Reprint ISSN 0973-9424

INTERNATIONAL JOURNAL OF MATHEMATICAL SCIENCES AND ENGINEERING APPLICATIONS

(IJMSEA)

ASCENT PUBLICATION @ ASCENT PUNE, INDIA

www.ascent-journals.com

International J. of Math. Sci. & Engg. Appls. (IJMSEA)
ISSN 0973-9424, Vol. 18 No. II (December, 2024), pp. 11-25

SOME DEMEANOUR BASED ON UTILITY AND MUTUAL INFORMATION OF DISCRETE MEMORY-LESS CHANNEL CAPACITY FOR MULTI IDENTICAL CASCADED CHANNELS

Rohit Kumar Verma
Associate Professor, Department of Mathematics,
Bharti Vishwavidyalaya, Durg, C.G., India
E-mail-rohitkverma73@rediffmail.com

M. Bhagya Laxmi
Ph.D. Research Scholar, Department of Mathematics,
Bharti Vishwavidyalaya, Durg, C.G., India
E-mail-bhagyalaxmi9039@gmail.com

Abstract

In the present paper we shall discuss the information capacity of a Gaussian Channel through utility-Based Channel Capacity and derive utility distribution in information theory and also have been derived Cascaded channel through the utility distribution .The outcome holds true for every category of channels (with or without memory, either discrete or continuous), as long as the necessary capacities are there. The term "capacity" refers to the highest of those rates for It uses block coding for a long enough delay to achieve arbitrarily high reliability. It should be noted that "channel capacitance" is typically defined in two ways. The first definition uses the capacity of the channel.as the sum of the "information" that the channel processes, where "information" is the difference between the "equivocation" at the output and the "uncertainty" at the input. According to the second definition, which is the one we apply here, the channel capacity is the highest "error free rate." These two definitions have been shown to be equal for specific classes of channels (such as memory-less channels and finite state indecomposable channels). Actually, the Fundamental Theorem of Information Theory is based on this equivalency.

Keywords: Shannon entropy, channel capacity, utility, cascaded channel etc.

1. Introduction

C. E. Shannon [34] examined Gaussian channel for the first time in 1948 in his original publication. According to his definition, the Gaussian channel is nothing more than a temporal discrete alphabet channel that can be transformed into a discrete four-input channel by applying a four-level input signal. Similar concepts are applied to transform a continuous channel into a discrete channel in certain real-world modulation techniques. A discrete channel's primary benefit is its simplicity in processing the output signal for error correction; yet, quantization results in the loss of some information. Although Pinsker [26] conducted a detailed analysis of the colored noise Gaussian channel's capacity, he provided the water-filling solution for it.

Shannon's 1948 development of channel capacity is a fundamental idea in information theory. Although it involves a free choice of input distribution, it is a crucial indicator of channel performance. However, in reality, a number of independent criteria for instance, financial, time, and energy limitations may limit this option, in which case channel performance must be taken into account. Simon (1970) examined a general technique for figuring out the channel capacity of N cascading identical discrete memory less non-singular channels. We go over the same method and expand on its uses when the source alphabet has both probabilities and utility.

Channel capacity refers to the maximum pace at which data may be transmitted across a channel with an infinitely low likelihood of error. It is an inherent function of dependable communication. Theorems of channel coding and its opposites demonstrate that the capacity C may be described as a formula that relies on the probabilistic representations of the channels as follows:

$$C = \max_{x} I(X;Y) \tag{1.1}$$

The capacity of memory-less channels, where X is the channel input, Y is the channel output, and $I(X;Y)$ is the average mutual information between X and Y, is (1.1), as demonstrated by Shannon (1948). The limiting equation (1.1) was reached using the expression

$$C = \lim_{x} sup \frac{1}{n} I(X^n;Y^n) \tag{1.2}$$

If a channel with memory exists, its capacity is equal to the maximum of average mutual information, where input is a sequence of length n and output is a comparable sequence. The channel capacity was demonstrated by Dorbushin (1964) indicated by equation (1.2) is for information stable channels; nevertheless, the formula (1.2) does not converse to a limit for information unstable channels. Channels such as averages memory less, stationary regular decomposable, and stationary non-anticipatory are examples of information unstable channels by Verma (2023). It therefore becomes unclear if there is a comprehensive universal formula for channel capacity that does not rely on presumptions like causality, stationary, memoryless-ness, or information stability.A has been defined by Verdu and Han (1994). use the following general formula to calculate channel capacity:

$$C = \sup_{X} \bar{I}(X;Y),$$

Where $X = \sum_{n=1}^{\infty} X_n^{(n)}$ denote as input process in the form of a sequence of finite- dimensional distribution and $Y = \sum_{n=1}^{\infty} Y_n^{(n)}$ is the corresponding output sequence of finite dimensional distributions and $\bar{I}(X;Y)$ is the information rate or mutual information between X and Y. Here Y is induced by X via the channel $W = \sum_{n=1}^{\infty} W_n^{(n)}$ where $W = \left\{P\left(\frac{Y^n}{X^n}\right) : A^n \to B^n\right\}$, It is an arbitrary series of n-dimensional conditional output distributions from A^n to B^n, where A and B stand for the input and output alphabets, respectively. The channel capacity is provided by .

1. The channel capacity is non negative i.e. $C \geq 0$ because

$$C = max\, I(X;Y) \geq I(X;Y) \geq 0$$

2. $C \leq \log|Y|$, $since\, max\, I(X;Y) \leq \max H(X) = \log|Y|$

3. $C \leq \log|Y|$, $since\, max\, I(X;Y) \leq \max H(Y) = \log|Y|$

4. $I(X;Y)$ is concave and continuous function

Discrete Channel Capacity

Think of n distinct channels in a cascade. We can presume that these channels share a similar alphabet of m symbols as they are all required to convey the same message. The right signals are connected to each channel. Emblem. Our assumptions are that all of the signals in a given channel have the same duration and that the intermediate channel functions as a "maximum a-posterior probability detector." Because the receiver needs to receive the entire signal before it can calculate the a posteriori probabilities, under these circumstances, there will be a delay on the i^{th} channel of at least Ti in addition to the propagation time.

