Bayes' Theorem Applied to the Selection of Models in PAC*

Lars Hemmingsen and Steen Hansen

Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK-1871 FRB C, Denmark

Z. Naturforsch. **51a**, 422–426 (1995); received September 26, 1995

Using Bayes' formula, the parameters in a given model fitting data from perturbed angular correlation of γ-rays (PAC) experiments are all integrated out, giving as the result the total probability for the model. Experimental examples are given from PAC for models containing one and two nuclear quadrupole interactions, respectively. It is shown how the Bayesian formulation leads to a quantification of the probability of the correctness of the models. Furthermore, this method leads to a transparent and intuitively appealing criterion for model selection. Examples are given using PAC measurements on two proteins: liver alcohol dehydrogenase and azurin.

Key words: PAC, Bayesian methods, model selection.

1. Introduction

Bayesian methods are becoming more and more popular. This is partly due to their simple and appealing nature, but mainly to the reliable results the methods give (see e.g. [1] for a review of Bayesian methods). Bayesian methods use prior probabilities for the fitting parameters, a procedure which might seem unreasonable. But prior probabilities are always used when analysing experimental data. This is done explicitly when a model is chosen, or implicitly (or perhaps even inconsciously) when a prior probability enters the calculations without appearing as such. Bayesian methods are also useful for the selective integration over parameters of a given model to improve the precision of the remaining parameters of the model, as has been shown by Bretthorst [2] for nuclear magnetic resonance experiments. In perturbed angular correlation (PAC)-spectroscopy one often wants to know whether a model assuming only one nuclear quadrupole interaction (NQI) is sufficient to fit the measured data. In this paper it is shown that the Bayesian method helps deciding which of two proposed models should be chosen.

2. Theory

2.1. Bayesian model selection

The most likely model is found by calculating the total probability for each model. This is demonstrated below, using Bayes' formula and Gaussian probability densities for the parameters of the models. The total probability of each model is then calculated by integration over the parameters.

When applied to the problem of finding which hypothesis H_i (or model) is most plausible after data $\boldsymbol{d}=(d_1,\ldots,d_M)$ have been measured, Bayes' theorem gives for the probability $P(H_i|\boldsymbol{d})$:

$$P(H_i|\boldsymbol{d})=P(\boldsymbol{d}|H_i)\cdot P(H_i)/P(\boldsymbol{d}), \tag{1}$$

where $P(\boldsymbol{d}|H_i)$ is the probability of the data $\boldsymbol{d}$, assuming that the model H_i is correct, $P(H_i)$ denotes the prior probability for the model H_i which is assumed constant for "reasonable" hypotheses (i.e. different prior probabilities should not be ascribed to different reasonable hypotheses), and $P(\boldsymbol{d})$ is the probability for measuring the data, but this term is reduced to a renormalization constant after the data have been measured.

For a given model H_i having parameters $\boldsymbol{f}=(f_1,\ldots,f_N)$ the most likely $\boldsymbol{f}$ is found using Bayes' theorem again:

$$P(\boldsymbol{f}|H_i,\boldsymbol{d})=\frac{P(\boldsymbol{d}|H_i,\boldsymbol{f})\,P(\boldsymbol{f}|H_i)}{P(\boldsymbol{d}|H_i)}. \tag{2}$$

The normalising constant $P(\boldsymbol{d}|H_i)$ is usually called the *evidence* for the hypothesis H_i.

When the most likely set of *parameters* for a given model H_i into be determined, the evidence can be ignored. This is often referred to as "the first level of inference" [1]. However, to determine which of several *models* is the most likely (at "the second level of infer-

* Presented at the XIIIth International Symposium on Nuclear Quadrupole Interactions, Providence, Rhode Island, USA, July 23–28, 1995.

Reprint requests to L. Hemmingsen.

0932-0784 / 96 / 0500-0422 $ 06.00 © – Verlag der Zeitschrift für Naturforschung, D-72072 Tübingen

ence"), (1) and 2 lead to

$$P(H_i|\boldsymbol{d}) \propto P(\boldsymbol{d}|H_i) = \int P(\boldsymbol{d}|H_i, \boldsymbol{f})\, P(\boldsymbol{f}|H_i)\, \mathrm{d}^N f, \quad (3)$$

where the likelihood is

$$P(\boldsymbol{d}|H_i, \boldsymbol{f}) = \exp(-L)/Z_L, \quad (4)$$

$$Z_L = \int \exp(-L)\, \mathrm{d}^M d\,. \quad (5)$$

For the usual case of Gaussian errors, $L = \chi^2/2$ and $Z_L = \prod(2\pi\sigma_i^2)^{1/2}$, where σ_i is the standard deviation of the Gaussian noise at data point i. χ^2 is defined in the conventional manner, i.e.

$$\chi^2 = \sum_{i=1}^{M} \frac{(d_i - y_i)^2}{\sigma_i^2}, \quad (6)$$

where $\boldsymbol{y}(\boldsymbol{f})$ denotes the modelled data.

