ZAGADNIENIA
NAUKOZNAWSTWA
1–4 (215–218), 2018
PL ISSN 0044 – 1619

Jakub Mielczarek
Institute of Physics, Jagiellonian University
Marek Szydłowski
Astronomical Observatory, Jagiellonian University
Adam Krawiec
Institute of Economic, Finance and Management, Jagiellonian University
*Paweł Tambor**
Faculty of Philosophy, The John Paul II Catholic University of Lublin

Bayesian reasoning in cosmology

Abstract. We discuss epistemological and methodological aspects of the Bayesian approach in astrophysics and cosmology. The introduction to the Bayesian framework is given for a further discussion concerning the Bayesian inference in physics. The interplay between the modern cosmology, Bayesian statistics, and philosophy of science is presented. We consider paradoxes of confirmation, like Goodman's paradox, appearing in the Bayesian theory of confirmation. As in Goodman's paradox the Bayesian inference is susceptible to some epistemic limitations in the logic of induction. However, Goodman's paradox applied to cosmological hypotheses seems to be resolved due to the evolutionary character of cosmology and the accumulation of new empirical evidence. We argue that the Bayesian framework is useful in the context of falsifiability of quantum cosmological models, as well as contemporary dark energy and dark matter problem.

Key words: cosmology, bayesianism, logic of induction

Bayesowskie wnioskowanie w kosmologii

Abstrakt. W artykule poddano dyskusji epistemologiczne i metodologiczne aspekty bayesowskiego podejścia w praktyce badawczej astrofizyki i kosmologii. Dokonano najpierw ogólnego wprowadzenia do metodologii bayesowskiej, by następnie odnieść przedstawione narzędzia pojęciowe do fizyki. Przedstawiono wzajemne oddziaływanie między współczesną kosmologią, stystystyką bayesowską i filozofią nauki. W szczególności rozważono paradoksy konfirmacji, jak paradoks Goodmana, mające swoje odzwierciedlenie w bayesowskiej teorii konfirmacji. Podano kosmologiczną wersję paradoksu Goodmana i propozycję jego rozwiązania, biorąc pod uwagę specyfikę kosmologii (jej ewolucyjny charakter i jednostkowość przedmiotową – Wszechświat, który jest obiektem badania). Zaprezentowano argumenty za tym, że metodologia bayesowska jest użyteczna w kontekście problemu falsyfikowalności kwantowych modeli kosmologicznych oraz współczesnej dyskusji nad problemem ciemnej energii i ciemnej materii.

Słowa kluczowe: kosmologia współczesna, logika indukcji, statystyka bayesowska

1. Introduction

In everyday experience, even when we do not realize it, we use our intuition to draw inferences. For example, when we hear doorbell, we immediately ask ourselves "Who has come?". On the way to the door we consider many different

* For correspondence: ul. Jana Pawła II 7, 25-025 Kielce, e-mail: pawel.tambor@gmail.com

possibilities. Maybe it is our friend, who said some days before that he visits us. Maybe it is our neighbor, who came to say that we should lower the music or maybe somebody came to inform us that we have won 1,000,000 EURO in the lottery and so on and so on. All of these possibilities are more or less probable, with some outcomes which we cannot figure out *ex ante*. But every bit of information influences our expectations and we stick on outcome which seems to be most probable (subjectively) to us. So, we look through the window and see a car on the street, that looks like our friend's, we are almost sure that it is he in fact. Or if we really play music very loudly, we can expect that sooner or later somebody will be angry. When we do not expect somebody we know, we assume that it may be the postman or someone who got the wrong address. In fact, we always choose the simplest case. It means that we never assume that Superman is standing behind the door, even when there is a reason why he should visit us. This intuitive feeling to choose the simplest solution is called Occam's razor (Rodriguez-Fernández 1999) and it states formally: "Accept the simplest explanation that fits the data".

In our example the data means in fact what we already know. This example is very easy, and our faculties manage very well to solve problems of this kind. However, there is a vast variety of problems for which there are uncountable possibilities and such problems can be treated only in an approximated manner. On the other hand, there are problems, which are too tedious to be solved by humans in reasonable time and we should employ to do job. So, can we enclose this intuitive knowledge in the form of mathematically defined theory and use it instead of mind? The answer is "Yes", this exciting idea is embodied in the form of Bayesian inference (MacKay 2003).

Below we introduce Bayesian inference and show how it works in practice. We start from general considerations which lead us to the connection with thermodynamics. Subsequently we show how to write down problems on computer with use of the Monte-Carlo approach (the Metropolis algorithm). We describe possible applications in the modern cosmology. There is a huge number of places in cosmology where Bayesian inference can often be applied. We have residual observations and a lot of theories. Some of them are easy, some are pure models, some are brilliant new ideas. This is as with our example with the doorbell. We hear the bell and we must predict who is at the door. Without any information we cannot predict who is ringing, because the sound of the bell is always the same. However, some people ring only once and some of them more times. Therefore, we must listen very carefully when something is ringing in cosmology. In the last section we try to put Bayesian inference into a larger perspective: containing epistemological aspects of the method as well as suggested limitations.

2. Bayesian inference and thermodynamics

Our goal is to present a way to choose among alternative theories taking into account their conformity with data. Of course, these theories can have different basis. They can be connected with everyday experience, data analysis, biology, physics etc. Because we want to apply finally Bayesian inference in physics, we can restrict ourselves now, without loss of generality, to physical theories. So, let us consider some unknown physical phenomenon and let us say we possess K theories $\{H_1, \ldots, H_K\}$ that can potentially describe it. All these theories differ from one another. Information about phenomenon investigated is contained in the collection of experimental data D. Both theories and experimental data we may consider as elements of the same set space Ω so $\{H_1, \ldots, H_K, D\} \in \Omega$. The set Ω together with measure P and σ-algebra F build probability space Ω, F, P. In such a well-defined theory of probability, a natural concept of conditional probability occurs. So, the probability of a given theory H_i, when we have data D is defined as

$$P(H_i|D) = \frac{P(H_i \cap D)}{P(D)} \ . \tag{1}$$

This probability tells us which theory describes experimental data D better and is called posterior probability. On the other hand, we can ask about probability of outcomes D when theory H_i is the true one

$$P(D|H_i) = \frac{P(D \cap H_i)}{P(H_i)} \ . \tag{2}$$

This probability tells us about different predictions D from the theory H_i and is a marginal likelihood, commonly called evidence. Because $P(H_i \cap D) = P(D \cap H_i)$ we can combine equations (1) and (2) what give us:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{P(D)} \ . \tag{3}$$

This equation is the famous Bayes theorem. The probability $P(H_i)$ in this equation is called prior probability and it is in fact hard to describe this number. It describes our initial beliefs about a given theory. It is a human factor to choose this number and can be non-objective. Sometimes one theory is chosen because of its mathematical beauty although a more probable alternative exists. The other factor is that some theory can well describe variety of others similar phenomena. However, if we do not have strong motivation to introduce some initial selection of the theories, then the most natural choice is to assume a homogeneous distribution of the prior probability

$$P(H_i) = \frac{1}{K} \ . \tag{4}$$

Then none of theories is favored. The probability $P(D)$ is a simple normalization constant, what we can calculate thanks to the normalization condition:

