Rastko Vuković

PHYSICAL INFORMATION

in social phenomena and mathematics

publisher

Economic Institute Banja Luka 2019

RS

May, 2019

publisher

Rough translation into English

Rastko Vuković
PHYSICAL INFORMATION – in social phenomena and mathematics

https://www.scribd.com/document/408301124/Physical-Information

Original in Serbian:

Rastko Vuković
FIZIČKA INFORMACIJA – u društvenim pojavama i matematici

https://www.scribd.com/document/406574702/

Lektori i recenzenti:

Duško Milinčić, informatika
Aleksandra Radić, fizika
Dragana Galić, matematika

Goran Dakić, Srpski jezik

CIP - Каталогизација у публикацији
Народна и универзитетска библиотека
Републике Српске, Бања Лука

519.72

ВУКОВИЋ, Растко
Физичка информација у друштвеним појавама и математици / Растко Вуковић. - Бања Лука : Економски институт, 2019 ([б.м. : б.и.]). - 103 стр. : илустр. ; 30 cm

Напомене и библиографске референце уз текст. - Библиографија: стр. 99-100. - Регистар.

ISBN 978-99938-854-6-7

COBISS.RS-ID 8404248

Foreword

This book consists of two clearly separated works. The first part of the first part are my columns in the daily local media[1] that go out on Thursdays. It is a portal for news from current practice, mostly of a political nature from the perspective of the opposition, whose editors somehow tolerated my texts which are somewhat inappropriate. In the following, I supplemented them with similar texts that should "dilute" the book and bring it closer to readers who deal with philosophy more easily than with mathematics.

The second part of the book is a more stringent consideration, definitions, attitudes and evidence in relation to physical information that I worked on a little earlier or in parallel with the above texts, and somewhat published on free sites like readgur.com or academia.edu. There is no need to mention those 5-10-page fragments, often untidy, uninteresting attempts and mistakes, but who during the elaboration of their thematic ideas did an important job for me. In the time before the discovery, such annoying search is always exciting.

I deliberately miss the third part. There I was planning to "paraphrase" some of my findings on power law, partly related to Barabási Scale-Free Networks research, and then from the Theory of Games, but the question is who would read it. Writing about nodes in nets and choosing options for example the winning, or a link between these two theories, would be equally mathematically dry as well as calculating the distributions that follow here, to deter even the few potential readers of this book. It may be a chance for such posts, but even if it would not, there is no harm.

R. Vuković, April 2019.

[1] http://izvor.ba/

Contents

Chapter 1

Society

Introduction

Physical information is a new theory. Semi-popular with a predominant reference to physics, I presented it in the book "Multiplicities" (see [3]), and before in some other my works that I do not mention here. It relies on classical information to follow intuitive assumptions and experimental results on the substance and its communication, while remaining close to Shannon's definition.

The first property of the *physical information* that does not support Shannon's definition of information is the *law of conservation*. The information is a specially defined amount of data that remains constant while the data are transformed. It is analogous to energy that goes from shape to shape, kinetic in potential or chemical or some third. Then there is a resistance to nature that I call the *principle of information*. More likely random events are more often realized but such are less informative, so we conclude that there is scrimp of information about it. This is the principle of minimalism. The third property is *uniqueness*, only mentioned here, because it mostly belongs to physics.

Otherwise, the first part of the book is the development of the consequences of applying the new theory of information to social phenomena. For the sake of simplicity, the terms of the physical information are not specified, mathematics is not used, and from the standpoint of truth this story becomes an impossible mission. A correctly interpreted theorem can only be a new theorem or a lemma or a corollary, and not a popular column. A small initial error of interpreting further steps easily evolves into drastically misconceptions, analogous to the *chaos theory* that deals with systems whose small changes in initial conditions escalate into large differences in the final results. Therefore, the chaotic development of falsehood can sometimes be corrected by new untruths, in order to compromise with populism.

Interpretations of freedom, truth, feminization, and others that I am writing here, that arose from my still unknown theories of physical information, are not exactly the same as what we intuitively implied under these notions, but they are almost similar. Why then do I publish it? First of all, because what we thought about the same things was rarely been consistently and often are suspicious. There is nothing more practical than good theory, and endless old debates about the terms redefined here did not have a good practical development. In a few hours about them would be said more "trues" than have all the Euclidean geometry theory, so we would become the lords of the galaxy by the production of so much knowledge, and we did not. Then, because by conquering the truth, the world is really progressing and will not wait for us.

1.1 Freedom

Have you ever wondered what is *freedom* without having to look for a political or philosophical dictionary? To take it simply, like "two plus two is four", without plunging into a deep epistemological or cognitive meaning? Freedom can be viewed and simply as something that has more when there are more possibilities.

We love freedom because we love what we want. Closed without the right to move or prevented in wanting to eat ice cream because the sweet hurts us or in poverty because we do not have money, freedom is deprived from us for the rule of law, our health, and bad economy. We do not like when we have no freedom, because "it never knows" what we would like to have, so we want it simply because we love to choose. Walking through shops, life among challenges, and harassment – all is freedom.

Have you noticed that precisely because it's a pleasant choice are hard? The muscle effort that works and the changes we make are the consequences of our choices, and ultimately, they are the result of our freedom. A deeper analysis of the "amount of possibilities" would show that it is in line with physical action, but I do not only think of this possible inconvenience of doing, but also the reluctance to deal with dilemmas.

Research will show that the brain does not like uncertainty, at least as much as it likes it. Here I am thinking of the feeling of anxiety about the problems of options, when it seems to us that there is no solution or that patience will abandon us before we find it and if we would like to let ourselves like a piece of wood down the water. We love order when we do not like opportunities.

We are looking for the law and the legal order because of the security they offer, and then because of the efficiency we hope for even more security. As the security, freedom is a dual phenomenon because we love it less when we have it more. We therefore like to organize ourselves because we do not like options, and we like options in order to better organize ourselves. Thus we arrive at the notion of freedom as something essential for development, and then the need for development in order to have security, and also efficiency, in order to have less freedom.

The phenomenon of freedom is inseparable from the occurrence of uncertainty, and then the real freedom does not exist without a fear of development, or, on the contrary, without the feeling of satisfaction in governing the possibilities. We admire the skill of solving the problem, as well as the ease of life, for the same reason. Because of the fear of freedom, that is, of the joy of being able to deal with it. These are the same emotions that put the state in dictatorship or development.

Freedom as the "amount of opportunity" necessarily leads to the conclusion that security is the opposite of freedom. Replacing a piece of freedom for a piece of security is a loss the both of them, because by depriving ourselves of freedom we reduce some dilemmas and become safe from the right to face life. Of what is the tiger safe in the cage?

Freedom is the right to options, to development, to the possibility of comfort, and without guarantees. It is such a stressful phenomenon!

http://izvor.ba/
March 7, 2019.

1.2 Liberalism

Liberalism is a movement of French revolutionaries (liberté, égalité, fraternité) in the 1800's for freedom and its protection by the rule of law. It was a battle cry for the liberation of the aristocracy of the pernicious potentials of equal and, on the horrors of the then monarchs, a call to the fraternal struggle against the hierarchy. We know that everyone has good and bad sides, but we have not dealt with the demons of liberalism yet.

One of the dark-light sides of liberalism will come to us from the knowledge that *equality* generates conflicts. The frenzy of sports competitions comes from fair rules, the sharper is the competition of equal, the greater the uncertainties of equal chances, and it turns out that nature does not like equality.

God does not build a long line, he can barely produce two equal snowflakes and he has never been able to think of something like Ford's production line, but he imagines so much! Yet, in much more difficult things, people have helped him, so the dilemma remains: will liberalism make a war with that absolute or will it only ignore the poor fellow? It's good for him to keep quiet and hide. I paraphrased.

In societies of the *equal* it is easier to sprout hierarchy, because – now we know – nature does not like equality. We see this in the domination of American democracy by corporations, communism from their lifelong president, revolutionary France from Emperor Napoleon, or equal before God by the Inquisition. Hierarchies are goody with equal people. There they are like sharks in the sea of small fish, ready for further mutual fighting.

Because the equality is right for them, the new hierarchies of liberalism have moved towards a global society, not realizing that the equating the world (in their favor) costs more and more, and in turn they get less and less, because nature does not like equality. We who do not follow these costs can instead notice the growth of *legal systems*. From the definition of a liberal state, which protects the liberties of citizens, it follows that the state must racketeering the citizens; confiscate their autonomy in the name of freedom.

It needs more and more repression against the resistance of equality (which generates conflicts), and then more equality as the basic fabric of the law. More law creates a greater need for the law to a spontaneous, unstoppable and in that senses a stable way. Those who are most capable of using freedom become its biggest victims, and countries with the highest laws (regulation, administration, bureaucracy) by the time are in the greater risk of slowing down with development.

The legally regulated state has been increasingly effective in protecting its oligarchs over time, better than any state in the past. Thus, we get an even smaller percentage of those who have an increasing percentage of everything. Unpleasant growing inequality arises, similar to the previous one, precisely in the course of alleged equality. It could also be shown as a great conspiracy of nature against small ordinary people, possibly in favor of technocrats lurking from the side.

If liberalism does not stop the mentioned absurdity, nature will continue its pursuit. States will strive to resemble on the tree trunks of living cells or some similar complex living beings. To them and us will be common the birth of universal units that are growing into specialized then die, which evolve into less free, less smart at the expense of the whole and less reproductive.

The law will more easily overcome obsolete hierarchies like family, customs, religion, but new fighters who have brought from liberalism. First of all, I am thinking of the modern owners of money and the power that will sooner or later want to master it. Initially cautiously, let's say with naive lobbies or the occasional responding of the rule of law on to the corruption, when we see it as a conflict of good and evil, then the balance of yin-yang, female and male principles, in the end, it could become a struggle for a life in which we're support the fighting against legal restrictions.

Liberalism is the ideology of freedom which in the name of liberation gives rise to the control of the majority by the systems and systems by the few. Still, by trembling from other trumps that nature hides in their sleeves, because of the pressure of the powerful, and ultimately because of habit and ignorance, we remain in liberalism.

March 14, 2019.

1.3 Truth or lie

Some asks me anew how to distinguish *truth from lies* on today's media. The questions of truth and untruth are the greatest secrets of the universe, and every new move in this field reverses the appearance of our civilization. It is not possible to have this "universal key" for truth unless you are the master of the universe. I think so, but I still say that there are certain ways.

The greatest wonder is that the truth is at all available to us, and this miracle was discovered by the ancient Greeks. They first understood the mystical relationship between the assumption, the implications, and the consequences. For each claim, only two states were established – true or false. No third. They noticed that from the false assumption by (good) deduction both the true and the false consequences result, and from the true only the true one, and it was the beginning of mathematics. When we can prove that a certain claim is both true and false then we have *contradiction*, and when we have it, then the assumption is false. So, the false assumption means that its negation is true. And that is that. Deduction has since been a useful addition to contradiction, but its glow fades.

Therefore, the supreme evidence for the discovery of lies is contradiction. Soft versions are based on suspicion, such as the oppressive ease of "proving". Others come from game theory. The tools of winning games are a desire for victory, an aggressive move and a lie (cunning). The power of this third decreases with disclosure, which makes it difficult for us to check it, because it is better "packed". If a competition is in progress, somewhere there are lies, but the player on the compromise, the good-good strategy does not go to victory, and it further complicates things. Lack of lies is the virtue of "goodness"; unlike the "villain" who could even conquer with lose-lose game (progression by the victim). An aggressive person (institution) in social competitions that strives for success or power is daring, rude, must lie. And that already is something in reveal?

The method of contradiction is not seen daily because it is heavy and repulsive. It's the game for the rare. Mathematics is the body in which such put the claims, and which returns it to us if they are true. We can imagine its abstract body as a robot that repeats only the correct sentences for us. It is clear that in the body of truth there is no room for a "barber that shaves all those, and those only, in its place, who do not shave themselves"? The concerned robot cannot say "I'm lying", but it can say, "I cannot say a lie". He can say once, "I cannot say a lie twice", but this cannot be said twice. If you understand, then you are a talented mathematician and I have nothing to talk about. But if you are not, then you can see how difficult the contradiction method is!

In examples like these, Bertrand Russell discovered in 1903 the *paradoxes* in the then-so-called naive theory of sets which was later corrected. In 1931, Gödel proved the *theorem of incompleteness*, which says that it is not possible a truth structure (say arithmetic) that could prove all its truths by itself. This further reduced the importance of deduction, as well as the power of axiomatic theories. Shortly, there is no set of all sets, no theory of all theories, and no formula of all formulas. Moreover, even the truth only move us (substance) directly, it can also be made of lies, for example, by negation or implication.

Contradiction is much harder than deduction, and therefore debates are more popular than geometry. That is why we are greatly contemplating and producing statistically significant percentages of wrongfully convicted prisoners, naively trying to fix things in the ignorance that we are actually victims of nature's skimping in providing information. The Greeks have noticed the open door to truth, and we that nature does not give it easily. We recognize its stinginess in the aggressive ease of polemic (from incorrect assumptions), in the greater attractiveness of lies and semi-truths than the truth, in the faster the spread of disinformation on the Internet, or in the fact that it is easier to encode than decode. Because *information* is larger when less likely, more likely events are realized more often. The difficulties with the truth, therefore, are not just our thing, but it is a universal principle.

This universal principle, that nature does not love the truth, even though it is only the truth that instantly initiates, makes life easier for liars, competitors, manipulators. It makes the world more interesting and reveals some deeper links between data, information, action and interaction. But we'll talk about it on some other occasion.

March 21, 2019.

1.4 Feminization

Local and humorous in a private correspondence with colleagues I used the term *feminization* for physical processes that give up of the outside world. Look your own business, and don't worry for others – commented it well they to me spreading the meaning, but advised me later not to develop ideas and applications. Not to waste time with the nonsense.

The second law of thermodynamics, it is known, speaks of the spontaneous transition of heat energy from the body of higher to the adjacent body of the lower temperature. By Boltzmann in 1905, the steps of Clausius, Gibbs, and Carnot, this law was observed in the form of a spontaneous growth of *entropy* (disorder) in the uniform distribution of air molecules in the room.

The amorphous, impersonal molecules of gas as if to hide information about themselves, reduces its emissions toward outside and deals with interior arrangement, so it is convenient to say that it feminizes. Try not to understand the same process now as increasing disorder (entropy), but on the contrary – as an increase in the inner regulatory and you have the need for a better term. There is also the discovery of the reverse side of the entropy or disorder.

This can be seen even more generally. Uncertainty before throwing coins, listening to news, accident, is realized in the information after. We exchange news with communication, as well as particles the consequences of their interactions. Due to the multitude of possibilities, interactions have no end, but what comes arrives. We think that the information is plastic like energy, so that it also transforms itself from shape to shape, neither arising from nothing nor disappearing into anything, otherwise the proof of the experiment would not be valid.

From this maintenance law, however, follows the finality of each property of the information. Namely, *infinite* is the set that is (by amount) equal to some of its proper subset. From the same conservation (maintenance) law we understand that the uncertainty is kind of information too. When inside is less uncertain than the outside, the editing of the interior is less risky than the outer march, it is more predictable and more meekly. Turning to that side of the unknown is feminization.

If you understand this, then you will be able to transfer "feminism" to *living beings* because their main properties are just options and decisions. Population less threatened from the external hazards becomes adapted and feminized. Initiatives of herbivores, collectors or vultures are less rude toward the outside than in the beast.

If it is not in an aggressive environment, the introvert species can evolve towards the optimum when every change would only deprave it. This perfection has its price in lagging and outdated in relation to the dynamic environment. That's why have had overcome the two-sex species, better adapted to complex relationships and environments, with male sex that would take on suffering or success in risk situations?

The question mark stands because this is an unexplored terrain in biology. Nevertheless, we accept that individuals in their youth are in rush for a fall and in acquire for experience achieving its maturity, and that they slow down with adventures in their old age. In this sense the very organizations of living beings are like living beings, so the use of a new determinant reaches, and, say, to social phenomena.

Each of about the 30 well-known civilizations that feminized, previously were in the expansion, often brutal. Feminization did not mark the end of every one civilization, but it never happened the other way, that it signifies the imperial rise of some. We recognize the greater chances for drastic changes through external uncertainties, and internal editing as a reward and pacification which was not worth without the first part.

Consequently, we see Suleiman the Magnificent (so-called Lawmaker) as the turning point of the rise of the Ottoman Empire and the beginning of the "sweet fall" of that later the "empire governed by women", as well as other successful civilizations that through their voracious rise and enrichment arrive on safety at their end. Further, I wonder if the working term "feminization" may not only be a homonym to the one known in everyday speech?

The questions are actually much more. Is the recommended useful work some of the most useful things that we could do, what would practice without good theories, do (legal) reductions of aggression really build a better future?

March 28, 2019.

1.5 Life

Here's what is *life* (I quote an elderly lady): you enter one door, you go out on another, and that's all! Could it be said on some more exact way? It certainly can, but for the start, forgets all about the usual debates about the meaning of life and "science" about it.

Life can also be defined by means of communication and alternatively by means of action – energy and duration products. This second is in science more familiar, so let's consider it first. There is no part of the known matter without at least some energy and duration, and yet nature is as if it wants them as little as possible. For every well-known trajectory of physical movement, the *principle of least action* is valid! It is the basic tool of theoretical physics and still without significant further application.

Waves move through the middles in different speeds, reflecting or refracting, constantly spending the smallest time between the two places on the road. Particle interactions with the environment are in the least possible energy exchange. This stinginess is confirmed by the famous Euler-Lagrange equations (1750), from which we then perform geodesic lines of movement in various fields of physics, including relativistic and quantum mechanics. The various minimalisms are regularly confirmed by the experiment. It is precisely this need of a substance to not-work, its selfishness, however, which allows it to possess excessive actions.

The substance proves its existence by action, but on the other hand by communication. The exchanged information is bigger as it is less likely, and less likely is the rarer happend, so the previous logic of minimalism of the action is now being repeated with the information. Natural savings of information becomes a newly discovered principle that triggers the accumulation of information as well.

The substance with a surplus of action and information that we do not call physical, called the "living being", in contrast to the one that does not have that surplus. Dead matter has minimal actions and communications, and then it means that living beings have them more: the living cell contains more information than the non-living substance it consists of, it has more options available and makes more decisions. It is revealed by the motion beyond the solution of the Euler-Lagrange equations.

From the new principle – that the nature is stingy with the emission of information – we come to the definition of living beings with information in excess. It would be easier for its surplus for the life to reject them into the surrounding, if the substance is not filled already, and therefore it must be solved by the interaction on rates, even by organizing. It is not possible anyone to communicate with everyone, and that fact leads the substance to the association among suitable.

Similarity is an opportunity for "organizing", and that is actually an abdication of personal freedoms, of surplus options, of information and action overbalances. The collective is created by taking accumulated individual accessions. We renounce part of our freedoms for the benefit of the rule of law, for the sake of safety and efficiency, as living individuals sometimes evolve into a higher structure. The individuals are embedded, limiting themselves to the benefit of the collective, and consistent with the new definition of life, we say that the very organization of living being is also some kind of *living being*.

The similarity of the living tissue goes to the initial universality of living cells. Growing ones specialize, contributing to the efficiency of living tissue. It is a replica of the principle of minimalism. *Intelligence* is the ability to use options (revised definition), so better organized may also be less intelligent. This is expected by the evolution of the organization.

For the complex organism the obedient is better, also those in the narrow segment of jobs, that is, the specialized cells which are less autonomous, no more unpredictable scandals are required, nor some too smart, but also those that can uncontrollably reproduce. In this direction is the development of social systems going?

The described view of life is not unknown to biology, but it gives it a deeper meaning here. Darwin's evolution lasts not by accidental selection, which due to the abundance of possibilities will tended to the disorder, but also by the principle of information leading it to the organization. The life of the individual, colony, and species also becomes like a cyclone in the ocean. After the storm shine the sun and everything calms down, as if the tornado was an unwanted disorder, and its calmness was a success, but in addition, we see some attraction in the unrest. The beauty of life moves us.

April 4, 2019.

1.6 Equality

It tells me a colleague to say something about *equality*. It can, but it is known that she, the nature, does not like great leveling in any sense and that we are raping her in that regard.

Nature hates equality and will do everything to hinder the ideology of our legal systems.

Barabási, an American mathematician of Hungarian origin, recently explored the networks (since 2000). Internet, power lines, popularity, money flow, are the applications of his discoveries that once established become a universal thing.

Networks become denser by increasing the number of nodes, for example, new users of the web or intersections of roads, and equality is expressed by randomly linking a new connection with already existing lines. Thus, however, stand out centers say *concentrators*, with many links, simply due to the increasing connectivity probability. They are aggregates, monopolies of cash flows, databases, command posts, always small number of them and in the function of saving their network communications. They are characterized by the so-called *power law* distribution.

The *six degrees of separation* rule is the result of Barabási distribution. The free-made acquaintances with someone who knows someone who knows someone and so on in (about) six steps from-to shall connect (almost) any couple of people of whatever great world. The ease of connectivity is again achieved through rare individuals with many acquaintances. Insisting on a different equality, for example by connecting just about every pair of nodes, leads to rapid network congestion, and the attachment of only adjacent, as in the *deaf phones* children's game (with a whisper from ear to ear), to slowdown and unreliability.

Synergy (Greek: $\sigma\upsilon\nu\epsilon\rho\gamma\acute{o}\zeta$ – work together), a state in which the whole is somewhat larger and different from its parts, now means accumulation of information. The resulting bonus becomes the livelihood of the free development of the network and the spontaneous separation of the concentrators. It may be shown that the same form is followed by *Nash equilibrium*.

This is the theory of John Nash, the American mathematician and winner of the Nobel Prize for Economics (1994), known from the biographical film "A Beautiful Mind" in the role of Russell Crowe. He worked on equilibrium in games, places of competitors in which the individual participant cannot gain advantage by leaving the group. It is typical for teamwork. Basically, it is a mathematical model, and therefore is widely applicable.

The free market builds Nash's equilibriums like vortices from which companies in the struggle for profit cannot easily get away. The economy is dynamic, so it's attractive spots eventually weak to the complete collapse of the participants or to the creation of a critical mass willing to move for the better. Capital crises have emerged that have been discovered by *Karl Marks* calling them the transitions of quantity into quality. Now we see them as spontaneous organization, network efficiency and savings of information.

The Communists tried to avoid the free market by the Étatism of economy (state-run) by compromises of the desired and possible, thus losing the part of both, which led to a known lag. Only recently, after the appearance of seemingly incompatible models, Barabási and Nesh, we can further understand that the links of the network nodes and strategy games have a common form, and then why and how their efficiency is aligned.

Common to these cases is saving interactions and running away from equality. Let nature leave at will it will work for us as we will not like, it will defy our ideologies and irritate us to remove it from duty. *Equality* costs, and in that cost there is a decrease in efficiency too, so we have more and more lawyers, more state regulations, less freedom and always too many mistakes and abuse of rights as long as we have with what to pay our dogma.

Because we want to believe that the idea of equality is not contradictory or that equality can be approached more and more (which is the same), and in order to justify the work of our legislators, nature would be better if there were equal persons or at least equal conditions. Because the nature has to bow before us, and we are who she has to worship!

April 11, 2019.

1.7 Dunning–Kruger effect

Dunning–Kruger effect is a cognitive bias in the field of psychology according to which people of lower abilities have an illusion of superiority and are wrongly evaluated as being smarter and more capable than they are.

Darwin (Charles Darwin, 1809-1882) once wrote that ignorance generates more self-confidence than knowledge. Because of this combination of poor self-consciousness and a low cognitive ability that leads to overestimation of oneself, *Russell* (Bertrand Russell, 1872-1970) said that the problem of this world is that smart people are in doubt, and stupid people know everything.

In investigating the effect named after them, Dunning and Kruger, among others (1999), asked respondents to evaluate different jokes, both their own and others. Uncompleted people have proved to be not only bad performers, but were also less able to recognize the quality of their work than others. It is not uncommon for students who were worse on the exam to feel that they "deserved" a better grade.

Let us now consider this phenomenon from the point of view of the theory of information in general in the collective (people, living and non-living beings, random events). A mass of equal individuals can be seen as an amorphous impersonal set, maximum entropy (mess) for which we can add that it has the smallest possible emission of information to the outside world and in that sense it is best "feminized" (turned to itself). Reducing external communication is simply called a smaller "mind" of a group of equal ones, which would have to be less than the average "mind" of its individuals. This tampering of individuals in the mass applied to living beings means that the D-K effect does not come from the less intelligence (IQ) of the respondents, but from their equality!

The mind of the mass can be stated in referendum questions in similar tests as well as the individual's minds, and the result should be different from other abilities, for example, that more them do more work. I have no doubt that such a measurement would confirm an analogy from the theory of information – the conclusion that the mass of equals is less effective than unequal ones, than hierarchy, of course, in relation to the outside world.

This efficiency can be defined as a security, economic or some third of the game theory, it does not matter. In any case, efficiency is the surplus of information[1] to the outside (outside of the collective), which in fact means loss of information, cost, because information like energy changes form (kinetic, potential, thermal, chemical), but not the total amount. Information is a measure of (variable) data.

The loss of information in efficiency is reflected in the reduction in the ability of communication of individuals, the limitation of their control of options, the reduction of individual freedoms in favor of maintaining the efficiency of the collective, then by time in the lagging of the collective in the changes, or becoming it obsolete in relation to the outside world. The consequence of efficiency is a lack of development (which is rarely known). On the one hand, there are many options, freedom and development, and on the contrary there are security and other efficiencies, and societies in a harmonious development tend to optimize all these opposites.

What is not mentioned in the psychology books about the D-K effect is the understanding of its absence. These are the situations of individuals in hierarchies. If this is confirmed

[1] Not known in the game theory, for now.

(by some future measurement), then it is reasonable to say that it is more effective for team work to choose the individuals of different specialties. The same comes from the thesis that nature “does not like” equality, and both comes from the “principles of information” – that the nature is stingy with the emission of information, that is, it is more likely to realize more probable events (which are less informative).

Therefore, it is wrong to estimate people with a surplus of self-confidence as some kind of stupid. Their appearance should be viewed wider – as the success of the democracy in which they live.

April 18, 2019.

1.8 Bystander effect

Bystander effect (or bystander apathy) is a social and psychological phenomenon in which individuals in the larger group show less interest in helping the victim of a miserable event. The more observers are there, the smaller the chance that someone will intervene.

American psychologist John Darley discovered this phenomenon by publishing an attachment to the New York Times in 1964 for the killing of 28-year-old boy Kitty Genovese, claiming that 38 witnesses watched the attack, but none of them called the police or offered help. A lot of professional books have since been written about “Genovese syndrome”, but, I believe, not much about what I’m about to tell you now.

The principle of least action (the product of energy and time) is one of the basic tools of theoretical physics. However, as far as it was unmistakable in predicting the trajectory of the motion of the substance, both in classical physics and in relativistic or quantum, it always persisted without wider application outside science. Until today.

If we know that physical information is an expression (equivalent) of a physical action, that principle of minimalism becomes also the *stingy principle* of information, it’s non-splurge. The reluctance in communication rules everywhere around us – from the reflection of light on the shortest path, through the falling of the body in the gravitational field along the paths of the least energy consumption, to the spontaneous processes of “feminization”, that is, the tendency of the physical system to reduce the “editing” of the outside world at the expense of its. “Genovese syndrome” is, simply said, the consequence of such processes. Here’s how.

By saving, matter can be grouped and created by surpluses of actions (unrest) and surplus information (life), and more than that. Living beings are thus further organized into more complex systems, at even more levels of life, unselfishly surrendering their parts of information (degree of freedom) or action to a common order. We also spontaneously become the cells of some tissue of our future ever more complex organization, by evolution formally similar to many other living tissues: the universality of birth (savings of communication mode), the specialization in the mature age (savings of the mode of action), the decline of freedom (saving information) the decline of intelligence (communication saving) and the ability to reproduce (saving the action).

The “Genovese syndrome”, consistently, represents a surrender of itself, the incorporation of its surpluses into a higher hierarchy. As for us today, this is above all a legal system. People adapt to their own social order become less prone to personal incentives outside and are increasingly willing to rely on the system to solve the problem. By developing in such currents, movements and individuals, we see less and less of things in personal responsibility, and more and more in the legislation (stop here not to argue with politics). When we believe in the system, then the environment like barracks becomes an ideal and then really weakens the "bystander effect".

In other words, individuals of a regulated society left to the situation of equality become blocked. Their apathy is a confirmation of the advanced democracy in which they live and their faith in the legal system. If this thesis is correct, then the respondents convinced that the system is "value" and brought into equality conditions will show greater bystander apathy. This can be tested, for example, with anonymous travelers on a bus, passers-by on the street, viewers in the theater.

Valid is also vice versa. If you do not want the incident anonymous look apathetic, disrupt the equality, produce a hierarchy by imposing yourself temporarily as a leader. Take the initiative in the burning theater decisively, commanding the departure of the room and you'll save many. In the event of an aggression on the passerby, boldly vow, "People, he / she attacks that person! Call the police!" And the new authority will launch the audience. Remember, Napoleon was not physically stronger than his soldiers, but he had the ability of domination and he would become a master of turning the apathy into killing tools and weapons.

In case you are not in the crowd, decline the chances that you are reliant on the system.

April 25, 2019.

1.9 Dictatorship

The biggest problem with the dictators is that they are not able to see the benefit of truth and freedom. Encouraged by the enthusiasm that comes with the order, and then because of prejudice, they overlook the outcry of obsolescence.

According to this definition of the *dictatorship* it is a lawful society that accelerate first and then slow down, such as the recent *fascism* been in which the state was the proclaimed master, and not the servant of the people. Or *communism* (dictatorship of the proletariat), and tomorrow maybe *liberalism* because of its foundation in the need to take on the rights of individuals in the name of their alleged freedom.

If we are bees or if we have shallow wisdom, less perception, modest curiosity and weaker impulses for development, we could have been born and die more than hundreds of millions of years, provided that the resources last. Evolution does not stand and we would slowly adapt to a reduced number of complications. This is likely happening in the nature, and the development of intelligence is so incredible – we are the one of countless millions of failed processes – that the directing of our species in this direction leads to a reasonable doubt of some biology setting, its principled questions.