Based on the decoding process and the noise statistics for each channel,The probability that a specific signal, say x_i, will be broadcast and another, say y_j, will be received is known as the transition probabilities, and it can be determined for each channel based on the noise statistics and the decoding process. Allow this probability can be expressed as follows for the i^{th} channel: $P\left(\frac{x_i}{y_j}\right)$

Classification of Channel Capacity

Channel capacity given by Shannon in equation (1.11) can be classified as follows

1. Deterministic Channel:

For every i , j, that is, if Y is determined by X or, equivalently, if $H\left(\frac{Y}{X}\right) = 0$ for all input distributions, or if $p\left(\frac{y_j}{x_i}\right) = 1\ or\ 0$.A channel illustrationis a device that takes as input X the name of a playing card selected from a standard 52-card deck, and outputs the card's suit as $H(Y) - H\left(\frac{Y}{X}\right) = H(Y) = \log 4$ is the information processed if the card is chosen at random so that all values of X (and thus of Y) are equally likely.

2. Lossless Channel:

if for every input distribution $H\left(\frac{Y}{X}\right) = 0$. Put differently, a lossless channel is distinguished by the absence of transition faults because the input is dictated by the output.

3. Symmetric Channel:

If the channel matrix $p\left(\frac{y_j}{x_i}\right)$ has the same set of numbers $p_1, p_2, \ldots\ldots, p_n$ in each row, and if the same set of numbers $q_1, q_2, \ldots\ldots, q_n$ in each column,

For example the matrix

$$\begin{array}{c} \quad y_1 y_2 y_3 y_4 \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{array}$$

reflects channels that are symmetrical. With the exception of permutation, the channel matrix's rows and columns are identical. The definition of a symmetric channel has the direct effect of In this type of channel, $H\left(\frac{Y}{X}\right)$ solely depends on the channel probabilities $p\left(\frac{y_j}{x_i}\right)$ and is not affected by the input distribution p(x).

It may be noted that if $X = x_i$ the probabilities associated with the output alphabets $y_1, y_2, y_3, \ldots\ldots\ldots, y_n$ are $p(y_1), p(y_2), \ldots\ldots\ldots, p(y_n)$.

Hence $H\left(\frac{Y}{X}\right) = -\sum_{i=1}^{n} p(y_j) \log p(y_j), \forall\, j = 1,2,3, \ldots\ldots, n$

Therefore $H\left(\frac{Y}{X}\right) = \sum_{i=1}^{n} p(y_j) H\left(\frac{Y}{X}\right)$

For any input distribution p(x).

Cascaded Channel :

It is fairly common in practice to use cascaded channels. Since the signal shrinks, they become essential because some waves, such as electromagnetic waves and microwave links, do not follow the earth's curvature. When the distance grows, it becomes prohibitive. Under such circumstances, the channel's entire design had to be divided into a series of smaller channels. "Intermediate channel" is another name for it. A number of authors have written about how well channels perform when they cascade with itself.

Many authors have written on how well channels perform when they cascade with one another. Ash (1965), Cover, and others examine the scenario of a binary symmetric channel cascading with itself n times. Mc-Eliece (1977) and Thomas (1991). Gallager (1968) demonstrates that the capacity reduces to zero for extended cascades of fixed channels at least as quickly as a specific exponentially declining function. In a preliminary study published in 1955, Silverman compares Z channels and BSCs with the same capacity and investigates which would have a larger capacity when cascaded with itself (i.e., a cascade of length 2). Such cascades are taken into account for the class of very noisy channels by Majani (1988), who additionally looks at the impact of an inverter positioned in between the channels. Simon (1970) provides a calculation for the capacity given specific presumptions.

There are significant differences between the issues of information transmission via a single channel and information transmission over a cascade of channels. The transmitter possesses all the information that needs to be conveyed in the first scenario. In the second scenario, however, the information that each transmitter has access to—specifically, the information that the first transmitter transmitted—takes the form of a collection of a posterior probability rather than a symbol. This is true for all transmitters except the first one. Therefore, it is crucial for the operation of cascaded channels that we anticipate the way the intermediate channel would function.

2. Our Result:

Utility-Based Channel Capacity

Information is transferred and processed in communication systems with the intention of achieving a particular goal—that is, a message that must be effective. The system's aim is to achieve the goal, which implies that a logical block exists that is to be able to differentiate between different signals' quality based on the specified criterion. These standards for a qualitative distinction of signals are predicated on the importance, relevance, or utility of the data that they are conveying.

Claim 2.1

Let $X = \{x_1, ..., x_n\}$ be a random variable with $u(x_i)$ and $p(x_i)$ $\forall\, i = 1, ..., n$ be the utility and probability based distribution such that $u(x_i) > 0$ and $p(x_i) > 0$ for all i and $\sum_{i=1}^{n} p(x_i) = 1$. Then utility and probability distribution for two variables Bhakar and Hooda [ ] the weighted channel capacity for memory less channel's will be

$$C(U) = \max_{x} I(X, Y; U)$$

Proof: Let $X = \{x_1, x_2, ..., x_n\}$ be a random variable .let $u(x_i)\ and\ p(x_i)\ \forall\ i = 1,2, ..., n$ be utility and probability based distribution respectively ,where $u(x_i) > 0\ and\ p(x_i) > 0\ \forall\ i$ and $\sum_{i=1}^{n} p(x_i) = 1$. Belis and Guiassu introduced a qualitative measures of information

