DOI: 10.20472/BMC.2015.001.008

SIMONA HAŠKOVÁ

University of Economics in Prague, Faculty of Management, Czech Republic

NORMATIVE AND FREQUENCY FORMS OF BAYESIAN CALCULATIONS IN MANAGERIAL STUDIES

Abstract:

The calculations based on Bayesian statistics are a part of many decision-making problems. Their solution relies in most cases on an exact approach based on Bayesian formula, in which the values of occurrences of examined phenomena in the form of prior and posterior probabilities are supplied. The findings of evolutionary biologists confirm that such an approach is less psychologically acceptable for untrained individuals in the field of mathematics and logic, and results in many wrong solutions. The paper proposes the way how to increase the correctness of solutions based on conditional probabilities by means of the calculation procedure of a frequency form represented by Venn diagram. This is based on an adaptation of the task presentation in a way more natural to human understanding. Its demonstration is performed in a managerial task focusing on the evaluation of the probability model of the project relating to the potential oil field deposit. The Bayesian formulas are here replaced by simpler Laplace probabilities that are comprehensible to managers and economists without prejudice to the accuracy of the result.

Keywords:

Bayesian formula, normative form, frequency form, Venn diagram, probability model of a project

JEL Classification: C51, D81, A12

1 Introduction

The results of managerial solutions are not nearly as reliable on average (and cannot by definition be) as the results of solutions in technical fields (Barber et al., 2000; Standing et al., 2006; Pinto & Mantel, 1990). There are many causes behind it. Besides the complexity factor, which reduces the reliability of solutions by bringing into play the *known* “*unknowns*” (consequences of which the managerial calculations do not take into account) and also the *unknown* “*unknowns*” (Rumsfeld (2002): “The ones we don´t know we don´t know”), and that by the nature of things are present in management more often than in technology, it has other causes.

These include, for instance, the fact that technology as the application of natural science consistently relies on *robust* natural law. In contrast to it, the various managerial disciplines are based on specific purpose-built applications of general microeconomics that studies the behaviour of people in conditions of scarcity (Frank et al., 2007); and people tend to use every bit of knowledge they acquire to their advantage. The diversity of people’s interests is thus reflected in the establishment in advance unknown deviations from the standard assumptions, from which managers proceed within their calculations, which brings an element of uncertainty to their conclusions. These causes of lower reliability of managerial calculations are the *objective* causes. With them, however, we can hardly do anything else than to try to improve the prediction of what is at least partially predictable (Gaskill, 1993).

The object of this paper is to show possible way for reducing the *subjective* causes of the incorrect managerial solutions regarding the managerial probabilistic estimates.

The subjective causes are mainly regarded to be the *cognitive distortions at the mental level of thinking* of a problem solver profoundly analyzed by e.g. Cohen (1992), Kahneman & Tversky (1996), Gigerenzer (1996). These distortions often emerge if problem solvers enter the environment of uncertainty, which significantly concerns the judgments based on conditional probabilities. Within this context the paper introduces the alternative procedure for the probability calculations that allows “constructing” the solutions by means of the easy to follow steps. This procedure facilitates the task understanding and therefore increases the likelihood that the problem solver will not commit false considerations leading to an incorrect result.

In order to achieve this, we first analyse the normative form to the problem solution presented by the Bayesian formula. Then the abstract Bayesian formula will be interpreted by means of the instrument of Venn diagram, the procedure of which corresponds to the frequency form in the sense of Gigerenzer´s idea (Gigerenzer & Hoffrage, 1995; Meder & Gigerenzer, 2014). Both procedures (in the normative and frequency form) will be demonstrated in probability calculations of the probabilistic model of a managerial project.

2 Methodology

Generally, there are two approaches to a rational solution of specific tasks: one of them proceeds from “above” (the exact approach); the second from the “bottom” (the “case-based reasoning” approach). The process from “above” starts with the model of more general tasks (general formula), which is adapted to the particular problem by an appropriate choice of free parameters where necessary (Hašková & Zeman, 2014). The procedure from the “bottom”, on the other hand, builds on the knowledge of the specifics of a particular task that are adequate to the chosen perspective on the problem solved (Ackoff, 1973).

In both cases, the problem can be presented in:

- the normative form (e.g. probability theory or Bayesian statistics, etc.);
- the frequency form.