"Better theory gives a better meaning to facts" – the motto is to redefine natural phenomena. So here we first declare that there are *choices* and that it is possible to *make decisions*. Then, let *intelligence* be ability to control options, and *freedom* their amount similar to classic information. With this we are in the field of mathematics where there is more truth than we can imagine and this is already the first unsolvable problem of dictatorship that does not want to lag behind. There is no best criterion (*Arrow's theorem*), there is no truth in all truths (*Gödel's theorem*), there is no set of all sets (*Russell paradox*). Mathematical-based theories create insurmountable obstacles with which there is no compromise, but in turn, they enable relatively easy and unlimited assimilation of exact constructions. They emphasize difficulties and clarity.

On the other hand, the assumption that we believe in physical experiments leads to the conclusion that information cannot be created from nothing and cannot disappear to null. It is plastic, like energy, changes shape and keeps the amount. Perceptions are then not only interactions, but also communication, and information, freedom, action, and truth are equivalents. It's weird, but the handy side of the magic of physics is that its phenomena, for which the law of conservation (energy, mass, moments, spin) is valid, are mutually reciprocal.

For example – in analogy with the increase in pressure by reducing the surface area, saving one's own action on one thing leaves them more for something else. A manager with simple daily routines has more creativity in a more important job, which is also a chance for less smart ones who usually better tolerate constraints.

Every action of freedom (perception) has its reaction in an obstacle (environment) and vice versa, and the total freedom is the result of conflicting (many) possibilities and limitations. This stimulates enthusiasm for those who are eager for freedom that in the limitations of the dictatorship see the liberation, but dual, and to those who are scared by open opportunities. Freedom is here the total *information of perception* which is the sum of the product's among ability of the individual and the corresponding impediments to the environment.

Without entering into the algebra, it is possible to feel that the norms absurdly do the both, and define the wishes and mean the lack of freedom, and for this second we say – because the material options are consumable. The essence of freedom is those possibilities, and of them are the uncertainties and of these are originalities. The essence of all of them is unpredictability.

There are no real discoveries on the trails, so the brake of creativity is order. What the legislator can think of as the rules of the game in the economy, education, street walking, is always too narrow for a genius who could appear there. That could be one of those few who create something new, the most needed to dictatorship.

Like a ship that is driven by powerful engines in a outlined direction, and which safely and efficiently follows the route, so the dictation in its best way leads its passengers to certainty, ignoring a lot, perhaps even something better along the way. The blindness to development is one of the prices of over-stretching, and the second is dumbing down, because we as a species are also adaptable. When we evolve in one way we are dwarfed in some other way – the curse is of good orderliness, because the development directions are always too many.

So we understand that the big problem of the dictatorship is their blocking of our mission to be human beings, to remain a curious and intelligent species on this planet.

May 3, 2019.

1.10 Crime and Penalty

There is no valid evidence that the extension of legal *penalties* reduces the percentage of *crime* in society. It is known that saturation of society by laws reduces the chances of development, so the question arises as to why we are striving to more and more durable laws? Why are we extinguishing out the fire with the oil in the alleged fight for justice?

What will happen now – some years ago they asked me about the proposal of the Family Violence Act (Official Gazette of RS, No. 94/2016). The *law* will come out and violence will be increased (perhaps significantly), and then it will be an opportunity for further intensification of penal policy. Today we know that all these three predictions have been confirmed, but that the same arguments have not gained popularity. A greater threat of *imprisonment* will provoke a greater affects, but it will also bring the *killings* to its alleged dismissal: I better beat it dead, but it laughs to me while I'm slaving in penal servitude. Behind this loose interpretation, of course, there are deeper reasons.

There are a number of statistical certificates, other experimental studies, but also some points of mathematical *theory of games* (Heiko Rauhut: Higher Punishment, Less Control? — Experimental Evidence On the Inspection Game; July 20, 2009) that prove that there is no accurate "rational conclusion" the greater penalties lead to lesser crime. Because a precise analysis requires a smart and objective interlocutor, we realize that our demands for "higher order" do not come from scientific sobriety.

The same was true with my recent criticism of the so-called Tiana's law[2], now voted in the Serbian Parliament, and maybe soon imitated in the Republika Srpska. The state must not deal with retaliation, I state, and then from the legislation, as well as from the investigation or court decisions, it is necessary to exclude personally interested persons. However, instead of exemptions in the implementation of the Tiana Law, the personally concerned persons were given the main role. Moreover, with about 160,000 signatories, citizens are captured by fraud (who will say that it is against the "protection" of children) in a way that many other laws can be duped, which is then considered politically correct.

The decline in credibility of legislation is always threatened by political manipulation. In this case, they are revealed in the official support of the positive sides of the new law without mentioning those other parts. In flirting with emotions that tend to *retaliate*, in the understanding that in the event of a collision of force and truth, the truth is easily lost. In the trepidation of politicians being able to make other "two legs" of the state subdued under their own and discover that it is only a matter of their grace that they will. In the doubt to the politicians do they want such kind of *dictatorship* which would inevitably result us in slowing down the development.

Professional and fair *lawyers* will recognize that this law is not completely in line with our Constitution or with international law, but let's say it does not matter here. From the point of view of the wider truth, it is irrelevant whether laws are like this or that, but whether they hinder the creativity of the society. This should be equally important priority of legislation, comparable to the principles of human rights or the mentioned separation from politics.

Honored lawyers could admit that the expected maniacs or criminals returnees often do not get into prisons as much as in madhouse, and Tiana's law targets (almost) into the empty, that is, it is statistically less significant than the collateral damage it has a chance of doing. Also, it is fair to say that legislation made far more errors than acknowledging that it is willing to admit and that the austerity increases its injustice. The mistakes of the law are worsening as more *capable* people, whose agility, entrepreneurship and wittiness are most needed, are more likely to fall into its slaves as innocent victims of the judiciary.

Therefore, the danger of increasing crime in society comes from those who would allegedly suppress it, but that's not all. Our *politicians* are not only ours, and therefore they must lie and manipulate. It reveals, for example, the strengthening of the powers of the Executors who seem to be there to discipline small rogues by confiscation the large amount of their assets. But the silence of the authorities for cry of the little people that in this way goes deeper into homeless throws doubt on the IMF[3] and its reliance on big capital. Are we making for this strongman the easier prey here?

[2]Initiative for tougher penalties for killer of children.

[3]International Monetary Fund

The impact of large capital has two faces. The trend is for the *rich* to be even richer, that they are (in percentage) less and they have more and more (of all), while others are becoming weaker and harmless. The rich tighten their hierarchies and by the rule of law ensure and increase the acquired values, but the state of disorder corresponds more closely to the population. So they do, and the *poor* who seek retribution get what they went for, even though the both did not want the same. This is again an interpretation behind which there are deeper reasons.

Namely, the *physical information* is true (it is impossible to communicate for which we have proved it is not possible), the action (it changes us, at least infinitesimal), matter (all substances as bulbs consist only of information). "Minimalism Information Principle" is valid because more likely events are more likely to be realized and they are less informative. That's why we are attracted to a lie (especially if it resembles the truth), so the "principle of least action" is everywhere in physics (the action is the product of energy and duration). This second is generally known, first so-and-so, and the following is not: therefore, the universe is expanding.

In short, we are striving for ever more denser and more stringent laws, primarily because of the "principles of information", but also because information is also the both, uncertainty and aggression. The skimping of nature leads to its accumulation and creation of living beings, to our hierarchies, and of turning towards ourselves. This reversal in the physical world is called a spontaneous growth of *entropy* ($\acute{\varepsilon}\nu\tau\rho o\pi\acute{\eta}$ – craft inward), in the society I call it *feminization*. Both are types of escape from freedom, from which we can not escape, because this world is only of its terrible tissue made.

1.11 Uniqueness

Physical substance is just what we can in some way interfere with. Even the *dark matter* of the universe according to this definition is a physical substance, because it gravitates to galaxies, to stars and other celestial bodies, and they act upon us. Also, there are no actions without information, and in addition, both are *truths* in the sense: it is not possible the physical act for which could be proved that it is not possible.

I answer the question asked, "Why are we not equal?" in which alludes that injustice in the world allegedly could be solved by the free proclamation of *equality* of all of us. We have overcomes the question that there is no identical faces and the same conditions, and therefore that equality is an impossible mission of legal doctrine, and then I am expected to present the deeper causes of this inequality. OK, I say, but the answer from the standpoint of today's trends is so unexpected that we have to go slowly, in steps.

Theorems we discover are like nodes of *free networks*, with a small number of so-called *concentrators* that have a large number of links opposite to a large number of others, and the idea of information, the discovery of this very theory, is one of the denser places of truth and deduction networks. Some of the *principles of information* are also thick: conservation, minimalism and uniqueness, and as I will explain are consequence of so-to say all three principles is the answer to the question raised.

The first is the "conservation" principle that talks about the law of maintaining (physical) information and is generally known to physics. If I need to say something new about it, let it be that the information is proportional to (physical) action, and we know that there is a quantum of action (Planck's constant), which is why the information is already indestructible. Different proof of "conservation" is communication itself. That's why we can communicate because the information sent to us cannot just disappear. Therefore, we must communicate, I will add, because we do not have everything – because we are different.

The second principle is "minimalism", a direct consequence of the "principle of least action" already known in (theoretical) physics. To it, generally, we can come up with non-speculative way by using (mathematical) probability theory. More likely events are less informative, and they are more common. Therefore, because it is greater news that a man has bitten a dog, then the news that a dog has bitten a man, we know that nature is shy with information. Then we know that in a set of equal outcomes the singular will have the least probability, so that nature does not like equality. The throwing of a fair coin has greater uncertainty; the outcome of "heads" or "tails" in a fair case is more informative.

The third principle is "uniqueness". We talked the least about it, because it is mostly a matter of physics. This principle can be understood as the announcement of the so-called *Mach's principle*, as Einstein once named, after the most famous physicist and philosopher of the 19th century, the influence of the mass of the whole universe on the water in the rotating washbowl in relation to those masses. Water is spilled due to the centrifugal force generated by the relative movement of water relative to the universe, and vice versa would not be possible. By the way, the same experiment with the washbowl also used Newton proving "absolute space".

Analogously to the Mach's principle, us defines our past. Each particle we consist of has its own history in which it is unique, and thus, each particle of the universe is *unique*. The substance is defined by information, including information on preserved information about it. Consequently, in the famous quantum-mechanical experiment *double slit* there occurs the interference of particle-waves (all matter is wavy) through two slits, even when these particles encounter one at a time separated by long periods of time. All the appropriate particles that have ever been passed, from the creation of the universe to the present, through the given space – interfere with that particle-wave now.

What we see today is like a wave on the surface of the sea which is also the *interference* of all the layers of water below. If nature allows equality, it would allow at least this phantom equality the history of at least two particles, and the aforementioned experiment "double slit" would not be possible. However, this equality it does not allow for any particle of the same quantum state in relation to only four known quantum numbers. What I am talking about is well known to chemists, who from this so-called *Pauli Exclusion Principle*, derives from the earlier known elements of the Mendeleev Periodic System.

Here is one place where it stuck the *legal doctrine* with its idea of equality. If you notice that in the legal practice it have found or proved two equal beings in this universe, or at least two truly equitable situations, I would like you to tell me as soon as possible. It will be a great earthquake in exact science – I answered at the end of the letter.

1.12 Free will

If there are options, then there is no *determinism* and, accordingly, it is said in another letter, we have free will and responsibility for our actions? It was a question to me from a colleague who, in accepting coincidence sees the possibility in consciousness to control our destiny. But things are not so simple.

Determinism is a philosophical idea of events and moral choices fully determined by some previous causes. It excludes free will, assuming that people can not act differently than they do. However, if sometimes we have really random events, then there is no idea of determinism. But again we are not free to manage our destiny, which is then left to uncertain outcomes.

As if they were aware of this paradox, some ancient thinkers limited any uncertainty to people, and their controls were attributed to *gods.* Today we can go a little further and calculate the more weird conclusions.

We know that there is infinity of natural numbers, we say *countably many.* Equally infinitely have integers or fractions. They make the so-called discrete (abstained) infinite sets. Unlike the finite, infinite sets can be equal to their proper part. Accordingly, all physical phenomena, for which the conservation law applies, are finally divisible.

For example, the smallest amount of action (energy and time products) is Planck's constant, and this is the smallest interaction, and, as far as I am concerned, the smallest carrier of the substance's communication. All mutual actions and physical communications, as well as all the atoms of our body, and even the universe, can be transformed into one at most the countable infinite series. Because of the wave nature (every form) of matter, we can always numerate the positions in some wavelengths, and the duration by the blinking, and the space-time of any given physical events remains a discreet set. All programs of modern (classic) computers can be so aligned, and hence, any material structure can be represented by countable, discrete codes.

In contrast, real numbers have an uncountable, the *continuum many.* There are so many points of the plane, the points of the line, the points of one segment, because the continuum is infinity, so it can be equal to its proper part. The irrational numbers, which in numeral notation have infinitely many non-periodic digits behind the comma, have as many as real. The very positions of these digits make up a series, but their variations are more than that. This impossibility of placing the continuum in discrete gives us an idea for a deeper understanding of our consciousness.

The multiplicity of our thoughts indicates their uncountability, although they always follow some (countable) sequence of moments. If the substance itself is (infinite) discrete, the world of the ideas that explains it is a continuum. Therefore, with the accurate cloning of a man by pure copying of his atoms, we will not transfer consciousness; it can not be done by classical programming, but can perhaps by quantum, since quantum states of matter are superposition of coincidences.

Superposition is generally a property of the linearity of connected phenomena, when twice more one means twice the other. Here, in particular, we collect the probability, as when we double the chance of winning a prize, by purchasing two ticket lots. Each random outcome realizes the information exactly equal to the amount of previous uncertainty, and, in analogy, the superposition by interacting *collapse* into a new *quantum state* without changing the corresponding quantities. Each interaction made the quantum system to *evolve* into its new reality, giving up of all possibilities that could happen but did not happen, which we call *pseudo-realities.* In pseudo-realms the same laws apply, but mutual (physical) communication with such is not possible because of the law of information conservation.

Thus, quantum superpositions constitute a continuum, un-countable many of possibilities, although the number of realized outcomes is always not more than countable infinite. Our "free will" passes through the continuum of the *multiverse* idea, through realized options of *parallel realities* within the same laws of physics are in place, and which do not communicate one by other, so that our physical body and all surrounding matter remain discreet.

The *quantum mechanics* is a highly consistent representation of abstract algebra and so far probably the most exact and experimentally proved branch of physics. Perhaps it is precisely the reason that the discoveries of quantum physics are so devastating to its experts, which are more in the spheres of an abstract than physical. Beginning from its superposition, for which Einstein, otherwise one of the founders of quantum mechanics, said in unbelief that "good God does not play a dice", and to the multiverse "whom God is not needed", as is criticized by modern theoreticians of theology, it in spite of its scientific reliability persistently remained at the margin of acceptability and somewhat on the other side of reason.

If one really could choose his own paths, with full awareness of the exact consequences, then he would actually manage the universe with his choices, he would change the entire material universe with his own desires, decisions and will. The question from the beginning is, do we really have so much power?

1.13 Repetitions

How come that the events are unrepeatable, and our genes are *repeating*? Why are the celestial bodies circling, and we walk on the sidewalks using only a few templates (which mathematicians are just discovering), and we claim that matter is made up of a huge number of unique interactions? Why do political parties of one society increasingly resemble one another, although there are no equal faces or equal conditions? These are interesting questions for the theory of information, which pretends to be more general than the classical, Shannon's.

Information is the true, but also it is an action and interaction, a physical matter and comes from uncertainty. Contrary to the usual belief that we know the past and the future is hopeful, we see only the consequences but not the causes. Only by substituting the thesis we arrive at the "conclusion" that similar events produce similar consequences, and then due to the finite partition of each property of information, and therefore the final number of their combinations, we perform (hypothetically) the thesis that everything, but really everything, in the material phenomena – is *periodic.*

In extremely simple systems like *photon* (particle-wave of electromagnetic radiation and especially lightness), the growth and fall of an electric field in one plane reduces with the rise and fall of the magnetic field in the perpendicular plane, alternately on the upper and lower left and right half-plane of the photon path, while it travels by encountering constantly to other places and other times.

The tendency of repetitiveness of matter increases the parsimony of the information. Similar to *free networks* of Barabási, which for the sake of efficiency tends to create a small number of nodes of concentrators with a large number of network connectors, large information systems favor fewer ones. Such a dominant "force" of the Roman Empire was its attraction that stemmed from wealth, orderliness, security, and from which sprouted barrenness and weakness that could not resisted to the waves of the barbarians. The descendants of the then-settlers, the Visigoths, today's advanced Westerners, may have similar fate of Romans in the pattern of rise and fall. History, of course, knows how to "return" even in shorter cycles.

Genomes of descendants also transmit features such as moral ones that are not physical, such as potatoes, but are material and expressions of physical information. They are also necessarily repetitive, within one species today, but also through their generations.

The theory of deterministic *chaos* is a new branch of mathematics inspired by meteorology and *recursions* (parts that repeat the whole). It deals with the phenomena of small initial differences that escalate and which is therefore a good test of the above theses. This is the *butterfly effect* whose wings movement in Mexico can cause a hurricane in Texas, or *critical mass* that is smaller than the majority but can trigger the whole system. As some chaotic and periodic things have already been discovered, for example in repetitions of the storm or structure of the tree trunks in the crown and leaves, and their forms are called *attractors*, now we just add that for all other natural currents will be found suitable "attractors". Due to the principle of uniqueness and finality of information, its cycles and recursions are never identical, but are always limited and mutually no more than similar.

The examples of the period in local climbs and falls we can find also in *economy*. Say, for your new product as a monopoly, demand may grow, your income grows, but the new way of earning is appealing to the competition that is imitating you and the supply rises. The customers are finally a lot, so the revenues go through its zenith and begin to decline, which could motivate you to start a similar cycle with a new idea.

It is interesting to note that mathematical analysis, which deals with the continuum (not characteristic to the substance), knows the theorem (*Fourier series*) claiming that each function can be sufficiently approximated by periodic sinusoids. Moreover, any fragment of a trajectory wills approximate any other trajectory given in advance (similar to physical) in increments with increasing accuracy. These attitudes indicate that the form is not as important as the mere repetition, and they, in my opinion, speak of the connection of the material and immaterial (abstract) world of truths. That the first is deductible from the other, that the other is the envelope of the first.

It is not a novelty to know that there are similar periods of material occurrence, but it is the discovery of the assertion of the impossibility of their non-periodic behavior. When we think a bit more, we will encounter the principles of information in various of our everyday routines, its unpredictability and uniqueness, its stinginess and the law of conservation.

1.14 Emmy Noether

Is there a woman in mathematics? Of course there is. I will assure you of one important discovery of one of them, the brilliant Ammy *Noether*[4] which, because of the theorem, is called a kind of icon of algebra and physics, but whose significance will only grow, I hope for the theory of information I'm just doing.

Emmy was a Jewish-born German raised to be a teacher of English and French in girls' schools, but instead she went to study mathematics at *Erlangen* University, where she worked with her father, mathematician Max Noether. Women were allowed to be on classes, but only in the presence of instructors, and its instructors are today widely known theorists *Hilbert*[5], *Klein*[6], *Minkowski*[7] and *Schwarzschild*[8]. She received her doctorate in 1907 on algebraic *invariants*.

The way Emmy treated these invariants became the subject of admiration for the first of Hilbert, Clay and *Einstein*[9], and then of many others capable of understanding Noether's mathematics, her own "poetry of logical ideas". Emmy's the main work is of 1915, which we often call "the most beautiful theorem in the world," I will try to explain in a popular way.

---

[4] Amalie Emmy Noether (1882-1935), German mathematician.
[5] David Hilbert (1842-1943), German mathematician.
[6] Felix Klein (1849-1925), German mathematician.
[7] Hermann Minkowski (1864-1909), German mathematician.
[8] Karl Schwarzschild (1873-1916), German physicist and astronomer.
[9] Albert Einstein (1879-1955), German-born theoretical physicist.

At a time when physics still discovered the law of energy maintenance – according to which the sum of the kinetic and potential energy of the body is constant (the energy of motion and rest, realized and unrealized) – *Euler*[10] and *Lagrange*[11] were miles ahead of their contemporaries. They considered the difference between kinetic and potential energy, which we call the *Lagrangian* today. They assumed that this difference in spontaneous situations did not change over time and in 1750 came to the partial *differential equations* of the second order named after them. Lagrange discovered the *principle of least action* brilliantly interpreted it and in 1788 laid the foundations of classical mechanics.

The Euler-Lagrange equations are related to any motion, in as much *generalized coordinates* as to give trajectories of the least time consuming in reflection or refraction of light between two points, through swinging pendulum, the spring vibration, and, say, the least energy consumption in the classical, relativistic, and finally in quantum physics. Generalized trajectories are "paths" of the evolution of physical systems of the unaltered Lagrangian, which we call *symmetries* or invariants.

When you stand in front of the mirror and observe your *reflection*, then you participate in the plane symmetry, the reflection in relation to the mirror plane. Each triangle with its reflection has equal sides, the same area, although the opposite orientation. Reflection is the axial symmetry in relation to the axis, the given real or central symmetry in relation to some point. In the same category of "immutability within the transformation" are included the *translations*, parallel shifts of figures without distorting the distance between the inner points. In the first grades of secondary schools, in *geometry* we already have learned these so-called *isometrics* (Greek for "having equal measurement"), where we could learn that geometry does not have much of the symmetries and that each of them can be reduced to one or two *rotations*.

Emmy Noether noticed that one of the two Euler-Lagrange equation's items is a change in the Lagrangian (energy) over the generalized trajectory and that this change in the case of symmetry disappears, and that the remaining one, which represents a change in the amount of the corresponding physical system over time – also disappears. She resolutely concluded that the presence of "immutability" means some symmetry and it the conservation of the corresponding physical value.

It is further clear why Noether's theorem will delight and Einstein, who once struggled with the understanding of invariant movements, the inertial straight-lined and the body in a free fall in the gravitational field, believing that in all such laws of physics they remain the same. The Noether's theorem also guarantees and the stability of the gravitational field.

In quantum mechanics, we know *Heisenberg's*[12] *uncertainty relations* according to which the product of the uncertainty of energy and the duration of any real physical process or particle cannot be smaller than the known constants (approximately Planck's). Analogously applies to momentum and length. It is in the way that for (almost) any pre-given little duration we can always have enough energy and have a physically realistic system, which is a condition of the differentiability of the mentioned equations. The quantum world's symmetry is, above all, the reversibility of all the quantum processes, which is reflected around the (current) present, and then it is somehow valid for macro-world.

In short, whenever we have some kind of immutability, we have adequate stability. For example, the water that goes around in the cup will look the same to us, and that is

[10]Leonhard Euler (1707-1783), Swiss-Russian mathematician.
[11]Joseph-Louis Lagrange (1736-1813), Italian mathematician.
[12]Werner Heisenberg (1901-1976), German theoretical physicist.

rotational symmetry – and we have a rotational law of conservation (angular momentum). The whirligig once started will continue to rotate until a force (friction) is stopped it. The body in *inertial movement* does not change and we have a well-known law of inertia of straight line motion. In quantum physics, as I have said, all processes, the *evolutions*, are described by regular operators, reversible, which is a type of symmetry, and then for the flows of information, the law of conservation is valid.

Let's step up forward shortly in front of Noeter's theorem and notice that infinite sets can be equal (in quantity) to their proper subset (part), and that the principle of conservation (my theorem) does not apply to such. In other words, if for the given physical property is Neter's theorem, then for this property conservation is valid, and then it is finally divisible, in mathematics we say is discrete. Therefore, all forms of the substance are atomized, quantized, quarked, and the physical information is also always finally split, say *discrete*.

On the other hand, from the Euler-Lagrange equations, with the zero of the item which denotes the change of the physical state by time, we see the presence of the corresponding symmetry. This is the reverse course of the conclusion of Noether, which now in a different way gives us evidence of the *periodicity* of the material phenomena we have previously discussed.

Freedom, the amount of options measured by physical information, is also discrete and consumable. Such are our originality, our discoveries, and hence the development of society. If we would measure the legal restrictions analogously, the same would apply to the judicial system, which is consistent with the theory of information I have already explained. This takes us a step away from game theory, an important part of "informatics", but I will talk about later.

So there are women in mathematics and their contribution is not sporadic. They did appear less often, but they can be very bright.

1.15 Balance games

Should the balance of events be constantly maintained and should not be overestimated? Sometimes a colleague, a peacemaker, advises in this way, and then he continues that it is not worth to push the things, because everything will come back to its own, so he says, the truth will eventually win and why to try? There is something in it, however, in the politics of defeatism there is not enough truth. Here is my opportunity to explain this.

The theory of information I deal with is the mathematics of choice. Game theory is the mathematics of deciding. When we set them one by another, it will be shown that these are two related areas with much in common. The theory of games was created in 1928 when John von *Neumann*[13] discovered the *minimax theorem*. The proof of this theorem I'll now try to retell, but it should be known that it is very difficult in the original and that it is worth trying to understand it at least partially because its story is very abstract and therefore very universal. If you experience a single piece, you will see its reflections in many places around you!

Let's imagine that we have a player that has more options, tactics or strategies, each of which has some worst outcome. Suppose the opponent (one or more of them) can (with some probability) recognize the worst outcomes. Then it is best for a player to decide on a strategy with the most favorable outcome for him. That's some value, his "maximin".

[13] John von Neumann (1903-1957), Hungarian-American mathematician.

Conversely, symmetrically, the opponent works. He decides on his strategy and his own optimal "minimax" score.

If one of the values, maximin or minimax were better than the other, then that side would win, regardless of the impeccable game of the opponent. Such a jerk would be unfair, and the connoisseurs would be boring, and so to say the game is meaningless. On the contrary, if the values of maximin and minimax are equal, then the game is fair, it has symmetry, we say it is in equilibrium. The outcome is uncertain and the game can in the right sense last.

This great discovery of von Neumann now adds Emmy *Noether*'s theorem to the symmetric system of the corresponding law of conservation, maintenance of quantity, during other changes. Then, let us also consider (my recent) discovery that the property subject to the law of conservation must be *discreet.* The conclusion is that equilibrium games (when maximin and minimax are equal) must be discrete. Even when they at first glance look like analogous (continuous) ones, they actually have clearly separate moves in some of their micro worlds. Therefore, any physical information is discrete, freedom is discrete, prohibition of freedom is discrete, not only legal prohibition, but all that we have from direct natural laws are discrete.

This last is strange, since the minimax theorem assumes that the strategies are defined on *compact sets* of values, meaning closed domains, those containing all their limit values. However, the physical space due to Heisenberg's relations of uncertainty is "scarcely" such, since it can be divided by an arbitrary but predetermined division to the infinitesimals. Further, the information exchanges in the physical world are always in balance.

It is impossible to transmit information from something into nothing, and it is not possible the moving of, say, photon through a *vacuum* without communication. Namely, when the vacuum communication did not take away the photon's information, the photon information would have grown unlimited for "information about its past information" (quotation from my new book). That's why physical reality should be understood as a continuous rally of the physical, with all the "lots" of the substances playing (multiple) always in discrete moves and in the constant Neumann equilibria. The claim that the information provided is equal to the received is expressed by the law of conservation physical information. At the same time, physical communication is also a two-sided physical action.

With the law of energy conservation, the above question arises in the suspicion that "everything will come again" because energy is the work of the force on the road, and the energy has the power to change things. The body thrown up may be coming back to us, in its "path of truth", but forces can create disorders that have gone too far. The constants in these actions are, I paraphrase, that the biggest storms on the sea are just some of the *balance game* of nature, and that in these "game moves" of nature only truths are always exchanged.

The players who compete for the win have: the desire for victory, initiative and cunning; and more likely the winner is who leads the opponent to *defeatism* and to this extent realizes the own *aggressiveness*, that is, those who extract from the random game more information, that is, the *action of tactics.* The theory of games is further overwritten for such a short text, even for each of the mentioned indications of this theory.

Nature seeks non-action, through the *principles of least action* of theoretical physics, that is, through the *principle of least information*, because there is nothing that has no action, information or truth (they are synonyms). There is not even the smallest part of the nature without aggression, yet again it seems as if it is complaining about it. Do not fool yourself my friend with the defeatism, comes to me to say, because to keep only one side of the two is unnatural.

1.16 Parrondo's paradox

Sometimes by combining losing processes we can get a winner. It is a paradox of the theory of games discovered by Spanish physicist Juan Parrondo in 1996, after which he was named. *Parrondo's paradox* will be used to answer the recently asked question: is there any deeper connection between "that's yours" theory of information and the theory of games, that is, among them and us the ordinary living world?

Let us define two simple losing games and form a third winner of them, and then note that similar complex games the nature play all around us constantly. Then we connect this with the *principles of minimalism* of information and actions, noting besides that not all are the games on the victory.

Let's imagine the first game so that our player in any move unconditionally loses one euro. Just like that. If he has one hundred euros at the beginning of the game, after hundred moves he will not have them. Clearly, this is a losing game, and its simplicity makes it easier for us to continue the story. In the second game, count the amount of money the player has, so if the number is even add to him three euros, and if it is odd, he lose five euros. It's not hard to notice that this game is also a losing for itself. For two consecutive moves, the player loses two euros (because of three gains and loses five), so the starting hundred euros he also loses in a hundred moves.

Let's agree that our player alternately plays the second than the first game. In this combination, with an initial 100 euros, he is in the second game and earns three euros and rises to odd 103 euros, then loses one euro in the first game and falls to the even 102 euros. He plays the second game and increases his earnings by three euros to odd 105 to lose one in the first and stand at even 104 euros. In every two consecutive moves he is richer for two euros.

I hope that you do not bother to add three euros and subtract one during two moves and you can notice that our player is thus richer for two euros each time. After such a couple of moves, he is always on an even number of money, and he gets two euros all the time. This is a pure win game with alternate substitutions of two simple loser games. It is an abstract example of the aforementioned paradox of game theory, but which facilitates can help us in understanding the promised answer.

The first game is recognized in the conditions of regulation, in the stability, safety and efficiency of, say, companies or societies, viewed in the long run. A better organization, a pervasive hierarchy, can mean greater instant success of the company in competition, but greater stability is generally more static, and it is by the time a cause of lagging in relation to the changing environment, in relation to some "others" that appear and whose significance grows in time.

