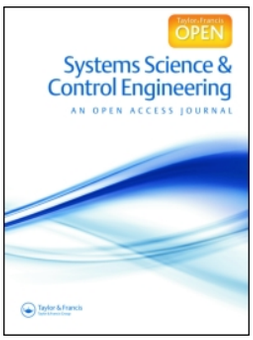


Systems Science & Control Engineering

An Open Access Journal

ISSN: (Print) 2164-2583 (Online) Journal homepage: https://www.tandfonline.com/loi/tssc20

The paradigm of complex probability and Claude Shannon's information theory

Abdo Abou Jaoude

To cite this article: Abdo Abou Jaoude (2017) The paradigm of complex probability and Claude Shannon's information theory, Systems Science & Control Engineering, 5:1, 380-425, DOI: 10.1080/21642583.2017.1367970

To link to this article: https://doi.org/10.1080/21642583.2017.1367970

Published online: 29 Aug 2017.

SYSTEMS SCIENCE & CONTROL ENGINEERING: AN OPEN ACCESS JOURNAL, 2017
VOL. 5, NO. 1, 380–425
https://doi.org/10.1080/21642583.2017.1367970

Taylor & Francis
Taylor & Francis Group

The paradigm of complex probability and Claude Shannon's information theory

Abdo Abou Jaoude

Department of Mathematics and Statistics, Faculty of Natural and Applied Sciences, Notre Dame University-Louaize, Zouk Mosbeh, Lebanon

ABSTRACT

Andrey Kolmogorov put forward in 1933 the five fundamental axioms of classical probability theory. The original idea in my complex probability paradigm is to add new imaginary dimensions to the experiment real dimensions which will make the work in the complex probability set totally predictable and with a probability permanently equal to one. Therefore, adding to the real set of probabilities $\mathscr{R}$ the contributions of the imaginary set of probabilities $\mathscr{M}$ will make the event in $\mathscr{C} = \mathscr{R} + \mathscr{M}$ absolutely deterministic. It is of great importance that stochastic systems become totally predictable since we will be perfectly knowledgeable to foretell the outcome of all random events that occur in nature. Hence, my purpose here is to link my complex probability paradigm to Claude Shannon's information theory that was originally proposed in 1948. Consequently, by calculating the parameters of the new prognostic model, we will be able to determine the magnitude of the chaotic factor, the degree of our knowledge, the complex probability, the self-information functions, the message entropies, and the channel capacities in the probability sets $\mathscr{R}$ and $\mathscr{M}$ and $\mathscr{C}$ and which are all functions of the message real probability subject to chaos and random effects.

ARTICLE HISTORY
Received 7 June 2017
Accepted 12 August 2017

KEYWORDS
Complex set; complex probability; probability norm; degree of our knowledge; chaotic factor; self-information; message entropy; channel capacity

Nomenclature

$\mathscr{R}$	Real probability set of events
$\mathscr{M}$	Imaginary probability set of events
$\mathscr{C}$	Complex probability set of events
i	The imaginary number where $i = \sqrt{-1}$
EKA	Extended Kolmogorov's Axioms
CPP	Complex Probability Paradigm
P_{rob}	Probability of any event
P_r	Probability in the real set $\mathscr{R}$ = message real probability
P_m	Probability in the imaginary set $\mathscr{M}$ corresponding to the real probability in $\mathscr{R}$ = message complementary probability in $\mathscr{M}$
P_m/i	Message complementary probability in $\mathscr{R}$
Pc	Probability of an event in $\mathscr{R}$ with its associated event in $\mathscr{M}$, it is the message probability in the complex set $\mathscr{C}$
Z	Complex probability number and vector, it is the sum of P_r and P_m
DOK	$= \lvert Z\rvert^2 =$ Degree of Our Knowledge of the random message, it is the square of the norm of Z.
Chf	Chaotic Factor of the random message
MChf	Magnitude of the Chaotic Factor of the random message
I_2	Surprisal self-information in base 2
RI_2	Rescaled surprisal in base 2
$\bar{I}_2$	Expectancy self-information in base 2
$R\bar{I}_2$	Rescaled expectancy in base 2
Φ	Simulation rescaling factor
H	Information entropy
H_b	Information binary entropy
H_b^R	$= H_b =$ Information binary entropy in $\mathscr{R}$
$\bar{H}_b^R$	Information complementary binary entropy in $\mathscr{R}$
$NegH_b^R$	Information negative binary entropy in $\mathscr{R}$
H_b^M	Information binary entropy in $\mathscr{M}$
H_b^C	Information binary entropy in $\mathscr{C}$
$-\text{logit}_2$	First derivative of the binary entropy in base 2
C	Channel capacity
BSC	Binary Symmetric Channel
C_{BSC}^R	The real capacity in $\mathscr{R}$ of a *BSC*
$\bar{C}_{BSC}^R$	The complementary real capacity in $\mathscr{R}$ of a *BSC*
C_{BSC}^M	The complex capacity in $\mathscr{M}$ of a *BSC*
C_{BSC}^C	The capacity in $\mathscr{C}$ of a *BSC*

1. Introduction

Firstly, information theory studies the quantification, storage, and communication of information. It was originally proposed by Claude Elwood Shannon in 1948 to find

CONTACT Abdo Abou Jaoude abdoaj@idm.net.lb

fundamental limits on signal processing and communication operations such as data compression, in a landmark paper entitled 'A Mathematical Theory of Communication'. Now this theory has found applications in many other areas, including statistical inference, natural language processing, cryptography, neurobiology (Rieke, Warland, van Steveninck, & Bialek, 1997), the evolution (Huelsenbeck, Ronquist, Nielsen, & Bollback, 2001) and function of molecular codes (Allikmets et al., 1998), model selection in ecology (Burnham & Anderson, 2002), thermal physics (Jaynes, 1957), quantum computing, linguistics, plagiarism detection (Bennett, Li, & Ma, 2003), pattern recognition, and anomaly detection (David & Anderson, 2003).

A key measure in information theory is 'entropy'. Entropy quantifies the amount of uncertainty involved in the value of a random variable or the outcome of a random process. For example, identifying the outcome of a fair coin flip (with two equally likely outcomes) provides less information (lower entropy) than specifying the outcome from a roll of a die (with six equally likely outcomes). Some other important measures in information theory are mutual information, channel capacity, error exponents, and relative entropy (Fazlollah, 1994 [1961]; Ash, 1990 [1965]; Gibson, 1998; Shannon, 1948; Hartley, 1928).

Applications of fundamental topics of information theory include lossless data compression (e.g. ZIP files), lossy data compression (e.g. MP3s and JPEGs), and channel coding (e.g. for Digital Subscriber Line (DSL)) (Kelly Jr, 1956; Kolmogorov, 1968; Landauer, 1993).

Moreover, the field is at the intersection of mathematics, statistics, computer science, physics, neurobiology, and electrical engineering. Its impact has been crucial to the success of the Voyager missions to deep space, the invention of the compact disc, the feasibility of mobile phones, the development of the Internet, the study of linguistics and of human perception, the understanding of black holes, and numerous other fields. Important subfields of information theory include source coding, channel coding, algorithmic complexity theory, algorithmic information theory, information-theoretic security, and measures of information. (Arndt, 2004; Ash, 1990; Gallager, 1968; Landauer, 1961; Timme, Alford, Flecker, & Beggs, 2012).

Furthermore, information theory studies the transmission, processing, utilisation, and extraction of information. Abstractly, information can be thought of as the resolution of uncertainty. In the case of communication of information over a noisy channel, this abstract concept was made concrete in 1948 by Claude Shannon in his paper 'A Mathematical Theory of Communication', in which 'information' is thought of as a set of possible messages, where the goal is to send these messages over a noisy channel, and then to have the receiver reconstruct the message with low probability of error, in spite of the channel noise. Shannon's main result, the noisy-channel coding theorem showed that, in the limit of many channel uses, the rate of information that is asymptotically achievable is equal to the channel capacity, a quantity dependent merely on the statistics of the channel over which the messages are sent (Rieke et al., 1997; Cover & Thomas, 2006; Csiszar & Korner, 1997; Goldman, 1968; MacKay, 2003; Mansuripur, 1987).

Information theory is closely associated with a collection of pure and applied disciplines that have been investigated and reduced to engineering practice under a variety of rubrics throughout the world over the past half century or more: adaptive systems, anticipatory systems, artificial intelligence, complex systems, complexity science, cybernetics, informatics, machine learning, along with systems sciences of many descriptions. Information theory is a broad and deep mathematical theory, with equally broad and deep applications, amongst which is the vital field of coding theory (McEliece, 2002; Pierce, 1961; Reza, 1961).

Additionally, coding theory is concerned with finding explicit methods, called *codes*, for increasing the efficiency and reducing the error rate of data communication over noisy channels to near the channel capacity. These codes can be roughly subdivided into data compression (source coding) and error-correction (channel coding) techniques. In the latter case, it took many years to find the methods Shannon's work proved were possible. A third class of information theory codes are cryptographic algorithms (both codes and ciphers). Concepts, methods and results from coding theory and information theory are widely used in cryptography and cryptanalysis. Information theory is also used in information retrieval, intelligence gathering, gambling, statistics, and even in musical composition (Shannon & Weaver, 1949; Stone, 2014; Yeung, 2002).

The landmark event that *established* the discipline of information theory and brought it to immediate worldwide attention was the publication of Claude E. Shannon's classic paper 'A Mathematical Theory of Communication' in the *Bell System Technical Journal* in July and October 1948. Prior to this paper, limited information-theoretic ideas had been developed at Bell Labs, all implicitly assuming events of equal probability. Harry Nyquist's 1924 paper, *Certain Factors Affecting Telegraph Speed*, contains a theoretical section quantifying 'intelligence' and the 'line speed' at which it can be transmitted by a communication system, giving the relation $W = K \log m$ (recalling Ludwig Boltzmann's constant), where

W is the speed of transmission of intelligence, m is the number of different voltage levels to choose from at each time step, and K is a constant. Ralph Hartley's 1928 paper, *Transmission of Information*, uses the word *information* as a measurable quantity, reflecting the receiver's ability to distinguish one sequence of symbols from any other, thus quantifying information as:

$$H = \log S^n = n \log S,$$

where S was the number of possible symbols, and n the number of symbols in a transmission. The unit of information was therefore the decimal digit, much later renamed the hartley in his honour as a unit or scale or measure of information. Alan Turing in 1940 used similar ideas as part of the statistical analysis of the breaking of the German second world war Enigma ciphers (Brillouin, 1962 [2004]; Gleick, 2011; Yeung, 2008).

Much of the mathematics behind information theory with events of different probabilities were developed for the field of thermodynamics by Ludwig Boltzmann and Josiah Willard Gibbs. Connections between information-theoretic entropy and thermodynamic entropy, including the important contributions by Rolf Landauer in the 1960s, are explored in *Entropy in thermodynamics and information theory* (Khinchin, 1957; Leff & Rex, 1990; Logan, 2014).

In addition, in Shannon's revolutionary and groundbreaking paper, the work for which had been substantially completed at Bell Labs by the end of 1944, Shannon for the first time introduced the qualitative and quantitative model of communication as a statistical process underlying information theory, opening with the assertion that:

> "The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point."

With it came the ideas of:

- the information entropy and redundancy of a source, and its relevance through the source coding theorem;
- the mutual information, and the channel capacity of a noisy channel, including the promise of perfect loss-free communication given by the noisy-channel coding theorem;
- the practical result of the Shannon–Hartley law for the channel capacity of a Gaussian channel; as well as
- the bit—a new way of seeing the most fundamental unit of information.

Also, some applications of information theory to other fields are:

1.1. Intelligence uses and secrecy applications

Information theoretic concepts apply to cryptography and cryptanalysis. Turing's information unit, the ban, was used in the Ultra project, breaking the German Enigma machine code and hastening the end of World War II in Europe. Shannon himself defined an important concept now called the unicity distance. Based on the redundancy of the plaintext, it attempts to give a minimum amount of ciphertext necessary to ensure unique decipherability. Information theory leads us to believe it is much more difficult to keep secrets than it might first appear. A brute force attack can break systems based on asymmetric key algorithms or on most commonly used methods of symmetric key algorithms (sometimes called secret key algorithms), such as block ciphers. The security of all such methods currently comes from the assumption that no known attack can break them in a practical amount of time. Information theoretic security refers to methods such as the one-time pad that are not vulnerable to such brute force attacks. In such cases, the positive conditional mutual information between the plaintext and ciphertext (conditioned on the key) can ensure proper transmission, while the unconditional mutual information between the plaintext and ciphertext remains zero, resulting in absolutely secure communications. In other words, an eavesdropper would not be able to improve his or her guess of the plaintext by gaining knowledge of the ciphertext but not of the key. However, as in any other cryptographic system, care must be used to correctly apply even information-theoretically secure methods; the Venona project was able to crack the one-time pads of the Soviet Union due to their improper reuse of key material (Campbell, 1982; Seife, 2006; Siegfried, 2000).

1.2. Pseudorandom number generation

Pseudorandom number generators are widely available in computer language libraries and application programmes. They are, almost universally, unsuited to cryptographic use as they do not evade the deterministic nature of modern computer equipment and software. A class of improved random number generators is termed cryptographically secure pseudorandom number generators, but even they require random seeds external to the software to work as intended. These can be obtained via extractors, if done carefully. The measure of sufficient randomness in extractors is min-entropy, a value related to Shannon entropy through Rényi entropy; Rényi entropy is also used in evaluating randomness in cryptographic systems. Although related, the distinctions among these measures mean that a random variable

with high Shannon entropy is not necessarily satisfactory for use in an extractor and so for cryptography uses (Escolano, Francisco, & Pablo, 2009; Theil, 1967).

1.3. Seismic exploration

One early commercial application of information theory was in the field of seismic oil exploration. Work in this field made it possible to strip off and separate the unwanted noise from the desired seismic signal. Information theory and digital signal processing offer a major improvement of resolution and image clarity over previous analog methods (Haggerty, 1981).

1.4. Semiotics

Concepts from information theory such as redundancy and code control have been used by semioticians such as Umberto Eco and Ferruccio Rossi-Landi to explain ideology as a form of message transmission whereby a dominant social class emits its message by using signs that exhibit a high degree of redundancy such that only one message is decoded among a selection of competing ones (Noth, 1981).

1.5. Miscellaneous applications

Information theory also has applications in gambling and investing, black holes, and bioinformatics. (Wikipedia, the free encyclopedia, *Information theory*).

Finally, and to conclude, this research paper is organised as follows: After the introduction in section I, the purpose and the advantages of the present work are presented in section II. Afterward, in section III, the extended Kolmogorov's axioms and hence the complex probability paradigm with their original parameters and interpretation will be explained and illustrated. In section IV, Shannon's information theory is summarised and reviewed. Moreover, in section V, the surprisal and expectancy self-information functions are defined. In section VI the complex probability paradigm axioms are applied to the concept of binary entropy which will be extended to the imaginary and complex sets. Additionally, in section VII, the BSC capacity is also extended to the sets $\mathscr{M}$ and $\mathscr{C}$. Also, in section VIII, all the CPP parameters new model are presented. In section IX, the flowchart of this current study is shown. Furthermore, the simulations of the novel model for various discrete and continuous probability distributions are illustrated in section X. In section XI, a final analysis will be done. Finally, we conclude the work by doing a comprehensive summary in section XII, and then present the list of references cited in the current research work.

2. The purpose and the advantages of the present work

All our work in classical probability theory is to compute probabilities. The original idea in this paper is to add new dimensions to our random experiment which will make the work totally deterministic. In fact, probability theory is a nondeterministic theory by nature that means that the outcome of the stochastic events is due to chance and luck. By adding new dimensions to the event occurring in the 'real' laboratory which is $\mathscr{R}$, we make the work deterministic and hence a random experiment will have a certain outcome in the complex set of probabilities $\mathscr{C}$. It is of great importance that stochastic systems become totally predictable since we will be perfectly knowledgeable to foretell the outcome of all chaotic and random events that occur in nature like for example in statistical mechanics, in all stochastic processes, or in the well-established field of information theory. Therefore, the work that should be done is to add to the real set of probabilities $\mathscr{R}$, the contributions of $\mathscr{M}$ which is the imaginary set of probabilities that will make the event in $\mathscr{C} = \mathscr{R} + \mathscr{M}$ absolutely deterministic. If this is found to be fruitful, then a new theory in stochastic sciences would be elaborated and this to understand deterministically those phenomena that used to be random phenomena in $\mathscr{R}$. This is what I called 'The Complex Probability Paradigm' that was initiated and elaborated in my nine previous papers (Abou Jaoude, 2013a, 2013b; Abou Jaoude, 2014; Abou Jaoude, 2015a, 2015b; Abou Jaoude, El-Tawil, & Kadry, 2010; Abou Jaoude, 2016a, 2016b; Abou Jaoude, 2017).

Moreover, information theory laws firstly introduced by Claude Shannon are very well known and established. An updated follow-up of the message behaviour with time which is subject to chaotic and non-chaotic effects is done by the message probability due to its definition that evaluates the flips chances in the transmitted and received message.

Furthermore, my purpose in this current work is to link the complex probability paradigm to Claude Shannon's information theory. In fact, the system message probability derived from information theory will be included in and applied to the complex probability paradigm. This will lead to the novel and original prognostic model illustrated in this paper. Hence, by calculating the parameters of the new prognostic model, we will be able to determine the magnitude of the chaotic factor, the degree of our knowledge, the complex probability, the message surprisal and expectancy self-information functions, the message entropies, and the channel capacities in the probability sets $\mathscr{R}$ and $\mathscr{M}$ and $\mathscr{C}$ and which are all

functions of the message real probability subject to chaos and random effects.

Consequently, to summarise, the objectives and the advantages of the present work are to:

(1) Extend classical probability theory to the set of complex numbers, hence to relate probability theory to the field of complex analysis in mathematics. This task was initiated and elaborated in my nine previous papers.
(2) Do an updated follow-up of the system behaviour with time which is subject to chaos. This follow-up is accomplished by the message real, imaginary, and complex probabilities due to their definitions that evaluate the flips chances in the transmitted and received message in $\mathscr{R}$, $\mathscr{M}$, and $\mathscr{C}$, and hence to relate probability theory to information theory in an original and a new way.
(3) Apply the new probability axioms and paradigm to information theory; thus, I will extend the concepts of information theory to the complex probability set $\mathscr{C}$.
(4) Prove that any random and stochastic phenomenon can be expressed deterministically in the complex set $\mathscr{C}$.
(5) Quantify both the degree of our knowledge and the chaos magnitude of the random message and channel.
(6) Draw and represent graphically the functions and parameters of the novel paradigm associated to a random message and channel.
(7) Show that the classical concept of the message entropy is always equal to 0 in the complex set; hence, no chaos, no disorder, no unpredictability, and no ignorance exist in $\mathscr{C}$ (complex set) $= \mathscr{R}$ (real set) $+ \mathscr{M}$ (imaginary set).
(8) Prove that by adding supplementary and new dimensions to any random experiment whether it is a random channel or message or any other stochastic system we will be able to do prognostic in a deterministic way in the complex set $\mathscr{C}$.
(9) Pave the way to apply the original paradigm to other topics in statistical mechanics, in stochastic processes, and to the theory of information. These will be the subjects of my subsequent research papers.

To conclude and to summarise, compared with existing literature, the main contribution of this research paper is to apply my original complex probability paradigm to the concepts of self-information, of random entropy, and of channel capacity thus to Claude Shannon's information theory. We emphasise that it is the first time that we link the complex probability paradigm to Claude Shannon's information theory; hence, the motivation of the current work will be to extend the classical information model to the complex set of numbers by adding supplementary imaginary dimensions to the information system. Consequently, all the Shannon's random quantities will be expressed deterministically in the novel paradigm. The following figure summarises the objectives of the current research paper (Figure 1).

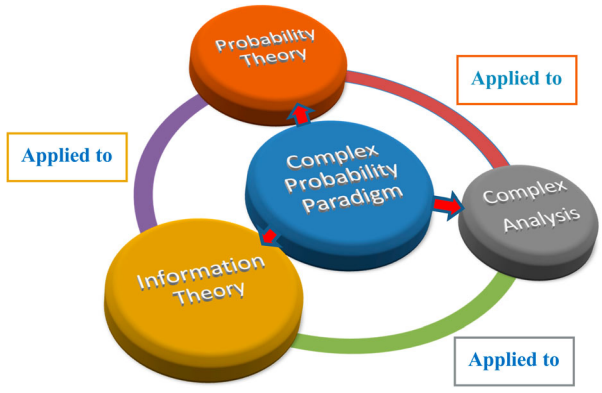


Figure 1. The diagram of the principal objectives of this research work.

3. The extended set of probability axioms

In this section, the extended set of probability axioms of the complex probability paradigm will be presented.

3.1. *The original Andrey Nikolaevich Kolmogorov set of axioms*

The simplicity of Kolmogorov's system of axioms may be surprising. Let E be a collection of elements $\{E_1, E_2, \ldots\}$ called elementary events and let F be a set of subsets of E called random events. The five axioms for a finite set E are (Benton, 1966a, 1966b; Feller, 1968; Montgomery & Runger, 2003; Walpole, Myers, Myers, & Ye, 2002):

Axiom 1: F is a field of sets.
Axiom 2: F contains the set E.
Axiom 3: A non-negative real number $P_{rob}(A)$, called the probability of A, is assigned to each set A in F. We have always $0 \leq P_{rob}(A) \leq 1$.
Axiom 4: $P_{rob}(E)$ equals 1.
Axiom 5: If A and B have no elements in common, the number assigned to their union is:

$$P_{rob}(A \cup B) = P_{rob}(A) + P_{rob}(B)$$

hence, we say that A and B are disjoint; otherwise, we have:

$$P_{rob}(A \cup B) = P_{rob}(A) + P_{rob}(B) - P_{rob}(A \cap B)$$

And we say also that:

$$P_{rob}(A \cap B) = P_{rob}(A) \times P_{rob}(B/A) = P_{rob}(B) \times P_{rob}(A/B)$$

which is the conditional probability. If both A and B are independent then:

$$P_{rob}(A \cap B) = P_{rob}(A) \times P_{rob}(B).$$

Moreover, we can generalise and say that for N disjoint (mutually exclusive) events $A_1, A_2, \ldots, A_j, \ldots, A_N$ (for $1 \leq j \leq N$), we have the following additivity rule:

$$P_{rob}\left(\bigcup_{j=1}^{N} A_j\right) = \sum_{j=1}^{N} P_{rob}(A_j)$$

And we say also that for N independent events $A_1, A_2, \ldots, A_j, \ldots, A_N$ (for $1 \leq j \leq N$), we have the following product rule:

$$P_{rob}\left(\bigcap_{j=1}^{N} A_j\right) = \prod_{j=1}^{N} P_{rob}(A_j)$$

3.2. Adding the imaginary part $\mathscr{M}$

Now, we can add to this system of axioms an imaginary part such that:

Axiom 6: Let $P_m = i \times (1 - P_r)$ be the probability of an associated event in $\mathscr{M}$ (the imaginary part) to the event A in $\mathscr{R}$ (the real part). It follows that $P_r + P_m/i = 1$ where i is the imaginary number with $i = \sqrt{-1}$.

Axiom 7: We construct the complex number or vector $Z = P_r + P_m = P_r + i(1 - P_r)$ having a norm $|Z|$ such that:

$$|Z|^2 = P_r^2 + \left(\frac{P_m}{i}\right)^2.$$

Axiom 8: Let Pc denote the probability of an event in the complex probability universe $\mathscr{C}$ where $\mathscr{C} = \mathscr{R} + \mathscr{M}$. We say that Pc is the probability of an event A in $\mathscr{R}$ with its associated event in $\mathscr{M}$ such that:

$$Pc^2 = \left(P_r + \frac{P_m}{i}\right)^2 = |Z|^2 - 2iP_rP_m$$

We can see that the system of axioms defined by Kolmogorov could be hence expanded to take into consideration the set of imaginary probabilities by adding three new axioms (Abou Jaoude, 2013a, 2013b; Abou Jaoude, 2014; Abou Jaoude, 2015a, 2015b; Abou Jaoude et al., 2010; Abou Jaoude, 2016a, 2016b; Abou Jaoude, 2017).

3.3. The purpose of extending the axioms

It is apparent from the set of axioms that the addition of an imaginary part to the real event makes the probability of the event in $\mathscr{C}$ always equal to 1. In fact, if we begin to see the set of probabilities as divided into two parts, one is real and the other is imaginary, then understanding will follow directly. The random event that occurs in the real probability set $\mathscr{R}$ (like tossing a coin and getting a head), has a corresponding probability P_r. Now, let $\mathscr{M}$ be the set of imaginary probabilities and let $|Z|^2$ be the Degree of Our Knowledge (*DOK* for short) of this phenomenon. P_r is always, and according to Kolmogorov's axioms, the probability of an event.

A total ignorance of the set $\mathscr{M}$ makes:

$$P_r = 0.5$$

and $|Z|^2$ in this case is equal to:

$$1 - 2P_r(1 - P_r) = 1 - (2 \times 0.5) \times (1 - 0.5) = 0.5.$$

Conversely, a total knowledge of the set in $\mathscr{R}$ makes:

$$P_{rob}(\text{event}) = P_r = 1$$

and

$$P_m = P_{rob}(\text{imaginary part}) = 0.$$

Here we have $|Z|^2 = 1 - (2 \times 1) \times (1 - 1) = 1$ because the phenomenon is totally known, that is, its laws and variables are completely determined, hence; our degree of our knowledge of the system is $1 = 100\%$.

Now, if we can tell for sure that an event will never occur i.e. like 'getting nothing' (the empty set), P_r is accordingly $= 0$, that is the event will never occur in $\mathscr{R}$. P_m will be equal to:

$$i(1 - P_r) = i(1 - 0) = i,$$

and

$$|Z|^2 = 1 - (2 \times 0) \times (1 - 0) = 1$$

because we can tell that the event of getting nothing surely will never occur; thus, the Degree of Our Knowledge (*DOK*) of the system is $1 = 100\%$. (Abou Jaoude et al., 2010).

We can infer that we have always:

$$0.5 \leq |Z|^2 \leq 1, \forall P_r : 0 \leq P_r \leq 1$$

and

$$|Z|^2 = DOK = P_r^2 + (P_m/i)^2,$$

where

$$0 \leq P_r, \frac{P_m}{i} \leq 1 \tag{1}$$

And what is important is that in all cases we have:

$$Pc^2 = \left(P_r + \frac{P_m}{i}\right)^2 = |Z|^2 - 2iP_rP_m = [P_r + (1 - P_r)]^2$$
$$= 1^2 = 1 \tag{2}$$

In fact, according to an experimenter in $\mathscr{R}$, the game is a game of chance: the experimenter doesn't know the output of the event. He will assign to each outcome a probability P_r and he will say that the output is nondeterministic. But in the universe $\mathscr{C} = \mathscr{R} + \mathscr{M}$, an observer will be able to predict the outcome of the game of chance since he takes into consideration the contribution of $\mathscr{M}$, so we write:

$$Pc^2 = \left(P_r + \frac{P_m}{i}\right)^2$$

Hence *Pc* is always equal to 1. In fact, the addition of the imaginary set to our random experiment resulted to the abolition of ignorance and indeterminism. Consequently, the study of this class of phenomena in $\mathscr{C}$ is of great usefulness since we will be able to predict with certainty the outcome of experiments conducted. In fact, the study in $\mathscr{R}$ leads to unpredictability and uncertainty. So instead of placing ourselves in $\mathscr{R}$, we place ourselves in $\mathscr{C}$ then study the phenomena, because in $\mathscr{C}$ the contributions of $\mathscr{M}$ are taken into consideration and therefore a deterministic study of the phenomena becomes possible. Conversely, by taking into consideration the contribution of the set $\mathscr{M}$ we place ourselves in $\mathscr{C}$ and by ignoring $\mathscr{M}$ we restrict our study to nondeterministic phenomena in $\mathscr{R}$ (Bell, 1992; Boursin, 1986; Dacunha-Castelle, 1996; Dalmedico-Dahan & Peiffer, 1986; Dalmedico-Dahan, Chabert, & Chemla, 1992; Ekeland, 1991; Franklin, 2001; Freund, 1973; Gleick, 1997; Gullberg, 1997; Science Et Vie, 1999; Srinivasan & Mehata, 1988; Stewart, 2002; Van Kampen, 2006; Wikipedia, the free encyclopedia, *Probability*; Kuhn, 1970; Warusfel & Ducrocq, 2004; Wikipedia, the free encyclopedia, *Probability theory*; Wikipedia, the free encyclopedia, *Probability distribution*; Abrams, 2008; Barrow, 1992; Daston, 1988; David, 1962; Gorrochum, 2012; Greene, 2003; Hacking, 2006; Jeffrey, 1992; Poincaré, 1968; Stewart, 1996; Stewart, 2012; Von Plato, 1994).

Moreover, it follows from the above definitions and axioms that (Abou Jaoude et al., 2010):

$$2iP_rP_m = 2i \times P_r \times i \times (1 - P_r)$$
$$= 2i^2 \times P_r \times (1 - P_r) = -2P_r(1 - P_r) \tag{3}$$
$$= Chf$$

$2iP_rP_m$ will be called the Chaotic factor in our experiment and will be denoted accordingly by '*Chf*'. We will see why we have called this term the chaotic factor; in fact:

In case $P_r = 1$, that is the case of a certain event, then the chaotic factor of the event is equal to 0.

