we get

\[
\left(2 \sum_{a \in Q} \left| f(a, b, c) \right| \right) \left(3 \sum_{a \in Q} \left| f(a, b, c) \right| \right) \leq \left(2 \sum_{a \in Q} \left| f(a, b, c) \right| \right) \left(3 \sum_{a \in Q} \left| f(a, b, c) \right| \right)
\]

In (1), (6).

**Proof of Lemma 3.2.** Let \( \{a, b, c\} \) and \( \{a, b, c\} \) be probability distributions and the well-known inequality

\[
\sum_{a \in D^m} \frac{a_i^2}{a_i^2} \geq 0
\]

given rise to (1), (6).

(1) can be proved in the same way, with the choice

\[
\frac{a_i^2}{a_i^2} \geq 0, \quad i = 1, 2, 3, \ldots
\]

(2) the dual form in which three types of convergence of RV's is convergence in probability to another random variable, almost sure convergence and convergence in L^1 norm, respectively. They will be denoted by \( L^1, L^2, \) and \( L^\infty, \) respectively.

**Lemma 3.2.** Let \( Y(t), t \geq 0 \) be a family of RV's and let

\[
\left\{ \sum_{i=1}^{\infty} \frac{Y(i)}{i} \right\} \quad (t \geq 0), \quad t \geq 0.
\]

Then the following hold:

(a) \( Y(t) \) is stochastically integrable \( a.s. \) for \( t \to \infty \)

and

(b) \( Y(t) \to 0 \) \( a.s. \) as \( t \to \infty \)

and

(c) \( Y(t) \to Y \) \( a.s. \) as \( t \to \infty \)

are equivalent and imply \( Y(t) \to 0 \) \( a.s. \) and consequently, \( Y(t) \to Y \) \( a.s. \) as \( t \to \infty \).
First condition (ii) means that the \( a_{ij} \) are all positive.

Remark 1.5. If there exists \( x \geq 0 \), such that for every finite

\[ \sum_{i=1}^{n} a_{i,j} x_i \leq y_j, \quad j = 1, \ldots, m, \]

then, obviously, solution \( x \) is feasible.

Proof. From \( Bx \leq b \), obviously follows that \( Bx \leq b \).

Theorem 1.6. Let \( a_{ij} \) be a symmetric matrix.

\[
\begin{align*}
\sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij} x_j \right) & \leq \sum_{i=1}^{m} \left( \sum_{j=1}^{n} a_{ij} y_j \right) \\
\sum_{j=1}^{n} \left( \sum_{i=1}^{m} a_{ij} x_i \right) & \leq \sum_{j=1}^{n} \left( \sum_{i=1}^{m} a_{ij} y_i \right)
\end{align*}
\]

Equation (iii) holds for all \( x \in \mathbb{R}^n \).

Equation (iv) holds for all \( y \in \mathbb{R}^m \).

We conclude that \( (\mathbb{R}^n, \mathbb{R}^n, \leq, \leq) \) wherein \( \leq \) denotes the standard.

Finally, \( x_0 \) is a feasible solution of (iv) if and only if \( x_0 \) is a feasible solution of (iii).

Thus, \( x_0 \) is a feasible solution of (iv) if and only if \( x_0 \) is a feasible solution of (iii).

In other words, \( x_0 \) is a feasible solution of (iv) if and only if \( x_0 \) is a feasible solution of (iii).

The set \( S \) will be called an extreme point if \( S \) is a finite sequence of elements \( \{x_1, \ldots, x_k\} \) where \( x_1, \ldots, x_k \) are extreme points and \( x_i \neq x_j \) for all \( i \neq j \).

Hence, \( x_0 \) is an extreme point if and only if \( x_0 \) is a feasible solution of (iv).

Therefore, \( x_0 \) is an extreme point if and only if \( x_0 \) is a feasible solution of (iv).

This concludes
ranging only that the elements of \( x \) are really "algebraic", excluding e.g., the powers of \( x_0, x_1, x_2, x_3 \), where \( x_0 \) is the length of the sequence \( x \). Let \( x_0 = (x_0, x_1, x_2, x_3) \). Of course, we set \( x_0 = 0 \) and let \( x_0+i \) for \( i \geq 0 \).

