ARCHIWUM FILOZOFII PRAWA
I FILOZOFII SPOŁECZNEJ
2015/1

Bartosz Janik, Paweł Banaś[1]
Jagiellonian University

Some Remarks on the Application of Bayesian Analysis in Law

Abstract

This paper discusses the use of Bayesian analysis in law. It introduces the key concepts of Bayesian analysis by giving some common examples of criminal cases. It focuses on the advantages of Bayesian analysis over some other probability interpretations (mainly the frequentist one). The last part of the text discusses the general notion of truth in legal proceedings and its possible interpretations within the probabilistic framework – given the Bayesian subjectivists-objectivists discussion.

1. Introduction

1.1. General overview

To reason under uncertain conditions, we should use statistical analysis. To choose the proper statistical calculi (i.e. certain way of understanding statistics[2]), we should answer some basic questions about the characteristics of those uncertain conditions. In this paper, we argue that Bayesian analysis is a proper way of understanding the statistics used for reasoning in conditions set in the courtroom.[3] Firstly, we introduce Bayesian analysis and explain how it should be interpreted. Secondly, we elaborate on the advantages of Bayesian approach that should be, as we believe, of particular interest to lawyers. Thirdly, we discuss some philosophical issues concerning Bayesian analysis by considering the problem of truth in legal proceedings. In the article examples are given in order to show how Bayesian analysis may be applied in real-life situations.

---

[1] This paper is a part of a project funded by National Science Centre of Poland (UMO-2012/07/N/HS5/00999).

[2] The main problem is to understand the slight difference between frequentist statistical inference and Bayesian inference. The former tells us that p(A) is long-run frequency with which A occurs in identical and ideal repeats of an experiment and the latter tells us that p(A|B) is a real number measure of plausibility conditional on truth of the information in B. In the frequentist view, we are operating on a set of random variables, in Bayesian inference we could use any logical proposition (P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support*, Cambridge University Press 2005).

[3] See E. Heit, *A Bayesian analysis of some forms of inductive reasoning*, in: M. Oaksford, N.Chater (eds), *Rational models of cognition*, Oxford University Press 1998, pp. 248–274.

1.2. Probability

Generally speaking, there are two main interpretations of probability: physical and evidential.[4] The former perceives probability as a "chance", i.e. events are predictable given all sufficient information. Speaking more vividly, frequentists say that the probability of getting a heads is 1/2, not because there are two equally likely outcomes but because repeated series of large numbers of trials demonstrate that the empirical frequency converges to the limit 1/2 as the number of trials goes to infinity. Within this approach we see probability as some kind of truth about the world – that the world is so that in an infinite number of coin tossing trials half of them will be heads and half of them will be tails. The problem with this interpretation is that in reality it is impossible to perform an infinite number of trials so as to "discover" the "true" probability ("chance") of the event to happen. Evidential (or "Bayesian") interpretation of probability is, on the other hand, the way to represent the degree to which a statement is plausible given the available evidence.

Within Bayesian framework a number of interpretations exist. Among them: the logical (e.g. Keynes) and the epistemic/inductive (e.g. Ramsey). The former presupposes that there is an "objective" (logical) relation between knowledge and probabilities. Keynes in his *Treatise on Probability* defended a position where two people having exactly the same knowledge of Bayesian analysis would hold the same belief. Ramsey, on the other hand, argued that this is not the case. We will come back to this problem in the last part of this article.

2. Bayesian analysis

2.1. Introduction to simple analysis

In this part of the paper, we give a formal introduction to simple Bayesian analysis. We shall use the term "simple" to make a difference between some basic properties of the Bayes formula and its applications and advanced Bayesian model selection. Simple Bayesian analysis serves as an introduction to more advanced statistics while still offering a great amount of tools for working with beliefs. We think that simple analysis will be sufficient for legal argumentation and reasoning in the courtroom (at least as far as the discussed examples are considered). It is, however, important to keep in mind that more sophisticated, "full-blooded", statistical models for Bayesian data analysis do exists if one needs them.

The most important compound of Bayesian analysis is the Bayes formula. To formally introduce the Bayes formula, we should use some advanced results from analysis and algebraic logic. For now, it will be sufficient to say that simple Bayesian analysis is an enrichment of propositional logic. The basic formulas which enrich the basic logical calculus are: sum rule and product rule.