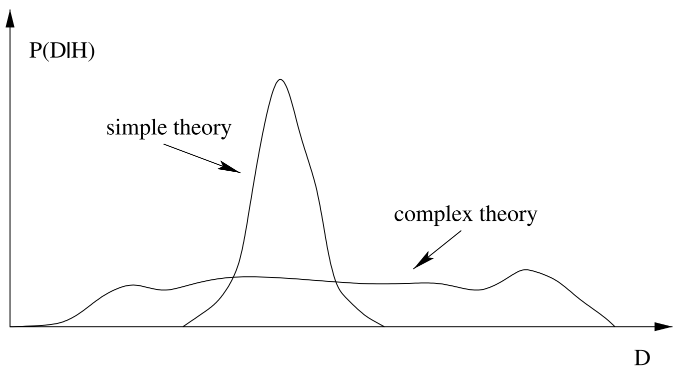
$$\sum_{i=1}^{K} P(H_i|D) = 1 \tag{5}$$

which together with the Bayes theorem (3) give us (Trotta 2008):

$$P(D) = \sum_{i=1}^{K} P(D|H_i)P(H_i) \quad . \tag{6}$$

Each of the theories $\{H_1, \ldots, H_K\}$ contains a number of parameters described by the vector θ_i for a particular theory. The simple theories (simple mathematically) contain in general a small number of parameters. The main increase of number of parameters enlarges the complexity of theories. This complexity can be in some cases accepted due to intrinsic beauty of a mathematical structure of theory. Nevertheless, the theory which has one factor to explain a phenomenon is preferable over the theory which employ many factors for description of it. Effective theories belong to the type of simple theories.

Figure 1. Evidence for simple and complex theories



$$P(D|H_i) = \int d\theta_i P(D|\theta_i, H_i)P(\theta_i|H_i). \tag{7}$$

Parameters θ_i of the model H_i are elements of the set space Ω. The values of these parameters may be fixed or significantly bounded by a theory. But when no limits are put on these parameters (there is no prior knowledge) the evidence is calculated as the marginal probability integrated over all the allowed range of values of the parameters of the model.

Now we can go back to the Bayes theorem and explain the idea of Bayesian inference. Considering the Bayes theorems for two models i and j and dividing the respective equations (3) by sides we obtain:

$$B_{ij} = \frac{P(D|H_i)}{P(D|H_j)} = \frac{P(H_i|D)P(H_j)}{P(H_j|D)P(H_i)}, \quad (8)$$

which is called the Bayes factor. If the priors $P(H_i)$ for all i are equal, then the Bayes factor reduces to the ratio of evidences $B_{ij} = P(H_i|D)/P(H_j|D)$. The values of B_{ij} can be interpreted as follows: if $0 < \ln B_{ij} < 1$ then inference is inconclusive, if $1 < \ln B_{ij} < 2{,}5$ we have weak, if $2{,}5 < \ln B_{ij} < 5$ we have moderate and if $5 < \ln B_{ij}$ we have strong evidence in favor of a model indexed by i over the model indexed by j (Jeffreys 1961). So, the main problem is now to calculate the evidence. The direct calculation is generally impossible. That is the reason is to use the Monte Carlo methods to do it. First, we introduce the parameter λ and redefine evidence to the form:

$$P(D|H_i)_\lambda = \int d\pi_i P^\lambda(D|\theta_i, H_i) \quad (9)$$

where

$$d\pi_i = d\theta_i P(\theta_i, H_i). \quad (10)$$

The $P(D|\theta_i, H_i)$ is in fact likelihood and we denote it by L. So

$$\frac{d \log P(D|H_i)_\lambda}{d\,\lambda} = \frac{\int d\pi_i L^\lambda \log L}{\int d\pi_i L^\lambda} = \langle \log L \rangle_\lambda \quad (11)$$

and

$$P(D|H_i) = exp\left[\int_0^1 d\lambda \frac{d \log P(D|H_i)_\lambda}{d\lambda}\right] = exp \int_0^1 d\lambda \langle \log L \rangle_\lambda. \quad (12)$$

Now we can show a connection between our approach and thermodynamics. Introducing:

$$E = -\log L \quad (13)$$

$$\frac{1}{T} = \lambda \quad (14)$$

$$Z = \int d\pi L^\lambda \quad (15)$$

we obtain

$$Z = \int d\pi e^{-E/T} \quad (16)$$

and equation (11) takes a known form:

$$\langle E \rangle_T = \frac{\int d\pi E e^{-E/T}}{\int d\pi e^{-E/T}} \ . \tag{17}$$

This is the energy of the system in the temperature T. When we calculate it for different temperatures we can directly evaluate the integral in the expression (12) and hence the evidence. As we see, to perform the Bayesian inference we have to calculate the thermodynamical integral (17). This kind of integrals can be solved analytically only in case of very simple systems. Numerical methods to solve this kind of problems are known as the Monte Carlo. It is not the subject of this paper to describe how they work in detail. However, to make this paper self-contained we add a short Appendix A introducing basics of the Monte Carlo methods. We also present experimental demonstration of property of ergodicity which is important in the context of Monte Carlo simulation (see Appendix B). An interested reader can find more on Monte Carlo simulations e.g. in (MacKay 2003).

3. A simple example

Now we have all theoretical equipment to show this approach in action. In this example we show how to perform the Bayesian inference in a very simple case. We consider a very simple kind of theories and a small sample of data-points to make computer computation short. We also design it for clarity and better understanding. However, generalizations to more advanced problems are straightforward. In the next section we will mention how to apply Bayesian methods to more complicated problems. Let us consider some experiment in which we perform measurements of some physical variable y for six different values of parameter x. In the experiment we also measure standard error of the outcomes y. In fact, we one can repeat many times measurements of y for a given x value. Then one can obtain the mean values of parameter y together with its dispersion. These data points we present in Table 1.

Table 1. In the table we collect the exemplary pairs (x, y) together with the uncertainty of y. The uncertainty Δy can be the result of the instrumental resolution.

x	y	Δy
1	1	7
2	3	3
3	5	4
4	7	6
5	10	3
6	15	1

The phenomenon which we instigate is still not undetermined, but we possess three polynomial models to describe them. We list these models below:

Model 1: $$y_1(x) = \alpha_0 + \alpha_1 x \tag{18}$$

Model 2: $$y_2(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 \tag{19}$$

Model 3: $$y_3(x) = \alpha_0 + \alpha_2 x^2 \tag{20}$$

Models 1 and 3 look simpler because each is described by two parameters when model 2 contains three unknown parameters.