In case $P_r = 0$, that is the case of an impossible event, then $Chf = 0$. Hence, in both two last cases, there is no chaos since the outcome is certain and is known in advance.

In case $P_r = 0.5$, $Chf = -0.5$.

So we notice that: $-0.5 \leq Chf \leq 0, \forall P_r: 0 \leq P_r \leq 1$. (Figures 2–4).

What is interesting here is thus we have quantified both the degree of our knowledge and the chaotic factor of any random event and hence we write now:

$$Pc^2 = |Z|^2 - 2iP_rP_m = DOK - Chf \tag{4}$$

Then we can conclude that:

Pc^2 = Degree of our knowledge of the system – Chaotic factor = 1, therefore $Pc = 1$ permanently.

This directly means that if we succeed to subtract and eliminate the chaotic factor in any random experiment, then the output will always be with a probability equal to 1 (Bogdanov & Bogdanov, 2012; Bogdanov & Bogdanov, 2013; Hawking, 2005; Penrose, 1999; Bogdanov & Bogdanov, 2010); Bogdanov & Bogdanov, 2009; Seneta, 2016; Davies, 1993; Hawking, 2011; Hawking, 2002; Aczel, 2000; Pickover, 2008; Reeves, 1988; Ronan, 1988; Boltzmann,

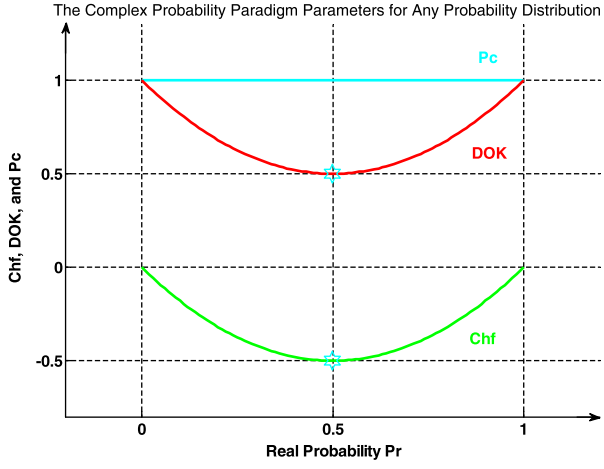


Figure 2. *Chf*, *DOK*, and *Pc* for any probability distribution in 2*D*.

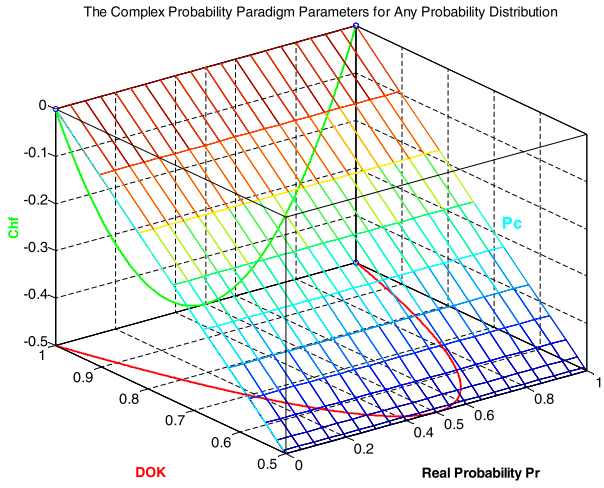


Figure 3. *DOK*, *Chf*, and *Pc* for any probability distribution in 3*D* with $Pc^2 = DOK - Chf = 1 = Pc$.

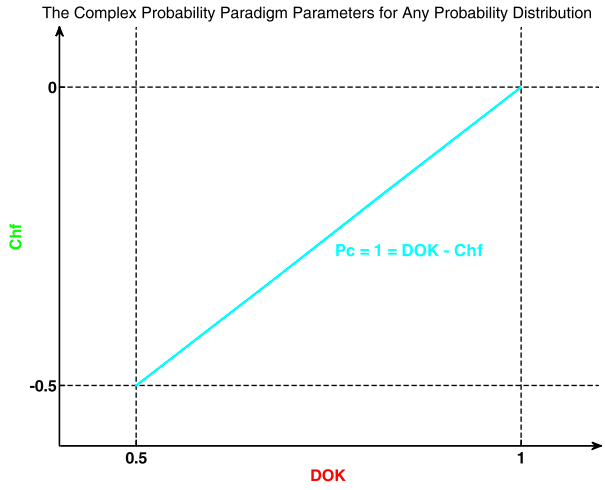


Figure 5. Graph of $Pc^2 = DOK - Chf = 1 = Pc$ for any probability distribution.

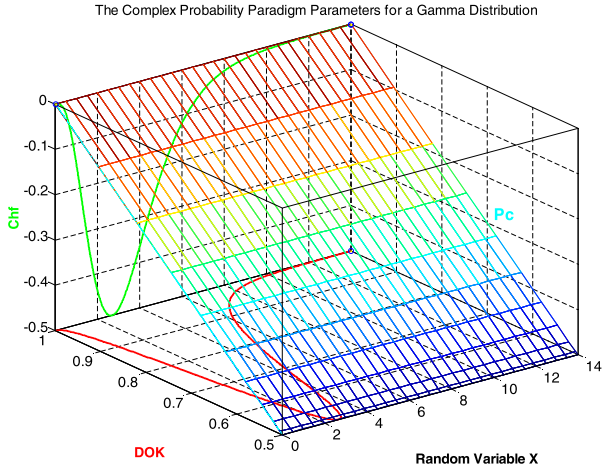


Figure 4. *DOK*, *Chf*, and *Pc* for a gamma probability distribution in 3*D* with $Pc^2 = DOK - Chf = 1 = Pc$.

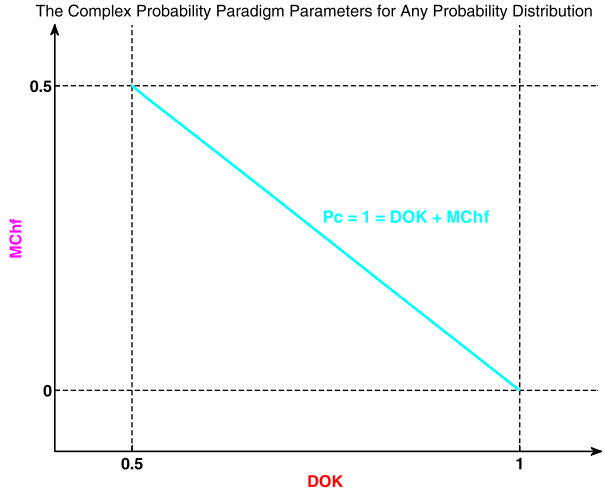


Figure 6. Graph of $Pc^2 = DOK + MChf = 1 = Pc$ for any probability distribution.

1995; Cercignani, 2010; Planck, 1969; Salsburg, 2001; Hald, 2003; Hald, 1998; Heyde & Seneta, 2001; Stigler, 1990; Mcgrayne, 1990; Bernstein, 1996; Ivancevic & Ivancevic, 2008; Balibar, 2002; Burgi, 2010; Hoffmann, 1975; http://www.statslab.cam.ac.uk/~rrw1/markov/M.pdf; Vitanyi, 1988; Moore, 1992).

The graph below shows the linear relation between both *DOK* and *Chf* (Figure 5).

Furthermore, we need in our current study the absolute value of the chaotic factor that will give us the magnitude of the chaotic and random effects on the studied message materialised by the real flips chances P_r and a probability density function, and which lead to an increasing system chaos in $\mathscr{R}$. This new term will be denoted accordingly *MChf* or Magnitude of the Chaotic factor (Abou Jaoude, 2015a, 2015b; Abou Jaoude, 2016a, 2016b; Abou Jaoude, 2017). Hence, we can deduce the following:

$$\begin{aligned} MChf &= |Chf| = |2iP_rP_m| = -2iP_rP_m \\ &= 2P_r(1-P_r) \geq 0, \forall P_r: 0 \leq P_r \leq 1, \end{aligned} \tag{5}$$

and

$$\begin{aligned} Pc^2 &= DOK - Chf \\ &= DOK + |Chf|, \quad \text{since} -0.5 \leq Chf \leq 0 \\ &= DOK + MChf = 1, \end{aligned} \tag{6}$$

$\Leftrightarrow 0 \leq MChf \leq 0.5$ where $0.5 \leq DOK \leq 1$.

The graph below (Figure 6) shows the linear relation between both *DOK* and *MChf*. Moreover, Figures 7–13 show the graphs of *Chf*, *MChf*, *DOK*, and *Pc* as functions of the real probability P_r for any probability distribution and for a gamma probability distribution.

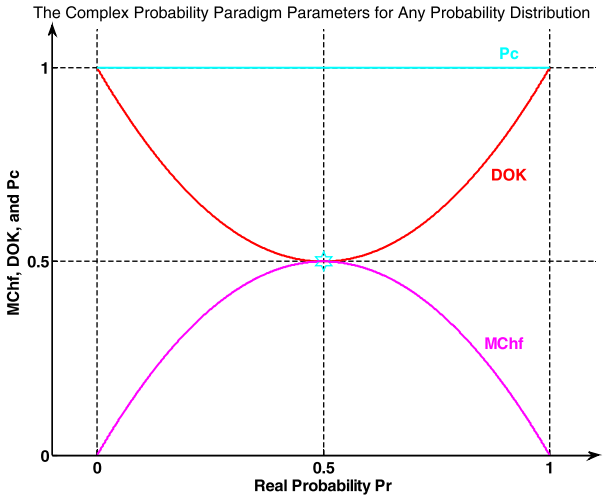


Figure 7. *MChf*, *DOK*, and *Pc* for any probability distribution in 2*D*.

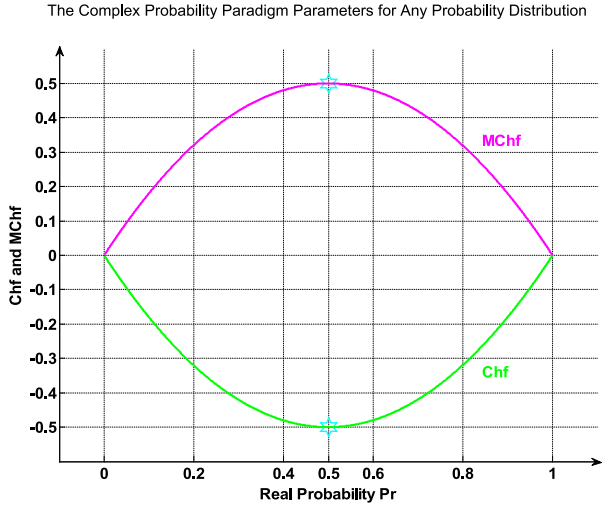


Figure 10. *Chf* and *MChf* for any probability distribution in 2*D*.

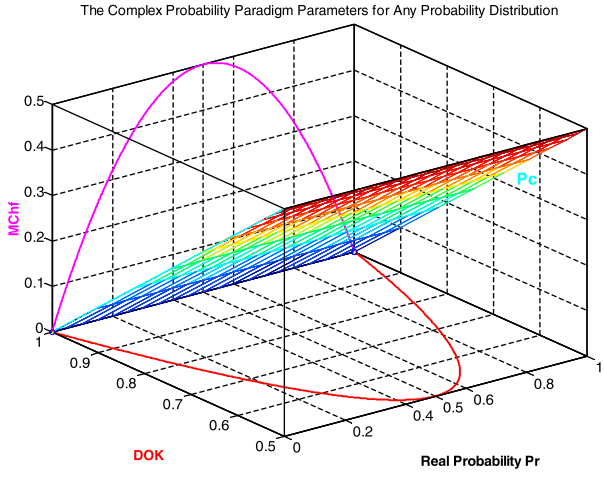


Figure 8. *DOK*, *MChf*, and *Pc* for any probability distribution in 3*D* with $Pc^2 = DOK + MChf = 1 = Pc$.

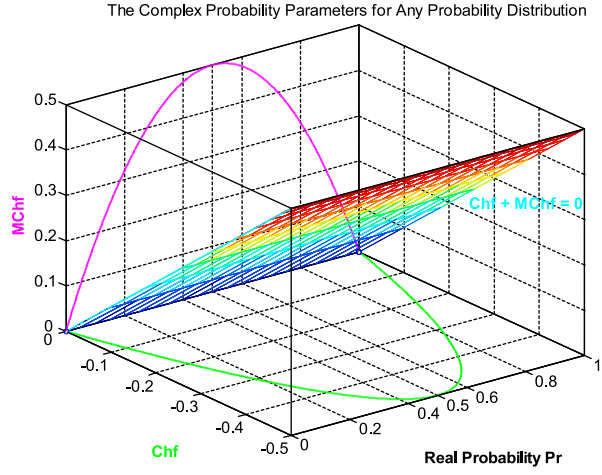


Figure 11. *Chf* and *MChf* for any probability distribution in 3*D* with $Chf + MChf = 0$.

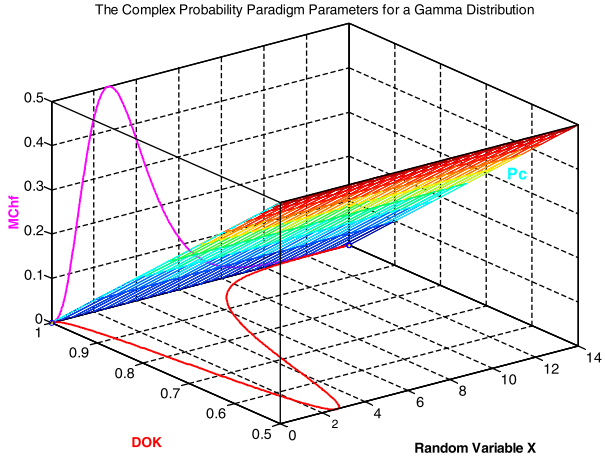


Figure 9. *DOK*, *MChf*, and *Pc* for a gamma probability distribution in 3*D* with $Pc^2 = DOK + MChf = 1 = Pc$.

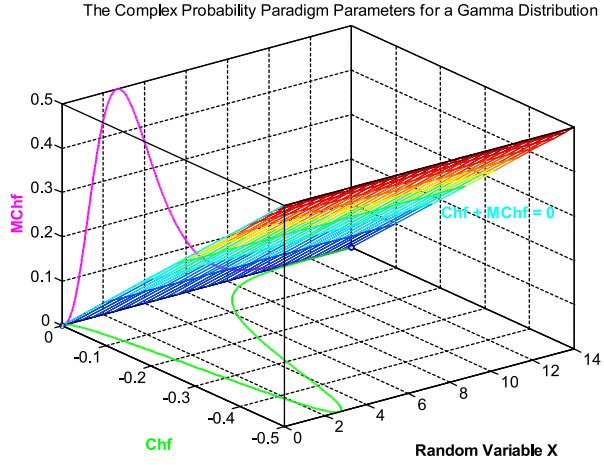


Figure 12. *Chf* and *MChf* for a gamma probability distribution in 3*D* with $Chf + MChf = 0$.

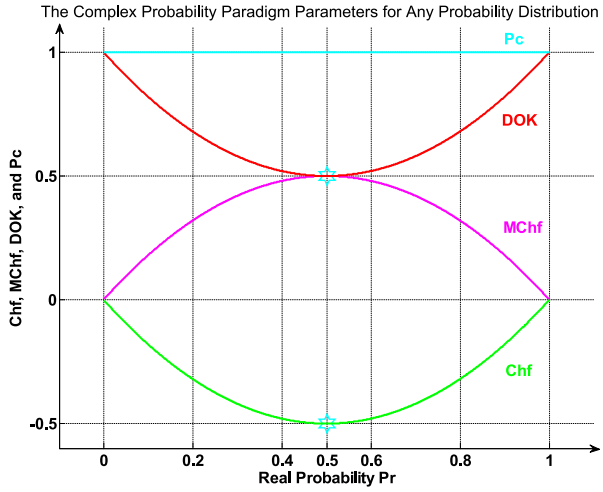


Figure 13. *Chf*, *MChf*, *DOK*, and *Pc* for any probability distribution in 2*D*.

To summarise and to conclude, as the degree of our certain knowledge in the real universe $\mathscr{R}$ is unfortunately incomplete, the extension to the complex set $\mathscr{C}$ includes the contributions of both the real set of probabilities $\mathscr{R}$ and the imaginary set of probabilities $\mathscr{M}$. Consequently, this will result in a complete and perfect degree of knowledge in $\mathscr{C} = \mathscr{R} + \mathscr{M}$ (since $Pc = 1$). In fact, in order to have a certain prediction of any random event, it is necessary to work in the complex set $\mathscr{C}$ in which the chaotic factor is quantified and subtracted from the computed degree of knowledge to lead to a probability in $\mathscr{C}$ equal to one ($Pc^2 = DOK - Chf = DOK + MChf = 1$). This hypothesis is verified in my nine previous research papers by the mean of many examples encompassing both discrete and continuous distributions (Abou Jaoude, 2013a, 2013b; Abou Jaoude, 2014; Abou Jaoude, 2015a, 2015b; Abou Jaoude et al., 2010; Abou Jaoude, 2016a, 2016b; Abou Jaoude, 2017). The Extended Kolmogorov Axioms (*EKA* for short) or the Complex Probability Paradigm (*CPP* for short) can be illustrated by the following figure (Figure 14).

4. A brief review of Claude Shannon's information theory

4.1. Quantities of information

Firstly, information theory is based on probability theory and statistics. Information theory often concerns itself with measures of information of the distributions associated with random variables. Important quantities of information are entropy, a measure of information in a single random variable, and mutual information, a measure of information in common between two random variables. The former quantity is a property of the probability distribution of a random variable and gives a limit on the rate at which data generated by independent samples with the given distribution can be reliably compressed. The latter is a property of the joint distribution of two random variables, and is the maximum rate of reliable communication across a noisy channel in the limit of long block lengths, when the channel statistics are determined by the joint distribution (Rieke et al., 1997; Huelsenbeck et al., 2001; Allikmets et al., 1998; Burnham & Anderson, 2002; Jaynes, 1957; Bennett et al., 2003; David & Anderson, 2003).

The choice of logarithmic base in the following formulae determines the unit of information entropy that is used. A common unit of information is the 'bit', based on the binary logarithm. Other units include the 'nat', which is based on the natural logarithm, and the 'hartley', which is based on the common logarithm (Fazlollah, 1994 [1961]; Ash, 1990 [1965]; Gibson, 1998; Shannon, 1948; Hartley, 1928).

In what follows, an expression of the form $p \log p$ is considered by convention to be equal to zero whenever

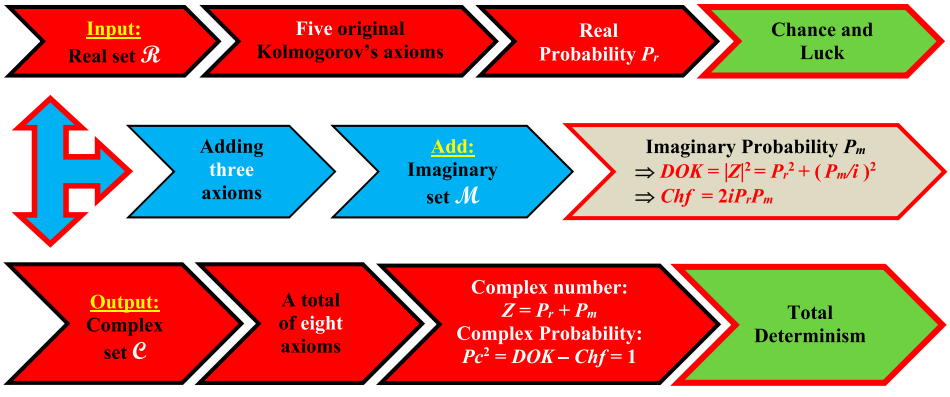


Figure 14. The *EKA* or the *CPP* diagram.

$p = 0$. This is justified because $\lim_{p \to 0+} p \log p = 0^-$ for any logarithmic base by L'Hôpital's rule.

4.1.1. Self-Information

Shannon derived a measure of information content called the self-information or 'suprisal' of a message x:

$$I(x) = \log\left[\frac{1}{p(x)}\right] = -\log[p(x)] \tag{7}$$

where $p(x) = P_{rob}(X = x)$ is the probability that message x is chosen from all possible choices in the message space X. The base of the logarithm only affects a scaling factor and, consequently, the units in which the measured information content is expressed. If the logarithm is base 2, the measure of information is expressed in units of bits (Kelly Jr, 1956; Kolmogorov, 1968; Landauer, 1961; Landauer, 1993; Timme et al., 2012).

Information is transferred from a source to a recipient only if the recipient of the information did not already have the information to begin with. Messages that convey information that is certain to happen and already known by the recipient contain no real information. Infrequently occurring messages contain more information than more frequently occurring messages. This fact is reflected in the above equation – a certain message, i.e. of probability 1, has an information measure of zero. In addition, a compound message of two (or more) unrelated (or mutually independent) messages would have a quantity of information that is the sum of the measures of information of each message individually. That fact is also reflected in the above equation, supporting the validity of its derivation (Arndt, 2004; Ash, 1990; Cover & Thomas, 2006; Gallager, 1968; Goldman, 1968).

An example: The weather forecast broadcast is: 'Tonight's forecast: Dark. Continued darkness until widely scattered light in the morning.' This message contains almost no information. However, a forecast of a snowstorm would certainly contain information since such does not happen every evening. There would be an even greater amount of information in an accurate forecast of snow for a warm location, such as Miami. The amount of information in a forecast of snow for a location where it never snows (impossible event) is the highest (infinity) (Csiszar & Korner, 1997; MacKay, 2003; Mansuripur, 1987).

This measure has also been called surprisal, as it represents the 'surprise' of seeing the outcome (a highly improbable outcome is very surprising). This term was coined by Myron Tribus in his 1961 book *Thermostatics and Thermodynamics* (McEliece, 2002; Pierce, 1961; Reza, 1961).

The information entropy of a random event is the expected value of its self-information.

4.1.1.1. Examples.

- On tossing a coin, the chance of 'tail' is 0.5. When it is proclaimed that indeed 'tail' occurred, this amounts to I('tail') $= \log_2 (1/0.5) = \log_2 2 = 1$ bit of information.
- When throwing a fair die, the probability of 'four' is 1/6. When it is proclaimed that 'four' has been thrown, the amount of self-information is I('four') $= \log_2 (1/(1/6)) = \log_2 (6) = 2.585$ bits.
- When, independently, two dice are thrown, the amount of information associated with {throw 1 = 'two' & throw 2 = 'four'} equals I('throw 1 is two & throw 2 is four') $= \log_2 (1/\text{P(throw 1 = 'two' \& throw 2 = 'four')}) = \log_2 (1/(1/36)) = \log_2 (36) = 5.170$ bits. This outcome equals the sum of the individual amounts of self-information associated with {throw 1 = 'two'} and {throw 2 = 'four'}; namely $2.585 + 2.585 = 5.170$ bits.
- In the same two dice situation we can also consider the information present in the statement 'The sum of the two dice is five' I('The sum of throws 1 and 2 is five') $= \log_2 (1/\text{P('throw 1 and 2 sum to five')}) = \log_2 (1/(4/36)) = 3.17$ bits. The (4/36) is because there are four ways out of 36 possible to sum two dice to 5. This shows how more complex or ambiguous events can still carry information.

4.1.2. Entropy of an information source (Brillouin, 1962 [2004]; Gleick, 2011; Khinchin, 1957; Leff & Rex, 1990; Logan, 2014; Shannon & Weaver, 1949; Siegfried, 2000; Stone, 2014; Yeung, 2002; Yeung, 2008)

Based on the probability mass function of each source symbol to be communicated, the Shannon entropy H, in units of bits (per symbol), is given by:

$$H = -\sum_i p_i \log_2 (p_i) \tag{8}$$

where p_i is the probability of occurrence of the i-th possible value of the source symbol. This equation gives the entropy in the units of 'bits' (per symbol) because it uses a logarithm of base 2, and this base-2 measure of entropy has sometimes been called the 'shannon' in his honour. Entropy is also commonly computed using the natural logarithm (base e, where e is Leonhard Euler's number), which produces a measurement of entropy in 'nats' per symbol and sometimes simplifies the analysis by avoiding the need to include extra constants in the formulas. Other bases are also possible, but less commonly used. For example, a logarithm of base $2^8 = 256$ will produce a measurement in bytes per symbol, and a logarithm of base 10 will produce a measurement in decimal digits (or

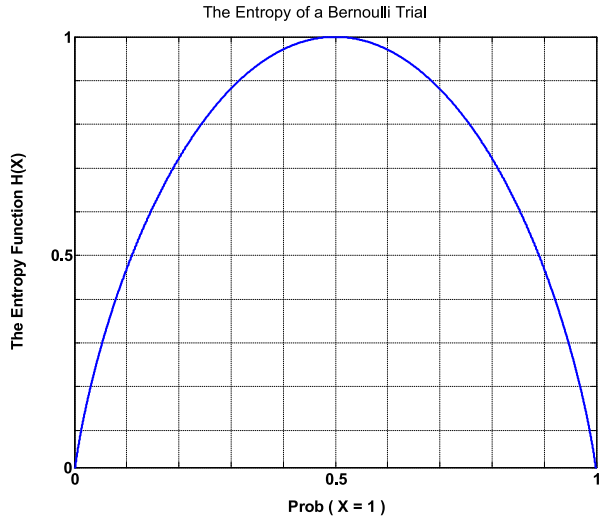


Figure 15. The entropy for a Bernoulli probability distribution.

hartleys) per symbol. Intuitively, the entropy H_X of a discrete random variable X is a measure of the amount of uncertainty associated with the value of X when only its distribution is known.

The entropy of a source that emits a sequence of N symbols that are independent and identically distributed (iid) is $N \times H$ bits (per message of N symbols). If the source data symbols are identically distributed but not independent, the entropy of a message of length N will be less than $N \times H$.

The entropy of a Bernoulli trial as a function of success probability, often called the binary entropy function $H_b(p)$. The entropy is maximised at 1 bit per trial when the two possible outcomes are equally probable, as in an unbiased coin toss (Figure 15).

Suppose one transmits 1000 bits (0s and 1s). If the value of each of these bits is known to the receiver (has a specific value with certainty) ahead of transmission, it is clear that no information is transmitted. If, however, each bit is independently equally likely to be 0 or 1, 1000 shannons of information (more often called bits) have been transmitted. Between these two extremes, information can be quantified as follows. If Ω is the set of all messages $\{x_1, x_2 \ldots, x_n\}$ that X could be, and $p(x)$ is the probability of some $x \in \Omega$, then the entropy, H, of X is defined:

$$H(X) = E_X[I(x)] = -\sum_{x \in \Omega} p(x) \log p(x) \tag{9}$$

(Here, $I(x)$ is the self-information, which is the entropy contribution of an individual message, and E_X is the expected value.) A property of entropy is that it is maximised when all the messages in the message space are equiprobable $p(x) = 1/n$; i.e. most unpredictable, in which case $H(X) = \log n$.

The special case of information entropy for a random variable with two outcomes is the binary entropy function, usually taken to the logarithmic base 2, thus having the shannon (Sh) as unit:

$$H_b(p) = -p\log_2 p - (1-p)\log_2(1-p) \tag{10}$$

4.1.3. Joint entropy

The joint entropy of two discrete random variables X and Y is merely the entropy of their pairing: (X, Y). This implies that if X and Y are independent, then their joint entropy is the sum of their individual entropies. For example, if (X, Y) represents the position of a chess piece — X the row and Y the column, then the joint entropy of the row of the piece and the column of the piece will be the entropy of the position of the piece.

$$H(X, Y) = E_{X,Y}[-\log p(x, y)] = -\sum_{x,y} p(x, y) \log p(x, y) \tag{11}$$

Despite similar notation, joint entropy should not be confused with cross entropy.

4.1.4. Conditional entropy (equivocation)

The conditional entropy or conditional uncertainty of X given random variable Y (also called the equivocation of X about Y) is the average conditional entropy over Y:

$$H(X|Y) = E_Y[H(X|y)] = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y)$$

$$= -\sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(y)} \tag{12}$$

Because entropy can be conditioned on a random variable or on that random variable being a certain value, care should be taken not to confuse these two definitions of conditional entropy, the former of which is in more common use. A basic property of this form of conditional entropy is that:

$$H(X|Y) = H(X, Y) - H(Y) \tag{13}$$

4.1.5. Mutual information (transinformation)

Mutual information measures the amount of information that can be obtained about one random variable by observing another. It is important in communication where it can be used to maximise the amount of information shared between sent and received signals. The mutual information of X relative to Y is given by:

$$I(X; Y) = E_{X,Y}[SI(x, y)] = -\sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \tag{14}$$

where SI (*S*pecific mutual *I*nformation) is the pointwise mutual information.

A basic property of the mutual information is that

$$I(X;Y) = H(X) - H(X|Y) \tag{15}$$

That is, knowing Y, we can save an average of $I(X;Y)$ bits in encoding X compared to not knowing Y. Mutual information is symmetric:

$$I(X;Y) = I(Y;X) = H(X) + H(Y) - H(X,Y) \tag{16}$$

4.1.6. Kullback-Leibler divergence (information gain)

The Kullback–Leibler divergence (or information divergence, information gain, or relative entropy) is a way of comparing two distributions: a 'true' probability distribution $p(X)$, and an arbitrary probability distribution $q(X)$. If we compress data in a manner that assumes $q(X)$ is the distribution underlying some data, when, in reality, $p(X)$ is the correct distribution, the Kullback–Leibler divergence is the number of average additional bits per datum necessary for compression. It is thus defined

$$\begin{aligned} D_{KL}(p(X)||q(X)) &= \sum_{x\in X} -p(x)\log q(x) - \sum_{x\in X} -p(x)\log p(x) \\ &= \sum_{x\in X} p(x)\log\frac{p(x)}{q(x)} \end{aligned} \tag{17}$$

Although it is sometimes used as a 'distance metric', KL divergence is not a true metric since it is not symmetric and does not satisfy the triangle inequality (making it a semi-quasimetric).