Product rule and sum rule are the simple algebraic formulas, telling us how to manipulate probabilities:

$$p(A|B) + p(\overline{A}|B) = 1$$

$$p(A,B|C) = p(A|C)p(B|A,C) = p(B|C)p(A|B,C)$$

[4] For more interpretations and different problems regarding them, see e.g. D. Gilles, *Philosophical Theories of Probability*, Routledge 2000, and W. Załuski, *Skłonnościowa Interpretacja Prawdopodobieństwa*, [Eng. Propensity Interpretation of Probability] OBI-Biblos 2008.

Changing notation to more convenient, we could introduce the Bayes formula as a consequence of the product rule and sum rule:

$$p(H_i|D,I) = \frac{p(H_i|I)p(D|H_i, I)}{p(D|I)}$$

The Bayes formula tells us that the probability of obtaining a particular hypothesis (represented by proposition H, asserting the truth of that hypothesis) from given data and background information ($p(H|D,I)$ – posterior probability of H_i) is a result of dividing the product of probability of obtaining this hypothesis with background information and probability of obtaining those data from background information and particular hypothesis settled true ($p(H_i|I)$ – prior probability and $p(D|H_i,I)$ – likelihood function $L(H_i)$) by probability of obtaining data from background information alone ($p(D|I)$ – normalization factor, ensuring that $\Sigma p(H_i|D,I) = 1$). We could skip the "background information" compound, but it's far more convincing to introduce the full model with background information, hypotheses and obtained data just as in classical philosophy of science. To elaborate on this topic, we should say that the Bayes rule gives us probability of posterior hypotheses on the basis of acquired data and prior information. We are using the discrete hypothesis space.

2.2. Example 1

Before dealing with the logical foundation of Bayesian analysis and introducing more advanced concepts, we should present a simple example of how Bayesian analysis is used to deal with a simple misunderstanding of statistical data interpretation. We should first set up the following question. Given the information about the incidence of disease and the information concerning possible errors in data reporting of a particular test, assuming that the test came back positive, what is the probability of having the disease? The obvious answer is that the ratio of true positive results is the probability of having the disease. So, when the probability of true positive results is set to be 80% and the test came back positive, then it is 80% probable that you have the disease. This answer is wrong, because it ignores crucial information concerning the incidence of the disease. The simple frequentist analysis is able to give us the sole data about the sensitivity of the test and the incidence of the disease, but fails to merge those two in the single case (with no statistical meaning).

To show how Bayesian analysis as a merging factor works, we should consider the following example.[5] The probability of a false positive result of the test is estimated to be 2.3% and the probability of a false negative result of the test is estimated to be 1.4%[6]. From the statistical data available, we know that the incidence of the disease is 1:10000. The last information is in fact a frequentist one. In fact, data interpreted in a classical way are very useful under certain conditions. The question will remain the

[5] Taken from P. Gregory, *Bayesian*….

[6] To be precise about terminology, we should talk about sensitivity of the test, i.e. the ratio of true positive results $p(T+|H+)$ and specificity of the test i.e. the ratio of true negative results $p(T-|H-)$. The provided percentage data stand for $p(T+|H-) = 2.3\%$ and for $p(T-|H+) = 1.4\%$. For the sake of simplicity, we offer a slightly different notation. We will employ more precise notation in the next example.

same, given the data and the information that the test came back positive, what is the probability of having a disease?

To represent the data, we will use the following sentences:

H = "You have the disease."
$\overline{H}$ = "You don't have the disease."
D_1 = "Positive test."
I_1 = "No known cause for the disease."
$p(D_1|H, I_1) = 0.986$
$p(D_1|\overline{H}, I_1) = 0.023$

From the observation that the normalizing factor in the Bayes formula could be expressed as the sum of all possible hypotheses and having only two hypotheses, we could say that:

$$p(D_1|I_1) = p(H|H, I_1) + \mathrm{p}(\overline{H}|I_1)p(D_1|\,\overline{H, I_1})$$

In this case, the Bayes formula could be expressed as follows:

$$p(H|D_1,I_1) = \frac{p(H|I_1)p(D_1|H, I_1)}{p(H|I_1)p(D_1|H, I_1) + p(\overline{H}|I_1)p(D_1|\overline{H}, I_1)}$$

and noticing that p(H|I$_1$) signifies incidence of the disease, we could write p(H|I$_1$) = 1-p(H|I$_1$). From the given numerical values, we obtain:

$$p(H|D_1, I_1) = 0.0042$$

So, the probability that you have the disease, based on no known cause and positive result of the test is 0.4% which is quite unintuitive.