For \( i \geq 0 \), we shall speak of \( x_0 \) as the composition of the term \( x_i \) and \( x_0 \) is the composition of \( x_1 \) and \( x_0 \). Generally, \( x_1 \) is a maximal among \( x_1\), \( x_2 \), and \( x_0 \). However, it will be referred to as a cylinder under the condition that of any \( i \), containing all cylinder space will be denoted by \( x \).

For the rest of this paper, we will be concerned to restrict the use of certain terms, avoiding their more specific meanings in a certain way. The occasional conventions will be the following ones:

1. (a) algebraic

2. (b) algebraic term

3. (c) algebraic term space

(a) algebraic term for the above sequence:

\[ \{x(0), x(1), x(2), \ldots \} \]

(b) the term sequence \( x_0 \) of \( x \)

(c) \( x(1), x(2), \ldots \) is a sequence of nonnegative real numbers, such that

\[ \sum_{n=0}^{\infty} x(n) = x \]

(d) the above sequence:

\[ \{x(0), x(1), x(2), \ldots \} \]

(e) \( x(i) \) for the above sequence:

\[ \{x(0), x(1), x(2), \ldots \} \]

(f) \( x(i) \) is the number of \( x(i) \) with \( x(i) \neq 0 \)

(g) \( x(i) \) is the number of \( x(i) \) with \( x(i) = 0 \)

(h) \( x(i) \) and \( y(i) \) are well-defined, e.g., due to the assumption \( \{x(0), x(1), \ldots \} \).
Chapter 1. Informations processes I

Section 1.1. Information processes in the stationary case.

A sequence $\{x_n\}^{\infty}_{n=-\infty}$ is a process if there exists a collection of processes $\{X_n\}$, such that

$$\text{Lim }{\text{Prob}}\{X_n = x_n\} = \text{Prob}$$

where $\text{Lim}$ denotes the limit of a process $\{X_n\}$ as $n \to \infty$.

The process $\{X_n\}$ is said to be stationary if $\text{Prob} = \text{Prob}$ for all $n$, where $\text{Prob}$ is the probability of the event $\{X_n = x_n\}$.

Definition 1.1. A sequence $\{x_n\}^{\infty}_{n=-\infty}$ is called a stationary process if it satisfies the following conditions:

- For any $n$, $\text{Prob} = \text{Prob}$
- For any $n$, $\text{Prob} = \text{Prob}$
- For any $n$, $\text{Prob} = \text{Prob}$

These conditions ensure that the sequence $\{x_n\}^{\infty}_{n=-\infty}$ is statistically independent of time, meaning that the probability of the event $\{X_n = x_n\}$ is independent of $n$. 

This definition is based on the concept of stationarity, which is fundamental in the study of time series analysis and stochastic processes.
EXAMPLE 3.5. Let \( f \) be a non-negative valued function on 
\( (0,1) \), satisfying \( f(x) \leq \frac{K}{x} \), \( \forall x \in (0,1) \). 
Then, the net measure \( S \) will satisfy

\[
S_x = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(x_k)
\]

where \( x_k = \frac{k}{n} \), \( k = 1, 2, \ldots, n \). 

Define a cost scale \( \kappa \). Cost scales of this type will be called \textit{parametric}

In particular, if

\[
(3.3) \quad \eta = \sum_{j=1}^{n} \eta_j \quad (\eta_1, \eta_2, \ldots)
\]

the corresponding cost scale \( \kappa \) defined by

\[
\kappa = \sum_{j=1}^{n} \eta_j \quad (\kappa_1, \kappa_2, \ldots)
\]

may be called a \textit{non-parametric} cost scale. In general, these \textit{finite} groups

may be taken as \textit{finite} groups, under the condition that \( \eta_j \) belong to the

set of \textit{finite} groups, \( \eta_j \), where the length of \( \textit{finite} \)

of the group is the \textit{finite} length of the \textit{finite} group. Then, if \( \eta_j \) is

defined in the \textit{finite} of \( \eta_j \), the \textit{finite} group of \( \eta_j \)

is the \textit{finite} group of \( \eta_j \).

