AIC, BIC, Bayesian evidence and a notion on simplicity of cosmological model

Marek Szydłowski*

Department of Theoretical Physics, Catholic University of Lublin, Al. Racławickie 14, 20-950 Lublin, Poland and Mark Kac Complex Systems Research Centre, Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland

Aleksandra Kurek†

Astronomical Observatory, Jagiellonian University, Orla 171, 30-244 Kraków, Poland

Abstract

Recent astronomical observations indicate that the Universe is in the phase of accelerated expansion. There are many cosmological models which explain this phenomenon, but should we prefer those models over the simplest one – ΛCDM model? According to the Occam's razor principle if all models describe the observations equally well we should prefer the simplest one. We consider the model comparison methods which involve such rules: the Akaike information criterion (AIC), Bayesian information criterion (BIC) and Bayesian evidence to compare the ΛCDM model with its generalisation where the interaction between dark matter and dark energy is allowed. The analyses based on the AIC and Bayesian evidence indicate that there is only a weak evidence in favour of the ΛCDM model over its generalisation, while those based on BIC quantity indicate the strong evidence in favour the simpler model. We also calculate some quantity which measure the effective number of model parameters that the given data can constrain. This value is used to compare the concordance LCDM model with its generalization basing on the extended interpretation of continuity condition-interacting ΛCDM cosmology. We conclude that data set are not enough informative to constrain all parameters allows to vary in the models.

*Electronic address: `uoszydlo@cyf-kr.edu.pl`

†Electronic address: `alex@oa.uj.edu.pl`

I. INTRODUCTION

Recent observations of type Ia supernova (SNIa) provide the main evidence that current Universe is in an accelerating phase of expansion [1]. Cosmic microwave background (CMB) data indicate that the present Universe has also negligible space curvature [2]. Therefore if we assume the Friedmann-Robertson-Walker (FRW) model in which effects of nonhomogeneities are neglected, then the acceleration must be driven by dark energy component X (matter fluid violating the strong energy condition $\rho_{\mathrm{X}} + 3p_{\mathrm{X}} \geq 0$). This kind of energy represents roughly 70% of the matter content of the current Universe. Because the nature as well as mechanism of the cosmological origin of the dark energy component are unknown some alternative theories try to eliminate the dark energy by modifying the theory of gravity itself. The main prototype of this kind of model is covariant brane models based on the Dvali-Gabadadze-Porratti (DGP) model [3] as generalized to cosmology by Deffayet [4]. The simplest explanation of a dark energy component is the cosmological constant with effective equation of state $p = -\rho$ but appears the problem of its smallness and hence its relatively recent dominance. Although the ΛCDM model offers possibility of explanation of observational data it is only effective theory which contain the enigmatic theoretical term – the cosmological constant Λ. Other numerous candidates for dark energy description have also been proposed like to evolving scalar field [5] usually referred as quintessence, the phantom energy [6, 7], the Chaplygin gas [8] etc. Some authors believed that the dark energy problem belongs to the quantum gravity domain [9].

These theoretical models are consistent with the observations, they are able to explain the phenomenon of the accelerated expansion of the Universe. But should we really prefer such models over the ΛCDM one? To answer this question we should use some model comparison methods to confront existing cosmological models having observations at hand.

Let us assume that we have N pairs of measurements (y_i, x_i) and that we want to find the relation between the y and x quantities. Suppose that we can postulate k possible relations $y \equiv f_i(x, \bar{\theta})$, where $\bar{\theta}$ is the vector of unknown model parameters and $i = 1, \ldots, k$. With the assumption that our observations come with uncorrelated gaussian errors with mean $\mu_i = 0$ and standard deviation σ_i the goodness of fit for the theoretical model is measured by the

χ^2 quantity given by

$$\chi^2 = \sum_{i=1}^{N} \frac{(f_l(x_i, \bar{\theta}) - y_i)^2}{2\sigma_i^2} = -2 \ln L, \tag{1}$$

where L is the likelihood function. For the particular family of models f_l the best one minimize the χ^2 quantity, which we denote $f_l(x, \hat{\bar{\theta}})$. The best model from our set of k models $f_1(x, \hat{\bar{\theta}}), \ldots, f_k(x, \hat{\bar{\theta}})$ could be the one with the smallest value of χ^2 quantity. But this method could give us misleading results. Generally speaking for more complex model the value of χ^2 is smaller, thus the most complex one will be choose as the best from our set under consideration.

