

# ON THE CONCEPTS MEANING, THEIR MEASUREMENTS AND USE IN PRACTICE OF RISK ASSESSMENT

Boyan Dimitrov

•  
Gnedenko e-Forum current president  
Kettering University Professor Emeritus,  
[dimitrob@gmail.com](mailto:dimitrob@gmail.com)

## Abstract

*In this talk we discuss the basic concepts used in axioms of sciences, their meaning explained through the axioms and measurement that follow afterwards. There are situations where meaning comes first, then axioms then measure. And there are situations where measurements come first and meanings – later. Things may look elementary, but they are supposed to build a solid background when make a system to analyze the risk. We illustrate this approach on examples from classic sciences as geometry and algebra, from examples of recent times as for uncertainty and probability and focus our attention on the concept of risk ant its measurements.*

**Keywords:** Axioms, concepts, dependence, meaning, measurement, probability, risk, uncertainty.

## I. Introduction

In this article I present my personal understanding about the terms listed in the above key words. These words are used in the everyday conversations between people and researchers, media, and publications. But my focus is on the scientific world. Basic concepts in sciences are usually names of some imaginary objects whose meaning can be understood after they are related to some explanations through appropriate axioms. For example, let look at the basic concepts in geometry: point, straight line, circle, angle; or at the algebraic concepts like number, zero, infinity; ... When a theory is built on the base of axioms, then measurements could be introduced and many other consequences related to measurements, involving more than two areas of science. For instance, in geometry we can talk about lengths, angle's measurements, comparison, areas, volumes, in algebra – about real and complex numbers, etc.

In this talk I will not go to details. An extended version can be presented to anyone, who would like to have it.

Without axioms and without basic rules that relate their basic elements it is impossible to use their consequences. And since our focus is on this work on uncertainty and risk, we continue with these basic concepts and related measurements to the uncertainty.

## II. Uncertainty and probability

Risk is related to uncertainty. Here we explain how and why we need to understand and model the uncertainty to correctly introduce its measures, namely the probability associated with the risk. Examples we give for uncertainty models and ways to introduce probability are very important. They determine the probability space where we work on the risk. Otherwise, probability spaces needed to assess the risk cannot be understood.

Both concepts, uncertainty and probability are basic concepts. No definitions can be given. But axioms, examples, additional basic concepts and rules, together with serious practice will help for easy work and use the Probability Theory.

## 1.1. Introduction to the Uncertainty and its meaning

**The Set theory** lies in the foundation of any axiomatic setups. It uses a *basic concept of a set*. There is no definition of set (as any other basic concept in any theory). A set can be informally described as collections of objects possessing some common features. Although objects of any kind can be collected into a set, here we see replacement of a name “set” by another name “collection”. Many times, we observe explanations of a basic concept with the help (in terms) of others through the axioms, and we will mark such facts in the sequel. Here we refer to the *Set Theory* just because it introduces some symbolic notations and meaning of operations with sets. These are needed to understand most mathematical fields.

Set theory begins with a fundamental relation between an object  $\omega$  and a set  $A$ . If  $\omega$  belongs to the set  $A$  (or  $\omega$  is an *element* of  $A$ ), the notation  $\omega \in A$  is used. Objects can be undefined concepts which could be explained through the axioms (see e.g., Euclidian axioms). A set is described by listing its elements when possible, or by a characterizing property of its elements, or just by description in words. Sets are also objects; the membership relation explain sets as well. A derived binary relation between sets is the subset relation, also called *set inclusion*. If all the elements of set  $A$  are also elements of set  $B$ , then  $A$  is subset of  $B$ , denoted  $A \subseteq B$ . As implied by this definition, a set is a subset of itself. And if  $A \subseteq B$ . and  $B \subseteq A$ , then the two sets are identical. Relationships and operations between sets are also important and used in many models based on the set theory. No axioms are known in this relation, but that the sets do exist and satisfy some relationships, which were imitating operations with numbers. The following is a partial list of these:

- The union of the sets  $A$  and  $B$ , denoted as  $A \cup B$ , is the set of all elements that are members of  $A$ , of  $B$ , or of both.
- The intersection of the sets  $A$  and  $B$ , denoted as  $A \cap B$ ,<sup>[2]</sup> is the set of all elements that are members of both  $A$  and  $B$ .
- Cartesian product of  $A$  and  $B$ , denoted as  $A \times B$ , is the set whose members are all ordered pairs  $(a, b)$ , where  $a$  is a member of  $A$  and  $b$  is a member of  $B$ .
- The empty set is the unique set containing no elements. For it the usual symbolic notation is  $\emptyset$ .
- The **power set** (or powerset) of a **Set**  $A$  is the **set** (actually, *the number*) of all subsets of  $A$  including the **set**  $A$  itself and the empty **set**.

These operations with the sets are extended to unions and intersections between more than 2 sets.

Examples of sets helped to create theory of numbers, real numbers, continuity, the Boolean algebra, Discrete Mathematics, and many other theories used in computer sciences.

To start with **Uncertainty**, we need to introduce some more basic concepts for which there are no definitions, like the concepts of points, lines, numbers referred above.

First is in Uncertainty the concept of *Experiment*. This word has many determinations, and no description According to Probability Theory *an experiment is just a collection of conditions that produce results*. This is not a definition, just a description. We denote this concept by the symbol  $E$ .

Another basic concept is the *outcome*  $\omega$ , a *simple result* from  $E$ . It is called a *simple event*.

Then we can define the set of all possible simple events under the  $E$ , and denote it with  $S$ . This set  $S$  contains no more, nor less but every outcome  $\omega$ , *obtained as a simple result* from  $E$ . And we call  $S$  is a *sure event*.

Every subset  $A$  of  $S$  is called *random event*. A *random event*  $A$  occurs only when  $\omega \in A$ .

The empty set  $\emptyset$  is called *impossible event*.

The specific situation allows to introduce some additional extra random events (subsets of  $S$ ) – the *complement*  $\bar{A}$  to a random event  $A$ . This is *the event that happens only and always when  $A$  does not happen*.

Also, the operations with events become specific interpretations:

- The *union* of the events  $A$  and  $B$ , denoted  $A \cup B$  is the set of all elementary outcomes  $\omega$  that favor  $A$ ,  $B$ , and that are members of  $A$ , or  $B$ , or both. It is read “*at least one of the two events occurs*”. Meanings matter.
- The *intersection* of the events  $A$  and  $B$ , denoted  $A \cap B$  is the set of all elementary outcomes  $\omega$  that favor both  $A$  and  $B$ . It is read as “*both events occur*” in  $E$ .
- If  $A \cap B = \emptyset$ , then the events  $A$  and  $B$  are called *mutually exclusive*, and their union will be symbolized as  $A + B$  (meaning it is *union of mutually exclusive events*)

These meanings and operations with symbols will be used in full scale in the next. For instance,

$$\bigcup_{k=1}^n A_k = A_1 \cup A_2 \cup \dots \cup A_n$$

means “at least one event in the list will occur in  $E$ ”. Also

$$\bigcap_{k=1}^n A_k = A_1 \cap A_2 \cap \dots \cap A_n$$

means that each one event in the list will occur in  $E$ .

This is how the operations with events stated in the set theory, become important meaning. It is used through the entire axiomatic of the Uncertainty where sets become meaning. For simplicity, we introduced specific symbols in operations with sets that do not change the meaning of unions of events and include knowledge about their mutual intersections. Operations between sets (events) and their meaning should be known to everyone familiar with the basics of Probability theory.

## 1.2. Axioms of Uncertainty and Examples

Uncertainty is related only with certain experiment, frequently understood in a broad and imaginary sense. For instance, a human being  $\omega$  can be considered random outcome because of many circumstances (conditions) at the place he/she is born, country, parents, growth, friends, ethnicity, education, politics during growth, etc. Experiment is not well defined but results  $\omega$  are clear and may serve as a set of outcomes undoubtedly.

The axioms of uncertainty should be now clear: This is a system  $F$  of random events  $\omega$  which satisfies the following rules (axioms):

1. The sure event  $S$  and the impossible event  $\emptyset$  belong to  $F$ ;
2. If the random events  $A, B, \dots, C$  belong to  $F$ , then their unions  $A \cup B, A \cup C, \dots$ , intersections  $A \cap B, A \cap C, \dots$ , and complements  $\bar{A}, \bar{B}, \dots, \bar{C}$  also belong to  $F$ .

The following models will illustrate the flexibility of this presentation of the uncertainty. Each model corresponds to the resolution of what one can recognize as a random event in an experiment. In definition of the risk corresponding explicit model is of high importance.