The latter form is here presented by means of Venn diagram, in which the frequencies correspond to the numbers of occurrences of considered phenomena. The Venn diagram structure is described by formal logic e.g. in (Shin, 1994). The normative form is processed in terms of Kahneman´s & Tversky´s idea of Bayesian inference (Griffin & Tversky, 1992). Both forms will be showcased in the probability calculations of a managerial probabilistic model of a project regarding the decision making about the optimal exploration method of the potential oil field and the strategy of oil production.

3 Introduction to a managerial study

A giant oil field along the Sierra Leone coast was confirmed in 2009 by a consortium of energy companies led by the US firm Anadarko Pertroleum (Simba Energy, 2012). Before any offshore drilling takes place, mining company performs exploratory drilling or seismic survey to get information about the oil amount presence in the sea.

In this connection a complex study was elaborated to determine the optimal exploration method of the potential oil field and the strategy of oil production (Hašková & Kolář, 2013). This contribution focuses on the formal probabilistic model construction of this project that consists in the probabilistic evaluation of the possible project´s scenarios.

The estimates of the probabilities of phenomena are a normal part of managerial life. Economic theory often relies on Bayesian statistics that views all uncertain phenomena as conditional probabilities. The Bayesian statistical analysis starts by assigning prior (initial) probabilities to the occurrences of the examined phenomena; they reflect what we think about these phenomena on the basis of our indigenous knowledge or subjective ideas, without knowing anything specific (e.g. Fleiss et al., 2003; Smith & Gelfand, 1992). The posterior (conditional) probability is the „revised“ probability of an event occurring after taking into account new information.

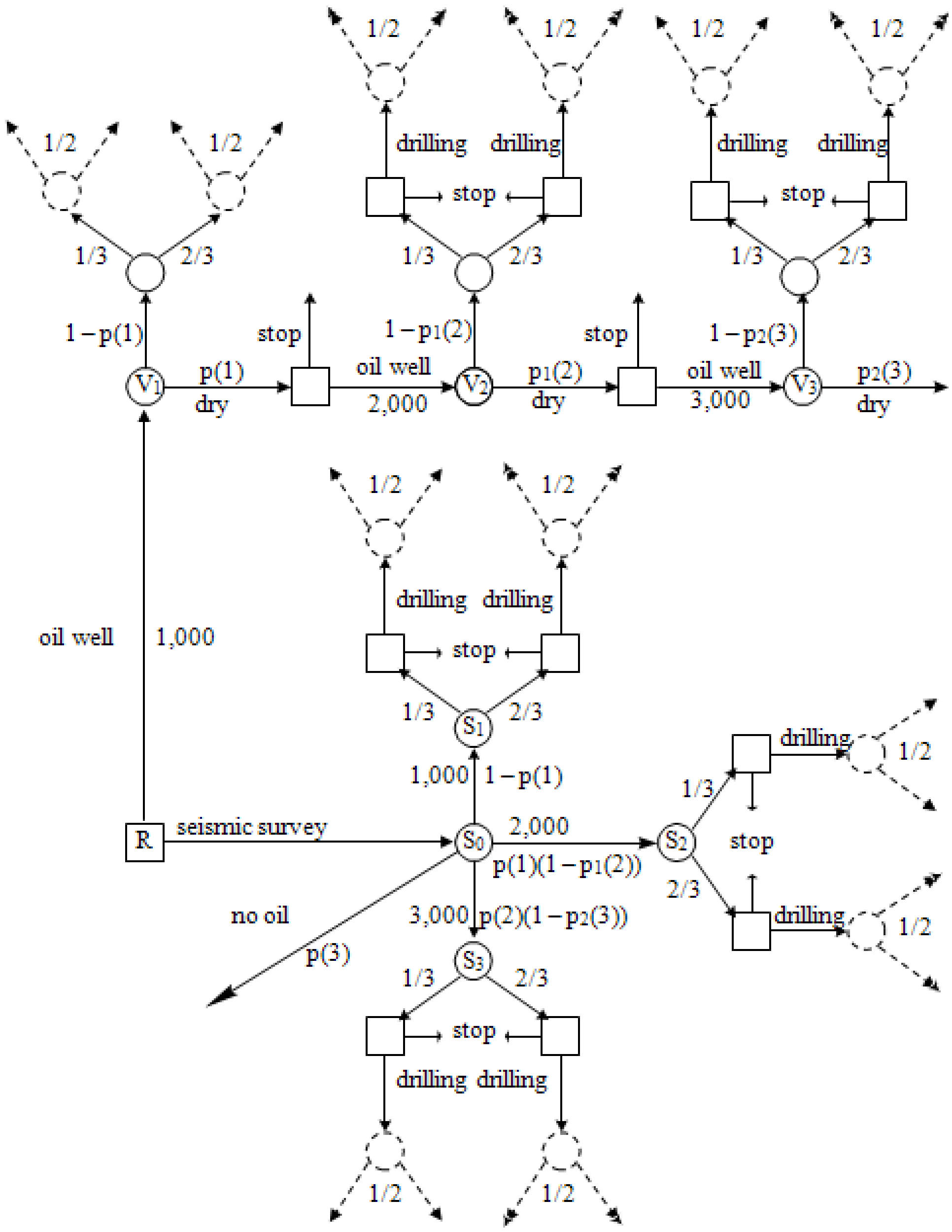
Any error in an intuitive estimate of the prior probability values or an error in the formulations of intuitive relations for the posterior probabilities can lead to an erroneous decision and thus to “sink” a good project or, conversely, to lead to investment in a poor project.