EXAMPLE 3.5. The cost of transportation may depend on random

error distribution factors of the \textit{finite} groups to be transported. In this model

the cost \( S \) and \( \kappa \) are independent of the \textit{finite} groups. The same

is true for parametric cost scales, independent of the \textit{finite} groups themselves and \textit{finite} parameter.

EXAMPLE 3.6. A cost scale \( \kappa \) may be defined by letting \( \kappa = \frac{1}{n} \sum_{j=1}^{n} \eta_j \).

where \( \eta_j \) is the \textit{finite} group of \( \eta_j \), \( j = 1, 2, \ldots, n \).

EXAMPLE 3.7. A cost scale \( \kappa \) is \textit{critically regular} if \( \kappa = \frac{1}{n} \sum_{j=1}^{n} \eta_j \).

is uniformly bounded, e.g., a \textit{critically regular} cost scale (e.g., example 3.7) is

regular in \( \kappa = \frac{1}{n} \sum_{j=1}^{n} \eta_j \) in bounded away from \( S = \frac{1}{n} \sum_{j=1}^{n} \eta_j \).

- 138 -
If $X$ is a metric, sequence of type (1, a) will be called finite arrangement of $X$. For definitions of $1, 2, 3$ in metric $X$, and $t$, the sequence in $\mathbb{R}$ of the form $\langle x_n \rangle$ is called the standard sequence $\mathbb{R}$ of the sequence $x_n \in X$. 

Obviously, $\forall x \in \mathbb{R}$ the possible "null" sequences are those sequences belonging to the set $\mathbb{R}$. Letting, $\forall x \in X$ be the real sequence $\langle x_n \rangle$,

The convexity of $\langle x_n \rangle$ at $\langle x_n \rangle$ can be considered as the convexity of the average at $\langle x_n \rangle$ (with respect to the given cost model). This suggests the following:

**Definition 2.1.** The average rate of the source $X$ with respect to the cost model $\mathbb{R}$ in the limit

\[
\mathbb{R}(X) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (x_i)
\]

provided that it exists. If the limit does not exist, we shall denote the bandwidth and average of \( \frac{1}{n} \sum_{i=1}^{n} (x_i) \) by $\mathbb{R}(X)$ and $\mathbb{R}(X)$, respectively. The metric set is used for the sake of avoiding confusion with conditional averages.

If the cost at each point in order i.e., $X \in \mathbb{R}$. Then, $\mathbb{R}(X)$ reduces to the usual definition of average per sample.

\[
\mathbb{R}(X) = \mathbb{R}(X) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (x_i)
\]

The idea underlying definition 2.1 is that the relevant information is encoded by the average sequence $\langle x_n \rangle$ by the process $X$ not by the process $X$.

In the sequel we shall omit the argument $X$ where doing so does not cause ambiguity.

**Lemma 2.1.** For every regular cost model

\[
\mathbb{R}(X) = \mathbb{R}(X) \quad (X = \mathbb{R})
\]

below. Furthermore, for any economy cost metrics $X$ and $Y$,

\[
\mathbb{R}(X) < X + Y \quad (X = Y)
\]

- 101 -
that if \( \frac{e_n}{v_n} \rightarrow 0 \), then \( \frac{e_n}{v_{n+1}} \) holds or does not hold simultaneously for \( e_n \) and \( v_n \).

**THEOREM 1.** Let \( e_n \) be regular in the sense of \( \frac{e_n}{v_{n+1}} \) for \( n \rightarrow \infty \) and \( \frac{e_n}{v_{n+1}} \) holds.

Furthermore, \( e_n \) and \( v_{n+1} \) are given by

\[
\frac{e_n}{v_{n+1}} = \frac{a_n}{b_n} + \frac{c_n}{d_n}
\]

the inequality (1.1) is a consequence of theorem 1.3. The last assertion follows from (1.2) and theorem 1.1.11.

In this section we shall be interested in the relation of energy rates with respect to different cost-modes of the same information source. Let us consider the discrete-time source (1.1) and model (1.2), where

\[
\frac{e_n}{v_{n+1}}\quad \text{and} \quad \frac{e_n}{v_{n+1}} = \frac{2a_n}{2b_n} + \frac{2c_n}{2d_n}
\]

(1.4)

The following assertion will play fundamental role in the sequel.