3.  $F_0 = \{S, \emptyset\}$ . It corresponds to the situation where one sees only the sure and impossible events. This is the *model of certainty*.
4.  $F_A = \{S, \emptyset, A, \bar{A}\}$ . This model corresponds to a situation, where one sees only when a random event  $A$  occurs or not (then occurs the complement  $\bar{A}$ ).
5. The sure event  $S$  is partitioned in a particular cases (mutually exclusive events  $A_1, A_2, \dots, A_n$ ) so that  $S = A_1 + A_2 + \dots + A_n$ . This means that in the experiment only some finite, or countable number of mutually exclusive random events can be recognized. The uncertainty system  $F$  consists of any union of events from these in the above partition. As an example, in roulette game there are 37 positions, but colors, odds, neighbors make combination of this partition. In natural numbers (is considered as a particular case, odds, evens, multiple to an integer number  $k$ , are combined events. Here are many models we do not consider in detail, and they represent a better resolution in the set of outcomes of the experiments.

6.  $F_\infty$  - the uncertainty consists of any particular outcome  $\omega$  in the experiment and contains every subset of the sure event  $S$ . This is the maximal resolution one can see in the outcomes of an experiment.

### 1.3. Measures of Uncertainty. Axioms of Probability

The uncertainty as introduced is just to register existence and what is the resolution of events in the experiments. To measure and study its properties comes the mathematics with all its power. And mathematics comes with another basic concept still not defined. This is the concept of probability, also introduced with axioms [1]. There is no definition of probability, again, it has just an explanation:

*Probability is a measure of the chance that a random event will occur under the conditions  $E$  of an experiment. For a random event  $A$  probability is denoted as  $P(A)$ .*

*The word measure introduces mathematics in the analysis of uncertainty. And any time when one introduces this measure should have in mind the conditions  $E$  of this experiment and the uncertainty model  $F$  associated with this measure. Uncertainty model  $F$  is necessary whenever probability is introduced. Probability is introduced for what is visible according to the uncertainty model  $F$ .*

Therefore, we need the uncertainty model  $F$  and only within a model the probability measures can be introduced. Then the axioms of probability can be applied. They are:

1. For any event  $A \in F$  the probability  $P(A) \geq 0$ ;
2.  $P(S) = 1$ ;
3. If  $A = A_1 + A_2 + \dots + A_n + \dots$ , then  $P(A) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$

The triplet  $\{S, F, P\}$  is then called *probability space*. The rules that are derived from the axioms work only in this probability space.

These axioms [2] do not explain how the probability must be calculated. But they are sufficient to derive all the mathematics for Probability Theory and all the rules and properties which govern any application of this theory. Axioms do not assign any numeric value for  $P(A)$ . But to make sure that any method used to assign such numeric values for which the relationships, like these

$$0 \leq P(A) \leq 1; \quad P(\emptyset) = 0; \quad P(A) = 1 - P(\bar{A}); \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

are valid. An important interpretation: The probabilities  $P(A)$  and  $P(B)$  are called *marginal probabilities*, while  $P(A \cap B)$  is called *joint probability*. This meaning is essential later when we introduce measures for dependence and the strength of dependence.

### 1.4. Methods for establishing numeric values of probability $P(A)$

Axioms do not provide ways to assign any numeric values of  $P(A)$  in any probability space  $\{S, F, P\}$ . However, practical needs and theoretical results of probability theory created the following approaches: Classic Definition, Statistical approach, and Subjective approach. Their use depends on certain assumptions (0mpoortant inn risk assessments), and we explain it shortly.

#### Classical probability definition

It is used when the following assumptions are legal to be applied:

The uncertainty model of an experiment shows that the sure event  $S$  is equivalent to the occurrence of *finite number of mutually exclusive and equally likely* outcomes  $\omega_1, \omega_2, \dots, \omega_n$ , i.e.  $S = \omega_1 + \omega_2 + \dots + \omega_n$  as in Model 3 of the uncertainty  $F$ . Every random event is equivalent to the occurrence of some  $k$  of these outcomes, i.e.  $A = \omega_{i_1} + \omega_{i_2} + \dots + \omega_{i_k} \in F$ . Then  $P(A)$  is determined as the ratio of the numbers  $k$  of the outcomes that favor  $A$  and the total number  $n$  of all possible outcomes i.e.

$$P(A) = k/n = \#(A)/\#(S).$$

Here I use the sign “#” as a symbol for the counts of the distinct elements in a set. Then the Classical definition of probability works. However, if  $S$  is a *finite geometric space* (line segment, plane region or a multidimensional region *with finite volume*), and the outcomes of the experiment are equally likely to appear as random points in  $S$ , then thinking  $\#(S)$  and  $\#(A)$  as respective *geometric measures* (length, area, volume), the same Classic definition applies, and the probability -- is then called *Geometric Probability*. And all rules of combinatorics (counting techniques) and the geometry measures can be used in numeric calculations of the measures of the chances random events to occur in a single experiment.

### Statistical determination of probability

This approach is the most real one. It can be used when *the experiment E can be repeated* without changed conditions many times. When repeated  $N$  times and a random event  $A$  is observed  $N(A)$  times its relative frequency  $P(A)$  is defined as the ratio of number of times  $A$  occurs and the total number  $N$  of experiments performed.

$$f_N(A) = N(A)/N$$

The probability  $P(A)$  is then defined as the limit of  $f_N(A)$  when  $N$  tends to infinity, i.e.

$$P(A) = \lim_{N \rightarrow \infty} f_N(A) = \lim_{N \rightarrow \infty} N(A)/N,$$

The legality of this approach is proven based on axioms of Probability, known as the *law of large numbers*. It has been used many times in practice to discover frauds in hazard games. And it is widely used in simulations.

### Subjective determination of probability

This approach usually is applied when any experiment cannot be used. It is subjected future events with known or guessed conditions. The experts are used to give their forecasts. And the user may use then the statistical approach to manipulate the expert's opinion. In absent of expert, just an opinion can be used. But any consequences should be delivered only when the axioms of the probability theory are verified and fulfilled.

We leave this discussion and assume, basic rules and meanings of probability (conditional probability, multiplication rules, total probability, and Bayes rule) are known to the audience, as well as definitions of random variables, their numeric summaries (mean values, variance, conditional expectations, regression equations, etc.) are familiar to the readers. Ways of calculations of probability and its determination are very, very important in measuring the risks. This is our reason to focus your attention on this necessary initial step.

In the next part of the talk, we will focus your attention on some issues, which might be less familiar and discussed mostly in scientific publications.

## 2. Important and less known approaches and results

### 2.1. Independence and dependence strength measured

Let stop for a minute on the concept of independence, its definition and on the ways of measuring dependence, and their use. There are many trials to define this very complicated concept (Declaration of Independence in USA, memorandum of independence from various countries in the World, independent people, independent students, etc.). But the only definition of independency in this world of uncertainty, that can be verified is given through the Probability theory. For two random events  $A$  and  $B$  independence is defined as satisfying the mathematical equation:

$$P(A \cap B) = P(A)P(B).$$

It means that the joint probability of the two events under the same experimental condition is equal to the product of their marginal probabilities. Independence is symmetric. It is **mutual**. Moreover, any pair  $A$  and  $\bar{B}$  (the complement of  $B$ ),  $\bar{A}$  and  $B$ ,  $\bar{A}$  and  $\bar{B}$  are mutually independent too. And all textbooks stop discussions here. But it can be continued. In a series of publications [5-7] I used an idea of Bulgarian mathematician N. Obreshkov [8] to introduce several measures of discovering dependence and to measure its strength for the case of pairs of random events. Then it was extended to measure the local dependence between pairs of random variables, no matter of their types (numeric or non-numeric, or mixed). I see here a huge area of future studies, since joint distributions between random variables and their marginal distributions are defined as probabilities of random events. What is worked out for relationships between random events will be valid for random variables too.

Let us see the measures of dependence between two random events, as proposed by Obreshkov [8].

*Connection* between random events. The number

$$\delta(A, B) = P(A \cap B) - P(A)P(B)$$

is called *connection between events A and B*.