$\text{H}(P;U) = -\sum_{i=1}^{n} u_i\ (x_i)\ p_i\ (x_i) \log p_i\ (x_i)$, where $u_i > 0$ is the utility attached to the ith event which occurs with probability p_i and they found deep application in theory of questionnaires .the measures defined by equation

$$H(P;U) = -\sum_{i=1}^{n} u_i\ (x_i) p_i\ (x_i) \log p_i\ (x_i)$$

Was called useful information by Longo .the Shannon's measures defined by equation

$$H(P) = -\sum_{i=1}^{n} p_i(x_i) \log p_i\ (x_i)$$

Posses very nice mathematical properties which are very useful in many fields.Therefore, the range of the When examining its application, a measure can occasionally seem overwhelming. This step does failing to achieve the maximum value for any ( ) i p x and failing to meet the additivity requirement. Bhakar and Hooda (1993) suggested and defined the next two measures of helpful information in order to get around these obstacles:

$$H(P\colon Q;U) = \frac{\sum_{i=1}^{n} u_i\ (x_i)\ p_i\ (x_i) \log \left({}^{p_i}/_{q_i}\right)}{\sum_{i=1}^{n} u_i\ (x_i)\ p_i\ (x_i)} \tag{2.1.1}$$

and for all probability distribution P and Q having attached with utility distribution U, we propose the function:

$$H_\alpha(P\colon Q;U) = \frac{1}{1-\alpha}\left[\phi(1) - \phi\left(\frac{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\alpha}(x_i) q_i^{1-\alpha}(x_i)}{\sum_{i=1}^{n} u_i\ (x_i)\ p_i\ (x_i)}\right)\right] \tag{2.1.2}$$

If $\phi(x) = \log x$, then equations (2.1.2) reduces to

$$H_{(\alpha,\beta)}(P\colon Q;U) = \frac{1}{1-\alpha}\left[\log 0 - \phi\left(\frac{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\alpha}(x_i) q_i^{1-\alpha}(x_i)}{\sum_{i=1}^{n} u_i\ (x_i) p_i^{;\beta}(x_i)}\right)\right]$$

$$= \frac{-1}{1-\alpha}\left[\log\left(\frac{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\alpha}(x_i) q_i^{1-\alpha}(x_i)}{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\beta}(x_i)}\right)\right] \text{,where } \beta \geq 1, \alpha > 1 \tag{2.1.3}$$

Which is generalized relative information measures of order $\alpha\ and\ \beta$ characterized and studied by Hooda and Sharma.

If $\beta = 1$ equation (2.1.3) reduces to

$$H_\alpha(P;U) =-\frac{1}{1-\alpha}\left[\log\left(\frac{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\alpha}(x_i) q_i^{1-\alpha}(x_i)}{\sum_{i=1}^{n} u_i\ (x_i) p_i(x_i)}\right)\right] \tag{2.1.4}$$

Which is generalized useful relative information measures of order α studied by Bhaker and Hooda..

Corresponding to (2.1.1), the conditional useful information measure can bedefined as

$$H_\alpha\left(\frac{X}{Y};U\right) =-\frac{1}{1-\alpha}\left[\log\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} u_i\,(x_i,y_j) p_i^{\alpha}(x_i,y_j) q_i^{1-\alpha}(x_i,y_j)}{\sum_{i=1}^{n} u_i\ (x_i,y_j)\ \ p_i(x_i,y_j)}\right)\right] \tag{2.1.5}$$

Where $u_i(x_i, y_j)$ *and* $p_i(x_i, y_j)$ are joint utility and probability distributions of X and Y respectively and is their conditional probability distribution.

We have

$$\lim_{\alpha\to 1} H_{\alpha,\beta,j}(P,Q;U) = j\ \frac{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\beta}(x_i) \log \left({}^{p_i}/_{q_i}\right)}{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\beta}(x_i)} \tag{2.1.6}$$

Which is generalized useful relative information measures of type-β. If $j = 1,$ equation (2.1.6) reduces to

$$H(P,Q;U) = -\ \frac{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\beta}(x_i) \log \left({}^{p_i}/_{q_i}\right)}{\sum_{i=1}^{n} u_i\ (x_i) p_i^{\beta}(x_i)} \tag{2.1.7}$$

$$\mathrm{H}\left(\frac{\mathrm{X}}{\mathrm{Y}};\mathrm{U}\right) = -\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} u_i\ (x_i\ ,y_j) p_i^{\beta}(x_i, y_j) \log \mathrm{p}\left({}^{x_i}\ /_{y_j}\right)}{\sum_{i=}^{n}\sum_{j=1}^{n} u_i\ (x_{i,}\ y_j) p_i^{\beta}(x_i\ ,y_j)} \tag{2.1.8}$$

Let $u_i\ (x_i) p_i(x_i) = \sum_{j=1}^{m} u_i\ \left(x_i\ ,y_j\right) p_i\left(x_i, y_j\right)$, then we have

$$H(X;U) = -\ \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} u_i\ (x_i\ ,y_j) p_i^{\beta}(x_i, y_j) \log \mathrm{p}(x_i\ )}{\sum_{i=}^{n}\sum_{j=1}^{n} u_i\ (x_{i,}\ y_j) p_i^{\beta}(x_i, y_j)} \tag{2.1.9}$$

and similarly

$$H(Y;U) = -\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} u_i\ (x_i\ ,y_j) p_i^{\beta}(x_i, y_j) \log \mathrm{p}(y_j)}{\sum_{i=}^{n}\sum_{j=1}^{n} u_i\ (x_{i,}\ y_j) p_i^{\beta}(x_i, y_j)} \tag{2.1.10}$$

and we have, $H_{\alpha,\beta,j}(P;U) = \frac{1}{1-\alpha}\left[\left(\frac{\sum_{i=1}^{n} u_i p_i^{\alpha+\beta-1} n^{\alpha-1}}{\sum_{i=1}^{n} u_i\, p_i^{\beta}}\right)^{j} - (n^{\alpha-1})^{j}\right]$

$$= \frac{\left(n^{\alpha-1}\right)^{j}}{1-\alpha}\left[\left(\frac{\sum_{i=1}^{n} u_i p_i^{\alpha+\beta-1}}{\sum_{i=1}^{n} u_i\, p_i^{\beta}}\right)^{j} - 1\right] \tag{2.1.11}$$