A clue is given by so called Occam's razor principle: "If two models describe the observations equally well, choose the simplest one." This principle has aesthetic as well as empirical justification. Let us quote simple example which illustrate this rule [10] and present it in Figure 1. We see the black box and the white one behind it. One can postulate two models: 1. There is one box behind the black box, 2. There are two boxes of identical height and colour behind the black box. Both models explain our observations equally well. According to the Occam's razor principle we should accept the explanation which is simpler so that there is only one white box behind the black one. Is not it more probable that there is only one box than two boxes with the same height and colour?

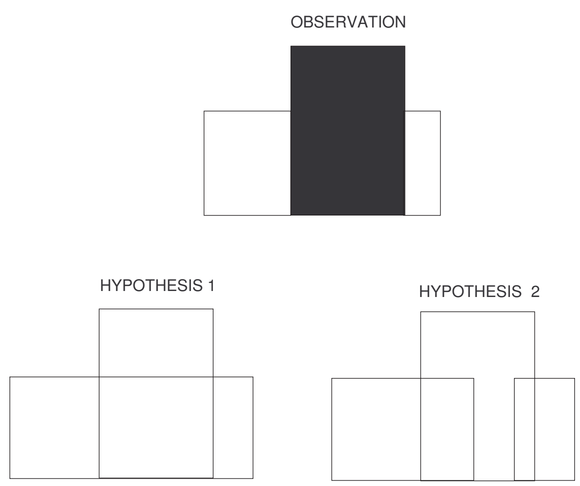

FIG. 1: Illustration to the example explaining the Occam's razor principle.

We could not use this principle directly because the situations when two models explain the observations equally well are rare. But in the information theory as well as in the bayesian theory there are methods for model comparison which include such rule.

In the information theory there are no true models. There is only reality which can be approximated by models, which depend on some number of parameters. The best one from the set under consideration should be the best approximation to the truth. The information lost when truth is approximated by model under consideration is measured by so called Kullback-Leibler (KL) information so the best one should minimizes this quantity. It is impossible to compute the KL information directly because it depends on truth which is unknown. Akaike [11] found approximation to the KL quantity which is called the Akaike information criterion (AIC) and is given by

$$\mathrm{AIC} = -2\ln\mathcal{L} + 2d, \tag{2}$$

where $\mathcal{L}$ is the maximum of the likelihood function and d is the number of model parameters. Model which is the best approximation to the truth from the set under consideration has the smallest value of the AIC quantity. It is convenient to evaluate the differences between the AIC quantities computed for the rest of models from our set and the AIC for the best one. Those differences (Δ_{AIC}) are easy to interpret and allow a quick 'strength of evidence' for considered model with respect to the best one. The models with $0 \leq \Delta_{\mathrm{AIC}} \leq 2$ have substantial support (evidence), those where $4 < \Delta_{\mathrm{AIC}} \leq 7$ have considerably less support, while models having $\Delta_{\mathrm{AIC}} > 10$ have essentially no support with respect to the best model.

It is worth noting that the complexity of the model is interpreted here as the number of its free parameters that can be adjusted to fit the model to the observations. If models under consideration fit the data equally well according to the Akaike rule the best one is with the smallest number of model parameters (the simplest one in such an approach).

In the Bayesian framework the best model (from the model set under consideration) is that which has the largest value of probability in the light of data (so called posterior probability) [12]

$$P(M_i|D) = \frac{P(D|M_i)P(M_i)}{P(D)}, \tag{3}$$

where $P(M_i)$ is a prior probability for the model M_i, $P(D)$ is normalization constant $[P(D) = \sum_{i=1}^{k} P(D|M_i)P(M_i)]$, D denotes data. $P(D|M_i)$ is the marginal likelihood, also called evidence

$$P(D|M_i) = \int P(D|\bar{\theta}, M_i)P(\bar{\theta}|M_i)d\bar{\theta} \equiv E_i, \tag{4}$$

where $P(D|\bar{\theta}, M_i)$ is likelihood under model i, $P(\bar{\theta}|M_i)$ is prior probability for $\bar{\theta}$ under model i.

Let us note that we can include the Occam's razor principle by assuming the greater prior probability for simpler model, but this is not necessary and rarely used in practice. Usually one assume that there is no evidence to favor one model over another which cause to equal value of prior for all models under consideration. It is convenient to evaluate the posterior ratio for models under consideration which in the case with flat prior for models is reduced to the evidence ratio (so called the Bayes factor B). The interpretation of the logarithm of Bayes factor values are as follows: $0 < \ln B \leq 2$ as a weak, $2 < \ln B \leq 6$ as a positive, $6 < \ln B \leq 10$ as a strong and $\ln B > 10$ as a very strong evidence in favor better model. This quantity involves the Occam's razor rule. Let us simplify the problem to illustrate how this principle works here [10, 13].