Assuming similar Gaussian distributions, the prior probability for the distribution $\boldsymbol{f}$ can be written as

$$P(\boldsymbol{f}|H_i) = \exp(S)/Z_S \quad (7)$$

with

$$S = -\sum_{j=1}^{N} \frac{(f_{j,\,\mathrm{prior}} - f_{j,\,0})^2}{2 s_j^2} \quad (8)$$

and

$$Z_S = \int \exp(S)\, \mathrm{d}^N f = \prod(2\pi s_j^2)^{1/2}, \quad (9)$$

where $\boldsymbol{f}_{\mathrm{prior}}$ and $\boldsymbol{f}_0$ are the estimates of the parameters in the model before and after the data are measured, respectively, and s_j defines the limits within which the parameter f_j is known to be found prior to the experiment. For the application of this method the exact choice of $\boldsymbol{f}_{\mathrm{prior}}$ and $\boldsymbol{s}$ is usually not very important.

It is noted that by using the specific choice for the prior probability in (8), S is equal to the second order approximation to the conventional expression for the entropy of a distribution when the distribution is non-negative and $s_j = \sqrt{f_{j,\,\mathrm{prior}}}$.

Calculating the evidence from the above equations, one obtains

$$P(\boldsymbol{d}|H_i) = \frac{\int \exp(S) \exp(-\chi^2/2)\, \mathrm{d}^N f}{Z_S\, Z_L} \approx \frac{\exp(S_0)\, f \exp(-\chi^2/2)\, \mathrm{d}^N f}{Z_S\, Z_L}, \quad (10)$$

assuming the prior distribution to be constant over the region where the posterior is maximal. For $\mathbf{B} = \nabla\nabla\chi^2/2$, (10) gives the expression for the evidence of a model:

$$P(\boldsymbol{d}|H_i) = \frac{\exp(S_0)\,(2\pi)^{N/2} \exp(-\chi_0^2/2) \det(\mathbf{B})^{-1/2}}{\prod(2\pi s_j^2)^{1/2} \prod(2\pi\sigma_i^2)^{1/2}} = \frac{\exp(S_0) \det(\mathbf{B})^{-1/2} \exp(-\chi_0^2/2)}{\prod(s_j) \prod(2\pi\sigma_i^2)^{1/2}}. \quad (11)$$

From (11) the normalised probability for a model can be calculated. For several possible models the one having the largest probability should be selected. In (11) the last factor, $\exp(-\chi_0^2/2)$, is proportional to the likelihood (the quality of the fit), and the $\prod(2\pi\sigma_i^2)^{1/2}$ is constant for a given data set. The remaining part of the fraction has been termed the "Occam factor" as it is this part which penalises the probability if more parameters are introduced in a model. Even though an increasing number of parameters will give a better fit to the data, the Occam factor will ensure that the total probability (as given by (11)) decreases if more than a "best" (most likely) number of parameters is introduced. If the parameters are almost independent, $\det(\mathbf{B})^{-1/2}$ can be approximated by the product of the posterior uncertainties of the parameters $\prod(\Delta_j)$ (given by the fitting procedure) which for $S_0 \approx 0$ gives the evidence as

$$P(\boldsymbol{d}|H_i) \propto \exp(-\chi_0^2/2) \prod\left(\frac{\Delta_j}{s_j}\right), \quad (12)$$

From (12) it is obvious how the Occam factor in the form of $\prod(\Delta_j/s_j)$ is balanced against the quality of the fit $\exp(-\chi_0^2/2)$. $\prod(\Delta_j/s_j)$ is the ratio of the accessible volume of the parameter space after and before the experiment, respectively. The negative logarithm of this ratio is equal to the information content of the experiment.

In the examples shown below, (11) has been used for calculations of the evidence.