4 Entrance data to the formal project model

The experience in geologically similar oil field areas as our considered potential oil field says that oil is encountered at a depth of 1,000 feet in 70 % of all exploratory wells; the oil is struck at a depth of 2,000 feet in the case of 10 % of all exploratory wells and in approximately with the same probability (i.e. 10 % of all exploratory wells) the oil is not struck even at a depth of 3,000 feet. If there is oil, it can be a more or less abundant source, which is subsequently referred to as being a “rich” or “poor” source of oil. A “rich” source of oil can be encountered about twice as often as a “poor” source.

The construction of the probabilistic model of the project regards to refine and formalize the obtained information about the occurrence and absence of oil at the depth of 1,000, 2,000 and 3,000 feet in connection with the formal model of the project shown in Fig. 1 (Hašková & Kolář, 2013). The formal model reflects all potentially possible sequences of sub-activities within the project including their possible outcomes formed into the oriented graph. The arrows (valued edges) emanating from the *decision nodes* (squares), where the manager considers the options of terminating the project either in the exploration stage or before the stage of drilling, reflect the results of those decisions. The edges exiting from the *situational nodes* (circles) marked as the solid line inform about the degree of “odds” (in terms of probability) of the presence of oil at the death of 1,000, 2,000 and 3,000 feet and/or of the quantity at a given depth. The edges leading to the leaves of the tree, in which the oil is drilled, are (unlike those in which it is not drilled) marked as a broken line (including edge nodes from which they exit). They distinguish the low oil demand from the high demand, while the possibilities are equally probable.

Figure 1 The structure of the formal model of the project



Source: Own elaboration

4.1 The exact approach to the probabilistic model construction: theoretical bases

Let us consider some phenomenon *A* with prior probability *p*(*A*) > 0. Then the subsequent awareness of the occurrence of some other phenomenon *B* with the prior probability *p*(*B*) > 0 changes prior probability *p(A)* to posterior (i.e. based on sensory experience) probability $p_B(A)$ – the conditional probability of event *A* if the event *B* occurred. In the same vein $p_A(B)$ is the conditional probability of event *B* if phenomenon *A* occurred. If $A \wedge B$ means the simultaneous occurrence of events *A* and *B*, then this applies

$$p(A \wedge B) = p(B) \cdot p_B(A) = p(A) \cdot p_A(B) \quad (1)$$

From (1) it can be calculated (I) $p_B(A)$ by means of $p(A \wedge B)$ and $p(B)$, (II) $p(A \wedge B)$ by means of $p(B)$ and $p_B(A)$ and (III) $p_B(A)$ by means of $p_A(B)$, $p(A)$ and $p(B)$ – Bayesian formula in the form

$$p_B(A) = p_A(B) \cdot p(A) / p(B) \quad (2)$$

The conditional probability represents the core of serious debate between the so-called Bayesian statisticians, who are its proponents, and conventional statisticians operating with classical instruments (to which, among other things p-values, permissible errors and significance levels belong), who are the opponents (Rosenthal, 2006). The conventional statisticians highlight the fact that Bayesian inference requires determination of the prior probability expressing one´s faith before the experiment begins, while it is not clear at all how to actually choose these probabilities (Camerer & Loewenstein, 2004, Saks & Uggerslev, 2010).