2.3. Parameter estimation and model selection

The formal derivation of the Bayes formula from propositional logic is due to Edwin T. Jaynes.[7] In the late 1980s he developed a program of probability as logic of science by showing that probability could be founded on logic. The proof of the theorem stating that probability calculus could be derived from propositional logic is quite interesting and founded on Jaynes', C.E. Shannon's and T. Cox's works.[8] Summarizing the whole point of Jaynes' works, we could say that probability is supposed to represent a rational

[7] We are aware that T. Jaynes' interpretation of Bayesian analysis yields some conceptual problems. The various interpretations of Bayesian analysis could be found in A. Gelman, C.R. Shalizi, *Philosophy and the practice of Bayesian statistics,* British Journal of Mathematical and Statistical Psychology 2000/66(1), pp. 8-38. We are not orthodox in using T. Jaynes approach. Every point of interpretation in our analysis will be explained.

[8] See E.T. Jaynes, *Bayesian Methods: General Background. An Introductory tutorial*, in: J.H. Justice (ed.), *Maximum Entropy and Bayesian Methods in Applied Statistics*, Cambridge University Press 1985, pp. 1–25; E.T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press 2003; C.E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal, 1948/27, pp. 379–423,623–656; R.T. Cox, *Probability, Frequency and Reasonable Expectation*, American Journal of Physics, 1946/1, pp. 1–13.

system of beliefs.[9] The main task of Jaynes' program is to show that probability could serve as the logic of science. The fundamental principles of Bayesian analysis, introduced by Jaynes, could be formulated as follows:[10]

1. Degrees of plausibility are represented by real numbers.
2. The measure of plausibility must exhibit qualitative agreement with rationality. This means that as new information supporting the truth of a proposition is supplied, the number which represents the plausibility will increase continuously and monotonically. Also, to maintain rationality, the deductive limit must be obtained where appropriate.
3. Consistency:
   - structural consistency: if a conclusion can be reasoned out in more than one way, every possible way must lead to the same result;
   - propriety: the theory must take account of all information, provided it is relevant to the question;
   - Jaynes' consistency: equivalent states of knowledge must be represented by equivalent plausibility assignments.

More advanced Bayesian analysis could be introduced by adding some formal mechanisms dealing with continuous probability distribution and then hypotheses should be estimated using the probability density function.[11] On that basis we could introduce a mechanism of marginalization which deals with free parameters of the theory and which could be introduced into discrete and continuous sample spaces. In this case, our hypothesis will be the value of estimated parameter. We are permitted to use any of the statistical tools that we know. For Bayesian analysis we could use marginalization. Marginalization will be simply an integration with regard to a certain parameter over the hypothesis space. Dealing with the certain parameter we will acquire a model with desired free parameters.

The second issue, connected with parameters and probability distribution is the necessity to decide how to distribute the prior probability. One of the many mechanisms which could be used is the principle of maximum entropy, introduced by C.E. Shannon in his communication theory.

$$H = \sum_{i=1}^{N} P(i|I) \log P(i|I)$$

The maximum entropy principle tells us that when we have some useful information, we could ascribe the probability density function simply by estimating its entropy and maximizing it according to this information constrains. Estimating parameters could also deal with the problem of measuring the intervals and estimating parameters in intervals of a particular width (simple or logarithmic).

9 See e.g. G.L. Bretthorst, *An Introduction To Parameter Estimation Using Bayesian Probability Theory,* in: P.F. Fourgere (ed.), *Maximum Entropy and Bayesian Methods*, The Netherlands: Kluwer Academic Publishers 1990, pp. 53–79.

10 See both E.T. Jaynes, *Probability…*, p. 24, and P. Gregory, *Bayesian…*, p. 30.

11 The important thing is to notice an interpretation. Bayesian probability density function is a measure of our state of knowledge of the value of the parameter, P. Gregory, *Bayesian…*, p. 7.

