Can information criteria fix the problem of degeneration in cosmology?

Marek Szydłowski*
Astronomical Observatory, Jagiellonian University,
Orla 171, 30-244 Krakow, Poland

and

Mark Kac Complex Systems Research Centre,
Jagiellonian University, Łojasiewicza 11, 30-348 Kraków, Poland

Adam Krawiec†
Institute of Economics, Finance and Management,
Jagiellonian University, Łojasiewicza 4, 30-348 Krakow, Poland

Paweł Tambor‡
Faculty of Philosophy, Catholic University of Lublin,
Al. Racławickie 12, 20-950 Lublin, Poland

Abstract

The information criteria are widely used to select the simple and proper specification of a model. In cosmology these criteria are used to select among the models with dark energy. These models are divided into two groups: the models with substantive dark energy and the models with a modified Friedmann equation. We show the advantages of using the consistent AIC over AIC in the problem considered. The cosmological model with cosmological constant – LCDM model – is favoured in the light of this criterion.

1. Introduction

Two trends are present, principally, in modern cosmology [1]. The first trend is called theoretical where different models are elaborated to explain observational facts, like acceleration, nature of dark matter and dark matter. All these attempts can be divided into three groups. First, we are looking for dark energy as a special form of substance (particles, fields). Second, we claim that theory of gravity which constitutes explanatory background should be modified. Third, we can postulate the violation of the cosmological principle as a source of acceleration (non-homogenous universe).

* e-mail address: marek.szydlowski@uj.edu.pl
† e-mail address: adam.krawiec@uj.edu.pl
‡ e-mail address: pawel.tambor@gmail.com

In the second trend, observational, we use observational methods to obtain astronomical data in order to estimate parameters of the theoretical models and probe theoretical hypothesis.

If we confront both trends, then we obtain many theoretical admissible models, which are in good agreement with observational data. This cognitive situation is called the degeneration problem. This degeneration problem has many faces in cosmology.

The same terms in model equations can mimic different effects. Objects of different ontologies produce the same term in the equations. For example, dark degeneracy means inability to resolve whether nature of dark energy is dynamical or is an effect of interaction between dark energy and dark matter [2]. This kind of degeneration is called the theoretical degeneration.

On the other side the observational degeneracy is a consequence of inability to discriminate between theoretical models on the ground of observational data. The statistical methods cannot resolve this problem definitely, but can restrict a class of admissible models. Here information criteria play their role [3, 4, 5, 6, 7, 8].

2. Choosing the right model

We can use some criteria in scientific practice to choose the best model:

- the oldest, the Middle Ages rule: Occam's razor principle — if two models describe the observations equally well choose the simpler one. This principle has aesthetical, as well as empirical justification.
- statistical criteria
  — goodness of fit (χ^2): favour the models with more parameters
  — information criteria (AIC and its generalization): realisation of Occam'srazor principle, because two models fit data equally well, the Akaike rule distinguishes model with smaller number of parameters.

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

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

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 choosen as the best from our set under consideration.

3. Akaike information criteria

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 minimize this quantity. It is impossible to compute the KL information directly, because it depends on truth, which is unknown. Akaike found approximation to the KL quantity, which is called the Akaike information criterion (AIC) and is given by [9]

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

where 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 ΔAIC

$$\Delta\mathrm{AIC}_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$$

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 evidence as the best model,
- $4 < \Delta\mathrm{AIC} \leq 7$ have considerably less support than the best model,
- $\Delta\mathrm{AIC} > 10$ have essentially no support with respect to the best model.

The AIC is not consistent because the true model among models considered is not always pointed out. Therefore, Bozdogan proposed the consistent version of AIC [10]

$$\mathrm{cAIC} = -2\ln L + d(\ln N + 1)$$

where L is a maximum of the likelihood function, d is a number of model parameters, N is a number of observations.

Taking into account the number of observations apart from a number of parameters, the cAIC "penalizes" stronger models with high number of parameters.

4. Models of dark energy

We consider 10 representative models divided into 2 groups. There are 5 models with substantial form of dark energy and 5 models with modified gravity. The former are presented in table 1 and the latter are presented in table 2.