4.1.7. Differential entropy

Differential entropy (also referred to as continuous entropy) is a concept in information theory that began as an attempt by Shannon to extend the idea of (Shannon) entropy, a measure of average surprisal of a random variable, to continuous probability distributions. Unfortunately, Shannon did not derive this formula, and rather just assumed it was the correct continuous analogue of discrete entropy, but it is not. The actual continuous version of discrete entropy is the limiting density of discrete points (LDDP).

Let X be a random variable with a probability density function f whose support is a set Ω. The differential entropy $h(X)$ or $h(f)$ is defined as

$$h(X) = -\int_{\Omega} f(x)\log f(x)\mathrm{d}x \tag{18}$$

As with its discrete analog, the units of differential entropy depend on the base of the logarithm, which is usually 2 (i.e. the units are bits). Related concepts such as joint, conditional differential entropy, and relative entropy are defined in a similar fashion.

4.1.8. Other quantities

Other important information theoretic quantities include Rényi entropy (a generalisation of entropy) and the conditional mutual information.

4.2. Coding theory (Campbell, 1982; Escolano et al., 2009; Haggerty, 1981; Noth, 1981; Seife, 2006; Theil, 1967; Wikipedia, the free encyclopedia, Information theory)

Coding theory is one of the most important and direct applications of information theory. It can be subdivided into source coding theory and channel coding theory. Using a statistical description for data, information theory quantifies the number of bits needed to describe the data, which is the information entropy of the source.

- Data compression (source coding): There are two formulations for the compression problem:
  lossless data compression: the data must be reconstructed exactly;
  lossy data compression: allocates bits needed to reconstruct the data, within a specified fidelity level measured by a distortion function. This subset of Information theory is called rate–distortion theory.
- Error-correcting codes (channel coding): While data compression removes as much redundancy as possible, an error correcting code adds just the right kind of redundancy (i.e. error correction) needed to transmit the data efficiently and faithfully across a noisy channel.

This division of coding theory into compression and transmission is justified by the information transmission theorems, or source–channel separation theorems that justify the use of bits as the universal currency for information in many contexts. However, these theorems only hold in the situation where one transmitting user wishes to communicate to one receiving user. In scenarios with more than one transmitter (the multiple-access channel), more than one receiver (the broadcast channel) or intermediary 'helpers' (the relay channel), or more general networks, compression followed by transmission may no longer be optimal. Network information theory refers to these multi-agent communication models.

4.2.1. Source theory

Any process that generates successive messages can be considered a source of information. A memoryless source is one in which each message is an independent identically distributed random variable, whereas the properties of ergodicity and stationarity impose less restrictive constraints. All such sources are stochastic. These terms

are well studied in their own right outside information theory.

4.2.1.1. Rate.

Information rate is the average entropy per symbol. For memoryless sources, this is merely the entropy of each symbol, while, in the case of a stationary stochastic process, it is

$$r = \lim_{n \to +\infty} H(X_n | X_{n-1}, X_{n-2}, X_{n-3}, \ldots) \tag{19}$$

that is, the conditional entropy of a symbol given all the previous symbols generated. For the more general case of a process that is not necessarily stationary, the *average rate* is

$$r = \lim_{n \to +\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \tag{20}$$

that is, the limit of the joint entropy per symbol. For stationary sources, these two expressions give the same result.

It is common in information theory to speak of the 'rate' or 'entropy' of a language. This is appropriate, for example, when the source of information is English prose. The rate of a source of information is related to its redundancy and how well it can be compressed, the subject of source coding.

4.2.2. Channel capacity

Communications over a channel—such as an ethernet cable—is the primary motivation of information theory. As anyone who's ever used a telephone (mobile or landline) knows, however, such channels often fail to produce exact reconstruction of a signal; noise, periods of silence, and other forms of signal corruption often degrade quality. How much information can one hope to communicate over a noisy (or otherwise imperfect) channel?

Consider the communications process over a discrete channel. A simple model of the process is shown below (Figure 16):

Here X represents the space of messages transmitted, and Y the space of messages received during a unit time over our channel. Let $p(y|x)$ be the conditional probability distribution function of Y given X. We will consider $p(y|x)$ to be an inherent fixed property of our communications channel (representing the nature of the noise of our channel). Then the joint distribution of X and Y is completely determined by our channel and by our choice of $f(x)$, the marginal distribution of messages we choose to send over the channel. Under these constraints, we would like to maximise the rate of information, or the signal, we can communicate over the channel. The appropriate measure for this is the mutual information, and this maximum mutual information is called the channel capacity and is given by:

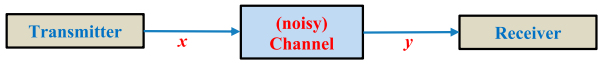


Figure 16. A simple model of the communications process over a discrete channel.

$$C = \max_f I(X;Y) \tag{21}$$

This capacity has the following property related to communicating at information rate R (where R is usually bits per symbol). For any information rate $R < C$ and coding error $\varepsilon > 0$, for large enough N, there exists a code of length N and rate $\geq R$ and a decoding algorithm, such that the maximal probability of block error is $\leq \varepsilon$; that is, it is always possible to transmit with arbitrarily small block error. In addition, for any rate $R > C$, it is impossible to transmit with arbitrarily small block error.

Channel coding is concerned with finding such nearly optimal codes that can be used to transmit data over a noisy channel with a small coding error at a rate near the channel capacity.

4.2.2.1. Capacity of particular channel models.

4.2.2.1.1. Continuous-time analog communications channel subject to Gaussian noise.

The Shannon–Hartley theorem states the channel capacity C, meaning the theoretical tightest upper bound on the information rate of data that can be communicated at an arbitrarily low error rate using an average received signal power S through an analog communication channel subject to additive white Gaussian noise of power N:

$$C = B \log_2 \left(1 + \frac{S}{N}\right) \tag{22}$$

where

- C is the channel capacity in bits per second, a theoretical upper bound on the net bit rate (information rate, sometimes denoted I) excluding error-correction codes;
- B is the bandwidth of the channel in hertz (passband bandwidth in case of a bandpass signal);
- S is the average received signal power over the bandwidth (in case of a carrier-modulated passband transmission, often denoted C), measured in watts (or volts squared);
- N is the average power of the noise and interference over the bandwidth, measured in watts (or volts squared); and
- S/N is the signal-to-noise ratio (SNR) or the carrier-to-noise ratio (CNR) of the communication signal to the noise and interference at the receiver (expressed as a linear power ratio, not as logarithmic decibels).

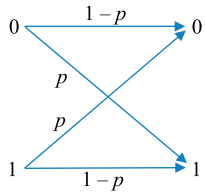


Figure 17. A Binary Symmetric Channel.

4.2.2.1.2. A Binary Symmetric Channel. A binary symmetric channel (BSC) with crossover probability p is a binary input, binary output channel that flips the input bit with probability p. The BSC has a capacity of $1 - H_b(p)$ bits per channel use, where H_b is the binary entropy function to the base 2 logarithm (Figure 17).

5. The surprisal and expectancy self-information functions (Jaynes, 1957; Bennett et al., 2003; David & Anderson, 2003; Fazlollah, 1994 [1961]; Ash, 1990 [1965]; Gibson, 1998; Shannon, 1948; Hartley, 1928)

5.1. Definitions

Shannon derived a measure of information content called the self-information or 'suprisal' of a message x:

$$I(x) = \log\left[\frac{1}{p(x)}\right] = -\log[p(x)]$$

where $p(x) = P_{rob}(X = x)$ is the probability that message x is chosen from all possible choices in the message space X. The base of the logarithm only affects a scaling factor and, consequently, the units in which the measured information content is expressed. If the logarithm is base 2, the measure of information is expressed in units of bits. Therefore, in base 2 the self-information or 'suprisal' of a message x is:

$$I_2(x) = \log_2\left[\frac{1}{p(x)}\right] = -\log_2[p(x)]$$

We define now a new function that we can call the 'expectancy' of the same message x which is in base 2:

$$\bar{I}_2(x) = \log_2\left[\frac{1}{1 - p(x)}\right] = -\log_2[1 - p(x)] \qquad (23)$$

In fact if the probability of a message x is $p(x)$ then the complementary probability of the complement event is $1 - p(x)$. Hence $I_2(x)$ corresponds to $p(x)$ and $\bar{I}_2(x)$ corresponds to $1 - p(x)$. $I_2(x)$ measures the surprisal of a message x that means

$$I_2(\text{impossible event}) = I_2[p(x) = 0] = -\log_2(0) \to -(-\infty) = \infty,$$

and

$$I_2(\text{certain event}) = I_2[p(x) = 1] = -\log_2(1) = 0$$

Therefore

$$\bar{I}_2(\text{impossible event}) = I_2[1 - p(x) = 1 - 0 = 1] = -\log_2(1) = 0,$$

and

$$\bar{I}_2(\text{certain event}) = I_2[1 - p(x) = 1 - 1 = 0] = -\log_2(0) \to -(-\infty) = \infty$$

Consequently, $\bar{I}_2(x)$ is the opposite of $I_2(x)$ and will measure the self-information acquired from the message expectancy or likeliness to occur.

Additionally, we have

$$I_2(x) + \bar{I}_2(x) = -\log_2[p(x)] - \log_2[1 - p(x)] = -\log_2\{p(x)[1 - p(x)]\}$$

And the binary entropy is

$$H_b[p(x)] = -p(x)\log_2[p(x)] - [1 - p(x)]\log_2[1 - p(x)] = p(x)I_2(x) + [1 - p(x)]\bar{I}_2(x)$$

5.2. Discrete and continuous message domains

In the discrete case, let $\Omega = \{\text{the set of messages } X\} = \{x_1, x_2, \ldots, x_n\}$. It can be ordered by a one-to-one correspondence between Ω and the discrete countable ordered interval $\Omega_Z = [L_b, U_b]$ having L_b as the lower bound and U_b as the upper bound and where Ω_Z is a subset of $\mathbf{Z}$ (the set of integers). Thus, $x_1 = L_b$, $x_2 = a + 1$, $x_3 = L_b + 2, \ldots, x_{n-1} = a + n - 2$, $x_n = L_b + n - 1 = U_b$. Knowing that Ω and Ω_Z have the same cardinal or number of elements, thus $|\Omega| = |\Omega_Z|$. So we can write without any confusion: $X \leq x$, that means all the messages in Ω_Z less or equal to x. Hence $X \leq x_3$ means $X = \{x_1, x_2, x_3\}$. We can also assign to each X in Ω_Z a probability measure $p_1 = P_{rob}(X = x_1)$, $p_2 = P_{rob}(X = x_2), \ldots, p_n = P_{rob}(X = x_n)$ such that $\sum_{k=1}^{n} p_k = 1$. Moreover, we can write

$$P_{rob}(X \leq x_j) = P_{rob}(X = x_1) + P_{rob}(X = x_2) + \ldots + P_{rob}(X = x_j) = \sum_{k=1}^{j} p_k$$

Which denotes the sum of all messages probabilities less than or equal to the message x_j in Ω_Z. As examples of a

discrete random variable X are the outcome of tossing a coin or of throwing a die, or the sum of two thrown dice.

If $F(x)$ is the discrete probability cumulative distribution function (CDF) of the random variable of messages X, then let

$$p(x) = P_{rob}(X \le x) = F(x) = \sum_{X \le x} p_k$$

where $p(x)$ denotes the sum of all messages probabilities less than or equal to the message x in Ω_Z.

The complement probability is:

$$1 - p(x) = P_{rob}(X > x) = 1 - F(x) = \sum_{X > x} p_k$$

and denotes the sum of all messages probabilities greater than the message x in Ω_Z.

The continuous case is an extension of the discrete case where Ω_Z is replaced by the continuous uncountable dense ordered message interval $\Omega_R = [L_b, U_b]$ which is a subset of R (the set of real numbers). Therefore,

$$p(x) = P_{rob}(X \le x) = F(x) = \int_{L_b}^{x} f(t)\mathrm{d}t$$

where $p(x)$ denotes the sum of all messages probabilities less than or equal to the message x in Ω_R. And

$$1 - p(x) = P_{rob}(X > x) = 1 - F(x) = \int_{x}^{U_b} f(t)dt$$

denotes the sum of all messages probabilities greater than the message x in Ω_R.

Hence $F(x)$ is the continuous CDF and $f(x)$ is the probability density function (PDF) of the random variable of messages X that can follow any possible continuous random distribution. As examples of a continuous random variable X are the lifetime of a light bulb, or the height of a building, or the length of a rod, or the body weight.

5.3. Rescaled self-information functions

Moreover, if $p(x) \in [0, 1]$ then $I_2(x)$ and $\bar{I}_2(x)$ belong to the interval $[0, \infty)$. We can rescale both of them in the new simulation domain which is $x \in [L_b, U_b]$ to make them belong to $[0, 1]$. Thus, $p(L_b) = P_{rob}(X \le L_b) = F(L_b) = 0$ and $p(U_b) = P_{rob}(X \le U_b) = F(U_b) = 1$ which is the complement to the probability $p(L_b)$. Let $RI_2(x)$ be the rescaled $I_2(x)$, $R\bar{I}_2(x)$ be the rescaled $\bar{I}_2(x)$ and Φ the simulation rescaling factor. Therefore:

$$RI_2(x) = I_2(x)/\Phi \text{ and } R\bar{I}_2(x) = \bar{I}_2(x)/\Phi$$

$$RI_2(x = L_b) = 1 \text{ and } R\bar{I}_2(x = L_b) = 0$$

$$RI_2(x = U_b) = 0 \text{ and } R\bar{I}_2(x = U_b) = 1.$$

Consequently:

$$\begin{aligned} &RI_2(x = L_b) + R\bar{I}_2(x = L_b) \\ &= I_2(x = L_b)/\Phi + \bar{I}_2(x = L_b)/\Phi \\ &= [I_2(x = L_b) + \bar{I}_2(x = L_b)]/\Phi \\ &= \{-\log_2[p(L_b)] - \log_2[1 - p(L_b)]\}/\Phi \\ &= \{-\log_2[p(L_b)] - \log_2[p(U_b)]\}/\Phi \\ &= RI_2(x = L_b) + RI_2(x = U_b) = 1 + 0 = 1 \end{aligned}$$

And

$$\begin{aligned} &RI_2(x = U_b) + R\bar{I}_2(x = U_b) \\ &= I_2(x = U_b)/\Phi + \bar{I}_2(x = U_b)/\Phi \\ &= [I_2(x = U_b) + \bar{I}_2(x = U_b)]/\Phi \\ &= \{-\log_2[p(U_b)] - \log_2[1 - p(U_b)]\}/\Phi \\ &= \{-\log_2[p(U_b)] - \log_2[p(L_b)]\}/\Phi \\ &= RI_2(x = U_b) + RI_2(x = L_b) = 0 + 1 = 1 \end{aligned}$$

and for $\forall x : L_b < x < U_b$

$$\begin{aligned} RI_2(x) + R\bar{I}_2(x) &= I_2(x)/\Phi + \bar{I}_2(x)/\Phi = [I_2(x) + \bar{I}_2(x)]/\Phi \\ &= \{-\log_2[p(x)] - \log_2[1 - p(x)]\}/\Phi \\ &= -\log_2\{p(x)[1 - p(x)]\}/\Phi \end{aligned}$$

If the message distribution is a symmetric distribution and if $x = (L_b + U_b)/2 = Md =$ the Median $=$ the Mean $=$ the Mode of this distribution then

$$P_{rob}\left(x = Md = \frac{L_b + U_b}{2}\right) = p(Md) = 1 - p(Md) = 0.5$$

and

$$\begin{aligned} &RI_2(x = Md) + R\bar{I}_2(x = Md) \\ &= \{-\log_2[p(Md)] - \log_2[1 - p(Md)]\}/\Phi \\ &= \{-\log_2[p(Md)] - \log_2[p(Md)]\}/\Phi \\ &= -2\log_2[p(Md)]/\Phi \\ &= RI_2(x = Md) + RI_2(x = Md) = 2RI_2(x = Md) \end{aligned}$$

Therefore

$$\begin{aligned} RI_2(x = Md) = R\bar{I}_2(x = Md) &= -\log_2[p(Md)]/\Phi \\ &= -\log_2(0.5)/\Phi = 1/\Phi \text{ bits} \end{aligned}$$

And

$$\begin{aligned} &RI_2(x = Md) + R\bar{I}_2(x = Md) \\ &= -\log_2\{p(x = Md)[1 - p(x = Md)]\}/\Phi \\ &= -\log_2\{0.5[1 - 0.5]\}/\Phi = 2/\Phi \text{ bits} \end{aligned}$$

As a first example, consider a continuous uniform distribution and take $[L_b, U_b] = [-2, 2]$ and $\Phi = 8.7$ then

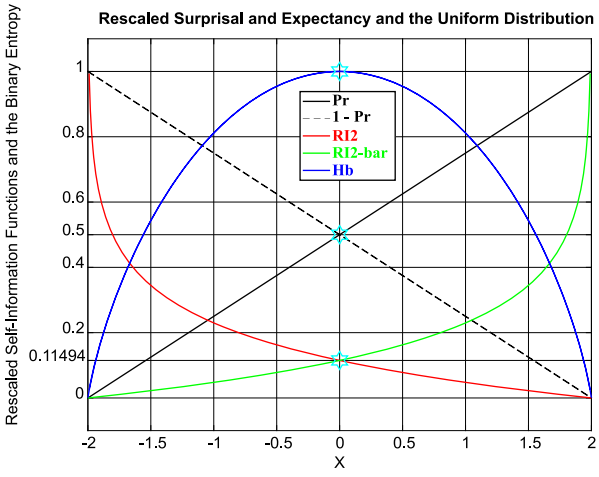


Figure 18. The rescaled surprisal and expectancy functions with the entropy of a uniform probability distribution.

$Md = (-2+2)/2 = 0$ and

$$RI_2(x=-2) + R\bar{I}_2(x=-2) = RI_2(x=-2) + RI_2(x=2)$$
$$= 1 + 0 = 1 \text{ bit}$$
$$RI_2(x=2) + R\bar{I}_2(x=2) = RI_2(x=2) + RI_2(x=-2)$$
$$= 0 + 1 = 1 \text{ bit}$$
$$RI_2(x = Md = 0) = R\bar{I}_2(x = Md = 0) = -\log_2[p(0)]/\Phi$$
$$= -\log_2(0.5)/8.7 = 1/8.7 = 0.11494 \text{ bits}$$

The figure above illustrates all these calculations. (Figure 18).

As a second example, consider the continuous standard Gaussian normal distribution and take $[L_b, U_b] = [-4, 4]$ and $\Phi = 15$ then $Md = (-4+4)/2 = 0$ and

$$RI_2(x=-4) + R\bar{I}_2(x=-4) = RI_2(x=-4) + RI_2(x=4)$$
$$= 1 + 0 = 1 \text{ bit.}$$
$$RI_2(x=4) + R\bar{I}_2(x=4) = RI_2(x=4) + RI_2(x=-4)$$
$$= 0 + 1 = 1 \text{ bit.}$$
$$RI_2(x = Md = 0) = R\bar{I}_2(x = Md = 0)$$
$$= -\log_2[p(0)]/\Phi$$
$$= -\log_2(0.5)/15$$
$$= 1/15 = 0.066667 \text{ bits}$$

The figure above illustrates all these calculations (Figure 19).

If the message distribution is not a symmetric distribution and if $x = Md =$ the Median of this distribution that divides it into two equal probability parts then

$$P_{rob}(x = Md) = p(Md) = 1 - p(Md) = 0.5$$

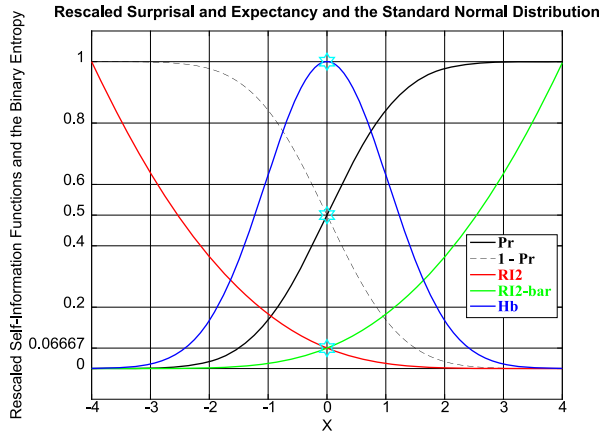


Figure 19. The rescaled surprisal and expectancy functions with the entropy of the standard Gaussian normal distribution.

and

$$RI_2(x = Md) + R\bar{I}_2(x = Md)$$
$$= \{-\log_2[p(Md)] - \log_2[1 - p(Md)]\}/\Phi$$
$$= \{-\log_2[p(Md)] - \log_2[p(Md)]\}/\Phi$$
$$= -2\log_2[p(Md)]/\Phi$$
$$= RI_2(x = Md) + RI_2(x = Md) = 2RI_2(x = Md)$$

Then

$$RI_2(x = Md) = R\bar{I}_2(x = Md)$$
$$= -\log_2[p(Md)]/\Phi = -\log_2(0.5)/\Phi$$
$$= 1/\Phi \text{ bits.}$$

And

$$RI_2(x = Md) + R\bar{I}_2(x = Md)$$
$$= -\log_2\{p(x = Md)[1 - p(x = Md)]\}/\Phi$$
$$= -\log_2\{0.5[1 - 0.5]\}/\Phi = 2/\Phi \text{ bits}$$

As an example, consider a continuous exponential distribution and take $[L_b, U_b] = [0, 7]$ and $\Phi = 10$ then $Md =$ Median $= 0.6931$ and

$$RI_2(x=0) + R\bar{I}_2(x=0) = RI_2(x=0) + RI_2(x=7)$$
$$= 1 + 0 = 1 \text{ bit.}$$
$$RI_2(x=7) + R\bar{I}_2(x=7) = RI_2(x=7) + RI_2(x=0)$$
$$= 0 + 1 = 1 \text{ bit.}$$
$$RI_2(x = Md = 0.6931) = R\bar{I}_2(x = Md = 0.6931)$$
$$= -\log_2[p(0.6931)]/\Phi$$
$$= -\log_2(0.5)/10 = 1/10$$
$$= 0.1 \text{ bits}$$

The figure below illustrates all these calculations (Figure 20).

Additionally, at the point $x = Md =$ the Median of the message distribution and for any probability distribution we have $p(Md) = 1 - p(Md) = 0.5$, therefore the binary entropy is maximum since it is equal to:

$$\begin{aligned} H_b[x = Md] &= -p(Md)\log_2[p(Md)] \\ &\quad - [1 - p(Md)]\log_2[1 - p(Md)] \\ &= -0.5\log_2[0.5] - 0.5\log_2[0.5] = -\log_2[0.5] \\ &= -\frac{Ln(1/2)}{Ln(2)} = \frac{Ln(2)}{Ln(2)} = 1 \end{aligned} \tag{24}$$

At this point we can notice that the message surprisal equals to the message expectancy and $H_b[x = Md]$ equals to 1 shannon (Figures 18–20).

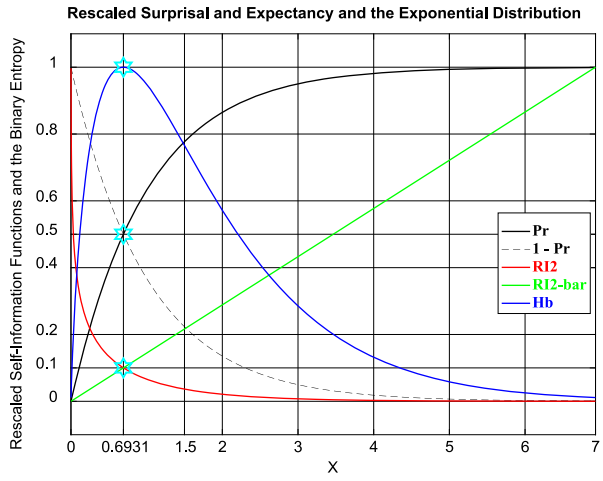


Figure 20. The rescaled surprisal and expectancy functions with the entropy of an exponential probability distribution.

6. The complex probability paradigm and the binary entropies (Abou Jaoude, 2013a, 2013b; Abou Jaoude, 2014; Abou Jaoude, 2015a, 2015b; Abou Jaoude et al., 2010; Abou Jaoude, 2016a, 2016b; Abou Jaoude, 2017; Abou Jaoude, 2004; Abou Jaoude, 2005; Abou Jaoude, 2007; Bidabad, 1992; Chan Man Fong, De Kee, & Kaloni, 1997; Fagin, Halpern, & Megiddo, 1990; Ognjanović, Marković, Rašković, Doder, & Perović, 2012; Stepić & Ognjanović, 2014; Cox, 1955; Wei, Qiu, Karimi, & Wang, 2014; Wei, Qiu, & Fu, 2015; Weingarten, 2002; Youssef, 1994)

6.1. The real binary entropy $H_b^R = H_b$

6.1.1. The real binary entropy H_b^R as a function of all the CPP parameters

In the real probability set $\mathscr{R}$ we have:

$$H_b^R(p) = H_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

And from CPP we have

$$Pc^2 = DOK - Chf = DOK + MChf = 1 = Pc$$

$$Chf = 2iP_rP_m = 2ip \times i(1-p) = -2p(1-p)$$

$$MChf = -Chf$$

$$DOK = |Z|^2 = P_r^2 + (P_m/i)^2 = 1 + Chf = 1 - 2p(1-p)$$

Then

$Chf = -2p + 2p^2 \Rightarrow 2p^2 - 2p - Chf = 0$ which is a second degree equation function of p. So the discriminant is: $\Delta = 4 + 8Chf$. Since $-0.5 \le Chf \le 0$ then $0 \le \Delta \le 4$, therefore the two real roots are:

$$p_1 = \frac{2 - \sqrt{4 + 8Chf}}{4} = \frac{1 - \sqrt{1 + 2Chf}}{2}$$

and

$$p_2 = \frac{2 + \sqrt{4 + 8Chf}}{4} = \frac{1 + \sqrt{1 + 2Chf}}{2}$$

Knowing that:

$$\begin{aligned} p_1 + p_2 &= \frac{1 - \sqrt{1 + 2Chf}}{2} + \frac{1 + \sqrt{1 + 2Chf}}{2} \\ &= 1 = Pc \Rightarrow p_2 = 1 - p_1 \end{aligned}$$

and

$$p_1 = 1 - p_2.$$

Therefore,

$$\begin{aligned} H_b^R(p) &= -p_1\log_2 p_1 - (1-p_1)\log_2(1-p_1) \\ &= -p_2\log_2 p_2 - (1-p_2)\log_2(1-p_2) \\ &= -p_1\log_2 p_1 - p_2\log_2 p_2 \end{aligned}$$

So,

$$\begin{aligned} H_b^R(p) = &-\left(\frac{1 - \sqrt{1 + 2Chf}}{2}\right)\log_2\left(\frac{1 - \sqrt{1 + 2Chf}}{2}\right) \\ &-\left(\frac{1 + \sqrt{1 + 2Chf}}{2}\right)\log_2\left(\frac{1 + \sqrt{1 + 2Chf}}{2}\right) \end{aligned}$$

$$\Rightarrow H_b^R(p) = -\left(\frac{1-\sqrt{1+2Chf}}{2}\right) \times \left[\log_2\left(1-\sqrt{1+2Chf}\right) - \log_2 2\right] - \left(\frac{1+\sqrt{1+2Chf}}{2}\right) \times \left[\log_2\left(1+\sqrt{1+2Chf}\right) - \log_2 2\right]$$

since $\log_2(x/y) = \log_2 x - \log_2 y$. But $\log_2 2 = 1$, Hence,

$$\Rightarrow H_b^R(p) = -\frac{1}{2}\left[\left(1-\sqrt{1+2Chf}\right)\log_2\left(1-\sqrt{1+2Chf}\right) + \left(1+\sqrt{1+2Chf}\right)\log_2\left(1+\sqrt{1+2Chf}\right)\right] + 1$$

But $\log_2(xy) = \log_2 x + \log_2 y$ and $\log_2(x/y) = \log_2 x - \log_2 y$

$$\Rightarrow H_b^R(p) = -\frac{1}{2}\log_2\left[\left(1-\sqrt{1+2Chf}\right) \times \left(1+\sqrt{1+2Chf}\right)\right] - \left(\frac{\sqrt{1+2Chf}}{2}\right) \times \log_2\left(\frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}}\right) + 1$$

But $(a-b)\times(a+b) = a^2 - b^2$.

Then,

$$\left(1-\sqrt{1+2Chf}\right) \times \left(1+\sqrt{1+2Chf}\right) = 1 - (1+2Chf) = -2Chf = 2MChf$$

and

$$\frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}} = \frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}} \times \frac{1+\sqrt{1+2Chf}}{1+\sqrt{1+2Chf}} = \frac{\left(1+\sqrt{1+2Chf}\right)^2}{2MChf}$$

But $(a+b)^2 = a^2 + 2ab + b^2$.