We define the second game as hasty innovation, over-accelerated, and not for four years, which could cost the company but can be exploited. This rush because of the excess of costs and lack of revenue will lead the company to losses. Contrary to the hustle and bustle of the second game and the constant slowing down of the first game, their combination, an innovation with periods of exploitation, of two losers would make the winners third game.

Random withdrawal of arbitrary game moves, which usually does not lead to gain, is defined as the *zero state* of the given game, and we compare the difference between mastery and randomness with physical information. It becomes the measure of *action of tactics*, the level of mastery, because each the physical information is a physical action (energy in duration). The definition of such a measure is also worthwhile in equilibrium games, where everyone gets, or everyone loses, because randomness is a universal "zero state".

The consequence of the new definition of tactics is, for example, a different view of *aggression* as a positive initiative – to the *optimum* of the inherent information. Namely, due to the objectivity of uncertainty, some kind of emission of information is inevitable, but because of the stinginess of information, they have their own optimums. In the theory of probability the uncertainty is a topic, and the *principle of parsimony* is seen in the more frequent realization of more likely events (which are less informative). In physics, where micro-effects are a permanent phenomenon and cannot be eradicated, the principle of parsimony can be seen in the need for force to cause macro-action, which is a non-spontaneous thing.

Analogously to the previous, an *initiative* of one company in competition with another means now a threat, an *action* that seeks a *reaction*, without which it makes the victory easier for the first competitor. Now the absence of opposition increases the chances of defeat, and this was not so explicit in classical theory. A similar example is an occupier that is coming to a new territory, which may be in interested in acting partially friendly and partially aggressively taking over more of the host. This example even more clearly emphasizes the importance of optimality.

Parrondo's paradox is mirrored in another way, again in *dualities* arising from the principle of minimalism (information and actions). Now we know that nature does not like excessive emissions of information analogous to spontaneous movement of the body along trajectories with minimal energy and time consumption, and further notice that its accumulation capacity results from this parsimony. The accumulation stimulates the evolution of living beings, and life itself is torn between leaking and acquiring information, that is the actions. We can say that the animate and inanimate worlds are captured by such streams, with the guards mentioned principles.

Accordingly, the life of ordinary mortals is "playing information". Whether or not we are aware of it, the competitions and communication make us.

1.17 Thermodynamics

The word *entropy* ($\acute{\varepsilon}\nu\tau\rho o\pi\acute{\eta}$ – turn inward) was introduced in physics by a German mathematician *Clausius* (Rudolf Clausius, 1822-1888). He analyzed the *Carnot cycle* (Sadi Carnot, 1796-1832), the French officers, engineers and physicists whose work is based on *thermodynamics*. Carnot's cycle is a theoretical physical process that observes circular changes in temperature and fluid pressure in a closed heat engine. The idea of entropy was further developed by *Boltzmann* (Ludwig Boltzmann, 1844-1906), interpreting it in 1870 as a measure of *uncertainty* in statistical mechanics, followed by the work of the American scientist *Gibbs* (Willard Gibbs, 1839 -1903) responsible for the transformation of physical chemistry into a rigorous inductive science.

Carnot devised an ideal heat machine, a thermodynamic cycle of maximum efficiency, with a cycle in four strokes. The first is isothermal (constant temperature) compression of the fluid (liquid or gas), then adiabatic (without heat change) compression, next isothermal expansion and the fourth stroke is adiabatic expansion. Adiabatic processes cannot be achieved in real terms, since at least small heat exchange with the external environment must exist, and with each such cycle, part of the energy of the system is lost irreversibly.

Not only in the imaginary ideal conditions, the optimal *work* of the heat machine

$$W = Q_2 - Q_1 \tag{1.1}$$

the difference is the largest (Q_2) and the least (Q_1) of the heat of the fluid in states respectively the largest (T_2) and at least (T_1) temperature. The specific change ($\Delta T = T_2 - T_1$) of the temperature in these two extreme states of the imagined cycle

$$\eta_0 = \frac{\Delta T}{T_2} = 1 - \frac{T_1}{T_2}, \tag{1.2}$$

is called Carnot's efficiency. *Kelvin* (Lord Kelvin, 1824-1907) showed by his works that the maximal work of a heat machine can produce the product of this coefficient (η_0) and the greatest heat (Q_2) of the fluid

$$W = \left(1 - \frac{T_1}{T_2}\right) Q_2, \tag{1.3}$$

and thus, by equating (1.1) and (1.3), Clausius found

$$\frac{Q_2}{T_2} - \frac{Q_1}{T_1} = 0, \tag{1.4}$$

that in the ideal process heat and temperature are directly proportional. He called the entropy the coefficient of heat and temperature

$$S_i = \frac{Q_i}{T_i}, \quad i = 1, 2 \tag{1.5}$$

which is in ideally conditions constant ($S_2 - S_1 = 0$).

In real terms, the optimal operation of the heat machine is less than imagined

$$Q_2 - Q_1 < \left(1 - \frac{T_1}{T_2}\right) Q_2, \tag{1.6}$$

and hence, orderly:

$$\frac{T_1}{T_2} Q_2 < Q_1, \tag{1.7}$$

more heat is transferred to the cold reservoir than in the Carnot cycle and

$$S_2 < S_1, \tag{1.8}$$

the entropy leaving the system is greater than the one that remains. Clausius' entropy in real terms spontaneously grows.

That's a historical look. For Clausius, the entropy was merely a quotient, a convenient substitution in his mathematical analysis of the Carnot cycle. Yet it gave the name for the entropy (turning inwards) alluding to the "something" that remains and increases as energy leaks out, which over time get more sense.

1.18 Statistical mechanics

In the development of entropy, Boltzmann introduced the hypothesis of elementary particles, atoms and molecules that move rapidly around their central points, vibrating by pushing each other and spreading the fluid as much as the vessels allow, occupying uniform positions. He noted that for a uniform arrangement of microstates (balls in boxes) there are more possibilities than any uneven distribution, and assuming that all combinations are equal, he found that the evenly are the most likely. The logarithm of these schedules Boltzmann recognized as Gibbs's *uncertainty*. The changes in the thermodynamic circular process he recognized as a change in Clausius' heat and temperature ratios. Today, in science, we are largely remembered him by this logarithm of the number of thermodynamic microstates, called Boltzmann's entropy, or as one of the founders of statistical mechanics.

In an *ideal cycle* of a heat engine, how many times the *heat* (energy) is reduced so many times the *temperature* of the fluid decreases, because their quotient is constant entropy, but also vice versa, when the heat increases the temperature increases proportionally.

In the real cycle, we have energy losses of *oscillating molecules* by transferring their higher oscillations to the lower oscillations of the cooler walls of the vessel. Consistent with the increase in the Clausus entropy, the quotients of heat and temperature (1.5), we now consider this as a decrease in temperature, more than as a decrease in heat. In addition to the kinetic energy of the molecule, otherwise the only one in an ideal cycle, in the reality there is also a potential energy bond between the molecules, overall less dominant than the temperature drop. Again, the thermal energy (numerator) of the cycle decreases, but the Clausius entropy (quotient) increases because the temperature (denominator) decreases faster.

Consider the same from the point of view of the law of conservation information. Imagine a larger system with the cycle and an environment so that the total information is closed. The total information is constant, so as much of the internal increases by (order) so much the external decreases (disorder).

This is another novelty. From the outside looking at an internal uniform arrangement we consider it impersonal, amorphous, less informative, which we perceive as the absence of order, and therefore we see an increase in entropy as an increase in disorder. It is at the expense of greater internal orderliness of the system! Molecules within the given cycle are arranged like lining up soldiers or evenly arranging balls into boxes.

The explanation of the aggregate information, internal and external, even if it is not the same before and after the process, does not impair the conformity of Boltzmann's statistical explanation of entropy with the Clausius definition. However, I believe that this view is true and that this will be reflected well in the continued application of entropy. Here I will only refer to the theory of relativity, especially the special one.

The coordinate system K' moves at a uniform velocity v with respect to the coordinates K. In each of the two systems, one as proper (own) observer which is in that system stationary and perceives the other in relative motion. In proportion to the Lorentz coefficient

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{1.9}$$

where $c \approx 3 \cdot 10^8$ m/s is the speed of light in a vacuum, relative energy rises, relative time slows, and relative lengths in the direction of motion shorten.

For example, if proper energy is E_0 (observed at rest), then relative (in motion) is

$$E = \gamma E_0 \approx E_0 \left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) = E_0 + \frac{1}{2} m_0 v^2 = E_p + E_k, \tag{1.10}$$

where $m_0 = E_0/c^2$ is its proper mass (at rest) and $E_p = E_0$ and $E_k = \frac{1}{2} m_0 v^2$ are the potential and kinetic energies of the given body. The Lorentz coefficient (1.9), taken as a function of the velocity quotient ($v^2/c^2 \to 0$), was developed in series and the sums of higher degrees of this quotient are neglected, since we consider the case of velocities v negligible with respect to speed of light c.

It is shown that the "total energy" E increases with motion due to the increase in kinetic energy, which gives us the right to (hypothetically) assume that the thermal energy in Clausius' entropy (1.5) does not change with motion, $Q = Q_0$. Moreover, in accordance with the above explanation of the "oscillation leakage" from the cycle vessel to the colder walls, we also add the assumption that with increasing energy of the oscillations increase the relative temperature of a given body $T = \gamma T_0$, so we conclude that the relative entropy is smaller than the proper ($S < S_0$), more precisely that $S = S_0/\gamma$.

Such interpretations are unusual but not contradictory. They have recently appeared in similar form to other authors (see [4]) and, of course, in my previous works (see [6]).

The increase in temperature can be "defended" by the Doppler Effect, which in the special theory of relativity has an additional transversal increase in relative wavelengths. They are equal to the arithmetic mean of the relative wavelengths of the incoming and outgoing sources and are proportional to the Lorentz coefficient (1.9). It's easy to check, so I'm not repeating it here.

The consequence of less relative entropy than the proper is the *law of inertia*. The body will not spontaneously transit from a state of greater entropy to a state of lower entropy and therefore remains in a state of relative rest and does not go into the moving system. It sees relative entropy of moving body as smaller and in the Boltzmann's sense too, because contraction is only in the direction of motion and not perpendicular to that direction, which disturb homogeneity.

Similar is the observation of less relative entropy in the gravitational field from the point of view of the weightless state of the free-fall satellite moving along geodesic lines. In a room that is stationary with respect to the gravitational field, lower air molecules are denser due to the attractive gravitational force, which will create the impression of *disturbance of homogeneity* and decrease of entropy. That is why the body in free fall does not abandon its path, as it would spontaneously switch from higher to lower entropy. This is also consistent with the Doppler effect of general relativity.

Let's summarize. By stopping abruptly, at the moment of collision with the obstacle, the body temperature is like just before the collision, the kinetic energy goes into heat and the entropy of the body increases. A glass in flight breaks down only when it hits an obstacle according to the increase in disorder due to the increase in entropy.

Chapter 2

Formalism

Here we are in the field of mathematics, mathematical theory of physical information. We will assume that the classical information theory is well-founded. Then we'll separate and emphasize some distribution cases that can be subdivided into additive parts. These parts will have information that by simple summing gives "physical information" to larger entities, which no longer has Shannon's form, although it may look so at the beginning. Such resolution is not possible in all information systems, for not all are separable to independent entities, and in addition, neither *continuum* is not a physical matter. These are guidelines. Here are just some of the probability distributions as the beginning and the model of the future theory.

The root of all is Hartley's information, as it is, and then seeking the expression of physical information as close to Shannon's definition as possible. What was the goal is information as an additive function of independent probability distributions of a similar form. This is an idea that would be in keeping with the law of conservation, and as you will see, it is elaborated in the case of a well-known binomial distribution and a few less popular with it. Indicators and the free walk of the probability theory, discussed here, are not new, but I've gone so fast to the unknown positions, and, above all, I've proved them "superfluous" in different ways, that the text might be tense even for mathematicians. That's why there is not much of them.

On the other hand, the topic itself is unpopular. Now in trendy is the entropy not considered as information, which has its strong reasons and maybe disproportionate number of followers. For example, the logarithm of the number of equal micro-states of thermodynamics (Boltzmann's formula of entropy) increases with the number of micro-states and decreases with the probability of one, so it seems that spontaneous entropy growth is equivalent to a more frequent occurrence of less probable states. This, of course, is not true, but the precondition of this deduction is also not true, since "equal distribution" is special and most likely of many.

Moreover, in my theory of information, entropy is not only a statistical measure of the "disorder" of a particle system, but it is also a measure of the "feminization" of the physical information system. Editing the inside of the system shows the facelessness to the outside, and this way of looking at the spontaneous growth of entropy is more general than the transfer of heat from the body of higher temperatures to the colder ones, that it will be invisible to physicists for years. Step by step, when investigating all (in) possibilities of modern entropy, I believe that there will be emerge a new aspect. Either way, do not give up because of the entropy for it's just in the hint here.

2.1 Hartley's information

When we have $N = 1, 2, 3, \dots$ outcomes with equal opportunities, theirs *information* is

$$H = \log_b N, \tag{2.1}$$

where the logarithm base ($b > 0$ and $b \not{1}$) determines the information unit. When $b = 2$ the information is measured in bit, when $b = 10$ the information is in decit, and when $b = e \approx 2.71828$ we express it in the nat. Measurement of information of equilibrium events by logarithm was established by *Hartley* (Ralph Hartley, 1888-1970) in 1828.

In the case of $b = 2$ at an interval from $N = 4$ to $N = 16$, the Hartley function (2.1) is approximately equal to the root ($H \approx \sqrt{N}$) as it is seen on the figure 2.1. The blue is the logarithmic function ($y = \log_2 x$), the red is root ($y = \sqrt{x}$).

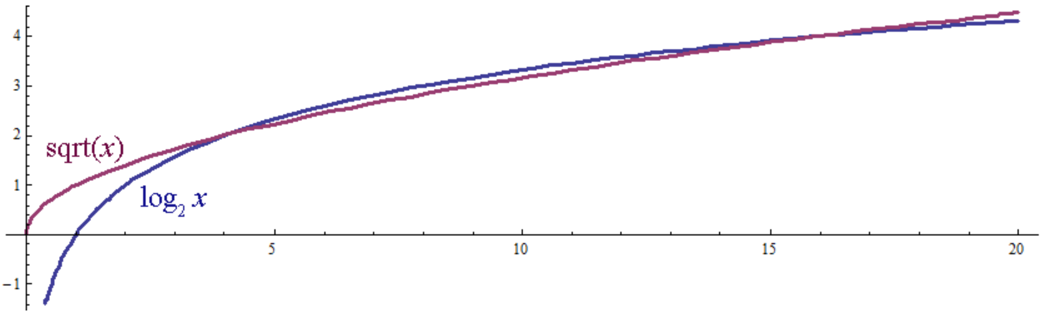

Figure 2.1: Logarithmic and root functions.

This coincidence, between the logarithmic and the root functions, uses our senses to collect *stimulus* (stimuli) that are further interpreted into information by means of receptor surfaces. A *threshold* stimulus is the smallest amount of energy (mechanical, light, thermal, chemical) that the receptor registers and on which organism can reacts. The differential threshold is the smallest change in the irritation which we observe[1].

The German doctor *Weber* (Ernst Heinrich Weber, 1795-1878) and his student *Fechner* (Gustav Fechner, 1801-1887) noted that the differential threshold was proportional to the energy already given. When we hold a 50 gram stone in our hand and it is necessary to add at least 1 gram to see the difference in weight, then a 100 gram stone has a differential threshold of 2 grams.

Weber's quotient (differential energy threshold divided by energy) for load lifting is 1:50, for pressure (touch on the skin) it is 1:7, for heat on the skin 1:30, for vision 1:60, for volume 1:10, for smell (tires) 1:4, and for taste (kitchen salt) 1: 3. Adding (integrating) changes yields a sum of all observations that a given sense can have in the form of Hartley's logarithm (2.1).

Unlike an irrigation surface whose length can be used for an approximate estimation of information, the logarithm expresses a deeper feature of information and its relation to probability. An important feature of physical information is the *conservation law*, that it is plastic and like energy that it can be transformed from shape to shape without changing the amount. This feature supports the logarithm of its so-called additivity.

Namely, the logarithm of the product is equal to the sum of the logarithms

$$\log_b MN = \log_b M + \log_b N, \tag{2.2}$$

[1]quotes from the book [3]

so if independent events individually have $M, N = 1, 2, 3, \dots$ outcomes then they will have a MN outcome in the pair. If their outcomes are equally probable, then this term expresses the law of maintaining (conservation) information. For example, throwing a coin (with two outcomes) in a pair with a throw-outs dice (with six outcomes) will have $12 = 2 \cdot 6$ of outcomes, and due to

$$\log 12 = \log 2 + \log 6$$

we can say that the information of the independent events is summed up.

The following example also confirms the law of information maintenance, but also points out that the amount of uncertainty (measured by Hartley's formula) is a type of information. Have a box containing $N \in \mathbb{N}$ of equal balls given. When we draw them one by one, the probability of pulling one when they were $n = 1, 2, \dots, N$ was $P_n = 1/n$ at the time of drawing, and the amount of uncertainty before the draw-down is equal to the information after, and both amounts $H_n = -\log_b P_n$. Total extracted information is:

$$H = H_N + H_{N-1} + \dots + H_2 =$$

$$= \log_b N + \log_b(N-1) + \dots + \log_b 2 + \log_b 1$$

$$= \log_b[N \cdot (N-1) \dots 2 \cdot 1],$$

and this is exactly the same as in (2.1), since $H = \log_b(N!)$, which is the total information of all possible pull-outs of one-to-one given balls from the box.

The next important feature of physical information is *principle of information*, which says that less informative events are more preferable, because they are more likely, because more likely events are more frequent. We can hardly notice this property even though its consequences are everywhere around us. Information is a matter of truth, and the truth can also be obtained from a lie (negation, implication), so the Internet falsity spreads several times faster than the truth. Fairy tales, in general fiction, but also discussions, are more interesting than geometric theorems, because even *deduction* can be accurate when starting from an incorrect assumption and leading to an inaccurate consequence. Encoding is easier than decoding.

Physical information is also a matter of *action*, and the physical world is ruled by the so-called the "principle of least action". Due to the equivalence of these two, this well-known principle of theoretical physics now becomes the "principle of least information". Presented in this form can also be extended to physics from biology and sociology.

In the technique we have something similar since 1948, when *Shannon* (Claude Shannon, 1916-2001) estimated the maximum amount of *information* that can pass through a channel without error (see [13]), by formula

$$C = B \log_2\left(1 + \frac{S}{N}\right). \tag{2.3}$$

There is C *channel capacity* in bits per second, the theoretical upper limit of the correct transmission. B the *bandwidth* is the frequency of the amount of data that can be transmitted in a fixed amount of time in hertz. The S signal is the average received signal strength over the given width, measured in watts (W), i.e. in joules per second, or volt amps. N the noise and interference power is the average unwanted "sound" judged to be unpleasant, loud or disruptive to hearing at a given bandwidth, also in watts. The S/N quotient is the signal-to-noise ratio of a given communication channel.

In order to double the transfer, the channel capacity (C), it is not enough to double the signal (S), it is necessary to square the expression in the brackets of the logarithm. By exponentiation the numerous, capacity increases, but only in linear form, which is also a type of interference in the transmission of information.

2.1.1 Examples

Example 2.1.1 (Guessing the number)**.** *Someone chooses an integer of x from 1 to N. We can ask him the question "Is this number greater than ...?" What is the minimum number of questions needed to identify the required number x?*

Solution. When $N = 2$ it is enough only one question: "Is x greater than one?". If the answer is "no" then $x = 1$, and if "yes", $x = 2$.

When $N = 16$ there are four questions enough. Let's say that the imaginary number is $x = 7$. The first question is, "Is the imaginary number greater than 8?". The answer is "no". The second question is: "Is the imaginary number larger than 4?". The answer is "yes". The third question is: "Is the imaginary number greater than 6?". The answer is "yes". The fourth question is: "Is the imaginary number larger than 7?". The answer is "no". Therefore, the imaginary number is $x = 7$.

In general, the smallest number of questions required by this *binary search* to detect an imaginary number $1 \le x \le N$ is $\log_2 N$ bits. It's Hartley's information. □

Example 2.1.2 (Searching damaged)**.** *Among the N equal coins there is one that is damaged, so it is lighter than the others. We have balance scale to compare the weight of the two groups of coins, so that we find out if the first group is lighter, equal or heavier than the other group. How much is the least weighing needed to find the damaged coin?*

Solution. When $N = 3$ one weighing is enough. We put the first and second coins on different plates, so if the first one is easier, this is the one, if the other is easier it is, and if the same weight is the third, that is.

When $N = 81$, there are enough four weights. In the first weighing we divide all coins into three groups of 27, so we find a lighter group in the previous way. In the second weighing, this lightweight group is divided into three groups of 9 coins and in the same way we find it easier. In the third weighing, the group we divided into three groups of 3 coins, so we use the same method to find easier. In the fourth weighing, this group is divided into three coins and found easier in the described way.

In general, the smallest number of weights required to find the corrupted coin by this *triad search* is $\log_3 N$. This is Hartley's information in the base units $b = 3$. □

Example 2.1.3 (Notation numbers)**.** *How many position numbers do you need to write a N in n-are notation?*

Solution. When is $n = 2$ we write binary with basic digits 0 and 1. We have $2^1 = 2$ options for the first two numbers, zero and one. We have $2^2 = 4$ of a two-digit number: 00, 01, 10 and 11, and these are zero, one, two and three. It is also understandable that n binary positions contain 2^n numbers, and if this is the number N, then $n = \log_2 N$.

When $n = 3$ we write in base 3 with the basic digits 0, 1 and 2. The single-digit numbers then have $3^1 = 3$, two-digits $3^2 = 9$, and in general n-to digits $3^n = N$ where one-digit, two-digit etc. are included, because on the left they can be zero, so $n = \log_3 N$.

When $n = 10$ we write decadal. It has $10^1 = 10$ decade single-digit numbers, and these are known digits 0, 1, ..., 9. It has $10^2 = 100$ decades (to the highest) two-digit numbers. In general, it has $10^n = N$ decades (up to) n-to digits, and hence $n = \log N$. It is Hartley's information of N of equal outcomes in decits.

Analogously, in the n base number for the N record, a $\log_n N$ position is required. □

2.2 Shannon information

When the quotient $x = S/N$, signal and noise in the Shannon's formula (2.3), of the given channel (signal-to-noise ratio) is smaller than one, transmission of *information* is considered bad. Only then, from the *Taylor's* (Brook Taylor, 1685-1731) development of the logarithmic function

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \dots, \quad -1 < x \le 1, \tag{2.4}$$

we find that the transfer (C) increases (approximately) linearly with the average received signal strength (S). However, this does not contradict the aforementioned principled nature of information skimping, to which should be added tolerance or inevitability in some small information emissions. This last complies with the conservation law and the assumption that there is some kind of *objective coincidence*.

Namely, if in every order of magnitude the nature would deduct with information, then the division of information into less and less parts could go into *infinity*. Then the law of conservation would not be valid, since an infinite set by definition (in quantity) is equal to its proper (or strict) subset, such is set of natural numbers proper subset of set of integers ($\mathbb{N} \subset \mathbb{Z}$). In particular, if any information could be shredded to infinite, then the assumption of the existence of an event without a cause is difficult to sustain.

From our general point of view, an event that has no cause and an event to which we can never find the cause are equivalent. Moreover, from an individual's point of view, such a random event is equivalent to an event to which the individual cannot know the cause. In other words, the *uncertainty* is a relative phenomenon. Individuals (person, particle) communicate by exchanging what they do not have, in particular, taking what they could not have in a given ambiance (in a given place at a given moment, in the given environment) without a current exchange. It is the meaning of communication that I mean not only in the above-mentioned event devoid of cause.

Consider, for example, a random event that can be realized as a "favorable" with the probability $P \in [0,1]$ or in "unfavorable" with the probability $Q = 1 - P$. It is formally equivalent to randomly extracting one of $N \in \mathbb{N}$ of equal balls, of which $K = 1, 2, \dots, N$ are designated as favorable and the remaining NK as unfavorable. The likelihood of pulling a "good" ball is $P = K/N$, and "not-good" $Q = (N-K)/N$, and again $P + Q = 1$. Hartley's information of the first and the second, respectively, is:

$$H_P = -\log_b P, \quad H_Q = -\log_b Q, \tag{2.5}$$

so the mathematical expectation (mean) of them is

$$S = -P\log_b P - Q\log_b Q. \tag{2.6}$$

It's Shannon *binary information*.

When we put $P = x$, $Q = 1 - x$, $S = y$ and $b = 2$, the expectation of the given event is the function $y = y(x)$ expressed in bits whose graph in Cartesian rectangle coordinate system Oxy is represented in the figure 2.2. It is seen that this information is maximal ($y = 1$) when the choices of favorable and unfavorable events are equal ($x = 0.5$), and in the limit cases it turns out that $y \to 0$ when $x \to 0$ or $x \to 1$.

In general, arbitrary logarithm bases $b > 0$ and $b \ne 1$, for $x \in (0,1)$ the corresponding Shannon binary function is

$$y = -x\log_b x - (1-x)\log_b(1-x). \tag{2.7}$$

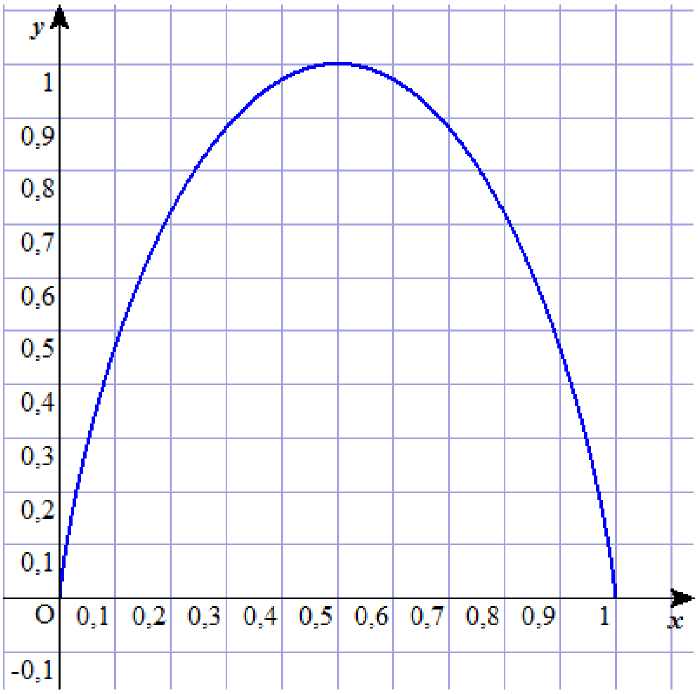

Figure 2.2: Shannon binary information.

That its graph is also symmetric to the line $x = 0.5$ is seen from the invariant on the substitution $x \to 1-x$. The continuity follows from the continuity of the logarithmic function, and by derivative we can find the maximum $y(0.5) = \log_b 2$.

Shannon's information of the discrete (at most countable) infinite set of random probability events $p_k \in (0,1)$ for $k = 1,2,\dots,n$ is

$$S = -\sum_{k=1}^{n} p_k \log_b p_k, \quad \sum_{k=1}^{n} p_k = 1, \tag{2.8}$$

where it can be $n \to \infty$ if the specified series converge. Maximum information $S = H = \log_b n$ is achieved when all the outcomes are equally probable.

We also denote these forms of information with $S(p_1,\dots,p_n)$. We take an arbitrary index r, an integer from 1 to n, then divide all the given events into two: the first of probability $q_1 = p_1 + \dots + p_r$ and the second of probability $q_2 = p_{r+1} + \dots + p_n$. It is then

$$S(p_1,\dots,p_n) = S(q_1,q_2) + q_1 S(\frac{p_1}{q_1},\dots,\frac{p_r}{q_1}) + q_2 S(\frac{p_{r+1}}{q_2},\dots,\frac{p_n}{q_2}). \tag{2.9}$$

Namely,

$$S(q_1,q_2) = -q_1 \log_b q_1 - q_2 \log_b q_2,$$

$$q_1 S(\frac{p_1}{q_1},\dots,\frac{p_r}{q_1}) = -p_1 \log_b p_1 - \dots - p_r \log_b p_r + q_1 \log_b q_1,$$

$$q_2 S(\frac{p_{r+1}}{q_2},\dots,\frac{p_n}{q_2}) = -p_{r+1} \log_b p_{r+1} - \dots - p_n \log_b p_n + q_2 \log_b q_2,$$

and the sum of these three equations is the given (2.9).

Theorem 2.2.1. *Property (2.9) together with the property*

$$S(\frac{1}{n},\dots,\frac{1}{n}) = H(n) = \log_b n, \tag{2.10}$$

defines Shenon's information (2.8).

Иначе, Хартлијева информација $H(n)$ је максимална вредност Шенонове информације (2.8) када је свака од датих вероватноћа константна, $p_k = \frac{1}{n}$ редом за $k = 1,\dots,n$.

Proof. We use the method of *mathematical induction.* For $n = 2$ and $p_1 = p_2 = \frac{1}{2}$ from (2.9) follows

$$S(\frac{1}{2},\frac{1}{2}) = S(\frac{1}{2},\frac{1}{2}) + S(1),$$

from where $S(1) = 0$, which means that the statement is true for $n = 1$.

From (2.9) follows, in the rows:

$$S(\frac{1}{n},\dots,\frac{1}{n}) = S(q_1,q_2) + q_1 H(nq_1) + q_2 H(nq_2),$$

$$S(q_1,q_2) = S_0(n) - q_1[H(n) + H(q_1)] - q_2[H(n) + H(q_2)],$$

$$S(q_1,q_2) = -q_1 H(q_1) - q_2 H(q_2),$$

where $H(q) = \log_b q$ according to (2.10). This is the statement of the theorem proved for $n = 2$.