The first step of Bayesian inference is to fit these models to experimental data. We can use for example method of least squares. We obtain:

Model 1: $$\hat{\alpha}_0 = -2{,}47 \pm 1{,}07;\ \hat{\alpha}_1 = 2{,}66 \pm 0{,}27, \quad \text{SSE} = 1{,}148;\ \bar{R}^2 = 0{,}949; \tag{21}$$

Model 2: $$\hat{\alpha}_0 = 0{,}7 \pm 1{,}02;\ \hat{\alpha}_1 = 0{,}28 \pm 0{,}67;\ \hat{\alpha}_2 = 0{,}34 \pm 0{,}09, \quad \text{SSE} = 0{,}571;\ \bar{R}^2 = 0{,}987; \tag{22}$$

Model 3: $$\hat{\alpha}_0 = 1{,}10 \pm 0{,}33,\ \hat{\alpha}_2 = 0{,}38 \pm 0{,}02, \quad \text{SSE} = 0{,}509;\ \bar{R}^2 = 0{,}990 \tag{23}$$

where $\hat{\alpha}_i$ are estimated values of parameters α with their errors; the sum of squared errors (SSE) and the adjusted coefficient of determination $\bar{R}^2$ are given for each model.

The important ingredient in the Bayesian inference is a choice of priors. It is the choice of the intervals and probability distribution for parameters. The parameter intervals should be specified, because we must perform the integration (look for the solution) in a finite parameter space. Standard errors of the parameters give us intervals necessary for Bayesian inference. Of course, the different choices of the parameter intervals can lead to the different values of the posterior probabilities. In our case, we choose the intervals as the $1 - \sigma$ confidence interval for the parameters. Namely, $[\alpha - \sigma, \alpha + \sigma]$, where α is the best fit value and σ is the standard error. Next, we assume that outcomes are from the Gaussian distribution and the likelihood function has then a form:

$$L \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{N} \frac{(y(x_i)-y_i)^2}{\sigma_i^2}\right] . \tag{24}$$

Now we apply theory described in the previous section. With use of the Metropolis algorithm we calculate energies $\langle E\rangle_\lambda$ for models considered. We perform calculations for the values of α from the range (0,1) as it is necessary to calculate the integral in (12). We show these results in Fig. 2, 3, 4.

Now with use of this data we can perform integrals in the form

$$I_i = \int_0^1 d\lambda \langle E\rangle_\lambda \tag{25}$$

From equation (12) we see that:

$$P(D|H_i) = exp \int_0^1 d\lambda \, \langle \log L\rangle_\lambda = e^{-I_i} \tag{26}$$

Obtained values for three models considered are:

$$I_1 = 5{,}8;\ I_2 = 7{,}6;\ I_3 = 4{,}8 \tag{27}$$

Now we can directly calculate Bayes factors:

$$B_{12} = 6{,}06;\ \ln B_{12} = 1{,}80 \tag{28}$$

$$B_{32} = 16{,}67;\ \ln B_{32} = 2{,}81 \tag{29}$$

$$B_{31} = 2{,}78;\ \ln B_{31} = 1{,}01 \tag{30}$$

Based on the obtained values of $\ln B_{ij}$ one can conclude:

- Because $1 < \ln B_{12} < 2{,}5$, we have weak evidence that the linear model is favored over the three-parametric quadratic model.
- Because $2{,}5 < \ln B_{32} < 5$, the two parametric quadratic model is moderately preferred over the the three-parametric quadratic model.
- Because $\ln B_{31} \approx 1$ there is no preferences between the two-parametric models, linear and quadratic.

We can also directly calculate posterior probabilities:

$$P(H_1|D) = 0{,}26 \tag{31}$$

$$P(H_2|D) = 0{,}04 \tag{32}$$

$$P(H_3|D) = 0{,}70 \tag{33}$$

Figure 2. $\langle E \rangle_\lambda$ dependence for the first model

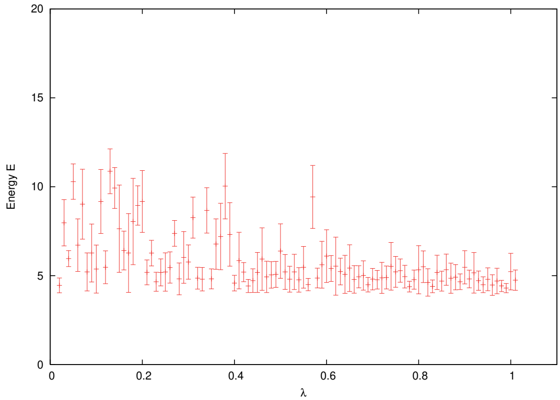


Figure 3. $\langle E \rangle_\lambda$ dependence for the second model

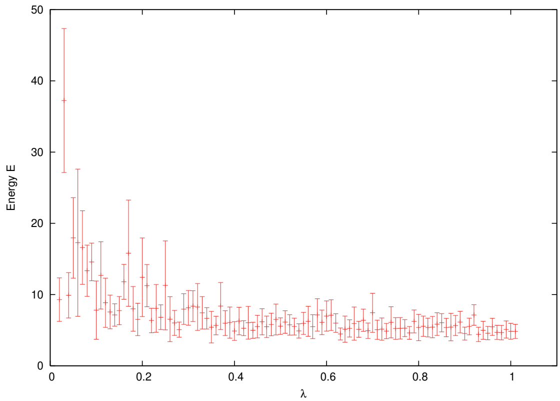
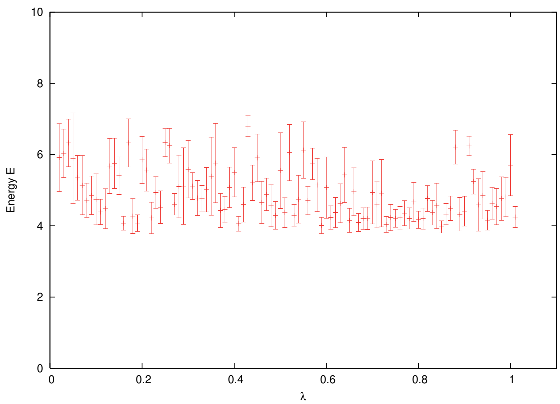


Figure 4. $\langle E \rangle_\lambda$ dependence for the third model



What we see is that the posterior probability suggest that the last model explains the experimental data in the best way. Even that it may look more complicated than linear model number one. The second model can be discarded. This model contains remaining models inside and naturally fits better to the experimental data because has more degrees of freedom. But comparing to the other models he is too complicated and not necessarily properly explains experimental data. We see that however first and last model possess the same number of degrees of freedom the last function seems to explain in the better way the nature of the investigated physical process. On the other hand, from the Bayes factor criterion, one cannot see such a preference. It is because the borders in the Bayes factor criterion were established empirically in the very careful manner.

The example presented here illustrates that the different criteria give us stronger or weaker conclusions. In particular, note that posterior probability criterion favors the model H_3 over the model H_1. On the other hand, the Bayes factor criterion indicates that there is no preference between the models H_3 and H_1. Therefore, this simple example gives us here the first lesson to be very careful interpreting statistical inference results.