For the independent distributions X and Y having U and V utilities distribution respectively, then from equation (2.1.11)

We have from equation (2.1.11) reduces to

$$H_{\alpha,\beta,j} = \frac{1-\alpha}{\left(n^{\alpha-1}\right)^{j}} H_{\alpha,\beta,j}\ (P;U) + 1 = \left(\frac{\sum_{i=1}^{n} u_i p_i^{\alpha+\beta-1} n^{\alpha-1}}{\sum_{i=1}^{n} u_i\, p_i^{\beta}}\right)^{j} \tag{2.1.12}$$

$$H(X.Y;U.V) = \frac{\left(n^{\alpha-1}\right)^{j}}{1-\alpha}\left[\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} u_i \cdot v_j\ \ (x_i, y_j)^{\alpha+\beta-1}}{\sum_{i=1}^{n}\sum_{j=1}^{m} u_i \cdot v_j\ \ (x_i, y_j)^{\beta}}\right)^{j} - 1\right]$$

$$= \frac{\left(n^{\alpha-1}\right)^{j}}{1-\alpha}\left[\left(\frac{\left[\sum_{i=1}^{n} u_i\,.(x_i)^{\alpha+\beta-1} . \sum_{j=1}^{m} v_j\,.(y_j)^{\alpha+\beta-1}\right]}{\sum_{i=1}^{n} u_i\,.(x_i)^{\beta}\ \ . \sum_{j=1}^{m} v_j (y_j)^{\beta}}\right)^{j} - 1\right]$$

$$= \frac{(n^{\alpha-1})^j}{1-\alpha}\left[\left(\frac{\sum_{i=1}^n u_i.(x_i)^{\alpha+\beta-1}}{\sum_{i=1}^n u_i.(x_i)^\beta}\right)^j.\left(\frac{\sum_{j=1}^m v_j.(y_j)^{\alpha+\beta-1}}{\sum_{j=1}^m v_j(y_j)^\beta}\right)^j - 1\right]$$

$$= \frac{(n^{\alpha-1})^j}{1-\alpha}\left\{\left[\frac{1-\alpha}{(n^{\alpha-1})^j} H_{\alpha,\beta,j}\ (X;U) + 1\right].\left[\frac{1-\alpha}{(n^{\alpha-1})^j} H_{\alpha,\beta,j}\ (Y;U) + 1\right] - 1\right\}$$

$$= \frac{(n^{\alpha-1})^j}{1-\alpha}\left\{\left(\frac{1-\alpha}{(n^{\alpha-1})^j}\right)^2 H_{\alpha,\beta,j}\ (X;U).H_{\alpha,\beta,j}\ (Y;U) + \frac{1-\alpha}{(n^{\alpha-1})^j} H_{\alpha,\beta,j}\ (Y;U) + \frac{1-\alpha}{(n^{\alpha-1})^j} H_{\alpha,\beta,j}\ (X;U) + 1 - 1\right\}$$

$$= \frac{1-\alpha}{(n^{\alpha-1})^j} H_{\alpha,\beta,j}\ (X;U).H_{\alpha,\beta,j}\ (Y;U) + H_{\alpha,\beta,j}\ (Y;U) + H_{\alpha,\beta,j}\ (Y;U) \tag{2.1.13}$$

This is well known functional equation and that measures defined by equation non additive.

and also we have from equation (2.1.7) reduces to

$$\mathrm{H}\left(\frac{\mathrm{Y}}{\mathrm{X}};\mathrm{U}\right) = -\frac{\sum_{i=1}^n\sum_{j=1}^m u_i\ (x_i\ ,y_j)p_i^\beta(x_i,y_j)\log \mathrm{p}\left({}^{y_i}/_{x_j}\right)}{\sum_{i=}^n\sum_{j=1}^n u_i\ (x_{i,}\ y_j)p_i^\beta(x_i,\ y_j)} \tag{2.1.14}$$

Now we consider

$$H(X,Y;U) = -\ \frac{\sum_{i=1}^n\sum_{j=1}^m u_i\ (x_i\ ,y_j)p_i^\beta(x_i,y_j)\log \mathrm{p}(x_i\ ,y_j)}{\sum_{i=}^n\sum_{j=1}^n u_i\ (x_{i,}\ y_j)p_i^\beta(x_i,y_j)}$$

Or

$$H(X,Y;U) = -\ \frac{\sum_{i=1}^n\sum_{j=1}^m u_i\ (x_i\ ,y_j)p_i^\beta(x_i,y_j)\log p(x_i)\log \mathrm{p}\left(\frac{y_j}{x_i}\right)}{\sum_{i=}^n\sum_{j=1}^n u_i\ (x_{i,}\ y_j)p_i^\beta(x_i,y_j)}$$

Or

$$H(X,Y;U) = -\ \frac{\sum_{i=1}^n\sum_{j=1}^m u_i\ (x_i\ ,y_j)p_i^\beta(x_i,y_j)\log p(x_i)}{\sum_{i=}^n\sum_{j=1}^n u_i\ (x_{i,}\ y_j)p_i^\beta(x_i,y_j)} - \frac{\sum_{i=1}^n\sum_{j=1}^m u_i\ (x_i\ ,y_j)p_i^\beta(x_i,y_j)\log \mathrm{p}\left(\frac{y_j}{x_i}\right)}{\sum_{i=}^n\sum_{j=1}^n u_i\ (x_{i,}\ y_j)p_i^\beta(x_i,y_j)}$$