The last interesting point is to introduce an objective method for choosing the better model.[12] In a situation where we have two or more models, we are forced to choose one of them. The problem of choosing one model over another is in fact the problem of choosing one hypothesis over another. Each model is build around the particular hypothesis that should explain the situation of obtaining particular data. Bayesian analysis provides a mechanism called the odds ratio which introduces a method for comparing different models. In Bayesian analysis, usually simpler models are favored. The mentioned ratio could be computed with the use of the following formula (for simple models):

$$O_{ij} = \frac{p(M_i|I)p(D|M_i, I)}{p(M_j|I)p(D|M_j, I)} = \frac{p(M_i|I)}{p(M_j|I)} B_{ij}$$

In the odds equation we have *a priori* information and the Bayes factor which is simply the ratio of likelihood for both models. The problem arises when we want to deal with more complicated models with dozens of parameters and different characteristics. The universal formula is to compute the global likelihood and to compare it across different models. This subject will not be pursued here. To avoid problems with large Bayesian models, we could introduce Bayesian networks that will be more suitable in simple cases for inference based on certain propositional data. The core of Bayesian network is to connect simple hypotheses and represent its structure as an acyclic graph. The reasoning in a simple Bayesian network (i.e. without proper statistical inference) is based on external probabilities of hypotheses and simple calculation within the network.

2.4. Example 2

To show how intuitive and simple Bayesian analysis can be, we will present an example concerning reasoning as a process. Suppose that you are a doctor and you have a patient Adam who is complaining about a headache and nausea. You want to examine the hypothesis that your patient has a very rare fatal disease H+ with the incidence of 2% (we are not discussing the causal relation between headache, nausea and the disease). You order the first test (with $p(T+|H+) = 90\%$ and $p(T-|H-) = 95\%$) which came back positive. We can count the overall probability of H+ by using the single Bayes formula:

$$p(H+|T+) = \frac{p(T+|H+)p(H+)}{p(T+|H+)p(H+) + p(T+|H-)p(H-)} = \frac{0.90 \cdot 0.02}{0.90 \cdot 0.02 + 0.05 \cdot 0.98} = 0.269 = 27\%$$

Despite the change of probability, you order another test (with $p(T-|H-) = 98\%$ and $p(T+|H+) = 95\%$) and this one came back positive, too. We add this to the already computed $p(H+) = 0.27$.

$$p(H+|T+) = \frac{p(T+|H+)p(H+)}{p(T+|H+)p(H+) + p(T+|H-)p(H-)} = \frac{0.95 \cdot 0.269}{0.95 \cdot 0.269 + 0.02 \cdot 0.731} = 0.9449 = 94.5\%$$

[12] For more information, see e.g. the already mentioned P. Gregory, *Bayesian…*, or E.T. Jaynes, *Bayesian…*.

The whole procedure gives us a conclusive outcome, which is – Adam, unfortunately, has the disease and is going to die. A very important thing to notice is that there are situations (as we will argue later, quite common in the courts) which lack empirically verified and specific tests to falsify or corroborate hypotheses connected with them.

3. Advantages of Bayesian analysis

3.1. Limits of frequentist analysis

Having introduced the Bayes formula, we can now discuss why Bayes analysis should actually be of particular interest to lawyers.[13] The usual type of questions lawyers must face in the courtroom is how probable it is, given all the available evidence, that X happened. Within the frequentist approach, the questions would be interpreted as follows: in how many identical situations, out of an infinite number, with exactly the same evidence available, the X would happen. The obvious problem with this approach, however, is that usually most (if not all) of the available evidence is of non-repetitive nature.[14] This means that it is impossible to "test" the hypothesis in a frequentist manner – all we have is one trial only, a once-in-a-lifetime phenomenon that cannot be fully reproduced.[15]

Let us consider a case where a judge must decide whether pre-trial detention is justified. Article 258 Polish Code of Criminal Procedure states that such an action would be appropriate if e.g. there is a justified concern that the suspect may try to escape. A frequentist could suggest that a judge should consider in how many similar situations (out of a statistically adequate number) a suspect has tried to escape. However, there are no such statistics available, and even if there are "similar situations", they are usually not even close to "identical".[16]

For Bayesian analysis it is perfectly plausible to answer questions about the probability of future outcomes given non-repetitive evidence. Considering the aforementioned example with pre-trial detention, let us say that the suspect, Adam, is male, studies philosophy and has already graduated from a law school. We can now imagine that in the history of this particular court, among all the male suspects 10% tried to escape and the same is true of 90% of all the philosophy student suspects and 80% of all the law school graduate suspects. At the same time, Adam is the second suspect ever who is all three at the same time (the last one did try to escape). Moreover, in the history of the court there have been 10,000 male suspects, 50 philosophy student suspects and 200 law graduate suspects.

