Further insight into Bayesian and Akaike information criteria of the EC-decay rate oscillations*

N. Winckler[1,2], *F. Bosch*[1], *P. Bühler*[3], *T. Faestermann*[4], *P. Kienle*[†4], *Yu.A. Litvinov*[1,2], *M.S. Sanjari*[1,5], *D. Shubina*[1,2], *and the Two-Body Weak Decay collaboration*[1]
[1]GSI, Darmstadt; [2]MPI-K, Heidelberg; [3]SMI, Vienna; [4]TU, Munich; [5]EMMI, Darmstadt

Introduction

We discuss the Akaike and Bayesian information criteria obtained from the model and data set presented in [1]. Both criteria are model-selection methods that consists in penalizing the log-likelihood as a function of the parameters: $\text{IC}_i = -2\log(L(\hat{\theta}_i|\text{data}, M_i)) + A_n K_i$. The Akaike information criterion (AIC) is a measure of the relative goodness of fit of a statistical model. Under appropriate conditions, the model that minimizes the AIC corresponds to the one that minimizes the Kullback-Leibler divergence with respect to the true unknown distribution.

The Bayesian Information Criterion (BIC) stems from Bayesian probabilities. If proper conditions are satisfied, it is twice the negative logarithm of the marginal likelihood, i.e. $-2\log(P(\text{data}|M_i))$. From the BIC difference of model M_i and M_j, i.e. ΔBIC_i, an approximation of the Bayes' factor : $B_{i,j} \approx \exp(-\Delta\text{BIC}_i/2)$ can be obtained. In this sense, the minimum BIC corresponds to the best model describing the data.

Results

The AIC and BIC values have been obtained from unbinned maximum likelihood method [2] and are thus free from any approximation. The AIC and BIC values as well as their differences and weights are listed in the table. Note that ΔAIC and ΔBIC values in the table are defined with respect to the minimum IC value : $\Delta\text{IC} = \text{IC}_i - \text{IC}_{\text{min}}$. The Bayesian and Akaike weights are both defined for two models M_i, M_j as $w_i = \frac{e^{-\Delta\text{IC}_i/2}}{e^{-\Delta\text{IC}_i/2} + e^{-\Delta\text{IC}_j/2}}$. We observe that the AIC and BIC model selection methods can lead to different conclusion depending on the data.

	EC data (245 MHz res.)	β^+ data (245 MHz res.)	EC data (cap. pick-up)
N	3616	2912	2989
AIC_0	28683.5	22711.2	23718.4
AIC_1	28674.7	22710.4	23689.4
ΔAIC_0	8.78	0.8	29
w_0	1.2 %	40.1 %	$5 \times 10^{-5}\%$
w_1	98.8%	59.9%	$100(1 - w_0)\%$
BIC_0	28689.67	22717.1	23724.4
BIC_1	28699.26	22734.3	23713.4
ΔBIC_0	0	0	11
ΔBIC_1	9.6	17	0
$B_{0,1}$	121.5	4914.7	0.0041
w_0	99.2%	99.98%	0.4%
w_1	0.8%	0.02%	99.6%
α_{AIC}	$\approx 20\%$	$\approx 20\%$	$\approx 20\%$
α_{BIC}	$6.1 \times 10^{-3}\%$	$8.2 \times 10^{-3}\%$	$7.96 \times 10^{-3}\%$

Discussion: AIC vs BIC

It is *a priori* not trivial to choose between the AIC and BIC as they rely on various assumptions and asymptotic approximations, which in both cases are considered unrealistic [3]. Usually AIC prefers complex models and BIC simpler ones [3]. Although information criteria stem from different paradigms, the decision making of choosing M_1 rather than M_0 is mathematically equivalent to a likelihood ratio test in rejecting the null when:

$$-2\log(L_0/L_1) > A_n(K_1 - K_0) \tag{1}$$

where the rejection region is given by the right hand term.The likelihood ratio distribution obtained from MC-toys is in very good agreement with a χ^2_4 distribution[1]. Thus, in the case of AIC, rejecting the null translates in the frequentist interpretation to a likelihood ratio test at an α level of $P(\chi^2_4 > 6) = 0.199$. In the case of BIC the α level depends on the sample size (c.f. α_{BIC} in table). Here we observe that AIC would have, in the frequentist interpretation, a Type I error rate of about 20% while for BIC the error rates decrease as the sample size increases and are all below $10^{-2}\%$. On the other hand, the Type II error rates are in general lower for AIC than for BIC [3]. Therefore in situations where Type I error has to be avoided, the BIC is usually prefered. And vice versa, AIC is recommanded against Type II error. In the case of the resonator EC-data M_1 is favoured if AIC is used and M_0 in the case of the BIC. In order to resolve this issue, a proper alternative would be the computation of the Bayes' factor from unbinned likelihood, which would significantly reduce the number of assumptions and approximations made.

References

[1] P. Kienle et al. PLB 726 (2013) 638
[2] N. Winckler et al. GSI report 2013
[3] J. Dziak *et al.*, Penn. State U.,Tech. Rep. Ser. 12-119 (2012)

* Work supported in part by the DFG cluster of excellence "Origin and Structure of the Universe" of the Technische Universität München, the Helmholtz-CAS Joint Research Group HCJRG-108, the Alliance Program of the Helmholtz Association (HA216/EMMI), and the HIC-for-FAIR through HGS-HIRE
† deceased at January 29, 2013

[1]From the Wilks theorem one expects a χ^2_3 distribution which is not obtained from MC toys. Assuming the Wilks theorem is valid in our case, α_{AIC} reduces to 11.1% and α_{BIC} reduce slightly as well.

doi:10.15120/GR-2014-1-PNI-AP-06