Then

$$\frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}} = \frac{1+1+2Chf+2\sqrt{1+2Chf}}{2MChf} = \frac{2+2Chf+2\sqrt{1+2Chf}}{2MChf} = \frac{1+Chf+\sqrt{1+2Chf}}{MChf}$$

But $Pc^2 = DOK - Chf = 1 \Rightarrow DOK = 1 + Chf$

Then

$$\frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}} = \frac{DOK+\sqrt{1+2Chf}}{MChf}$$

Therefore

$$H_b^R(p) = -\frac{1}{2}\log_2(2MChf) - \left(\frac{\sqrt{1+2Chf}}{2}\right) \times \log_2\left(\frac{DOK+\sqrt{1+2Chf}}{MChf}\right) + 1$$

$$\Rightarrow H_b^R(p) = \frac{1}{2} - \frac{1}{2}\left(1-\sqrt{1+2Chf}\right)\log_2(MChf) - \left(\frac{\sqrt{1+2Chf}}{2}\right)\log_2\left(DOK+\sqrt{1+2Chf}\right)$$

$$\Rightarrow H_b^R(p) = \frac{1}{2}\left[1 - \left(1-\sqrt{1+2Chf}\right)\log_2(MChf) - \left(\sqrt{1+2Chf}\right)\log_2\left(DOK+\sqrt{1+2Chf}\right)\right]$$

But $Pc^2 = 1 = Pc = p_1 + p_2$, therefore we get the final formula of H_b^R as a function of all the CPP parameters:

$$H_b^R(Chf, MChf, DOK, Pc) = \frac{1}{2}\left\{Pc - \left[\left(1-\sqrt{1+2Chf}\right)\log_2(MChf) + \left(\sqrt{1+2Chf}\right)\log_2\left(DOK+\sqrt{1+2Chf}\right)\right]\right\} \quad (25)$$

In fact and to check: if $p = 0$ or $p = 1$ then from the CPP equations above we have:

$Chf = 0 = MChf$ and $DOK = 1$ with $Pc = 1$always. Therefore:

$$\Rightarrow H_b^R(Chf, MChf, DOK, Pc) = H_b(0,0,1,1) = \frac{1}{2}\left\{1 - \left[\left(1-\sqrt{1+2\times 0}\right)\log_2(0) + \left(\sqrt{1+2\times 0}\right)\log_2\left(1+\sqrt{1+2\times 0}\right)\right]\right\}$$

$$= \frac{1}{2}\{1 - [(1-1)\log_2(0) + (1)\log_2(1+1)]\}$$

$$= \frac{1}{2}\{1 - [0\log_2 0 + \log_2 2]\}$$

By L'Hôpital's rule: $\lim_{p\to 0+} p\log_2 p = 0^-$, and $\log_2 2 = 1$ then:

$$\Rightarrow H_b^R(Chf, MChf, DOK, Pc) = H_b(0,0,1,1) = \frac{1}{2}\{1 - [0+1]\} = 0$$

Moreover, if $p = 0.5$ then from the CPP equations above we have:

$Chf = -0.5, MChf = 0.5$ and $DOK = 0.5$ with $Pc = 1$as always. Therefore:

$$\begin{aligned}
&\Rightarrow H_b^R(Chf, MChf, DOK, Pc) \\
&= H_b(-0.5, 0.5, 0.5, 1) \\
&= \frac{1}{2}\left\{1 - \left[\left(1 - \sqrt{1 - 2 \times 0.5}\right)\log_2(0.5)\right.\right. \\
&\quad \left.\left. + \left(\sqrt{1 - 2 \times 0.5}\right)\log_2\left(0.5 + \sqrt{1 - 2 \times 0.5}\right)\right]\right\} \\
&= \frac{1}{2}\{1 - [(1-0)\log_2(0.5) + (0)\log_2(0.5+0)]\} \\
&= \frac{1}{2}\{1 - \log_2 0.5\} = \frac{1}{2}\left\{1 - \log_2\left(\frac{1}{2}\right)\right\}
\end{aligned}$$

Since

$$\log_2\left(\frac{1}{2}\right) = -\log_2 2 = -1.$$

Then:

$$\begin{aligned}
\Rightarrow H_b^R(Chf, MChf, DOK, Pc) &= H_b^R(-0.5, 0.5, 0.5, 1) \\
&= \frac{1}{2}\{1 + \log_2 2\} = \frac{1}{2} \times 2 = 1
\end{aligned}$$

6.1.2. The real binary entropy $H_b^R = H_b$ as a function of chf alone:

As $Pc^2 = DOK - Chf = 1 \Rightarrow DOK = 1 + Chf$ and $MChf = -Chf$ and $Pc = 1$ always, then (Figure 21):

$$H_b^R(Chf) = \frac{1}{2}\left\{1 - \left[\left(1 - \sqrt{1 + 2Chf}\right)\log_2(-Chf) + \left(\sqrt{1 + 2Chf}\right)\log_2\left(1 + Chf + \sqrt{1 + 2Chf}\right)\right]\right\}. \quad (26)$$

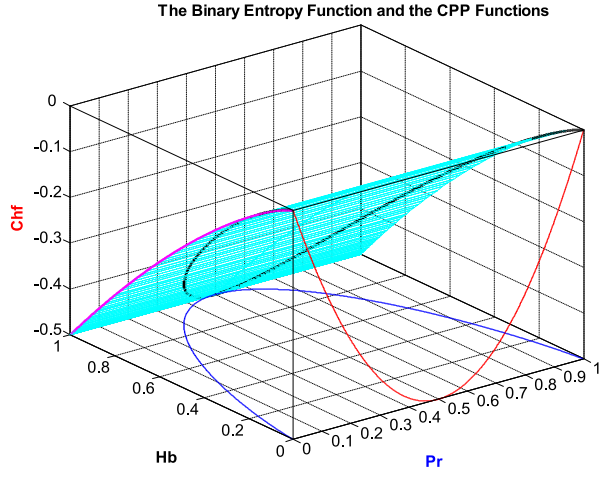


Figure 21. The graphs of $H_b = H_b^R(P_r, Chf)$ in black in the surface $H_b^R(Chf)$in cyan and of $H_b^R(P_r)$ in blue and of $Chf(P_r)$ in red.

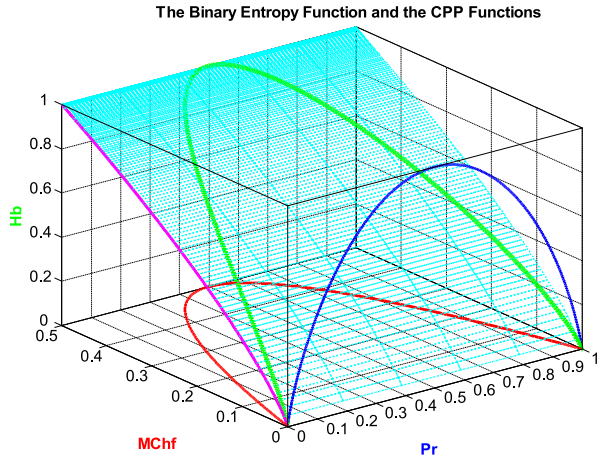


Figure 22. The graphs of $H_b = H_b^R(P_r, MChf)$ in green in the surface $H_b^R(MChf)$ in cyan and of $H_b^R(P_r)$ in blue and of $MChf(P_r)$ in red.

6.1.3. The real binary entropy $H_b^R = H_b$ as a function of MChf alone:

As $Pc^2 = DOK + MChf = 1 \Rightarrow DOK = 1 - MChf$ and $MChf = -Chf$ and $Pc = 1$ always, then (Figure 22):

$$H_b^R(MChf) = \frac{1}{2}\left\{1 - \left[\left(1 - \sqrt{1 - 2MChf}\right)\log_2(MChf) + \left(\sqrt{1 - 2MChf}\right)\log_2\left(1 - MChf + \sqrt{1 - 2MChf}\right)\right]\right\} \quad (27)$$

6.1.4. The real binary entropy $H_b^R = H_b$ as a function of DOK alone:

As $Pc^2 = DOK - Chf = 1 \Rightarrow Chf = DOK - 1$ and $1 + 2Chf = 2DOK - 1$.

And $Pc^2 = DOK + MChf = 1 \Rightarrow MChf = 1 - DOK$ and $Pc = 1$ always, then (Figure 23):

$$H_b^R(DOK) = \frac{1}{2}\left\{1 - \left[\left(1 - \sqrt{2DOK - 1}\right)\log_2(1 - DOK) + \left(\sqrt{2DOK - 1}\right)\log_2\left(DOK + \sqrt{2DOK - 1}\right)\right]\right\} \quad (28)$$

6.1.5. The real binary entropy $H_b^R = H_b$ as a function of MChf and DOK alone:

As $Pc^2 = DOK - Chf = 1 \Rightarrow Chf = DOK - 1$and $Pc = 1$ always, then:

$$H_b^R(MChf, DOK) = \frac{1}{2}\left\{1 - \left[\left(1 - \sqrt{2DOK - 1}\right)\log_2(MChf) + \left(\sqrt{2DOK - 1}\right)\log_2\left(DOK + \sqrt{2DOK - 1}\right)\right]\right\} \quad (29)$$

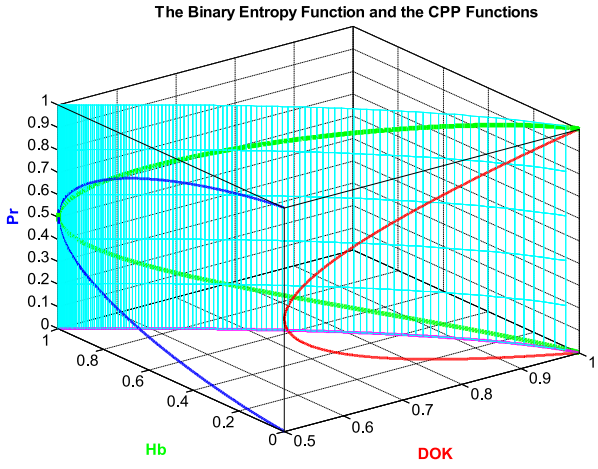


Figure 23. The graphs of $H_b = H_b^R(DOK, P_r)$ in green in the surface $H_b^R(DOK)$ in cyan and of $H_b^R(P_r)$ in blue and of $DOK(P_r)$ in red.

6.2. The complex binary entropy H_b^M:

6.2.1. Definition of H_b^M

From information theory we have:

$$H_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

In the real probability set $\mathscr{R}$ we have $p = P_r$ and $1-p = 1-P_r$ therefore

$$H_b^R(P_r) = -P_r\log_2 P_r - (1-P_r)\log_2(1-P_r)$$

In the imaginary probability set $\mathscr{M}$ we have $p = P_m = i(1-P_r)$.

If $P_r = p \Rightarrow P_m = i(1-P_r) = i(1-p) \Rightarrow P_m/i = (1-p)$.

Check that $P_r + P_m/i = p + (1-p) = 1$ which is true according to axiom 6.

If $P_r = 1-p \Rightarrow P_m = i(1-P_r) = i[1-(1-p)] = ip \Rightarrow P_m/i = p$.

Check that $P_r + P_m/i = (1-p) + p = 1$ which is also true according to axiom 6.

Therefore

$$H_b^M(P_m) = -i(1-p)\log_2[i(1-p)] - ip\log_2[ip]$$

Since $p = P_r$ and $1-p = 1-P_r$ then

$$H_b^M(P_m) = -i(1-P_r)\log_2[i(1-P_r)] - iP_r\log_2[iP_r]$$

But $P_m = i(1-P_r) = i - iP_r \Rightarrow iP_r = i - P_m$, hence we get the final expression of $H_b^M(P_m)$:

$$H_b^M(P_m) = -P_m\log_2 P_m - (i-P_m)\log_2(i-P_m) \qquad (30)$$

6.2.2. The relation between H_b^M and H_b^R:

We have:

$H_b^M = -i(1-p)\log_2[i(1-p)] - ip\log_2[ip]$ then

$$H_b^M/i = -(1-p)\log_2[i(1-p)] - p\log_2[ip]$$

Since $\log_2 xy = \log_2 x + \log_2 y$ then

$$\begin{aligned} H_b^M/i &= -(1-p)[\log_2 i + \log_2(1-p)] - p[\log_2 i + \log_2 p] \\ &= -(1-p)\log_2 i - (1-p)\log_2(1-p) - p\log_2 i - p\log_2 p \\ &= -(1-p)\log_2(1-p) - p\log_2 p - \log_2 i \\ &= H_b^R - \log_2 i \end{aligned}$$

Therefore, $H_b^M = iH_b^R - i\log_2 i = iH_b^R - \log_2 i^i$ since $\theta\log_2 x = \log_2 x^\theta$.

Now using Euler's formula: $e^{i\theta} = \cos\theta + i\sin\theta$ then for $\theta = \frac{\pi}{2} + 2k\pi$ where $k \in \mathbf{Z}$ (the set of all integers), we get

$$\begin{aligned} e^{i(\frac{\pi}{2}+2k\pi)} &= \cos\left(\frac{\pi}{2} + 2k\pi\right) + i\sin\left(\frac{\pi}{2} + 2k\pi\right) \\ &= 0 + i(1) = i. \text{ Therefore.} \end{aligned}$$

$$\begin{aligned} i^i &= \left[e^{i(\frac{\pi}{2}+2k\pi)}\right]^i = e^{i^2(\frac{\pi}{2}+2k\pi)} \\ &= e^{-(\frac{\pi}{2}+2k\pi)} \text{ since } i^2 = -1. \end{aligned}$$

Hence

$$-\log_2(i^i) = -\log_2\left[e^{-(\frac{\pi}{2}+2k\pi)}\right] = \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right)$$

since $\log_2 x^\theta = \theta\log_2 x$ and $\log_2 e = \dfrac{Lne}{Ln2} = \dfrac{1}{Ln2}$.

Note that for $k = 0 \Rightarrow -\log_2(i^i) = 2.26618$.

For $k = 1 \Rightarrow -\log_2(i^i) = 11.3309$.

For $k = -1 \Rightarrow -\log_2(i^i) = -6.79854$.

Thus we conclude that

$$\begin{aligned} H_b^M(p) &= iH_b^R(p) - i\log_2 i = iH_b^R(p) - \log_2 i^i \\ &= iH_b^R(p) + \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \text{ where } k \in \mathbf{Z}. \end{aligned} \qquad (31)$$

Note that $H_b^M(p)$ is a complex number and vector where:

$$\mathrm{Re}[H_b^M(p)] = \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) = \text{the real part of } H_b^M(p),$$

and

$$\mathrm{Im}[H_b^M(p)] = H_b^R(p) = \text{the imaginary part of } H_b^M(p).$$

Therefore $H_b^M(p)$ curve in the complex plane lies always in the constant real planes $\mathrm{Re}[H_b^M(p)]$ depending on the values of $k \in \mathbf{Z}$ and in these fixed planes it is equal to $iH_b^R(p)$.

6.3. The complementary real binary entropy $\bar{H}_b^R = H_b^R$:

The real probability in the probability set $\mathscr{R}$ is P_r and its real entropy is $H_b^R = H_b$.

According to axiom 6, the corresponding imaginary probability in the imaginary probability set $\mathscr{M}$ is $P_m = i(1 - P_r)$ and its complex entropy is H_b^M.

The related real probability to $\mathscr{M}$ in the set $\mathscr{R}$ is $P_m/i = (1 - P_r)$ and its real entropy is $\bar{H}_b^R$.

We have

$$H_b(p) = -p\log_2 p - (1-p)\log_2(1-p) \text{ and } p = P_m/i.$$

Hence

$$\begin{aligned}\bar{H}_b^R(P_m/i) &= -(P_m/i)\log_2(P_m/i) \\ &\quad - [1 - (P_m/i)]\log_2[1 - (P_m/i)]\end{aligned}$$

We have $P_m = i(1 - P_r)$ then $P_m/i = 1 - P_r$ and $1 - P_m/i = P_r$.

Therefore,

$$\begin{aligned}\bar{H}_b^R(P_m/i) &= -(1-P_r)\log_2(1-P_r) - P_r\log_2 P_r \\ &= -P_r\log_2 P_r - (1-P_r)\log_2(1-P_r) \\ &= H_b^R(P_r) \qquad (32)\end{aligned}$$

6.4. The real negative binary entropy $NegH_b^R$:

The real negative binary entropy is defined by:

$$NegH_b^R(P_r) = -H_b^R(P_r) = P_r\log_2 P_r + (1-P_r)\log_2(1-P_r) \qquad (33)$$

Note that, $0 \le H_b^R \le 1$ and $-1 \le NegH_b^R \le 0$ and if H_b^R is maximum then $NegH_b^R = -H_b^R$ is minimum and vice versa. So when $H_b^R = 1 =$ maximum for $P_r = 0.5$ then $NegH_b^R = -1 =$ minimum. Also, both of them are zero when $P_r = 0$ (impossible event) or $P_r = 1$ (sure event), so when $H_b^R = 0 =$ minimum then $NegH_b^R = 0 =$ maximum. Therefore, if H_b^R measures the amount of disorder, of uncertainty, of unpredictability, and of information gain in a message then since $NegH_b^R = -H_b^R$, that means the opposite of H_b^R, $NegH_b^R$ measures the amount of order, of certainty, of predictability, and of information loss in a message.

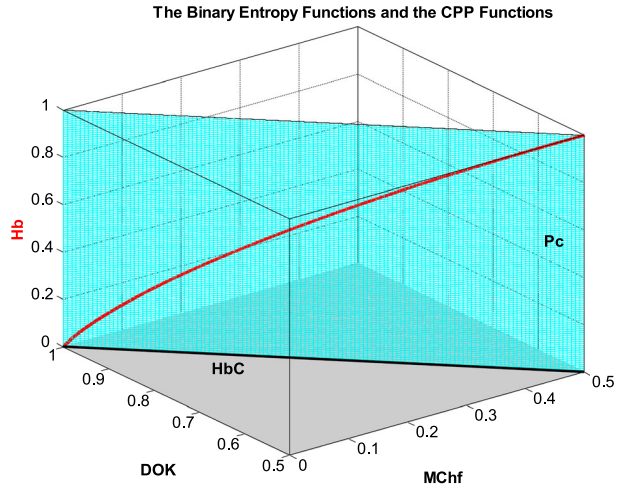


Figure 24. The graph of $H_b = H_b^R(MChf, DOK)$ in red and the graph of $H_b^C(MChf, DOK) = 0$in black in the plane $Pc^2 = DOK + MChf = 1 = Pc$ in cyan.

6.5. The binary entropy H_b^C in the set $\mathscr{C}$:

We have $H_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$.

In the probability set $\mathscr{C}$ we have $p = P_C = 1$, therefore

$$\begin{aligned}H_b^C(P_C) &= -P_C\log_2 P_C - (1-P_C)\log_2(1-P_C) \\ &= -1 \times \log_2 1 - (1-1)\log_2(1-1) \\ &= 0 - 0\log_2 0 = 0 \qquad (34)\end{aligned}$$

since by L'Hôpital's rule: $\lim_{p\to 0+} p\log_2 p = 0^-$, and $\log_2 1 = 0$. (Figure 24).

6.6. Relations between the binary entropies $H_b^R, \bar{H}_b^R, NegH_b^R, H_b^M$, and H_b^C:

Note that:

$$\begin{aligned}H_b^C(P_C) = H_b^R(P_r) - \bar{H}_b^R(P_r) = H_b^R(P_r) - H_b^R(P_r) = 0, \\ \forall P_r : 0 \le P_r \le 1, \forall P_m : 0 \le P_m \le i \qquad (35)\end{aligned}$$

$$\begin{aligned}H_b^C(P_C) = H_b^R(P_r) + NegH_b^R(P_r) = H_b^R(P_r) + [-H_b^R(P_r)] = 0, \\ \forall P_r : 0 \le P_r \le 1, \forall P_m : 0 \le P_m \le i. \qquad (36)\end{aligned}$$

$$\begin{aligned}H_b^C(P_C) &= H_b^M(P_r) - H_b^M(P_r) = H_b^M(P_r) - iH_b^R \\ &\quad - \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) = 0, \\ &\forall k \in \mathbf{Z}, \forall P_r : 0 \le P_r \le 1, \forall P_m : 0 \le P_m \le i \qquad (37)\end{aligned}$$

6.7. The derivatives of the binary entropy functions in terms of the CPP parameters (Wikipedia, the free encyclopedia, Information theory; Abou Jaoude, 2013a, 2013b; Abou Jaoude, 2014; Abou Jaoude, 2015a, 2015b; Abou Jaoude et al., 2010; Abou Jaoude, 2016a, 2016b; Abou Jaoude, 2017)

6.7.1. The first derivatives of the binary entropy functions

6.7.1.1. The first derivative of H_b^R. Since $H_b^R(p) = H_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$ then its derivative relatively to p is:

$$\frac{dH_b^R(p)}{dp} = -\log_2\left(\frac{p}{1-p}\right) = -\text{logit}_2(p)$$

We have $p = \frac{1+\sqrt{1+2Chf}}{2}$ then $1-p = \frac{1-\sqrt{1+2Chf}}{2}$.

Therefore $\frac{dH_b^R(p)}{dp} = -\log_2\left(\frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}}\right)$.

We have also

$$\begin{aligned}\frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}} &= \frac{1+\sqrt{1+2Chf}}{1-\sqrt{1+2Chf}} \times \frac{1+\sqrt{1+2Chf}}{1+\sqrt{1+2Chf}} \\ &= \frac{\left(1+\sqrt{1+2Chf}\right)^2}{1-(1+2Chf)} \\ &= \frac{1+(1+2Chf)+2\sqrt{1+2Chf}}{-2Chf} \\ &= \frac{2+2Chf+2\sqrt{1+2Chf}}{-2Chf} \\ &= \frac{1+Chf+\sqrt{1+2Chf}}{-Chf} \\ &= \frac{DOK+\sqrt{1+2Chf}}{MChf}\end{aligned}$$

since $DOK - Chf = 1 \Rightarrow DOK = 1 + Chf$ and $MChf = -Chf$.

Therefore:

$$\frac{dH_b^R(Chf, MChf, DOK)}{dp} = \begin{cases} -\log_2\left(\frac{DOK+\sqrt{1+2Chf}}{MChf}\right) & \text{if } 0.5 < p \le 1 \\ -\log_2\left(\frac{MChf}{DOK+\sqrt{1+2Chf}}\right) & \text{if } 0 \le p \le 0.5 \end{cases} \quad (38)$$

since $Chf(p)$, $MChf(p)$, and $DOK(p)$ are not monotonous functions of the strictly increasing variable $p \in [0, 1]$.

Moreover, since $0.5 \le DOK \le 1$ and $-0.5 \le Chf \le 0$ and $0 \le MChf \le 0.5$ for $\forall p \in [0, 1]$ then:

$$\begin{cases} \left(\frac{DOK+\sqrt{1+2Chf}}{MChf}\right) > 1 & \text{if } 0.5 < p \le 1 \\ \left(\frac{MChf}{DOK+\sqrt{1+2Chf}}\right) \le 1 & \text{if } 0 \le p \le 0.5 \end{cases}$$

Therefore

$$\begin{cases} \frac{dH_b^R(p)}{dp} < 0 & \text{if } 0.5 < p \le 1 \\ \frac{dH_b^R(p)}{dp} \ge 0 & \text{if } 0 \le p \le 0.5 \end{cases}$$

Hence

$$\begin{cases} H_b^R(p) \text{is decreasing} & \text{if } 0.5 < p \le 1 \\ H_b^R(p) \text{is increasing} & \text{if } 0 \le p \le 0.5 \end{cases}$$

Notice that, if $p = 0.5$ then $Chf = -0.5$ and $MChf = DOK = 0.5$, therefore

$$\begin{aligned}\frac{dH_b^R(p)}{dp} &= -\log_2\left(\frac{MChf}{DOK+\sqrt{1+2Chf}}\right) \\ &= -\log_2\left(\frac{0.5}{0.5+\sqrt{1+2(-0.5)}}\right) \\ &= -\log_2(1) = 0.\end{aligned}$$

Hence we have the maximum of $H_b^R(p)$ at this point which is absolutely true for any probability distribution.

If $p = 0$ then $Chf = MChf = 0$ and $DOK = 1$, therefore

$$\begin{aligned}\frac{dH_b^R(p)}{dp} &= -\log_2\left(\frac{MChf}{DOK+\sqrt{1+2Chf}}\right) \\ &= -\log_2\left(\frac{0}{1+\sqrt{1+2\times 0}}\right) \\ &= -\lim_{x\to 0+}\log_2(x) = -(-\infty) = \infty\end{aligned}$$

That means that at $p = 0$ the tangent to $H_b^R(p)$ is vertical and $H_b^R(p)$ is increasing.

If $p = 1$ then $Chf = MChf = 0$ and $DOK = 1$, therefore

$$\begin{aligned}\frac{dH_b^R(p)}{dp} &= -\log_2\left(\frac{DOK+\sqrt{1+2Chf}}{MChf}\right) \\ &= -\log_2\left(\frac{1+\sqrt{1+2\times 0}}{0}\right) \\ &= -\lim_{x\to 0+}\log_2\left(\frac{1}{x}\right) \\ &= -(\infty) = -\infty\end{aligned}$$

That means that at $p = 1$ the tangent to $H_b^R(p)$ is vertical and $H_b^R(p)$ is decreasing.

6.7.1.2. The first derivative of $\bar{H}_b^R$. Since $\bar{H}_b^R(p) = H_b^R(p)$ then we will reach all the same conclusions as for $H_b^R(p)$.

6.7.1.3. The first derivative of $NegH_b^R$. Since $NegH_b^R(p) = -H_b^R(p)$ then its derivative relatively to p is:

$$\frac{dNegH_b^R(p)}{dp} = -\frac{dH_b^R(p)}{dp} = \log_2\left(\frac{p}{1-p}\right) = \text{logit}_2(p)$$

Therefore:

$$\frac{dNegH_b^R(Chf, MChf, DOK)}{dp} = \begin{cases} \log_2\left(\dfrac{DOK+\sqrt{1+2Chf}}{MChf}\right) & \text{if } 0.5 < p \le 1 \\ \log_2\left(\dfrac{MChf}{DOK+\sqrt{1+2Chf}}\right) & \text{if } 0 \le p \le 0.5 \end{cases} \quad (39)$$

since $Chf(p)$, $MChf(p)$, and $DOK(p)$ are not monotonous functions of the strictly increasing variable $p \in [0, 1]$.

Moreover, since $0.5 \le DOK \le 1$ and $-0.5 \le Chf \le 0$ and $0 \le MChf \le 0.5$ for $\forall p \in [0, 1]$ then:

$$\begin{cases} \left(\dfrac{DOK+\sqrt{1+2Chf}}{MChf}\right) > 1 & \text{if } 0.5 < p \le 1 \\ \left(\dfrac{MChf}{DOK+\sqrt{1+2Chf}}\right) \le 1 & \text{if } 0 \le p \le 0.5 \end{cases}$$

Therefore

$$\begin{cases} \dfrac{dNegH_b^R(p)}{dp} > 0 & \text{if } 0.5 < p \le 1 \\ \dfrac{dNegH_b^R(p)}{dp} \le 0 & \text{if } 0 \le p \le 0.5 \end{cases}$$

Hence

$$\begin{cases} NegH_b^R(p) \text{ is increasing} & \text{if } 0.5 < p \le 1 \\ NegH_b^R(p) \text{ is decreasing} & \text{if } 0 \le p \le 0.5 \end{cases}.$$

Notice that, if $p = 0.5$ then $Chf = -0.5$ and $MChf = DOK = 0.5$, therefore

$$\frac{dNegH_b^R(p)}{dp} = \log_2\left(\frac{MChf}{DOK+\sqrt{1+2Chf}}\right) = \log_2\left(\frac{0.5}{0.5+\sqrt{1+2(-0.5)}}\right) = \log_2(1) = 0.$$

Hence we have the minimum of $NegH_b^R(p)$ at this point which is absolutely true for any probability distribution.

If $p = 0$ then $Chf = MChf = 0$ and $DOK = 1$, therefore

$$\frac{dNegH_b^R(p)}{dp} = \log_2\left(\frac{MChf}{DOK+\sqrt{1+2Chf}}\right) = \log_2\left(\frac{0}{1+\sqrt{1+2\times 0}}\right) = \lim_{x\to 0+}\log_2(x) = -\infty$$

That means that at $p = 0$ the tangent to $NegH_b^R(p)$ is vertical and $NegH_b^R(p)$ is decreasing.

If $p = 1$ then $Chf = MChf = 0$ and $DOK = 1$, therefore

$$\frac{dNegH_b^R(p)}{dp} = \log_2\left(\frac{DOK+\sqrt{1+2Chf}}{MChf}\right) = \log_2\left(\frac{1+\sqrt{1+2\times 0}}{0}\right) = \lim_{x\to 0+}\log_2\left(\frac{1}{x}\right) = \infty$$

That means that at $p = 1$ the tangent to $NegH_b^R(p)$ is vertical and $NegH_b^R(p)$is increasing.