The second lesson is to be careful about results of estimations. The results presented above, based on the $1 - \sigma$ interval choice of the parameters intervals. In case of the $3 - \sigma$ intervals the resulting Bayes factors are the following

$$\ln B_{12} = 0{,}4 \tag{34}$$

$$\ln B_{23} = 1{,}5 \tag{35}$$

$$\ln B_{13} = 1{,}9 \tag{36}$$

Therefore, there the first model is weakly favored with respect to the third one. Also, the second model is weakly favored with respect to the third one. The first and the second models are indistinguishable. The $3 - \sigma$ results differ from these performed previously.

The comparison between the $1 - \sigma$ and $3 - \sigma$ examples was performed to show explicitly the sensitivity of the Bayesian inference on the parameter intervals. This dependence on the priors becomes significant in case of the *weak data*, e.g. for the small sample size or significant error. For the *strong data* the impact from the priors becomes irrelevant. In the case considered here, the data sample is small, and the significant errors are present. Because of this, the inference is sensitive on the choice of the priors as observed.

4. Applications in cosmology

In this section we present some applications of Bayesian inference in modern cosmology. The Bayesian methods have been introduced to cosmology relatively late. The proliferation of dark energy models forced using the formal methods of model selection. Starting from the pioneering work of (John and Narlikar 2002) the Bayesian methods have become popular (Trotta 2008). The reason is that cosmology really needs these methods because our knowledge about the Universe is still very limited. We have a lot of theories about different stages of the Universe but only a modest number of observations to verify them. So, it is the right place for the Bayesian methods.

4.1. Dark energy and dark matter

The first case that we would like to talk about is connected with very mysterious behavior of the Universe, discovered in the end of the last decade. Namely, observations of the distant type Ia supernovae (SNIa) indicated that the Universe expansion accelerates (Riess et al. 1998). This discovery was apparently in the conflict with other observations and beliefs that Universe in filled with normal matter like dust, stars, planets etc. Given this assumption the Universe always decelerates. So, what is happening in the Universe? Why did it start to accelerate after a previous phase of deceleration? What is the mysterious component of the Universe that we call dark energy? There is presently an enormous number of possible answers for this question. The presence of the cosmological constant Lambda, phantom fluid, modified gravity, field of quintessence, quantum gravitational effects, brane world models, vacuum energy and so on and so on. The number of possible solutions is really impressive. So, which one of them is the real solution? Which one describes Nature in the right way? Or maybe none of them, maybe we still must be looking for new models. Let us neglect the possibility of building new models to explain acceleration of the Universe and limit ourselves only to some already proposed solutions. This is precisely what we can do thanks to Bayesian inference. Due to Bayesian methods one can obtain a ranking of accelerating models (Szydłowski, Kurek, and Krawiec 2006; Kurek and Szydłowski 2008). The analysis shows that standard cosmological model, the so-called LCDM model, is on the top off all theoretical propositions (Kurek and Szydłowski 2008).

The assumed model of the Universe is the Friedmann-Robertson-Walker model filled with dust matter, dark and baryonic, and dark energy. In Bayesian estimation of cosmological model parameters for these components we start from the cosmography which determines the luminosity function as the function of density parameters. The density parameters define the fraction of energy in the total energy budget of the universe. They are dimensionless, and their sum is equal one. The additional parameter is the Hubble constant which describes the rate of expansion of the current Universe. While the kinematic part of the cosmology is controlled by cosmography basing on analysis of trajectories of photons, the observations of CMB control the dependence of perturbation on time.

Among many theoretical propositions the simplest candidate for dark energy is the time-independent cosmological constant. The other propositions: the Chaplygin gas, models with varying coefficient of the equation of state, quintessence, phantoms, etc. These models are fitted using the astronomical data: the measurement of absolute magnitudes of high redshift type Ia supernovae, cosmic microwave background (CMB) radiation, the measurement of baryon acoustic peak (BAO) in galaxies correlation function, gas mass fraction value in galaxy clusters, gamma-ray bursts (GRB) which are at higher redshift than SNIa. In selection of cosmological models with different forms of dark energy different criteria are used. The simplest criteria are the information criteria AIC, BIC and Bayes factor, posterior probability. These methods allow us pick out the best model in the light of data at our disposal.

4.2. Can we distinguish quantum gravitational effects from observations?

This intriguing question corresponds to the presence of possible observational phenomena of quantum gravity. The quantum gravitational effects are predicted to be very small and unreachable by the present and any future generations of accelerators. However, the quantum gravitational effect can survive as the relict from the very early Universe in which quantum gravitational effects have been dominant. These effects can influence the spectrum of inflationary perturbations (Danielsson 2002; Mielczarek 2008). These primordial fluctuations then lead to the fluctuations of matter and finally to the structures formation in the Universe. So, can we deduce some information about quantum gravity from observations of microwave background radiation and large-scale structures? At first glance it can sound strange because quantum gravity describes microscopic property of gravitational field at a very deep level. However, the same effects were very important in the early universe and could influence its global properties. Is here a place for Bayesian inference? The answer is *Yes*. We have now a lot of predictions from quantum theories of gravity like Loop Quantum Gravity (Ashtekar and Lewandowski 2004) and still a very limited number of observations in the region where they can be important, e.g. non-gaussianity in primordial fluctuation in CMB or the primordial gravitational waves spectrum. The two examples presented in this section are very important, but they are not the only ones. There is a lot of other places in cosmology where Bayesian inference is and should be applied. For example, we still do not know what the dark matter (the second dominant component of the Universe) is. It may be an axion, higisno, gluino other super particles (Jungman, Kamionkowski, and Griest 1996) or just neutrinos etc. It is an ideal place for the

Bayesian inference to point out the best candidate. Unfortunately, still too little is known from observations.

5. Some general epistemic remarks on Bayesian inference

It is an obvious fact that Bayes' theorem is a theorem. Nevertheless, it should be explicitly explained, that to be a Bayesian is more than *to know and use the theorem*. The Bayesian theory of confirmation achieves a great success in such areas of human activity as physics, biology, medicine, cognitive science (decision theory) on the one hand and encounters serious limitations of the analyzed method on the other hand. A radical Bayesian would probably say that it can always be applied: we always have to do with a joint distribution and a numerical representation of belief is always possible. However, a Bayesian epistemology could be treated not only as competing with other methods of confirmation, but as a way of making a specific (not exclusive) model of beliefs and related evidences, which can be easily elaborated and understood (Talbott 2008).