It seems that with these data available the court could use the frequentist approach and the decision would be that it is quite likely that Adam will try to escape. For an orthodox frequentist, however, it is now impossible to measure probability based on all the evidence available. There was only one situation "identical" with that of Adam (being

[13] The aim of this short article is not to discuss the admissibility of Bayesian reasoning in a court setting. Instead we would like to present its general idea and show why it should be of interest for lawyers.

[14] In Polish criminal procedure there is an institution of a "court experiment". It is debatable whether it is methodologically appropriate to use a court experiment in order to measure frequentist probability.

[15] For a discussion of some practical differences between the application of Bayesian vs. frequentists calculi, see Z. Dienes, *Bayesian Versus Orthodox Statistics: Which Side Are You On?*, Perspectives on Psychological Science, 2011/4, pp. 274–290.

[16] There are tools that frequentists may adopt in order to "predict" future outcomes – which are often used in the so-called "risk assessment". Within this approach the correlation is measured between the appearance of risk factors and the situation in question.

a male, a philosophy student and a law school graduate at the same time) which is too little to build a satisfactory model. What a frequentist can now do is to see how a given risk factor (e.g. being a philosophy student) mediates another (e.g. being a male). They can then combine all the factors and decide that the simple fact that Adam has all three characteristics is of little importance given that being a male philosophy student, a male law school graduate and philosophy student who is a law school graduate – all these independently still make Adam quite likely to escape. We will come back to this court decision later.

Let us now imagine that the court finds out that the suspect also suffers from a serious disease that requires him to visit a local, highly specialized hospital ward every second day, which makes the escape much more unlikely. This is the first time that the court has been faced with such a situation. It is now impossible for a frequentist to incorporate it into the rest of the evidence available. It is also impossible to say that this information is of little importance for the court.

3.2. Bayesian analysis and non-repetitive trials

Bayesian analysis does not require huge amounts of data to be conducted. A single non-repetitive piece of evidence is as good as any statistical data. When a judge considers a factor like the suspect's disease, he/she takes into consideration that it is highly unlikely that the suspect will find medical assistance elsewhere and that he could not survive without medical care. For the Bayes formula this "high unlikelihood" must be, however, quantified, i.e. the judge may acknowledge that there is only one hospital providing the necessary help in the country and the mortality rate of the disease (without proper medical care provided) is near 100%. Therefore, the "high unlikelihood" will be assigned the weight $P(X) = 0{:}95$. The problem is that not always can a judge easily quantify the weights of the more or less subjective "likelihood".[17]

3.3. Bayes and the problem of objectivity

An important and frequently asked question is how Bayesian analysis can be objective or what it means to say that Bayesian analysis is objective.[18] Of course, saying that any data analysis is objective can be a bit misleading. Every data analyst must be involved in a set of subjective choices, e.g. selecting a model, interpreting data and choosing them for a particular analysis. Therefore, we define objectivity here, by saying that within Bayesian approach probability objectively measures the plausibility of propositions. For subjectivists probability corresponds to a "personal belief" (see the dispute between Keynes and Ramsey mentioned in the introduction), but rationality and coherence constrain the probabilities a subject may have, however, still allowing for a substantial variation within those constraints. The objective (non-informative, default) and

---

[17] For a reference, see e.g. E. Cheng, *Reconceptualizing the burden of proof*, Yale Law Journal, 2013/122, pp. 1254–1279.