6.7.1.4. The first derivative of H_b^M. Since

$$H_b^M(p) = iH_b^R(p) + \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right), \forall k \in \mathbf{Z}$$

then

$$\frac{dH_b^M(p)}{dp} = i \times \frac{dH_b^R(p)}{dp}$$

since

$$\frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right)$$

is a constant term so its first derivative equals 0, therefore the first derivative of $H_b^M(p)$ is similar to that of $H_b^R(p)$but in the complex plane. Hence, we will reach all the same conclusions as for $H_b^R(p)$.

6.7.1.5. The first derivative of H_b^C. Since $H_b^C(p) = 0$ then its first derivative relatively to p is:

$$\frac{dH_b^C(p)}{dp} = 0, \quad \forall p \in [0, 1];$$

$$\frac{dH_b^C(Chf)}{dp} = 0, \quad \forall Chf \in [-0.5, 0];$$

$$\frac{dH_b^C(MChf)}{dp} = 0, \quad \forall MChf \in [0, 0.5];$$

$$\frac{dH_b^C(DOK)}{dp} = 0, \quad \forall DOK \in [0.5, 1].$$

That means $H_b^C(p)$ is a horizontal line which is always equal to zero.

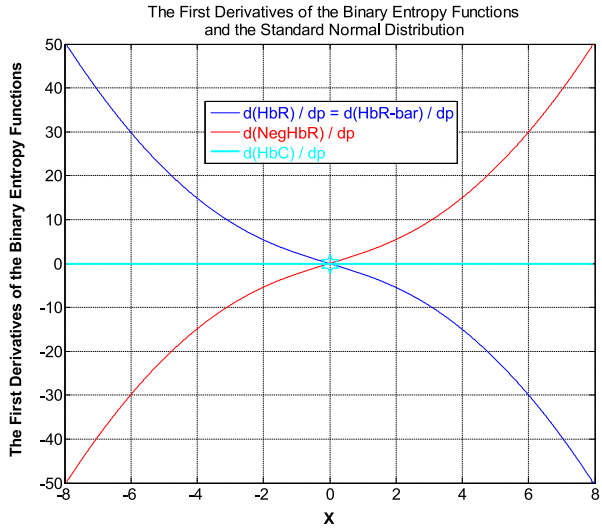


Figure 25. The first derivatives of the binary entropy functions for the standard Gaussian normal distribution.

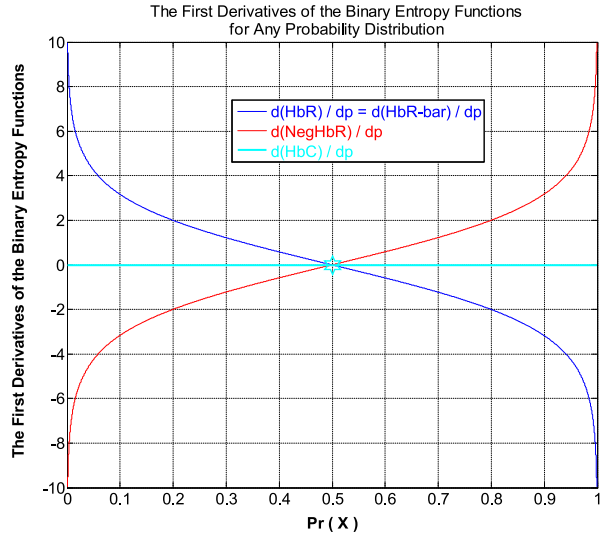


Figure 27. The first derivatives of the binary entropy functions for any probability distribution.

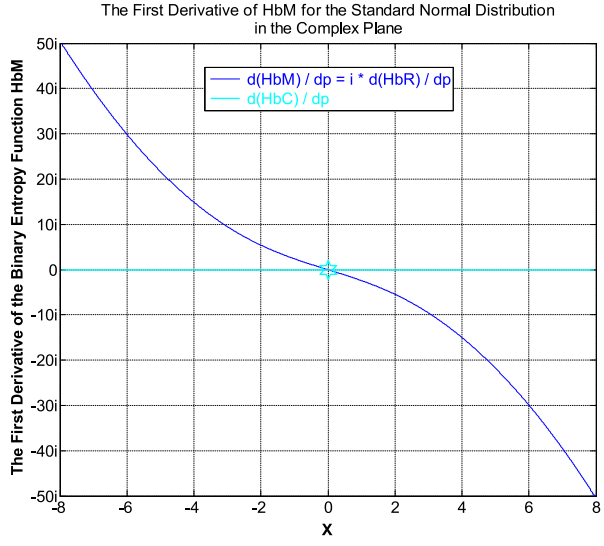


Figure 26. The first derivative of the entropy H_b^M in the complex plane for the standard Gaussian normal distribution.

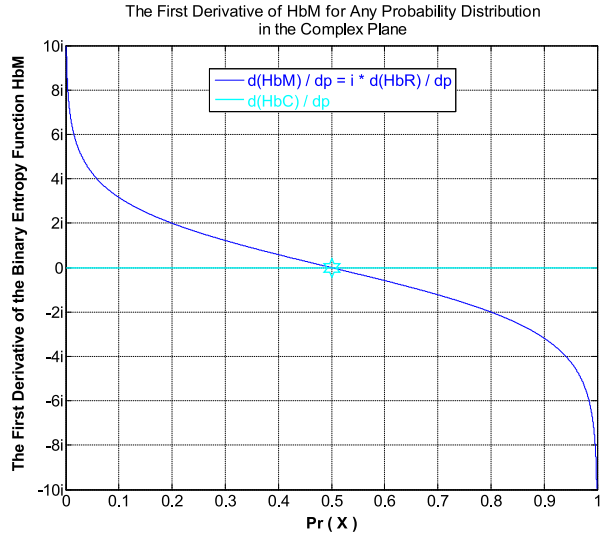


Figure 28. The first derivative of the entropy H_b^M in the complex plane for any probability distribution.

The figures above illustrate all these calculations. (Figures 25–28).

6.7.2. The second derivatives of the binary entropy functions

6.7.2.1. The second derivative of H_b^R. The second derivative of $H_b^R(p) = H_b(p)$ is:

$$\frac{d^2H_b^R(p)}{dp^2} = \frac{d}{dp}\left[-\log_2\left(\frac{p}{1-p}\right)\right]$$

$$= \frac{d}{dp}[-\text{logit}_2(p)]$$

$$= \frac{-1}{[p(1-p)]Ln2} \quad (40)$$

Since $p \in [0,1]$ then $p \geq 0$ and $1-p \geq 0$ as well as $Ln2 > 0$.

Therefore $\frac{d^2H_b^R(p)}{dp^2} < 0$, that means that $H_b^R(p)$is a curve concave down everywhere, which is absolutely true for any probability distribution.

We have $p = \frac{1+\sqrt{1+2Chf}}{2}$ then $1-p = \frac{1-\sqrt{1+2Chf}}{2}$, therefore

$$p \times (1-p) = \frac{1^2 - \left(\sqrt{1+2Chf}\right)^2}{4} = \frac{-2Chf}{4} = \frac{-Chf}{2} = \frac{MChf}{2} = \frac{(1-DOK)}{2},$$

since $DOK - Chf = 1 \Rightarrow -Chf = 1 - DOK$ and $MChf = -Chf$.

Consequently,

$$\frac{d^2H_b^R(Chf)}{dp^2} = \frac{2}{Chf \times Ln2},$$
$$\frac{d^2H_b^R(MChf)}{dp^2} = \frac{-2}{MChf \times Ln2},$$
$$\frac{d^2H_b^R(DOK)}{dp^2} = \frac{2}{(DOK-1) \times Ln2}.$$

Notice that $-0.5 \le Chf \le 0 \quad \forall p \in [0,1]$, then $\frac{d^2H_b^R(Chf)}{dp^2} < 0$ so the same conclusion as above.

In addition, $0 \le MChf \le 0.5 \quad \forall p \in [0,1]$, then $\frac{d^2H_b^R(MChf)}{dp^2} < 0$ so the same conclusion as above.

Also $0.5 \le DOK \le 1 \Rightarrow -0.5 \le DOK - 1 \le 0 \quad \forall p \in [0,1]$, then $\frac{d^2H_b^R(DOK)}{dp^2} < 0$ so the same conclusion as above.

6.7.2.2. The second derivative of $\bar{H}_b^R$.

Since $\bar{H}_b^R(p) = H_b^R(p)$ then we will reach all the same conclusions as for $H_b^R(p)$.

6.7.2.3. The second derivative of $NegH_b^R$.

Since $NegH_b^R(p) = -H_b^R(p)$ then the second derivative of $NegH_b^R(p)$ is:

$$\frac{d^2NegH_b^R(p)}{dp^2} = -\frac{d^2H_b^R(p)}{dp^2} = \frac{d}{dp}\left[\log_2\left(\frac{p}{1-p}\right)\right] = \frac{d}{dp}[\text{logit}_2(p)] = \frac{1}{[p(1-p)]Ln2} \quad (41)$$

Since $p \in [0,1]$ then $p \ge 0$ and $1-p \ge 0$ as well as $Ln2 > 0$.

Therefore $\frac{d^2NegH_b^R(p)}{dp^2} > 0$, that means that $NegH_b^R(p)$is a curve concave up everywhere, which is absolutely true for any probability distribution.

Consequently,

$$\frac{d^2NegH_b^R(Chf)}{dp^2} = \frac{-2}{Chf \times Ln2},$$
$$\frac{d^2NegH_b^R(MChf)}{dp^2} = \frac{2}{MChf \times Ln2},$$
$$\frac{d^2NegH_b^R(DOK)}{dp^2} = \frac{-2}{(DOK-1) \times Ln2}$$

Notice that $-0.5 \le Chf \le 0 \quad \forall p \in [0,1]$, then $\frac{d^2NegH_b^R(Chf)}{dp^2} > 0$ so the same conclusion as above.

In addition, $0 \le MChf \le 0.5 \quad \forall p \in [0,1]$, then $\frac{d^2NegH_b^R(MChf)}{dp^2} > 0$ so the same conclusion as above.

Also $0.5 \le DOK \le 1 \Rightarrow -0.5 \le DOK - 1 \le 0 \quad \forall p \in [0,1]$, then $\frac{d^2NegH_b^R(DOK)}{dp^2} > 0$ so the same conclusion as above.

6.7.2.4. The second derivative of H_b^M.

Since $H_b^M(p) = iH_b^R(p) + \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right), \forall k \in \mathbf{Z}$ then $\frac{d^2H_b^M(p)}{dp^2} = i \times \frac{d^2H_b^R(p)}{dp^2}$ since $\frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right)$ is a constant term so its second derivative equals 0, therefore the second derivative of $H_b^M(p)$ is similar to that of $H_b^R(p)$but in the complex plane. Hence, we will reach all the same conclusions as for $H_b^R(p)$.

6.7.2.5. The second derivative of H_b^C.

Since $H_b^C(p) = 0$ then its second derivative relatively to p is:

$$\frac{d^2H_b^C(p)}{dp^2} = 0, \quad \forall p \in [0,1];$$
$$\frac{d^2H_b^C(Chf)}{dp^2} = 0, \quad \forall Chf \in [-0.5,0];$$
$$\frac{d^2H_b^C(MChf)}{dp^2} = 0, \quad \forall MChf \in [0,0.5];$$
$$\frac{d^2H_b^C(DOK)}{dp^2} = 0, \quad \forall DOK \in [0.5,1];$$

That means $H_b^C(p)$ is a horizontal line which is always equal to zero and with zero concavity.

The figures below illustrate all these calculations (Figures 29–32).

6.8. The taylor series for all the binary entropies (Wikipedia, the free encyclopedia, Information theory)

The Taylor series of the binary entropies functions in a neighbourhood of 1/2 are

$$H_b^R(p) = 1 - \frac{1}{2Ln2}\sum_{n=1}^{\infty}\frac{(1-2p)^{2n}}{n(2n-1)}, \text{for } 0 \le p \le 1 \quad (42)$$

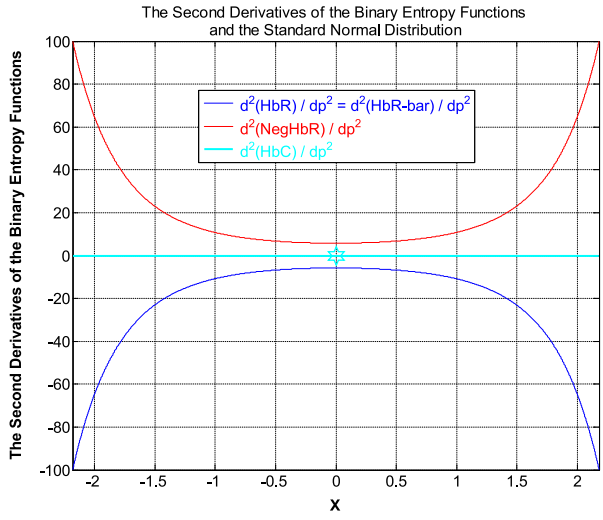


Figure 29. The second derivatives of the binary entropy functions for the standard Gaussian normal distribution.

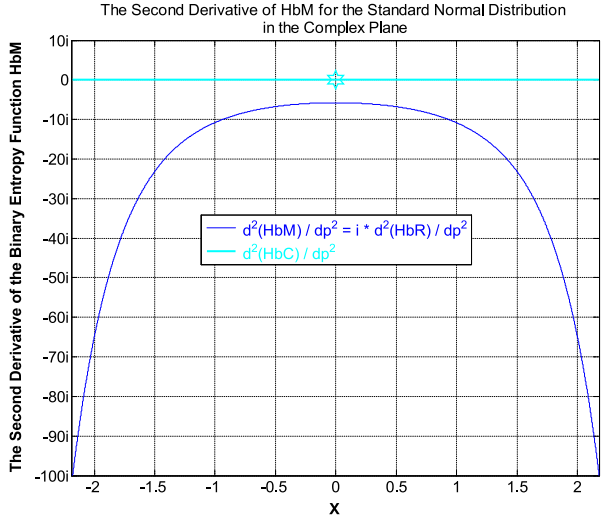


Figure 30. The second derivative of the entropy H_b^M in the complex plane for the standard Gaussian normal distribution.

$$\bar{H}_b^R(p) = 1 - \frac{1}{2Ln2}\sum_{n=1}^{\infty}\frac{(1-2p)^{2n}}{n(2n-1)} \tag{43}$$

$$NegH_b^R(p) = -1 + \frac{1}{2Ln2}\sum_{n=1}^{\infty}\frac{(1-2p)^{2n}}{n(2n-1)} \tag{44}$$

$$\begin{aligned} H_b^M(p) &= iH_b^R(p) + \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \\ &= i\mathrm{Im}[H_b^M(p)] + \mathrm{Re}[H_b^M(p)] \\ &= i\left[1 - \frac{1}{2Ln2}\sum_{n=1}^{\infty}\frac{(1-2p)^{2n}}{n(2n-1)}\right] \\ &\quad + \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right), \quad \forall k \in \mathbf{Z}. \end{aligned} \tag{45}$$

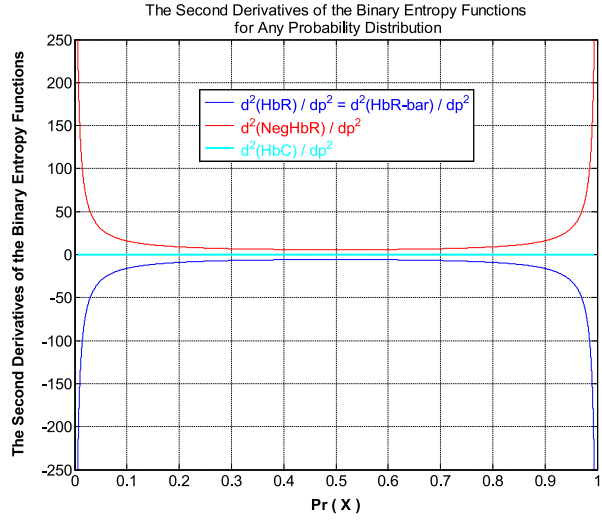


Figure 31. The second derivatives of the binary entropy functions for any probability distribution.

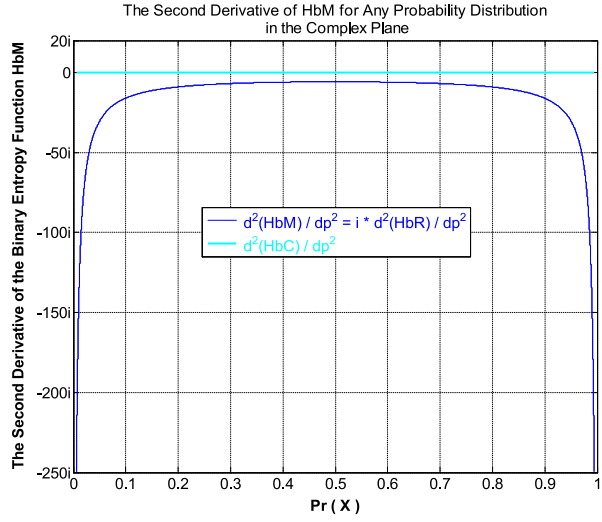


Figure 32. The second derivative of the entropy H_b^M in the complex plane for any probability distribution.

$$H_b^C(p) = H_b^R(p) + NegH_b^R(p) = 0 \tag{46}$$

The figures below illustrate all these calculations (Figures 33–35).

6.9. Graphical representations

The figures below verify and illustrate all the computations and calculations made (Figures 36–42).

6.10. Analysis

We have always: $0 \le p \le 1$, $0 \le 1-p \le 1$, $0 \le P_r \le 1$, and $0 \le P_m \le i$ then:

$0 \le H_b^R \le 1$ and $\psi \le H_b^M \le \psi + i$ where $H_b^M = \psi + iH_b^R$ and $\psi = \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right), k \in \mathbf{Z}$.

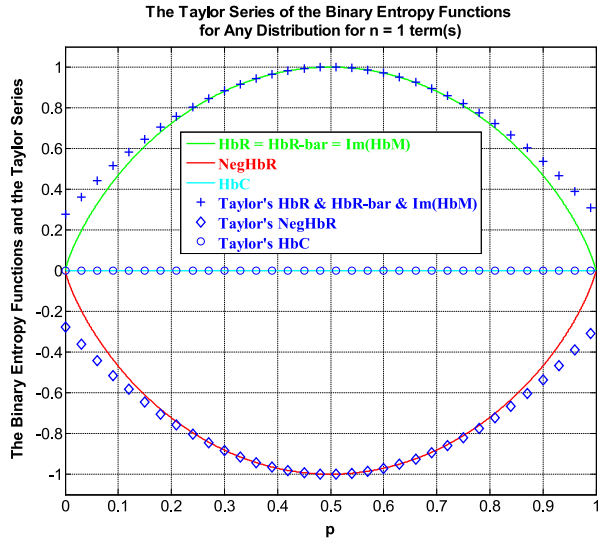


Figure 33. The Taylor series of the binary entropy functions for any probability distribution for $n = 1$ term.

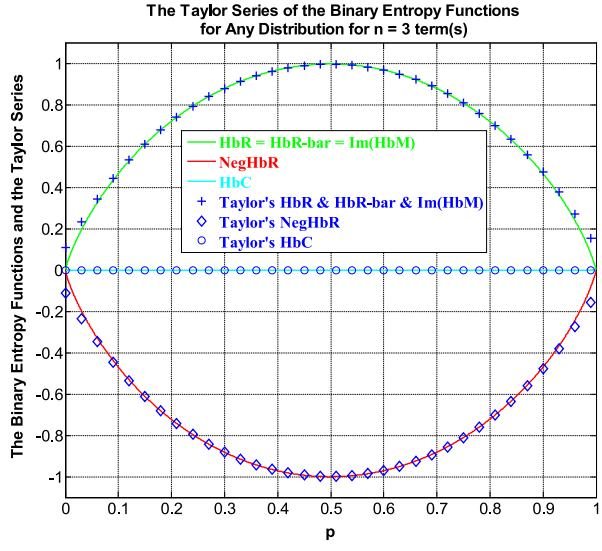


Figure 34. The Taylor series of the binary entropy functions for any probability distribution for $n = 3$ terms.

If $P_r = 0$ or $P_r = 1$ then $P_m = 0$ or $P_m = i$ therefore $H_b^R = 0$ and $H_b^M = \psi$.

If $P_r = 0.5$ then $P_m = 0.5i$ therefore $H_b^R = 1$ and $H_b^M = \psi + i$.

And this for any probability distribution of $P_r(X)$.

Moreover, we have always:

$$\mathrm{Re}(H_b^M) = \psi, \mathrm{Im}(H_b^M) = H_b^R, \text{ and } H_b^R = \bar{H}_b^R.$$

This is due to the symmetric nature in the expression of $H_b^R(P_r) = H_b(P_r)$ which leads to continuous compensations in its formula between p and $1 - p$ $\forall p : 0 \leq p \leq 1$ as well as between P_r and $P_m/i = 1 - P_r$ $\forall P_r, \forall P_m/i : 0 \leq P_r, P_m/i \leq 1$.

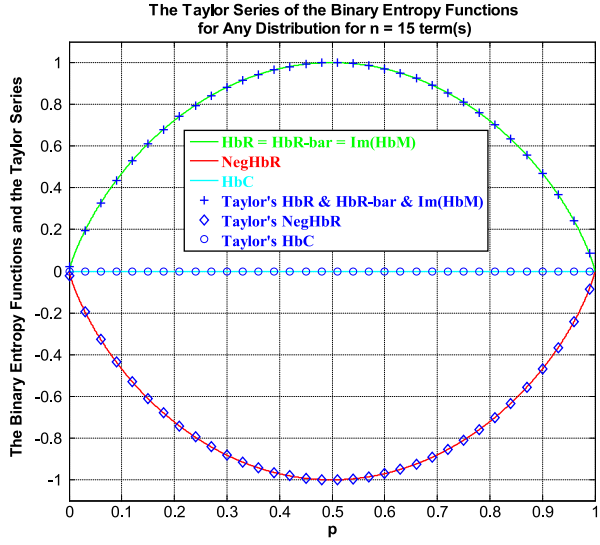


Figure 35. The Taylor series of the binary entropy functions for any probability distribution for $n = 15$ terms.

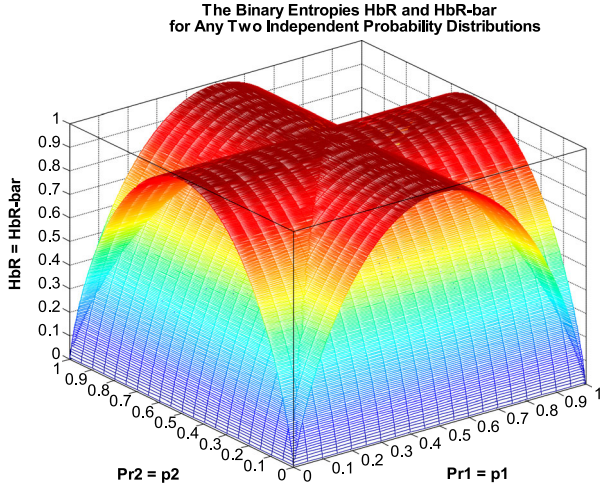


Figure 36. The graph of $H_b^R(p) = \bar{H}_b^R(p)$ for any two independent probability distributions.

7. The CPP and the channel capacities in $\mathscr{R}$, $\mathscr{M}$, and $\mathscr{C}$ (Campbell, 1982; Escolano et al., 2009; Haggerty, 1981; Noth, 1981; Seife, 2006; Theil, 1967; Wikipedia, the free encyclopedia, *Information theory*)

We have from Shannon's information theory $C_{BSC}(p) = 1 - H_b(p)$ where p is the probability of bits flips and BSC = binary symmetric channel.

The channel capacity in $\mathscr{R}$ corresponding to the real probability $P_r = p$ will be

$$C_{BSC}^R(p) = C_{BSC}(p) = 1 - H_b^R(p) \tag{47}$$

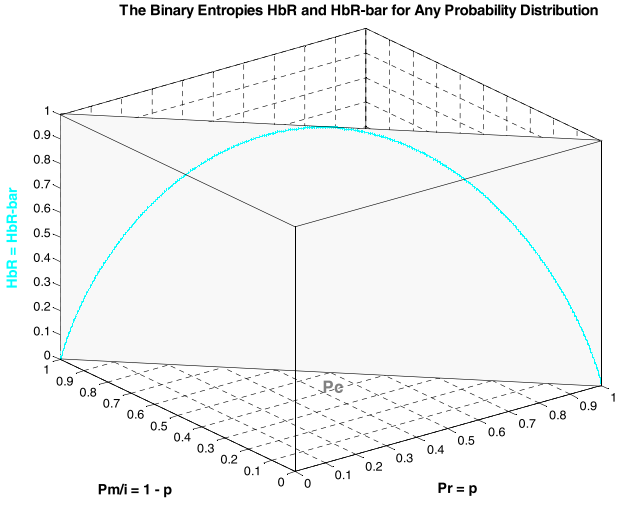


Figure 37. The graph of $H_b^R(p) = \bar{H}_b^R(p)$ in cyan in the plane $Pc(p) = P_r + P_m/i = 1$ in grey for any probability distribution.

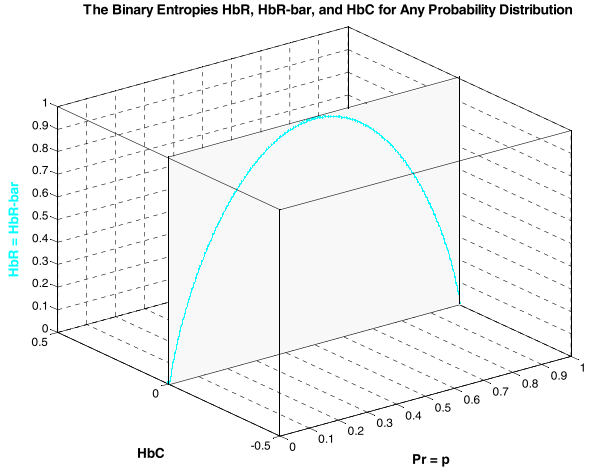


Figure 38. The graph of $H_b^R(p) = \bar{H}_b^R(p)$ in cyan in the plane $H_b^C(p) = 0$ in grey for any probability distribution.

Since $0 \le p \le 1$ then $0 \le H_b^R(p) \le 1$ and $0 \le C_{BSC}^R(p) \le 1$. Also when $H_b^R(p)$ is minimum then $C_{BSC}^R(p)$ is maximum and vice versa.

The channel capacity in $\mathscr{M}$ corresponding to the imaginary probability $P_m = i(1 - P_r) = i(1 - p)$ will be

$$C_{BSC}^M(p) = i - H_b^M(p) \tag{48}$$

Since we have $H_b^M(p) = iH_b^R(p) + \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right)$ where $k \in \boldsymbol{Z}$, consequently:

$$\begin{aligned} C_{BSC}^M(p) &= i - H_b^M(p) \\ &= i - iH_b^R(p) - \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \\ &= i[1 - H_b^R(p)] - \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \end{aligned}$$

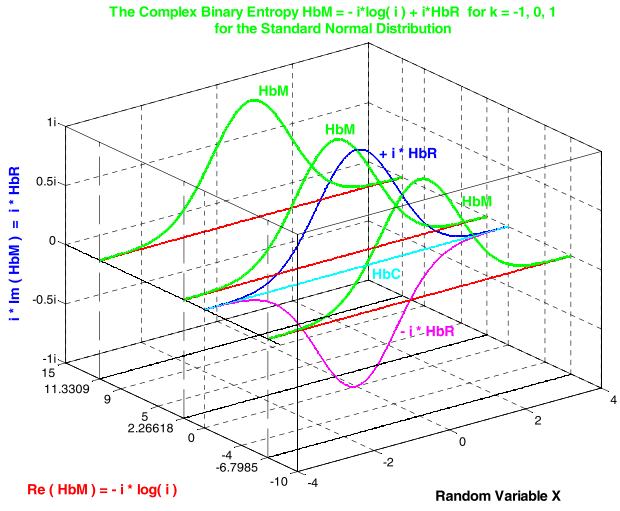


Figure 39. The graphs of $H_b^M = -i\log_2 i + iH_b^R$ for $k = -1, 0, 1$ in green with $-iH_b^R = iNegH_b^R$ and $H_b^R = \bar{H}_b^R$ for the standard Gaussian normal distribution.

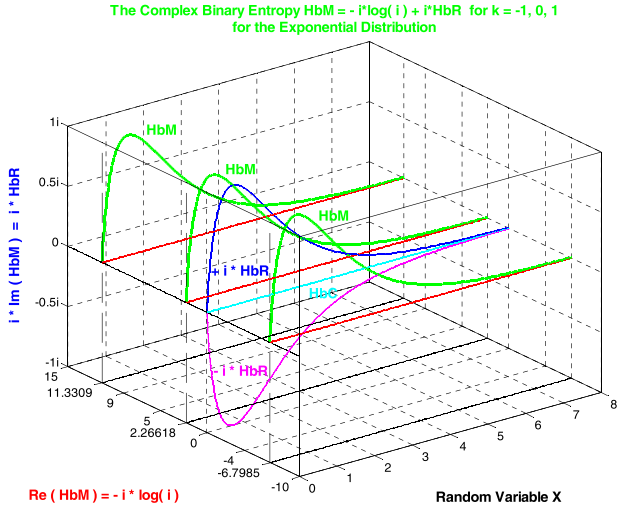


Figure 40. The graphs of $H_b^M = -i\log_2 i + iH_b^R$ for $k = -1, 0, 1$ in green with $-iH_b^R = iNegH_b^R$ and $H_b^R = \bar{H}_b^R$ for an exponential distribution.

$$\begin{aligned} &= iC_{BSC}^R(p) - \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \\ &= iC_{BSC}^R(p) - \psi, \\ \text{where } \psi &= \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \end{aligned} \tag{49}$$

Therefore $\mathrm{Im}[C_{BSC}^M(p)] = C_{BSC}^R(p)$ and $\mathrm{Re}[C_{BSC}^M(p)] = -\psi$ where $k \in \boldsymbol{Z}$.