Assume that $\bar{P}(\bar{\theta}|D, M)$ is the non normalized posterior probability for the vector $\bar{\theta}$ of model parameters. In this notation $E = \int \bar{P}(\bar{\theta}|D, M) d\bar{\theta}$. Suppose that posterior has a strong pick in the maximum: $\bar{\theta}_{\mathrm{MOD}}$. It is reasonable to approximate the logarithm of the posterior by its taylor expansion in the neighbour of $\bar{\theta}_{\mathrm{MOD}}$ so we finished with the expression

$$\bar{P}(\bar{\theta}|D, M) = \bar{P}(\bar{\theta}_{\mathrm{MOD}}|D, M) \exp\left[-(\bar{\theta} - \bar{\theta}_{\mathrm{MOD}})^T \mathrm{C}^{-1} (\bar{\theta} - \bar{\theta}_{\mathrm{MOD}})\right], \tag{5}$$

where $\left[\mathrm{C}^{-1}\right]_{ij} = -\left[\frac{\partial^2 \ln \bar{P}(\bar{\theta}|D,M)}{\partial\theta_i \partial\theta_j}\right]_{\bar{\theta}=\bar{\theta}_{\mathrm{MOD}}}$. Posterior is approximated by the gaussian distribution with the covariance matrix C and the mean $\bar{\theta}_{\mathrm{MOD}}$. The expression for the evidence take a form $E = \bar{P}(\bar{\theta}_{\mathrm{MOD}}|D, M) \int \exp\left[-(\bar{\theta} - \bar{\theta}_{\mathrm{MOD}})^T \mathrm{C}^{-1} (\bar{\theta} - \bar{\theta}_{\mathrm{MOD}})\right] d\,\bar{\theta}$. Because the posterior has a strong pick near the maximum, the most contribution to the integral comes from the neighbour close to $\bar{\theta}_{\mathrm{MOD}}$. Contribution from the other other region of $\bar{\theta}$ can be ignored, so we can expand the limit of the integral to whole R^d. With this assumptions one can obtain $E = (2\pi)^{\frac{d}{2}} \sqrt{\det \mathrm{C}}\bar{P}(\bar{\theta}_{\mathrm{MOD}}|D, M) = (2\pi)^{\frac{d}{2}} \sqrt{\det \mathrm{C}} P(D|\bar{\theta}_{\mathrm{MOD}}, M) P(\bar{\theta}_{\mathrm{MOD}}|M)$. Suppose that the likelihood function has sharp pick in $\hat{\bar{\theta}}$ and the prior for $\bar{\theta}$ is nearly flat in the neighbour of $\hat{\bar{\theta}}$. In this case $\hat{\bar{\theta}} = \bar{\theta}_{\mathrm{MOD}}$ and the expression for the evidence takes the form $E = \mathcal{L}(2\pi)^{\frac{d}{2}} \sqrt{\det \mathrm{C}} P(\hat{\bar{\theta}}|M)$. The $(2\pi)^{\frac{d}{2}} \sqrt{\det \mathrm{C}} P(\hat{\bar{\theta}}|M)$ quantity is called the Occam factor (OF). When we consider the case with one model parameter with flat prior $P(\theta|M) = \frac{1}{\Delta\theta}$ the Occam factor OF$= \frac{2\pi\sigma}{\Delta\theta}$ which can be interpreted as the ratio of the volume occupied by the posterior to the volume occupied by prior in the parameter space. The more parameter space wasted by the prior the smaller value of the evidence. It is worth noting that the evidence does not penalize parameters which are unconstrained by the data [14].

As the evidence is hard to evaluation an approximation to this quantity was proposed by

Schwarz [15] so called Bayesian information criterion (BIC) and is given by

$$\mathrm{BIC} = -2\ln\mathcal{L} + 2d\ln N, \tag{6}$$

where N is the number of the data points. The best model from the set under consideration is this which minimize the BIC quantity. It is also convenient to analyse the difference between BIC quantities for the rest models from our set with the BIC for the best one. These differences could be interpreted in the following way: $0 < \Delta_{\mathrm{BIC}} \leq 2$ as a weak, $2 < \Delta_{\mathrm{BIC}} \leq 6$ as a positive, $6 < \Delta_{\mathrm{BIC}} \leq 10$ as a strong and $\Delta_{\mathrm{BIC}} > 10$ as a very strong evidence in favour of a better model.

One can notice the similarity between the AIC and BIC quantities though they come from different approaches to model selection problem. The dissimilarity is seen in the so called penalty term: ad, which penalize more complex models (complexity is identified here as the number of free model parameters). One can evaluated the factor by which the additional parameter must improve the goodness of fit to be included in the model. This factor must be greater than a so equal to 2 in the AIC case and equal to $\ln N$ in the BIC case. Notice that the latter depends on the number of the data points.