From equations (2.1.9) and (2.1.14) above equation reduces to

$$H(X,Y;U)\ = \mathrm{H(X;U)} + \mathrm{H}\left(\frac{\mathrm{Y}}{\mathrm{X}};\mathrm{U}\right) \tag{2.1.15}$$

and similarly, we have

$$H(X,Y;U) = -\ \frac{\sum_{i=1}^n\sum_{j=1}^m u_i\ (x_i\ ,y_j)p_i^\beta(x_i,y_j)\log p(x_i)\log \mathrm{p}\left(\frac{y_j}{x_i}\right)}{\sum_{i=}^n\sum_{j=1}^n u_i\ (x_{i,}\ y_j)p_i^\beta(x_i,y_j)}$$

Or

$$H(X,Y;U) =- \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} u_i\ (x_i\ ,y_j)p_i^{\beta}(x_i,y_j)\log p(y_j)}{\sum_{i=}^{n}\sum_{j=1}^{n} u_i\ (x_{i,}\ y_j)p_i^{\beta}(x_i,y_j)} - \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} u_i\ (x_i\ ,y_j)p_i^{\beta}(x_i,y_j)\log p(y_j)\log p\left(\frac{x_i}{y_j}\right)}{\sum_{i=}^{n}\sum_{j=1}^{n} u_i\ (x_{i,}\ y_j)p_i^{\beta}(x_i,y_j)}$$

Or $$H(X,Y;U) = \mathrm{H(Y;U)} + \mathrm{H}\left(\frac{\mathrm{X}}{\mathrm{Y}};\mathrm{U}\right) \quad (2.1.16)$$

From (2.1.15) and (2.1.16) together we have

$$H(X,Y;U) = \mathrm{H(X;U)} + \mathrm{H}\left(\frac{\mathrm{Y}}{\mathrm{X}};\mathrm{U}\right) = \mathrm{H(Y;U)} + \mathrm{H}\left(\frac{\mathrm{X}}{\mathrm{Y}};\mathrm{U}\right)$$

Or $$\mathrm{H(X;U)} + \mathrm{H}\left(\frac{\mathrm{Y}}{\mathrm{X}};\mathrm{U}\right) = \mathrm{H(Y;U)} + \mathrm{H}\left(\frac{\mathrm{X}}{\mathrm{Y}};\mathrm{U}\right)$$

Or $$\mathrm{H(X;U)} - \mathrm{H}\left(\frac{\mathrm{X}}{\mathrm{Y}};\mathrm{U}\right) = \mathrm{H(Y;U)} - \mathrm{H}\left(\frac{\mathrm{Y}}{\mathrm{X}};\mathrm{U}\right)$$

Now, letting $X = \{x_1, x_2, \ldots\ldots, x_n\}$be a set of n-letter input alphabet and $Y = \{y_1, y_2, \ldots\ldots, y_m\}$be a set of m-letter output alphabet.Consider the probability distribution function to be $\mathrm{p}(x_i)$ and $\mathrm{p}(y_j)$, $\forall\ i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$ based on X and Y, respectively. For every i, let $u(x_i)$ be the utility that corresponds to $\mathrm{p}(x_i)$. Therefore, the definition of average beneficial mutual information is

$$I(X,Y;U) = \mathrm{H(X;U)} - \mathrm{H}\left(\frac{\mathrm{X}}{\mathrm{Y}};\mathrm{U}\right) \quad (2.1.17)$$

Thus, a memory-less channel's weighted channel capacity or channel capacity with utilities is determined by

$$C(U) = \max_{x} I(X,Y;U)$$

Hence the proof.

Utility and Mutual Information

Claim 2.2

Let $X = \{x_1, \ldots, x_n\}$ be a set of input alphabet of source X and $Y = \{y_1, \ldots, y_n\}$ be the set of output alphabet of N^{th} destination Y_N. Then relationship between the source distribution for useful mutual information and the output state for N identical cascaded sub-channels will be

$$H(Y_N, U) - \sum_{i=1}^{n} a_i \lambda_i^N r_i - 1 = N$$

and hence $C(U) = \max I(X, Y_N, U) - 1 = N$

Proof: Let $X = \{x_1, x_2, \ldots\ldots, x_n\}$ be a set of input alphabet of source X and $Y = \{y_1, y_2, \ldots\ldots, y_n\}$ be the set of output alphabet of N^{th} destination Y_N . Let $p(x_i), i = 1,2,\ldots,n$ and $p(y_j), j =$

$1,2, \ldots, n$, be the probability distribution function defined on X and Y_N respectively. Hence, the mutual information between source X and Y_N destination can be given as

$$I(X, Y_N) = -\sum_{j=1}^{n} p(y_j) \log p(y_j) - H^T p(X) \tag{2.2.1}$$

Let the transition matrix of the i^{th} sub-channel be represented by A_i. Given that Ai is presumed to be nonsingular, it must be square. Given that there are n letters to be sent, or A_i, the input and output state column vectors for the i^{th} sub-channel are $p(y_k)$ and $p(y_{k-1})$.

$$p(y_k) = {A_i}^T p(y_{k-1})$$

$p(y_0) \approx p(X)$, where $p(X)$ represents the probability distribution vector for the source X, is also true. The relationship betweenthe source distribution and the output state for N identical sub-channels is

$$p(X) = [(A^T)^{-1}]^N p(y_N) \tag{2.2.2}$$

Since all i^{th} sub-channels are assumed to be the same, the i^{th} subscript on A^T is ignored in this case. Kerline (1966) states that $p(X)$ can be written as follows if B represents A^{-1}:

$$p(X) = \sum_{i=1}^{m} a_i \lambda_i^N v_i \tag{2.2.3}$$

where the appropriate Eigen vector and i^{th} Eigen value of the matrix B^T are denoted by v_i and λ_i, respectively. An analysis of the set of equations yields the coefficients a_i.