The full list of **properties of the connection** can be found in [5]: I cite here the most important,

**δ1)** The connection  $\delta(A, B)$  equals to the covariance between the indicators of the two random events  $A$  and  $B$ ;

**δ2)** The connection  $\delta(A, B)$  equals to zero if and only if the events are independent (something not valid in case of r.v.'s);

**Δ3)** The probability for occurrence of one of the two events can be recalculated as the conditional probability after the occurrence of the other event when their connection is known. The following relation holds:

$$P(A | B) = P(A) + \frac{\delta(A, B)}{P(B)}.$$

*This equation* indicates that the knowledge of the connection is very important and *can be used for calculation of the posteriori probabilities similar to when we apply the Bayes' rule!*

Connection function just shows *direction of dependence* – positive or negative. But it is not good for measuring the strength of dependence. Next measures serve better for measuring the strength of dependence.

### Regression coefficients as measure of dependence strengths

Regression coefficient  $r_B(A)$  of the event  $A$  with respect to the event  $B$  is called the difference between the conditional probability for the event  $A$  given the event  $B$ , and the conditional probability for the event  $A$  given the complementary event  $\bar{B}$ , namely

$$r_B(A) = P(A | B) - P(A | \bar{B}).$$

This *measure of the dependence of the event  $A$  on the event  $B$ , is directed dependence.*

The regression coefficient  $r_A(B)$  of the event  $B$  with respect to the event  $A$ , is defined analogously:

$$r_A(B) = P(B | A) - P(B | \bar{A}).$$

**The whole list of statements about regression coefficients can be found in [5]. We present the most important here.**

(r1) The equality to zero  $r_B(A) = r_A(B) = 0$  holds if and only if the two events are independent.

(r2) The regression coefficients  $r_B(A)$  and  $r_A(B)$  are numbers with equal signs and this is the sign of their connection  $\delta(A, B)$ . The relationship

$$\delta(A, B) = r_B(A) P(B)[1 - P(B)] = r_A(B) P(A)[1 - P(A)]$$

holds.

The numerical values of  $r_B(A)$  and  $r_A(B)$  may not always be equal. There **exists an asymmetry in the dependence between random events, and this reflects the nature of real life.**

To be valid  $r_B(A) = r_A(B)$  it is necessary and sufficient to be fulfilled

$$P(A)[1 - P(A)] = P(B)[1 - P(B)].$$

(r3) The probability for occurrence of one of the two events can be recalculated as the conditional probability after the occurrence of the other event when their regression coefficients are known. The following relation holds:

$$P(A | B) = P(A) + r_B(A)[1 - P(B)],$$

and vice versa.

(r4) The regression coefficients  $r_B(A)$  and  $r_A(B)$  are numbers between  $-1$  and  $1$ , i.e. they satisfy the inequalities

$$-1 \leq r_B(A) \leq 1; \quad -1 \leq r_A(B) \leq 1.$$

(r4.1) The equality  $r_B(A) = 1$  holds only when the random event  $A$  coincides with (or is equivalent to) the event  $B$ . Then it is also valid the equality  $r_A(B) = 1$ ;

(r4.2) The equality  $r_B(A) = -1$  holds only when the random event  $A$  coincides with (or is equivalent to) the event  $\bar{B}$  - the complement of the event  $B$ . Then it is also valid  $r_A(B) = -1$ , and respectively  $\bar{A} = B$ .

**We interpret the properties (r4) of the regression coefficients in the following way:** As closer is the numerical value of  $r_B(A)$  to 1, "as denser inside within each other are the events  $A$  and  $B$ , considered as sets of outcomes of the experiment". In a similar way we also interpret the negative values of the regression coefficient: "As closer is the numerical value of  $r_B(A)$  to -1, as denser within each other are the events  $A$  and  $\bar{B}$  considered as sets of outcomes of the experiment".

The regression function possesses the property

$$r_B(A \cup C) = r_B(A) + r_B(C) - r_B(A \cap C).$$

These properties are anticipated to be used in simulation of dependent random events with desired values of the regression coefficients, and with given marginal probabilities  $P(A)$  and  $P(B)$ . Some restrictions must be satisfied when model dependent events by use of regression coefficients.

The asymmetry in this form of dependence of one event on the other can be explained by the different capacity of the events. Events with less capacity (fewer amounts of favorable outcomes will have less influence on events with larger capacity. Therefore, when  $r_B(A)$  is less than  $r_A(B)$ , the event  $A$  is weaker in its influence on  $B$ . We accept it as reflecting what indeed exists in the real life. By catching the asymmetry with the proposed measures, we are convinced about their flexibility and utility features.

**Correlation between two random events  $A$  and  $B$**  we call the number

$$R_{A,B} = \pm \sqrt{r_B(A) \cdot r_A(B)},$$

where sign, plus or minus, is the sign of the either of the two regression coefficients.

An equivalent representation of the correlation coefficient  $R_{A,B}$  in terms of the connection  $\delta(A, B)$  holds,

$$R_{A,B} = \frac{\delta(A, B)}{\sqrt{P(A)P(\bar{A})P(B)P(\bar{B})}} = \frac{P(A \cap B) - P(A)P(B)}{\sqrt{P(A)P(\bar{A})}\sqrt{P(B)P(\bar{B})}}.$$

We do not discuss the properties of the correlation coefficient  $R_{A,B}$  between the events  $A$  and  $B$ . Just notice that it equals to the formal correlation coefficient  $\rho_{I_A, I_B}$  between the random variables  $I_A$  and  $I_B$  - the indicators of the two random events  $A$  and  $B$ . This explains the terminology proposed by Obreshkov, 1963.

The knowledge of  $R_{A,B}$  allows how to calculate the posterior probability of one of the events under the condition that the other one occurred. For instance,  $P(A | B)$  will be determined by the rule

$$P(A | B) = P(A) + R_{A,B} \sqrt{\frac{P(\bar{B})P(A)P(\bar{A})}{P(B)}}.$$

This rule reminds again the Bayes' rule for posterior probabilities. The net increase, or decrease in the posterior probability compared to the prior probability equals to the quantity  $R_{A,B} \sqrt{\frac{P(\bar{B})P(A)P(\bar{A})}{P(B)}}$ , and depends only on the value of the mutual correlation  $R_{A,B}$  (positive or negative).

A new interesting approach in measuring the strength of dependence between random events, based on the use of entropy, was recently introduced to me by another Bulgarian



mathematician Dr. Valentin Iliev.

A preprint of his work can be found at <https://www.preprints.org/manuscript/202106.0100/v3>. No official publication of this work is known to me yet. However, lots of research in this direction is needed. One is about the dependence between two variables in many of the important situations. The other one is about the extension of the dependence measures between more than two variables. Good luck for those who start working on this. Use of these measures can play significant role in the studies of risk analysis!

### 3.2.1. On the axioms of statistics and classes of variables

Statistics also uses several basic concepts, most of which are explained in terms of sets. These concepts are *population, sample, variables, measurement, data set*. Most of the textbooks on Probability and statistics start their introductions to uncertainty with these concepts.

**Population** – this is the set that contains no more, nor less, but all the individuals targeted in a statistical study.

**Sample** – this is the set of  $n$  individuals selected from the population, whose data will be used for information about properties of various characteristics of the population.

**Measurements** – this is a sequence of characteristics  $\vec{x}_i = (x_1, x_2, \dots, x_k)_i$  one can get after recording the data inspecting the selected individuals  $i, i=1,2,\dots,n$ .

**Data set** – or *data matrix* is the matrix of all recorded observations  $\vec{X} = \{ (x_1, x_2, \dots, x_k)_i \}_{i=1,\dots,n}$ .

**Variables** – these are the recorded values of the measured/observed characteristics for individuals in the sample, located in any of the columns of the data set. These can be:

- Non-numerical or nominal, like names, symbols, labels (like gender, color, origin, etc);
- Ordinal, where categories could be ordered (good, better, best), like preference, age groups etc;
- Numeric (relative or absolute), like temperature, weight, scores, income, etc. Here digital records work.

No axioms of statistics are known. It works mainly with meanings and data. All statistical manipulations with the statistical data sets are based on assumed probability models for the data in a column, i.e., axioms and rules of probability theory start work and used. And here we observe a change of the concept of the population, called sample space. This is the set of all the possible values of the variables in a column (as a population) and parameters of the distributions assumably presenting what is going on in the original population, are called population parameters. Usually in statistics the targets are the estimation of these population parameters. Further various hypotheses are built and tested according to the huge set of algorithms developed in statistics. And there we face the concept of risk, related to many, if not to all decisions based on statistical data, or on probability models applied in the real life.