3. Application to PAC

3.1. PAC

In γ-ray PAC spectroscopy the NQI of the nuclear quadrupole moment with the electric field gradient (EFG) from the surrounding charge distribution is measured. In the PAC-experiments referred to in this work the isotope ^{111m}Cd was used. The "angular correlation" refers to the fact that the two γ-rays are not emitted in random directions with respect to each other. Furthermore, if the nucleus interacts with extranuclear fields during the decay, the angular correlation is perturbed. The perturbed angular correla-

tion can be measured and provides information about the local charge distribution around the PAC-isotope. By using appropriate combinations of coincidence count-rates in the detectors, the experimental data are described by the following time dependent function, (see e.g. [4]):

$$A_2^{\text{eff}}\, G_2(\omega_0, \eta, \delta, \lambda; t) + c_0 \qquad (13)$$
$$= A_2^{\text{eff}}\, e^{-(\lambda t)} \left(a_0 + \sum_{i=1}^{3} a_i \cos(b_i\, \omega_0\, t) \cdot \exp\left\{\frac{-(b_i \omega_0 t_{\text{res}})^2}{16 \ln 2}\right\} \exp\left\{-\tfrac{1}{2}(b_i \omega_0 \delta t)^2\right\} \right) + c_0 ,$$

where A_2^{eff} is the measured amplitude of the anisotropy, ω_0 is proportional to the numerically largest element of the diagonalised EFG-tensor and η is the so-called asymmetry parameter, also determined by the EFG-tensor [4]. The parameters a_i and b_i are dependant on η [3], and c_0 is the zero point level of the data. The Brownian motion of the molecules in solution gives rise to a rotational correlation time, τ_c, which is fitted through $\lambda = 1/\tau_c$. Small variations in the coordination geometry of the PAC-isotope from one molecule to the next give rise to a spread of ω_0-values, which is fitted with a Gaussian distribution centered at ω_0 having the width δ. The $\exp\left\{\frac{-(b_i \omega_0 t_{\text{res}})^2}{16 \ln 2}\right\}$-term takes into account the limited time resolution of the experimental setup, with $t_{\text{res}} = 1$ ns. Equation (13) is valid under the assumptions that the PAC-isotope occupies the same site in all the randomly oriented molecules and that the molecules move sufficiently slow to apply the socalled adiabatic approximation [4].

If the fit includes more than one NQI, it is given by a sum of terms. For example the function fitted to the data assuming two NQIs is $c_0 + A_{2,1}^{\text{eff}}\, G_2(\omega_{0,1}, \eta_1, \delta_1, \lambda_1; t) + A_{2,2}^{\text{eff}}\, G_2(\omega_{0,2}, \eta_2, \delta_2, \lambda_2; t)$.

A priori unrestrictive limits may be set for the parameters $\boldsymbol{f}$ to be fitted, and in this work the following values have been used: $-0.5 < c_0 < 0.5$, $0.0 < A_2^{\text{eff}} < 0.2$, $0.0 < \eta < 1.0$, $0.0\ \text{Mrad/s} < \omega_0 < 1000\ \text{Mrad/s}$, $0.0 < \delta < 0.5$, $0.0\ \text{ns}^{-1} < \lambda < 1.0\ \text{ns}^{-1}$. These values have been chosen on the basis of a very large number of PAC-experiments, spanning a wide range of experimental conditions, and they determine the s_j in (11).

3.2. One or two NQIs

The addition of an extra NQI leads to an additional five parameters in the fitting function. The precision of the fitted parameters is typically of the order of 1% (i.e. $\Delta_j/s_j \approx 0.01$ in (12)), leading to a difference in the Occam factor of the order of 0.01^5 between the two models. As the quality of the fit is usually expressed through the value of the reduced chi-square, χ_r^2, the criterion for acceptance of the increased complexity of the two-NQI model is that χ_r^2 has to decrease by about $\frac{2}{M} 5 \log 0.01$, where M is the number of data points. For the examples shown below, M is about 200. This leads to a reduction in χ_r^2 of about 0.2 to accept the additional NQI (or a reduction in $\chi^2/2$ – the exponent in (11) – of about 20).

However, if the two-NQI fit is much "broader" than the one-NQI fit, in the sense that the Occam factor is larger for the two-NQI fit, the two-form model should be preferred even if it lowers the χ^2 only marginally, as the "weaker" statement about the two-NQI model is less informative than the more precise statement about the one-NQI model. Clearly, this will rarely be the case when analysing PAC spectra.