If H is the column vector with i^{th} component

$$H\left(\frac{y_N}{x_i}\right) = -\sum_{i,j=1}^{n} p\left(\frac{y_j}{x_i}\right) \log p\left(\frac{y_j}{x_i}\right) \tag{2.2.4}$$

From equation (2.2.1) we have

$$I(X, Y_N) = -\sum_{j=1}^{n} (y_j p) \log p(y_j) - H^T p(X) \tag{2.2.5}$$

$$= -\sum_{j=1}^{n} p(y_j) \log p(y_j) - \sum_{i=1}^{m} H^T a_i \lambda_i^N v_i$$

Assume that the utility u_i is the same as the input probabilities $p(x_i)$. For the N cascaded channels, we are assuming that the utility for each sub-channel will be the same.
Next, let u_i be the utility that corresponds to the input probabilities $p(x_i)$. This defines the "useful" mutual information for N identical cascaded channels. For the N cascaded channels, we are assuming that the utility for each sub-channel will be the same.
Then, for N identical cascaded channels, the "useful" mutual information is defined as

$$I(X, Y_N, U) = -H(Y_N, U) + \sum_{i=1}^{m} a_i \lambda_i^N r_i \tag{2.2.6}$$

Where $r_i = -H^T v_i$

So the channel capacity of N identical cascaded channels is given by

$$C = \max_{Y_N} I(X; Y_N)$$

and $$H\left(\frac{y_N}{x_i}; U\right) = - \frac{\sum_{i,j=1}^{n} u_j\, p\left(\frac{y_j}{x_i}\right) \log p\left(\frac{y_j}{x_i}\right)}{\sum_{i,j=1}^{n} u_j\ p(x_i, y_j)} \tag{2.2.7}$$

Also we know that

$$H(Y_N\,; U) = - \frac{\sum_{j=1}^{n} u_j\, p(y_j) \log p(y_j)}{\sum_{j=1}^{n} u_j\ p(y_j)}$$

Or

$$H(Y_N\,; U) = - \frac{1}{k} \sum_{j=1}^{n} u_j\, p(y_j) \log p(y_j) \tag{2.2.8}$$

Where $k = \sum_{j=1}^{n} u_j\ p(y_j)$

For all non-negative real values of probability $p(x_i)$ and corresponding positive utilities u_i, the equation (2.2.6) can be considered defined. We must maximize (2.2.6) based on the requirement $\sum_{i=1}^{n} p(x_i) = 1$. We use the Lagrange's multiplier approach and assume that the answer contains no $p(x_i) < 0$. Aside from that,

$$\phi = I(X, Y_N, U) + N\left(\sum_{i=1}^{n} p(x_i) - 1\right)$$

$$= - H(Y_N, U) + \sum_{i=1}^{m} a_i \lambda_i^N r_i + N\left(\sum_{i=1}^{n} p(x_i) - 1\right)$$

$$= - \frac{1}{k} \sum_{j=1}^{n} u_j\, p(y_j) \log p(y_j) + \sum_{i=1}^{n} a_i \lambda_i^N r_i + \mathrm{N} \sum_{i=1}^{n} p(x_i) - N$$

$$= - \sum_{j=1}^{n} u_j\, p(y_j) \log p(y_j) + K \sum_{i=1}^{n} a_i \lambda_i^N r_i + NK \sum_{i=1}^{n} p(x_i) - N\,K$$

Or $$\sum_{j=1}^{n} u_j\, p(y_j) \log p(y_j) + NK - K = K \sum_{i=1}^{n} a_i \lambda_i^N r_i$$

Where $\sum_{i=1}^{n} p(x_i) = 1$

Or $$H(Y_N, U) + N - 1 - \sum_{i=1}^{n} a_i \lambda_i^N r_i = 0$$

Or $$H(Y_N, U) - \sum_{i=1}^{n} a_i \lambda_i^N r_i = 1 + N$$

Hence $$C(U) = max = I(X, Y_N, U) = 1 + N$$

and this completes the proof.

Claim 2.3

For the average useful mutual information in N identical cascade channels process the generalized logarithmic input probabilities holds the inequality

$$\sum_{j=1}^{n} \sum_{k=1}^{l} u_j p_0(y_j) p_k(y_j) \log p_k(y_j) \geq \sum_{j=1}^{n} \sum_{k=1}^{l} u_j p_k(y_j) p_0(y_j) \log p_k(y_j)$$

which implies

$$\sum_{j=1}^{n} \sum_{k=1}^{l} u_j p_0(y_j) p_k(y_j) \log p_k(y_j) \geq \sum_{j=1}^{n} \sum_{k=1}^{l} u_j \log p_k(y_j)$$

and ΔI is a convex function of the input probabilities.

Proof: Let us define input distribution

$$P_0(x) = \sum_{k=1}^{l} b_k P_k(x)$$

and let $b_1, b_2, \ldots\ldots\ldots b_l$ are non- negative number such that

$$\sum_{k=1}^{l} b_k = 1 \tag{2.3.1}$$

The average "useful" mutual information corresponding to $P_0(x)$ will then be demonstrated to satisfy

$$I_0(X, Y_N, U) \geq \sum_{k=1}^{l} I_k(X, Y_N, U)\, b_k$$

Where $I_k(X, Y, U)$ is the average useful mutual information corresponding to input distribution.$P_k(x)$.