## III. Uncertainty and related risks in applications

My extended searches for definition of the concept of risk faced different opinions. One of it <https://rolandwanner.com/the-difference-between-uncertainty-and-risk/> seems reasonable:

*A risk is the effect of uncertainty on certain objectives. These can be business objectives or project objectives. A more complete definition of risk would therefore be “an uncertainty that if it occurs could affect one or more objectives.*

A risk is the effect of uncertainty on certain **objectives**. These can be business objectives or project objectives. A more complete definition of risk would therefore be “an uncertainty that if it occurs could affect one or more objectives”. Objectives are what matters! I agree with this description. But how to measure the risk? How close are we at certain moment to some risk?

This recognizes the fact that there are other uncertainties that are irrelevant in terms of objectives, and these should be excluded from the risk evaluation. With no objectives, we have no risks. Linking risk with objectives makes it clear that every case in life is risky. Everything we do

aims to achieve objectives of some sort, including personal objectives, project objectives, and business objectives. Wherever objectives are defined, there will be risks to their successful achievement will be completed.

The PMBOK guide <https://www.amazon.com/Project-Management-Knowledge-PMBOK%C2%AE-Sixth/dp/1628251840> defines risk as an *uncertain event* or *set of circumstances*, and if it occurs has a positive or negative effect on achievement of objectives. Both definitions are alike. Well, but how to measure the effects? Moreover, conditions change during the time. Therefore, risk assessment also will change, depending on the current conditions. Risk assessment is a process to be observed and followed!

There are events and circumstances/conditions mentioned in these definitions. That's obviously something separated from event. There are *uncertain future events* and if they occur, they could affect achievement of objectives, and that's been in the focus. So what does it mean? Where events correspond to the event called risk? This distinction may still be strange. Moreover, conditions may change over the time. Therefore, risk assessment is a process that must be controlled over the time.

The above online references built in me the incentives, that *the risk is something objective, and it is accepted subjectively*. Without measures of the risks it does not make any sense to discuss it further. I hope, my examples below related to various understandings of the risk could explain my feelings, and will rise much of questions more to think about. Scientists use the concepts of risk in various situations and this is what I would like to present.

## 2.2. Risks in hypotheses testing and their measures

In testing hypotheses we have the following situation: a *null hypothesis*  $H_0$  is tested versus some *alternative hypothesis*  $H_a$  and some appropriate test, based on certain model assumptions are applied. After a test is performed, the statistician should take a decision, not reject the null hypothesis, or reject it. Since the nature (uncertainty) is not known, such personal decision contains some risk, which has measures, probability to admit error of type 1 (usually denoted by the symbol  $\alpha$ ) meaning to reject the correct null hypothesis. And if the statistician accepts the null hypothesis as being correct, there is another risk for this decision to be incorrect. Its measure is given by the probability  $\beta$  to admit this error of type 2.

Here we observe the risk of taking wrong decisions of any kind based on uncertain data. No more discussions but the two faces of a risk are in front of us. Let us continue with other informative values.

### The p-value

Whenever you run a statistical package with some statistical data set and use some probability model to test, at the output you get a list of numerical results. In it you will find lots of these labeled as p-values, either for the model itself, or **for parameters used in it. Each p-value corresponds to some specific pairs of hypotheses (formally as described above)** and taking decisions one needs to take is based on their meaning. The p-values in my understanding, are hidden measures of the risks in taking decisions. This meaning of the p-value is very important and needs to be understood. The book [9] of A. Vickers is an interesting place to look at. It offers a funny introduction to the fundamental principles of statistics, presenting the essential concepts in thirty-four brief, enjoyable stories. Drawing on his experience as a medical researcher, **Vickers** blends insightful explanations and humor, with minimal math, to help readers to understand and interpret the statistics they read every day. And there is no explanation what the p-value means.

*My own simple definition-explanation (MEANING) of the p-value* should be clear for any user and should be used in decisions:

*The p-value is the measure of the chance for the null hypothesis to survive in the conditions provided by the statistical data used in its evaluation.*

You cannot find the explanation of p-value concept in any textbook or books related to its discussion like in [9]. Lately the ASA (American Statistical Association) also made a huge discussion on this issue <https://www.bing.com/search?form=MOZLBR&pc=MOZD&q=asa+p-value> gives the following definition:

*A p-value is the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.*

Can you find any useable meaning in this definition? I cannot. Other definitions in the textbooks are focused on the rules of its numeric calculations (since it depends on the null and alternative hypotheses formulations, on the tests applied, targeted parameters, and many other things) and do not give any understandable meaning. *But meaning is the most stable something, that does not depend on any method of calculation and anything else. Measures of the meanings are numbers and should be used as measures of the risk taken in decisions (look at the chance that the null hypotheses are correct in this particular discussion, when statistical data is available and used).*

### 2.3. Other visions on the risk and their numeric measures

Risk is a complex concept. It is met in old times mainly in the hazard games. Then (and even in now days) it was measured in terms of "odds".

#### Odds

A true definition of the odds should be the ratio of the probability that a random event  $A$  will happen in the condition of an experiment, and the probability that another event  $B$  will occur. So.

$$\text{Odds of (A versus B)} = P(A)/P(B).$$

This ratio should be read as ratio of two integers, therefore these probabilities are usually multiplied by 10, or 100, or even by thousand (and then fractional parts are removed) for easy understanding on behalf of users not familiar to probability. Best is the example in case of classic probability

$$\text{Odds of (A versus B)} = \#(A)/\#(B).$$

According to the web information <https://www.bing.com/search?form=MOZLBR&pc=MOZD&q=odds> or <https://www.thefreedictionary.com/odds> odds provide preliminary information about bidders in games what are their chances to win after the experiment is performed if the player bet is for event  $A$ . There usually for  $B$  stays the complement of  $A$ . In my opinion, the odds are kind of hints given by experts to the players what is their risk when in games they bid for the occurrence of certain random future event. However, I think this measure is good also when people talk about risks. In my opinion, odds could be used for comparing chances of two events even when they are results in different experiments. However, the next measure also makes sense.

#### Relative Risk when tests are used

In medicine, biology, and in many actuarial, social, and engineering business the usual practices are to apply a test for establishing if an individual in the population possesses some property (let use terminology "category B ") or is not in this category. As a current example let me notice the popular PCR test for establishing if a person has a COVID 19 virus or not.

Therefore, let  $B$  and  $A$  are two events where  $A$  has the sense of test factor (for example, an environment, habitat, or type of training, temperature different from the normal, etc.), that an individual belongs to the category  $B$ . The relative risk (Relative Risk, denote it briefly RR) of event  $B$  with respect to event  $A$  is defined by the rule

$$RR(B \text{ versus } A) = \frac{P(B | A)}{P(B | \bar{A})}. \quad (*)$$

The point is that the larger the RR, the more test (risk factor  $A$  is increasing the probability of occurrence of  $B$ ) effects the category  $B$  to be true. For example, if we want to evaluate the influence of some risk factors (obesity, smoking, etc.) on the incidence of a disease (diabetes, cancer, etc.), we need to look at the value of the relative risk, when test  $A$  is applied and indicates such categorization of the risk to be a fact. It is kind of "odds measure" useful to know. We illustrate such situation with an example from biostatistics below.

### Tests for the Truth

It is well known that in medicine research and in other experiential sciences to discover the presence of certain diseases there are used specific tests. When the result of the test is positive it is considered that the object owns the quality of what it is tested. However, tests are not perfect. They are very likely to give a positive result when tested objects really have that quality, for which they are tested. Also, it happens to get a negative test although the object possesses that property. And there is another possibility, although unlikely, the test gives a positive result, even when the subject does not possess the property in question. These issues are closely related to the conditional probability. The biostatistics has established specific terminology in this regard which should be known and used. It is known that in carrying out various tests among the population that suffers from something it is possible to get a positive test (indicating that the person may be ill), or negative test. In turn, the latter is an indication that the person is not sick. Unfortunately, there are no tests that are 100% truthful. It is possible that the tested is sick, but the test may not show it, as well as being positive, although inspected subject is not sick. Here the concepts of conditional probabilities play a special role and are important in assessing the results of the tests.

### Screening tests for a risk

**Predictive value positive,  $PV^+$**  of a screening test is called the probability of a tested individual to possess tested quality, (such as being sick), when tested positive, ( $T^+$ ) i.e.

$$PV^+ = P(\text{sick} | T^+);$$

**Predictive value negative,  $PV^-$**  of a screening test is called the probability that the tested individual does not have the tested quality (is not sick), on condition that the test was negative ( $T^-$ ) i.e.

$$PV^- = P(\text{healthy} | T^-).$$

**The sensitivity of the test** is determined by the probability that the test will be positive, provided that the individual has the tested quality, i.e.

$$\text{Sensitivity} = P(T^+ | \text{sick}).$$

**The specificity of the test** is determined by the probability that the test gives a negative result, provided that the tested individual does not possess the quality for which is being checked, i.e.

$$\text{Specificity} = P(T^- | \text{not sick}).$$

In words Specificity = P (no symptom detected | no disease).