4.2 The Bayesian normative form of probability calculations

The expressions for the probability of the edges extending from some of the situational nodes (see Fig. 1) correspond to the unknown constants of the type $p(i)$, $i = 1, 2, 3$ and $p_{i-1}(i)$. Their specific values result from the following definitions and considerations:

Let us denote by a symbol i a depth of an exploratory well in thousands of feet and by a symbol $\boldsymbol{i}$ or respectively $\neg\boldsymbol{i}$ a phenomena consisting in the fact that in the depth i the exploratory well is dry or respectively wet. Then the probabilistic task can be solved by the exact approach (from the "above") in the normative form presented by Bayesian formula described by (2) as $p_{\boldsymbol{i-1}}(\boldsymbol{i}) = p_{\boldsymbol{i}}(\boldsymbol{i-1}) \cdot p(\boldsymbol{i}) / p(\boldsymbol{i-1})$ putting in relation the conditional probabilities of occurrence of two mutually dependent phenomena $\boldsymbol{i-1}$ and $\boldsymbol{i}$ at the given exploratory well (the index is the conditioning phenomenon, the argument is the conditional phenomenon) with their probabilities without index (unconditional). For every $i = 1,2,3$ applies $p_{\boldsymbol{i}}(\boldsymbol{i-1}) = 1$ a $p(\neg\boldsymbol{1}) = 0.7$, $p(\boldsymbol{3}) = 0.1$, $p(\boldsymbol{1} \wedge \neg\boldsymbol{2}) = 0.1$, where the symbol $\wedge$ signs for a binary operator of a simultaneous occurrence of both phenomena. The task is to calculate $p(\boldsymbol{1})$, $p_{\boldsymbol{1}}(\boldsymbol{2})$, $p(\boldsymbol{2})$ a $p_{\boldsymbol{2}}(\boldsymbol{3})$.

Solution:

Given that $p_{\boldsymbol{i}}(\boldsymbol{i-1}) = 1$ for $i = 1,2,3$ the Bayes formula degenerates to $p_{\boldsymbol{i-1}}(\boldsymbol{i}) = p(\boldsymbol{i}) / p(\boldsymbol{i-1})$. Since $p(\neg\boldsymbol{1}) = 1 - p(\boldsymbol{1}) = 0.7$, then $p(\boldsymbol{1}) = 1 - 0.7 = 0.3$. Furthermore, $p(\boldsymbol{1} \wedge \neg\boldsymbol{2}) = p(\boldsymbol{1}) \cdot p_{\boldsymbol{1}}(\neg\boldsymbol{2}) = p(\boldsymbol{1}) \cdot (1 - p_{\boldsymbol{1}}(\boldsymbol{2})) = 0.1$, from which $p_{\boldsymbol{1}}(\boldsymbol{2}) = 1 - (0.1 / p(\boldsymbol{1})) = 1 - 0.1 / 0.3 = 2 / 3$. By substituting $\boldsymbol{i} = \boldsymbol{2}$ to the Bayes formula we get $p_{\boldsymbol{1}}(\boldsymbol{2}) = p(\boldsymbol{2}) / p(\boldsymbol{1})$, from which $p(\boldsymbol{2}) = p_{\boldsymbol{1}}(\boldsymbol{2}) \cdot p(\boldsymbol{1}) = (2 / 3) \cdot 0.3 = 0.2$. In the same vein, for $\boldsymbol{i = 3}$ it applies $p_{\boldsymbol{2}}(\boldsymbol{3}) = p(\boldsymbol{3}) / p(\boldsymbol{2}) = 0.1 / 0.2 = 1 / 2$.

5 The frequency form as an alternative form of presentation of Bayesian model

The Bayesian approach advocated e.g. by E. T. Jaynes in (2003) is, in terms of his theory, the extension of the classical Aristotelian logic to the case of statements of truth values, which lie in the range between absolute truth and absolute untruth; therefore, it ranks it as fuzzy logic. However, for the untrained individual in logic and mathematics this is less psychologically acceptable, which implies that the human brain is not furnished with the "mental logic" that would help to solve similar types of problems. This thesis is supported by evolutionary geneticists, who explain why it is so from their point of view (Ridley, 2003).