It should be pointed out that presented model selection methods are widely used in context of cosmological model comparison [13, 14, 16–35].

We should keep in mind that conclusions based on such quantities depend on the data at hand. Let us mention again the example with the black box. Suppose that we made a few steps toward this box that we can see the difference between the height of the left and right side of the white box. Our conclusion change now.

Let us quote example taking from [25]. Assume that we want to compare the Newtonian and Einsteinian theories in the light of the data coming from laboratory experiment where general relativistic effects are negligible. In this situation Bayes factor between Newtonian and Einsteinian theories will be close to unity. Whereas comparing the general relativistic and Newtonian explanations of the deflection of a light ray that just grazes the sun's surface give the Bayes factor $\sim 10^{10}$ in the favor of the first one (and even greater with more accurate data).

Having this in mind an interesting supplement to the above considerations seems to be quantity which measure the effective number of model parameters (C_b) that the given data set can constrain [31, 34]. It can be computed using the relation

$$\mathrm{C}_b = \overline{\chi^2(\bar{\theta})} - \chi^2(\hat{\bar{\theta}}), \tag{7}$$

where the mean is taken over the posterior pdf for $\bar{\theta}$ and $\hat{\bar{\theta}}$ is the mode of the posterior pdf (different choices are also possible). As have been shown this quantity correspond to the number of parameters for which the width of the posterior probability distribution is significantly narrower than the width of the prior probability distribution [31]. These parameters can be considered to have been well measured by the data given our prior assumption in the model. It helps to determine if the data is informative enough to measure the parameters under consideration.

We share with George Efstathiou opinion [36–38] that there is no sound theoretical basis for considering the dynamical dark energy, where as we are beginning to see an explanation for a small cosmological constant emerging from more fundamental theory. In our opinion the ΛCDM model has the status of satisfactory effective theory. Estathiou argued why the cosmological constant should be given higher weight as a candidate for dark energy description than dynamical dark energy. In this argumentation Occam's razor is used to point out a more economical model explaining the observational data. On the other hand Biesiada advocated the use of Akaike information criterion which favour rather a dynamical model of dark energy (quintessence model) [27]. In our opinion quantification of Occam's razor by computing Bayesian evidence and Bayesian information criterion is more suitable as the AIC is unadequate for many statistical problems [39].

The main aim of this paper is to compare the simplest cosmological model – ΛCDM model – with its generalisation where the interaction between dark energy and dark matter sector is allowed using methods described above.

II. INTERACTING MODEL

The interacting interpretation of continuity condition (conservation condition) was investigated in the context of the coincidence problem since the paper Zimdahl [40], for recent developments in this area see Olivares et al. [41, 42], see also Le Delliou et al. [43] for discussion recent observational constraints.

Let us consider two basic equations which determine the evolution of FRW cosmological models

$$\frac{\ddot{a}}{a} = -\frac{1}{6}(\rho + 3p) \tag{8}$$

$$\dot{\rho} = -3H(\rho + p). \tag{9}$$

Equation (8) is called the accelerated equation and equation (9) is the conservation (or adiabatic) condition. Equation (8) can be rewritten to the form analogous to the Newtonian equation of motion

$$\ddot{a} = -\frac{\partial V}{\partial a}, \tag{10}$$

where $V = V(a)$ is potential function of the scale factor a. To evaluate $V(a)$ from (10) via integration by parts it is useful to rewrite (9) to the new equivalent form

$$\frac{d}{dt}(\rho a^3) + p\frac{d}{dt}(a^3) = 0. \tag{11}$$

From (8) we obtain

$$\frac{\partial V}{\partial a} = \frac{1}{12}(\rho + 3p)d(a^2). \tag{12}$$

It is convenient to calculate pressure p from (11) and then substitute to (12). After simple calculations we obtain from (12)

$$\frac{\partial V}{\partial a} = -\frac{1}{6}\left[a^2\frac{d\rho}{da} + \rho d(a^2)\right]. \tag{13}$$

Therefore

$$V = -\frac{\rho a^2}{6}. \tag{14}$$

In formula (14) ρ means effective energy density of the fluid filled the Universe.

We find very simple interpretation of (8): the evolution of the Universe is equivalent to motion of the particle of unit mass in the potential well parameterized by the scale factor. In the procedure of reduction of the problem of FRW evolution to the problem of investigation dynamical system of Newtonian type we only assume that effective energy density satisfies the conservation condition. We do not assume the conservation condition for each energy component (or noninteracting matter sectors).