**False negative** is determined the outcome of the test where an individual, who was tested as a negative, is sick (possessing the tested quality).

To be effective a test prediction the disease should have high sensitivity and high specificity. The relative risk to be sick if tested positive is then the ratio

$$\text{RR}(\text{risk to be sick when tested positive}) = P(\text{sick} | T^+) / P(\text{sick} | T^-) = PV^+ / [1 - PV^-].$$

The use of RR in terms of odds is a very good and useful idea in risk assessment ia tests.

If we replace here the word "sick" by the word "risk" in technical items or systems, we will see how good these methods for estimation fit in a wider area of the life.

## 2.4. Reliability and risk

Possibly for many natural reasons, the **Risk concept** is most related to another complex concept - the Reliability. This is an exciting area of discussions, and in my opinion, not finished yet, probably will never be finished. I studied the web opinions and have had detailed discussions with my Gnedenko Forum colleagues, and still did not come to any determined conclusions. Here are some brief results of my research.

In my modest opinion, first and foremost important thing is to understand what the risk is? Then discuss ways to assess or measure it. According to me, the risk is an objective-subjective feeling that something undesirable, dangerous event may happen under certain conditions. While one can explain in words what it is to everyone else, this is not sufficient without some general frames and numeric measure of that risk.

Let see what the wise sources about the risk concept are talking.

### 4.3.1. Definition of Risk Related to Reliability (from the web)

Creating a reliable product that meets customer expectations is risky.

What is risk and how does one go about managing risk? The recent set of ISO (International Organization for Standards) updates and elevates risk management. Here are given some details:

ISO 9000:2015 includes the definition of risk as "the effect of uncertainty on an expected result. ISO 31000:2009 includes the definition of risk as "the effect of uncertainty on objectives."

The origin of the English word 'risk' traces back to the 17<sup>th</sup> century French and Italian words related to *danger*.

A dictionary definition says "the possibility that something unpleasant or unwelcome will happen."

Risk from a business point of view may need a bit more refinement. The notes in the ISO standards expand and bound the provided definition as *definition away from unwanted outcomes to include the concept of a deviation from the expected*.

Surprise seems to be an appropriate element of risk. Surprise may include good and bad deviations from the expected.

For the purposes of the ISO standards, *risk includes considerations of financiality, operations, environmental, health, safety, and may impact business or society objectives at strategic, project, product or process levels*.

The discussion about a specific risk should include the events and consequences. While we may discuss the risk of an event occurring, we should include the 'so what' element as well. If an event occurs, then this consequence is the result. Of course, this can get complex quickly as events and associated consequences rarely have a one-to-one relationship.

Finally, the ISO notes on risk as including a qualitative element to characterizing risk.

As reliability professionals, these ISO definitions may seem familiar and comfortable. We have long dealt with the uncertainty of product or process failures. We regularly deal with the probability of unwanted outcomes. We understand and communicate the cost of failures.

What is new is the framework described by the ISO standards for the organization to identify and understand risk as it applies to the organization and to the customer?

Reliability risk now has a place to fit into the larger discussions concerning business, market, and societal risk management. I agree that reliability risk is a major component of the risks facing an organization in regard of its products and processes of operations. We witness the news making recalls in recent years (nuclear plant's accidents, plane crashes). We are threatened that something may happen, e.g. some companies may be ruined. As reliability professionals, we use the tools to identify risks, the tools to mitigate or eliminate risks, and the tools to estimate future likelihoods and consequences of risks. How do we view the connection between risk and reliability? Why car companies recall some vehicles from users to make a free replace of some parts in order to prevent unexpected failures?

#### 4.3.2. Discussions on Risk Related to Reliability (at Gnedenko Forum)

As reliability professionals, these ISO definitions may seem familiar and comfortable. We have long dealt with the uncertainty of product or process failures. We regularly deal with the probability of unwanted outcomes. We understand and consider also the cost of failures.

What is new is the framework described by the ISO standards for the organization to identify and understand risk as it applies to the organization and to the customer?

Reliability risk now has a place to fit into the larger discussions concerning business, market, and societal risk in management. I agree that reliability risk is a major component of the risks facing an organization in regard of its products and processes of operations. Witness the news making recalls in recent years (nuclear plant's accidents, plane crashes, product recalls). We are threatening what sometimes happen, when companies are ruined. As reliability professionals, we use the tools to identify riskiest events, the tools to mitigate or eliminate risks, and the tools to estimate future likelihoods and consequences of risks. How do we view the connection between risk and reliability? Why car companies recall some vehicles from users to replace some parts in order to prevent unexpected failures?

We usually connect the risk to some specific vision on characteristic in the reliability terms. Here I present some points of view of my Gnedenko Forum associates, and this is not an exhaustive review.

##### **The vision of I. A. Ushakov - the founder of Gnedenko Forum on the risk issues**

It is based on the following items:

- *Choice* – start or do not start the process of steps moving towards a target with some risk;
- *Event* – a one time threat at  $t_1$  realized once, when something happened;
- *Phenomenon* – an event of some duration  $t_2 - t_1$  when qualitative changes in state take place;
- *Uncertainty* - some unknown time during which the change occurs, place of manifestation, duration and force of impact;
- *State* - in each state the object is characterized by a number of attributes (properties).
- *Reliability* - the property of an object is securely functioning without fails with a 100% level of efficiency. Refusals are considered. The criterion for refusal is determined as YES / NO.;
- *Operational efficiency* of functioning  $t$  (the term introduced by Ushakov and Dzirkala) is a system that functions continuously, albeit with a reduced level of input parameters. Faithfully, this is the same reliability without a strict failure criterion, but with several levels of quality / efficiency;

- *Stability* - the property of the system to return (within a reasonable time) to the previous 100% level of functionality after the breakdown of some separate components;
- *Vitality* - the property of the object to continue function within acceptable limits after the failure of some separate components;
- *Safety* is a property of the object to perform its functions without causing any damages to the operating personnel, the environment, and so on.

*The Risk* is a measure of the quality of a particular process to achieve the goal, taking into account the uncertainties on the road. The risk for an object (process) is a value proportional to the deviation of its current state from a certain standard of quality for that object (process).

I see here – *risk is identified as a measure* (a number), without meaning. But I am looking for a meaning, like the meaning of probability as the measure of the chance something to happen. Possibly this “something” is the risk. In my opinion there are too many conditions until some risk is identified and measured. The measure seems to me not clearly set up.

#### **Professor V. A. Kashtanov – Risk is a property of a real and modeled process**

Slightly different opinion is expressed by **V. Kashtanov, another member of the Gnedenko Forum advisory board**. Here I present his vision. Everything is based on the possibility to present the situations by an appropriate mathematical process in development.

Let me first present his overall opinion about the risk.

#### **SOME METHODOLOGICAL ASPECTS OF CONSTRUCTION AND ANALYSIS MATHEMATICAL RISK MODELS**

Before talking about the features and construction technology mathematical models of risk, specifics of the mathematical apparatus should be used in the analysis of such models. It is necessary to define the concept of risk. The proposed definition of risk is based on the obvious existence in everyday understanding of the relationship between the risk and the concept of danger. The risk situation in everyday practice is associated with the occurrence of danger events. Note that the danger should be understood in a broad sense, from danger (trouble) of losing a wallet with a small amount of money up to the global danger of a thermonuclear war. Attentive reader will give numerous examples of the presence of the above connections from various life situations. All life situations are developed (change, evolve, vary, etc.) depending on time. They represent processes in which time acts as an independent variable.

Building a model of a process, it is necessary to determine the set of states of this process and consider the process as wandering (changing states) over this set. Further, observing any real life (not mathematical) process, we can talk about changing on time FEELINGS for risk or danger. This change in emotions is since the state of the observed process changes over time. Risk (danger) is a PROPERTY of the states of the real life or model process. Therefore, we can give such a qualitative definition.

Risk (danger) is a PROPERTY in the real life or in a model process. The discussed approaches so far concern only qualitative definitions and assessments

If we understand a safe situation as a situation of no risk, then the opposite of the above concept of risk should be considered no risk concept, or safety concept (no danger). Therefore, with this approach (identifying risk and danger) could not use the term “risk” and will be limited to terms “danger” and “safety”. The concept of risk uses only a measure of hazard assessment (this will be discussed below). In some works, safety is also defined as a property of the process functioning in the system.