[18] Bayesian analysis is sometimes criticized for being empty, i.e. sometimes the prior probability cannot be properly ascribed and thus the starting point cannot be rationalized. We think that this argumentation is only partially acceptable. The focal point of this critique is due to the lack of restraint in the process of selecting the priors and the lack of the threshold for the acceptance of the hypothesis. In legal settings, however, the threshold is set by legal standards and the priors are partially governed by the admissibility of evidence. For more information, see M. Albert, *Bayesian Rationality and Decision making: A Critical Review*, Analyse & Kritik, 2003/25, pp. 101–117 and M. Albert, *Why Bayesian Rationality Is Empty, Perfect Rationality Doesn't Exist, Ecological Rationality Is Too Simple, and Critical Rationality Does the Job*, in: M. Baurmann, B. Lahno (eds), *Perspectives in Moral Science*, RMM Volume 0 2009, pp. 49–65.

subjective (like Ramsey's approach) variants of Bayesian probability differ mainly in their interpretation and construction of the prior probability.[19]

Prior probability distribution, often called simply the "prior", of an uncertain quantity x (like a proportion of suspects who will try to escape in the future) is the probability distribution that would express one's uncertainty about x before the data can be actually gathered. It is meant to attribute uncertainty rather than randomness (or physical "chance"). This means that the prior "measures" how uncertain we are rather than the frequentist probability that the quantity in question will be x.

The use of Bayesian probability involves specifying a prior probability. Prior probability may be obtained in many ways. We may know e.g. that recent studies show that 10% of all the suspects try to escape. This is a perfect situation where we have the data which are "physical-like" probability of the event in question where our uncertainty can mirror this "physical-like" probability: we are perfectly sure that out of 10,000 suspects 10% will escape. Statistical data are, however, not the only possible source of the prior. This may be obtained through the consideration of whether the required prior probability is greater or lesser than the reference probability associated with a thought experiment (e.g. having no previous statistical knowledge, we assume that it is very possible that about 10% of suspects will escape [reference prior] and we can be much less sure that it will be 50% who will try to escape [required prior]). The obvious problem here is that for a given problem, multiple thought experiments could be easily suggested and different people may disagree which is the best and which prior to assign (the problem is known as the reference class problem). To illustrate it consider the problem presented by Pierre-Simone Laplace: what is the probability that the sun will rise tomorrow?[20] It turns out that the plausibility that the sun will rise tomorrow increases with the number of days on which the sun has risen so far. In other words, the Bayesian probability that the sun will rise tomorrow will differ if one considers the horizon of one person, of humanity and of the earth as a planet. This also means the if we consider two persons – one living a century earlier from the other, both considering the earth horizon – for the latter it will be more probable that the sun will rise tomorrow (even if that person lives much closer to the scientifically predicted end of our planet). The conclusion is that Bayesian probability is a conditional probability given what one knows and may vary from one person to another.

So how can Bayesian analysis be considered objective? Let us consider two possible solutions. The first one is connected with using uninformative priors. The term "uninformative (objective) prior" means only that this is a prior that has not been elicited subjectively. This, however, is problematic and some claim that, in fact, no such things as objective priors exist. In legal situations it would probably be impossible to look for an objective prior for every case in question. The other way of understanding Bayesian analysis as objective (as far as legal proceedings are concerned) stands on the idea that Bayesian analysis presents the decision-making process in an elegant, formal manner. After agreeing on certain prior weights the decision must be drawn in accordance with the Bayes formula and will always be the same. The defendant, who is not happy with the court's decision, may either look for some new evidence or prove that the prior weight of some evidence in question is implausible (and justify another one, using e.g.

---

[19] The problem of priors is frequently discussed in the literature on probability. For a reference, see e.g. A. Gelman, C.R. Shalizi, *Philosophy…* or, for a possibly more "approachable" source, the already mentioned Z. Dienes, *Bayesian…*.

[20] E.T. Jaynes, *Probability…*, pp. 387–391 and D. Howie, *Interpreting Probability: Controversies And Developments In The Early Twentieth Century*, Cambridge University Press 2002.

Bayes formula again). It is, hence, not objective in a sense that there can be only one possible interpretation of the evidence available but because one set of data (evidence with its priors) will always lead to the same conclusion.