Equations (8) and (9) admit the first integral which usually called the Friedmann first integral. This first integral has a simple interpretation in the particle-like description of the FRW cosmology, namely energy conservation

$$\frac{\dot{a}^2}{2} + V(a) = E = -\frac{k}{2}, \tag{15}$$

where k is the curvature constant and V is given by formula (14).

Let us consider the universe filled with two components fluid

$$\rho = \rho_{\mathrm{m}} + \rho_X, \quad p = 0 + w_X \rho_X, \tag{16}$$

where ρ_{m} means energy density of usual dust matter and ρ_X denotes energy density of dark energy satisfying the equation of state $p_X = w_X \rho_X$, where $w_X = w_X(a)$. Then equation (11) can be separable on dark matter and dark energy sectors which in general can interacts

$$\frac{d}{dt}(\rho_{\mathrm{m}} a^3) + 0 * \frac{d}{dt}(a^3) = \Gamma$$
$$\frac{d}{dt}(\rho_X a^3) + w_X(a)\rho_X \frac{d}{dt}(a^3) = -\Gamma$$

In our previous paper [44] it was assumed that

$$\Gamma = \alpha a^n \frac{\dot{a}}{a}, \tag{17}$$

which able us to integrate (17) which gives

$$\rho_m = \frac{C}{a^3} + \frac{\alpha}{n} a^{n-3} \tag{18}$$

$$\frac{d\rho_X}{da} + \frac{3}{a}(1 + w_X(a))\rho_X = -\alpha a^{n-4}. \tag{19}$$

The solution of homogeneous equation (19) can be written in terms of average $\overline{w_X}(a)$ as

$$\rho_X = \rho_{X,0} a^{-3(1+\overline{w_X}(a))}, \tag{20}$$

where

$$\overline{w_X}(a) = \frac{\int w_X(a) d(\ln a)}{d(\ln a)}. \tag{21}$$

The solution of nonhomogeneous equation (19) is

$$\rho_X = -\left[\int_1^a a^{n-1+3\overline{w_X}(a)} da\right] a^{-3(1+\overline{w_X}(a))} + \frac{C_X}{a^{3(1+\overline{w_X}(a))}}. \tag{22}$$

Finally we obtain

$$\rho_{\mathrm{eff}} \equiv 3H^2 + 3\frac{k}{a^2} = \rho_m + \rho_X = \tag{23}$$
$$\frac{C_m}{a^3} + \frac{\alpha}{n+1} a^{n-3} + \frac{C_X}{a^{3(1+\overline{w_X}(a))}} - \left[\int_1^a a^{n-1+3\overline{w_X}(a)} da\right] a^{-3(1+\overline{w_X}(a))} + \frac{C_X}{a^{3(1+\overline{w_X}(a))}}.$$

The second and last terms origin from interaction between dark matter and dark energy sectors.

Let us consider the simplest case of $\overline{w_X}(a)$ =const= $w_X(a)$. Then integration of (22) can be performed and we obtain

$$\rho_{\text{eff}} = \frac{C_m}{a^3} + \frac{C_X}{a^{3(1+w_X)}} + \frac{C_{\text{int}}}{a^{3-n}} \tag{24}$$

where $C_{\text{int}} = \frac{\alpha}{n} - \frac{\alpha}{n-3w_X}$. In this case we obtain one additional term in ρ_{eff} or in the Friedmann first integral scaling like a^{2-n}. It is convenient rewrite the Friedmann first integral to the new form using dimensionless density parameter. Then we obtain

$$\left(\frac{H}{H_0}\right)^2 = \Omega_{\text{m},0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_{\text{int}}(1+z)^{3-n} + \Omega_{X,0}(1+z)^{3(1+w_X)}. \tag{25}$$

Note that this additional power law term related to interaction can be also interpreted as the Cardassian or polytropic term [45, 46] (one can easily show that the assumed form of interaction always generates a correction of type $a^m, m = 1-n$, in the potential of the ΛCDM model and vice versa). Another interpretation of this term can origin from Lambda decaying cosmology when Lambda term is parametrized by the scale factor [47].

In the next section we draw a comparison between the above model with the assumption that $\overline{w_X}(a) = \text{const} = -1$ and the ΛCDM model.

III. RESULTS

To compare the ΛCDM model with the interacting ΛCDM model we use the SNIa data, constraints from the CMB shift parameter, constraints from SDSS parameter A as well as $H(z)$ observational data.