Assume that the theorem is valid for given $n = 1,2,\dots$ and prove that it is also valid for $n+1$. We put $q_1 = p_1 + \dots + p_n$ and $q_2 = p_{n+1}$. According to (2.9) it is then

$$S(p_1,\dots,p_n,p_{n+1}) = S(q_1,q_2) + q_1 S(\frac{p_1}{q_1},\dots,\frac{p_n}{q_1}) + q_2 S(1).$$

We have proved that $S(1) = 0$ and $S(q_1,q_2) = -q_1\log_b q_1 - q_2 \log_b q_2$, and it is assumed that

$$S(\frac{p_1}{q_1},\dots,\frac{p_n}{q_1}) = -\frac{p_1}{q_1}\log_b\frac{p_1}{q_1} - \dots - \frac{p_n}{q_1}\log_b\frac{p_n}{q_1}.$$

Taking all this into account we get

$$S(p_1,\dots,p_n,p_{n+1}) = -p_1\log_b p_1 - \dots - p_n \log_b p_n - p_{n+1}\log_b p_{n+1},$$

which means that the theorem is true for $n+1$, and according to the principle of mathematical induction it is valid for every natural number n. □

These are known[2] the properties of Shannon information, and they are inherently interesting, but from which it follows that it does not support the properties of physical information. Shannon's information expresses the average value of Hartley's information within a certain distribution, and only within that framework. I recall that if $\omega_1,\dots,\omega_n$ are random events of some distribution, with the corresponding probabilities $p_1,\dots,p_n$, then they make a complete set Ω so that $p_1 + \dots + p_n = 1$.

Therefore, one (any) of the given probabilities can be expressed by others, which in itself diminishes the uncertainty of the event ω_k, and hence the Shannon information. The

[2]see [11], 3rd theorem

expression (2.8) is the underestimated value of physical information in cases of the final $n > 1$. We will see more of this later.

We will retain the method of calculating mathematical expectations by defining it also in relation to some distribution, but Hartley's information of its outcomes will be chosen differently from that of Shannon. Specifically, we'll consider the discrete case $n \to \infty$, but also the continuum, or the probability density, so that in addition to the conservation law we could adopt the principle of information as part of the new formula.

2.2.1 Entropy

As we know from statistical physics, Hartley's information (2.1) is proportional to *Boltzmann* (Ludwig Boltzmann, 1844-1906) *entropy*, when it is assumed that N is the number of the micro-states of the thermodynamic system. That is why classical information, and therefore Shannon's, is often referred to in literature as entropy (see [8]). We will keep this name here as well whenever this doesn't leads to confusion with the same expression of physics.

Shannon introduced the labeling of average *mutual information* between the two processes, the random variables X and Y

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \tag{2.11}$$

as the sum of two own entropies minus the entropy of the pair. He pointed to this on the basis of the theorem on encoding and communicating multiple separate random processes through a channel with noise and general coding theorems. The first theorem focuses on the detection of transmission errors, the other on an analog-to-digital conversion, and on data compression. Special cases of both coding are in Shannon's original work [14].

The average mutual information is also defined by conditional probability, then also by conditional entropy $H(X|Y) = H(X,Y) - H(Y)$, from where

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{2.12}$$

In this form, the mutual information is the difference between the information of the main process and the process contained therein, when the other one is known. It follows that the information of the random variable does not change by repetition, $H(X,X) = H(X)$, and because of (2.11) and

$$I(X,X) = H(X), \tag{2.13}$$

why entropy is considered a special case of average mutual information.

For classical theory, our *principle of information* is a novelty, although it is obvious. Since the probability theory has been made, it is assumed that more likely random events are more likely to be realized, then it is noticed that the greater the news is it is rarer, or less likely, and yet it has been missed by researchers to notice that more informative events are happening less often. This simple observation, that nature is stingy with the emission of information, is the principle of information. Could it be that it is neglected by giving greater importance to entropy in the interpretation of information?

Boltzmann's entropy is paradoxical. The second law of thermodynamics, which says that heat spontaneously passes from the body of higher temperatures to the body of a lower temperature, which is equivalent to a spontaneous growth of entropy, which grows with the logarithm of the number of micro-states ($\log N$), at first glance is as if we were talking about an increase of entropy with a decrease in probability ($1/N$). So rapidly watching is ignoring the fact that a higher number of micro-states (N) is achieved by a more uniform arrangement (molecules, particles, balls in boxes) which is actually more likely[3]. It's a matter of combinatorics. Accordingly, Boltzmann's entropy and information are striving to more likely outcomes; they both follow the principle of information.

This observation is a warning that it should not be easy to join Hartley's information formula with Boltzmann's entropy. On the other hand, their connection cannot be ignored either. The two of them calm the formula (2.13) of increasing entropy with the increase in average mutual information, that is, about *feminization* of the growth of entropy. It speaks about the rise of internal communication at the expense of external.

2.2.2 Examples

Example 2.2.2 (Facsimile)**.** *What is the information of one page of facsimile?*

Solution. Page[4] for the transfer consists of the points represented by the binary digits (1 for black, 0 for the white dot). The resolution is 200 dots per inch (2.54 cm), or 4×10^4 dots per square inch. Therefore, the number of binary digits needed to represent the page is $8.5 \times 11 \times 4 \times 10^4 = 3.74$ Mbit.

With a 14.4 kbps modem (kilobytes per second), the transmission of such a page takes 4 minutes and 20 seconds. Thanks to coding techniques, transmission time can be reduced to 17 seconds! □

Example 2.2.3 (Music)**.** *How many hours of MP3 music contains one CD ROM?*

Solution. One CD ROM, which has a capacity of 650 Mbytes (Mega Byte), contains more than 10 hours of MP3 stereo music.

Namely, the music analogue CD quality signal, with left and right channels, is sampled at 44.1 Khz (kilo hertz), and each plot has 16 bit (bits). One second of stereo music generates $44.1 \times 10^3 \times 16 \times 2 = 1,411$ Mbit. One byte is eight bits, one minute is 60 seconds, one hour has 60 minutes, so calculate. □

Example 2.2.4 (Cards)**.** *Two playing cards are simultaneously drawn from the 32-card deck. Let A be an event that at least one card is red, and B does that one of them is the king spade. How much is the information $I(A,B)$?*

Solution. The probability of event A and probability of A under condition B are:

$$P(A) = \frac{16}{32}\cdot\frac{15}{31} + 2\cdot\frac{16}{32}\cdot\frac{16}{31} = \frac{47}{62}, \quad P(A|B) = \frac{16}{31},$$

and because of (2.1) and (2.12) we obtain:

$$I(A,B) = \log_2\frac{62}{47} - \log_2\frac{31}{16} = -0,5546 \text{ bit}.$$

Because event B makes A less likely, the reciprocal information is negative. □

Example 2.2.5 (Alfabet)**.** *Find entropy H, in bits, text with alphabet letters of probability $p_1 = \Pr(A) = \frac{1}{4}$, $p_2 = \Pr(B) = \frac{1}{8}$, $p_3 = \Pr(C) = \frac{1}{2}$ and $p_4 = \Pr(D) = \frac{1}{8}$.*

[3] see [9], Figure 3.2 and explanation.

[4] More about this on the site `http://www-public.imtbs-tsp.eu/`

Solution. We use the Shannon formula (2.8), we get:

$$H = -\sum_{k=1}^{4} p_k \log_2 p_k = -\frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{2}\log_2\frac{1}{2} - \frac{1}{8}\log_2\frac{1}{8} = 1,75 \text{ bit},$$

because the $\log_2 2^x = x$. $\square$

2.3 Binomial distribution

Bernoulli (Jacob Bernoulli, 1654-1705) or *binomial distribution* simply states the probability of success or failure of the outcome of a random experiment that is repeated. It is a type of distribution with two possible outcomes, so the prefix "bi" means two or twice.

It's not a novelty that every *polyvalent logic* (true, maybe, false) can be reduced to a dual point (true, false) and that any random event can be reduced to a "favorable" and "unfavorable" outcomes. When we consider such a case, in the probabilities respectively p and q, where $p + q = 1$, then we can speak of the simple binomial distribution $\mathcal{B}(1, p)$. The complex binomial distribution $\mathcal{B}(n, p)$ is obtained by repeating the same $n = 1, 2, 3, \dots$ times with constant probability $\Pr(\text{favorable}) = p$ of favorable outcome and probability $q = 1 - p$ of unfavorable. We assume that the solo $\mathcal{B}(1, p)$ in the complex complex $\mathcal{B}(n, p)$ binomial distributions are independent events[5].

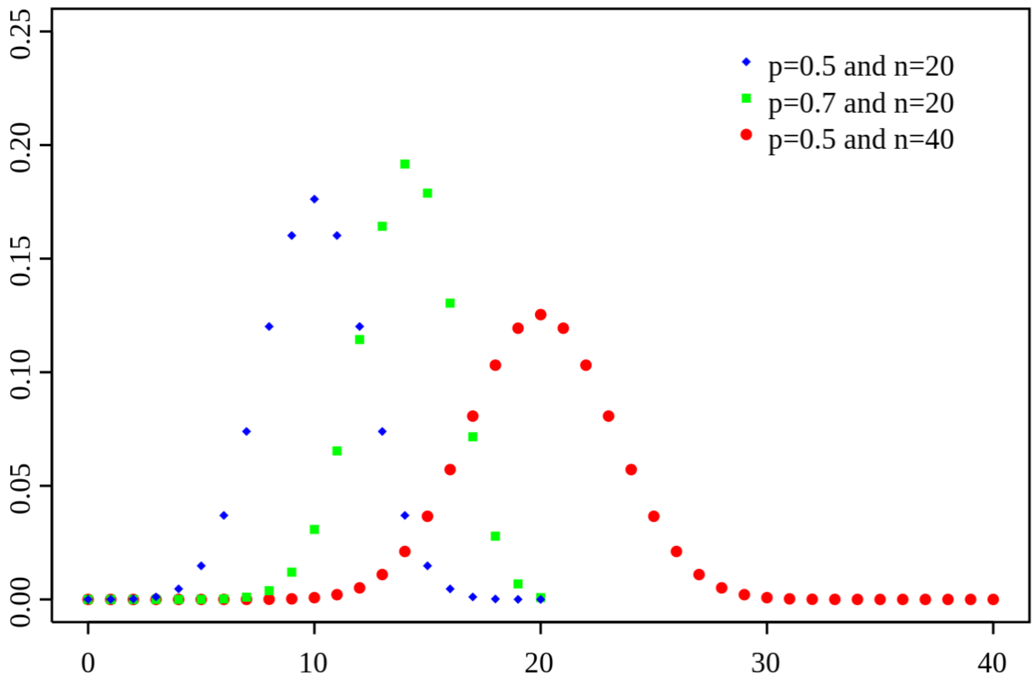

Figure 2.3: Binomial distributions $\mathcal{B}(n, p)$.

In the binomial distribution $\mathcal{B}(n, p)$, the random experiment repeats $n \in \mathbb{N}$ times independently with a constant probability $p \in (0, 1)$. A discrete random variable X represents the number of successes, the favorable outcomes in that repetition series, each of a particular probability p. The probability of one of the sequences with $k = 0, 1, \dots, n$ of success in n realizations is

$$p_k = p^k q^{n-k}, \tag{2.14}$$

where $q = 1 - p$ is the probability of a particular failure. The probability of all sequences with exactly k successes in n realization is

$$P_k = \binom{n}{k} p^k q^{n-k}, \tag{2.15}$$

[5]We follow the attachment [2].

where the binomial coefficient $\binom{n}{k} = n!/k!(n-k)!$ is the number of subsets of k elements taken from the set of n elements. That the numbers (2.15) make the distribution of the probability follows from

$$\sum_{k=0}^{n} P_k = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1, \quad P_k \in (0,1). \tag{2.16}$$

The probability P_k form graphs in the figure 2.3. That the numbers (2.14) do not make the distribution is following from:

$$\sum_{k=0}^{n} p_k = q^n + pq^{n-1} + p^2 q^{n-2} + \dots + p^n = \frac{q^n - p^n}{q-p} \neq 1, \tag{2.17}$$

although they obviously are some probabilities.

The mean value, or *mathematical expectation* distribution of the X variable whose value x_k is realized with the probability p_k, is the

$$\langle X \rangle = \sum_k x_k p_k, \tag{2.18}$$

in Dirac's notation, popular in quantum mechanics. The following theorem holds for the expectation of the binomial distribution.

Theorem 2.3.1. *Mathematical expectation $\mathcal{B}(n,p)$ is*

$$\mu = \langle k \rangle = \sum_{k=0}^{n} k P_k = np,$$

where the probability of the random variable k is $\binom{n}{k} p^k q^{n-k}$, according to (2.15).

Proof. Calculate, orderly:

$$\mu = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} p \frac{\partial}{\partial p} (p^k q^{n-k}) =$$

$$= pq^n \frac{\partial}{\partial p} \sum_{k=0}^{n} \binom{n}{k} \left(\frac{p}{q}\right)^k = pq^n \frac{\partial}{\partial p} \left(1 + \frac{p}{q}\right)^n = npq^n \left(1 + \frac{p}{q}\right)^{n-1} \frac{1}{q}$$

$$= npq^n \frac{(p+q)^{n-1}}{q^{n-1}} \cdot \frac{1}{q} = np(p+q)^{n-1} = np,$$

because $p + q = 1$. By this the theorem is proved. □

Variance $\sigma^2 = \langle (X-\mu)^2 \rangle$ is the mean square deviation of the random variable X from its expectation μ. In the general case, the following lemma (small theorem) applies.

Lemma 2.3.2. *Let X be a discrete random variable. It was then*

$$\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2,$$

i.e. the variance is also the expectation of the square X minus the expectation of X squared.

Proof. Followed by:

$$\sigma^2 = \sum_x (x-\mu)^2 \Pr(X=x) = \sum_x (x^2 - 2\mu x + \mu^2)\Pr(X=x) =$$

$$= \sum_x x^2 \Pr(X=x) - 2\mu \sum_x x \Pr(X=x) + \mu^2 \sum_x \Pr(X=x)$$

$$= \sum_x x^2 \Pr(X=x) - 2\mu\cdot\mu + \mu^2 \cdot 1$$

$$= \langle X^2\rangle - \langle X\rangle^2.$$

By this the lemma is proven. □

In the case of the binomial distribution $\mathcal{B}(n,p)$ with the random variable $k = 0,1,2,\dots,n$ and the natural number n of the experiments in which in each favorable outcome has a constant probability $p \in (0,1)$ and the unfavorable probability $q = 1-p$, the following theorem holds.

Theorem 2.3.3. *For $\mathcal{B}(n,p)$ the variance is $\sigma^2 = npq$.*

Proof. First we calculate:

$$\langle X^2\rangle = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} kn\binom{n-1}{k-1} p^k q^{n-k} = np\sum_{k=1}^{n} k\binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}.$$

Putting $j = k-1$ and $m = n-1$ we find further:

$$\langle X^2\rangle = np\sum_{j=0}^{m}(j+1)\binom{m}{j}p^j q^{m-j} =$$

$$= np\left[\sum_{j=0}^{m} j\binom{m}{j}p^j q^{m-j} + \sum_{j=0}^{m}\binom{m}{j}p^j q^{m-j}\right] = np\left[\sum_{j=0}^{m} m\binom{m-1}{j-1}p^j q^{m-j} + \sum_{j=0}^{m}\binom{m}{j}p^j q^{m-j}\right]$$

$$= np\left[(n-1)p\sum_{j=1}^{m}\binom{m-1}{j-1}p^{j-1}q^{(m-1)-(j-1)} + \sum_{j=0}^{m}\binom{m}{j}p^j q^{m-j}\right]$$

$$= np[(n-1)p(p+q)^{m-1} + (p+q)^m] = np[(n-1)p+1] = n^2p^2 + np(1-p).$$

How is (lemma 2.3.2):

$$\sigma^2 = \langle X^2\rangle - \langle X\rangle^2 = np(1-p) + n^2p^2 - (np)^2 = npq,$$

so the theorem is proved. □

Dispersion (standard deviation) of the binomial distribution $\mathcal{B}(n,p)$ is

$$\sigma = \sqrt{npq}. \tag{2.19}$$

The dispersion is the name for the root of variance (mean square deviation). In the extreme case, say $n > 50$, the binomial distribution for a small expectation ($\mu = np < 10$) goes to Poison's, and otherwise ($\mu \ge 10$) into the normal, called Gaussian, distribution.

In the classical theory of information, the logarithm of the dispersion ($S = \ln\sqrt{2\pi\sigma^2}$) is the Shannon's normal distribution, which means that in the limit case this dispersion acts as a set of N in Hartley's information (2.1). When it passes to the continuum of probability it is calculated using the integral of density, and physical information loses its meaning. Due to the law of conservation, that every property of physical information is discreet.

2.3.1 Approximations

Let us consider that the Bernoulli (binomial) distribution $\mathcal{B}(n,p)$ is given by the number $n = 1,2,3,\dots$ repetition of independent events whose favorable outcome has the probability p, and unfavorable $q = 1-p$. The number of favorable outcomes $k = 0,1,2,\dots,n$ in a given number of repetitions of the random test has the probability

$$\Pr(B,k) = \binom{n}{k} p^k q^{n-k}, \tag{2.20}$$

with the mean value and variance:

$$\mu_B = np, \quad \sigma_B^2 = npq. \tag{2.21}$$

When n is large, let's say $n > 50$ and p small, so $np < 10$ then Bernoulli is reduced to the *Poisson distribution*, otherwise it goes over to Gauss's (normal) distribution.

When n is large and p is so small that $np < 10$, then we have:

$$\lim_{n\to\infty} \binom{n}{k} p^k q^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}, \tag{2.22}$$

where $\lambda = np$.

This is proved by:

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k},$$

$$\Pr(B,k) \approx \begin{cases} \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n & \text{if } k \text{ small compared to } n, \\ \frac{\lambda^k}{k!} e^{-\lambda} & \text{if } n \text{ is big.} \end{cases} \tag{2.23}$$

Exponential factor comes from:

$$\ln\left(1-\frac{\lambda}{n}\right)^n = n\ln\left(1-\frac{\lambda}{n}\right) = n\left(-\frac{\lambda}{n} - \frac{1}{2}\frac{\lambda^2}{n^2} - \dots\right) \approx -\lambda,$$

if $\lambda/n \approx 0$. Thus Bernoulli's distribution is reduced to Poisson, in both cases, when $p \approx 0$, as well as when $p \approx 1$, since then $q \approx 0$.

In general, when $p > 0$ and when $n \to \infty$, then

$$\Pr(B,k) \to \frac{1}{\sqrt{2\pi npq}} \exp(-\frac{x^2}{2}), \quad x = \frac{j-np}{\sqrt{npq}}, \tag{2.24}$$

uniformly in x at each finite interval. This is de Moivre–Laplace *approximation* of Bernoulli probabilities (see [12]).

Namely, $j = np + x\sqrt{npq} \to \infty$ and $k = nj = nq - x\sqrt{npq} \to \infty$, when $n \to \infty$, for x remains in the finite interval. Applying *Stirling's approximation*

$$m! \approx \sqrt{2\pi m} m^m e^{-m}, \tag{2.25}$$

for (2.20) we find:

$$\Pr(B,k) \approx \frac{\sqrt{2\pi n} n^n e^{-n}}{\sqrt{2\pi j} j^j e^{-j} \sqrt{2\pi k} k^k e^{-k}} p^j q^k = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{jk}} \left(\frac{np}{j}\right)^j \left(\frac{nq}{k}\right)^k.$$

How it is

$$\frac{jk}{n} = n\left(n + x\sqrt{\frac{pq}{n}}\right)\left(q - x\sqrt{\frac{pg}{n}}\right) \approx npq,$$

so

$$\Pr(B,k) \approx \frac{1}{\sqrt{2\pi npq}}\left(\frac{np}{j}\right)^j\left(\frac{np}{k}\right)^k.$$

Using the development of a logarithmic function in a series

$$\ln(1+t) = t - \frac{t^2}{2} + O(t^3),$$

we have:

$$\ln\left(\frac{np}{j}\right)^j\left(\frac{nq}{k}\right)^k = -(np + x\sqrt{npq})\ln\left(1 + x\sqrt{\frac{q}{np}}\right) - (np - x\sqrt{npq})\ln\left(1 - x\sqrt{\frac{p}{nq}}\right) =$$

$$= -\frac{x^2}{2} + O\left(\frac{1}{\sqrt{n}}\right),$$

with which is (2.24) proven.

Similarly, the other part of the Moivre–Laplace approximation is proved

$$\Pr\{a \le \frac{X - np}{\sqrt{npq}} \le b\} \to \frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{x^2}{2}\right) dx, \quad n \to \infty, \tag{2.26}$$

where the random variable X denotes the number of k of favorable outcomes of probability p, and hence defines the number of $n - k$ of unfavorable outcomes of probability $q = 1 - p$, in the constant number of n repetitions in the Bernoulli distribution. The term right is known *Riemann integral* which defines Gaussian distribution.

It's also known

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = 1, \tag{2.27}$$

which means that the Riemann integral well defines *Gaussian distribution* probability. By the change of (2.26) the density of the Gaussian probability is written in the form of

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \tag{2.28}$$

where $\mu = np$ and $\sigma^2 = npq$ is the mean value and variance of Bernoulli distribution, according to Moivre–Laplace approximations and Gaussian. Of course it is

$$\int_{-\infty}^{\infty} \rho(x)\, dx = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1. \tag{2.29}$$

Note that for larger dispersions ($\sigma > 1/\sqrt{2\pi}$) the density coefficient in front of the exponent becomes smaller than one, which means that the distribution density is then less probable, and that the Hartley's information are bigger.

As according to (2.13) entropy grows with average mutual information, so with the increasing dispersion of normal distribution, internal information grows. This not only does not have to be accompanied by an increase in external communication, but it is not. Namely, the spontaneous monotonous, amorphous or faceless distribution of particles reduces their emission of information to the outside, as if with the increase in complexity information systems gives up of the outside world (is "feminized").

2.3.2 Information

Hartley's probability information (2.14) and (2.15) are:

$$h_k = -\log_b p_k, \quad H_k = -\log_b P_k, \quad k = 0, 1, \dots, n, \tag{2.30}$$

where the choice of the logarithm database ($b > 0$, $b \neq 1$) defines the unit of information measure. When $b = 2$ this unit is *bit*, when $b = 10$ is the *decit*, and when we have a natural logarithm of the base ($b = e \approx 2.71828$) the unit information is *nat*. In any case, it is:

$$H_k = -\log_b[\binom{n}{k} p_k] = -\log_b \binom{n}{k} - \log_b p_k = -\log_b \binom{n}{k} + h_k, \tag{2.31}$$

so $H_k \le h_k$.

When the series of information h_k is given with probabilities in the order P_k, which we know to form a binomial distribution, then their mean value is denoted

$$L_n = \sum_{k=0}^{n} P_k h_k \tag{2.32}$$

and we call *physical information* of the binomial distribution $\mathcal{B}(n,p)$. When the series of information H_k is given with probabilities ordered by P_k, then their mean value denotes

$$S_n = \sum_{k=0}^{n} P_k H_k \tag{2.33}$$

and we call Shannon's binomial distribution information. Because (2.31) is $S_n \le L_n$, where equality applies only to $n = 1$, and the difference increases with the increase of n.

In the case of $n = 1$, both of these information are the same

$$L_1 = S_1 = -p \log_b p - q \log_b q, \tag{2.34}$$

where $P_0 = p$ and $P_1 = q = 1 - p$. Note that these probabilities P_0 and P_1 are dependent. In general, in any distribution, because $P_0 + P_1 + \dots + P_n = 1$ will be P_0 depending on the other n probability. This dependence reduces the information that the distribution carries, so Shannon's information is less than physical.

Lemma 2.3.4. *For $\mathcal{B}(2,p)$ the physical information is $L_2 = 2L_1$, where L_1 is the physical information of the single binomial distribution $\mathcal{B}(1,p)$.*

Proof. Followed by:

$$L_2 = -p^2 \log_2 p^2 - 2pq \log_2 pq - q^2 \log_2 q^2 =$$

$$= -2p^2 \log_2 p - 2pq \log_2 p - 2pq \log_2 q - 2q^2 \log_2 q$$

$$= -2p(p+q) \log_2 p - 2q(p+q) \log_2 q,$$

and hence $L_2 = 2L_1$ □

Lemma 2.3.5. *For $\mathcal{B}(3,p)$ the physical information is $L_3 = 3L_1$.*

Proof. Follows from:

$$\begin{aligned} L_3 &= -p^3\log_2 p^3 - 3p^2q\log_2 p^2q - 3pq^2\log_2 pq^2 - q^3\log_2 q^3 = \\ &= -3p^3\log_2 p - 3p^2q\cdot 2\log_2 p - 3p^2q\log_2 q - 3pq^2\log_2 p - 3pq^2\cdot 2\log_2 q - 3q^3\log_2 q \\ &= -3p^2(p+q)\log_2 p - 3pq(p+q)\log_2 p - 3pq(p+q)\log_2 q - 3q^2(p+q)\log_2 q \\ &= -3p(p+q)^2\log_2 p - 3q(p+q)^2\log_2 q \\ &= 3(-p\log_2 p - q\log_2 q), \end{aligned}$$

so $L_3 = 3L_1$. □

From these lemmas it is clear why (2.32) and not (2.33) we called *physical information*, because n repetition of independent simple binomial distributions $\mathcal{B}(1,p)$ defines a complex $\mathcal{B}(n,p)$, and only (2.33) expresses the conservation law $L_n = nL_1$. The following theorem proves this in the general case.

Theorem 2.3.6. *Physical information $\mathcal{B}(n,p)$ is $L_n = nL_1$, where L_1 is the information of the pure binomial distribution $\mathcal{B}(1,p)$.*

Proof. Using (2.32) calculate:

$$\begin{aligned} L_n &= -\sum_{k=0}^{n}\binom{n}{k}p^{n-k}q^k\log_2 p^{n-k}q^k = \\ &= -\sum_{k=0}^{n}\binom{n}{k}(n-k)p^{n-k}q^k\log_2 p - \sum_{k=0}^{n}\binom{n}{k}p^{n-k}kq^k\log_2 q \\ &= -p(\frac{\partial}{\partial p}\sum_{k=0}^{n}p^{n-k}q^k)\log_2 p - q(\frac{\partial}{\partial q}\sum_{k=0}^{n}\binom{n}{k}p^{n-k}q^k)\log_2 q \\ &= -p[\frac{\partial}{\partial p}(p+q)^n]\log_2 p - q[\frac{\partial}{\partial q}(p+q)^n]\log_2 q \\ &= -np(p+q)^{n-1}\log_2 p - nq(p+q)^{n-1}\log_2 q \\ &= n(-p\log_2 p - q\log_2 q), \end{aligned}$$

i.e. L_n is n times bigger than L_1, which has to be proved. □

We communicate because we do not have everything we need, and then because the information does not come out of nothing and does not just disappear, but it is transmitted and can be used. The confirmation of the law on the conservation of information also comes from the possibility of proof by measurement. Hartley's definition of information (2.1) is consistent with its indestructibility because it defines information by realizing the certainty of a previously identical amount of uncertainty. It thus supports the idea that uncertainty is kind of information. Logarithm is the only function $f : N \to \mathbb{R}^+$ with the property $f(xy) = f(x) + f(y)$, so Hartley's choice is the only candidate for the respondent physical information (of equal outcomes).

Because of the law of conservation, all the properties of physical information must be final, we said, because of the definition of infinite sets[6]. Hence, each property of physical information is finally divisible, so it makes sense to talk about the smallest information, the smallest interaction and the slightest action. Consequently, the theorem 2.3.6 states that more complex physical systems have more information.

[6]Infinite sets are, as far as quantity is concerned, equal to their proper part.

2.3.3 Examples

For the binomial distribution, the following three criteria are valid:

1. Number of repetitions $n \in \{1, 2, 3, \dots\}$ is fixed.
2. Every repetition $k \le n$ is an independent random event.
3. The probability of a favorable outcome p is the same in every repetition.

Example 2.3.7. *Fair coin is thrown 10 times. What is the likelihood of falling exactly 6 heads?*

Solution. The number of repetitions $n = 10$, the number of favorable outcomes is $k = 6$, the probability of a favorable outcome is $p = {}^1/_2$ as well as the unfavorable $q = 1 - p = {}^1/_2$. The probability of the required number of outcomes, according to (2.15), is:

$$P_6 = \binom{10}{6} p^6 q^{10-6} = \frac{10!}{6! \cdot (10-6)!} \cdot 0,5^{10} = 210 \cdot 0,000976563 = 0,205078$$

on about six decimal places. □

Example 2.3.8. *About 70% of sports car customers are men. If at random we select 12 owners of sports cars, find the probably that exactly 9 of them are men.*

Solution. The number of repetitions $n = 12$, the number of favorable outcomes is $k = 9$, the probability of a favorable outcome is $p = 0.7$ and the unfavorable $q = 1-p = 0.3$. The required probability is:

$$P_9 = \binom{12}{9} p^9 q^{12-9} = \frac{12!}{9! \cdot (12-9)!} \cdot 0,7^9 \cdot 0,3^3 = 220 \cdot 0,0403536 \cdot 0.027 = 0,239700$$

approximately on 6 decimals. □

Example 2.3.9. *The probability that one product is defective is 0.01, and 100 products are taken from a large warehouse. What is the probability that among these 100 products:*
I) be exactly 5 defective;
II) the number of defected is not greater than 10.

Solution. Given $n = 100$, $p = 0.01$ and $q = 0.99$, we find two probabilities, orderly:

$$P_I = P\{X_{100} = 5\} = \binom{100}{5} \cdot 0,01^5 \cdot 0,99^{95} \approx 0,003$$

$$P_{II} = \sum_{k=0}^{10} P\{X_{100} = k\} = \sum_{k=0}^{10} \binom{100}{k} \cdot 0,01^k \cdot 0,99^{100-k} \approx 1$$

It's almost certain that defected are no more than 10. □

Example 2.3.10. *With 14400 throws of fair coins, the tail fell 7428 times. Define the likelihood of such and larger deviating tail drop than* $np = 14400 \cdot 0,5 = 7200$.

Solution. Use $\sqrt{npq} = \sqrt{14400 \cdot 0,5 \cdot 0,5} = 60$, so the probability:

$$P\{7428 - 7200 \le X_{14400} - 7200 \le 1440\} = P\{3,8 \le \frac{X_{14400} - 7200}{60} \le 120\} =$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{3,8}^{+\infty} e^{-\frac{x^2}{2}}\,dx = \Phi(+\infty) - \Phi(3,8) = 1 - 0,99993 = 0,00007.$$

Laplase function values $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}}\,dt$ are also found in the tables. □

2.4 Distribution density

Normal approximation, $\mathcal{N}(np, npq)$, of the binomial distributions $\mathcal{B}(n, p)$ defines the probability of the form

$$\Pr\{a \le \frac{X_n - np}{\sqrt{npq}} \le b\} \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\,dx, \quad \text{when } n \to \infty. \tag{2.35}$$

In example 2.3.10 we tied it to the once-often tabulated Laplace function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt, \tag{2.36}$$

which makes the area on left to abscise x below Gaussian curve in the figure 2.4.