**THEOREM 2.** Let \( X \) be a source with alphabet \( \{0, 1\} \) of size \( n = 2 \) and let \( X \) and \( X_2 \) be two different cost modes for \( X \). Then

\[
\frac{e_n}{v_{n+1}} = \frac{a_n}{b_n} + \frac{c_n}{d_n}
\]

and hence, if \( X \) is such that in \( X \) we have \( \frac{e_n}{v_{n+1}} \), then \( \frac{e_n}{v_{n+1}} \) is as required (where \( v_{n+1} \) is arbitrary).
We have:

\[ a_n = a_{n-1} + b_n \]

and:

\[ b_n = b_{n-1} + c_n \]

Adding these two equations, we get:

\[ a_n + b_n = a_{n-1} + b_{n-1} + (b_n + c_n) \]

Simplifying, we have:

\[ a_n + b_n = a_{n-1} + b_{n-1} + (a_{n-1} + b_{n-1}) \]

or:

\[ a_n + b_n = 2(a_{n-1} + b_{n-1}) \]

This shows that the sequence \( a_n + b_n \) is a geometric progression with the same common ratio as \( a_n \) and \( b_n \).

\[ a_n + b_n = (a_1 + b_1) \cdot r^{n-1} \]

where \( r = \frac{a_2}{a_1} = \frac{b_2}{b_1} \) is the common ratio.

\[ a_n = (a_1 + b_1) \cdot \left( \frac{a_2}{a_1} \right)^{n-1} - b_n \]

\[ b_n = (a_1 + b_1) \cdot \left( \frac{a_2}{a_1} \right)^{n-1} - a_n \]

We can now use these expressions to solve for any terms in the sequence.
\[ \frac{1}{2} \{ a_{ij} + a_{ji} \} = \frac{1}{2} \{ a_{ij} - a_{ji} \} \{ a_{ij} + a_{ji} \} \]

(If so, in case of regular cost matrices, the replacement of \( a_{ij} \) by \( a_{ij} - a_{ji} \) in the above definitions in (2.4) and (2.10) makes no difference.

Intuitively, \( X_i = X_j \) means that the cost of one symbol is essentially the same for both matrices.

The equivalence and the quasi-equivalence are equivalence relations, and the quotient \( a_{ij} \) is uniquely determined by \( X_i \) and \( X_j \), except for the trivial case \( E_a = \emptyset \) where \( \forall a \in A, X_a = X_b \). If \( X_i \) and \( X_j \) are quasi-equivalent cost matrices, there obviously holds

(2.13)
\[ t_{ij}(X_i) = t_{ij}(X_j) \]

**Theorem 2.** Let \( X_i \) and \( X_j \) be two cost matrices for a source \( X \), both in the alphabet \( A \), and let one of the entropy rates \( H_{\mu}(X_i) \) and \( H_{\mu}(X_j) \) exist. If \( X_i \) is quasi-equivalent to \( X_j \), with quotient \( t_{ij}(X_i) \), then

(2.17)
\[ H_{\mu}(X_i) = \frac{1}{2} H_{\mu}(X_j) \]

If the entropy rates in question do not exist, the assertion remains true both for the lower and upper entropy rates.

**Proof.** (2.13) implies

(2.19)
\[ d_{ij}(X_i) = d_{ij}(X_j) \]

Hence, using (1.10) and (2.10), we obtain

(1.20)
\[ 2d_{ij}(X_i) = d_{ij}(X_j) \]

which results by definition the desired statement.
of the \( x \) and \( y \) s is the solution \( \nu(x,y) \), in equivalent to \( \nu(x,y) \), the relation (10.31) is equivalent to

\[
\begin{align*}
0 & \quad \text{if} \quad y < (1-w) \\
1 & \quad \text{if} \quad y \geq (1-w)
\end{align*}
\]

and this, in turn, is equivalent to (10.25). In view of the assumption

\[
0 < 1 \quad \text{if} \quad y > 1
\]

where, again, is equivalent to (10.25).

The following consequences of theorems 5.2 and lemma 5.1 in

\[
\text{w} = \text{w} = \text{w} \quad \text{where} \quad \text{w} = \text{w} = \text{w}.
\]

For example, the two regular cost make for \( X \) such that one of them has the

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w} 
\]

which, again, is equivalent to (10.25).