We use $N_1 = 192$ SNIa data [48–50]. In this case the likelihood function has the following form

$$L_{SN} \propto \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{N_1} \frac{(\mu_i^{theor} - \mu_i^{obs})^2}{\sigma_i^2}\right)\right], \tag{26}$$

where σ_i is known, $\mu_i^{\text{obs}} = m_i - M$ (m_i–apparent magnitude, M–absolute magnitude of SNIa), $\mu_i^{theor} = 5\log_{10} D_{Li} + \mathcal{M}$, $\mathcal{M} = -5\log_{10} H_0 + 25$ and $D_{Li} = H_0 d_{Li}$, where d_{Li} is the luminosity distance, which with the assumption $k = 0$ is given by $d_{Li} = (1+z_i)c\int_0^{z_i} \frac{dz'}{H(z')}$.

We also include information obtained from the CMB data. Here the likelihood function has the following form

$$L_R \propto \exp\left[-\frac{(R^{\text{theor}} - R^{\text{obs}})^2}{2\sigma_R^2}\right], \tag{27}$$

where R is the so called shift parameter, $R^{\rm theor} = \sqrt{\Omega_{\rm m,0}} \int_0^{z_{dec}} \frac{H_0}{H(z)} dz$, and $R^{\rm obs} = 1.70 \pm 0.03$ for $z_{\rm dec} = 1089$ [2, 51].

As the third observational data we use the measurement of the baryon acoustic oscillations (BAO) from the SDSS luminous red galaxies [52]. In this case the likelihood function has the following form

$$L_A \propto \exp\left[-\frac{(A^{\rm theor} - A^{\rm obs})^2}{2\sigma_A^2}\right], \tag{28}$$

where $A^{\rm theor} = \sqrt{\Omega_{\rm m,0}} \left(\frac{H(z_A)}{H_0}\right)^{-\frac{1}{3}} \left[\frac{1}{z_A}\int_0^{z_A} \frac{H_0}{H(z)} dz\right]^{\frac{2}{3}}$ and $A^{\rm obs} = 0.469 \pm 0.017$ for $z_A = 0.35$.

Finally we used the observational $H(z)$ data ($N_2 = 9$) from [53] (see also [54, 55] and references therein). This data based on the differential ages ($\frac{dt}{dz}$) of the passively evolving galaxies which allow to estimate the relation $H(z) \equiv \frac{\dot{a}}{a} = -\frac{1}{1+z}\frac{dz}{dt}$. Here the likelihood function has the following form

$$L_H \propto \exp\left(-\frac{1}{2}\left[\sum_{i=1}^{N_2} \frac{(H(z_i) - H_i(z_i))^2}{\sigma_i^2}\right]\right),$$

where $H(z)$ is the Hubble function, H_i, z_i are observational data.

The final likelihood function is given by

$$L = L_{SN} L_R L_A L_H, \tag{29}$$

with the number of data points $N = 192 + 1 + 1 + 9$.

To obtain the values of AIC and BIC quantities we perform the $\chi^2 = -2\ln L$ minimization procedure after marginalization over the H_0 parameter in the range $< 60, 80 >$. The values of Bayesian evidence as well as effective number of model parameters were obtained using the MCMC algorithm called nested sampling [56] which implementation to the cosmological case is available as a part of the CosmoMc code [57, 58] called CosmoNest [18, 19, 32, 59] which was changed for our purpose. We assume flat prior probabilities for the model parameters in the range: $H_0 \in < 60, 80 >$, $\Omega_{\rm m,0} \in < 0, 1 >$, $\Omega_{\rm int} \in < -1, 1 >$, $3 - n \equiv m \in < -10, 10 >$. The values of evidence and C_b were averaged from the eight chains. Results are presented in Table I.

Analysing the value of difference between the AIC quantity one can conclude that there is substantial support for the interacting ΛCDM model with respect to the ΛCDM model while the difference between the BIC quantity to the conclusion that there is strong evidence against the more complex model. The logarithm of the Bayes factor quantity confirm the first

TABLE I: Values of ΔAIC, ΔBIC, $\ln B$ with respect to the ΛCDM model and C_b for both models.

model	ΔAIC	ΔBIC	$\ln B$	C_b
ΛCDM	0	0	0	0.7 ± 0.1
Int ΛCDM	1.28	7.9	1.7 ± 0.4	1.5 ± 0.2

conclusion there is only weak evidence in favor of the ΛCDM model. Analysing the effective number of model parameters we can conclude that the data sets used in evaluation are not enough informative to constrain all model parameters: in the ΛCDM case there is nearly one parameter which is well constrained by the data while in the case of its generalisation the number of effective parameters is equal to 1.5 (whereas number of all considered parameters is equal to 4 here).