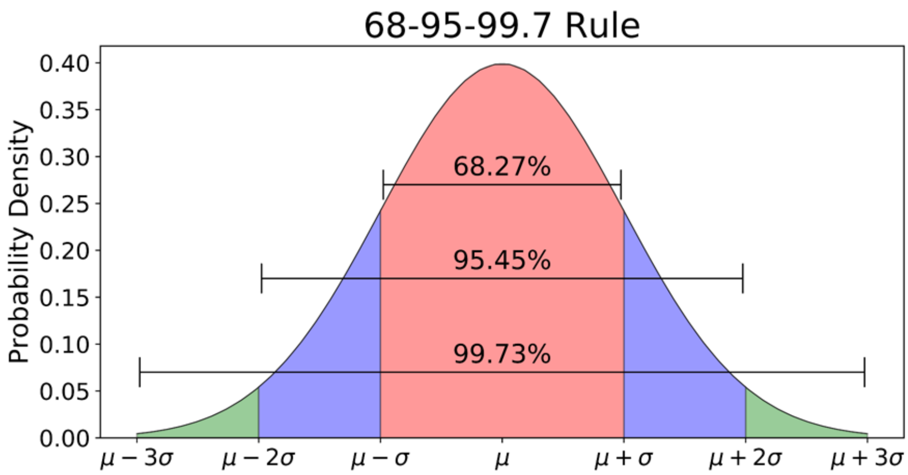

Figure 2.4: Density of normal distribution $\mathcal{N}(\mu, \sigma^2)$.

The surface below the Gaussian density function in the figure represents the limit values of random variables

$$X_n^* = \frac{X_n - \mu}{\sigma}, \quad n \to \infty \tag{2.37}$$

binomial distributions in the interval $[a, b] \subset \mathbb{R}$, with the expectation $\mu = np$ and the variance $\sigma^2 = npq$. This is the normal distribution $\mathcal{N}(\mu, \sigma^2)$. If we introduce a random variable X^* as

$$\Pr\{a \le X^* \le b\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\,dx = \Phi(b) - \Phi(a), \tag{2.38}$$

we get a random variable of continuous type, because for each real number x_0 it is

$$\Pr\{X^* = x_0\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}}\,dx = 0. \tag{2.39}$$

Otherwise, the random variable X is *continuous type* if there is a function $\rho(x) \ge 0$, $x \in \mathbb{R}$, such that

$$\Pr\{a \le X \le b\} = \int_a^b \rho(x)\,dx. \tag{2.40}$$

The function $\rho(x)$ is called *density* of (probability distribution) the random variable X. The figure 2.4 is the density graph of the *normal distribution*.

It should be noted that $\rho(x)$ does not represent any probability. The name "density" comes from the probability that X takes the value in a small interval $[x, x + \Delta x]$ which is proportional to the length of the interval Δx and the value of $\rho(x)$. The normal distribution with expectation 0 and variance 1, which we call $\mathcal{N}(0,1)$, has a probability density (2.36). The following theorem claims that this is indeed a distribution.

Theorem 2.4.1. *Density of normal distribution*

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{2.41}$$

represents the distribution of probability.

Proof. We calculate the integral of a given function, in the order:

$$\int_{-\infty}^{\infty} \rho(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{x^2}{2}}\,dx =$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty e^{-\frac{x^2}{2}}\,dx \cdot \int_0^\infty e^{-\frac{x^2}{2}}\,dx} = \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty e^{-\frac{x^2}{2}}\,dx \cdot \int_0^\infty e^{-\frac{y^2}{2}}\,dy}$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty \int_0^\infty e^{-\frac{x^2+y^2}{2}}\,dydx}, \quad y = xt,$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty \int_0^\infty \exp(-\frac{x^2+x^2t^2}{2})\,xdtdx}$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty \int_0^\infty \exp(-\frac{x^2(1+t^2)}{2})\,xdxdt}$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty \left[-\frac{1}{1+t^2} \exp\left(-\frac{1}{2}x^2(1+t^2)\right)\right]_0^\infty dt}$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty \left(0 + \frac{1}{1+t^2}\right) dt}$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\int_0^\infty \frac{1}{1+t^2}\,dt} = \frac{2}{\sqrt{2\pi}} \sqrt{[\arctan(t)]_0^\infty}$$

$$= \frac{2}{\sqrt{2\pi}} \sqrt{\arctan(\infty) - \arctan(0)} = \frac{2}{\sqrt{2\pi}} \sqrt{\frac{\pi}{2} - 0} = 1,$$

and that was to be proven. □

By supstitution $X \to \frac{X-\mu}{\sigma}$ directly from this theorem it follows that

$$\rho^*(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2.42}$$

is also distribution density. More precisely, with the change of the integral we find, in the following order:

$$\int_{-\infty}^{\infty} \rho(x)\,dx = 1$$

$$\int_{-\infty}^{\infty}\rho(\frac{x-\mu}{\sigma})\,d\frac{x-\mu}{\sigma}=1$$
$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp[-\frac{(x-\mu)^2}{2\sigma^2}]\,d\frac{x-\mu}{\sigma}=1$$
$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp[-\frac{(x-\mu)^2}{2\sigma^2}]\,dx=1$$

that is

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx=1, \tag{2.43}$$

which means that besides $\rho(x)$ we have also the distribution of (continuous) probability density $\rho^*(x)$. These are respectively the density of the *standard and general* normal distribution. The first (ρ) is sometimes called canonical, and the other (here ρ^*) is the default.

2.4.1 Expectation and variance

Theorem 2.4.2. *Mathematical expectation of standard density (2.41) is* $\langle x\rangle=0$.

Proof. follows from:

$$\langle x\rangle=\int_{-\infty}^{\infty}x\rho(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}xe^{-\frac{1}{2}x^2}\,dx=$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}xe^{-\frac{1}{2}x^2}\,dx+\frac{1}{\sqrt{2\pi}}\int_{0}^{\infty}xe^{-\frac{1}{2}x^2}\,dx$$
$$=\frac{1}{\sqrt{2\pi}}\left(-e^{-\frac{1}{2}x^2}\right)_{-\infty}^{0}+\frac{1}{\sqrt{2\pi}}\left(-e^{-\frac{1}{2}x^2}\right)_{0}^{\infty}$$
$$=\frac{1}{\sqrt{2\pi}}(-1+0)+\frac{1}{\sqrt{2\pi}}(0+1)$$
$$=\frac{1}{\sqrt{2\pi}}-\frac{1}{\sqrt{2\pi}}=0,$$

which was to be proved. □

Theorem 2.4.3. *The variance of the standard density probability (2.41) is* $\langle x^2\rangle=1$.

Proof. We work without the lemma 2.3.2 and get directly:

$$\langle x^2\rangle=\int_{-\infty}^{\infty}x^2\rho(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2e^{-\frac{1}{2}x^2}\,dx=$$
$$=\frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{0}x\left(xe^{-\frac{1}{2}x^2}\right)dx+\int_{0}^{\infty}x\left(xe^{-\frac{1}{2}x^2}\right)dx\right]$$
$$=\frac{1}{\sqrt{2\pi}}\left[\left(-xe^{-\frac{1}{2}x^2}\right)_{-\infty}^{0}+\int_{-\infty}^{0}e^{-\frac{1}{2}x^2}\,dx+\left(-xe^{-\frac{1}{2}x^2}\right)_{0}^{\infty}+\int_{0}^{\infty}e^{-\frac{1}{2}x^2}\,dx\right]$$
$$=\frac{1}{\sqrt{2\pi}}\left[(0-0)+(0-0)+\int_{-\infty}^{0}e^{-\frac{1}{2}x^2}\,dx+\int_{0}^{\infty}e^{-\frac{1}{2}x^2}\,dx\right]$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^2}\,dx=1,$$

and that was to be proven. □

By the substitution in the density (2.42) we get the distribution (2.43) with the expectation μ and the variance σ^2. It's easy to check. It can also be done different, say from

$$X = \frac{X^* - \mu}{\sigma} \iff X^* = \mu + \sigma X,$$

follow:

$$\langle X^* \rangle = \langle \mu + \sigma \cdot X \rangle = \mu + \sigma \langle X \rangle = \mu + \sigma \cdot 0 = \mu,$$
$$\langle (X^* - \mu)^2 \rangle = \langle [(\mu + \sigma X) - \mu]^2 \rangle = \langle (\sigma X)^2 \rangle = \sigma^2 \langle X^2 \rangle = \sigma^2.$$

Thus

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \mu \tag{2.44}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \sigma^2. \tag{2.45}$$

The two integrals define respectively *expectation and variance* general normal, so-called Gaussian distribution. We can get the same results differently.

Example 2.4.4. *Check the expectation of general normal distribution (2.44) by direct integration.*

Check. Integrate the left side of equality (2.44):

$$\langle x \rangle = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp[-\frac{(x-\mu)^2}{2\sigma^2}]\, dx =$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x+\mu) \exp[-\frac{x^2}{2\sigma^2}]\, dx =$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mu \exp[-\frac{x^2}{2\sigma^2}]\, dx$$
$$= I_1 + I_2.$$

We calculate the first of these two integrals:

$$I_1 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx =$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{0} x \exp[-\frac{x^2}{2\sigma^2}]\, dx + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx$$
$$= -\frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{-\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} (-x) \exp[-\frac{(-x)^2}{2\sigma^2}]\, dx + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx$$
$$= -\frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} x \exp[-\frac{x^2}{2\sigma^2}]\, dx = 0.$$

Therefore, the first of the above integrals $I_1 = 0$. Calculating the second, I_2, using the theorem 2.4.1. We find:

$$\langle x \rangle = I_1 + I_2 = 0 + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mu \exp[-\frac{x^2}{2\sigma^2}]\, dx =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu \exp[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2] \, d\frac{x}{\sigma} = \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[-\frac{1}{2}x^2] \, dx = \mu.$$

So, $\langle x \rangle = \mu$, and this should have been checked. □

Example 2.4.5. *Check the variance of general normal distribution (2.45) by direct integration.*

Check. Integrate the left side (2.45), using the previous one:

$$\langle (x-\mu)^2 \rangle = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp[-\frac{(x-\mu)^2}{2\sigma^2}] \, dx = \dots$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 \exp[-\frac{x^2}{2\sigma^2}] \, dx, \quad x \to \sigma\sqrt{2}x,$$

$$= \frac{\sigma\sqrt{2}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\sigma\sqrt{2}x)^2 \exp[-\frac{(\sigma\sqrt{2}x)^2}{2\sigma^2}] \, dx$$

$$= \sigma^2 \frac{4}{\sqrt{\pi}} \int_0^{\infty} x^2 \exp[-x^2] \, dx.$$

Let $d = x^2$, then $x = \sqrt{t}$ and $dt = 2xdx = 2\sqrt{t}dx$, so $dx = (2\sqrt{t})^{-1}dt$. Substitute:

$$\langle (x-\mu)^2 \rangle = \sigma^2 \frac{4}{\sqrt{\pi}} \int_0^{\infty} (\sqrt{t})^2 (2\sqrt{t})^{-1} e^{-t} \, dt = \sigma^2 \frac{4}{\sqrt{\pi}} \frac{1}{2} \int_0^{\infty} t^{\frac{3}{2}-1} e^{-t} \, dt =$$

$$= \sigma^2 \frac{4}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{3}{2}\right) = \sigma^2 \frac{4}{\sqrt{\pi}} \frac{1}{2} \frac{\sqrt{\pi}}{2} = \sigma^2,$$

where $\Gamma()$ is a gamma function. Thus, the normal distribution variance is really σ^2. □

2.4.2 Triangular distribution

When we consider examples of the normal distribution $\mathcal{N}(\mu, \sigma^2)$ as an approximation of the binomial $\mathcal{B}(n,p)$, for $n \to \infty$, it is often assumed that the initial distribution is binomial, and that it is not actually. Such is an example of a discrete *triangular distribution*.

Example 2.4.6. *The sum of twice randomly chosen numbers from the set $\{1, 2, \dots, n\}$ does not make a binomial distribution.*

Check. The random variable X of this distribution takes values $1+1, 1+2\dots, 1+n$, then $2+1, 2+2, \dots, 2+n$, and so up to $n+1, n+2, \dots, n+n$. It is assumed that each of these $n \times n$ values is equally probable, that is, each of the matrix components

$$\mathbf{N} = \begin{pmatrix} 2 & 3 & \dots & n+1 \\ 3 & 4 & \dots & n+2 \\ \dots & & & \\ n+1 & n+2 & \dots & 2n \end{pmatrix} \tag{2.46}$$

has the same probability. On sloping diagonals, there are equal numbers, one 2, two 3 and in general have $k \in \{1, 2, \dots, n\}$ of numbers $x = k+1$, up to the number $x = n+1$ that is repeated n times. Then the number $x = n+2$ repeats $n-1$ times, the number $x = n+3$ repeats $n-2$ times and in general the number $x = n+k$ repeats $n-k+1$ times.

The probability distribution of this matrix is

$$\Pr(X=x)=\begin{cases}\frac{n-|x-(n+1)|}{n^2} & x\in\{2,3,\dots,2n\}\\ 0 & x\notin\{2,3,\dots,2n\}\end{cases} \tag{2.47}$$

Thus, for six numbers of cube ($n=6$) we have the probability $\Pr(2)=\frac{1}{36}$, $\Pr(3)=\frac{2}{36}$, $\Pr(4)=\frac{3}{36}$, $\Pr(5)=\frac{4}{36}$, $\Pr(6)=\frac{5}{36}$, $\Pr(7)=\frac{6}{36}$, $\Pr(8)=\frac{5}{36}$, $\Pr(9)=\frac{4}{36}$, $\Pr(10)=\frac{3}{36}$, $\Pr(11)=\frac{2}{36}$ и $\Pr(12)=\frac{1}{36}$. The sum of these probabilities is one that means they make the distribution.

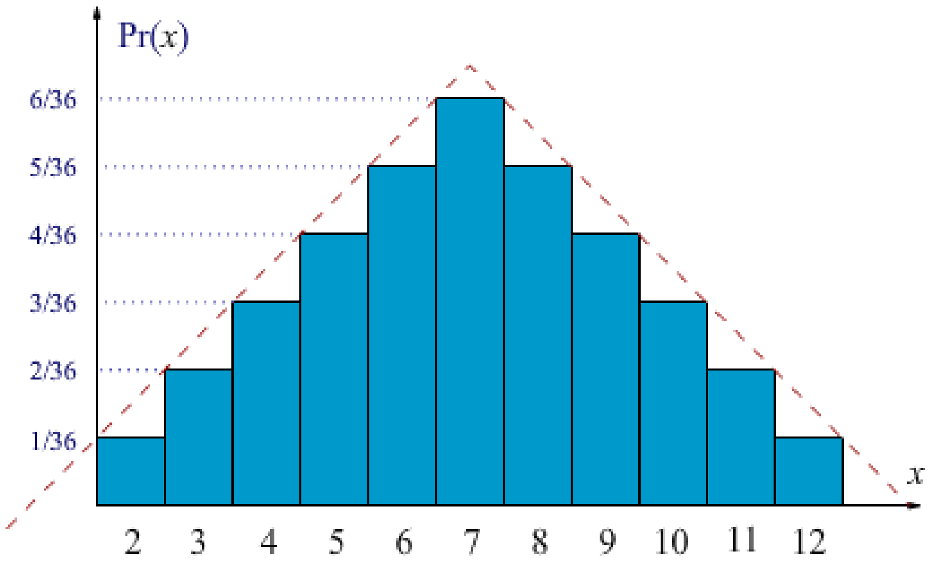

Figure 2.5: Discrete triangular distribution of probabilities.

In the figure 2.5 we see that the probabilities (2.47) make a “triangle” distribution, as opposed to the “bell-shaped” normal seen in the figure 2.4. It retains its triangular form and when $n\to\infty$, which means it can not be approximated by the normal one. This is because the distribution itself (2.47) is not binomial.

Indeed, the general binomial distribution $\mathcal{B}(n,p)$ have the probabilities

$$\Pr(X=x)=\begin{cases}\binom{n}{x}p^x(1-p)^x & x\in\{0,1,2,\dots,n\}\\ 0 & x\notin\{0,1,2,\dots,n\}\end{cases} \tag{2.48}$$

so for $n=6$ there is no such $p\in(0,1)$ that will give (2.47). □

When calculating discrete triangular distributions, it happened the adding the powers of integers, forms

$$F_p(n)=\sum_{k=1}^{n}k^p=1^p+2^p+3^p+\dots+n^p,\quad p=1,2,3,\dots \tag{2.49}$$

which we call *Faulhaber's* (Johann Faulhaber, 1580-1635) sums. For the first few powers, these sums can be obtained elementally and, fortunately, they are the most common in use.

These are:

$$\begin{aligned}
F_1(n) &= 1+2+\dots+n = \frac{1}{2}(n^2+n) = \frac{1}{2}n(n+1) \\
F_2(n) &= 1^2+2^2+\dots+n^2 = \frac{1}{6}(2n^3+3n^2+n) = \frac{1}{6}n(n+1)(2n+1) \\
F_3(n) &= 1^3+2^3+\dots+n^3 = \frac{1}{4}(n^4+2n^3+n^2) = \frac{1}{4}n^2(n+1)^2 \\
F_4(n) &= 1^4+2^4+\dots+n^4 = \frac{1}{30}(6n^5+15n^4+10n^3-n)
\end{aligned} \tag{2.50}$$

For larger powers, the expressions of the Faulhaber's formulas become more and more complicated.

Example 2.4.7. *Show that $\mu = n+1$ and $\sigma^2 = (n^2-1)/6$ in the order are the expectation and variance of the discrete triangular distribution (2.47).*

Solution. Write the probabilities (2.47) in the following way

$$\Pr(X=x) = \begin{cases} \frac{x-1}{n^2} & x \in \{2,3,\dots,n\} \\ \frac{2n-x+1}{n^2} & x \in \{n+1,n+2,\dots,2n\} \\ 0 & x \notin \{2,3,\dots,2n\} \end{cases} \tag{2.51}$$

The mathematical expectation is:

$$\langle x\rangle = \sum_{x=2}^{2n} x\Pr(x) = \sum_{x=2}^{n} x\Pr(x) + \sum_{x=n+1}^{2n} x\Pr(x) =$$

$$= \sum_{x=2}^{n} x\frac{x-1}{n^2} + \sum_{x=n+1}^{2n} x\frac{2n-x+1}{n^2}$$

$$= \frac{1}{n^2}\left(\sum_{x=2}^{n} x^2 - \sum_{x=2}^{n} x\right) + \frac{2n+1}{n^2}\sum_{x=n+1}^{2n} x - \frac{1}{n^2}\sum_{x=n+1}^{2n} x^2$$

$$= \frac{1}{n^2}\left(\sum_{x=2}^{n} x^2 - \frac{(n-1)(n+2)}{2}\right) + \frac{2n+1}{n^2}\frac{n(3n+1)}{2} - \frac{1}{n^2}\left(\sum_{x=2}^{2n} x^2 - \sum_{x=2}^{n} x^2\right)$$

$$= \frac{2}{n^2}\sum_{x=2}^{n} x^2 + \frac{1}{2n^2}[n(2n+1)(3n+1)-(n-1)(n+2)] - \frac{1}{n^2}\sum_{x=2}^{2n} x^2$$

$$= \frac{2}{n^2}\left(\frac{1}{3}n^3+\frac{1}{2}n^2+\frac{1}{6}n-1\right) + \frac{1}{n^2}[3n^3+2n^2+1] - \frac{1}{n^2}\left(\frac{1}{3}(2n)^3+\frac{1}{2}(2n)^2+\frac{1}{6}(2n)-1\right)$$

$$= \frac{1}{n^2}(n^3+n^2) = n+1.$$

Therefore, the mathematical expectation of the triangular distribution is $\mu = n+1$.

Calculate the variation using the previous and the lemma 2.3.2:

$$\langle (x-\mu)^2\rangle = \langle x^2\rangle - \langle x\rangle^2 =$$

$$= \sum_{x=2}^{n} x^2\Pr(x) + \sum_{x=n+1}^{2n} x^2\Pr(x) - (n+1)^2$$

$$= \sum_{x=2}^{n} x^2 \frac{x-1}{n^2} + \sum_{x=n+1}^{2n} x^2 \frac{2n-x+1}{n^2} - (n+1)^2$$

$$= \frac{1}{n^2}\sum_{x=2}^{n} x^3 - \frac{1}{n^2}\sum_{x=2}^{n} x^2 + \frac{2n+1}{n^2}\sum_{x=n+1}^{2n} x^2 - \frac{1}{n^2}\sum_{x=n+1}^{2n} x^3 - (n+1)^2$$

$$= \frac{1}{n^2}\left(2\sum_{x=2}^{n} x^3 - \sum_{x=2}^{2n} x^3\right) + \frac{1}{n^2}\left[(2n+1)\sum_{x=2}^{2n} x^2 - (2n+2)\sum_{x=2}^{n} x^2\right] - \frac{n^2(n+1)^2}{n^2}$$

$$= \frac{1}{n^2}\left[2\left(\frac{n^2(n+1)^2}{4} - 1\right) - \left(\frac{(2n)^2(2n+1)^2}{4} - 1\right) - n^2(n+1)^2\right] +$$

$$+ \frac{1}{n^2}\left[(2n+1)\left(\frac{1}{3}(2n)^3 + \frac{1}{2}(2n)^2 + \frac{1}{6}(2n) - 1\right) - (2n+2)\left(\frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n - 1\right)\right]$$

$$= -\frac{1}{n^2}\left[\frac{n^2(n+1)^2}{2} + n^2(2n+1)^2 + 1\right] + \frac{1}{n^2}\left(\frac{14n+6}{3}n^3 + \frac{6n+2}{2}n^2 + \frac{2n}{6}n + 1\right)$$

$$= -\frac{1}{n^2}\left(\frac{9}{2}n^4 + 5n^3 + \frac{3}{2}n^2 + 1\right) + \frac{1}{n^2}\left(\frac{14}{3}n^4 + 5n^3 + \frac{4}{3}n^2 + 1\right)$$

$$= \frac{1}{n^2}\left(\frac{1}{6}n^4 - \frac{1}{6}n^2\right) = \frac{1}{6}(n^2 - 1).$$

Therefore, the variance of the triangular distribution ise $\sigma^2 = (n^2-1)/6$. □

Triangular probability distributions may also be *continual*. Then, for the general case, the lower bound a is usually taken, the upper bound b and the mod c, where $a < c < b$ and are denoted by $\mathcal{T}(a,b,c)$. In the figure 2.6 we see one such triangle of area $\frac{1}{2}(b-a)\frac{2}{b-a} = 1$, of the well-defined distribution. The probability density is given by the function

$$\rho(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & x \in [a,c] \\ \frac{2(b-x)}{(b-a)(b-c)} & x \in [c,b] \\ 0 & x \notin [a,b] \end{cases} \tag{2.52}$$

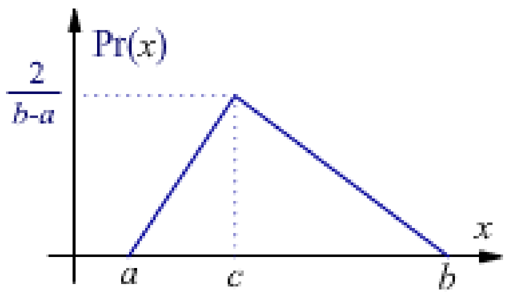

Figure 2.6: Continuous triangular distribution.

Example 2.4.8. *Show that are respectively:*

$$\mu = \frac{a+b+c}{3}, \quad \sigma^2 = \frac{1}{18}(a^2+b^2+c^2-ab-bc-ca), \tag{2.53}$$

the expectation and variance of the continuous triangular distribution (2.52).

Proof. By definition, the mathematical expectation is:

$$\langle x\rangle = \int_{-\infty}^{\infty} x\rho(x)\,dx = \int_a^c x\frac{2(x-a)}{(b-a)(c-a)}\,dx + \int_c^b x\frac{2(b-x)}{(b-a)(b-c)}\,dx =$$

$$= \left.\frac{\frac{2}{3}x^3 - ax^2}{(b-a)(c-a)}\right|_a^c + \left.\frac{bx^2 - \frac{2}{3}x^3}{(b-a)(b-c)}\right|_c^b$$

$$= \frac{(c-a)(2c+a)}{3(b-a)} + \frac{(b-c)(b+2c)}{3(b-a)}$$

$$= \frac{(b-a)(a+b+c)}{3(b-a)} = \frac{a+b+c}{3}.$$

Thus the mathematical expectation $\mu = \langle x\rangle$ is proved.

By lemma 2.3.2 the variance is:

$$\sigma^2 = \langle x^2\rangle - \langle x\rangle^2 =$$

$$= \int_{-\infty}^{\infty} x^2\rho(x)\,dx - \left(\int_{-\infty}^{\infty} x\rho(x)\,dx\right)^2$$

$$= \int_a^c x^2\frac{2(x-a)}{(b-a)(c-a)}\,dx + \int_c^b x^2\frac{2(b-x)}{(b-a)(b-c)}\,dx - \mu^2$$

$$= \left.\frac{2(\frac{1}{4}x^4 - a\frac{1}{3}x^3)}{(b-a)(c-a)}\right|_a^c + \left.\frac{2(b\frac{1}{3}x^3 - \frac{1}{4}x^4)}{(b-a)(b-c)}\right|_c^b - \left(\frac{a+b+c}{3}\right)^2$$

$$= \frac{(c-a)(3c^2+2ac+a^2)}{6(b-a)} + \frac{(b-c)(b^2+2bc+3c^2)}{6(b-a)} - \frac{(a+b+c)^2}{9}$$

$$= \frac{1}{6}(a^2+b^2+c^2+ab+bc+ca) - \frac{1}{9}(a^2+b^2+c^2+2ab+2bc+2ca)$$

$$= \frac{1}{18}(a^2+b^2+c^2-ab-bc-ca).$$

This is the second requested result. □

Continuous triangular probabilities become simpler in the case of an equilateral triangle, when $a = 0$, $b = 2\alpha$, $c = \alpha$, with the height $1/\alpha$, of the triangle $\mathcal{T}(0, 2\alpha, \alpha)$. Again, we have a distribution (the area of the triangle is 1). Substituting these data in (2.53) we find the expectation $\mu = \alpha$ and the variance $\sigma^2 = \mu^2/6$.

Density (2.52) is then

$$\rho(x) = \begin{cases} \frac{x}{\alpha^2} & x \in [0,\alpha] \\ \frac{2\alpha - x}{\alpha^2} & x \in [\alpha, 2\alpha] \\ 0 & x \notin [0, 2\alpha] \end{cases} \tag{2.54}$$

so mathematical expectation and variance can be calculated directly as in the following example, analogous to the previous one.

Example 2.4.9. *Show that respectively:*

$$\mu = \alpha, \quad \sigma^2 = \alpha^2/6, \tag{2.55}$$

are the expectation and variance of continuous triangular distribution (2.54).

Proof. By definition, the mathematical expectation is:

$$\langle x\rangle = \int_{-\infty}^{\infty} x\rho(x)\,dx =$$

$$= \int_0^{\alpha} \frac{x^2}{\alpha^2}\,dx + \int_{\alpha}^{2\alpha} \frac{2\alpha x - x^2}{\alpha^2}\,dx$$

$$= \left.\frac{x^3}{3\alpha^2}\right|_0^{\alpha} + \left.\frac{\alpha x^2 - \frac{x^3}{3}}{\alpha^2}\right|_{\alpha}^{2\alpha}$$

$$= \frac{\alpha}{3} + \frac{2\alpha}{3} = \alpha,$$

so the expectation is $\mu = \langle x\rangle = \alpha$, which should have been proven.

According the lemma 2.3.2 the variance is:

$$\sigma^2 = \langle x^2\rangle - \langle x\rangle^2 =$$

$$= \int_{-\infty}^{\infty} x^2\rho(x)\,dx - \mu^2$$

$$= \int_0^{\alpha} \frac{x^3}{\alpha^2}\,dx + \int_{\alpha}^{2\alpha} \frac{2\alpha x^2 - x^3}{\alpha^2}\,dx - \alpha^2$$

$$= \left.\frac{\frac{1}{4}x^4}{\alpha^2}\right|_0^{\alpha} + \left.\frac{\frac{2}{3}\alpha x^3 - \frac{1}{4}x^4}{\alpha^2}\right|_{\alpha}^{2\alpha} - \alpha^2$$

$$= \frac{\alpha^2}{4} + \frac{11\alpha^2}{12} - \alpha^2 = \frac{\alpha^2}{6},$$

and that's the variance we should get. □

The consensus of the discrete and continuous triangular distribution results is striking, especially the expectations and variances from the example 2.4.7 and this last 2.4.7. Now let's look at several of their applications in the store.

Example 2.4.10 (Store)**.** *The dealer is planning a new sale place. How could a future sales model look like with a triangular distribution?*

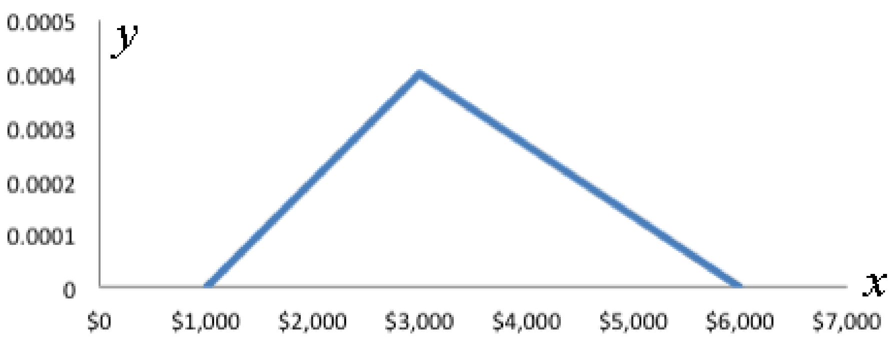

Figure 2.7: Estimation of sales.