The application of Bayes' theorem is of course only correct to the extent that the underlying assumptions are correct, most importantly, for example, that the noise of the experiment as well as the prior for the parameters have been estimated correctly. If the noise level of the experiment is estimated incorrectly, this will influence the choice of model as the absolute value of the chi-square is used in the calculations. E.g. a 20% error in the estimate of the standard deviation for the measured data will change the difference in the chi-squares between two models by 44%. This implies that the reduction in $\chi^2/2$ necessary to accept an additional NQI is changed (from 20 as mentioned above) by about 9. Similarly, changing the width of the prior for the parameters by a factor of two changes the Occam factor by exp(3.5). Consequently, if the standard deviation for the measured data is known to about 20% precision and the prior for the parameters is assumed to be known within a factor of two, then the log(evidence) is only known with a precision of 10–15 and the difference in log(evidence) between two models should be larger than this to distinguish between the models. For PAC this means that if the difference in log(evidence) is smaller than 10–15, the simpler model having fewer parameters should be preferred.

3.3. Examples

For the case of PAC a box function will impart the prior information about the fitting parameters better

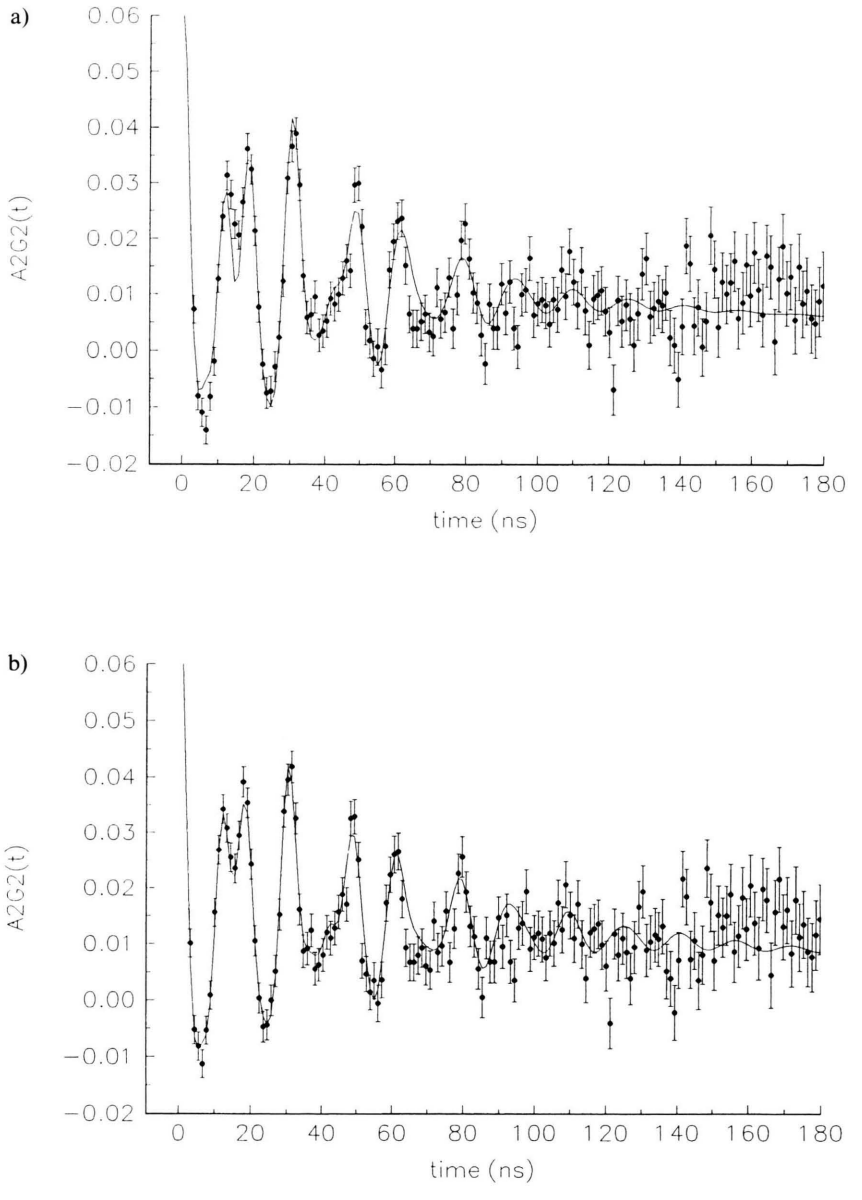

Fig. 1.
a) The experimental PAC-spectrum for the Met121Asn mutant of azurin and the fit (full line) assuming only one NQI to be present.
b) The experimental PAC-spectrum for the Met121Asn mutant of azurin and the fit (full line) assuming two NQIs to be present. The first three data points are not shown nor fitted to, because of the presence of large systematic errors for these points.

than a Gaussian, and consequently a box function of width s_j has been used for S in (11) and the following examples.