The consequence of this is a high rate of incorrect solutions relating to the Bayesian types of tasks (see e.g. Kahneman, 2011; Mlodinow, 2008). Gigerenzer says to this that the reason for erroneous judgments may not be the "tendency" of solvers to systematically succumb to illusions, but rather the fact that (Gigerenzer, 1998): "*If a human reasoning system enters an environment where statistical information is formatted differently from that encountered in the environment in which humans evolved, the reason algorithms may fail.*" And as he further says, the problem can be solved by the adaptation of mental algorithms to their environment. Evolution equipped us with gathering and counting information obtained on the basis of case by case involving observation in the form of frequencies rather than with formal statistical skills and deductive reasoning (Gigerenzer, 1998; Gigerenzer & Gaissmaier 2011).

5.1 The transformation of the normative form of the Bayesian model presentation in the frequency form

The question arises what other tools then the Bayesian approach to offer to the category of solvers like practising managers and economists, who seldom belong to trained mathematicians or logicians so that the their results were reliable (in the sense of correctness)?

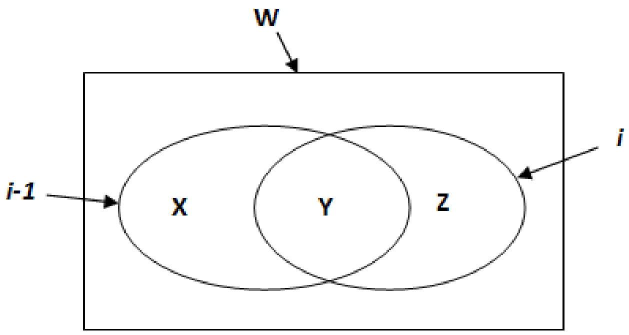
In the next we will follow and develop this idea while bearing in mind the evolutionary tendency of the problem solvers towards the observation and reasoning in the form of frequencies. Its nature lies in presenting a problem in the structure of individual easy to follow steps, which facilitates understanding of the task and thus contributes to its correctness.

The construction of a solution in this form requires from the solvers analytical abilities and creativity of a different kind then procedures of Bayesian statistics. One of the useful frequency forms of statistical problems presentation is an adequately interpreted Venn diagram that mediates the accessible vision on the problem to those who are not familiar with conditional probabilities.

5.2 Venn diagram

Numbers *X, Y, Z* or respectively *W* are the frequencies (numbers of exploratory wells) in which phenomena $\boldsymbol{X} = \neg \boldsymbol{i} \wedge (\boldsymbol{i-1})$, $\boldsymbol{Y} = \boldsymbol{i} \wedge (\boldsymbol{i-1})$ and $\boldsymbol{Z} = \boldsymbol{i} \wedge \neg(\boldsymbol{i-1})$ occur (see Fig. 2), or respectively the universe of consideration ***W*** (consisting of all contemplated exploratory wells); the equivalent of the relation $p_i(\boldsymbol{i-1}) = 1$ is $\boldsymbol{Z} = \boldsymbol{\emptyset}$ in the diagram, or $Z = 0$. This is reflected in the diagrams in Fig. 3 proceeding from the same universe ***W*** ($W = 100$), which are the frequency equivalents of the degenerated Bayesian formulas $p_{i-1}(\boldsymbol{i}) = p(\boldsymbol{i}) / p(\boldsymbol{i-1})$ for $i = 1,2,3$.

Figure 2 Venn diagram as a frequency form of presentation of the Bayesian managerial problem