Thus $C_{BSC}^M(p)$ curve in the complex plane lies always in the constant real planes $\mathrm{Re}[C_{BSC}^M(p)] = -\psi$ depending on the values of $k \in \boldsymbol{Z}$ and in these fixed planes it is equal to $iC_{BSC}^R(p)$.

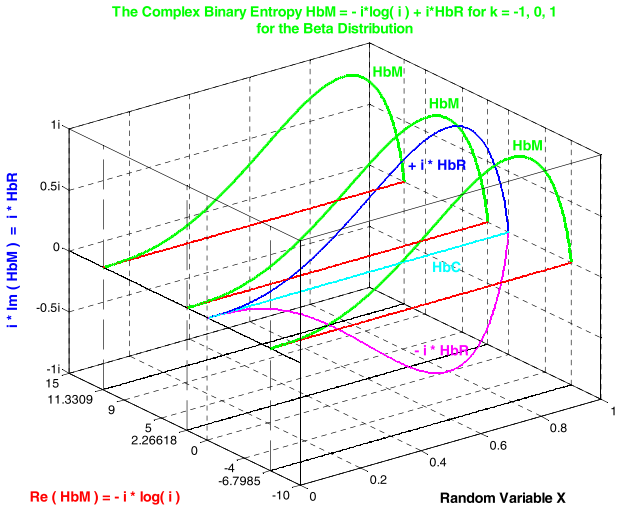


Figure 41. The graphs of $H_b^M = -i\log_2 i + iH_b^R$ for $k = -1, 0, 1$ in green with $-iH_b^R = iNegH_b^R$ and $H_b^R = \bar{H}_b^R$ for a beta distribution.

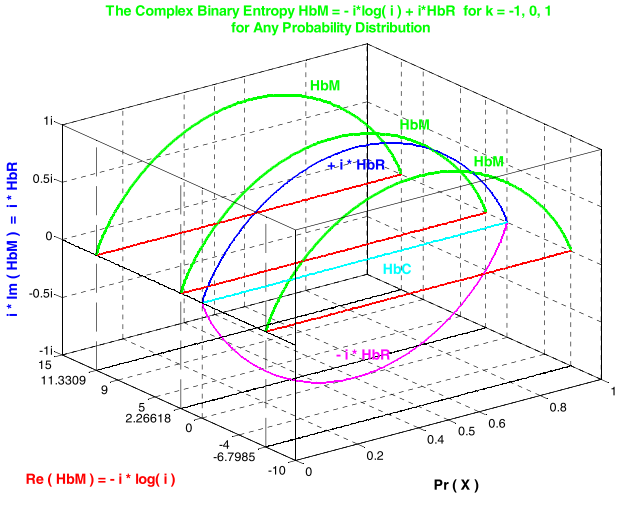


Figure 42. The graphs of $H_b^M = -i\log_2 i + iH_b^R$ for $k = -1, 0, 1$ in green with $-iH_b^R = iNegH_b^R$ and $H_b^R = \bar{H}_b^R$ for any probability distribution.

The channel capacity in $\mathscr{R}$ corresponding to the complementary real probability $P_m/i = 1 - P_r = 1 - p$ will be

$$\bar{C}_{BSC}^R(p) = 1 - \bar{H}_b^R(p) = 1 - H_b^R(p) = C_{BSC}^R(p)$$
$$\text{since } \bar{H}_b^R(p) = H_b^R(p). \tag{50}$$

The channel capacity in $\mathscr{C}$ corresponding to the probability $Pc = p = 1$ will be

$$C_{BSC}^C(p) = 1 - H_b^C(p) = 1 - 0 = 1$$
$$\text{since } H_b^C(p) = 0,\ \forall P_r : 0 \le P_r \le 1,\ \forall P_m : 0 \le P_m \le i. \tag{51}$$

The figures below illustrate all the computations and formulas deduced (Figures 43–46).

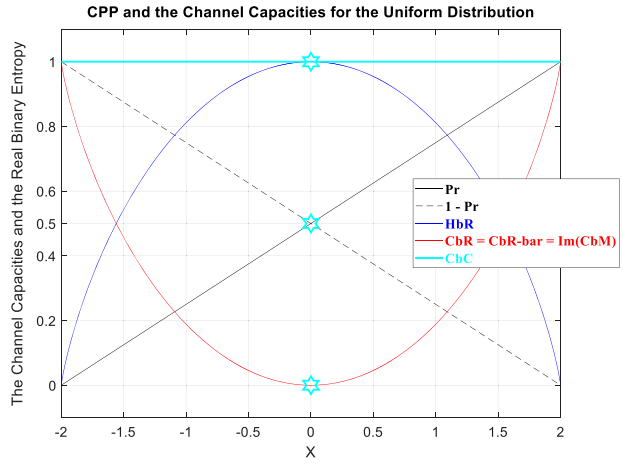


Figure 43. The complex probability paradigm and the BSC capacities for a uniform probability distribution.

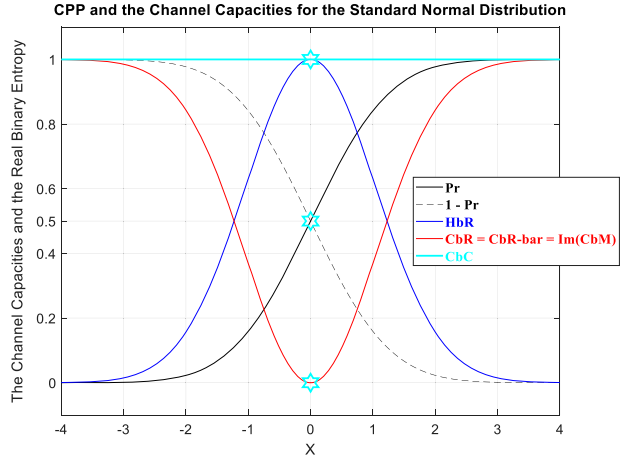


Figure 44. The complex probability paradigm and the BSC capacities for the standard Gaussian normal probability distribution.

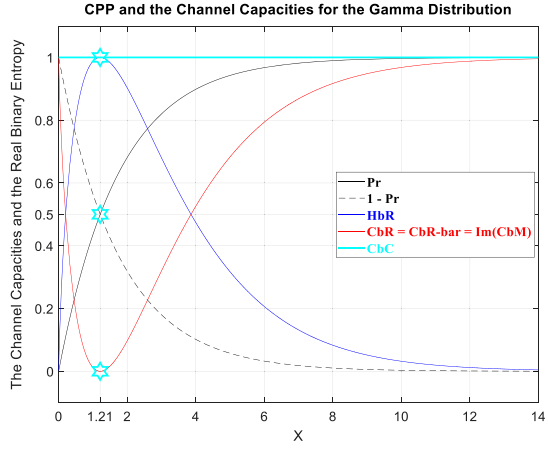


Figure 45. The complex probability paradigm and the BSC capacities for a gamma probability distribution.

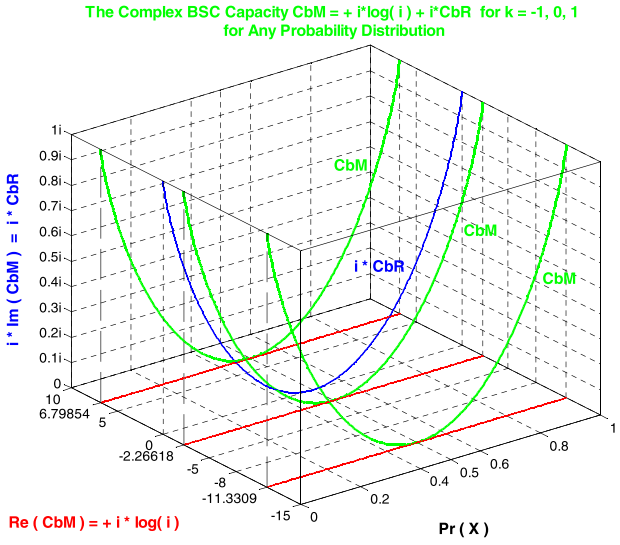


Figure 46. The graphs of $C_{BSC}^{M} = \mathrm{Re}(C_{BSC}^{M}) + i\mathrm{Im}(C_{BSC}^{M}) = i\log_2 i + iC_{BSC}^{R}$ for $k = -1, 0, 1$ in green for any probability distribution.

8. The evaluation of the new paradigm parameters

The cumulative distribution function (*CDF*) of the discrete or continuous random variable of the message x is denoted by $F(x)$. Then the new *CPP* parameters are the following:

The real probability:

$$P_r(x) = P_{rob}(X \le x) = F(x) \tag{52}$$

The imaginary probability:

$$P_m(x) = i \times [1 - P_r(x)] = i \times [1 - F(x)] \tag{53}$$

The real complementary probability:

$$P_m(x)/i = [1 - P_r(x)] = [1 - F(x)] \tag{54}$$

The complex random vector:

$$Z(x) = P_r(x) + P_m(x) = P_r(x) + i \times [1 - P_r(x)] \tag{55}$$

The Degree of Our Knowledge (*DOK*):

$$DOK(x) = |Z(x)|^2 = P_r^2(x) + [P_m(x)/i]^2 = P_r^2(x) + [1 - P_r(x)]^2 = 1 - 2P_r(x) + 2P_r^2(x) \tag{56}$$

The Chaotic Factor (*Chf*):

$$Chf(x) = 2iP_r(x)P_m(x) = -2P_r(x)[P_m(x)/i] = -2P_r(x)[1 - P_r(x)] = -2P_r(x) + 2P_r^2(x) \tag{57}$$

The Magnitude of the Chaotic Factor (*MChf*):

$$MChf(x) = |Chf(x)| = -2iP_r(x)P_m(x) = 2P_r(x)[P_m(x)/i] = 2P_r(x)[1 - P_r(x)] = 2P_r(x) - 2P_r^2(x) \tag{58}$$

For any value of the random variable x, the probability expressed in the complex set $\mathscr{C}$ is:

$$Pc^2(x) = [P_r(x) + P_m(x)/i]^2 = |Z(x)|^2 - 2iP_r(x)P_m(x) = DOK(x) - Chf(x) = DOK(x) + MChf(x) = 1 \tag{59}$$

then, $Pc(x) = P_r(x) + P_m(x)/i = P_r(x) + [1 - P_r(x)] = 1$ always.

Hence, the prediction of the outcome of the message random variable x in $\mathscr{C}$ is permanently certain and absolutely deterministic.

The surprisal of the message x in base 2 is:

$$I_2(x) = \log_2\left[\frac{1}{P_r(x)}\right] = -\log_2[P_r(x)] \tag{60}$$

The rescaled surprisal of the message x in base 2 is:

$$RI_2(x) = I_2(x)/\Phi \tag{61}$$

The expectancy of the same message x in base 2:

$$\bar{I}_2(x) = \log_2\left[\frac{1}{1 - P_r(x)}\right] = -\log_2[1 - P_r(x)] \tag{62}$$

The rescaled expectancy of the same message x in base 2:

$$R\bar{I}_2(x) = \bar{I}_2(x)/\Phi \tag{63}$$

The real binary entropy in $\mathscr{R}$ is:

$$H_b^R[P_r(x)] = -P_r(x)\log_2 P_r(x) - [1 - P_r(x)]\log_2[1 - P_r(x)] \tag{64}$$

The complex binary entropy in $\mathscr{M}$ is:

$$H_b^M[P_r(x)] = iH_b^R[P_r(x)] + \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \text{ where } k \in \mathbf{Z}. \tag{65}$$

The real complementary binary entropy in $\mathscr{R}$ is:

$$\bar{H}_b^R[P_r(x)] = H_b^R[P_r(x)] \tag{66}$$

The real negative binary entropy in $\mathscr{R}$ is:

$$NegH_b^R[P_r(x)] = -H_b^R[P_r(x)] = P_r(x)\log_2 P_r(x) + [1 - P_r(x)]\log_2[1 - P_r(x)] \tag{67}$$

The real binary entropy in $\mathscr{C}$ is:

$$H_b^C[P_r(x)] = 0, \forall P_r(x) : 0 \le P_r(x) \le 1, \quad \forall P_m(x) : 0 \le P_m(x) \le i \tag{68}$$

The real BSC capacity in $\mathscr{R}$ is:

$$C_{BSC}^R[P_r(x)] = 1 - H_b^R[P_r(x)] \tag{69}$$

The complex BSC capacity in $\mathscr{M}$ is:

$$C_{BSC}^M[P_r(x)] = iC_{BSC}^R[P_r(x)] - \frac{1}{Ln2}\left(\frac{\pi}{2} + 2k\pi\right) \text{ where } k \in \mathbf{Z} \tag{70}$$

The real complementary BSC capacity in $\mathscr{R}$ is:

$$\bar{C}^{R}_{BSC}[P_r(x)] = C^{R}_{BSC}[P_r(x)] \tag{71}$$

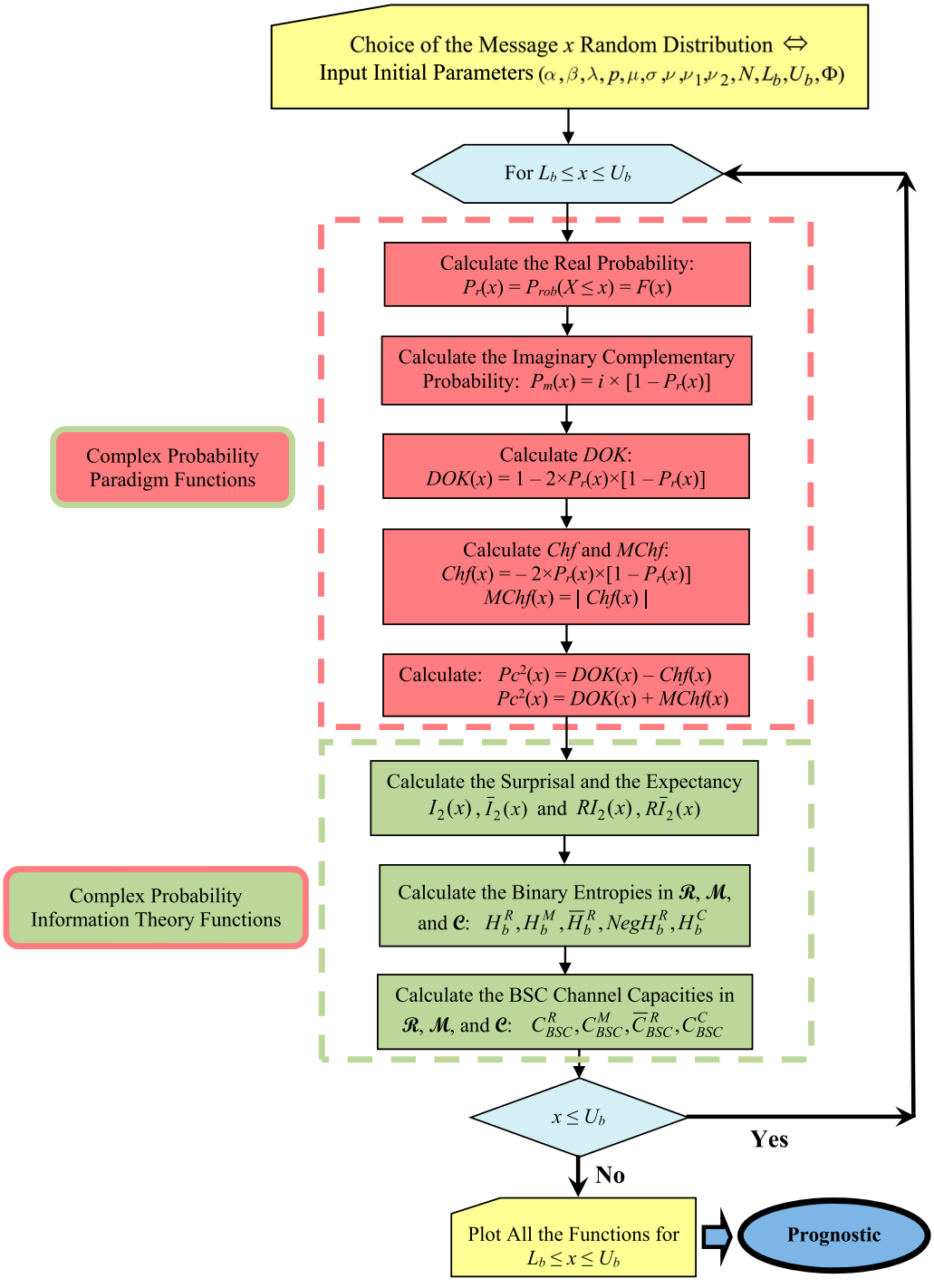
The real BSC capacity in $\mathscr{C}$ is:

$$C^{C}_{BSC}[P_r(x)] = 1, \quad \forall P_r(x) : 0 \le P_r(x) \le 1, \quad \forall P_m(x) : 0 \le P_m(x) \le i \tag{72}$$

Let us consider thereafter different discrete and continuous probability distributions to simulate the probability cumulative distribution function $P_r(x) = F(x)$ and to draw, to visualise, as well as to quantify all the information theory new paradigm parameters.

9. Flowchart of the complex probability information theory paradigm

The following flowchart summarises all the procedures of the proposed complex probability prognostic model for L_b (Lower bound) $\le$ (Message x) $\le U_b$ (Upper bound):



10. The new paradigm applied to various discrete and continuous probability distributions

In this section, the simulation of the novel CPP model for various discrete and continuous random distributions will be done. Note that all the numerical values found in the paradigm functions analysis for all the simulations were computed using the 64-Bit MATLAB version 2017 software. It is important to mention here that a few important and well-known probability distributions were considered although the original CPP model can be applied to any random distribution beside the ten probability cases below. This will lead to similar results and conclusions. Hence, the new paradigm is successful with any discrete or continuous random case (refer to the definitions, graphs, and axioms in section III).

10.1. Simulation of discrete probability distributions

10.1.1. The binomial probability distribution

The probability density function (*PDF*) of this discrete distribution is:

$$f(x) = {}_NC_x p^x q^{N-x}, \quad \text{where } p+q=1 \text{ and } 0 \le x \le N \tag{73}$$

Taking in our simulation $N = 16$ and $p + q = 1$, then:

The mean of this binomial random distribution is: $\mu = Np = 16 \times 0.5 = 8$.

The standard deviation is: $\sigma = \sqrt{Npq} = \sqrt{16 \times 0.5 \times 0.5} = \sqrt{4} = 2$.

The Median is $Md = \mu = 8$.

The cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \sum_{k=0}^{x} f(k) = \sum_{k=0}^{x} {}_NC_k p^k q^{N-k} \tag{74}$$

I have taken the domain for the binomial random variable to be: $x \in [L_b = 0, U_b = N = 16]$ and $\Delta x = x_k - x_{k-1} = 1$, then: $x = 0, 1, 2, \ldots, 16$.

The real probability$P_r(x)$ is:

$$P_r(x) = F(x) = \sum_{k=0}^{x} {}_NC_k p^k q^{N-k}$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x) = 1 - \sum_{k=0}^{x} {}_NC_k p^k q^{N-k}$$

$$= \sum_{k=x+1}^{N=16} {}_NC_k p^k q^{N-k}$$

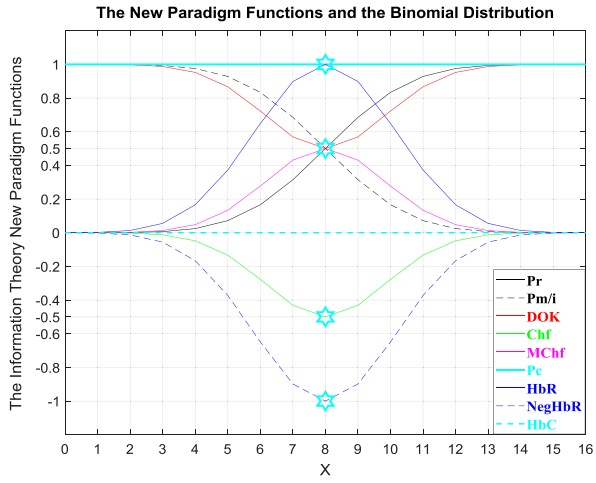


Figure 47. The new paradigm functions and the binomial distribution.

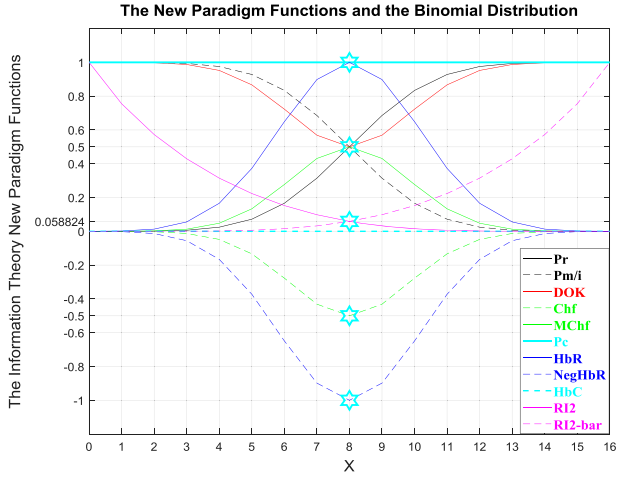


Figure 48. The new paradigm functions with the rescaled surprisal and expectancy and the binomial distribution.

The rescaled surprisal and expectancy self-information functions with the simulation rescaling factor $\Phi = 17$ are:

$$RI_2(x = Md = 8) = \bar{RI}_2(x = Md = 8) = -\log_2[p(8)]/\Phi$$

$$= -\log_2(0.5)/17 = 1/17 = 0.058824 \text{ bits}.$$

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 47–49).

10.1.2. The poisson probability distribution

The probability density function (*PDF*) of this discrete distribution is:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ where } 0 \le x < \infty. \tag{75}$$

and the cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \sum_{k=0}^{x} f(k) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!} \tag{76}$$

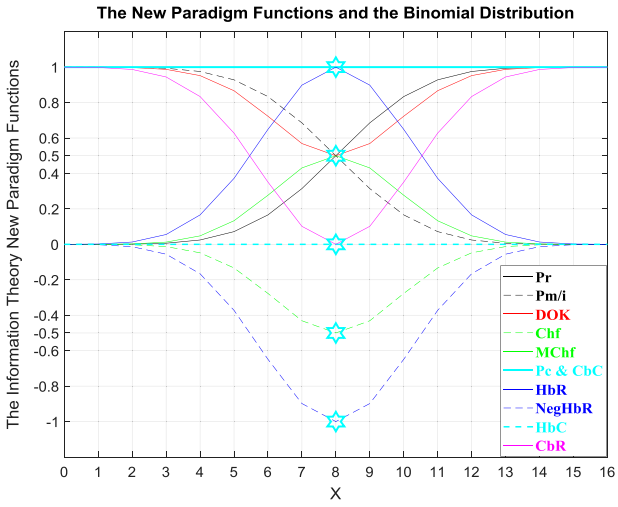


Figure 49. The new paradigm functions with the BSC capacities and the binomial distribution.

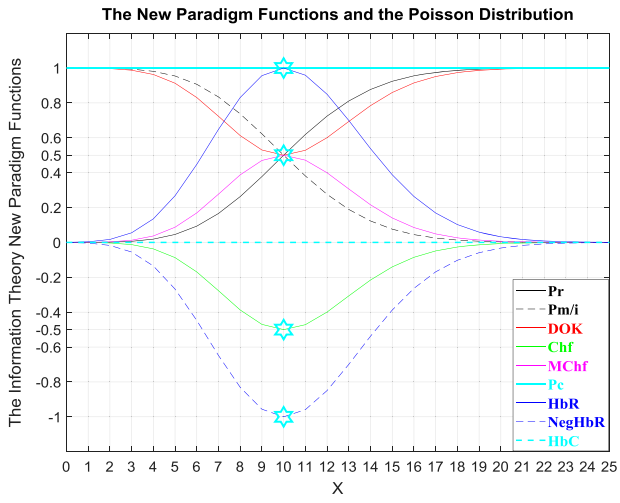


Figure 50. The new paradigm functions and the Poisson distribution.

For the Poisson random variable: $x \in [L_b = 0, \infty)$ and $\Delta x = x_k - x_{k-1} = 1$, then $x = 0, 1, 2, \ldots, \infty$.

I have taken in the simulation the domain for the Poisson random variable to be: $x \in [L_b = 0, U_b = 25]$ and $\Delta x = x_k - x_{k-1} = 1$, then: $x = 0, 1, 2, \ldots, 25$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x) = 1 - \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

$$= \sum_{k=x+1}^{25} \frac{e^{-\lambda}\lambda^k}{k!}$$

The mean of this Poisson random distribution is: $\mu = \lambda = 10.6685$.

The standard deviation is: $\sigma = \sqrt{\lambda} = \sqrt{10.6685} = 3.266267$.

The median Md is $\approx \lfloor \lambda + 1/3 - 0.02/\lambda \rfloor = 10$.

The rescaled surprisal and expectancy self-information functions with the simulation rescaling factor $\Phi = 14.82$ are:

$$RI_2(x = Md = 10) = \bar{RI}_2(x = Md = 10)$$
$$= -\log_2[p(10)]/\Phi = -\log_2(0.5)/14.82 = 1/14.82$$
$$= 0.067476 \text{ bits}$$

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 50–52).

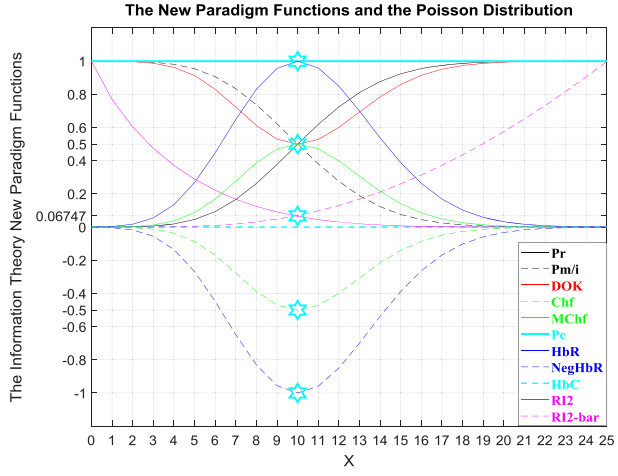


Figure 51. The new paradigm functions with the rescaled surprisal and expectancy and the Poisson distribution.

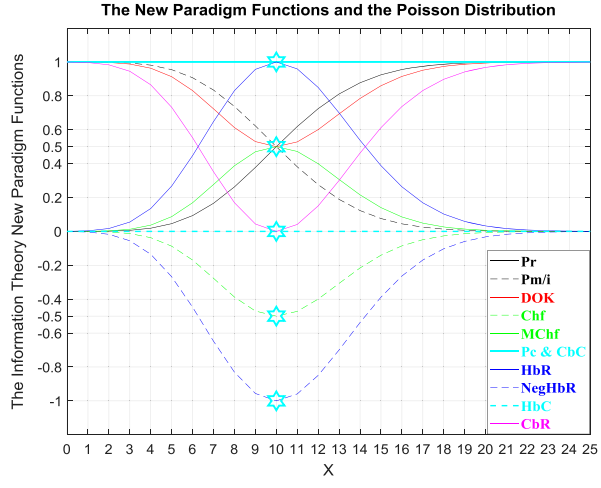


Figure 52. The new paradigm functions with the BSC capacities and the Poisson distribution.

10.2. Simulation of continuous probability distributions

10.2.1. The uniform probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \dfrac{1}{U_b - L_b} & \text{if} \quad L_b \le x \le U_b \\ 0 & \text{elsewhere} \end{cases} \tag{77}$$

and the cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_{-\infty}^{x} f(t)dt = \int_{L_b}^{x} f(t)dt$$

$$= \begin{cases} \dfrac{x - L_b}{U_b - L_b} & \text{if } L_b \le x \le U_b \\ 0 & \text{elsewhere} \end{cases} \tag{78}$$

I have taken the domain for the continuous uniform random variable to be: $x \in [L_b = -2, U_b = 2]$ and $dx = 0.01$.

Then $F(x) = \frac{x+2}{4}$ with $-2 \le x \le 2$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \frac{x+2}{4} \quad \text{with} \quad -2 \le x \le 2$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x) = 1 - \frac{x+2}{4}$$

The mean of this continuous uniform random distribution is:

$$\mu = \frac{L_b + U_b}{2} = \frac{-2+2}{2} = 0.$$

The standard deviation is:

$$\sigma = \frac{|L_b - U_b|}{\sqrt{12}} = \frac{|-2-2|}{\sqrt{12}} = \frac{4}{\sqrt{12}} = 1.1547.$$

The Median is $Md = 0$.

The rescaled surprisal and expectancy self-information functions with the simulation rescaling factor $\Phi = 8.7$ are:

$$RI_2(x = Md = 0) = \bar{RI}_2(x = Md = 0) = -\log_2[p(0)]/\Phi$$

$$= -\log_2(0.5)/8.7 = 1/8.7 = 0.11494 \text{ bits.}$$

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 53–55).