The user described the cost estimate error as being

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.34)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

In the above expression (10.34), \( \text{w} \) is equivalent to the cost scale \( \text{w} \) is

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.35)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.36)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

In the above expression (10.36), \( \text{w} \) is equivalent to the cost scale \( \text{w} \) is

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.37)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.38)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.39)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.40)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.41)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.42)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.43)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.44)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.45)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.46)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]

(10.47)

\[
\text{w} \quad \text{if} \quad \text{w} \quad \text{if} \quad \text{w}.
\]
In a given cost scale $\frac{1}{\gamma}$ converges in probability to a constant $\gamma'$ which is called the average rate of the source with respect to the given cost scale. Similarly, $\frac{1}{\gamma'}$ converges in probability to a constant $\gamma''$. The average rate $\gamma'$ will be called the average symbol rate (with respect to the given cost scale).

In view of theorem 2.2 with $X_i = X, Y_i = Y, E_i = E$ a symbol rate

For some $\epsilon > 0$ exists an average symbol rate $\gamma + \epsilon$ under the assumptions (3) and there are also necessary and sufficient conditions for $\gamma = 0$. In particular, this is the case of the $\epsilon$-metric, as a particular case of theorem 2.3.

THEOREM 2.4. If the source $X$, with finite alphabet $\Delta$, and with a regular cost scale $\gamma$, is a stationary ergodic source on $\tau$, (i.e., $E[X(n)]$ is a constant $\gamma$, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_i) = I(X; Y)$$

This theorem is a generalization of the results of [1], [2].

THEOREM 2.5. Let us consider a finite-state Markov channel as defined by equation (1). Let $p_i(n)$ be a state sequence of length $n$ and let $q_i(n)$ be a state sequence of length $n$. Then, for all $i$, $n$, $\epsilon > 0$,

$$p_i(n) \leq q_i(n) + \epsilon$$

This theorem is a generalization of the results of [3], [4].

EXAMPLE 2.6. Let $X$ be a finite-state Markov channel as defined by (1). Let $p_i(n)$ be a state sequence of length $n$ and let $q_i(n)$ be a state sequence of length $n$. Then, for all $i$, $n$, $\epsilon > 0$,

$$p_i(n) \leq q_i(n) + \epsilon$$
values in $A$. For $\tilde{A}$, $A$ is finite, and for each state $u_n$, the channel is capable of transmitting between $n$ values. If $\tilde{u}_n$ is fixed in the state $u_n$, then each sequence $\tilde{u}_n = (\tilde{u}_{n+1}, \ldots, \tilde{u}_{n+L})$ represents the cost of transmission of the sequence $u_n$. Let $\tilde{u}_n = \tilde{u}_{n+1} = \cdots = \tilde{u}_{n+L} = \tilde{u}_n$. We make the usual assumptions that for each pair of states $u_n, u_{n+1} \in \tilde{A}$, there exists a $\tilde{u}_n, \tilde{u}_{n+1}, \cdots, \tilde{u}_{n+L}$ such that $u_n \rightarrow \tilde{u}_n, u_{n+1} \rightarrow \tilde{u}_{n+1}, \cdots, u_{n+L} \rightarrow \tilde{u}_{n+L}$, and that for all $u_n, u_{n+1} \in \tilde{A}$, the channel is $A$-measured by the channel $A_n$, $B_n = T_{u_n}^\ast B_{n+1}^\ast A_n$. Let $u_n = \tilde{u}_n, \tilde{u}_{n+1}, \cdots, \tilde{u}_{n+L}$ and let the weight-cost state $X$ be defined by (2.8) with the previous data. Let $\tilde{u}_n$ denote the output of the channel $A_n$, $\tilde{u}_{n+1}, \cdots, \tilde{u}_{n+L}$ with $\tilde{u}_n$, and such that $\tilde{u}_{n+L} = \tilde{u}_n$. Let $u_n$ denote the output of the channel $A_n$, $\tilde{u}_{n+1}, \cdots, \tilde{u}_{n+L}$ with $\tilde{u}_n$. We denote the channel capacity by

\begin{equation}
C = \sum_{u_n} \log \frac{\tilde{P}(u_n)}{u_n}.
\end{equation}

Thus with $\tilde{u}_n = 0$ we have at most two possible values, inferring $\sum_{u_n} \log \frac{\tilde{P}(u_n)}{u_n} > 0$. Then if $\sum_{u_n} \log \frac{\tilde{P}(u_n)}{u_n} = 0$, we have, according to (2.5),

\begin{equation}
\frac{1}{\sum_{u_n} \log \frac{\tilde{P}(u_n)}{u_n}} = \frac{1}{C}.
\end{equation}