In the Bayesian approach probability is attributed to hypotheses that are being confirmed. This confirmation can be interpreted both qualitatively and quantitatively, since inference is based on empirical data and there are relations between hypotheses, theories and observations to be explicated. It is indeed a crucial point in understanding Bayesian inference – the meaning which is ascribed to probability. Using probabilistic methods, one can measure two things: how often a specific event occurs and how strong evidence (confirming our beliefs) is. Let us generally state that Bayesianism can be treated as an epistemic theory which examines the relation between beliefs and empirical evidence as to measure the strength of the beliefs. As it has been shown the most important concept used to gain that goal is the notion of conditional probability (Strevens 2006). Using the Bayesian inference, we not only measure the strength of beliefs but also propose the method for rational estimating a change of the beliefs under the influence of a new evidence.

Sometimes, among scientists and philosophers of science, there is a bit of hesitation about exclusiveness of such an approach. S. Okasha in his study on van Frassen's conception of induction wrote (Okasha 2000, 693): "He accepts the Bayesian representation of opinion in terms of degrees-of-belief, and he agrees that synchronic probabilistic coherence is a necessary condition of rationality. However, he does not accept the Bayesian thesis that conditionalization is the only rational way to respond to new evidence; though he allows that it is a rational way." It can be said in that sense that Bayesianism offers a solution to old problems of induction. We have got an approximately coherent and reasonable model for probability corrections. Of course, it is possible if having initial probability and evidence (priors). The proposed solution has its weakness: its method often tells nothing how to estimate these probabilities.

The crucial point in Bayesian inference lies in the fact that it is able to deal with the problem, only if we manage the input of some probabilities. It makes sense since we never start reasoning with absolutely no knowledge. The result we achieve – $P(H|E)$ – are always "conditional": it reveals a property of H which is not objective, but related to E and certain knowledge, called *background knowledge*. A subtler epistemic analysis can be carried out with reference to the types of background knowledge. P. Wang studied the problem of underlying knowledge and discerned two types of conditions which influence the evaluation of probability distribution function (Wang 2004, 98–99), in the formulas, as follows:

- Explicit conditions:

$$P(H|K_1) = P(H|E \wedge K_0) = \frac{P(E|H \wedge K_0)P(H|K_0)}{P(E|K_0)}$$

where E is a binary proposition, belongs to proposition space (then we can define $P_0(E)$), $P_0(E) > 0$;

- Implicit conditions

$$P_{K_1}(H) = P_{K_0}(H|E) = \frac{P_{K_0}(E|H)P_{K_0}(H)}{P_{K_0}(E)}$$

– non-binary propositions allowed,
– there may be statements outside the proposition space,
– "Even if a proposition is assigned to a prior probability of zero according to one knowledge source, it is possible for the proposition to be assigned a non-zero probability according to another knowledge source".

All these discernments are not trivial since we have to answer the question: *whether all the background knowledge can be probabilistic-valued?* This is one of the most important epistemic questions of Bayesian Theory of Confirmation, beside the others:

- *Are there degrees of belief?* The answer, maybe, lies in an attempt to distinguish "rational" degrees of belief from belief in general. Are the corrections in probability, gained in the Bayesian procedure, just new probabilistic information or do they deliver *new reason to believe* that the proposition considered is true?
- When are we actually updating our belief and when there is just a revision of probability (known problem of old evidence)?

While elaborating empirical data in cosmology, one can use classical or Bayesian statistics. In a classical approach we rely on the classical definition of likelihood but doing the same *Bayesian way* we are dealing with *a priori* and *a posteriori* probability, respectively. There is an opinion among some cosmologists that the standard approach (*the standard maximum-likelihood technique*) is satisfactory indeed, which is reasonable if there is a single model under consideration. When we have several competitive hypotheses, Bayesian statistics would be a better choice. It must be said also, that degree of belief cannot offer us the true model of the universe with dark energy (Kurek et al. 2009). There are two groups among Bayesians who differ from each other with respect to criteria used in choosing of priors: objective Bayesians: (Jaynes and Bretthorst 2003; Jeffreys 1961; Rosenkrantz 2012) and subjective Bayesians: (De Finetti 1975; Howson and Urbach 1989).

It is not only the problem (or problems) of induction, that Bayesian inference tries to deal with, but also the problem of finding a justification of induction inference itself, which can be explicated in several schemas:

- Inductive Generalization

Nobody denies that a finite number of experimental data cannot deliver an exhaustive proof to a universal statement but, according to induction, empirical evidence confirms generalization (Carnap 1962; Reichenbach 1971). An example of that is enumerative induction which principle explicates, as follow:

Several crows are black.
Therefore, all crows are black.

Every load added so far has not damaged this truck.
Therefore, the next piece of load will not damage this truck.

- Hypothetical Induction

It occurs when some hypothesis deductively entails the evidence.

An evidence confirms hypothesis, if the evidence is a logical consequence of the hypothesis.

In the case of existing multiple competing hypotheses, one can try to show that the falsity of the hypothesis entails the falsity of the evidence or use additional criteria of hypotheses' selection, like simplicity or inference to the best explanation. However, these proposals create *market of hypotheses* (for example cosmological models) with rules for successful selection but in fact they cannot give any rational explanation to the evidence. It may be a truism, but the difference between explanation and confirmation should be treated with special care. The more so because there is not a unity among Bayesians concerning representation of the degree to which evidence supports a hypothesis. The most popular are three options:

1) a difference measure: $P(H|E) - P(H)$;

2) a normalized difference measure: $P(H|E) - P(H|\neg E)$;

3) a likelihood measure: $\frac{P(H|E)[1-P(H)]}{[1-P(H|E)]\,P(H)}$.

The inductive generalization, which has the simple pattern: extrapolation from particular data do general conclusions, suffers several problems called paradoxes of confirmation. Goodman's paradox, know in the literature as problem of "grue", is particularly interesting (Goodman 1983). Especially the question of its counterpart in the field of cosmology. In a traditional version:

- We have two hypotheses: (1) all emeralds are green and (2) all emeralds are grue (green if examined until some time t and blue otherwise).
- Evidence: *found emerald is green* confirms both: (1) and (2).

Any satisfying resolutions to the paradox propose additional assumptions; for example, pointing out on "green" as a natural kind term instead of "grue".

In a search for possible cosmological version of the paradox we can compare for example two related models the cold dark matter cosmological model (CDM model) and the Lambda cold dark matter cosmological model with the positive cosmological constant term (LCDM model). The latter seems to be the simplest candidate for the dark energy description. The Bayesian method of confirmation dedicated to select between these two models reveals a quite opposite verdict while used in the 90s and currently. Using the sample of (Perlmutter et al. 1998) there is not enough information to distinguish these models. The extended sample with additional 42 high z SNIa (Perlmutter et al. 1999) gives a weak evidence to favor the LCDM model over the CDM one. However, in our opinion it is a misunderstanding to treat this study case as a paradox in Goodman's sense. It becomes obvious, because when new observational data confirm better the LCDM model in comparison with the CDM model, the latter simply disappears out the stage. The paradox of confirmation would occur when related to a certain family of models there will be the same degree of confirmation (the same time and evidence) assigned to hypotheses differing from each other for example with regard to foreseeable future scenarios of Universe evolution[1].