Let $\Delta I = I_0(X, Y_N, U) - \sum_{k=1}^{l} I_k(X, Y_N, U)\, b_k$ (2.3.2)

We have from equations (2.2.3) and (2.3.1) values substituting in equations (2.3.2) ,we get

$$\Delta I = H_0(Y_N, U) - H_0\left({}^{Y_N}/_{X}, U\right) + \sum_{k=1}^{l}\left[H_k(Y_N, U) - H_k\left({}^{Y_N}/_{X}, U\right)\right]\sum_{k=1}^{l} b_k$$

$$= -\frac{\sum_{j=1}^{n} u_j\, p_0(y_j)\log p_0(y_j)}{\sum_{j=1}^{n} u_j\, p_0(y_j)} - H_0\left({}^{Y_N}/_{X}, U\right) + \sum_{k}^{l} H_k(Y_N, U) - \sum_{k=1}^{l} H_k\left({}^{Y_N}/_{X}, U\right) \quad (2.3.3)$$

Since $p_0\left(\frac{y_j}{x_i}\right)$is the sum of the channel probabilities $p_k\left(\frac{y_j}{x_i}\right)$,$k = 1,2, \ldots.., l$. Therefore, we have

$$H_0\left({}^{Y_N}/_{X}, U\right) = -\frac{\sum_{i,j=1}^{n} u_j\, p_0\left(\frac{y_j}{x_i}\right)\log p_0\left(\frac{y_j}{x_i}\right)}{\sum_{i,j=1}^{n} u_j\, p_0(x_i, y_j)}$$

$$= -\sum_{k=1}^{l} H_k\left({}^{Y_N}/_{X}, U\right) \tag{2.3.4}$$

From equation (2.3.3),we get

$$\Delta I = H_0(Y_N, U) - H_0\left({}^{Y_N}/_{X}, U\right) + \sum_{k=1}^{l} H_k(Y_N, U) - \sum_{k=1}^{l} H_k\left({}^{Y_N}/_{X}, U\right)$$
$$= H_0(Y_N, U) - H_0\left({}^{Y_N}/_{X}, U\right) + \sum_{k=1}^{l} H_k(Y_N, U) + H_0\left({}^{Y_N}/_{X}, U\right)$$
$$= H_0(Y_N, U) + \sum_{k=1}^{l} H_k(Y_N, U)$$

$$= -\frac{\sum_{j=1}^{n}\sum_{k=1}^{l} u_j\, p_0(y_j)\log p_0(y_j)}{\sum_{j=1}^{n} u_j\, p_0(y_j)} - \frac{\sum_{j=1}^{n}\sum_{k=1}^{l} u_j\, p_k(y_j)\log p_k(y_j)}{\sum_{j=1}^{n} u_j\, p_k(y_j)}$$

$$= -\frac{\sum_{j=1}^{n}\sum_{k=1}^{l} u_j\, p_0(y_j) p_k(y_j)\log p_0(y_j) - .\sum_{j=1}^{n}\sum_{k=1}^{l} u_j\, p_k(y_j) p_0(y_j)\log p_k(y_j)}{\sum_{j=1}^{n} u_j\, p_k(y_j) p_0(y_j)}$$

By the generalized Shannon’s inequality, we have

$$\sum_{j=1}^{n}\sum_{k=1}^{l} u_j\, p_0(y_j) p_k(y_j) \log p_k(y_j) \geq \sum_{j=1}^{n}\sum_{k=1}^{l} u_j\, p_k(y_j) p_0(y_j) \log p_k(y_j)$$

with equality only if $p_0(y_j) = p_k(y_j)$for each j, it implies

$$\sum_{j=1}^{n}\sum_{k=1}^{l} u_j \log p_k(y_j) \geq \sum_{j=1}^{n}\sum_{k=1}^{l} u_j \log p_k(y_j) \qquad (2.3.5)$$

Hence (2.3.5) gives$\Delta I \geq 0$, since $\sum_{j=1}^{n}\sum_{k=1}^{l} u_j\, p_k(y_j) \geq 0$
is always non-negative. Thus, the theorem.

Conclusion:

The channel capacity and its various forms have been covered in this paper. Additionally, we have talked about the classification of the Discrete Memory-less channel. We have a model for the discrete channel capacity that memory-less channel having a proven claims and utility. We have examined the many kinds of cascaded channels in this paper. Additionally, we have established claims and defined the channel capacity of N identical cascaded channels. Finding the best arrangement for n randomly selected binary channels is still an unresolved challenge. Nevertheless, we have presented a class of information theoretic problems that estimate the channel capacity with utility and deal with channel performance in cascade.
Communication over a single channel differs from the difficulty of communication through cascaded channels in that the transmitter in the latter scenario is fully aware of what it should send.

References :

1. Abu Dayya, A. and Beaulieu,N.C.(1994): "Analysis of switched diversitysystems on generalized-fading channels," IEEE Trans. Commn., vol. 42, pp. 1813-1831, Nov.
2. Abu Dayya, A,and Beaulieu, N. C. (1954): "Switched diversity on microcellularRician channels," IEEE Trans. Veh. Technol., vol. 43, pp. 970-976, Nov.
3. Adamchik, V. S. and Marichev,O. I. (1990): "The algorithm for calculatingintegrals of hypergeometric type functions and its realization in REDUCEsystem," in Proc. Internat'l Conf. Symbolic and Algebraic Computation,", pp. 212-224.
4. Ahlswede, R. (1998): "The weak capacity of averaged channels," Z.Wahrschhkhkeitstheorie uenv. Geb.; vol. 11, pp. 61-73.
5. Alouini, M. S and Simon,M. K. (2003):"Postdetection switched combining: asimple diversity scheme with improved BER performance," IEEE Trans.Commun., vol. 51, pp. 1591-1602..
6. Alouini, M. S. and Goldsmith, A. (1999): "Capacity of Rayleigh fading channelsunder different adaptive transmission and diversity-combining techniques", IEEE Trans. Veh. Technol., vol. 48, pp. 1165-1181.

7. Cover,T. M. and J. A. Thomas,J. A. (1991): Ekmnts of Information Theory.New York Wiley.

8. Cover,T. M. :"Broadcast channels," IEEE Tmns. Inform. Theory,

9. Dobrushin,R. L.(1964): "General formulation of Shannon's main theorem in information theory," Amer. Math. Soc. T"., vol. 33, pp. 323-438, AMs, Providence, RI..