In the Law of the Russian Federation “On Safety” in determining the quality the term “sensation” is used in the concept of security. When building mathematical models with emotionality it is impossible to work with any concepts. Required measure for the defined above property of the process functioning in the system under study we spell it as follows:

DEFINITION (measure). Risk is a quantitative indicator that evaluates the danger.

If we are talking about assessment, then there must be an observer, implementing this observation process. When assessing a hazard (as in in many other cases) opinions may differ, that



some situation (condition) is very dangerous, for some it might be not. So, risk assessment could be subjective. Objective (consistent) assessment should be linked to a process describing the evolution (functioning) of the system under study. Hence, we come to the next

**DEFINITION.** Risk as a quantitative indicator that evaluates danger, is some functional (operator) defined on the set of trajectories of the process describing the evolution (functioning) of the system under study. Let us note some features of the processes describing the functioning of the system under study:

- An observer can interfere with the operation of the system, that is managing the process. In this case, you can and should put the problem of optimization and search for an optimal strategy management.
- Uncertainty factor (randomness, limited information, the inability to observe the state of the process or observe them with errors).

**OUTPUT.** A controlled random process on the trajectories of which the functional is defined and for which the optimization problem is set (search for the optimal control strategy) can be interesting and a promising model for researching the problems of risk analysis.

Let me resume and present details in this discussion:

**A qualitative definition:** *The risk (danger) is a property of a real or modeled process.*

It is common to talk on political, social, economic, technologic, and other processes possessing risks. Then in modeling such processes it is needed to determine sets of states of each process and look at it as a random walk within the set of all states. Then the above definition of the risk makes sense and can be understood. The danger is when the process passes into the set of risky states. Safety is when the process is out of there risky states. Such an approach allows to **understand the concept of risk and the ways to assess the risk (author's note) as probability to pass from a safety set into the risky ones with a use of respective model.** Risk (maybe catastrophe) is to be in that set of risky states. Assessment of the risk is to measure the chance to get there from a one that you currently are.

Therefore, without mathematical models we may not be able to give good definitions for concepts widely used in our scientific and social life. According to Kashtanov the following definition should be valid:

*Risk as a quantitative indicator* assessing the danger of some functional operator calculated on the set of trajectories of the process that describes the evolution (the functional behavior) of the studied system.

By the way, a controller may interfere the system work, and have some control on the process. Uncertainty factors (randomness, limited information, impossibility to observe process states or making measurements with errors) create additional challenges in the risk evaluation and control.

I like this approach since it explains the models of risky events and allows to measure the chance to get into it from the safety states. This opinion of our Gnedenko Forum panelist Dr. V. Kashtanov is close to the ISO understandings.

#### **Dr. A. Bochkov – Risk as a value proportional to the deviation current state from the ideal**

Here is the opinion of **A. Bochkov**, our Gnedenko Forum <https://gnedenko.net/> corresponding secretary and since many years the motor in issuing 4 quarterly refereed and free of charge publications of our journal "Reliability – Theory & Applications" (RTA, please visit <http://www.gnedenko.net/Journal/index.htm>):

*Risk and safety* relate to the humanity. Namely the human assesses the degree of safety and risk **considering** to his own actions during the life, or the reliability of used systems as a source of potential danger, or risk that such thing may happen. *Tools, machines, systems, technical items do not feel the risk. Risk is felt by people.* How do people estimate the risk is a different question? *Safety* in most tools that people use is depending on its ergonomic design, its instructions for use, and the care of organizations and producers of items for common use. There are ergonomic decisions that



help such use and users to follow these instructions, which are always in process of improvement.

*Safety* is also a kind of complicated concept. It does not have unique definition neither some unique measure. Everything is specific and kind of mixture between objective and personal. There are no numeric measures for the safety. But I (BD) agree with the statement that *as higher is the safety, as less is the risk*. And still the measurement for the risk as (one that could introduce mathematics in its study) is missing. This is a number which should make aware people and companies, and governments to pay attention on some existing coming in the future risks but is still not clear. According to Bochkov, the *risk* should be measured as a *degree of difference of the current estimated state from the ideal*. Degrees are usually measured in percentages, or in points on some scale (BD). Here is no probability to be used. Humans usually assess the probability. But humans always do not understand what they risk with. The maximal price of the risk for humans is their own life. For the one who takes such risk the price has no numeric expression. Such a loss cannot be compensated. However, when a risk is not related to loses of life's, the maximum loses are estimated by the means available by those who take decisions on actions against risks. And usually, the human lives are priced by the money insurance agencies are paying for the lost lives.

The described approach assumes the availability of retrospective information on the implementation of the risk. The risk of an object (process) is defined as a value proportional to the deviation of its current state from a certain standard of quality of the object (process). In fact, risk is a measure of the quality of an object (process). The measure of the risk itself is the threat of changes in the composition or properties of the object (process) or its environment, or the appearance of changes associated with the occurrence of undesirable processes caused by arbitrary influences. The measure of the threat of failure to achieve the goal is considered in this case as a variable that represents a function relative to the current state of the object (process): it increases when the estimated situation approaches a certain acceptable limit, after which the object (process) cannot achieve the corresponding goals.

The problem is formulated as follows. Let there be a set of features  $X$  (risk factors), a set of acceptable implementations of situations  $O=(o_1, \dots, o_D)$ , For example, the risk was realized or not realized, and there is an objective function  $o^*: X \rightarrow O$ , whose values  $o_i=o^*x_i$  are known only on a finite subset of features  $(x_1, \dots, x_l) \subset X$ . The "sign – response" pairs  $(x_i, o_i)$  are called precedents. The set of pairs  $X_l = (x_i, o_i)_{i=1}^l$  will make a training sample. It is required to restore the dependence  $o^*$  on the sample  $X_l$ , that is, to construct a separating function  $o/H\pi\pi: X \rightarrow O$ , which would approximate the objective function  $o^*(x)$ , and not only on the objects of the training sample, but also on the entire set  $X$ . The separating function  $\pi(o)$  is called the logical selection function:  $\pi(o) \rightarrow \{0,1\}$ , indicating that the situation  $\pi(o)$  is selected into some subset of  $\pi(o): (\pi(o)=1)$ , or not ( $\pi(o)=0$ ).

In general, the separating function can be arbitrary, but for its use to be correct, it is necessary to impose several restrictions on the form  $\pi(o)$  (axioms of choice).

Axiom 1 (inheritance): if  $O' \subseteq O$ , then  $\pi(O') \supseteq (\pi(O) \cap O')$ .

Axiom 2 (consent):  $\prod_i \pi(O_i) \subseteq \pi(\cup_i O_i)$ .

Axiom 3 (of discarding):  $(\pi(O) \subseteq O' \subseteq O) \Rightarrow (\pi(O') = \pi(O))$ .

Axiom 4 (path independence):  $\pi(X_1 \cup X_2) = \pi(\pi(X_1) \cup \pi(X_2))$

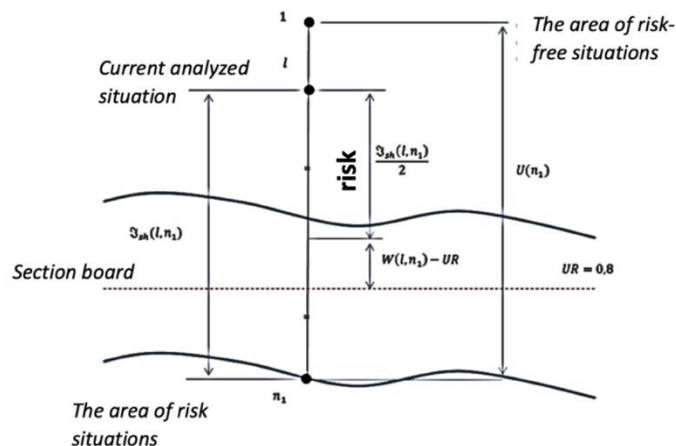
The accepted axiomatic assumptions show that the constructed separating rule should be a monotone function with respect to the set of situations identified as risk-free. As a result, the resulting situation classifier monotonically turns into a product of rules. This important property can be used so as not to retrain the classifier when data about new situations arrives.