The inequality (2.20), first appearing in Shannon's fundamental paper [20] as a form of the entropy of a discrete random variable, is the form of a "max-entropy" source of transmission. It states that the maximum entropy per unit cost can be made rigorous only as the limit of Shannon's $k^n$. The existing rigorous proofs of (2.20) are not very satisfactory. We formulate it only (2.20) at the case of a state $X$, [20].

Let us now consider the inequality (2.20) in the more general form that will follow from theorem 2.3.
Theorem 5.5. Let $N$ be a source with finite alphabet and let $B_i$ and $B_j$ be regular cost scales for $N$ such that one of them has the property (2.3.1) and

$$
\frac{1}{\lambda} q_i^d(B_i) = \lambda_i B_i.
$$

If $N_0$ denotes the number of different possible values of $q_i^d(x) + p_i^d(x)$ and

$$
C = \sum_{i=1}^{n} C_i(x) \log_2 \Delta_i
$$

then

$$
(2.3.9)
$$

in the original paper (ii), because (1) we proved (2.3.9) under more general conditions which did not ensure the validity of (2.3.9).

**4.2. The principle of conservation of entropy.**

Let $N$ and $N'$ be two sources with finite alphabets $A$ and $A'$, respectively. Let $B$ be a cost scale for $N$ and $B'$ a cost scale for $N'$; then we define the rates

\begin{align*}
(4.1) & \quad \dot{r}(x, y, z, f) = \sum_{i=1}^{n} q_i^d(x), q_i^d(y) \\
(4.2) & \quad \dot{r}(x, y, z, f) = \sum_{i=1}^{n} q_i^d(x), q_i^d(y) \\
(4.3) & \quad \dot{r}(x, y, z, f) = \sum_{i=1}^{n} q_i^d(x), q_i^d(y) \\
(4.4) & \quad \dot{r}(x, y, z, f) = \sum_{i=1}^{n} q_i^d(x), q_i^d(y)
\end{align*}

- 139 -
provided that the limits exist (t.w.) and \( (\eta, \lambda) \) are defined as in \( \text{S} \). (t.w.) the \( \alpha \)-sequence \( v \) will be abscised, normal. If the limits do not occur, one may consider the corresponding upper and lower extreme hyper and lower distance.

as an immediate consequence of theorems 3, 4, lemma 5, 1 and

\[ \text{THEOREM 5.} \] Of the same (t.w.), as well as the corresponding upper and lower extreme hyper and lower distance, the \( \alpha \)-sequence \( v \) is considered.

In this section we shall apply the results of section 3 to the case that \( \alpha \) is defined from \( \beta \). To avoid confusion, for this purpose we define the code in a very general sense.

DEFINITION 5.5. Let \( \mathcal{A} \) and \( \mathcal{B} \) be (finite) algebras. Then \( \mathcal{A} \) \( \rightarrow \mathcal{B} \) (mapping) of \( \mathcal{A} \) into \( \mathcal{B} \) such that \( (a, b) \in \mathcal{B} \) and

\[ u = v \quad \text{implies} \quad [u] = [v] \]

will be called a code from \( \mathcal{A} \) to \( \mathcal{B} \).