10. Feinstein,A.(1954): "A new basic theorem of information theory," IRETmns. Inform. Theory, vol. IT-4, pp. 2-22,1954.

11. Gradshteyn I,S.and Ryzhik,I.M.: Table of Integrals, Series, and Kieffer,J.C. "A general formula for the capacity of stationary non-anticipatory channels," Ifomt. Contr., vol. 26, pp. 381-391,

12. Gray ,R. M and Omstein,D. S.(1979): "Block coding for discrete stationary d-bar continuous noisy channels," IEEE Trans. Inform. Theory,vol. IT-25, pp. 292-306.

13. Gray ,R. M. and Davisson,L. D.,(1977): Ergodic and Information Theory.Stroudsbourg, PA: Dowden, Hutchinson & Ross.

14. Gray,R. M. : Entropy and Information Theory. New York

15. Han T. S. and Verdu,S.(1993):"Approximation theory of output statistics," IEEE Tmns. Inform. lhory, vol. 39, pp. 752-772.

16. Hill F.S., Jr. and M. A. Blanco ,M.A.(1973): ",Random geometric series andinter-symbol interference," IEEE Trans. Inform. Theory, vol. IT-19,pp. 326-335.

17. Lazarakis et al.,F..(1994): "Average channel capacity in a mobile radio environment with Rician statistics," IEICE Trans. Commun., vol. E77-B, pp.971-977.

18. Lee,W. C. Y. (1990): "Estimate of channel capacity in Rayleigh fading environment," IEEE Trans. Veh. Technol., vol. 39, pp. 187-189, Aug. 1990.

19. Nedoma,J.(1957): "The capacity of a discrete channel," in Pmc. 1st& gue Conf. Information Theory, Statist. Decision Functions, Random, Pmessa, Prague, 1957, pp. 143-181.

20. Nedoma,J.(1960): "On nonergodic channels," in Pmc. 2nd hgue Conf.Inform. Theory, Statist. Deciswn Functions, Random Processes,Prague, 1960, pp. 363-395.

21. Parthasarathy,K. R.(1963):‘‘Effective entropy rate and transmission ofinformation through channels with additive random noise,"Sankhya, vol. A25, no. 1, pp. 75-84.

22. Pinsker,M. S. (1964): Information and Information Stabiliw of RandomVariables and Processes. San Francisco: Holden-Day.

23. Sagias et al.,N. C.(2003): "Performance analysis of switched diversity inWeibull fading," Electron. Lett., vol. 39, pp. 1472-1474.

24. Sagias et al.,N. C.(2004):"Channel capacity and second order statistics inWeibull fading," IEEE Commun. Lett., vol. 8, pp. 377-379.

25. Sagias et al.,N. C.(2004): "Performance analysis of dual selection diversity incorrelated Weibull fading channels," IEEE Trans. Commun., vol. 52, pp.1063-1067.

26. Shannon C. E.(1990):" A mathematical theory of communication", Bell System Technical Journal,27(1948): 379-423, 623-659

27. Simon ,M. K. and M.S. Alouini,M.S.(2001): Digital Communication over FadingChannels, 1st ed. New York: Wiley.

28. Tellambura et al,,C.(2001): "Unified analysis of switched diversity systems inindependent and correlated fading channels," IEEE Trans. Commun., vol.49, pp. 1955-1965, Nov.

29. Verma, R. K. (2023): Some Fuzzy Noiseless Coding Theorems Connected with Non-Shannon Information Measures and Their Bounds, International Journal of Fuzzy Mathematics and Systems (IJFMS), 2023, Vol. 13(1), pp. 1-9.

30. Verma, R. K. (2023): Optimal Code Word Length Via Modified Verma Information Measures for Discrete Noiseless Channel in Intuitionistic Fuzzy Environment, International Journal of Fuzzy Mathematics and Systems (IJFMS), 2023, Vol. 13(1), pp. 11-20.

31. Verma, R. K. (2023): Modified Version of Verma Measures of Information and Their Kinship with Past Information Measures, International Journal of Pure and Applied Mathematical Sciences (IJPAMS), 2023, Vol. 16(1), pp. 17-24.

32. Verma, R. K. (2023): On Optimization Policy for Verma Entropy by Dynamic Programming, Asian Journal of Probability and Statistics (AJPAS), 2023, Vol. 21 (4), pp. 14-21, DOI: 10.9734/AJPAS/2023/v21i4469.

33. Verma, R. K. (2023): On Optimal Channel Capacity Theorems via Verma Information Measure with Two-Sided Input in Noisy State, Asian Journal of Probability and Statistics (AJPAS), 2023, Vol. 22 (2), pp. 1-7, DOI:10.9734/AJPAS/2023/v22i2478.

34. Verma, R. K. (2023): Possible Innovatory Propagation Models and Their Corresponding Models in Pythagorean Fuzzy Set Via Modified Verma Hybrid Shannon Entropy, Bulletin of Mathematics and Statistics Research (BOMSR), 2023, Vol.11.Issue.2, pp.127-132, DOI:10.33329/bomsr.11.2.127.

35. Verma, R. K. (2023): On Quantile Approach of Generalized Quasi-Log Probabilistic Theoretic Measure with Edge Detection and Object Pixel Threshold Grey Values, Global Journal of Mathematics and Mathematical Sciences (GJMMS), Vol. 13 (2), pp. 93-102.

36. Ko et al.,Y.C. (2000): "Analysis and optimization of switched diversity systems,"IEEE Trans. Veh. Technol., vol. 49, pp. 1813-1831.

37. Ziv,J, "The Capacity of the General Time-Discrete Channel withFinite Alphabet," Information and Control, vol. 14, pp. 233-251.