To identify the features that need to be considered in the separating rule, it is proposed to use the Hamming metric. The value of this metric – the distance between one-dimensional objects of the same type (rows, columns) is measured by the number of pairs that do not match them. Since the problem of risk identification can be limited to a natural class of monotone functions, not all mismatches in a pair are interesting, but only "ordered" ones, i.e. use the semi-Hamming metric, which reflects only the number of "successful" features for which the value "correct" (equal to 1) in the description of the first of the compared situations, and "erroneous" (equal to 0) in the second

situation.

For risk-free situations, the risk measure is chosen as the minimum of the half-Hamming metric, which determines the distance to the interface of the analyzed situations, and for risky situations, respectively, the maximum.

Graphically, the algorithm is illustrated by the scheme shown in Fig. 1.



**Figure 1.** Illustration of the definition of the threat measure of the failure of an object (process) to reach the target state for situations with positive dynamics

#### **Prof. Isaak B. Russman – Risk as measure of the achievability of the target**

The number and variety of Professor Russman scientific works is striking - more than 100 in total, from astrophysics, to modeling structural changes in the economy, trading in securities and assessing the quality of knowledge. Possessing a systematic view of the essence of objects, he preferred to highlight a common basis in all the variety of objects and processes, the modeling and description of which he was engaged in. In this systematic approach, Dr. Russman used many original ideas, for example, allowing one to formalize control tasks using the mechanism of "difficulties in achieving the goal" (a description of the idea in a popularized form can be found here), or graphically present the control task using risk zones and the object's trajectory to goals.

These ideas, one way or another, were used in many of his works and used in the works of his followers. We hope that this page [https://www.adeptis.ru/russman/scientific\\_heritage.html](https://www.adeptis.ru/russman/scientific_heritage.html) will serve not only the memory of Isaak Borisovich, but will also serve as a serious source of new ideas, both for theorists and for practitioners working in the field of modeling and optimization of control processes for various systems. The main work of this author (in Russian) concerning the risk issues is [Об оценке интегрального риска инвестиционных проектов \(геометрический подход\)](#) In English "On the assessment of the integral risk of investment projects (geometric approach)", published in *Научно-практический вестник "Энергия"*, №3(41), 2000, с. 111-116(Russian) (Scientific and practical bulletin "Energy", No. 3 (41), 2000, p. 111-116).

Dr. Russman opinion deserves a bit more detailed description.

The risk estimates the difficulty of obtaining the declared result  $d_k$  with the existing estimates of the quality of the resource ( $\mu_k$ ), and the requirements for this quality ( $\varepsilon_k$ ). The concept of the difficulty of achieving a goal for a given quality and requirements for the quality of the resource and the result follows from the considerations that it is more difficult to obtain a result of a certain quality, the lower the quality of the resource ( $\mu_k$ ) or the higher the requirements for its quality ( $\varepsilon_k$ ).

From general considerations, the (leading measure of the risk, BD) difficulty  $d_k$  of obtaining the result should have the following basic properties:

- at  $\mu_k = \varepsilon_k$  be maximal, i.e. equal to one (indeed, the difficulty of obtaining a result is maximum at the lowest permissible quality value);

- for  $\mu_k = 1$  and  $\mu_k > \varepsilon_k$ , be minimal, that is, equal to zero (for an extremely high possible value of quality, regardless of requirements (for  $\varepsilon_k < 1$ ), the difficulty should be minimal);
- for  $\mu_k > 0$  and  $\varepsilon_k = 0$  be minimal, i.e. equal to zero (obviously, if no requirements are imposed on the quality of the resource component, and  $\mu_k$  is greater than zero, then the difficulty of obtaining a result for this component should be minimal).

For these three conditions, for  $\varepsilon_k < \mu_k$ , a function of the form is valid:

$$d_k = \frac{\varepsilon_k(1 - \mu_k)}{\mu_k(1 - \varepsilon_k)}$$

We also assume that  $d_k = 0$  for  $\mu_k = \varepsilon_k = 0$  and  $d_k = 1$  for  $\mu_k = \varepsilon_k = 1$ .

The functioning of a reliable system is characterized by maintaining its main characteristics within the established limits. The actions of such a system are aimed at minimizing deviations of its current state from some given ideal-goal. In relation to the system, the goal can be considered as the desired state of its outputs, i.e. not only the value of its target function.

Consider a system in the process of achieving a goal, moving from its current state to some future result, the quantitative expression of which is  $A_{pl}$ . Let's assume that the goal is completed in time  $t_{pl}$ . We also assume that there is a minimum speed  $v_{min}$  of movement to the target in time and a maximum speed  $v_{max}$ . It is most convenient to measure the quantitative expression of the result and the time required to achieve it in dimensionless quantities; for this purpose, we assume  $A_{pl}$  and  $t_{pl}$  equal to one or 100%. In Figure 1, the minimum and maximum velocity trajectories of the system correspond to the OD and OB lines.

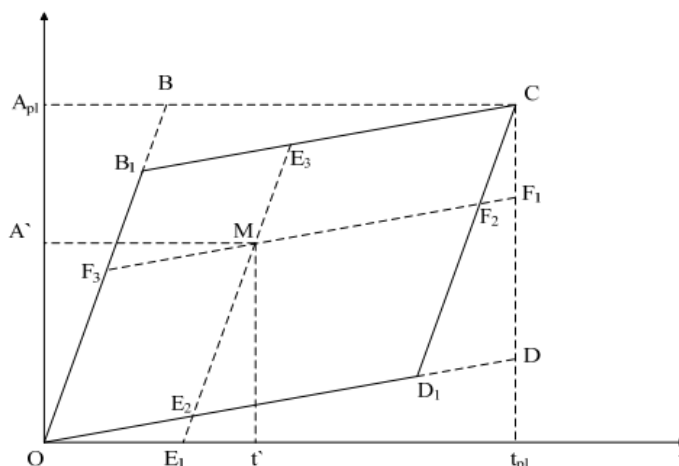


Figure 2. Geometric interpretation of the system's movement towards the goal

The polyline  $OD_1C$  is the boundary of the restricted area, and for any point  $M$  with coordinates  $(t', A')$  that describes the position of the system on an arbitrary trajectory to the target within the parallelogram  $OB_1CD_1$ , the distance is taken as the risk of not reaching the target:

$$r(M) = \max \left\{ \ln \frac{1}{1 - d_1}, \ln \frac{1}{1 - d_2} \right\},$$

where  $d_1 = \frac{\varepsilon_1(1 - \mu_1)}{\mu_1(1 - \varepsilon_1)}$ ,  $d_2 = \frac{\varepsilon_2(1 - \mu_2)}{\mu_2(1 - \varepsilon_2)}$ ,  $\varepsilon_1 = \frac{|E_1E_2|}{|E_1E_3|}$ ,  $\mu_1 = \frac{|E_1M|}{|E_1E_3|}$ ,  $\varepsilon_2 = \frac{|F_1F_2|}{|F_1F_3|}$ ,  $\mu_2 = \frac{|F_1M|}{|F_1F_3|}$

### Dr. V. V. Rykov – Risk as a two-dimensional process

One more opinion of our RTA chief editor V. V. Rykov, an established expert in reliability is briefly shown here. It is well presented in his book [10]. He relates almost each system reliability characteristic to some respective risk, with an emphasis on the economic consequences when a risky event happens. And probability models are used in each situation. My brief review (BD) follows.

Variety of risks accompanied individual people during their life. Same things happen with various industrial product lines, agricultural, financial companies, biological, environmental

systems, and many other units. Risk appears due to the uncertainty of some events that may occur with respect of which corresponding decision and actions should be taken. The mathematical Risk Theory is generated and developed on the actuarial science, where the risk is that an insurance company will be ruined and is measured by the *probability of ruin*.

However, nowadays the understanding of risk is related to the occurrence of some “risky” event and with related to it consequences in terms of material or monetary losses to restore. Numerous examples support this position one can read in the Rykov’s book [10]. But we focus on situations related to reliability. Before that let see his mathematical definition of the risk given there.

Risk is related to the occurrence of some random (risky) event  $A$ , which probability  $P_t(A)$  varies in the time  $t$ , depending on current conditions. Occurrence of such event at time  $t$  generates some damages measured by a value  $X_t$ . In this way the risk is characterized by two variables  $(T, X_T)$ , where  $T$  is the time of the occurrence of the risky event  $A$ , and  $X_T$  is the measure of the damages then. Both components can be random.

In Reliability theory  $T$  can be the time of use of a technical system, that may vary according to the reliability model of management of this system and of its structure. Application of this approach is demonstrated in analysis of technological risk for popular reliability systems in the frames of some known failure models. The focus in [110] is on the measured risk.