To illustrate this situation let us consider two hypotheses: 1) The Universe decelerates; 2) The Universe decelerates until some time t and accelerates afterward. The CDM model is valid with the first hypothesis and the LCDM model is in agreement with the second one. From the 60's it was known that the Universe is expanding with the decelerating rate. So, we have a paradox here. However, the evidences of accelerating Universe due to SNIa data falsified the first model.

[1] Historically these two hypotheses have never coexisted in the same time. Until the late 90s the hypothesis of the CDM model was accepted by cosmologists, but SNIa observations made that the new hypothesis of the LCDM model was necessary to be formulated.

And the paradox is naturally solved. This example teaches us that paradoxes of Goodman's type (in the logic of induction) are common in evolutionary sciences but they are not dangers because we hope that new evidences (which appear due to science development) we discriminate between two hypotheses.

In Goodman's paradox there is only one kind of evidence; we need to draw an emerald and check its color. In the case of cosmological hypotheses, we are not left with only one evidence. A new evidence appears and resolves the paradox in favor of one of the hypothesis. It comes from new observations. We know that this evidence will appear eventually because we, scientists look for it. The reason that there is no paradox after a new evidence appears, is that one hypotheses is falsified (the CDM model) and only one hypothesis (the LCDM model) becomes in agreement with this new evidence.

It is often said that a scientific theoretical research means achieving two specific goals: (1) finding a model which approximates a phenomenon best and (2) constructing a hypothesis that offers best prediction. It is a good example to show how in this context two criteria of model selection are being compared: the Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Liddle et al. 2007). Although these model comparison methods are put together as competitors, they in fact try to ask different questions (Szydłowski et al. 2015). The AIC estimates predictive power of an elaborated hypothesis, while the BIC – goodness-of-fitting (Sober 2002). M. Forster and E. Sober have explained this nuance with respect to the fitting problem (Malcolm Forster 1994, 5–9): "Even though a hypothesis with more adjustable parameters would fit the data better, scientists seem to be willing to sacrifice goodness-of-fit if there is a compensating gain in simplicity. (…) Since we assume that observation is subject to error, it is overwhelmingly probable that the data we obtain will not fall exactly on that true curve. (…) Since the data points do not fall exactly on the true curve, such a best-fitting curve will be *false*. If we think of the true curve as the 'signal' and the deviation from the true curve generated by errors of observation as 'noise', then fitting the data perfectly involves confusing the noise with the signal. It is overwhelmingly probable that any curve that fits the data perfectly is false."

The general comments of this section can be summed up by a statement that Bayesian inference is a method dedicated to specific goals in scientific practice (Linder and Miquel 2008). With respect to cosmology, the mentioned LCDM – CDM models comparison reveals in Bayesian inference context another problem. It strictly concerns currently changing concept of the model in physics (Morrison 2005). At present there is a special emphasis placed on effectiveness and mediating function of models in physics. This status of scientific models is determined by the way they are designed: they are not simply derived from the underlying theory, nor fixed by the evidence only. Their "nature" is determined by a mediating role (between a theory and phenomena). Morrison states, as follows (Morrison

1998, 67): "Although they are designed for a specific purpose these models have an autonomous role to play in supplying information, information that goes beyond what we are able to derive from the data/theory combination alone."

In cosmology built on general relativity, the solutions of Einstein equations can be treated as the geometrical models of the Universe. A construction of a model starts from assuming specific idealizations (symmetries, etc). It means in a practice that we reduce degrees of freedom (all apart gravitational ones are neglected). For example, assumption of spacial homogeneity means that the Einstein equations which constitute the system of non-linear partial differential equation system reduce to the ordinary differential equation system in the cosmological time. It is said that those formulations of scientific laws are certain approximations of the investigated phenomena. There has been recently quite an important and interesting discussion about validity of application the Bayesian inference to idealization itself (Shaffer 2001; Jones 2006). The problem concerns idealized hypotheses and a question of assignment probability to them since they can be treated as counterfactuals. What is a posterior probability of the ideal gas law or the law of motion for simple pendulum? Jones showed that solution lies exactly in the understanding of the procedure of elaborating a model. If we treat the model idealizations not as a result of abstraction but as a distortion, the methodological consequences may exclude for Bayesian inference (Jones 2006, 3): "Given that most scientific hypotheses are idealized in some way, Bayesianism seems to entail that most scientific hypotheses cannot be confirmed. Bayesians thus confront an apparent trilemma: either develop a coherent proposal for to assign prior probabilities to counterfactuals; or embrace the counterintuitive result that idealized hypotheses cannot be confirmed; or reject Bayesianism."

The general Bayesian conception of empirical evidence can be put into three main statements/consequences:

- Less probable evidence delivers best confirmation to hypothesis;
- Evidence confirms better those hypotheses in context of which it is more probable.
- If the hypothesis' probability is very little, it can be confirmed only by very strong evidence.

After the last two decades cosmology come down on the side of the LCDM model where dark energy is accepted as the force behind the Universe acceleration without true understanding what dark energy is. The disappointing consequence of this state of affairs is that cosmological research has proliferated with various theoretical models of dark energy. The Bayesian approach turned out to be effective method to distinguish the most favored model from a set of theoretical models (hypotheses of dark energy). However, the true hypothesis can be absent among the hypotheses considered. The Bayesian methods give only indications related to the set of hypotheses. Therefore, cosmology should be open on new theories, models. Such an attitude is so-called the open-minded Bayesianism (Wenmackers and Romeijn 2015).

6. Summary

In this paper we have presented basics of Bayesian inference and showed how to use it in practice. We have introduced some mathematical background and formulated a problem in the similarity with thermodynamics. As a case study we choose three simple models. Then the known Monte Carlo methods and Metropolis algorithm were used to select the best model in the light of data. The general remark which can be derived from these considerations is to be careful in evaluation of the models in the light of the data and in using the complementary indicators. The Bayesian methods started to be popular due to new discoveries in cosmology at the beginning of XXI century. We presented the areas of cosmology where Bayesian inference has been applied, namely problem of dark energy, dark matter, and testing quantum effects by astronomical data. Subsequently we have studied epistemological aspects of the Bayesian confirmation theory in the context of problems of modern cosmology where the Bayesian approach offers not only the estimation of model parameters from the observational data but also methods of the comparison of models (selection). We have demonstrated that the Bayesian inference is based on some assumptions of philosophical character. The philosophical issues of inference in context of cosmological models on the example of models without and with the dark energy component (the cosmological constant) are discussed. We pointed out that Goodman's famous paradox does not appear in the cosmology reconstructed using Bayesian methodology. The reason for this we are looking for new evidences which falsify one hypothesis such that only one hypothesis becomes in agreement with observational data. Note that the Bayesian framework enable us to test and select between competing hypotheses so one can construct the ranking of cosmological models explaining acceleration of the current Universe. Therefore, we obtain the best model favored by data.