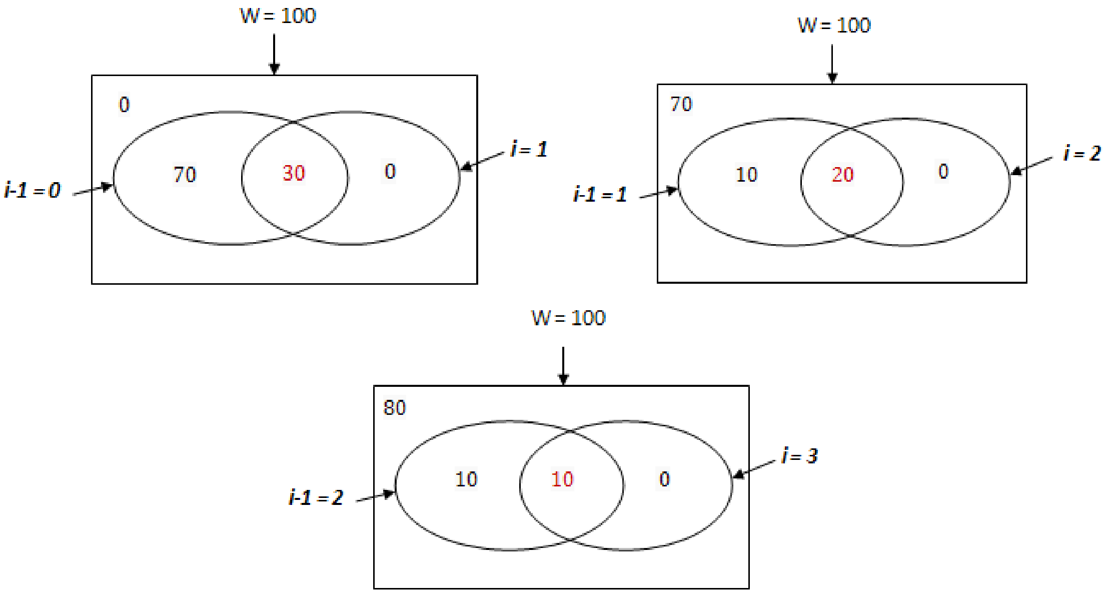
Source: Own elaboration

Therein the marked ellipses associate the exploratory wells of the property ***i*** or respectively ***i-1***. Therefore, if they appear repeatedly in two diagrams (in our case there is always the left ellipse of the following diagram the repetition of the right ellipse of the previous diagram) the total number of the wells associated in them must remain the same in both diagrams. On this consideration the following solution method is based (see Fig. 3):

- We first record what we know from the task assignment in the trio of diagrams (black numbers);
- This will be completed with the red numbers in the relevant diagrams calculated in the way that the sum of all wells is 100 in the diagram or in the way that the numbers of wells of the identical ellipses are the same in different diagrams;
- In such supplemented diagrams we enumerate the probability $p(\boldsymbol{i})$ as a ratio of the number of wells in the ellipse ***i*** to *W* (thus $p(\boldsymbol{1}) = 30 / 100 = 0.3$ and $p(\boldsymbol{2}) = 20 / 100 = 0.2$) and $p_{i-1}(\boldsymbol{i})$ as a ratio of the numbers of wells in the ellipse ***i*** to the numbers of wells in the ellipse ***i-1*** (i.e. $p_1(\boldsymbol{2}) = 20 / 30 = 2 / 3$ and $p_2(\boldsymbol{3}) = 10 / 20 = 1 / 2$).

The mutual correspondence between the calculations of the frequency form carried out in Venn diagrams and the formulas of Bayesian statistical calculus of a normative form is apparent from the comparison of the two solution methods.

Figure 3 Venn diagram as a tool for calculating the unknown probabilities $p_1(2)$, $p(2)$ and $p_2(3)$



Source: Own elaboration

6 Discussion and conclusion

The discussed conditional probability calculations are the subject of many scientific analyses. The majority notice the descriptive side of things, which they interpret; the minority put forward proposals, which should lead to a reduction of the problems that often occur in the solutions of these tasks. From the perspective of an exact approach by Bayesian formula, the poor results of solvers are due to their inclination to various forms of distortion of judgment and the lack of effective comprehension of the Bayesian procedures. To avoid this requires an initial total understanding of a given problem.

The purpose of this paper was to show a different way how to cope with the probability task solution. Pointing out that if the problem solver is an individual who is not a trained mathematician or a logician, it is desirable to introduce such solution procedures that are more comprehensible for them (this category of solvers often involves economists and managers). This way is purposed in the paper by means of presenting the probability task by Venn diagram, which corresponds to the task solution in the frequency form. Such a procedure facilitates the cognitive processes of understanding the Bayesian tasks and allows the feed-back control of the solution reached.

Reference

ACKOFF, R. L. (1973). Science in the Systems Age: Beyond IE, OR, and MS. In: *Operations Research,* Vol. 21, No. 3, pp. 661-671.