IV. CONCLUSION

We presented the methods of model comparison coming from information as well as bayesian theory. Both of them include the Occam's razor principle which states that if two models describe the observations equally well we should choose the simpler one. According to Akaike and Schwarz rule the model complexity is interpreted in the term of number of free model parameter while according to the Bayesian evidence more complex model wast greater volume of the parameter space. Finally we present the quantity which measure the effective number of model parameters which the data used in analysis can constrain. This quantity is also called the Bayesian complexity and reduce to the number of free model parameters when the data set is highly informative. We use those methods to answer the question if we should prefer the generalisation of the simplest ΛCDM model where the interaction between dark matter and dark energy sector is allowed. The AIC and bayesian evidence give similar results: there is only weak evidence to favor the ΛCDM model over the interacting ΛCDM one. This is with contrary with the conclusion from the analysis of the BIC quantity: here there is a strong evidence to favor the ΛCDM model. The analysis of the Bayesian complexity gives that data sets considered carry not enough information to constraint all parameters allows to vary in the considered models.

Acknowledgments

This work has been supported in part by the Marie Curie Actions Transfer of Knowledge project COCOS (contract MTKD-CT-2004-517186).

---

[1] A. G. Riess et al. (Supernova Search Team), Astron. J. **116**, 1009 (1998), astro-ph/9805201.

[2] D. N. Spergel et al. (WMAP) (2006), astro-ph/0603449.

[3] G. R. Dvali, G. Gabadadze, and M. Porrati, Phys. Lett. **B485**, 208 (2000), hep-th/0005016.

[4] C. Deffayet, Phys. Lett. **B502**, 199 (2001), hep-th/0010186.

[5] P. J. E. Peebles and B. Ratra, Astrophys. J. **325**, L17 (1988).

[6] R. R. Caldwell, Phys. Lett. **B545**, 23 (2002), astro-ph/9908168.

[7] M. P. Dabrowski, T. Stachowiak, and M. Szydlowski, Phys. Rev. **D68**, 103519 (2003), hep-th/0307128.

[8] A. Y. Kamenshchik, U. Moschella, and V. Pasquier, Phys. Lett. **B511**, 265 (2001), gr-qc/0103004.

[9] E. Witten (2000), hep-ph/0002297.

[10] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, 2003).

[11] H. Akaike, IEEE Trans. Auto. Control **19**, 716 (1974).

[12] H. Jeffreys, *Theory of Probability* (Oxford University Press, Oxford, 1961), 3rd ed.

[13] R. Trotta, Mon. Not. Roy. Astron. Soc. **378**, 72 (2007), astro-ph/0504022.

[14] A. R. Liddle, P. Mukherjee, D. Parkinson, and Y. Wang, Phys. Rev. **D74**, 123506 (2006), astro-ph/0610126.

[15] G. Schwarz, Annals of Statistics **6**, 461 (1987).

[16] M. P. Hobson and C. McLachlan, Mon. Not. Roy. Astron. Soc. **338**, 765 (2003), astro-ph/0204457.

[17] M. Beltran, J. Garcia-Bellido, J. Lesgourgues, A. R. Liddle, and A. Slosar, Phys. Rev. D **71** (2005), astro-ph/0501477.

[18] P. Mukherjee, D. Parkinson, and A. R. Liddle, Astron. J. **638** (2006), astro-ph/0508461.

[19] P. Mukherjee, D. Parkinson, P. S. Corasaniti, A. R. Liddle, and M. Kunz, Mon. Not. Roy.

Astron. Soc. **369**, 1725 (2006), astro-ph/0512484.

[20] A. Niarchou, A. H. Jaffe, and L. Pogosian, Phys. Rev. D **69** (2004), astro-ph/0308461.

[21] A. R. Liddle, Mon. Not. Roy. Astron. Soc. **351**, L49 (2004), astro-ph/0401198.

[22] T. D. Saini, J. Weller, and S. L. Bridle, Mon. Not. Roy. Astron. Soc. **348**, 603 (2004), astro-ph/0305526.

[23] M. V. John and J. V. Narlikar, Phys. Rev. **D65** (2002), astro-ph/0111122.

[24] D. Parkinson, S. Tsujikawa, B. A. Bassett, and L. Amendola, Phys. Rev. **D71**, 063524 (2005), astro-ph/0409071.

[25] M. V. John, Astrophys. J. **630**, 667 (2005), astro-ph/0506284.

[26] P. Serra, A. Heavens, and A. Melchiorri, Mon. Not. Roy. Astron. Soc. **379**, 169 (2007), astro-ph/0701338.

[27] M. Biesiada, JCAP **0702**, 003 (2007), astro-ph/0701721.