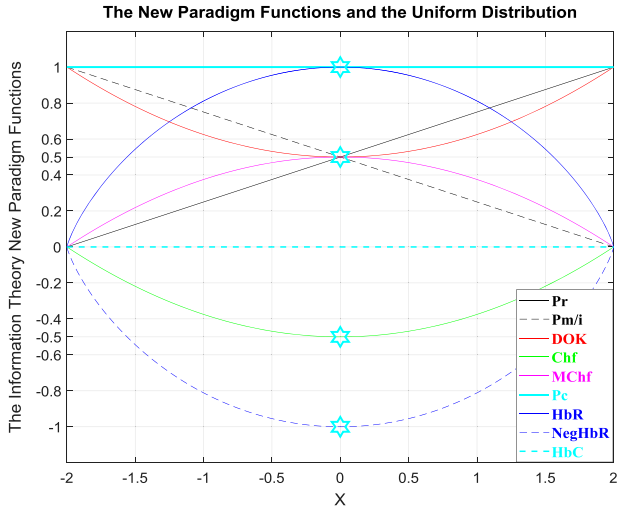


Figure 53. The new paradigm functions and the uniform probability distribution.

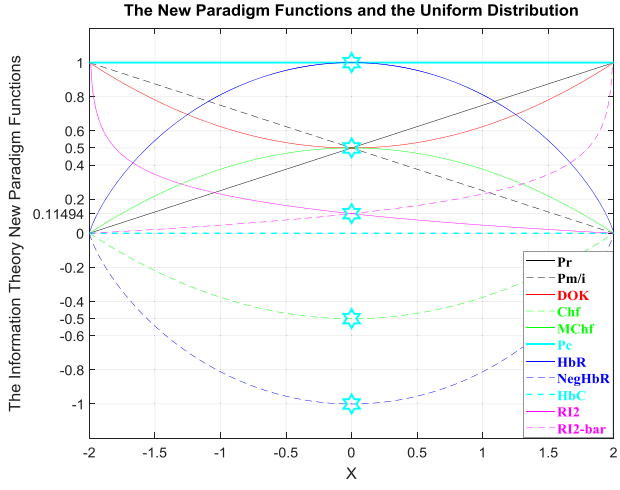


Figure 54. The new paradigm functions with the rescaled surprisal and expectancy and the uniform distribution.

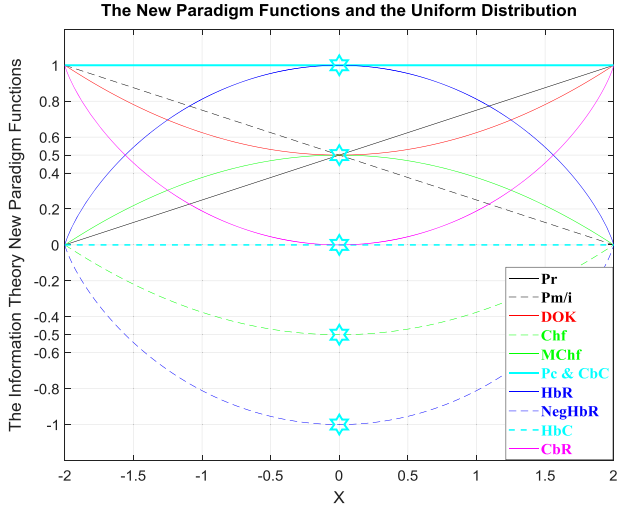


Figure 55. The new paradigm functions with the BSC capacities and the uniform distribution.

10.2.2. The standard Gaussian normal probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \text{for } -\infty < x < \infty \tag{79}$$

and the cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \tag{80}$$

The domain for this standard Gaussian normal variable is: $x \in [L_b = -4, U_b = 4]$ and I have taken $dx = 0.001$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x) = 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt$$

In the simulations, the mean of this standard normal random distribution is $\mu = 0$.

The standard deviation is $\sigma = 1$.

The median is $Md = 0$.

The rescaled surprisal and expectancy self-information functions with the simulation rescaling factor $\Phi = 15$ are:

$$RI_2(x = Md = 0) = \bar{RI}_2(x = Md = 0) = -\log_2[p(0)]/\Phi = -\log_2(0.5)/15 = 1/15 = 0.066667 \text{ bits.}$$

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 56–58).

10.2.3. The exponential probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right), \text{for } 0 \le x < \infty \tag{81}$$

Where μ is the parameter of the distribution and is equal to 1 here.

The cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_{0}^{x} f(t)dt = \int_{0}^{x} \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right) dt \tag{82}$$

The domain for this exponential variable is: $x \in [L_b = 0, U_b = 7]$ and I have taken $dx = 0.001$.

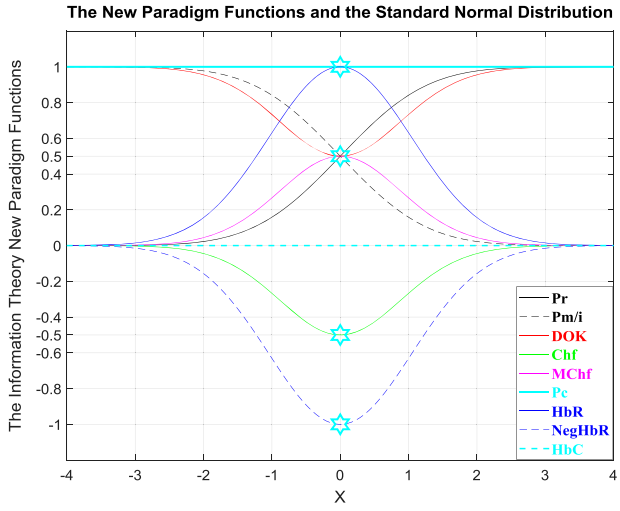


Figure 56. The new paradigm functions and the standard Gaussian normal distribution.

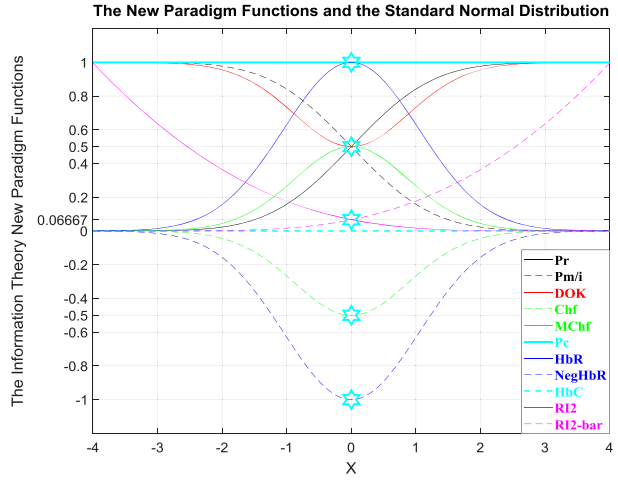


Figure 57. The new paradigm functions with the rescaled surprisal and expectancy and the standard Gaussian normal distribution.

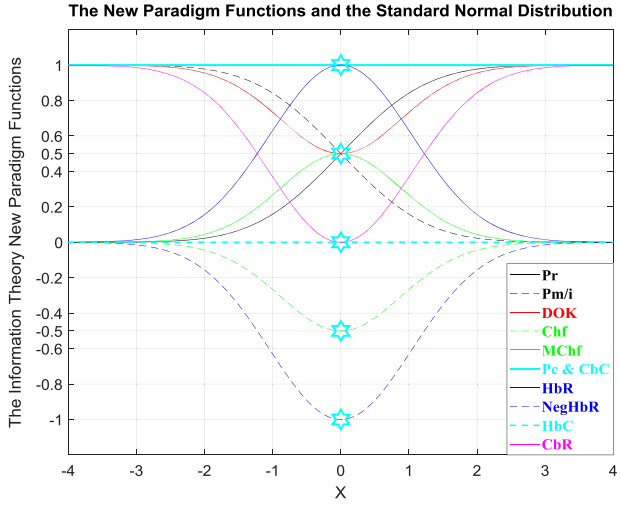


Figure 58. The new paradigm functions with the BSC capacities and the standard Gaussian normal distribution.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \int_0^x \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right) dt$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x)$$

$$= 1 - \int_0^x \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right) dt$$

$$= \int_x^\infty \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right) dt$$

In the simulations, the mean of this exponential random distribution is $\mu = 1$.

The standard deviation is $\sigma = 1$.

The median is $Md = Ln2 = 0.693147$.

The rescaled surprisal and expectancy self-information functions with the simulation rescaling factor $\Phi = 10$ are:

$$RI_2(x = Md = Ln2) = \bar{RI}_2(x = Md = Ln2)$$

$$= -\log_2[p(Ln2)]/\Phi = -\log_2(0.5)/10 = 1/10$$

$$= 0.1 \text{ bits}$$

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 59–61.)

10.2.4. The gamma probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right),$$

$$\text{for } 0 < x < \infty, \text{and } \alpha, \beta > 0 \qquad (83)$$

Where α is the shape parameter $= 1$ and β is the scale parameter $= 1.75$. $\Gamma(\alpha)$ is a complete gamma function.

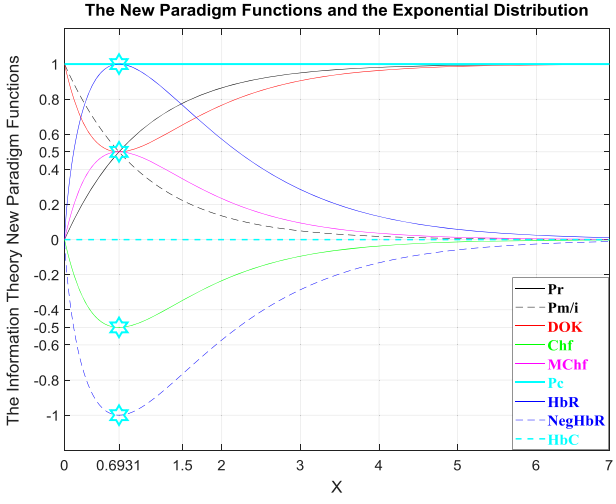


Figure 59. The new paradigm functions and the exponential distribution.

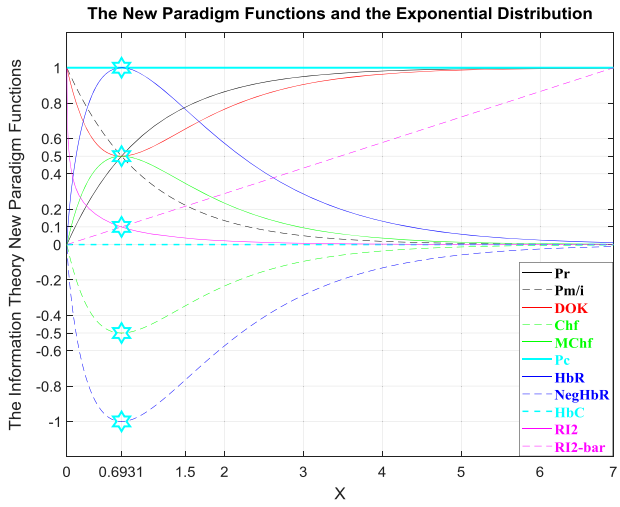


Figure 60. The new paradigm functions with the rescaled surprisal and expectancy and the exponential distribution.

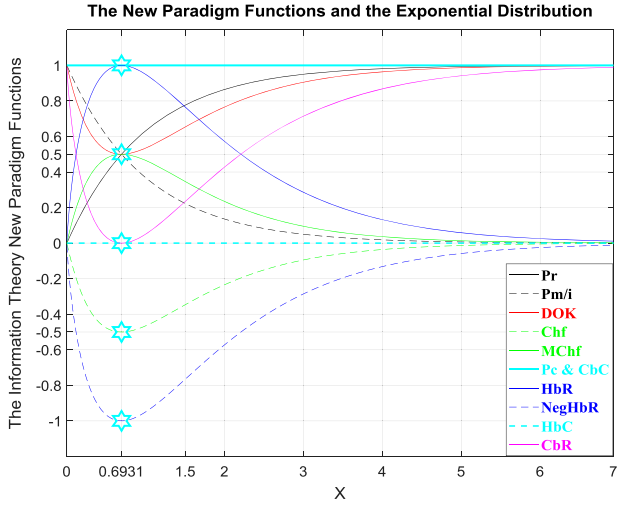


Figure 61. The new paradigm functions with the BSC capacities and the exponential distribution.

And the cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_0^x f(t)dt$$

$$= \int_0^x \frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} \exp\left(-\frac{t}{\beta}\right) dt \qquad (84)$$

The domain for this gamma variable is: $x \in (L_b = 0, U_b = 14]$ and I have taken $dx = 0.01$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \int_0^x \frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} \exp\left(-\frac{t}{\beta}\right) dt$$

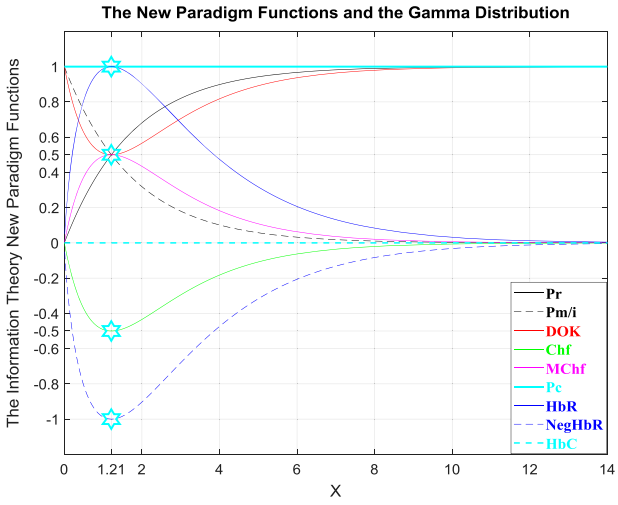


Figure 62. The new paradigm functions and the gamma distribution.

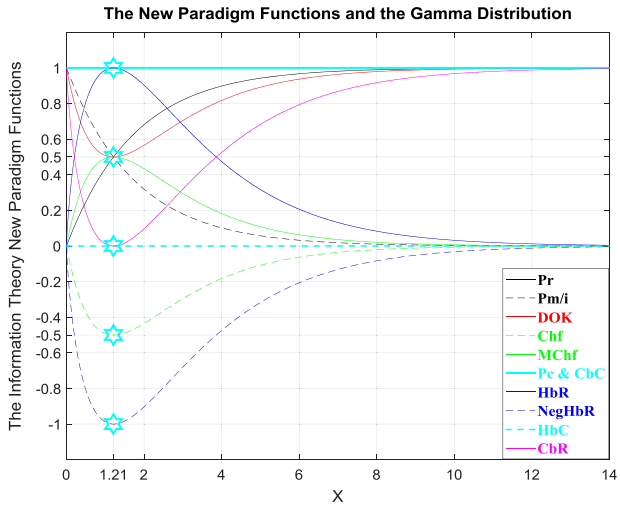


Figure 63. The new paradigm functions with the BSC capacities and the gamma distribution.

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x)$$
$$= 1 - \int_0^x \frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} \exp\left(-\frac{t}{\beta}\right) dt$$
$$= \int_x^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} t^{\alpha-1} \exp\left(-\frac{t}{\beta}\right) dt$$

In the simulations, the mean of this gamma random distribution is $\mu = \alpha\beta = 1.75$.

The standard deviation is $\sigma = \sqrt{\alpha\beta^2} = 1.75$.

There is no simple closed form for the median *Md* of the gamma distribution. Hence, from the simulations we have $Md = 1.21$.

A graph for the surprisal and expectancy self-information functions for this distribution can be drawn that is similar to the previous graphs for other probability distributions. The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 62 and 63).

10.2.5. The beta probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \text{for } 0 \le x \le 1 \tag{85}$$

Where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and the shape parameters $\alpha, \beta > 0$. $\alpha = 3$ and $\beta = 1.2$ here.

The cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_0^x f(t)dt$$
$$= \int_0^x \frac{1}{B(\alpha,\beta)} t^{\alpha-1}(1-t)^{\beta-1} dt \tag{86}$$

The domain for this beta variable is: $x \in [L_b = 0, U_b = 1]$ and I have taken $dx = 0.0001$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \int_0^x \frac{1}{B(\alpha,\beta)} t^{\alpha-1}(1-t)^{\beta-1} dt$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x)$$
$$= 1 - \int_0^x \frac{1}{B(\alpha,\beta)} t^{\alpha-1}(1-t)^{\beta-1} dt$$
$$= \int_x^1 \frac{1}{B(\alpha,\beta)} t^{\alpha-1}(1-t)^{\beta-1} dt$$

In the simulations, the mean of this beta distribution is $\mu = \frac{\alpha}{\alpha+\beta} = 3/1.2 = 2.5$.

The standard deviation is $\sigma = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = 0.198107$.

The median *Md* of the beta distribution is $\approx \frac{\alpha-1/3}{\alpha+\beta-2/3} = 0.75035$.

The rescaled surprisal and expectancy self-information functions with the simulation rescaling factor $\Phi = 39.4$ are:

$$RI_2(x = Md) = \bar{RI}_2(x = Md) = -\log_2[p(0.75035)]/\Phi$$
$$= -\log_2(0.5)/39.4 = 1/39.4 = 0.02538 \text{ bits}$$

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 64–66).

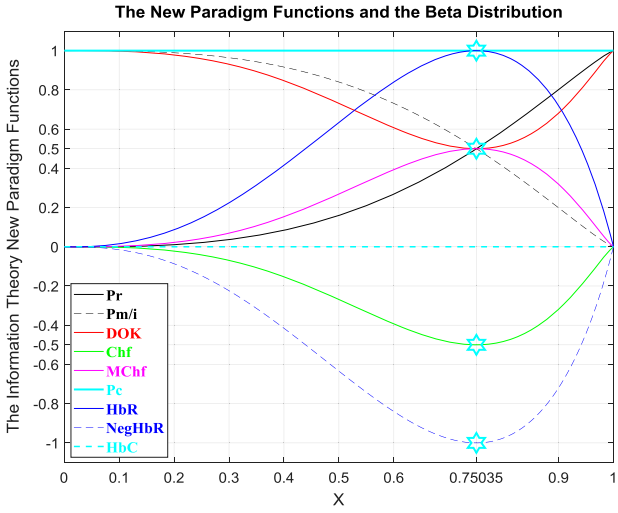


Figure 64. The new paradigm functions and the beta distribution.

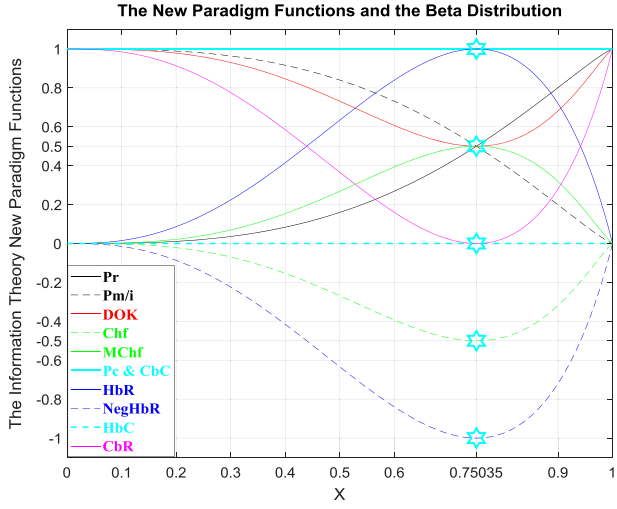


Figure 66. The new paradigm functions with the BSC capacities and the beta distribution.

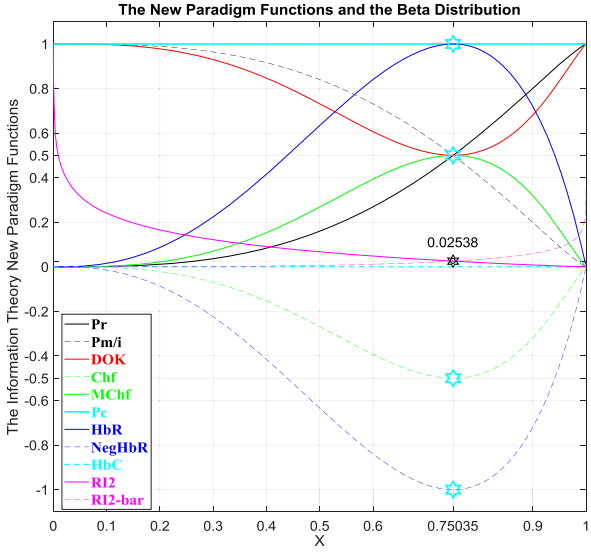


Figure 65. The new paradigm functions with the rescaled surprisal and expectancy and the beta distribution.

10.2.6. The chi2 probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \frac{x^{(\nu-2)/2}}{2^{\nu/2}\Gamma(\nu/2)} \exp\left(-\frac{x}{2}\right), \text{for } 0 \le x < \infty \tag{87}$$

Where $\nu > 0$ is the number of the degree of freedom. $\nu = 4$ here.

The cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_0^x f(t)dt = \int_0^x \frac{t^{(\nu-2)/2}}{2^{\nu/2}\Gamma(\nu/2)} \exp\left(-\frac{t}{2}\right) dt \tag{88}$$

The domain for this Chi2 variable is: $x \in [L_b = 0, U_b = 20]$ and I have taken $dx = 0.001$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \int_0^x \frac{t^{(\nu-2)/2}}{2^{\nu/2}\Gamma(\nu/2)} \exp\left(-\frac{t}{2}\right) dt$$

The complementary probability $P_m(x)/i$ is:

$$\begin{aligned} P_m(x)/i &= 1 - P_r(x) = 1 - F(x) \\ &= 1 - \int_0^x \frac{t^{(\nu-2)/2}}{2^{\nu/2}\Gamma(\nu/2)} \exp\left(-\frac{t}{2}\right) dt \\ &= \int_x^\infty \frac{t^{(\nu-2)/2}}{2^{\nu/2}\Gamma(\nu/2)} \exp\left(-\frac{t}{2}\right) dt \end{aligned}$$

In the simulations, the mean of this Chi2 distribution is $\mu = \nu = 4$.

The standard deviation is $\sigma = \sqrt{2\nu} = 2.828427$.

The median *Md* of the Chi2 distribution is $\approx \nu[1 - 2/(9\nu)]^3 = 3.369684$.

A graph for the surprisal and expectancy self-information functions for this distribution can be drawn that is similar to the previous graphs for other probability distributions.

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 67 and 68).

10.2.7. The F probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \frac{\Gamma\left[\frac{(\nu_1+\nu_2)}{2}\right]}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{x^{\frac{\nu_1-2}{2}}}{\left[1 + \left(\frac{\nu_1}{\nu_2}\right)x\right]^{\frac{\nu_1+\nu_2}{2}}},$$

$$\text{for } 0 \le x < \infty \tag{89}$$

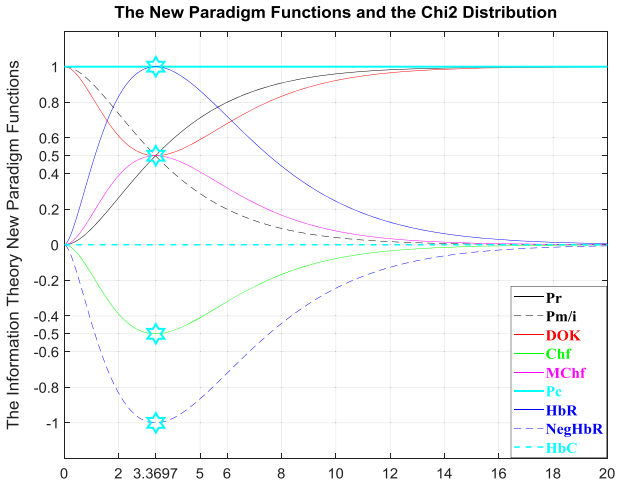


Figure 67. The new paradigm functions and the Chi2 distribution.

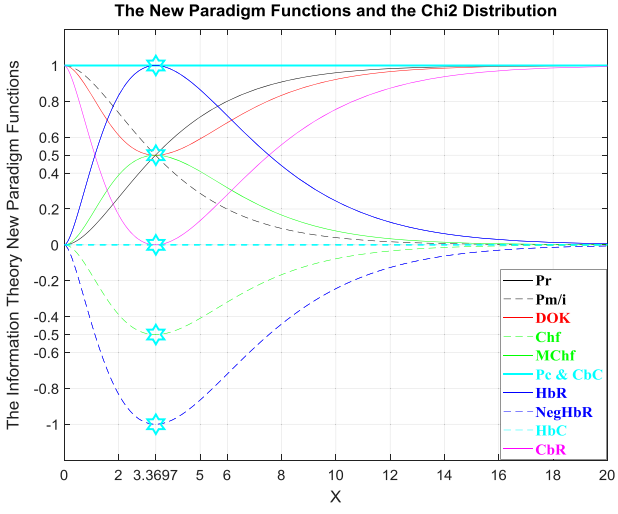


Figure 68. The new paradigm functions with the BSC capacities and the Chi2 distribution.

where $\nu_1, \nu_2 > 0$ are the numbers of the degrees of freedom. $\nu_1 = 7$ and $\nu_2 = 18$ here. $\Gamma(\cdot)$ is the gamma function.

And the cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_0^x f(t)dt = \int_0^x \frac{\Gamma\left[\frac{(\nu_1+\nu_2)}{2}\right]}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \times \frac{t^{\frac{\nu_1-2}{2}}}{\left[1+\left(\frac{\nu_1}{\nu_2}\right)t\right]^{\frac{\nu_1+\nu_2}{2}}}dt \quad (90)$$

The domain for this F variable is: $x \in [L_b = 0, U_b = 7]$ and I have taken $dx = 0.001$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \int_0^x \frac{\Gamma\left[\frac{(\nu_1+\nu_2)}{2}\right]}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \times \frac{t^{\frac{\nu_1-2}{2}}}{\left[1+\left(\frac{\nu_1}{\nu_2}\right)t\right]^{\frac{\nu_1+\nu_2}{2}}}dt$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x) = 1 - \int_0^x \frac{\Gamma\left[\frac{(\nu_1+\nu_2)}{2}\right]}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{t^{\frac{\nu_1-2}{2}}}{\left[1+\left(\frac{\nu_1}{\nu_2}\right)t\right]^{\frac{\nu_1+\nu_2}{2}}}dt = \int_x^\infty \frac{\Gamma\left[\frac{(\nu_1+\nu_2)}{2}\right]}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{t^{\frac{\nu_1-2}{2}}}{\left[1+\left(\frac{\nu_1}{\nu_2}\right)t\right]^{\frac{\nu_1+\nu_2}{2}}}dt$$

In the simulations, the mean of this F distribution is $\mu = \nu_2/(\nu_2 - 2) = 1.125$.

The standard deviation is $\sigma = 0.770759$.

The median *Md* of the F distribution is ≈ 0.942.

A graph for the surprisal and expectancy self-information functions for this distribution can be drawn that is similar to the previous graphs for other probability distributions. The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 69 and 70).

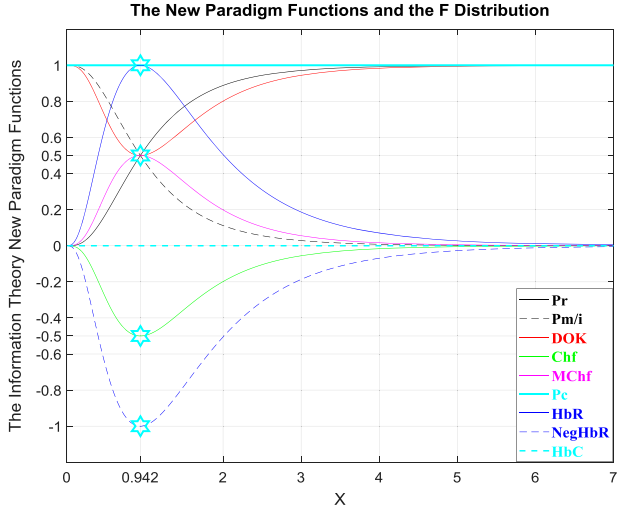


Figure 69. The new paradigm functions and the F distribution.

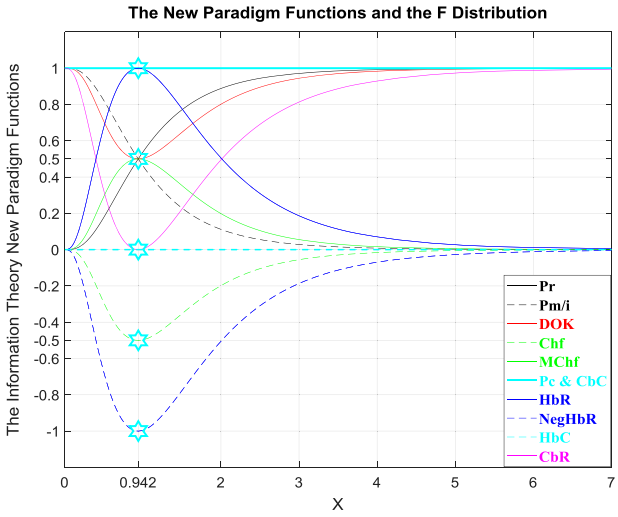


Figure 70. The new paradigm functions with the BSC capacities and the F distribution.