Solution. Let's estimate the minimum weekly sales at around \$1000, the maximum of \$6000, and the expected about \$3000. Then at abscissa (x-axis) we take $a = 1000$, $b = 6000$ and $c = 3000$, and on ordinates (y-axes) its density. The graph looks like a triangle in the figure 2.7. The maximum density is at the top of the triangle, with ordinate 0.0004 of abscissa $c = 3000$.

When a trader wants to determine the likelihood that the weekly sales will be less than \$2000, he seeks the area of the triangle from abscise 1000 to 2000, half way to the top of a given triangle, and then half the maximum ordinate. Surface, half-product of base and height, such a triangle has the area

$$\Pr(x < 2000) = \frac{1}{2} \times 1000 \times 0.0002 = 0.1$$

represents the likelihood of this small weekly sales. If his weekly sales were lower than this amount, he would have poor prospects to cover the costs, but the likelihood of such a thing occurring is very small.

These estimates for the weekly sales of x in the interval from x_1 to x_2 are obtained by using the density formula (2.52) by calculating the integral

$$\Pr(x_1 < x < x_2) = \int_{x_1}^{x_2} \rho(x)\,dx$$

with given parameter values a, b and c. □

Example 2.4.11 (Voting)**.** *Voting for school representatives has been completed, but votes have not yet been counted. Make a model of the triangular distribution of candidate's K expectations.*

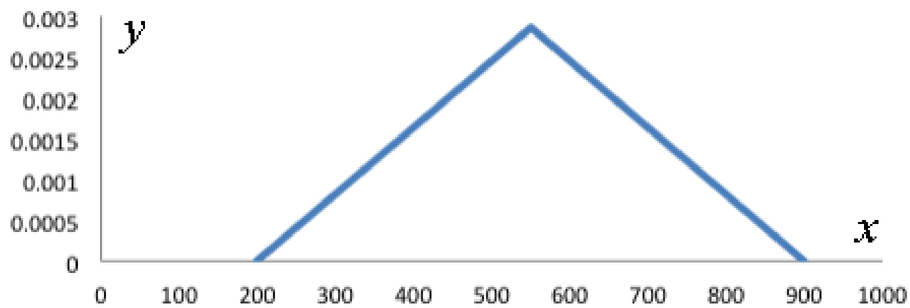

Figure 2.8: Assessment of voting.

Solution. Candidate K is considering the number of votes he could get. He believes that the apparent value is about 550, but he could get up to 900, or drop down to 200. For him, the simplest is a triangular distribution with the parameters $a = 200$, $b = 900$ and $c = 550$. The density graph of such a distribution, in the figure 2.8, reaches the maximum of $y_0 = 0.002857$ with the abscissa $c = 550$.

If the candidate wants to determine the probability of having more than 450 votes, he should find the probability $\Pr(X > 450)$ with the given distribution density. He can find it by integrating the corresponding function (2.52), from 450 to 900, but the distribution of triangles can be done much easier. Here's how.

In the figure 2.8 of the main triangle, note the triangle of abscissa 200 to 450, the outcome that the candidate does not want. It is part of the main triangle to the left of the top, with the lower cathets at $250 : 350$, so in the same ratio $h : y_0$ must also be heights. Hence

$$h = \frac{250 \times y_0}{350} = \frac{250 \times 0.002857}{350} = 0.00204,$$

and the surface of this triangle is a half-product of the base and height

$$p = \frac{1}{2} \times 250 \times h = 0.25509,$$

and this area is the probability of an unwanted event. Therefore, the desired event has the probability $\Pr(X > 450) = 1 - p = 0.74491$. □

2.4.3 Uniform distribution

Uniform distribution (continuous uniform, rectangular, or distribution that has constant probability) is the distribution of so symmetric probabilities that all intervals of the same length are of the same probability. The density of probability of uniform distribution is

$$\rho(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & x \notin [a,b] \end{cases} \tag{2.56}$$

where $a < b$. We denote the uniform distribution $\mathcal{U}(a,b)$.

Example 2.4.12. *Show that expectation and variance of the uniform distribution:*

$$\mu = \frac{b+a}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}.$$

Solution. Mathematical expectation is:

$$\langle x \rangle = \int_{-\infty}^{\infty} x\rho(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{b+a}{2},$$

so $\mu = \langle x \rangle$.

The variance is:

$$\langle (x-\mu)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = \int_a^b \frac{x^2}{b-a}\,dx - \mu^2 =$$

$$= \frac{1}{3}\frac{x^3}{b-a}\bigg|_a^b - \left(\frac{a+b}{2}\right)^2 = \frac{b^2+ba+a^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{b^2-2ab+a^2}{12},$$

and this is the required value σ^2. □

These results are intuitively understandable. The arithmetic mean of the interval $[a,b]$ is the expectation μ, and the length of the interval $b-a$ is proportional to the dispersion $\sigma = \sqrt{\sigma^2}$.

Although it is relatively simple and perhaps just because of that, the examples of the uniform distribution are everywhere. These are serial numbers of randomly selected banknotes, the number obtained by throwing a dice, the numbers drawn at the lottery.

Example 2.4.13 (Quiz). *There are 18 contestants on the quiz. The question is asked for all 18, and the allowed response time is 30 seconds. How many competitors will respond in the first 5 seconds?*

Solution. We assume a uniform distribution of responses. Then $a = 0$ and $b = 30$, and the interval of the desired outcome is 0 to 5 seconds. We look at the probability density rectangle with a base of 30 and the height of $1/30$ so that the total surface is one.

The outcome we are looking for is its cut on the left, a base of 5 and the same height. The area of that segment is $5 \times \frac{1}{30} = \frac{1}{6}$ is the required probability, and the number of competitors is $18 \times \frac{1}{6} = 3$. In the first 5 seconds, 3 competitors will respond with a response. □

Example 2.4.14 (Airport). *Airplane landing was announced in an interval of 30 minutes. What is the probability of arrival of the plane between 25 and 30 minutes?*

Solution. We assume a uniform distribution of arrivals. Parameters are $a = 0$ and $b = 30$, and a favorable outcome is an interval of 25 to 30 minutes. We are looking for the probability $\Pr(25 < X < 30)$ as the area of the rectangle base $30 - 25 = 5$ and height $1/30$, the scrap from rectangular of length 30 and the same height. That's sixth; the probability of arrival of the flight between 25 and 30 minutes is $\frac{1}{6} \approx 0.16$. □

2.4.4 Information of density

For a well-defined function of the probability distribution density, it is valid that it is everywhere nonnegative and that it's integral over the entire probability space

$$\int_\Omega \rho(\omega)\,d\omega = 1. \tag{2.57}$$

This is a general definition, but we mainly consider the one-dimensional space Ω and this is one axis of numbers, abscissa, the x-axis, at an interval from minus to plus infinite, and so we the infinitesimal change instead $d\omega$ write dx.

When the continuous distribution of probability is given by its density $\rho(x)$, then the classic *information* is defined by the expression

$$S = -\int_{-\infty}^{\infty} \rho(x) \log_b \rho(x)\,dx, \tag{2.58}$$

which we call the Shannon information, or *entropy*. The information unit is determined by the logarithm's base ($b > 0$, $b \ne 1$). With $b = 2$ we work in bits, with $b = e$ we work with natural logarithms in the nats.

For the uniform distribution (2.56), the classical *information* density is

$$S_U = \ln(b - a), \tag{2.59}$$

nat. Namely:

$$S_U = -\int_{-\infty}^{\infty} \frac{1}{b-a} \ln \frac{1}{b-a}\,dx = \frac{x}{b-a} \ln(b-a)\Big|_a^b = \ln(b-a).$$

It's Hartley's information of evenly probable outcomes. As previously mentioned, it is physical information because it supports the physical law of information conservation.

Theorem 2.4.15. *For the triangular distribution (2.52), the classical information of the density is*

$$S_T = \ln\frac{(b-a)\sqrt{e}}{2}, \tag{2.60}$$

in nats.

Proof. We calculate the classic density information:

$$S_T = -\int_{-\infty}^{\infty}\rho(x)\ln\rho(x)\,dx =$$

$$= -\int_a^c \frac{2(x-a)}{(b-a)(c-a)}\ln\frac{2(x-a)}{(b-a)(c-a)}\,dx - \int_c^b \frac{2(b-x)}{(b-a)(b-c)}\ln\frac{2(b-x)}{(b-a)(b-c)}\,dx$$
$$= I_1 + I_2.$$

The first of the integrals is:

$$I_1 = -\frac{1}{2}(b-a)(c-a)\int_0^{2/(b-a)} t\ln t\,dt, \quad t = \frac{2(x-a)}{(b-a)(c-a)},$$

$$= -\frac{1}{2}(b-a)(c-a)\int_0^{2/(b-a)} \ln t\,d\frac{t^2}{2}$$

$$= -\frac{1}{2}(b-a)(c-a)\left(\frac{t^2}{2}\ln t - \int\frac{t^2}{2}\,d\ln t\right)_0^{2/(b-a)}$$

$$= \frac{c-a}{b-a}\ln\frac{b-a}{2} + \frac{1}{2}(b-a)(c-a)\int_0^{2/(b-a)}\frac{t^2}{2}\frac{1}{t}\,dt$$

$$= \frac{c-a}{b-a}\ln\frac{b-a}{2} + \frac{1}{2}(b-a)(c-a)\left(\frac{t^2}{4}\right)_0^{2/(b-a)}$$

$$= \frac{c-a}{b-a}\cdot\ln\frac{b-a}{2} + \frac{1}{2}\cdot\frac{c-a}{b-a}$$

$$= \frac{c-a}{b-a}\ln\frac{(b-a)\sqrt{e}}{2}.$$

The second one was:

$$I_2 = \frac{1}{2}(b-a)(b-c)\int_{2/(b-a)}^{0} t\ln t\,dt, \quad t = \frac{2(b-x)}{(b-a)(b-c)},$$

$$= \frac{1}{2}(b-a)(b-c)\int_{2/(b-a)}^{0}\ln t\,d\frac{t^2}{2}$$

$$= \frac{1}{2}(b-a)(b-c)\left(\frac{t^2}{2}\ln t - \int\frac{t^2}{2}\,d\ln t\right)_{2/(b-a)}^{0}$$

$$= -\frac{b-c}{b-a}\ln\frac{2}{b-a} - \frac{1}{2}(b-a)(b-c)\int_{2/(b-a)}^{0}\frac{t^2}{2}\frac{1}{t}\,dt$$

$$= \frac{b-c}{b-a}\ln\frac{b-a}{2} - \frac{1}{2}(b-a)(b-c)\left.\frac{t^2}{4}\right|_{2/(b-a)}^{0}$$

$$= \frac{b-c}{b-a}\ln\frac{b-a}{2} + \frac{1}{2}\frac{b-c}{b-a}$$
$$= \frac{b-c}{b-a}\ln\frac{(b-a)\sqrt{e}}{2}.$$

The sum of these integrals is:

$$S = I_1 + I_2 =$$
$$= \frac{c-a}{b-a}\ln\frac{(b-a)\sqrt{e}}{2} + \frac{b-c}{b-a}\ln\frac{(b-a)\sqrt{e}}{2}$$
$$= \ln\frac{(b-a)\sqrt{e}}{2},$$

which is exactly the result(2.60). □

By comparing triangular and uniform information we find

$$S_T < S_U, \tag{2.61}$$

because:

$$\ln\frac{(b-a)\sqrt{e}}{2} < \ln(b-a),$$
$$\frac{(b-a)\sqrt{e}}{2} < b-a,$$
$$\sqrt{e} < 2,$$

since all these inequalities are equivalent to each other (equally true), and the last is equivalent to $e < 4$, which is true since $e = 2.71828\ldots$ is the natural logarithm base.

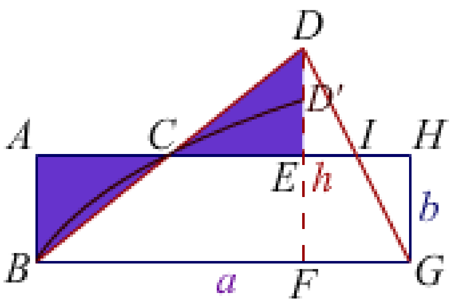

Figure 2.9: Uniform and triangular distribution.

The same can be understood with the figure 2.9. The $ABGH$ quadrilateral has the base $a = BG$ and such a side edge $b = GH$ that its surface is $ab = 1$, because it represents a uniform distribution. The BGD triangle has the same base a but twice the height $h = FD$ of b, since the area of the triangle is $\frac{1}{2}ah = 1$, and it represents a triangular distribution. Therefore, the charted triangles are congruent, $\Delta CAB \cong \Delta CED$.

What does this have to do with inequality (2.61)? The triangle ΔCAB is at a lower altitude than the ΔCED, has smaller ordinates, which means that its points give less probability, and therefore more information. Because the lower triangle lies in the low probability area, it carries more information to the top of the upper, although both have the same surface. Hence, by integrating, inequality arises (2.61).

Continuing, if instead of the triangular distribution, instead of the straight line $B-C-D$ it goes along the curve line $B-C-D'$, as in the figure 2.9, forming a "curvilinear triangle" distribution, the transfer of points from the lower region to the upper would be smaller than in the triangular distribution, so that the corresponding information would have a value between the triangular and the uniform. A different example is the normal distribution, which in spite of its "bell-shaped" form that reminds us of the triangle (figure 2.4) for its unlimited base (from $-\infty$ to $+\infty$) and for that the low probability, which means high information, can have total information greater than the uniform distribution.

It is the another question whether the information (2.58) should be regarded as physical in the case of normal distribution, the way we accepted it in the case of Hartley's definition? If information of both uniform and even triangular distribution can still be considered sufficiently "Hartley", in order to be physical, the question arises: where is the limit after which we have to give up of the Shannon definition (2.58), so that we can count on the physical law of information conservation? These are dilemmas that make interesting use of the α parameter in the next theorem (in the proof only) to look at both options.

We work with the general normal distribution (2.43) in finding the mean values of the corresponding Hartley information. However, for Hartley's information we do not take the probability of the normal distribution itself, but we act as in the case of physical information of the binomial distribution L_B, permitting the possibility that the coefficient $1/\sqrt{2\pi\sigma^2}$ has some other value α.

Theorem 2.4.16. *The information of the normal distribution is* $S_N = \ln\sqrt{2\pi e\sigma^2}$.

Proof.

$$I_\alpha = -\int_{-\infty}^{\infty} \rho(x)\ln\rho_\alpha(x)\,dx =$$

$$= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln[\alpha e^{-\frac{(x-\mu)^2}{2\sigma^2}}]\,dx$$

$$= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln\alpha\,dx - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

$$= -(\ln\alpha)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

$$= -\ln\alpha - \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[\frac{x-\mu}{2}\, d\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right]$$

$$= -\ln\alpha - \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} \frac{x-\mu}{2}\, de^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= -\ln\alpha - \frac{1}{\sqrt{2\pi\sigma^2}}\left[\frac{x-\mu}{2}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, d\frac{x-\mu}{2}\right]$$

$$= -\ln\alpha + \frac{1}{2},$$

According

$$I_\alpha = \ln(\alpha^{-1}\sqrt{e}), \tag{2.62}$$

so submitting $\alpha^{-1} = \sqrt{2\pi\sigma^2}$, we get the requested claim. □

If we understand the dispersion of the normal distribution as the "base", similar to the base of the uniform and triangular distribution, then in all three cases the bases behaves analogously to the number N of equally likely outcomes of Hartley's information. In this way, we already have physical information in the classical one. However, to really consider the physical information of the continuum is contrary to the nature of such information.

Infinite sets are those that are (by quantity) equal to some its proper subsets, and this is not possible if the conservation law is valid. Due to the law of conservation of the physical information, continuous information cannot be physical, and accordingly, the continuous values that we have reached are beyond the end of the story.

On the other hand, the question arises as to why there is some information beyond which physical information is not available? I discussed this issue in more detail in the book "Multiplicities" [3], and here I will transfer only a part. The physical action (energy and time product) and physical information are the proportional sizes. There are no actions without information and vice versa. A well-known "principle of the least action" (skimping with actions) and a new "principle of information" (information skimping) are related. Both, physical action and information, are kind of the true statements. Only something that is "true" can stand behind an arbitrary physical phenomenon, but truth can also be obtained from falsehood, such as negation or implication.

Money functions as a fiction, because people believe that for an amount of money they can get the appropriate counter value. Religions or sects will gather their followers on the fictitious idea of their beliefs. Pythagoras' theorem helps build physical objects, although it is not itself a physical matter. Fictions that produce effects in physical reality, and therefore physical information, I called pseudo-real. They provide a one-way flow of information, without the possibility of a mutual effect, such as the past or *parallel reality*, in addition to the above.

In that book, the mathematical concept of an abstract "reality" is presented, whose "objectivity" is indisputable, simply because the laws of mathematics cannot be changed. They are "more objective than the most objective" in that sense. In other words, in the infinite mathematical substance is always the finite physical substance in all its properties. This construction is similar to Plato's "world of ideas", except that now we know that there is no "set of all sets" (Russell's paradox of sets), nor a "formula of all formulas" (Gödel's incompleteness theorems), which we just add the finiteness of each property of a physical substance.

How is the continuum possible at all? It is also one of the difficult questions discussed in the book [3]. An infinite discrete sequence of events is a potential possibility, fiction, because none of its realizable part is infinite. By allowing an objective coincidence, which is the principle of the mentioned book, and the meaning of quantum mechanical superposition, some events have two or more implementation options. If in the sequence of events infinitely there are different possibilities of realization, then the cardinal number, i.e. the number of set elements, these options have *continuum* ($\mathfrak{c}$), although each potential sequence of physical events is at most *countable infinite* ($\aleph_0$). There is the continuum of the parallel realities, although we communicate only with one of its discrete series.

If our thought was one-dimensional, discrete, like a series of all atoms of our body or a set of all atoms of visible space, then it would be possible a human to simply encode and get a duplicate of him. Such a person could not think about a continuum thoughts. However, our thought is ambiguous, so we can get to know parallel realities, as we can discover theorems, but we can not physically communicate with them in the way we act on them just as they can change us.

2.5 Discrete probability

A discrete distribution describes the probability of occurrence of certain values of a discrete random variable. A discrete random variable has countable values, finitely or infinitely like a set of natural numbers.

The set of probability values (2.51) of the discrete triangular distribution is finite, having a total of $2n-1$. We can reorganize them into collection with sum

$$\frac{n}{n^2} + 2\sum_{k=1}^{n-1} \frac{k}{n^2} = 1, \tag{2.63}$$

which is easy to check. Note that these and earlier probabilities are the same, but that the random variables of those in relation to the indices of these probabilities are greater for one.

The classic *information* of this, the final triangle distribution is

$$\begin{aligned} S_T(n) &= -\frac{n}{n^2}\log_b \frac{n}{n^2} - 2\sum_{k=1}^{n-1} \frac{k}{n^2}\log_b \frac{k}{n^2} \\ &= \frac{n}{n^2}\log_b \frac{n}{n^2} - 2\sum_{k=1}^{n} \frac{k}{n^2}\log_b \frac{k}{n^2} \\ &= -\frac{\log_b n}{n} - \frac{2}{n^2}\sum_{k=1}^{n} k\log_b k + \frac{4\log_b n}{n^2}\sum_{k=1}^{n} k \\ &= -\frac{\log_b n}{n} - \frac{2}{n^2}\sum_{k=1}^{n} k\log_b k + \frac{4\log_b n}{n^2}\frac{1}{2}n(n+1), \end{aligned} \tag{2.64}$$

so

$$S_T(n) = \frac{2n+1}{n}\log_b n - \frac{2}{n^2}\sum_{k=1}^{n} k\log_b k. \tag{2.65}$$

Approximately:

$$\sum_{k=1}^{n} k\log k \approx \int_1^{n+1} x\log x\,dx = \Theta(n^2\log n), \tag{2.66}$$

so the upper series asymptotically[7] is approaching the logarithm of the number n

$$S_T(n) = \Theta(\log_b n). \tag{2.67}$$

Physical information can be estimated by the same

$$L_T(n) = \Theta(\log_b n). \tag{2.68}$$

Like continuous, this distribution has the form of Hartley's information.

The discrete distributions are much to be handled here. I will not do their list either, but I will just mention the basic idea that for some discrete distribution it makes no sense to separate the physical from the classical information, for some it has. These meaning are hardly in (2.67), but there are and many more turbid cases. Below I will deal with only a few well-known infinite discrete distributions in which there is a clear difference between the two types of information, where is a simple formula of physical information, which in addition is sufficiently instructive, or at least inspirational, to be worth it. Above all there are an indicator of one, indicator of two, or more times that repeatedly give "favorable" events in a series of the same random tests, and a random walk.

[7] Big-Θ (Big-Theta) notation.

2.5.1 Indicators

We throw the cube until it hits the "six" for the first time, or we throw it to another sixth, or up to a third or up to n of them. We repeat a random experiment with a constant probability of a favorable outcome of p and an unfavorable $q = 1 - p$, until a favorable outcome appears n times. It is a discrete (countable infinite) distribution, a complement to a binomial. In the first works I called it the *draw-to-win*, but this type of distribution can also be called *indicators*.

Indicator once

We are reviewing the repetition of the same random experiment until the first appearance of a "favorable" event, for example, throwing a cube before the first "sixth" drop. The probability of a drop of "sixth" after throwing a cube is $p = \frac{1}{6}$, as opposed to the non-falling sixth with the probability $q = \frac{5}{6}$. A favorable outcome can occur either in the first attempt, or in the second, third, ..., or k-th, with an infinite number of probabilities

$$p,\ pq,\ pq^2,\ pq^3,\ \dots,\ pq^{k-1},\ \dots \tag{2.69}$$

whose sum is one. For $k = 1, 2, 3, \dots$ this is the distribution of the probability $p_k = pq^{k-1}$, where we can add $p_0 = 0$ to simplify the formula.

Lemma 2.5.1. *The probability series (2.69) is the distribution.*

Proof. First of all, for any $k = 0, 1, 2, \dots$ is $p_k \ge 0$. Then, we calculate:

$$\sum_{k=0}^{\infty} p_k = p + qp + q^2p + q^3p + \dots = p \cdot (1 + q + q^2 + q^3 + \dots) = p \cdot \frac{1}{1-q} = 1,$$

because $1 - q = p$. Therefore, the given series (2.69) makes the distribution of probability. □

The mean, or *expected value* of the random variable X that takes the value x_i with the probability p_i of the given distribution, is long-term average, we say expected, in repetition of the same experiment. In physics it is denoted by Dirac's notation, with

$$\langle X \rangle = \sum_i p_i x_i, \tag{2.70}$$

In our case, the value of a random variable is an ordinal number of attempts, so x, or k, are natural numbers. For calculations which follow, the methods and results of calculation of the following known infinite sums are useful.

Lemma 2.5.2. *When* $0 < x < 1$ *then:*

1. $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$;
2. $\sum_{k=0}^{\infty} kq^k = \frac{q}{(1-q)^2}$;
3. $\sum_{k=0}^{\infty} k^2q^k = \frac{q(1+q)}{(1-q)^3}$;
4. $\sum_{k=0}^{\infty} k^3q^k = \frac{q(1+4q+q^2)}{(1-q)^4}$.

Proof. 1. Convergence of this series follows from the initial condition $|q| < 1$. If we denote its sum with $Z_1 = 1 + q + q^2 + \dots$, then $qZ_1 = q + q^2 + q^3 + \dots$, so $Z_1 - qZ_1 = 1$ hence $Z_1 = 1/(1-q)$, which is the first statement of the lemma.

2. Convergence follows from the initial condition and $kq^k \to 0$ when $k \to \infty$. Then we have:

$$Z_2 = \sum_{k=0}^{\infty} kq^k = \sum_{k=1}^{\infty} (k-1)q^{k-1} = \frac{1}{q}\sum_{k=1}^{\infty} kq^k - \sum_{k=1}^{\infty} q^{k-1} = \frac{1}{q}Z_2 - Z_1,$$

and hence $Z_2 = q/(1-q)^2$, which is the second claim of the lemma.

3. Convergence follows similar to the previous one. Then:

$$Z_3 = \sum_{k=0}^{\infty} k^2 q^k = \sum_{k=1}^{\infty} (k-1)^2 q^{k-1} = \sum_{k=1}^{\infty} (k^2 - 2k + 1)q^{k-1} =$$

$$= \sum_{k=1}^{\infty} k^2 q^{k-1} - 2\sum_{k=1}^{\infty} kq^{k-1} + \sum_{k=1}^{\infty} q^{k-1} = \frac{1}{q}Z_3 - \frac{2}{q}Z_2 + Z_1,$$

and hence $Z_3 = q(1+q)/(1-q)^3$, which is the third claim of the lemma.

4. Convergence follows in the previous way. Next we have:

$$Z_4 = \sum_{k=0}^{\infty} k^3 q^k = \sum_{k=1}^{\infty} (k-1)^3 q^{k-1} = \sum_{k=1}^{\infty} (k^3 - 3k^2 + 3k - 1)q^{k-1} =$$

$$= \sum_{k=1}^{\infty} k^3 q^{k-1} - 3\sum_{k=1}^{\infty} k^2 q^{k-1} + 3\sum_{k=1}^{\infty} kq^{k-1} - \sum_{k=1}^{\infty} q^{k-1}$$

$$= \frac{1}{q}Z_4 - \frac{3}{q}Z_3 + \frac{3}{q}Z_2 - Z_1,$$

and hence $Z_4 = q(1 + 4q + q^2)/(1-q)^4$, and this is the fourth claim of the lemma. □

Theorem 2.5.3. *Expectation of the number of attempts for the distribution (2.69) is* $\mu_1 = \frac{1}{p}$.

Proof.

$$\mu_1 = \langle k \rangle = \sum_{k=0}^{\infty} kp_k = p\sum_{k=1}^{\infty} kq^{k-1} = \frac{p}{q}\cdot\sum_{k=1}^{\infty} kq^k = \frac{p}{q}\cdot\frac{q}{(1-q)^2} = \frac{1}{p}.$$

□

Variance σ^2 is the expected value of the square of deviation the random variable X from its expected value μ, relative to the given distribution. More precisely

$$\sigma^2 = \langle (X-\mu)^2 \rangle = \sum_i (x_i - \mu)^2 p_i, \tag{2.71}$$

but for its calculation it is often more useful the form of the lemma 2.3.2. As we know, the root of the variance, the number $\sigma = \sqrt{\sigma^2}$, is called *dispersion.*

Theorem 2.5.4. *The variance of the number of attempts to distribute (2.69) is* $\sigma_1^2 = \frac{q}{p^2}$.

Proof. We use the lemma 2.3.2, then the lemma 2.5.2, we get:

$$\sigma_1^2 = \langle k^2 \rangle - \langle k \rangle^2 = \sum_{k=0}^{\infty} k^2 pq^{k-1} - \mu_1^2 =$$

$$= \frac{p}{q}\sum_{k=0}^{\infty} k^2 q^k - \left(\frac{1}{p}\right)^2 = \frac{p(1+q)}{(1-q)^3} - \frac{1}{p^2} = \frac{q}{p^2},$$

because $1 - q = p$. By this the theorem is proved. □

In the case of throwing out fair coins, the drops of "tails" and "heads" are equally probable events, $p = q = \frac{1}{2}$, so the expectation of the number of throws up for "tails" is $\mu_1 = 2$, and the dispersion, the expected dissipation around that number $\sigma_1 = \sqrt{2} \approx 1.41$. In the case of throwing a fair dice, we expect "sixth" after $\mu_1 = 6$ throw with a dispersion $\sigma_1 = \sqrt{30} \approx 5.48$ throws.

The Hartley event Information (2.69) is

$$h_k = -\log_b p_k, \quad p_k \in (0,1]. \tag{2.72}$$

Shannon (S) and physical (L) information in the simple situation of one throw of a coin, or a cube, or a random event with a favorable outcome probability $p \in (0,1)$ and an unfavorable $q = 1 - p$, we consider equals:

$$S = L = -p\log_b p - q\log_b q. \tag{2.73}$$

This means that in the following text, all three, h_k, S and L, are considered as physical information, the starting point of *conservation law* of the summation of information of independent events.

The Shannon information of the distribution (2.69) we calculate by definition. Let's mark it with S_1 and then define physical information (2.69) with $L_1 = S_1$.

Theorem 2.5.5. *The physical information of the distribution (2.69) is*

$$L_1 = \frac{1}{p}(-p\log_b p - q\log_b q).$$

Proof. Using (2.72) we find:

$$\begin{aligned}
L_1 = -\sum_{k=0}^{\infty} p_k \log_b p_k = -\sum_{k=0}^{\infty} q^k p \log_b q^k p = \\
= -\sum_{k=0}^{\infty} q^k p(k\log_b q + \log_b p) \\
= -p\sum_{k=0}^{\infty}(kq^k \log_b q + q^k \log_b p) = \\
= -p(\log_b q)\sum_{k=0}^{\infty} kq^k - p(\log_b p)\sum_{k=0}^{\infty} q^k = \\
= -p(\log_b q)\cdot\frac{q}{(1-q)^2} - p(\log_b p)\cdot\frac{1}{1-q} \\
= -\frac{q\log_b q + p\log_b p}{p},
\end{aligned}$$

and that was to be proven. □

In the base $b = 2$, in the case of throwing the coin, the distribution (2.69) has the information $L_1 = 2$ bits, and in the case of throwing the cube $L_1 \approx 3.90$ bits. Throwing the cube until the first "sixth" drop is a less predictable event than throwing coins until the first "tail" is dropped; it has a higher dispersion with a finishing with more information. The information L_1 is greater than L, formula (2.73), due to additional uncertainty in expectation of a favorable outcome.

Repeated throwing cubes (dices) are independent random events. It does not matter if we throw the same cube several times in succession, or we have more cubes that we throw at once. The expectation of falling two “six” must be the same in such two cases, and they must be the same information. That’s the next topic.