[28] W. Godlowski and M. Szydlowski, Phys. Lett. **B623**, 10 (2005), astro-ph/0507322.

[29] M. Szydlowski, A. Kurek, and A. Krawiec, Phys. Lett. **B642**, 171 (2006), astro-ph/0604327.

[30] M. Szydlowski and W. Godlowski, Phys. Lett. **B633**, 427 (2006), astro-ph/0509415.

[31] M. Kunz, R. Trotta, and D. Parkinson, Phys. Rev. **D74**, 023503 (2006), astro-ph/0602378.

[32] D. Parkinson, P. Mukherjee, and A. R. Liddle, Phys.Rev. **D73**, 123523 (2006), astro-ph/0605003.

[33] R. Trotta, Mon. Not. Roy. Astron. Soc. **378**, 819 (2007), astro-ph/0703063.

[34] A. R. Liddle, Mon. Not. Roy. Astron. Soc. Lett. **377**, L74 (2007), astro-ph/0701113.

[35] A. Kurek and M. Szydlowski (2007), astro-ph/0702484.

[36] G. Efstathiou (2007), arXiv:0712.1513 [astro-ph].

[37] S. Chongchitnan and G. Efstathiou, Phys. Rev. **D76**, 043508 (2007), arXiv:0705.1955 [astro-ph].

[38] M. Szydlowski, A. Krawiec, and W. Czaja, Phys. Rev. **E72**, 036221 (2005), astro-ph/0401293.

[39] D. L. Dowe, S. Gardner, and G. Oppy, Br. J. Philos. Sci. **58**, 709 (2007).

[40] W. Zimdahl, Int. J. Mod. Phys. **D14**, 2319 (2005), gr-qc/0505056.

[41] G. Olivares, F. Atrio-Barandela, and D. Pavon, Phys. Rev. **D74**, 043521 (2006), astro-ph/0607604.

[42] G. Olivares, F. Atrio-Barandela, and D. Pavon (2007), arXiv:0706.3860 [astro-ph].

[43] M. Le Delliou, O. Bertolami, and F. Gil Pedro, AIP Conf. Proc. **957**, 421 (2007),

arXiv:0709.2505 [astro-ph].

[44] M. Szydlowski, T. Stachowiak, and R. Wojtak, Phys. Rev. **D73**, 063516 (2006), astro-ph/0511650.

[45] K. Freese and M. Lewis, Phys. Lett. **B540**, 1 (2002), astro-ph/0201229.

[46] W. Godlowski, M. Szydlowski, and A. Krawiec, Astrophys. J. **605**, 599 (2004), astro-ph/0309569.

[47] F. E. M. Costa, J. S. Alcaniz, and J. M. F. Maia (2007), arXiv:0708.3800 [astro-ph].

[48] A. G. Riess, L.-G. Strolger, S. Casertano, H. C. Ferguson, B. Mobasher, B. Gold, P. J. Challis, A. V. Filippenko, S. Jha, W. Li, et al., ApJ **659**, 98 (2007), astro-ph/0611572.

[49] W. M. Wood-Vasey, G. Miknaitis, C. W. Stubbs, S. Jha, A. G. Riess, P. M. Garnavich, R. P. Kirshner, C. Aguilera, A. C. Becker, J. W. Blackman, et al., ApJ **666**, 694 (2007), astro-ph/0701041.

[50] T. M. Davis, E. Mortsell, J. Sollerman, A. C. Becker, S. Blondin, P. Challis, A. Clocchiatti, A. V. Filippenko, R. J. Foley, P. M. Garnavich, et al., ApJ **666**, 716 (2007), astro-ph/0701510.

[51] Y. Wang and P. Mukherjee, Astrophys. J. **650**, 1 (2006), astro-ph/0604051.

[52] D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H. Seo, M. Tegmark, Z. Zheng, et al., Astrophys. J. **633**, 560 (2005).

[53] J. Simon, L. Verde, and R. Jimenez, Phys. Rev. **D71**, 123001 (2005), astro-ph/0412269.

[54] L. Samushia and B. Ratra, Astrophys. J. **650**, L5 (2006), astro-ph/0607301.

[55] H. Wei and S. N. Zhang, Phys.Lett. **B644**, 7 (2007).

[56] J. Skilling, http://www.inference.phy.cam.ac.uk/bayesys/.

[57] A. Lewis, http://cosmologist.info/cosmomc/.

[58] A. Lewis and S. Bridle, Phys. Rev. **D66**, 103511 (2002), astro-ph/0205436.

[59] D. Parkinson, P. Mukherjee, and A. Liddle (http://www.cosmonest.org/).