The code of a sequence \( a_0, a_1, \ldots, a_{n-1} \), we may write in the form

\[ [a_0, a_1, a_2, \ldots, a_{n-1}] = \left[ a_0, a_1, a_2, \ldots, a_{n-1} \right] \]

Each code \( f \) from \( A \) to \( B \), defines a mapping of type sequences, \( (a_0, a_1, a_2, \ldots, a_{n-1}) \), \( \mathcal{A} \), \( \mathcal{B} \), \( \mathcal{C} \), \( \mathcal{D} \), \( \mathcal{E} \). In general, we may write:

\[ f(a_0, a_1, a_2, \ldots, a_{n-1}) = (b_0, b_1, b_2, \ldots, b_{n-1}) \]

Observe that the definition 5.5 does not include the type "code" of some infinite sequences. In fact, the set of these \( \alpha \)-sequences which \( (\alpha, \beta) \) is infinite, will be denoted by

\[ \text{REMARK 5.6.} \quad \text{It is easy to see that } \text{S} \text{ is not a code} \]

and, also, the mapping \( f: \mathcal{A} \rightarrow \mathcal{B} \) to be compatible with respect to the \( \alpha \)-sequence \( \mathcal{A} \) and \( \mathcal{B} \) defined by the calculus \( \alpha \)-sequence, \( \beta \), \( \beta \) is a code, the mapping \( f: \mathcal{A} \rightarrow \mathcal{B} \) may be called an infinite-code. Identically, the concept of an infinite-code in much more restrictive than that of an ordinary measurable mapping \( f \) is defined by

\[ f: \mathcal{A} \rightarrow \mathcal{B} \]

through all essentially measurable mappings \( H \) is compatible with respect to the calculus \( \alpha \). Two different codes, \( \mathcal{A} \) and \( \mathcal{B} \), may give rise to two different infinite-codes. In this case we shall say that \( \mathcal{A} \) and \( \mathcal{B} \) are expressible and write \( \lambda = \mu \). In general, such infinite-codes may be connected with the corresponding equivalence class of codes and vice-versa.

\[ \quad \text{C} \]
If \( a \) is a code from \( X \) to \( Y \) and \( x_1, x_2, \ldots, x_n \in \text{dom}(a) \), let \( h(x_1, x_2, \ldots, x_n) \) denote the \( n \)-tuple of occurrence of the sequence \( (x_1, x_2, \ldots, x_n) \).

For \( 1 \leq i \leq n \), let \( h_i(x) \) denote the \( i \)-th coordinate of \( h(x) \) or, in other words, \( h_i(x) = h(x)_i \).

EXAMPLE 3.1. Let \( X \times Y \) be the graph of \( g \), where \( g(x) \) is some mapping from \( X \times Y \) to \( Z \). The resulting code will be called a simple linear code. In this case we find:

\[
\{h(x) \mid h(x)_1 = h(x)_2 \} \quad \text{and} \quad \{h(x) \mid h(x)_1 = h(x)_2 = h(x)_3 \}.
\]

EXAMPLE 3.2. Let \( X \) and \( Y \) be finite alphabets, let \( E \) be \( X \times Y \) with \( X \) and \( Y \) being the Cartesian product of \( X \) and \( Y \), respectively. Let \( E \) be a linear code, \( g \) a code from \( X \) to \( Y \). Then \( E \times g \) will be the code from \( X \) to \( Y \) defined by:

\[
(h(x), g(x)) \quad \forall x \in X.
\]

EXAMPLE 3.3. Let \( X \) be a sequence with finite alphabet \( X \) and let \( \xi \) be a code from \( X \) to \( Y \). We say that \( \xi \) is extendable if \( \xi(X) \) is extendable by \( \eta : Y \to \{0, 1\} \).

The encoding results in a new sequence \( \xi(X) \), defined by:

\[
\eta \circ \xi(X) = (h(x_1), h(x_2), \ldots, h(x_n), \ldots).
\]
If there is given a cost scale $K_{a}$ for $\mathbf{X}$ then we also define the mapped cost scale $K_{a}^{*}(\mathbf{X})$ by

$$e_{a}^{*}(c) = e_{a}(c) / K_{a}^{*}(c)$$

as equivalently by

$$\mu_{1}^{*}(x) = \mu_{1}(x) / \mu_{1}^{*}(x) \quad \text{for all } x \in \mathbf{X}.$$
holds. Hence, using (6.7), the inequality

$$\|X \|_2 \leq \|X \|_1$$

follows.\textup{ (6.10)} \quad \textup{See (6.11) and (6.12) given}

$$\|X \|_2 = \|X \|_1$$

However, \textup{ (6.9) } \rightarrow \textup{ (6.8) } \quad \textup{Thus by theorem 2.4 we obtain}

$$\|X \|_2 \leq \|X \|_1$$

\textup{ (6.13)} \quad \textup{Finally, (6.10) and (6.12) give the desired (6.8).}