Indicator twice

Randomly, we draw[8] one of the ten numbers $0.1,\dots,9$, hopefully number 3. We repeat the pullout until we get two ($n=2$) of the desired outcome. Again, we have constant probability of a favorable and unfavorable outcome, respectively $p = 1/10$ and $q = 1-p = 9/10$, in each of the drawings. The probability of final winnings is:

$$p^2,\ 2p^2q,\ 3p^2q^2,\ 4p^2q^3,\ \dots,\ (k-1)p^2q^{k-2},\ \dots, \tag{2.74}$$

where we can write $p_0 = p_1 = 0$ and $p_k = (k-1)p^2q^{k-1}$, respectively for $k = 0,1,2,\dots$ draws.

Namely, when a second favorable event occurs in a series of k drawings, then the first favorable could happen to $k-1$ modes. The probability of such is $p_k = p^2q^{k-2}$, and the probability of all $k-1$ modes is $P_k = (k-1)p_k$. The second part of the above claim, that we have the distribution of probability, is contained in the next lemma.

Lemma 2.5.6. *The probability sum (2.74) is one.*

Proof. From $p_k \ge 0$ and $q^k \to 0$ when $k\to\infty$, then because:

$$\sum_{k=0}^{\infty} P_k = p^2 + 2p^2q + 3p^2q^2 + \dots = \frac{p^2}{q}\cdot(q + 2q^2 + 3q^3 + \dots) = \frac{p^2}{q}\cdot\frac{q}{(1-q)^2} = 1,$$

follows the lemma claim.. □

Now we can calculate the expected number of drawings of the indicator twice and the variance of this mean value in the usual way for the probability distribution.

Theorem 2.5.7. *Expectation of the distribution of indicators (2.74) is $\mu_2 = \frac{2}{p}$.*

Proof. In k-th drawing, the favorable outcome has the probability P_k, so it is:

$$\mu_2 = \langle k\rangle = \sum_{k=2}^{\infty} kP_k = \sum_{k=2}^{\infty} k(k-1)p^2q^{k-2} =$$

$$= p^2(2\cdot 1 + 3\cdot 2q + 4\cdot 3q^2 + \dots) = p^2\frac{d}{dq}(2q + 3q^2 + 4q^3 + \dots)$$

$$= p^2\frac{d}{dq}\left[\frac{d}{dq}(q^2 + q^3 + q^4 + \dots)\right] = p^2\frac{d}{dq}\left(\frac{d}{dq}\frac{q^2}{1-q}\right)$$

$$= p^2\frac{d}{dq}\frac{2q - q^2}{(1-q)^2} = \frac{2p^2}{(1-q)^3} = \frac{2}{p},$$

which was to be proved. □

[8] RNG: `https://www.mathgoodies.com/calculators/random_no_custom`

Example 2.5.8. *Let us prove the theorem 2.5.7 elementarily, without a derivative.*

Proof.

$$\mu_2 = \langle k\rangle = \sum_{k=2}^{\infty} kP_k = \sum_{k=2}^{\infty} k(k-1)p^2q^{k-2} =$$

$$= \frac{p^2}{q^2}\left[\left(-q + \sum_{k=1}^{\infty} k^2q^k\right) - \left(-q + \sum_{k=1}^{\infty} kq^k\right)\right] = \frac{p^2}{q^2}\left(\sum_{k=1} k^2q^k - \sum_{k=1}^{\infty} kq^k\right)$$

$$= \frac{p^2}{q^2}\left[\frac{q(1+q)}{(1-q)^3} - \frac{q(1-q)}{(1-q)^3}\right] = \frac{p^2}{q^2}\frac{2q^2}{(1-q)^3} = \frac{2}{p}.$$

□

Note that the expectation of this distribution is twice the expectation of the previous one, $\mu_2 = 2\mu_1$. This is exactly what we expect for physical information, that this can be understood as being repeated twice before. Consistent, it is not surprising that the variance of this distribution is twice as large as the previous one, and this is confirmed by the following theorem.

Theorem 2.5.9. *The variance of indicators distribution (2.74) is* $\sigma_2^2 = \frac{2q}{p^2}$.

Proof.

$$\sigma_2^2 = \langle k^2\rangle - \langle k\rangle^2 = \sum_{k=2}^{\infty} k^2P_k - \left(\sum_{k=2}^{\infty} kP_k\right)^2 =$$

$$= \sum_{k=2}^{\infty} k^2(k-1)p^2q^{k-2} - \mu^2 = p^2\frac{\partial}{\partial q}\left(\sum_{k=2}^{\infty} k^2q^{k-1}\right) - \mu^2$$

$$= p^2\frac{\partial}{\partial q}\left[\frac{\partial}{\partial q}\left(\sum_{k=2}^{\infty} kq^k\right)\right] - \mu^2 = p^2\frac{\partial}{\partial q}\left[\frac{\partial}{\partial q}\left(-q + \frac{q}{(1-q)^2}\right)\right] - \mu^2$$

$$= p^2\frac{\partial}{\partial q}\left[-1 + \frac{1+q}{(1-q)^3}\right] - \left(\frac{2}{p}\right)^2$$

$$= p^2\frac{4+2q}{(1-q)^4} - \frac{4}{p^2} = \frac{2q}{p^2}.$$

□

In the distribution (2.74), the probability p_k of a favorable event is less than the probability P_k of all favorable, same events realized on any of the $k-1$ modes. That is why Hartley's information of the first are larger than the corresponding of the second:

$$h_k = -\log_b p_k, \quad H_k = -\log_b P_k. \tag{2.75}$$

However, only the second represent a distribution, so the mean value of the first in relation to that distribution is called physical information and we denote it L_2, as opposed to the mean of the second by the same distribution that is Shannon's information.

Theorem 2.5.10. *Physical information of the indicator (2.74) is*

$$L_2 = \frac{2}{p}(-p\log_b p - q\log_b q).$$

Proof. We calculate in the order:

$$L_2 = \sum_{k=2}^{\infty} P_k h_k =$$

$$= -\sum_{k=2}^{\infty} P_k \log_b p_k$$

$$= -\sum_{k=2}^{\infty}(k-1)q^{k-2}p^2 \log_b q^{k-2}p^2$$

$$= -p^2 \sum_{k=2}^{\infty}(k-1)q^{k-2}(\log_b q^{k-2} + \log_b p^2) =$$

$$= -p^2(\log_b q)\sum_{k=2}^{\infty}(k-2)(k-1)q^{k-2} - p^2(2\log_b p)\sum_{k=2}^{\infty}(k-1)q^{k-2}$$

$$= -p^2(\log_b q)(1\cdot 2q + 2\cdot 3q^2 + 3\cdot 4q^3 + \dots) - p^2(2\log_b p)(1 + 2q + 3q^2 + \dots)$$

$$= -p^2(\log_b q)\frac{\partial}{\partial q}(q^2 + 2q^3 + 3q^4 + \dots) - p^2(2\log_b p)\frac{\partial}{\partial q}(q + q^2 + q^3 + \dots)$$

$$= -p^2(\log_b q)\frac{\partial}{\partial q}\Big[q^2\frac{\partial}{\partial q}(q + q^2 + q^3 + \dots)\Big] - p^2(2\log_b p)\frac{\partial}{\partial q}\frac{q}{1-q}$$

$$= -p^2(\log_b q)\frac{\partial}{\partial q}\Big(q^2\frac{\partial}{\partial q}\frac{q}{1-q}\Big) - p^2(2\log_b p)\frac{1}{(1-q)^2}$$

$$= -p^2(\log_b q)\frac{\partial}{\partial q}\frac{q^2}{(1-q)^2} - 2\log_b p$$

$$= -p^2(\log_b q)\frac{2q}{(1-q)^3} - 2\log_b p$$

$$= -\frac{2q}{p}\log_b q - 2\log_b p$$

$$= \frac{2}{p}(-p\log_b p - q\log_b q),$$

and that was to be proven. □

Because the information $L_2 = 2L_1$ we call physical, the Shannon's information of this distribution would have been

$$S_2 = \sum_{k=2}^{\infty} P_k H_k, \tag{2.76}$$

and it is less than physical, since $H_k < h_k$ for every $k > 2$. In order not to challenge the value of Shannon information, the excess L_2 in relation to S_2 should be interpreted in some physical way. I suggest that this be in line with the principle of least information.

It is precisely this skimping with the emission of information (from uncertainty) that makes the realization of more likely events more frequent. We can further note that the same principle applies to packing and hiding information, in various ways. Locked information in a complex event can be unlocked and unpacked, but it costs.

In the aforementioned random drawback with return, one by one of ten numbers from the set $\{0, 1, \dots, 9\}$, to the second draw of number 3, the basic probabilities are $p = 0.1$ and $q = 0.9$. The probability of the first twenty winnings (2.74) are in the table 2.1.

k:	2	3	4	5	6	7	8	9	10	11
P_k:	0.010	0.018	0.024	0.029	0.033	0.035	0.037	0.038	0.039	0.039
k:	12	13	14	15	16	17	18	19	20	21
P_k:	0.038	0.038	0.037	0.036	0.034	0.033	0.031	0.030	0.029	0.027

Table 2.1: Distribution of probability

We see that these probabilities grow first to $k = 10$ and then they decline. In general, the function of this probability $P(k)$, for $k \in \mathbb{R}$, has a maximum of $P(k_0)$ for $k_0 = 1 - \ln^{-1} q$. The expectation of distribution (2.74) is $\mu_2 = 20$, with dispersion $\sigma_2 \approx 13.416$. Physical information in base e is $L_2 \approx 6.502$, or $L_2 \approx 9.380$ bits in base 2 of the logarithm.

Indicator n times

Let us now observe a repeat of a random experiment with a favorable outcome of a constant probability $0 < p < 1$ and an unfavorable of probability $q = 1 - p$ until a favorable outcome occurs in three ($n = 3$) times. If this collection occurred in k reps ($k \ge n$), then in the previous $k - 1$ repetition a favorable outcome occurred two ($n - 1 = 2$) times, which can be achieved on $\binom{k-1}{2}$ of equal methods (ways).

In each of these equitable ways, a favorable outcome occurred three times, and unfavorable $k - 3$ times, because there were total k reps. The likelihood of one of these collections, the "three favorable" in the k repetition, is $p_k(3) = p^3 q^{k-3}$, and the probability of all equals is $P_k(3) = \binom{k-1}{2} \cdot p_k(3)$.

In general, the probability of only one collection in $n = 1, 2, 3, \dots$ of such favorable outcomes in $k = 0, 1, 2, \dots$ experiments, and the probability of all, are respectively:

$$p_k(n) = p^n q^{k-n}, \quad P_k(n) = \binom{k-1}{n-1} p_k, \tag{2.77}$$

where we mean $p_0(n) = \dots = p_{n-1}(n) = 0$. Also, $p_k = p_k(n)$ and $P_k = P_k(n)$, unless explicitly stated otherwise. It is clear that the infinite series P_k makes the distribution of probability, and that the series of probabilities p_k do not form a distribution.

Example 2.5.11. *Prove that the set $P_k(n)$ for $k = 1, 2, 3, \dots$ and $n = 3$ is a distribution.*

Proof. Followed from:

$$\sum_{k=1}^{\infty} P_k(3) = p^3 \sum_{k=1}^{\infty} \frac{(k-1)(k-2)}{2} q^{k-3} = \frac{p^3}{2} \frac{\partial^2}{\partial q^2} \sum_{k=0}^{\infty} q^k =$$

$$= \frac{p^3}{2} \frac{\partial^2}{\partial q^2} \frac{1}{1-q} = \frac{p^3}{2} \frac{\partial}{\partial q} \frac{1}{(1-q)^2} = \frac{p^3}{(1-q)^3} = 1,$$

because the $p + q = 1$. □

Let us note that similar proof that P_k is a distribution is valid for arbitrarily n, whereby for forming the numerator $(k-1)(k-2)\dots(k-n+1)$ we take $n-1$ partial derivatives of the geometric series, which are then abbreviated with the denominator $1 \cdot 2 \cdot 3 \dots (n-1)$. This is obvious, and then it is obvious that the next lemma is valid.

Lemma 2.5.12. *For $n,k \in \mathbb{N}$ and $k \ge n$ is*

$$\sum_{k=n}^{\infty}\binom{k-1}{n-1}q^{k-n} = \binom{n-1}{n-1} + \binom{n}{n-1}q + \binom{n+1}{n-1}q^2 + \dots = (1-q)^{-n},$$

where $q \in (0,1)$.

Calculating physical information for (2.77) analogously to theorems 2.5.10, with the help of the previous scheme, is slightly more complicated:

$$L_n = -\sum_{k=n}^{\infty}\binom{k-1}{n-1}q^{k-n}p^n \log_b q^{k-n}p^n =$$

$$= -p^n(\log_b q)\sum_{k=n}^{\infty}\binom{k-1}{n-1}(k-n)q^{k-n} - np^n(\log_b p)\sum_{k=n}^{\infty}\binom{k-1}{n-1}q^{k-n}$$

$$= -p^n(\log_b q)q\sum_{k=n}^{\infty}\binom{k-1}{n-1}(k-n)q^{k-n-1} - np^n(\log_b p)\cdot p^{-n}$$

$$= -qp^n(\log_b q)\frac{\partial}{\partial q}\sum_{k=n}^{\infty}\binom{k-1}{n-1}q^{k-n} - n\log_b p$$

$$= -qp^n(\log_b q)\frac{\partial}{\partial q}(1-q)^{-n} - n\log_b p$$

$$= -qp^n(\log_b q)\cdot n(1-q)^{-n-1} - n\log_b p$$

$$= -n\frac{q}{p}\log_b q - n\log_b p$$

$$= \frac{n}{p}(-p\log_b p - q\log_b q).$$

This proved the following theorem.

Theorem 2.5.13. *Physical information (2.77) is $L_n = nL_1$, where L_1 is the case for $n = 1$.*

I'm not sure if the distribution (2.77) is considered in mathematics, and I believe that it is unknown physics, so I will mention a few of its basic characteristics. Its significance for (my) theory of information, in particular its relation to physical action, can be demonstrated for an area analogous to the Second Kepler law[9] and information, as well as the relation of dispersion of this distribution with the spread of "random walk", but of everything listed in I will continue to demonstrate only this last.

Theorem 2.5.14. *Expectation of the distribution (2.77) is $\mu_n = \frac{n}{p}$.*

Proof.

$$\mu_n = \langle k\rangle = \sum_{k=n}^{\infty}kP_k = \sum_{k=n}^{\infty}k\binom{k-1}{n-1}p^nq^{k-n} =$$

$$= p^n\sum_{k=n}^{\infty}k\cdot\frac{(k-1)!}{(n-1)![(k-1)-(n-1)]!}q^{(k-1)-(n-1)}$$

$$= np^n\cdot\sum_{k=n}^{\infty}\frac{k!}{n!(k-n)!}q^{k-n} = np^n\cdot(1-q)^{-(n+1)} = \frac{n}{p}.$$

□

[9] see [3], Gravitation on page 59.

Theorem 2.5.15. *The variance of the distribution (2.77) is* $\sigma_n^2 = \frac{nq}{p^2}$.

Proof. First we calculate:

$$\langle k^2\rangle = \sum_{k=n}^{\infty} k^2 P_k = \sum_{k=n}^{\infty} k^2 \binom{k-1}{n-1} p^n q^{k-n} = np^n q^{-n} \sum_{k=n}^{\infty} k\binom{k}{n} q^k =$$

$$= np^n q^{-n+1} \frac{\partial}{\partial q} \sum_{k=n}^{\infty} \binom{k}{n} q^k = np^n q^{-n+1} \frac{\partial}{\partial q} \left[q^n \sum_{k=n}^{\infty} \binom{k}{n} q^{k-n} \right]$$

$$= np^n q^{-n+1} \frac{\partial}{\partial q} \left[q^n (1-q)^{-(n+1)} \right]$$

$$= np^n q^{-n+1} [nq^{n-1}(1-q)^{-n-1} + q^n(-n-1)(1-q)^{-n-2}(-1)]$$

$$= np^n q^{-n+1} q^{n-1} (1-q)^{-n-2} [n(1-q) + q(n+1)]$$

$$= np^{-2}(n - nq + nq + q) = \frac{n^2 + nq}{p^2}.$$

Then we find the variance:

$$\sigma_n^2 = \langle k^2\rangle - \langle k\rangle^2 = \frac{n^2+nq}{p^2} - \frac{n^2}{p^2} = \frac{nq}{p^2},$$

and this was to be proven. □

You will find the basic features of mathematical expectations and variances, and something about the indicators, in other places. For example, it the interesting lecture "Expectation & Variance" by Alberto Mayer (see [10]), or something similar.

2.5.2 Random walk of point

The point T is moving by abscise, x-axes, starting from the origin ($x = 0$). From the position x with one step it can only reach one of the two positions $x \pm 1$, with the probability of moving to the right $p \in (0,1)$ and to the left $q = 1 - p$. With $P_k(x)$ we denote the probability that the point T in the k-th step ($k = 0, 1, 2, \dots$) is found at the position x.

An analysis of this discrete *random walk* with the constant probability p, which follows, will show that in the case $p = q = \frac{1}{2}$ its mean value is the origin, and in the case of $p > q$ by the time the given point moves to the right. In any case, from the probability $P_k(x)$ it passes to the probability $P_k(-x)$ by replacing the values p and q (table 2.2).

$P_k(x)$	$x=-4$	$x=-3$	$x=-2$	$x=-1$	$x=0$	$x=1$	$x=2$	$x=3$	$x=4$
$k=0$					1				
$k=1$				q	0	p			
$k=2$			q^2	0	$2pq$	0	p^2		
$k=3$		q^3	0	$3q^2p$	0	$3p^2q$	0	p^3	
$k=4$	q^4	0	$4q^3p$	0	$6q^2p^2$	0	$4p^3q$	0	p^4

Table 2.2: Probability of steps abscise.

When the point T in the even number position $x = 0, \pm 2, \pm 4, \dots$ in the next step it will be found in an odd number position $x+1$ with probability p, or also odd $x-1$ with probability

q, but vice versa, after an odd position in the next step, will be found in the even. Therefore, starting from the even number of zero ($x = 0$), after the even number k of the steps the point will always be found in some even position x, and after an odd number of steps it will be found in the odd number.

From the table 2.2 it is seen that the distribution of the positions of the random walk of the point abscissa for the given number of steps is similar to the binomial (2.15), and little to the indicators (2.77). In order for the T point starting from the origin to arrive at the abscissa $x \in \{0, 1, 2, \dots\}$ after k steps, it must make a total of exactly x steps to the right (probability p) and exactly $k - x$ steps left (probability q), which is a choice of probability

$$p_k(x) = \begin{cases} p^x q^{k-x} & k, x \text{ -- equal parity,} \\ 0 & k, x \text{ -- different parity.} \end{cases} \tag{2.78}$$

These choices are the same as dials of x elements from a set of k elements, i.e.

$$\binom{k}{x} = \frac{k!}{x!(k-x)!}. \tag{2.79}$$

Therefore, the probability of finding a given point at the position x after the k step is

$$P_k(x) = \begin{cases} \binom{k}{x} p_k(x) & k, x \ = \ \text{parity,} \\ 0 & k, x \ \neq \ \text{parity.} \end{cases} \tag{2.80}$$

If we do not have to look at this parity-oddity, let's notice that it is

$$\lim_{\alpha \to 0} \alpha \ln \alpha = 0, \tag{2.81}$$

and that in the extreme case we can symbolically write $0 \ln 0 = 0$. This will make our formulas simpler.

Random walking information

From the previous it is clear that the inequality of probability is valid

$$p_k(x) \le P_k(x), \tag{2.82}$$

and that only (2.80) is distribution. Namely, for fixed $k \in \mathbb{N}$ we add all integers x (to absolute $|x|$ delete the sign of x) and we find:

$$\sum_x P_k(x) = \sum_x \binom{k}{|x|} p_k(x) = (p + q)^k = 1, \tag{2.83}$$

because $p + q = 1$. Then

$$-\ln p_k(x) \ge -\ln P_k(x). \tag{2.84}$$

Hartley's information is less if the probability is greater. Therefore, if we define the mean of this information in relation to the distribution (2.80), we have two different values:

$$\begin{cases} L_k = -\sum_x P_k(x) \ln p_k(x) & \text{-- physical information} \\ S_k = -\sum_x P_k(x) \ln P_k(x) & \text{-- classical information} \end{cases} \tag{2.85}$$

and $L_k \ge S_k$. The first is a new kind of information, and the second one is known by Shannon.

The name *physical information* complies with the conservation law

$$L_k = kL_1, \tag{2.86}$$

which we presume to be valid in the world of physics. Intuitively, if each of the steps k is equally coincidental, then the amount of uncertainty of the k step is exactly equal to k times the uncertainty of one step, so that must also be the definition of physical information. That (2.86) is indeed proved correctly is the case (theorem 2.3.6). However, because of (2.84), this means that Shannon's information is not "physical".

We will find a deeper sense of inequality (2.84) in the *principle of information*[10], that nature austere with the emission of information. Retaining the Shannon's definition of information, the physical one we can try to explain physically, I believe, by this principle. In increasingly complex physical systems (here in a number of steps) there is an increasing *surplus information* that is supposedly somehow to be either removed or locked. This rising surplus in the multitude is opposed to this principle of minimalism that requires reduction, and this is contrary to the law of conservation that says that information can only be transformed from one form to another but cannot change the total quantity.

So I assume that the organization itself, in particular, has the physical meaning of accumulated information. Consequently, the surplus of the structure means excess information, and then the surplus of choice and options, and therefore the surplus of interactions. When the principles of the least action of physics apply to the structure of information, they are then the structures of *inanimate beings*, and those who have excess of the information I call *living beings*. It has been written about this in my texts cited in the bibliography.

The steps expectation

The mathematical expectation (mean value) of the point T on the abscissa in the case of $k = 0$ of the step is $\mu_0 = 0 \cdot 1 = 0$. This, of course, is true, because the point does not move from the origin. In the case of one step ($k = 1$), the expected position at x-axis is $\mu_1 = -1 \cdot q + 1 \cdot p = p - q$, which is a little to the right of the origin if $p > q$. After two steps ($k = 2$) we expect the mean value of the aperture point:

$$\mu_2 = -2 \cdot q^2 + 0 \cdot 2pq + 2 \cdot p^2 = 2(p^2 - q^2) = 2(p-q)(p+q) = 2(p-q).$$

After three steps $\mu_3 = -3q^3 - 3q^2p + 3p^2q + 3p^3 = 3(p^3 - q^3) + 3(p^2q - q^2p) =$

$$= 3(p-q)(p^2 + pq + q^2) + 3pq(p-q) = 3(p-q)(p^2 + 2pq + q^2) = 3(p-q)(p+q)^2,$$

therefore $\mu_3 = 3(p-q)$. In general, after the k steps, the expected position of the T point is given in the following theorem.

Theorem 2.5.16. *The position expectation of the distribution (2.80) for given k is $\mu_k = k(p-q)$.*

Proof. We proceed from the definition of expectation:

$$\mu_k = \langle x \rangle = \sum_x x P_k(x) = -\sum_{x<0} |x| \binom{k}{|x|} p_k(x) + \sum_{x \ge 0} x \binom{k}{x} p_k(x).$$

[10] see in more detail in the book Multiplicities [3]

Switching $x \to -x$ in the probability $p_k(x)$ are replaced p and q. That is why:

$$\mu_k = \sum_{x\ge 0} x\binom{k}{x}(p^x q^{k-x} - q^x p^{k-x}) =$$

$$= p\frac{\partial}{\partial p}\sum_{x\ge 0}\binom{k}{x}p^x q^{k-x} - q\frac{\partial}{\partial q}\sum_{x\ge 0}\binom{k}{x}q^x p^{k-x}$$

$$= p\frac{\partial}{\partial p}(p+q)^k - q\frac{\partial}{\partial q}(p+q)^k$$

$$= kp(p+q)^{k-1} - kq(q+p)^{k-1},$$

then from the $p+q=1$ the requested statement is followed. □

The result $\mu_k = k\mu_1$ confirms the independence of the probability of the steps, and then the very idea of physical information (2.86). On the other hand, with each step, the average expected position of the material point T moves to the right for $p-q$. In analogy to the physical movement of the particle, here is the question of where the redundant information that is coming out then goes, assuming that it does not accumulate?

The moving particle communicates with the space, so much more that the incremental information $L_1 = -p\ln p - q\ln q$ is greater. This increment is greater when the probability difference between p and q is smaller, or when the particle velocity is smaller. In other words, the relative observer observes weaker communication of faster particles, which means that it observes a slower course of time of faster particles. We are again here in the thesis (see [3]) that the present is created by communication and that the perception of probability in the physical world is a relative matter.

Steps variance

The next important value of the distribution of the variable is the *variance* variable. We know that this is the expectation of the squares of the difference of the x variable and its expectation $\mu = \langle x\rangle$, that is, the difference in the expectation of the square of the variable and the square of its expectation (lemma 2.3.2).

Applied to the case $k=0$, the variance is $\sigma_0^2 = 0\cdot 1 - \mu_0^2 = 0$, which is logical because there is no movement. In the case of $k=1$ we have $\sigma_1^2 = (-1)^2 q + (+1)^2 p - \mu_1^2 = q+p-(p-q)^2 = 4pq$. When $k=2$ variance is $\sigma_2^2 = 4q^2 + 4p^2 - 4(p-q)^2 = 8pq$. For $k=3$ we have:

$$\sigma_3^2 = 9q^3 + 3q^2 p + 3p^2 q + 9p^3 - 9(p-q)^2 =$$

$$= 9(q^3+p^3) + 3pq(q+p) - 9(p^2 - 2pq + q^2) =$$

$$= 9(q^2 - pq + p^2) + 3pq - 9(p^2 - 2pq + q^2) = 12pq.$$

The following theorem confirms these special cases.

Theorem 2.5.17. *The variance of the distribution (2.80) for given k is $\sigma_k^2 = 4kpq$.*

Proof. Using the previous calculate:

$$\langle x^2\rangle = \sum_x x^2 P_k(x) = \sum_{x\ge 0} x^2\binom{k}{x}(p^x q^{k-x} + q^x p^{k-x}) =$$

$$
\begin{aligned}
&= p\frac{\partial}{\partial p}\sum_{x\ge 0} x\binom{k}{x}p^x q^{k-x} + q\frac{\partial}{\partial q}\sum_{x\ge 0} x\binom{k}{x}q^x p^{k-x}\\
&= p\frac{\partial}{\partial p}p\frac{\partial}{\partial p}\sum_{x\ge 0}\binom{k}{x}p^x q^{k-x} + q\frac{\partial}{\partial q}q\frac{\partial}{\partial q}\sum_{x\ge 0}\binom{k}{x}q^x p^{k-x}\\
&= p\frac{\partial}{\partial p}p\frac{\partial}{\partial p}(p+q)^k + q\frac{\partial}{\partial q}q\frac{\partial}{\partial q}(q+p)^k\\
&= p\frac{\partial}{\partial p}pk(p+q)^{k-1} + q\frac{\partial}{\partial q}qk(q+p)^{k-1}\\
&= kp[1+(k-1)p] + kq[1+(k-1)q]\\
&= 2kpq + k^2(p^2+q^2).
\end{aligned}
$$

Then form the variance:

$$
\sigma_k^2 = \langle x^2\rangle - \langle x\rangle^2 = 4kpq,
$$

and that's what it was supposed to prove. □

Analogy with binary distribution is only partial, because there is the expectation (np) greater than this, and the variance (npq) is smaller. The interpretations are also different. The material particle T would increase its information if it moves without communication in the environment, at least for the information of its past information. Communication with the vacuum keeps its information constant.

2.5.3 Front's random walk

We observe the integer abscise (x-axes) positions and the T material point that originated in $x = 0$ and made $k = 0, 1, 2, \dots$ steps. Each step is of the same length (± 1) with the probability $p \in (0, 1)$ to the right and $q = 1 - p$ to the left. When T has reached the abscise $x \in \mathbb{Z}$ its next abscise is a random number $x + 1$ of probability p or the number $x - 1$ of probability q. Assume that $p > q$, as in the previous one.

It is easy to understand that in an odd number of steps $k \in \{1, 3, 5, \dots\}$ the position of the given point will be an odd integer $x \in \{\pm 1, \pm 3, \pm 5, \dots\}$, and in even $k \in \{0, 2, 4, \dots\}$ is an even number $x \in \{0, \pm 2, \pm 4, \dots\}$. To the place $x \ge 0$ abscise point arrives after x steps to the right and $k - x$ steps left ($k \ge x$). This can be achieved on

$$
C_x^k = \binom{k}{x} = \frac{k!}{x!(k-x)!} \tag{2.87}
$$

ways, each of probability

$$
p_k(x) = p^x q^{k-x}, \tag{2.88}
$$

and the total probability

$$
P_k(x) = \binom{k}{|x|}p_k(x), \tag{2.89}
$$

where $|x| = x$ for $x \ge 0$ and $|x| = -x$ for $x < 0$. Thus we form the table 2.2, whereby we notice that by replacing $x \to -x$ in the probabilities $p_k(x)$ or $P_k(x)$ it is necessary to replace p with q. It is clear that $p_k(x) \le P_k(x)$ and that $P_k(x)$ is a binomial distribution.

After enough steps, such as the wave propagating from one center, the data point arrives to each preposition abscise $x \in \mathbb{Z}$. During its expansion, it has less and less probability of appearance both at the front of the wave and at its depth. Unlike the previous one, the consideration of random steps, where the distribution of the probability of a constant k over all possible x is discussed, we now consider the probability of the moving *wave phase*, which is moving away from the starting point by step k.

We will note that these are also some probability distributions and that they are formally identical to those of the *indicators* section (repetition-to-collection). Namely, throwing the dice until the first drop of "sixth" is equivalent to spreading the position of the given point T to the first reaching of the pre-given place x. Throwing the cube up to the second drop "sixth" is formally equivalent to the second waves position of the datum. In general, when we throw the cube until we collect k "sixth", it will be as if we are collecting k fronts of the wave of the spread of the position of the given point.

Let's do it in a slightly different way to notice new repercussions, first of all in the theory of physical information in a slightly new sense.

Distribution of the front line

The zero or frontal expansion of the position the material point T forms on abscissa $x \ge 0$ when in all previous $k = 0,1,2,\dots$ steps were constantly going to the right, except in the last. This is the event of probability $p^k q$, where $k = x$, which I call the *wave front* of moving the given point. Since these random positions always have some frontal boundary, this probability $qP_k(k)$ make some distribution:

$$q,\ qp,\ qp^2,\ \dots,\ qp^k,\dots \quad (k = 0,1,2,\dots). \tag{2.90}$$

And indeed:

$$\sum_{k=0}^{\infty} qp^k = q + qp + qp^2 + \dots = q(1 + p + p^2 + \dots) = \frac{q}{1-p} = 1, \tag{2.91}$$

because $p + q = 1$. This distribution is similar to that known to us (2.69), the distribution of the indicators of one appearance, with the expectation of the number of steps $\frac{1}{p}$.