Appendix

A. Monte Carlo method, Markov chain and Metropolis algorithm

In this appendix we show how to compute thermodynamical integral (17) with use of the Monte Carlo simulations. Our short introduction to this subject is based partially on this made in (Huang 2001). Let us consider state of the system labeled by Γ and corresponding energy $E(\Gamma)$. Our task is to compute integral:

$$\langle E\rangle_T = \frac{\int d\pi E(\Gamma)e^{-E(\Gamma)/T}}{\int d\pi e^{-E(\Gamma)/T}} \tag{37}$$

where integration is performed over all available states Γ. Since in numerical computations we always discretize the system, integration $\int d\pi$ is replaced by the summation. Our task now is to write a program which generates states Γ from the canonical ensemble given with the probability $e^{-E(\Gamma)/T}$. The crucial observation is that we do not have to generate all possible states to calculate (37). The main contribution to their value comes from the equilibrium states. Therefore, the idea is to find these equilibrium states and average over them. Starting from some arbitrary initial state we create a sequence of states:

$\grave{\Gamma}_1 \to \cdots \to \grave{\Gamma}_K$	$\to$	$\Gamma_1 \to \cdots \to \Gamma_K$
Non-equilibrium		equilibrium

finally finding ensemble of equilibrium states. Then one can calculate:

$$\langle E\rangle_T \simeq \frac{1}{K}\sum_{i=1}^{K} E(\Gamma_i) \tag{39}$$

In order to find equilibrium states the Markov chain method can be applied. We consider sequence of transitions $\Gamma_i \to \Gamma_{i+1}$ with probability $P(\Gamma_i|\Gamma_{i+1})$. Moreover, we assume

$$P(\Gamma_i|\Gamma_{i+1}) \geq 0 \tag{40}$$

$$\sum_{\Gamma_{i+1}} P(\Gamma_i|\Gamma_{i+1}) = 1 \tag{41}$$

$$e^{-E(\Gamma_i)/T}P(\Gamma_i|\Gamma_{i+1}) = e^{-E(\Gamma_{i+1})/T}P(\Gamma_{i+1}|\Gamma_i) \tag{42}$$

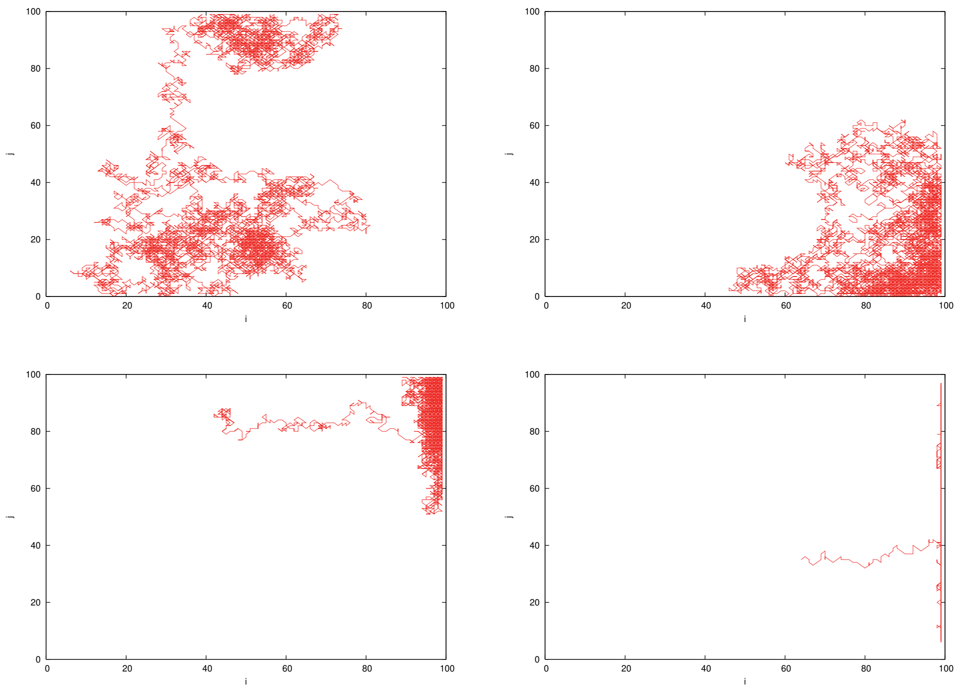
Practical realization of the above conditions is given by *Metropolis algorithm*. Namely it states:

- Take initial state Γ_i.

- Make some move to neighboring state Γ_{i+1}.
- If $E(\Gamma_{i+1}) < E(\Gamma_i)$, accept the change.
- If $E(\Gamma_{i+1}) > E(\Gamma_i)$, accept the change conditionally with the probability $e^{-[E(\Gamma_{i+1})-E(\Gamma_i)]/T}$.

All computations performed in Sec. 3 have been done applying this simple set of rules. As an example, we show here the Markov chains in the parameter space for the models considered there. In Fig. 5 we show sequence of moves for the first model considered in Sec. 3. We assume values of the parameter λ = 0,1; 1; 10; 100. Since $\frac{1}{\lambda} = T$ the higher value of λ corresponds to lower temperatures. We investigate here a broad range in λ, however for the calculations of the evidence only values of $\lambda \in [0,1]$ are required.

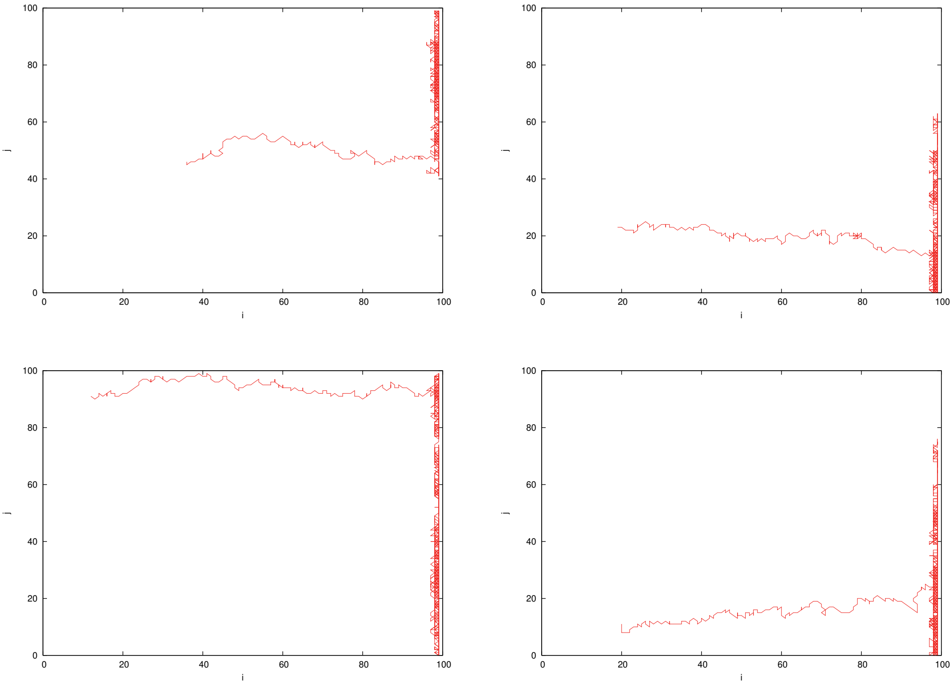
Figure 5. Top left: λ = 0,1; Top right: λ = 1; Bottom left: λ = 10; Bottom right: λ = 100