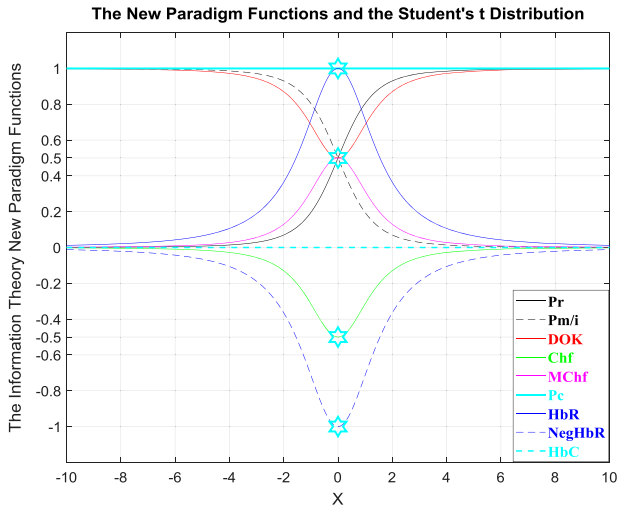


Figure 71. The new paradigm functions and the Student's t distribution.

10.2.8. The student's t probability distribution

The probability density function (*PDF*) of this continuous distribution is:

$$f(x) = \frac{dF(x)}{dx} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}},$$

$$\text{for } -\infty < x < \infty \tag{91}$$

Where $\nu > 0$ is the number of the degrees of freedom. $\nu = 3$ here. $\Gamma(\cdot)$ is the gamma function.

The cumulative distribution function (*CDF*) is:

$$F(x) = P_{rob}(X \le x) = \int_{-\infty}^{x} f(t)dt$$

$$= \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} dt \tag{92}$$

The domain for this student's t variable is: $x \in [L_b = -10, U_b = 10]$ and I have taken $dx = 0.001$.

The real probability $P_r(x)$ is:

$$P_r(x) = F(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} dt$$

The complementary probability $P_m(x)/i$ is:

$$P_m(x)/i = 1 - P_r(x) = 1 - F(x) = 1$$

$$- \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} dt$$

$$= \int_{x}^{\infty} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} dt$$

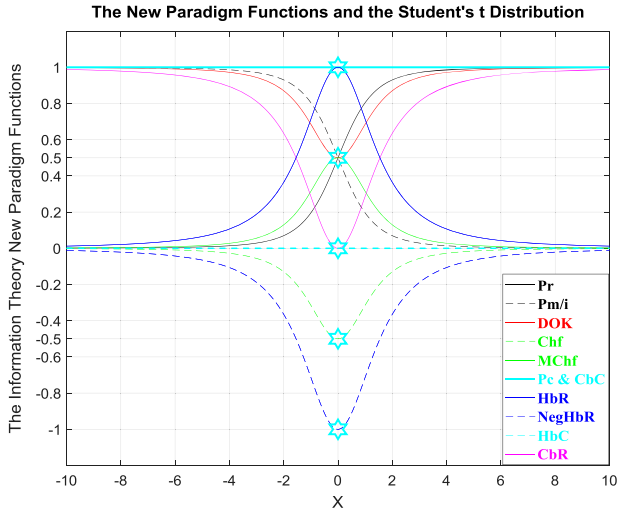


Figure 72. The new paradigm functions with the BSC capacities and the Student's t distribution.

In the simulations, the mean of this t distribution is $\mu = 0$.

The standard deviation is $\sigma = \sqrt{\nu/(\nu-2)} = 1.73205$.

The median *Md* of the t distribution is $= 0$.

A graph for the surprisal and expectancy self-information functions for this distribution can be drawn that is similar to the previous graphs for other probability distributions.

The other parameters are calculated from the *CPP* paradigm (refer to section VIII) (Figures 71 and 72).

11. Final analysis

In the complex set $\mathscr{C}$ we have the entropy always equal to 0, so no loss no gain but complete conservation of

Table 1. The complex probability paradigm prognostic functions for L_b (Lower bound) $\leq$ (Message x) $\leq U_b$ (Upper bound).

For Any Random Distribution	*DOK*	*Chf*	*MChf*	P_m/i	*Z*	*Pc*	RI_2	$\bar{RI}_2$
$x = L_b \Rightarrow P_r = 0$	$= 1$	$= 0$	$= 0$	$= 1$	$= i$	$= 1$	$= 1$	$= 0$
$0 < P_r < 0.5$	$\downarrow$	$\downarrow$	$\uparrow$	$\downarrow$	$\mathrm{Re}(Z)\uparrow$	$= 1$	$\downarrow$	$\uparrow$
$x \uparrow \Rightarrow P_r \uparrow$					$\mathrm{Im}(Z)\downarrow$			
$x =$ Median	$=$ Min	$=$ Min	$=$ Max	$= +0.5$	$= 0.5 + 0.5i$	$= 1$	$= \frac{1}{\Phi}$	$= \frac{1}{\Phi}$
$P_r = 0.5$	$= +0.5$	$= -0.5$	$= +0.5$					
$0.5 < P_r < 1$	$\uparrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\mathrm{Re}(Z)\uparrow$	$= 1$	$\downarrow$	$\uparrow$
$x \uparrow \Rightarrow P_r \uparrow$					$\mathrm{Im}(Z)\downarrow$			
$x = U_b \Rightarrow P_r = 1$	$= 1$	$= 0$	$= 0$	$= 0$	$= 1$	$= 1$	$= 0$	$= 1$

information. The Lavoisier principle in chemistry and science affirms that mass and energy are conserved. The Law of Conservation of Mass (or Matter) in a chemical reaction can be stated thus: In a chemical reaction, matter is neither created nor destroyed. Knowing that it was discovered by Antoine Laurent Lavoisier (1743–94) about 1785. Therefore, it applies also to information theory.

Moreover, in $\mathscr{M}$ we have parallel planes and parallel similar curves for entropy and channel capacity. In $\mathscr{R}$, we have disorder, uncertainty, and unpredictability. In $\mathscr{C}$ we have order, certainty, and predictability since $Pc = 1$ permanently and entropy $= 0$ constantly. Additionally, in $\mathscr{R}$ we have chaos and imperfect and incomplete knowledge or partial ignorance. In $\mathscr{C}$ we have chaos always equal to 0 and DOK $= 1$ continuously, thus complete and perfect and total knowledge of the random message and channel.

Furthermore, the extension of all random and nondeterministic phenomena in $\mathscr{R}$ to the set $\mathscr{C}$ leads to certain knowledge and sure events since DOK $= 1$ and $Pc = 1$. Consequently, no randomness exists in $\mathscr{C}$ and all phenomena are deterministic in this set. Therefore, in $\mathscr{C}$ prognostic is assured and definite.

Table 1 summarises the complex probability paradigm prognostic functions for any probability distribution ($\uparrow$ = increases and $\downarrow$ = decreases).

Table 2 summarises the complex probability paradigm prognostic entropies for any probability distribution.

Table 3 summarises the complex probability paradigm prognostic BSC capacities for any probability distribution.

Accordingly, at each instant in the novel prognostic model, the random entropy and channel capacity are certainly predicted in the complex set $\mathscr{C}$ with $Pc^2 = DOK - Chf = DOK + MChf$ maintained as equal to one through a continuous compensation between *DOK* and *Chf*. This compensation is from the instant $x = L_b$ until the instant $x = U_b$. We can understand also that *DOK* is the measure of our certain knowledge (100% probability) about the expected event, it does not include any uncertain knowledge (with a probability less than 100%). We can see that

Table 2. The complex probability paradigm prognostic entropies for L_b (Lower bound) $\leq$ (Message x) $\leq U_b$ (Upper bound).

For Any Random Distribution	H_b^R	$\bar{H}_b^R$	$NegH_b^R$	H_b^M	H_b^C
$x = L_b \Rightarrow P_r = 0$	$= 0$	$= 0$	$= 0$	$= \psi$	$= 0$
$0 < P_r < 0.5$	$\uparrow$	$\uparrow$	$\downarrow$	$\mathrm{Re}(H_b^M) = \psi$	$= 0$
$x \uparrow \Rightarrow P_r \uparrow$				$\mathrm{Im}(H_b^M) \uparrow$	
$x =$ Median	$=$ Max	$=$ Max	$=$ Min	$=$ Max	$= 0$
$P_r = 0.5$	$= +1$	$= +1$	$= -1$	$= \psi + i$	
$0.5 < P_r < 1$	$\downarrow$	$\downarrow$	$\uparrow$	$\mathrm{Re}(H_b^M) = \psi$	$= 0$
$x \uparrow \Rightarrow P_r \uparrow$				$\mathrm{Im}(H_b^M) \downarrow$	
$x = U_b \Rightarrow P_r = 1$	$= 0$	$= 0$	$= 0$	$= \psi$	$= 0$

Table 3. The complex probability paradigm prognostic BSC capacities for L_b (Lower bound) $\leq$ (Message x) $\leq U_b$ (Upper bound).

For Any Random Distribution	C_{BSC}^R	$\bar{C}_{BSC}^R$	C_{BSC}^M	C_{BSC}^C
$x = L_b \Rightarrow P_r = 0$	$= 1$	$= 1$	$= -\psi + i$	$= 1$
$0 < P_r < 0.5$	$\downarrow$	$\downarrow$	$\mathrm{Re}(C_{BSC}^M) = -\psi$	$= 1$
$x \uparrow \Rightarrow P_r \uparrow$			$\mathrm{Im}(C_{BSC}^M) \downarrow$	
$x =$ Median	$=$ Min	$=$ Min	$=$ Min	$= 1$
$P_r = 0.5$	$= 0$	$= 0$	$= -\psi$	
$0.5 < P_r < 1$	$\uparrow$	$\uparrow$	$\mathrm{Re}(C_{BSC}^M) = -\psi$	$= 1$
$x \uparrow \Rightarrow P_r \uparrow$			$\mathrm{Im}(C_{BSC}^M) \uparrow$	
$x = U_b \Rightarrow P_r = 1$	$= 1$	$= 1$	$= -\psi + i$	$= 1$

in computing Pc^2 we have eliminated and subtracted in the equation above all the random factors and chaos (*Chf*) from our random experiment, hence no chaos exists in $\mathscr{C}$, it only exists (if it does) in $\mathscr{R}$; therefore, this has yielded a 100% deterministic experiment and outcome in $\mathscr{C}$ since the probability *Pc* is continuously equal to 1. This is one of the advantages of extending $\mathscr{R}$ to $\mathscr{M}$ and hence of working in $\mathscr{C} = \mathscr{R} + \mathscr{M}$. Hence, in the novel prognostic model, our knowledge of all the parameters and indicators (I_2, $\bar{I}_2$, H_b, C, etc ...) is always perfect, constantly complete, and totally predictable since $Pc = 1$

permanently, independently of any probability profile or random factors.

12. Conclusion and perspectives

In the current paper we applied and linked the theory of Extended Kolmogorov Axioms to Claude Shannon's information theory. Hence, a tight bond between the new paradigm and quantities of information, entropies, and channel capacities was established. Thus, the theory of 'Complex Probability' was developed beyond the scope of my previous nine papers on this topic.

Moreover, as it was proved and illustrated in the new model, when $x = L_b$ or $x = U_b$ then the degree of our knowledge (*DOK*) is one and the chaotic factor (*Chf* and *MChf*) is 0 since the state of the random message and channel is totally known. During the process of message transmission [$L_b <$ (Message x) $< U_b$] we have: $0.5 < DOK < 1$, $-0.5 < Chf < 0$, and $0 < MChf < 0.5$. Notice that during this whole process we have always $Pc^2 = DOK - Chf = DOK + MChf = 1$, that means that the phenomenon which seems to be random and stochastic in $\mathscr{R}$ is now deterministic and certain in $\mathscr{C} = \mathscr{R} + \mathscr{M}$, and this after adding to $\mathscr{R}$ the contributions of $\mathscr{M}$ and hence after subtracting the chaotic factor from the degree of our knowledge. Furthermore, the probabilities of the message flips corresponding to each instance of x have been determined in the probability sets $\mathscr{R}$, $\mathscr{M}$, and $\mathscr{C}$ byP_r, P_m, and Pc respectively. Therefore, at each instance of x, the information theory parameters I_2, $\bar{I}_2$, H_b, C, etc ... are surely predicted in the complex set $\mathscr{C}$ with Pc maintained as equal to 1 permanently. Furthermore, using all these illustrated graphs and simulations throughout the whole paper, we can visualise and quantify both the system chaos (*Chf* and *MChf*) and the certain knowledge (*DOK* and *Pc*) of the information theory model. This is certainly very interesting and fruitful and shows once again the benefits of extending Kolmogorov's axioms and thus the originality and usefulness of this new field in applied mathematics and prognostic that can be called verily: 'The Complex Probability Paradigm'.

It is important to mention in the conclusion that a few important and well-known probability distributions were considered in the current research paper although the original CPP model can be applied to any random distribution. This will lead to similar results and conclusions and proves the success of my novel paradigm.

As a prospective and future work, it is planned to more develop the novel proposed prognostic paradigm and to apply it to a wide set of stochastic and random systems like the analytic prognostic of vehicle suspensions systems and of petrochemical pipelines (in their three modes: unburied, buried, and offshore) under the linear and nonlinear damage accumulation cases.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abou Jaoude, A. (2004). *Ph.D. thesis in applied mathematics: Numerical methods and algorithms for applied mathematicians*. Bircham International University. Retrieved August 1, 2004, from http://www.bircham.edu.

Abou Jaoude, A. (2005). *Ph.D. thesis in computer science: Computer simulation of Monte Carlo methods and random phenomena*. Bircham International University. Retrieved October 14, 2005, from http://www.bircham.edu

Abou Jaoude, A. (2007). *Ph.D. thesis in applied statistics and probability: Analysis and algorithms for the statistical and stochastic paradigm*. Bircham International University. Retrieved April 27, 2007, from http://www.bircham.edu

Abou Jaoude, A. (2013a). *The complex statistics paradigm and the law of large numbers*. *Journal of Mathematics and Statistics (JMSS), Science Publications*, *9*(4), 289–304.

Abou Jaoude, A. (2013b). The theory of complex probability and the first order reliability method. *Journal of Mathematics and Statistics, Science Publications*, *9*(4), 310–324.

Abou Jaoude, A. (2014). *Complex probability theory and prognostic*. *Journal of Mathematics and Statistics (JMSS), Science Publications*, *10*(1), 1–24.

Abou Jaoude, A. (2015a). The complex probability paradigm and analytic linear prognostic for vehicle suspension systems. *American Journal of Engineering and Applied Sciences*, *8*(1), 147–175.

Abou Jaoude, A. (2015b). *The paradigm of complex probability and the Brownian motion*. *Systems Science & Control Engineering*, *3*(1), 478–503.

Abou Jaoude, A. (2016a). *The paradigm of complex probability and analytic nonlinear prognostic for vehicle suspension systems*. *Systems Science & Control Engineering*, *4*(1), 334–378.

Abou Jaoude, A. (2016b). *The paradigm of complex probability and Chebyshev's inequality*. *Systems Science & Control Engineering*, *4*(1), 99–137.

Abou Jaoude, A. (2017). *The paradigm of complex probability and analytic linear prognostic for unburied petrochemical pipelines*. *Systems Science & Control Engineering*, *5*(1), 178–214.

Abou Jaoude, A., El-Tawil, K., & Kadry, S. (2010). *Prediction in complex dimension using Kolmogorov's set of axioms*. *Journal of Mathematics and Statistics (JMSS), Science Publications*, *6*(2), 116–124.

Abrams, W. (2008). *A brief history of probability*, second moment. Retrieved May 23, 2008, from http://www.secondmoment.org/articles/probability.php

Aczel, A. (2000). *God's equation*. New York: Dell Publishing.

Allikmets, R., Wasserman, W. W., Hutchinson, A., Smallwood, P., Nathans, J., Rogan, P. K., ... Dean, M. (1998). Organization of the ABCR gene: Analysis of promoter and splice junction sequences. *Gene*, *215*(1), 111–122.

Arndt, C. (2004). *Information measures, information and its description in science and engineering*. Springer Series: Signals and Communication Technology.

Ash, R. B. (1990 [1965]). *Information theory*. Dover Publications.
Ash, R. B. ([1965] 1990). *Information theory*. New York: Interscience. New York: Dover.
Balibar, F. (2002). *Albert Einstein: Physique, philosophie, politique* (1st ed.). Paris: Le Seuil.
Barrow, J. (1992). *Pi in the sky*. London: Oxford University Press.
Bell, E. T. (1992). *The development of mathematics*. New York: Dover Publications.
Bennett, C. H., Li, M., & Ma, B. (2003). Chain letters and evolutionary histories. *Scientific American*, *288*(6), 76–81.
Benton, W. (1966a). *Probability*, *encyclopedia Britannica* (Vol. 18). Chicago: Encyclopedia Britannica.
Benton, W. (1966b). *Mathematical probability*, *encyclopedia Britannica* (Vol. 18). Chicago: Encyclopedia Britannica.
Bernstein, P. L. (1996). *Against the Gods: The remarkable story of risk*. New York. Wiley.
Bidabad, B. (1992). *Complex probability and Markov Stochastic processes, Proc. First Iranian Statistics Conference, Tehran*. Isfahan University of Technology.
Bogdanov, I., & Bogdanov, G. (2009). *Au Commencement du Temps*. Paris: Flammarion.
Bogdanov, I., & Bogdanov, G. (2010). *Le Visage de Dieu*. Paris: Editions Grasset et Fasquelle.
Bogdanov, I., & Bogdanov, G. (2012). *La Pensée de Dieu*. Paris: Editions Grasset et Fasquelle.
Bogdanov, I., & Bogdanov, G. (2013). *La Fin du Hasard*. Paris: Editions Grasset et Fasquelle.
Boltzmann, L. (1995). *Lectures on gas theory*. New York: Dover.
Boursin, J.-L. (1986). *Les Structures du Hasard*. Paris: Editions du Seuil.
Brillouin, L. ([1956, 1962] 2004). *Science and information theory*. Mineola, NY: Dover.
Burgi, M. (2010). *Interpretations of negative probabilities*. Retrieved from https://arxiv.org/abs/1008.1287
Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer Science.
Campbell, J. (1982). *Grammatical man*. Touchstone/Simon & Schuster.
Cercignani, C. (2010). *Ludwig Boltzmann, the man who trusted atoms*. Oxford: Oxford University Press.
Chan Man Fong, C. F., De Kee, D., & Kaloni, P. N. (1997). *Advanced mathematics for applied and pure sciences*. Amsterdam: Gordon and Breach Science Publishers.
Cover, T., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). New York: Wiley-Interscience.
Cox, D. R. (1955). A use of complex probabilities in the theory of stochastic processes. *Mathematical Proceedings of the Cambridge Philosophical Society*, *51*, 313–319.
Csiszar, I., & Korner, J. (1997). *Information theory: Coding theorems for discrete memoryless systems* (2nd ed). Akademiai Kiado.
Dacunha-Castelle, D. (1996). *Chemins de l'Aléatoire*. Paris: Flammarion.
Dalmedico-Dahan, A., Chabert, J.-L., & Chemla, K. (1992). *Chaos Et Déterminisme*. Paris: Edition du Seuil.
Dalmedico-Dahan, A., & Peiffer, J. (1986). *Une Histoire des Mathematiques*. Paris: Edition du Seuil.
Daston, L. (1988). *Classical probability in the enlightenment*. Princeton: Princeton University Press.
David, F. N. (1962). *Games, Gods and gambling: The origins and history of probability and statistical ideas from the earliest times to the Newtonian Era*. London: Charles Griffin Co. Ltd.
David, R., & Anderson, D. R. (2003, November 1). *Some background on why people in the empirical sciences may want to better understand the information-theoretic methods*. Retrieved June 23, 2010, from https://www.jyu.fi/bioenv/en/divisions/biosciences/eko/coeevolution/events/itms/why.
Davies, P. (1993). *The mind of God*. London: Penguin Books.
Ekeland, I. (1991). *Au Hasard. La Chance, la Science et le Monde*. Paris: Editions du Seuil.
Escolano, R., Francisco, S. P., & Pablo, B. (2009). *Information theory in computer vision and pattern recognition*. Springer.
Fagin, R., Halpern, J., & Megiddo, N. (1990). A logic for reasoning about probabilities. *Information and Computation*, *87*, 78–128.
Fazlollah, M. R. (1994 [1961]). *An introduction to information theory*. New York: Dover Publications.
Feller, W. (1968). *An introduction to probability theory and its applications* (3rd ed.). New York: Wiley.
Franklin, J. (2001). *The science of conjecture: Evidence and probability before Pascal*. Baltimore: Johns Hopkins University Press.
Freund, J. E. (1973). *Introduction to probability*. New York: Dover Publications.
Gallager, R. (1968). *Information theory and reliable communication*. New York: John Wiley and Sons.
Gibson, J. D. (1998). *Digital compression for multimedia: Principles and standards*. Morgan Kaufmann.
Gleick, J. (1997). *Chaos, making a new science*. New York: Penguin Books.
Gleick, J. (2011). *The information: A history, a theory, a flood*. New York: Pantheon.
Goldman, S. (1968). *Information theory*. New York: Prentice Hall/Dove.
Gorrochum, P. (2012). Some laws and problems in classical probability and how Cardano anticipated them. *Chance Magazine*. Retrieved from http://chance.amstat.org/
Greene, B. (2003). *The elegant universe*. New York: Vintage.
Gullberg, J. (1997). *Mathematics from the birth of numbers*. New York: W.W. Norton & Company.
Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. New York: Cambridge University Press.
Haggerty, P. (1981). The corporation and innovation. *Strategic Management Journal*, *2*, 97–118.
Hald, A. (1998). *A history of mathematical statistics from 1750 to 1930*. New York. Wiley.
Hald, A. (2003). *A history of probability and statistics and their applications before 1750*. Hoboken, NJ: Wiley.
Hartley, R. V. L. (1928, July). Transmission of information. *Bell System Technical Journal*.
Hawking, S. (2002). *On the shoulders of giants*. London: Running Press.
Hawking, S. (2005). *God created the integers*. London: Penguin Books.
Hawking, S. (2011). *The dreams that stuff is made of*. London: Running Press.
Heyde, C. C., & Seneta, E. (2001). *Statisticians of the centuries*. New York. Springer.
Hoffmann, B. (1975). *In collaboration with Helen Dukas, Albert Einstein, Créateur et Rebelle* (1st ed). Paris: Editions du Seuil.
Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, *294*, 2310–2314.

Ivancevic, V. G., & Ivancevic, T. T. (2008). *Quantum leap: From Dirac and Feynman, across the universe, to human body and mind*. Singapore: World Scientific.
Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620–630.
Jeffrey, R. (1992). *Probability and the art of judgment*. Cambridge: Cambridge University Press.
Kelly Jr, J. L. (1956, July). Saratoga.ny.usA. New interpretation of information rate. *Bell System Technical Journal*, *35*, 917–926.
Khinchin, A. I. (1957). *Mathematical foundations of information theory*. New York: Dover.
Kolmogorov, A. (1968). Three approaches to the quantitative definition of information. *International Journal of Computer Mathematics*, *1*, 3–7.
Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: Chicago Press.
Landauer, R. (1961). IBM.com, "*Irreversibility and Heat Generation in the Computing Process*". *IBM Journal of Research and Development*, *5*(3), 183–191.
Landauer, R. (1993). IEEE.org, "*Information is Physical*" Proc. Workshop on Physics and Computation PhysComp'92 (IEEE Comp. Sci.Press, Los Alamitos) pp. 1–4.
Leff, H. S., & Rex, A. F. (Eds.). (1990). *Maxwell's demon: Entropy, information, computing*. Princeton, NJ: Princeton University Press.
Logan, R. K. (2014). *What is information? - propagating organization in the biosphere, the symbolosphere, the technosphere and the econosphere*. Toronto: DEMO Publishing.
MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.
Mansuripur, M. (1987). *Introduction to information theory*. New York: Prentice Hall.
McEliece, R. (2002). *The theory of information and coding*. Cambridge.
Mcgrayne, S. B. (1990). *The theory that would not die: How Bayes" rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. New Haven. Yale University Press.
Montgomery, D. C., & Runger, G. C. (2003). *Applied statistics and probability for engineers* (3rd ed.). New York: John Wiley & Sons.
Moore, W. J. (1992). *Schrödinger: Life and thought*. Cambridge. Cambridge University Press.
Noth, W. (1981). Semiotica. *Semiotics of Ideology*, (148).
Ognjanović, Z., Marković, Z., Rašković, M., Doder, D., & Perović, A. (2012). A probabilistic temporal logic that can model reasoning about evidence. *Annals of Mathematics and Artificial Intelligence*, *65*, 1–24.
Penrose, R. (1999). *Traduction Francaise: Les Deux Infinis et L'Esprit Humain*. Roland Omnès: Flammarion.
Pickover, C. (2008). *Archimedes to Hawking*. Oxford: Oxford University Press.
Pierce, J. R. (1961). *An introduction to information theory: Symbols, signals and noise* (2nd ed.). Dover. Reprinted by Dover 1980.
Planck, M. (1969). *Treatise on thermodynamics*. New York: Dover.
Poincaré, H. (1968). *La Science et l'Hypothèse* (1st ed.). Paris: Flammarion.
Reeves, H. (1988). *Patience dans L'Azur, L'Evolution Cosmique*. Paris: Le Seuil.
Retrieved from http://www.statslab.cam.ac.uk/ ~ rrw1/markov/M.pdf
Reza, F. (1961). *An introduction to information theory*. New York: McGraw-Hill. New York: Dover.
Rieke, F., Warland, D., van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: The MIT Press.
Ronan, C. (1988). *Traduction Française: Histoire Mondiale des Sciences*. Paris: Claude Bonnafont, Le Seuil.
Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*.
Science Et Vie. (1999). *Le Mystère des Mathématiques*. Numéro 984.
Seife, C. (2006). *Decoding the universe*. Viking.
Seneta, E. W. (2016). *"Adrien-Marie Legendre"* (*version* 9). StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies.
Shannon, C. E. (1948, July & October). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423 & 623–656.
Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication (PDF)*. Urbana, IL: University of Illinois Press.
Siegfried, T. (2000). *The bit and the pendulum*. Wiley.
Srinivasan, S. K., & Mehata, K. M. (1988). *Stochastic processes* (2nd ed.). New Delhi: McGraw-Hill.
Stepić, A. I., & Ognjanović, Z. (2014). Complex valued probability logics. Publications De l'institut Mathématique. *Nouvelle Série, Tome*, *95*(109), 73–86. doi:10.2298/PIM1409073I
Stewart, I. (1996). *From here to infinity* (2nd ed.). Oxford: Oxford University Press.
Stewart, I. (2002). *Does God play dice?* (2nd ed.). Oxford: Blackwell Publishing.
Stewart, I. (2012). *In pursuit of the unknown*. Oxford: Basic Books.
Stigler, S. M., *The history of statistics: The measurement of uncertainty before 1900*. Belknap Press/Harvard University Press.
Stone, J. V. (2014). *Chapter 1 of book "information theory: A tutorial Introduction"*. University of Sheffield.
Theil, H. (1967). *Economics and information theory*. Chicago: Rand McNally & Company.
Timme, N., Alford, W., Flecker, B., & Beggs, J. M. (2012). *Multivariate information measures: An experimentalist's perspective*. Retrieved from https://arxiv.org/abs/1111.6857
Van Kampen, N. G. (2006). *Stochastic processes in physics and chemistry. Revised and enlarged edition*. Sydney: Elsevier.
Vitanyi, P. M. B. (1988).*"Andrei Nikolaevich Kolmogorov"*. *CWI Quarterly*, *1*, 3–18. Retrieved January 27, 2016, from http://homepages.cwi.nl/ paulv/KOLMOGOROV.BIOGRAPHY.html
Von Plato, J. (1994). *Creating modern probability: Its mathematics, physics and philosophy in historical perspective*. New York: Cambridge University Press.
Walpole, R., Myers, R.,Myers, S., & Ye, K. (2002). *Probability and statistics for engineers and scientists* (7th ed.). Prentice, NJ: Prentice Hall.
Warusfel, A., & Ducrocq, A. (2004). *Les Mathématiques, Plaisir et Nécessité* (1st ed). Paris: Edition Vuibert.
Wei, Y., Qiu, J., & Fu, S. (2015). Mode-dependent nonrational output feedback control for continuous-time semi-Markovian jump systems with time-varying delay. *Nonlinear Analysis: Hybrid Systems*, *16*, 52–71.
Wei, Y., Qiu, J., Karimi, H. R., & Wang, M. (2014). H_∞ model reduction for continuous-time Markovian jump systems with incomplete statistics of mode information. *International Journal of Systems Science*, *45*(7), 1496–1507.

Weingarten, D. (2002). Complex probabilities on R^N as real probabilities on C^N and an application to path integrals. *Physical Review Letters*, *89*. doi:10.1103/PhysRevLett.89.240201

Wikipedia, the free encyclopedia. *Information theory*. Retrieved from https://en.wikipedia.org/

Wikipedia, the free encyclopedia. *Probability distribution*. Retrieved from https://en.wikipedia.org/

Wikipedia, the free encyclopedia. *Probability*. Retrieved from https://en.wikipedia.org/

Wikipedia, the free encyclopedia. *Probability theory*. Retrieved from https://en.wikipedia.org/

Yeung, R. W. (2002). *A first course in information theory*. Kluwer Academic/Plenum Publishers.

Yeung, R. W. (2008). *Information theory and network coding*. Springer.

Youssef, S. (1994). Quantum mechanics as complex probability theory. *Modern Physics Letters A*, *9*, 2571–2586.