As a basic measure it is proposed the two dimensional distribution function  $F(t, x) = P\{T \leq t, X_T \leq x\}$ . I like this approach. It admits lots of analytical analysis of some particular characteristics and cases, as demonstrated in this book [10], and used in many other new research projects.

But this is done just within theoretical frames and assumptions. Practical applications need sufficient data for such a function  $F(t, x)$  to be appropriately estimated. I do not refer to any particular result from [10]. Just note that the variable  $T$  varies depending on the reliability system model where a catastrophic event may happen. Then the value of the losses  $X_T$  will be known. And also, the value of losses which extend some critical value  $X_{critical}$  is achieved or is close to be achieved. Then the process should be stopped (times to stop are also interesting to be analyzed in such approach) to be able to *prevent the risk*. Such discussion is not used in that book but deserves attention.

And somehow, I could not see any satisfactory other measure of the risk shown, different from the amount of expected loses. Average characteristics are found in [10] but no clear measure of the risk which could be used as a BELL RING THAT THE RISK IS NEARBY. I believe that if the risk is measured in terms of odds, then when the odds reach certain level, the bell must ring loud to make an alarm for those who are supposed to take care about that risk. Somme odd’s addition to Rykov’s approach might be interesting and useful. However, it will be nice and useful to havve Dr. Rykov presentation at this Conference talks and discussians.

#### **Prof. N. Singpurwalla – A Bayesian point of view on the risks in reliability**

This point of view isdiscussed in the book [11]. I do not dare to get into details. JJust recommend everyone interested in that approach to look in the source. The authoor writes:

There is an increasing emphasis on what is commonly referred to as ‘risk’, and how best to manage it. The management of risk calls for its quantification, and this in turn entails the quantification of its two elements: the uncertainty of outcomes and the consequences of each outcome. The outcomes of interest here are adverse events such as the failure of an infrastructure element, e.g. a dam; or the failure of a complex system, e.g. a nuclearpower plant; or the failure of a biological entity, e.g. a human being. ‘Reliability’ pertains to the quantification of the occurrence of adverse events in the context of engineering and physical systems. In the biological context, the quantification of adverse outcomes is done under the label of ‘survival analysis’. The mathematical underpinnings of both reliability and survival analysis are the same; the methodologies could sometimes be very different. A quantification of the consequences of adverse events is done under the aegis of what is known as utility theory. The literature on reliability and survival analysis is diverse, scattered and plentiful. It ranges over outlets in engineering, statistics (to include

biostatistics), mathematics, philosophy, demography, law and public policy. The literature on utility theory is also plentiful, but it is concentrated in outlets that are of interest to decision theorists, economists and philosophers.

One of the aims of this book is to develop material in reliability with the view that the ultimate goal of doing a reliability analysis is to appreciate the nature of the underlying risk, and to propose strategies for managing it. The second is to introduce the notion of the 'utility of reliability', and to describe how this notion can be cast within a decision theoretic framework for managing risk.

The second aim of this book is to summarize the Bayesian methodology in its broader context. Much of this Bayesian methodology is directed toward predicting lifetimes of surviving units.

The focus has been on the use of concepts and notions in reliability theory that are germane to econometrics and finance, in particular the assessment of income inequalities and financial risk.

I (BD) am not in the position to judge the Bayesian approach and, concepts and models presented in [11]. I only notice that I could not find a clear definition of "what is the risk" and how one can measure how close a process is to a risky situation. And let me share with you an opinion about this book: "What I liked most about this book, however, is the way it blends interesting technical material with foundational discussion about the nature of uncertainty." (*Biometrics*, June 2008).

#### **My authors opinion on the risk and its measures**

During my work on this "risk issue" I have seen so many opinions and approaches, as most of you probably met too. Most of these are focused on particular situations, issues, or cases. That's why I got to some general definition of the risk in a random process, and the ways of its evaluation. My vision is as an opinion of a mathematician, experienced in Applied Probability, modeling and data analysis the theory and practice.

First of all, random processes need some realistic adequate probability model where the set of states  $\Omega$  of the process are defined (known, well described, observed and controlled) at any time, and dynamic probabilities of changes should be known. In other words, it should be defined a probability space  $\{\Omega, @_t, P_t\}$ . Here  $\Omega$  is the set of all possible states of the process, and their meaning cannot vary in the time;  $@_t$  is the uncertainty model at the time  $t$ , and  $P_t$  is the probability that works in that time. Let  $B_t$  be the set of undesired (adverse, risky) events. It is an element of  $@_t$ , and let  $A_t$  be the set of current states (also an element of  $@_t$ ).  $A_t$  Could be a result of a test performed at time  $t$ , or a current state.

**Definition:** Risk at time  $t$  is the set  $B_t$  of undesired (risky) events. Let  $A_t$  be the set of current states in the probability space  $\{\Omega, @_t, P_t\}$  that describes the evolution of the observed process. Measure of the risk at this moment is the relative risk

$$RR(B_t \text{ with respect to } A_t) = P(B_t | A_t) / P(B_t | \text{complement of } A_t)$$

in terms of odds.

This definition is a replicate borrowed from the measurements of risk in biostatistics presented above. By the way, in the cases discussed in Bochkov terms where safety and risky sets have nothing in common, this measure = 0; In Kashtanov's and Rykov's considerations it needs calculations. Please, notice that  $A_t$  and  $B_t$  may particularly overlap. As more  $A_t$  is covering  $B_t$ , the higher the risk is (please, look in my measures of dependence discussed above). I think, this is a natural measure. However, the regression coefficients also can be used as measures of dependence of the risk on the current situation.

I am not sure that the risks in reliability analysis should stop here. Reliability itself is defined as the probability that a system functions at a given time. Therefore, the fact that it does not work then it is a kind of risky event. The availability coefficient also is probability of an event,

that the system is able to function at a certain time. And many other functions, like failure rates, number of renewals, maintenance costs, effectiveness, expenses for supporting functionality, etc. have quantitative measures that could be used as risky variables. Each of these uses a construction of an appropriate probability space, where the above general definition and evaluation of the risk can be applied. Therefore, the issues for the risk assessments are not finished yet. There is a lot of open questions for researchers to work on.

According to me, the risk is a concept with many faces. Possibly the reliability may highlight some of it. Following to the above views and discussions I got to the following

**Proposition:** The risk (analogously the reliability) is a complex notion that is measured by numerous indexes depending on different applications and situations. But the basic measures should be unique.

#### IV. Conclusions

Axioms are the foundation in establishing new areas of theoretical research.

In each theory there arise numerous of new concepts that have meaning, may generate new axioms and areas of studies, which need explanation, metrics for evaluation useful in their better understanding and clear practical use and applications.

*Reliability and Risk* can most profitably to be used by practitioners and research workers in reliability and survivability as a source of information, reference, and rise new open problems.

Our article is tracing some ways of developing useful knowledge in studying the risky events and problems in the real-life.

I am sure that the risks in reliability analysis should not stop here

## References

- [1] Gnedenko, B. V. Theory of Probability, 6-th edition, CRS Press, Taylor & Francis group, Boca Raton, New York, London (2005)
- [2] Kolmogorov A. [1933]. Foundations of the theory of probability, *New York, USA: Chelsea Publishing Company (1950)*.
- [3] Feller, W. An Introduction to Probability Theory and Its Applications, Vol. 1, John Wiley and Sons (1968).
- [4] Dimitrov, B., and Yanev, N. *Probability and Statistics*, 3-rd edition, Sofia University "Kliment Ohridski", (2007).
- [5] Dimitrov, B., Some Obreshkov Measures of Dependence and Their Use, *Compte Rendus de l'Academie Bulgare des Sciences* Volume 63, No.1, (2010) pp. 15-18.
- [6] Dimitrov, B. Measures of dependence in reliability, *Proceedings of the 8-th International Conf. on Mathematical Methods in Reliability MMR2014*, Stellenbosch, South Africa, (2013) pp 65-69.
- [7] Dimitrov B. Dependence structure of some bivariate *distributions*, *Serdica J. Computing*, v.8, No 3, (2014) pp.101–122.
- [8] Obreshkov, N. Theory of probability, Nauka I Izkustvo, Sofia )1963) – in Bulgarian.
- [9] Vickers, A. What is the P-value anyway? 34 Stories to Help You Actually Understand Statistics Pearson UK Higher Education (2010).
- [10] Rykov, V. V. Reliability of Engineering Systems and Technological Risks, Wiley and Sons (2016).
- [11] Nozer D. Singpurwalla, Reliability and Risk A Bayesian Perspective, Wiley and Sons (2006).