We distinguish two types of random variables, one is the end position of an abscissa point (x), and the second is the number of steps (k). From the proof of the following theorem it will be clear that the expectation of the abscise and the number of steps, as well as the variance of these two frontal wave variables are equal, $\mu_0(x) = \mu_0(k)$ and $\sigma_0^2(x) = \sigma_0^2(k)$.

Theorem 2.5.18. *Distribution of the front (2.90) has its expectation and variance:*

$$\mu_0(k) = \frac{p}{q}, \quad \sigma_0^2(k) = \frac{p}{q^2},$$

in relation to positions abscissa.

Proof. Using lemma 2.5.2. The mathematical expectation of the step is:

$$\mu_0(k) = \langle k \rangle = \sum_{k=0}^{\infty} kqp^k = 0\cdot q + 1\cdot qp + 2\cdot qp^2 + \dots + xqp^x + \dots =$$

$$= q(p + 2p^2 + \dots + kp^k + \dots) = q\sum_{k=0}^{\infty} kp^k = q\frac{p}{(1-q)^2} = \frac{p}{q}.$$

The variance of the step is:

$$\sigma_0^2(k) = \langle (k-\mu_0)^2\rangle = \langle k^2\rangle - \langle k\rangle^2 = \sum_{k=0}^{\infty} k^2 qp^k - \mu_0^2(k) =$$

$$= q\sum_{k=0}^{\infty} k^2 p^k - \left(\frac{p}{q}\right)^2 = q\frac{p(1+p)}{(1-p)^3} - \frac{p^2}{q^2} = \frac{p}{q^2}.$$

These two had to be proven. □

Physical information L_0 does not depend on these numbers separately, except from the conditions $k = x$ and, of course, from the distribution itself (2.90). It is equal to Shannon's information S_0.

Theorem 2.5.19. *The physical information of the distribution front (2.90) is*

$$L_0 = -\frac{1}{q}(p\ln p + q\ln q).$$

Proof.

$$L_0 = -\sum_{k=0}^{\infty} qp^k \ln qp^k = -q\ln q\sum_{k=0}^{\infty} p^k - q\ln p\sum_{k=0}^{\infty} kp^k =$$

$$= -\frac{q\ln q}{1-p} - q(\ln p)\frac{p}{(1-p)^2} = -\frac{q\ln q}{1-p} - \frac{p\ln p}{1-p}$$

$$= \frac{1}{q}(-p\ln p - q\ln q),$$

and that was to be proven. □

Distribution the first line behind the front

In the next line of the spreading wave of the given point, the first behind the frontal, the number of steps k and abscise x must have the same parity (both are even or both are odd numbers). That is why this expansion of the achieved points even more resembles the physical waves. Additionally, until this line is reached, one additional movement is required, to the left, with all the movements to the right with which the wave front arrives.

Moving to the left in our case equals the probability $q = 1-p$ as the absence of movement to the right behind the front, and the probabilities of these lines of wave form the distribution:

$$q^2,\ 2q^2p,\ 3q^2p^2,\ \dots,\ kq^2p^{k-1},\ \dots,\quad (k = 1,2,3,\dots). \tag{2.92}$$

That this is really a probability distribution follows from $kq^2p^{k-1} \in (0,1)$ and the sum:

$$q^2 + 2q^2p + 3q^2p^2 + \dots + kq^2p^{k-1} + \dots = \frac{q^2}{p}\sum_{k=0}^{\infty} kp^k = \frac{q^2}{p}\frac{p}{(1-p)^2} = 1.$$

In other words, because this random walk always has this phase, positions behind the zero wave line, so this probability also makes some distribution. In the first line behind the front, as opposed to the frontal, now probability (2.92), they have different expected abscissa values (x) then the expected value of the step number (k), i.e. $\mu_1(x) \neq \mu_1(k)$, however, the variances of these variables are the same, $\sigma_1^2(x) = \sigma_1^2(k)$. This is the content of the following two theorems.

Theorem 2.5.20. *The expectation of abscissa and its phase variance, of the distribution (2.92), is:*

$$\mu_1(x) = -1 + \frac{2p}{q}, \quad \sigma_1^2(x) = \frac{2p}{q^2}.$$

Proof. The expectation is calculated (the lemma 2.5.2), orderly:

$$\mu_1(x) = \langle x\rangle = -1\cdot q^2 + 0\cdot 2q^2p + 1\cdot 3q^2p^2 + 2\cdot 4q^2p^3 + \dots + (k-1)(k+1)q^2p^k + \dots =$$

$$= q^2\sum_{k=0}^{\infty}(k-1)(k+1)p^k = q^2\sum_{k=0}^{\infty}(k^2-1)p^k = q^2\left(\sum_{k=0}^{\infty}k^2p^k - \sum_{k=0}^{\infty}p^k\right)$$

$$= q^2\left[\frac{p(1+p)}{(1-p)^3} - \frac{1}{1-p}\right] = \frac{p(1+p)}{q} - q = -1 + \frac{2p}{q}.$$

The variance apscise distribution (2.92) is:

$$\sigma_1^2(x) = \langle x^2\rangle - \langle x\rangle^2 =$$

$$= q^2\sum_{k=0}^{\infty}(k-1)^2(k+1)p^k - \mu_2^2(x) = q^2\sum_{k=0}^{\infty}(k^3-k^2-k+1)p^k - \left(-1+\frac{2p}{q}\right)^2 =$$

$$= q^2\left[\frac{p(1+4p+p^2)}{(1-p)^4} - \frac{p(1+p)}{(1-p)^3} - \frac{p}{(1-p)^2} + \frac{1}{1-p}\right] - \left(\frac{3p-1}{q}\right)^2 = \frac{2p}{q^2}.$$

These two had to be proven. □

From the table 2.2 we find that the expectations for abscise and the number of steps can be different, which is proved by the following theorem. In that other sloping line on the right, the triangle of the table, the first abscise is negative ($x = -1$), but this becomes irrelevant by squaring, so the variance of the abscise and the number of steps is the same.

Theorem 2.5.21. *Expectation of the number of steps and its distribution variance (2.92) are:*

$$\mu_1(k) = 1 + \frac{2p}{q}, \quad \sigma_1^2(k) = \frac{2p}{q^2},$$

where $q = 1 - p$.

Proof. Expectation is now:

$$\mu_1(k) = \langle k\rangle = 1\cdot q^2 + 2\cdot 2q^2p + 3\cdot 3q^2p^2 + 4\cdot 4q^2p^3 + \dots =$$

$$= \frac{q^2}{p}\sum_{k=0}^{\infty}k^2p^k = \frac{q^2}{p}\frac{p(1+p)}{(1-p)^3} = \frac{1+p}{1-p} = 1 + \frac{2p}{q}.$$

The variance of the steps is:

$$\sigma_1^2(k) = \langle (k-\mu)^2\rangle = \langle k^2\rangle - \langle k\rangle^2 =$$

$$= (1^2\cdot q^2 + 2^2\cdot 2q^2p + 3^2\cdot 3q^2p^2 + 4^2\cdot 4q^2p^3 + \dots) - \mu_2^2(k)$$

$$= \frac{q^2}{p}\sum_{k=0}^{\infty}k^3p^k - \left(1+\frac{2p}{q}\right)^2 = \frac{q^2}{p}\frac{p(1+4p+p^2)}{(1-p)^4} - \left(\frac{1+p}{q}\right)^2 =$$

$$= \frac{1+4p+p^2}{q^2} - \frac{1+2p+p^2}{q^2} = \frac{2p}{q^2}.$$

These are the two results that were to be proven. □

Example 2.5.22. *Prove the theorem 2.5.21 with the derivative.*

Proof. Expectation is also:

$$\mu_1(k) = \langle k\rangle = 1\cdot q^2 + 2\cdot 2q^2p + 3\cdot 3q^2p^2 + 4\cdot 4q^2p^3 + \dots =$$
$$= q^2\frac{\partial}{\partial p}(p + 2p^2 + 3p^3 + 4p^4 + \dots)$$
$$= q^2\frac{\partial}{\partial p}\left[p\frac{\partial}{\partial p}(p + p^2 + p^3 + p^4 + \dots)\right]$$
$$= q^2\frac{\partial}{\partial p}\left[p\frac{\partial}{\partial p}\frac{p}{1-p}\right]$$
$$= q^2\frac{\partial}{\partial p}\frac{p}{(1-p)^2} = q^2\frac{1+p}{(1-p)^3}$$
$$= \frac{1+p}{1-p} = 1 + \frac{2p}{q}.$$

For the variance of the steps we get:

$$\sigma_1^2(k) = \langle (k-\mu)^2\rangle = \langle k^2\rangle - \langle k\rangle^2 =$$
$$= (1^2\cdot q^2 + 2^2\cdot 2q^2p + 3^2\cdot 3q^2p^2 + 4^2\cdot 4q^2p^3 + \dots) - \mu_1^2(k)$$
$$= q^2\frac{\partial}{\partial p}\left[p\frac{\partial}{\partial p}\left(p\frac{\partial}{\partial p}\sum_{k=1}^{\infty}p^k\right)\right] - \mu_1^2(k)$$
$$= q^2\frac{\partial}{\partial p}\left[p\frac{\partial}{\partial p}\left(p\frac{\partial}{\partial p}\frac{p}{1-p}\right)\right] - \mu_1^2(k)$$
$$= q^2\frac{\partial}{\partial p}\left[p\frac{\partial}{\partial p}\frac{p}{(1-p)^2}\right] - \left(\frac{1+p}{1-p}\right)^2$$
$$= q^2\frac{\partial}{\partial p}\frac{p(1+p)}{(1-p)^3} - \frac{(1+p)^2}{(1-p)^2}$$
$$= \frac{1+4p+p^2}{(1-p)^2} - \frac{1+2p+p^2}{(1-p)^2}$$
$$= \frac{2p}{(1-p)^2} = \frac{2p}{q^2}.$$

This is the result of theorems. □

Example 2.5.23. *Prove the expectation of the abscissa $\mu_1(x) = -1 + \frac{2p}{q}$, not in the way of a theorem 2.5.20, but with the derivative.*

Proof. Expectations is calculated in order:

$$\mu_1(x) = \langle x\rangle = -1\cdot q^2 + 0\cdot 2q^2p + 1\cdot 3q^2p^2 + \dots =$$
$$= q^2\frac{\partial}{\partial p}(-1\cdot p + 0\cdot p^2 + 1\cdot p^3 + 2\cdot p^4 + \dots)$$
$$= q^2\frac{\partial}{\partial p}\left[-p + p^3\left(1 + 2p + \dots + xp^{x-1} + \dots\right)\right]$$

$$= q^2 \frac{\partial}{\partial p}\left[-p + p^3 \frac{\partial}{\partial p}(p + p^2 + \dots + p^x + \dots)\right]$$

$$= q^2 \frac{\partial}{\partial p}\left(-p + p^3 \frac{\partial}{\partial p}\frac{p}{1-p}\right)$$

$$= q^2 \frac{\partial}{\partial p}\left[-p + \frac{p^3}{(1-p)^2}\right] = q^2 \frac{\partial}{\partial p}\left[\frac{-p+2p^2}{(1-p)^2}\right]$$

$$= q^2 \frac{-1+3p}{(1-p)^3} = \frac{1}{q}[(-1+p)+2p] = -1 + \frac{2p}{q}.$$

This is the required result. □

As usual, the Shannon information S_1 is less than the physical L_1 and we are only interested in this the second. The following theorem speaks of it.

Theorem 2.5.24. *The physical distribution information (2.92) is*

$$L_1 = -\frac{2}{q}(p \ln p + q \ln q),$$

where is also $q = 1 - p$.

Proof.

$$L_1 = -\sum_{k=1}^{\infty} kq^2p^{k-1} \ln q^2p^{k-1} =$$

$$= -2q^2(\ln q)\sum_{k=1}^{\infty} kp^{k-1} - q^2(\ln p)\sum_{k=1}^{\infty} k(k-1)p^{k-1} =$$

$$= -2q^2(\ln q)\frac{\partial}{\partial p}\sum_{k=1}^{\infty} p^k - q^2(\ln p)p\frac{\partial}{\partial p}\left(\frac{\partial}{\partial p}\sum_{k=1}^{\infty} p^k\right)$$

$$= -2q^2(\ln q)\frac{\partial}{\partial p}\frac{p}{1-p} - q^2(\ln p)p\frac{\partial}{\partial p}\left(\frac{\partial}{\partial p}\frac{p}{1-p}\right)$$

$$= -2q^2(\ln q)\frac{1}{(1-p)^2} - q^2(\ln p)p\frac{\partial}{\partial p}\frac{1}{(1-p)^2}$$

$$= -2\ln q - q^2(\ln p)\frac{2p}{(1-p)^3}$$

$$= \frac{2}{q}(-p\ln p - q\ln q),$$

and that's what it was supposed to prove. □

Example 2.5.25. *Prove the theorem 2.5.24 elementarily, without a derivative.*

Proof.

$$L_1 = -2q^2(\ln q)\sum_{k=1}^{\infty} kp^{k-1} - q^2(\ln p)\sum_{k=1}^{\infty} k(k-1)p^{k-1} =$$

$$= -\frac{2q^2}{p}(\ln q)\sum_{k=1}^{\infty} kp^k - \frac{q^2}{p}(\ln p)\left(\sum_{k=1}^{\infty} k^2p^k - \sum_{k=1}^{\infty} kp^k\right)$$

$$= -\frac{2q^2}{p}(\ln q)\frac{p}{(1-p)^2} - \frac{q^2}{p}(\ln p)\left[\frac{p(1+p)}{(1-p)^3} - \frac{p}{(1-p)^2}\right]$$

$$= -2\ln q - (\ln p)\left(\frac{1+p}{1-p} - 1\right)$$

$$= -2\ln q - \frac{2p}{q}\ln p$$

$$= -2\frac{1}{q}(p\ln p + q\ln q).$$

So $L_1 = 2L_0$, the mark L_0 from the theorem 2.5.19. This is the result of the theorem 2.5.24. □

There is a lot similar proves in my previous work in this way, and there are many mistakes too. It's hard to be sure with a discovery that you do not have to compare with. Subsequently, and this is already now when we know that these claims are true, you can skip other ways to prove these same items.

Distribution of the n-th line behind the front

In general, the wave propagation phase, $n = 0, 1, 2, \dots$ lines behind the frontal, reaches abscise x in the number of steps k with which it has the same parity. Its probability $q\binom{k}{n}p^{k-n}q^n$, for $k = n, n+1, n+2, \dots$, also make us "almost a known" distribution:

$$\binom{n}{n}q^{n+1},\ \binom{n+1}{n}pq^{n+1},\ \binom{n+2}{n}p^2q^{n+1},\ \dots, \tag{2.93}$$

and that this really is the distribution of probability we can understand in the following way.

When we throw a fair cube, with the likelihood of $q = \frac{1}{6}$ falls "sixth", and with the probability $p = \frac{5}{6}$ not "six". Let's imagine that we throw this dice $k+1$ times and we wonder how likely it is that $n+1$ "sixth" was get. First of all, k and n are integers and $k \ge n \ge 0$. Secondly, in the last, $k+1$-th throw, the "sixth" fell, whereas the previous k fell them exactly n. We can distribute these previous sixths to $\binom{k}{n}$ modes, each with the probability $p^{k-n}q^n$, and we subjoin to this product another factor, q, for the last "sixth". These are again probabilities (2.93), but now we can easily understand that they belong to different events, one of which must be exactly and can be realized.

That probabilities (2.93) represent the distribution of probability (their sum is one) we can understand directly in this interpretation. The point T randomly moves forward and backwards in abscissa, and sooner or later comes to every position within the final, frontal, which we call here as zero lines. A set of these positions makes n the lines behind the front, whose any two different positions are different events that must occur.

Because (2.93) is the distribution, we have:

$$\sum_{k=n}^{\infty}\binom{k}{n}p^{k-n}q^{n+1} = 1,$$

and hence the sum

$$\sum_{k=n}^{\infty}\binom{k}{n}p^k = p^n q^{-n-1}, \tag{2.94}$$

where $q = 1-p$, and arbitrary $p \in (0,1)$ and $n \in \mathbb{N}$. We use this result in the following proves.

Theorem 2.5.26. *The expectation and variance of step k for the given n of the distribution (2.93) are:*

$$\mu_n = n + (n+1)\frac{p}{q}, \quad \sigma_n^2 = (n+1)\frac{p}{q^2},$$

where $p + q = 1$.

Proof. We calculate the mathematical expectation of the step:

$$\mu_n = \langle k \rangle = \sum_{k=n}^{\infty} k\binom{k}{n} p^{k-n} q^{n+1} =$$

$$= p^{1-n} q^{n+1} \sum_{k=n}^{\infty} \binom{k}{n} k p^{k-1} = p^{1-n} q^{n+1} \frac{\partial}{\partial p} \sum_{k=n}^{\infty} \binom{k}{n} p^k$$

$$= p^{1-n} q^{n+1} \frac{\partial}{\partial p} [p^n (1-p)^{-n-1}]$$

$$= p^{1-n} q^{n+1} [n p^{n-1} (1-p)^{-n-1} + p^n (-n-1)(1-p)^{-n-2}(-1)] =$$

$$= n + (n+1)\frac{p}{q}.$$

The variance of the step is:

$$\sigma_n^2 = \langle (k - \mu_n)^2 \rangle = \langle k^2 \rangle - \langle k \rangle^2 =$$

$$= \sum_{k=n}^{\infty} k^2 \binom{k}{n} p^{k-n} q^{n+1} = p^{-n} q^{n+1} \sum_{k=n}^{\infty} \binom{k}{n} k^2 p^k - \mu_n^2$$

$$= p^{1-n} q^{n+1} \frac{\partial}{\partial p} \sum_{k=n}^{\infty} \binom{k}{n} k p^k - \mu_n^2$$

$$= p^{1-n} q^{n+1} \frac{\partial}{\partial p} \left[p \frac{\partial}{\partial p} \sum_{k=n}^{\infty} \binom{k}{n} p^k \right] - \mu_n^2$$

$$= p^{1-n} q^{n+1} \frac{\partial}{\partial p} \left[p \frac{\partial}{\partial p} (p^n q^{-n-1}) \right] - \mu_n^2$$

$$= p^{1-n} q^{n+1} \frac{\partial}{\partial p} [n p^n q^{-n-1} + (n+1) p^{n+1} q^{-n-2}] - \mu_n^2, \quad q = 1 - p,$$

$$= p^{1-n} q^{n+1} [n^2 p^{n-1} q^{-n-1} + n(n+1) p^n q^{-n-2} + (n+1)^2 p^n q^{-n-2} + (n+1)(n+2) p^{n+1} q^{-n-3}] - \mu_n^2$$

$$= [n^2 + n(n+1) p q^{-1} + (n+1)^2 p q^{-1} + (n+1)(n+2) p^2 q^{-2}] - \mu_n^2$$

$$= (n+1)[(n+1) - n] p q^{-1} + (n+1)[(n+2) - (n+1)] p^2 q^{-2}$$

$$= (n+1) p q^{-1} (1 + p q^{-1})$$

$$= (n+1)\frac{p}{q^2}.$$

These two had to be proven. □

It is easy to verify that the previous expectations and variances of the step are special cases of this theorem, for $n = 0$ and $n = 1$. In addition, the analogy of the random walk of the point T should be noted with the expansion of the actual wave of the substance. Because we work with *physical information*, it deserves special attention.

For example, the physical information of the wave front L_0, probability (2.90), is constant regardless of the number of previous steps, and then the information of each subsequent inner wave line is equal to that constant multiplied by orderly n-th, the line number increased by one, $L_n = (n+1)L_0$. We will prove this last in the next theorem, but before that we can comment on one.

How is it possible to increase the number of random steps and the information L_n remains the same? Throwing the dice repeatedly requires more action and develops more uncertainty, it multiplies Hartley's information, and the question raised is not as meaningless as it seems at first glance; especially now when we work with physical information that is subject to the law of conservation and which we cannot just "push under the carpet". One solution I proposed earlier is pretty "obvious". Like a particle of physics, a material point in the random movement "communicates" with an empty space. It returns to uncertainty information that from uncertainty is realized again by every new random step!

One of the possibilities of these "communication" with empty space is communication with (in physics known) the virtual vacuum particles, and the other is (in physics unknown) communication with the past, about which I wrote in the book Quantum Mechanics [5], and here already mentioned in the section "1.10 Uniqueness" generalizing the *Mach's principle.*

Theorem 2.5.27. *The physical information of the distribution (2.93) is*

$$L_n = -\frac{n+1}{q}(p\ln p + q\ln q),$$

where $p + q = 1$.

Proof. We calculate Hartley's information probability $p^{k-n}q^{n+1}$, ignoring the repetition, and then we find their mean value in relation to the distribution (2.93). We get the physical information:

$$L_2 = -\sum_{k=n}^{\infty}\binom{k}{n}p^{k-n}q^{n+1}\ln p^{k-n}q^{n+1} =$$

$$= -\sum_{k=n}^{\infty}\binom{k}{n}p^{k-n}q^{n+1}[(k-n)\ln p + (n+1)\ln q] =$$

$$= -q^{n+1}(\ln p)\sum_{k=n}^{\infty}\binom{k}{n}(k-n)p^{k-n} - (n+1)p^{-n}q^{n+1}(\ln q)\sum_{k=n}^{\infty}\binom{k}{n}p^k$$

$$= -q^{n+1}(\ln p)p\frac{\partial}{\partial p}\sum_{k=n}^{\infty}\binom{k}{n}p^{k-n} - (n+1)p^{-n}q^{n+1}(\ln q)p^n q^{-n-1}$$

$$= -pq^{n+1}(\ln p)\frac{\partial}{\partial p}[p^{-n}\cdot p^n q^{-n-1}] - (n+1)\ln q$$

$$= -pq^{n+1}(\ln p)[(n+1)q^{-n-2}] - (n+1)\ln q$$

$$= -\frac{n+1}{q}(p\ln p + q\ln q),$$

and that's what it was supposed to prove. □

On the one hand, it is expected due to information conservation law that the information will grow with the depth of the waves, but let's say it is strange. It is strange and why it was possible to reduce these formulas to the mentioned law of conservation at all, and then it is even more interesting to have a wider picture, that the physical information is so accumulated when we know it is a time-limited phenomenon. These paradoxes can be resolved by extending the Mach's principle[11] to the accumulation of information into the past, and then on the influence of the pseudo-real past on the real present, which, after the Quantum Mechanics I written in the popular Multiplicities (see [3]).

Stacking Past

Information is even bigger as we expect it less. On the other hand, the information can not disappear into nothing or come from nothing, but can only be transformed from one form (of data) to another, always equal to the same total quantity. This absurd disappearance of information after its creation is partly explained by the communication of the substance with space, the oscillation of uncertainty-certainty in the motion of a particle, and the proclamation of uncertainty by the type of information.

But what if it's not enough? What if the duration of a random walk generates too much information so that it can only be consumed by oscillation? The particle in the next position accumulates the same information as the one from the previous state; the new one consumes the same significant amount of uncertainty as an old one, but also leaves its mark in history. Unlike *energy*, with which we do not have to worry about the "energy of energy" from the past, now we have to consider the "information of information" as something. This surplus, the *past*, cannot be ignored too.

With the constant emergence of the present, that is, by creating space, time and matter, huge amounts of information are constantly coming from a largely uncertain future and going mainly into the pseudo-real past. I draw this out of my previous texts, and it seems that such a view of reality is now functioning.

I underline that, information can be determined using the physical *action* (energy and duration products) and the changes it produces, then there are sources of knowledge that can change us, but it is not possible the reverse, that we can change say the math theorems. Information is an act, interaction, truth. Similarly, knowledge is a special kind of information that comes from pseudo-reality, such as Pythagoras theorem, or dogma, or the past and parallel reality, from which we get one-way incentives, unlike the realities that interact with us.

These are the preoccupations in which we can deepen ourselves, and maybe give the importance to the famous thesis that the past partly creates our present and future. The past is sediment of *presents* and, if it could fully determine our future, then we would have challenged the assumption of the uncertainty from which the information comes out. Also, if the current information could completely determine all its consequences, we would have a slightly different mismatch now with the assumption that the one generated information is further transformed from form to form without changing the amount, which would eventually be deducted.

Hence the conclusion that it is not possible to accurately predict *future*, it is not possible to accurately understand the present, nor it is possible to accurately perceive the past, perhaps even to the extent that they do not exist as something unique. All this is not a

[11] Existence of absolute rotation comes from a wider presence of matter.

novelty for physics, and therefore supports this theory. Moreover, it's not a novelty the claim that we are determined by what we have been. The novelty would be, for example, the *principle of diversity* that we could base on this, mentioned in the book Multiplicities. It would be the principle that in the nature of matter it is not the creation of long real lines, identical objects, completely equal destinies. It agrees with the thesis that the past defines us, and then with the new (hypo)thesis that every particle of the universe is unique because it has a unique past. You can also call it the *principle of uniqueness*.

I hope that it is not difficult for these elements of my previous considerations to be recognized in the random steps of the material point and the accumulation of the occurring random movements in the spread of its possible positions, and that it is not difficult the vice versa. We recognize the ideas that arise by considering random steps in a well-known experiment of *double-slit* quantum mechanics. In my book of Quantum Mechanics (see fig. 1.26), I explained why all the corresponding waves that were ever in place, from the creation of the universe to the present, interfere now.

The fact that from all of these past events we see only one particle, as a wave on the surface of the sea, formally results from the exclusion of dependent events and only such ones (theorem 1.4.47 of the same book), and then from the independence of the intervening events. This latter means that the word "interference" is unfortunately chosen in physics and that real interaction does not arise from that phenomenon, but by collision, by refusal, by turning in a field caused by another particle. Independent particles do not collide, do not attract, and do not refuse, like, for example, charged, but interfering like photons.

Epilogue

And that would be it for now. Parts of this book mostly are not from my previous books, yet again, all is as it were there. Consider them as novelty and good prediction, or the incompleteness of those and the excesses of these, but I published the articles simply because such works are written in small circulations and for rare readers who encounter.

Latter, when some notice that the information is everywhere around us, and by understanding its settings, legitimacy and scope, its applications will be the basis for the interpretation of society, I believe, such as Darwin's theory, chemistry, and physics itself are the basics of biology, this story will also need a third chapter. If our species is sustained, and the distant future generations would be confused that we dare to consider liberalism without understanding the information and similar with other sociological theories.

Widely looking at the phenomenon of information, it is clear that it is just scratched over its surface here. There is no such development, here the theory of physical information, which I once wrote in the book "Mathematical theory of information and communication" (see [11]), about the channel capacity, Markov chain, ergodic source, crypto codes, matrices, and it should be. There is no quantum mechanics, too, because the presence of information in the material world is so dominant that a book on most of its aspects is a mission impossible. It is, in fact, strange that today it seems too exposed in the book what will latter look insufficient.

Rastko Vuković, April 7, 2019, math prof. photographed for the panel with his students, graduates of the IV-4 general course of the Gimnazija Banja Luka.

Bibliography

[1] Rastko Vukovic: *MINIMALISM of INFORMATION* – physical information and applications; Economic Institute Banja Luka, 2020 (on hold); (https://www.scribd.com/document/421999486/Minimalism-of-Information)

[2] Растко Вуковић: *БИНОМНА РАСПОДЕЛА* – Физичка информација 4; 27. фебруар 2019. (https://www.academia.edu/38450558/Binomna_raspodela)

[3] Растко Вуковић: *MULTIPLICITIES* – Choices for animate and inanimate; Economic Institute Banja Luka, 2018. (https://archive.org/details/Multiplicity_Rastko)

[4] Cristian Farías, Victor A. Pinto & Pablo S. Moya: *What is the temperature of a moving body?*; Scientific Reports volume 7, Article number: 17657 (2017); (https://www.nature.com/articles/s41598-017-17526-4)

[5] Растко Вуковић: *QUANTUM MECHANICS* – the representation of continuum – of metrical and vector spaces; the book is finished on April 26, 2018. (https://www.scribd.com/document/377641503/Quantum-Mechanics)

[6] Растко Вуковић: *SPACE-TIME* – principle physics of chances; Economic Institute, May, 2017. (https://archive.org/details/Principles_Vukovic)

[7] Albert-László Barabási: *Network Science*, Cambridge University Press; 1 edition (August 5, 2016); ISBN-10: 1107076269

[8] Robert M. Gray: *Entropy and Information Theory*; First Edition, Corrected; Information Systems Laboratory; Electrical Engineering Department Stanford University; Springer-Verlag New York, 2000-2013. (https://ee.stanford.edu/~gray/it.pdf)

[9] Растко Вуковић: *INFORMATION of PERCEPTION* – freedom, democracy and physics; Economic Institute Banja Luka,June 2016. (https://www.scribd.com/document/360939276/Information-of-Perception)

[10] Albert R. Meyer: *Expectation & Variance*, Course Notes, Week 13, Massachusetts Institute of Technology, Mathematics for Computer Science, 2006. (https://www.cs.princeton.edu/courses/archive/fall06/cos341/handouts/variance-notes.pdf)

[11] Растко Вуковић: *МАТЕМАТИЧКА ТЕОРИЈА ИНФОРМАЦИЈЕ И КОМУНИКАЦИЈЕ*, Друштво математичара Републике Српске, Бања Лука 1995.

[12] Zoran A. Ivković: *Teorija verovatnoća sa matematičkom statistikom*, Univerzitet u Beogradu; Izdavačko preduzeće Građevinska knjiga, Beograd, 1976.

[13] AFTAB, CHEUNG, KIM, THAKKAR, YEDDANAPUDI: *Information Theory*, INFORMATION THEORY AND THE DIGITAL AGE; 6.933 – FINAL PAPER. (http://web.mit.edu/6.933/www/Fall2001/Shannon2.pdf)

[14] Claude E. Shannon: *A Mathematical Theory of Communication*, Reprinted with corrections from The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948. (http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf)

Index