Table 1: Models with a substantial form of dark energy

No.	Cosmological model
1	**ΛCDM** $w_{\mathrm{X}} = -1$ $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m},0}(1+z)^3 + (1-\Omega_{\mathrm{m},0})$
2	**constant E.Q.S.** $w_{\mathrm{X}} = w_0 < -1$ $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m},0}(1+z)^3 + (1-\Omega_{\mathrm{m},0})(1+z)^{3(1+w_0)}$
3	**dynamic E.Q.S.** $w_{\mathrm{X}} = w_0 + w_1(1-a)$ $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m},0}(1+z)^3 + (1-\Omega_{\mathrm{m},0})(1+z)^{3(w_0+w_1+1)}\exp[-\frac{3w_1 z}{1+z}]$
4	**quintessence** $\bar{w}_{\mathrm{X}}(a) = \int w_{\mathrm{X}}(a)\mathrm{d}(\ln a)/\int \mathrm{d}(\ln a) \equiv w_0 a^{\alpha}$ $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m},0}(1+z)^3 + (1-\Omega_{\mathrm{m},0})(1+z)^{3(1+w_0(1+z)^{-\alpha})}$
5	**oscillating E.Q.S.** $w_{\mathrm{X}}(z) = -1 + (1+z)^3\left\{C\cos(\ln(1+z))\right\}$ $\frac{H^2(z)}{H_0^2} = \Omega_{\Lambda,0}\exp\left((1+z)^3 D_2\cos(\ln(1+z))\right) + \Omega_{\mathrm{m},0}(1+z)^3$

Table 2: Models with modified gravity

No.	Cosmological model
6	**Interacting DE & DM** $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\mathrm{int},0}(1+z)^n + 1 - \Omega_{\mathrm{m},0} - \Omega_{\mathrm{int},0}$
7	**Bounce ΛCDM** $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m},0}(1+z)^3 - \Omega_{n,0}(1+z)^n + 1 - \Omega_{\mathrm{m},0} + \Omega_{n,0}$
8	**Cardassian** $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{r},0}(1+z)^4 + \Omega_{\mathrm{m},0}(1+z)^4\left[\frac{1}{1+z} + (1+z)^k\left(\frac{\Omega_{\mathrm{C},0}}{\Omega_{\mathrm{m},0}}\right)E(z)\right]$
9	**DGP** $\frac{H^2(z)}{H_0^2} = \left[\sqrt{\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\mathrm{rc},0}} + \sqrt{\Omega_{\mathrm{rc},0}}\right]^2, \quad \Omega_{\mathrm{rc},0} = \frac{(1-\Omega_{\mathrm{m},0})^2}{4}$
10	**Sahni-Shtanov brane I** $\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\sigma,0} + 2\Omega_{l,0} - 2\sqrt{\Omega_{l,0}}P(z)$

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

For the SNIa the likelihood function has the following form

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

where N_1 is the number of supernovae, σ_i is known, $\mu_i^{\rm obs} = m_i - M$ (m_i–apparent magnitude, M–absolute magnitude of SNIa), $\mu_i^{\rm theor} = 5\log_{10} D_{Li}$ $+\mathcal{M}$, $\mathcal{M} = -5\log_{10} H_0 + 25$ 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

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

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 $z_{\rm dec} = 1089$.

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

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

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. 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 N_2 is a number of observations, $H(z)$ is the Hubble function, H_i, z_i are observational data.

The final likelihood function is given by

$$L = L_{SN}L_RL_AL_H.$$

Table 3: Results of AIC and cAIC

Model	AIC	cAIC
1	217.88	**226.51**
2	219.89	232.83
3	**217.02**	234.27
4	217.49	234.74
5	220.01	237.26
6	219.16	236.41
7	221.88	239.13
8	218.88	231.82
9	226.87	235.50
10	221.88	239.13

The AIC and cAIC have been calculated for all models considered. The results of both information criteria are presented in table 3. In this context we can see the advantage of using the consistent AIC for discrimination of the best model. While the AIC indicated substantial support for four models: 3rd, 4th, 1st and 8th, the cAIC point out only the 1st model—the ΛCDM model.

Moreover, the cAIC exposes more models to have no support with respect to the best model. These are 5th, 7th and 10th by the cAIC, and only 5th by the AIC. The rest models have less support than the best model in the respective categories.

5. Conclusions

Our investigation shows that the degeneration problem of type I (theoretical) can be removed because the cAIC favoured standard cosmological model (ΛCDM model). Note that the "true" model of the Universe may be absent from the sample of models used in our analysis. Therefore, we see how important the theoretical trend is.

The detailed results are the following.

- The AIC shows that 4 models has substantial evidence (with model 3 being the best, models 1, 4 and 8 have AIC difference less than 2 with respect to model 3) — our degeneration problem is still not solved.
- The cAIC shows model 1 — the ΛCDM model — as the best model with no substantial support for any other model — the degeneration problem is solved.

References

[1] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).

[2] R. von Marttens, L. Lombriser, M. Kunz, V. Marra, L. Casarini, and J. Alcaniz, (2020), 2020.100490.

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

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

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

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

[7] M. Y. J. Tan and R. Biswas, Mon. Not. Roy. Astron. Soc. **419**, 3292 (2012), 1105.5745.

[8] M. Szydlowski, A. Krawiec, A. Kurek, and M. Kamionka, Eur. Phys. J. **C75**, 5 (2015), 0801.0638.

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

[10] H. Bozdogan, Journal of Mathematical Psychology **44**, 62 (2000).