BARBER, P.; GRAVES, A.; HALL, M.; SHEATH, D.; TOMKINS, C. (2000). Quality failure costs in civil engineering projects. In: *International Journal of Quality & Reliability Management*, *17*(4/5), pp. 479-492.

CAMERER C.; LOEWENSTEIN G. (2004) *Behavioral economics: Past, present, future*. Princeton: Princeton University Press.

COHEN, J. (1992). A power primer. In: *Psychological bulletin*, *Vol. 112(1),* pp. 155.

FLEISS, J.L.; LEVIN B.; PAIK, M.C. (2003) *Statistical methods for rates and proportions*. John Wiley & Son. Canada.

FRANK, R. H.; BERNANKE, B.; JOHNSTON, L. D. (2007). *Principles of economics*. New York: McGraw-Hill/Irwin.

GASKILL, L. R.; VAN AUKEN, H. E.; MANNING, R. A. (1993). A factor analytic study of the perceived causes of small business failure. In: *Journal of Small Business Management*, *31*(4), pp. 18.

GIGERENZER, G. (1996) On narrow norms and vague heuristics: A reply to Kahneman and Tversky. In: *Psychological Review*, Vol 103(3), pp. 592-596.

GIGERENZER, G. (1998) Ecological intelligence: An adaptation for frequencies. In: *The evolution of mind*, pp. 9-29.

GIGERENZER, G.; GAISSMAIER W. (2011) Heuristic Decision Making. In: *Annual Review of Psychology*, Vol. 62: pp. 451-482.

GIGERENZER, G.; HOFFRAGE U. (1995) How to improve Bayesian reasoning without instruction: Frequency formats. In: *Psychological Review*, Vol 102(4), pp. 684-704.

GRIFFIN, D.; TVERSKY A. (1992) The weighing of evidence and the determinants of confidence. In: *Cognitive Psychology*, Vol. 24, Issue 3, pp. 411–435.

HAŠKOVÁ, S.; KOLÁŘ, P. (2013) The mathematic modeling of process economization of natural resources. In: *China-USA Business Review*: vol. 12, No. 2, pp. 122-129.

HAŠKOVÁ, S.; ZEMAN, R. (2014) Analytical microeconomics as a search tool for the stock optimization. *Proceedings of the 40th International Conference on Mathematical Methods in Economics „MME 2014“*, pp. 269-274.

JAYNES, E.T. (2003) *Probability Theory - The Logic of Science*. Cambridge University Press, Cambridge.

KAHNEMAN, D. (2011) *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.

KAHNEMAN, D.; TVERSKY A. (1996) On the reality of cognitive illusions. In: *Psychological Review*, Vol 103(3), pp. 582-591.

MEDER, B.; GIGERENZER, G. (2014) *Statistical Thinking: No One Left Behind Probabilistic Thinking*. In: Advances in Mathematics Education, pp 127-148

MLODINOW, L. (2008). *The Drunkard's Walk: How Randomness Rules Our Lives*. Pantheon Books, New York.

PINTO, J. K.; MANTEL JR, S. J. (1990). The causes of project failure. In: *Engineering Management, IEEE Transactions on*, *37*(4), pp. 269-276.

RIDLEY, M. (2003). *The Red Queen: Sex and the Evolution of Human Nature.* Harper Perennial; Reprint edition.

ROSENTHAL, J. S. (2006). *Struck by Lightning:: The Curious World of Probabilities*. National Academies Press.

RUMSFELD, D. H. *News Transcript* [online]. (2002) [cit. 2015-02-20] Available from: http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636

SAKS, A. M.; UGGERSLEV, K. L. (2010). Sequential and combined effects of recruitment information on applicant reactions. In: *Journal of Business and Psychology*, *25*(3), pp. 351-365.

SHIN, SUN-JOO (1994) *The logical status of diagrams*. Cambridge University Press.

SIMBA ENERGY, Market Highlights [online]. (2012) [cit. 2015-02-20] Available from: http://www.simbaenergy.ca/investors/market-highlights.aspx

SMITH, A.F.M.; GELFAND A.E. (1992) Bayesian Statistics without Tears: A Sampling–Resampling Perspective. In: *The American Statistician*. Vol. 46, Issue 2.

STANDING, C.; GUILFOYLE, A.; LIN, C., & LOVE, P. E. (2006). The attribution of success and failure in IT projects. In: *Industrial Management & Data Systems*,*106*(8), pp. 1148-1165.