4. Bayes and "truth" beyond reasonable doubt

As pointed out in section 1.2 above, Bayesian analysis does not tell us any truth about the world in the way frequentist probability is supposed to. Another way to put it is to say that Bayes formula does not tell us how probable (in the "physical" meaning of the word) it is that the thing will happen – but how plausible it is. One could say that such a notion of plausibility is of little use in legal settings, where we should rather look for an "objective" or "material" truth: the question for a court should be not whether it is plausible to think that a suspect will try to escape but what the "real" chances are that he/she will do it. This is, however, a theory. Revealing an "objective" or "material" truth as the aim of legal proceedings seems like a reasonable demand put on a legal system.[21] But due to epistemological restraints, it is a matter of a well-justified decision (of a judge) whether something is or is accepted as true in the context of a given case. Even if all the evidence is "objective" (e.g. consisting of scientifically tested frequentist probabilities) the court must always make a decision. In section 3.1 we considered a situation where the court was able to run a risk factor analysis which employed the frequentist notion of probability. Still, however, on that basis the court had to decide, firstly, which data were insignificant and, secondly, what will be the most possible outcome. Any lawyer cannot know for sure what the future outcome will be or what happened in the past if he learns about it from indirect evidence. He must, however, make a decision, which requires him to choose the most plausible option – and Bayes formula is a tool that he will find very helpful.

Bayesian analysis is a mean of formalizing the decision-making process and presenting it in a clear and easy-to-follow way. It shows exactly why, given all the evidence, the court must have decided in a certain way. If one disagrees with the eventual results, she may either provide additional data or discuss the weight of any of the priors. The defendant who tries to prove that something did not happen, must in fact convince the judge that claiming that the thing in question did happen is implausible. While employing Bayesian analysis, he will know where to direct his efforts.

5. Summary

In this paper we have presented the Bayes formula as a tool for testing the evidential ("subjective") probability, i.e. the plausibility of a hypothesis given the available evidence. We argued that Bayesian analysis should be of particular interest for lawyers as it helps to assess the probability even if the available evidence is of non-repetitive nature. We also discussed how Bayesian analysis can be considered "objective": we claimed that its "objectivity" lies in the fact that for anyone who runs the analysis with the same data, the result will be exactly the same. Still, however, they may disagree what the data should look like in this particular case. Finally, we defended the idea that employing Bayesian analysis in the courtroom does not contradict the notion of truth in legal procedures.

[21] We do not intend here to take any position in an ongoing debate concerning the advantages of an adversarial system of law as we see Bayesian reasoning applicable to systems with as well as without the "material truth" principle.

BIBLIOGRAFIA / REFERENCES:

Albert, M. (2003). Bayesian Rationality and Decision making: A Critical Review. *Analyse & Kritik 25*, 101–117.

Albert, M. (2009). Why Bayesian Rationality Is Empty, Perfect Rationality Doesn't Exist, Ecological Rationality Is Too Simple, and Critical Rationality Does the Job. In M. Baurmann, B. Lahno (Eds.), *Perspectives in Moral Science*. Frankfurt: Frankfurt School Verlag.

Bretthorst, G.L. (1990). An Introduction To Parameter Estimation Using Bayesian Probability Theory. In P.F. Fourgere (Ed.), *Maximum Entropy and Bayesian Methods*. Dordrecht: Kluwer Academic Publishing.

Cheng, E. (2013). Reconceptualizing the burden of proof. *Yale Law Journal 122/5*, 1254–1279.

Cox, R.T. (1946). Probability, Frequency and Reasonable Expectation. *American Journal of Physics 14/1*, 1–13.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science 6/3*, 274–290.

Gelman, A., Shalizi, C.R. (2000). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology 66/1*, 8–38.

Gilles, D.A. (2000). *Philosophical Theories of Probability*. London–New York: Routledge.

Gregory, P.C. (2005). *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support*. Cambridge: Cambridge University Press.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford, N.Chater (Eds.), *Rational models of cognition*. Oxford: Oxford University Press.

Howie, D. (2002). *Interpreting Probability: Controversies And Developments In The Early Twentieth Century*. Cambridge: Cambridge University Press.

Jaynes, E.T. (1985). Bayesian Methods: General Background. An Introductory tutorial. In J.H. Justice (Ed.), *Maximum Entropy and Bayesian Methods in Applied Statistics*. Cambridge: Cambridge University Press.

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

Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal 27*, 379–423.

Załuski, W. (2008). *Skłonnościowa Interpretacja Prawdopodobieństwa*. Kraków: OBI-Biblos.