Example 1: The enzyme horse liver alcohol dehydrogenase (EC 1.1.1.1) catalyses the reversible oxidation of alcohols using NAD^+ as coenzyme. It contains a zinc ion at the active site which is bound to two cysteines, one histidine and one solvent molecule. In the resting state of the enzyme the solvent molecule is H_2O. The pK_a of this water molecule has been determined by PAC to be 11.0 ± 0.2 for the cadmium substituted enzyme [5]. This was done by recording spectra at several different pH values. At low pH only a single NQI is expected (corresponding to metalbound H_2O), but at pH values near the pK_a two NQIs should be found, and then again at high pH only a single NQI is expected (corresponding to metalbound OH^-). In

Table 1. LADH at different pH-values. The difference in χ_r^2 and log(evidence), respectively, between fits using 1 and 2 NQIs.

pH	$\Delta\chi_r^2$	Δlog(evidence)
7.9	0.00	−13
9.4	0.14	− 5
10.3	0.81	69
11.2	1.51	129
11.5	0.29	17

Table 2. Azurin and the Met121Asn mutant. Difference in χ_r^2 and log(evidence), respectively, between fits using 1 and 2 NQIs.

Sample	$\Delta\chi_r^2$	Δlog(evidence)
native azurin	0.02	−9
Met121Asn mutant of azurin	0.32	15

Table 1 the method presented in this paper has been applied to the experimental data recorded at different pH-values. The variation of Δlog(evidence) is in good agreement with expectation, i.e. the evidence for two NQIs peaks strongly around the pK_a. Furthermore, it is found that the criterion for preferring a fit with two NQIs to a fit with only a single NQI, is a change in χ_r^2 of 0.15–0.30, which is also in good agreement with experience.

Example 2: The protein azurin participates in electron transfer processes in some Gram negative bacteria [6]. It contains a copper ion which is bound to three strong ligands, one cysteine and two histidines. Furthermore, two groups, one glycine and one methionine, interact weakly with the copper ion [7]. In this work we have studied the PAC spectra of cadmium substituted native azurin as well as the Met121Asn mutant of it [8]. The PAC spectrum recorded for the Met121Asn mutant is shown in Fig. 1 along with the fits assuming one and two NQIs, respectively. The interesting question asked here is whether the metal ions all have the same coordination geometry, or if two different coordination geometries are present. The result of applying the Bayesian method is that there is only one NQI present for the native protein, a fact which has been shown previously [8], whereas it is likely that two NQIs are present for the Met121Asn mutant, see Table 2. This is in good agreement with the fact that the presence of two NQIs has been observed by PAC-spectroscopy for several other Met121Xxx (Xxx being an aminoacid) mutants of azurin.

4. Conclusion

It has been shown that the Bayesian framework can be used for model selection in perturbed angular correlation of γ-rays spectroscopy. Using prior knowledge about the acceptable range of the fitted parameters and comparing this to the standard deviation and the quality of the fit, a quantitative criterion for the acceptance of a more complex model is derived.

[1] D. J. C. MacKay, in Maximum Entropy and Bayesian Methods, Seattle 1991, edited by C. R. Smith et al, pp. 39–66 (1992). Kluwer Academic Publishers, Netherlands.
[2] G. L. Bretthorst, in Maximum-Entropy and Bayesian Methods in Science and Engineering (Vol. 1), edited by G. J. Erickson and C. Ray Smith, pp. 173–187 (1988). Kluwer Academic Publishers, Dordrecht.
[3] T. Butz, Hyperfine Int. **52,** 189 (1989).
[4] E. Danielsen and R. Bauer, Hyperfine Int. **62,** 311 (1990).
[5] L. Hemmingsen, R. Bauer, M. J. Bjerrum, M. Zeppezauer, H. W. Adolph, G. Formicka, and E. Cedergren-Zeppezauer, Biochemistry **34,** 7145 (1995).
[6] G. Karlsson, Ph.D.-thesis, Dep. of Biochem. and Biophys., Chalmers University of Technology and University of Göteborg, Sweden 1993.
[7] M. D. Lowery and E. I. Solomon, Inorg. Chim. Acta **198–200,** 233 (1992).
[8] E. Danielsen, R. Bauer, L. Hemmingsen, M.-L. Andersen, M. J. Bjerrum, T. Butz, W. Tröger, G. W. Canters, C. W. G. Hoitink, G. Karlsson, Ö. Hansson, and A. Messerschmidt, J. Biol. Chem. **270,** 573 (1995).