B. Numerical demonstration of property of *ergodicity*

The very important question related to Monte Carlo simulations is the "ergodicity" of the algorithm. It means that in the finite number of steps (finite time) the system must be freely close to any point in the phase space. This prevents the system to being trapped in a subset of states. In the Monte Carlo simulations it causes that we can always find a proper energy minimum, even for very low temperatures when fluctuations are small. To check it we performed Markov chains in the low temperature system. In such a system, when algorithm is not ergodic, a Markov chain cannot always lead to the proper minimum. In Fig. 6 we show Markov chain in the parameters space for the third model considered in Sec. 3. We show that starting from the different points in the parameter space system always go to the same region where the proper minimum is placed. This is a visual proof of ergodicity for a kind of function considered. It is possible that it is not true for more complicated kind of functions.

Figure 6. Top left: $\lambda = 100$; Top right: $\lambda = 100$; Bottom left: $\lambda = 100$; Bottom right: $\lambda = 100$



References

Ashtekar, A., Lewandowski, J., 2004, *Background Independent Quantum Gravity: a Status Report*. "Classical and Quantum Gravity" 21.15: R53–R152.

Carnap, R., 1962, *Logical Foundations of Probabilities*. Chicago: University of Chicago Press.

Danielsson, U.H., 2002, *Note on Inflation and Trans-Planckian Physics*. "Physical Review D" 66.2: 123501.

De Finetti, B., 1975. *Theory of Probability*. London: John Wiley & Sons.

Goodman, N., 1983, *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Howson, C., Urbach, P., 1989, *Scientific Reasoning*. La Salle, IL: Open Court Pub Co.

Huang, K., 2001, *Introduction to Statistical Physics, Second Edition*. London: Taylor & Francis.

Jaynes, E.T., Bretthorst, G.L., 2003, *Probability Theory*. Cambridge: Cambridge University Press.

Jeffreys, H., 1961, *Theory of Probability*. Third. Oxford: Oxford University Press.

John, M.V., Narlikar, J., 2002, *Comparison of Cosmological Models Using Bayesian Theory*, Physical Review D. 65.

Jones, N., 2006, *Resolving the Bayesian Problem of Idealization*.

Jungman, G., Kamionkowski, M., Griest, K., 1996, *Supersymmetric Dark Matter*.

Kurek, A., Szydłowski, M., 2008, *The ΛCDM Model in the Lead—a Bayesian Cosmological Model Comparison*. "The Astrophysical Journal" 675.1: 1–7.

Kurek, A., Kukier, Ł., Szydlowski, M., Tambor, P., 2009, *INTRODUCTION To The BAYESIAN METHODOLOGY of CONTEMPORARY COSMOLOGY* (*Wstep Do Bayesowskiej Metodologii Wspolczesnej Kosmologii*). "Zagadnienia Filozoficzne W Nauce (Philosophical Problems in Science)" 45: 62–96.

Liddle, A.R., Corasaniti, P.S., Kunz, M., Mukherjee, P., Parkinson, D., Trotta, R., 2007, *Comment on 'Tainted Evidence: Cosmological Model Selection Versus Fitting', by Eric v. Linder and Ramon Miquel* (Astro-Ph/0702542v2), March.

Linder, E.V., Miquel, R., 2008, *Cosmological Model Selection: Statistics and Physics*. "International Journal of Modern Physics D" 17.12: 2315–24.

MacKay, D.J.C., 2003, *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

Malcolm, F., Sober, E., 1994, *How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions*. "The British Journal for the Philosophy of Science" 45.

Mielczarek, J., 2008, *Gravitational Waves From the Big Bounce*. "Journal of Cosmology and Astroparticle Physics" 11 011.

Morrison, M.C., 1998, *Modelling Nature: Between Physics and the Physical World*.

Morrison, M.C., 2005, *Approximating the Real: the Role of Idealizations in Physical Theory*, Edited by M R Jones and N Cartwright. *Poznan Studies in the Philosophy of the Sciences and the Humanities* 86. Amsterdam: 145–72.

Okasha, S., 2000, *Van Fraassen's Critique of Inference to the Best Explanation*. "Studies in History and Philosophy of Science Part A" 31.4: 691–710.

Perlmutter, S., Aldering, G., Goldhaber, G., et al., 1999, *Measurements of Ω and Λ From 42 High-Redshift Supernovae*. "The Astrophysical Journal" 517.2: 565–86.

Perlmutter, S., Aldering, G., Della Valle, M., et al., 1998, *Discovery of a Supernova Explosion at Half the Age of the Universe*. "Nature" 391.6662: 51–54.

Reichenbach, H., 1971, *The Theory of Probability*. University of California Press.

Riess, A.G., Filippenko, A.V., Challis, P., et al., 1998, *Observational Evidence From Supernovae for an Accelerating Universe and a Cosmological Constant*. "The Astronomical Journal 116.3: 1009–38.

Rodriguez-Fernández, J.L., 1999, *Ockham's Razor*. "Endeavour" 23.3: 121–25.

Rosenkrantz, R.D., 2012, *Inference, Method and Decision*. Dordrecht: Springer Science & Business Media.

Shaffer, M.J., 2001, *Bayesian Confirmation of Theories That Incorporate Idealizations*. "Philosophy of Science 68.1: 36–52.

Sober, E., 2002, *Instrumentalism, Parsimony, and the Akaike Framework*. "Philosophy of Science" 69.S3: S112–23.

Strevens, M., 2006, *The Bayesian Approach to the Philosophy of Science*, Edited by D M Borchert. *Encyclopedia of Philosophy* 1. Detroit: 495–502.

Szydłowski, M., Krawiec, A., Kurek, A., Kamionka, M., 2015, *AIC, BIC, Bayesian Evidence Against the Interacting Dark Energy Model*. "The European Physical Journal C" 75.1: 5.

Szydłowski, M., Kurek, A., Krawiec, A., 2006, *Top Ten Accelerating Cosmological Models*. "Physics Letters B" 642.3: 171–78.

Talbott, W., 2008, "Bayesian Epistemology." Edited by N Zalta. *Stanford Encyclopedia of Philosophy*.

Trotta, R., 2008, *Bayes in the Sky: Bayesian Inference and Model Selection in Cosmology*. "Contemporary Physics" 49.2: 71–104.

Wang, P., 2004, *The Limitation of Bayesianism*. "Artificial Intelligence" 158.1: 97–106.

Wenmackers, S., Romeijn, J.W., 2015, *New Theory About Old Evidence*. "Synthese" 193.4: 1225–50.