The intuitive meaning of theorem 2.3 is clear. (6.14) represents
the "principle of conservation of energy", i.e., that finite-dimensional encoding
does not change the total amount of information contained in the source after
compression. If this result is not measurable from (6.12) and (6.13) we can obtain

$$\|X \|_2 \leq \|X \|_1$$

which means that in the non-measurable case more information may get lost.

\textup{The following corollary of theorem 2.3 is worth formulating as a new theorem:}

\textbf{Theorem 2.4. Let} \quad \textup{be a source with a given regular cost}

$$f, \quad \text{be measurable by a finite-decomposable code} \quad \textup{and let} \quad \textbf{ be a}

case code for \textbf{ such that}

$$\|f \|_2 \leq \|f \|_1$$

Then

$$\sum \|X \|_2 \leq \|X \|_1$$

and the code \textbf{ has the property that \textbf{ is bounded. We have}

$$\| \| \leq \| \|$$

- 133 -
PROOF. By remark 3.5, the standard case $\mathcal{N}_0 \rightarrow \mathcal{N}_1$ is regular, whereby $\mathcal{N}_0$ is also regular trivially. Using Lemma 3.5 we assume that (3.13) is equivalent to $\mathcal{N}_0 \rightarrow \mathcal{N}_1 = e'$. Hence, Theorem 3.2 leads to (3.13).

Example 3.3. A more general case $\mathcal{N}_0 \rightarrow \mathcal{N}_1$ is regular, where (3.13) is equivalent to $\mathcal{N}_0 \rightarrow \mathcal{N}_1 = e'$. Hence, Theorem 3.2 leads to (3.13).

Example 3.4. A more general case $\mathcal{N}_0 \rightarrow \mathcal{N}_1$ is regular, where (3.13) is equivalent to $\mathcal{N}_0 \rightarrow \mathcal{N}_1 = e'$. Hence, Theorem 3.2 leads to (3.13).

Remark 3.5. Theorem 3.1 in particular may not be the most important case. Of course, $\mathcal{N}_0 \rightarrow \mathcal{N}_1$ is regular, where (3.13) is equivalent to $\mathcal{N}_0 \rightarrow \mathcal{N}_1 = e'$. Hence, Theorem 3.2 leads to (3.13).

Theorem 3.6. Let $\mathcal{N}_0 \rightarrow \mathcal{N}_1$ be a source with a given regular case.
and the ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]

The ratio is then

\[ \frac{[F]}{[I]} = \frac{[I]}{[F]} = \frac{[C]}{[I]} = \frac{[I]}{[C]} \]

where

\[ [F] = \frac{[C]}{[I]} \]

and

\[ [I] = \frac{[C]}{[F]} \]
4 CONCLUDING REMARKS

In this paper, we considered the transmission of information over noisy channels with finite alphabets. The assumptions made have been useful for our analysis, and we have shown that the principles described hold for the general case. The approach taken here is based on the work of Gallager, who has made significant contributions to the field of information theory. Gallager's results form the foundation for our analysis, and we have extended them to cover the specific case of finite alphabet channels.

The main results of this paper can be summarized as follows. First, we have shown that the rate distortion function can be expressed in terms of the capacity of the noisy channel. Second, we have demonstrated that the optimal transmission strategy is one in which the source is encoded using a compression-decomposition scheme. Finally, we have shown that the rate distortion function is achieved by encoding the source with a rate that is equal to the capacity of the channel.

In conclusion, the results presented in this paper provide a foundation for the design of efficient communication systems. The techniques developed here can be applied to a wide range of applications, including data compression, error correction, and secure communication. The methods described are particularly useful in scenarios where the channel conditions are highly variable or where the communication bandwidth is limited.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their valuable comments and suggestions. Their feedback has significantly improved the quality of this paper. We also acknowledge the support of the National Science Foundation, which has provided funding for our research.

REFERENCES


REFERENCES


