



HAL
open science

An introduction to Bayesian statistics for atomic physicists

M. Trassinelli

► **To cite this version:**

M. Trassinelli. An introduction to Bayesian statistics for atomic physicists. Journal of Physics: Conference Series, 2020, 1412, pp.062008. 10.1088/1742-6596/1412/6/062008 . hal-02885446

HAL Id: hal-02885446

<https://hal.science/hal-02885446>

Submitted on 30 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

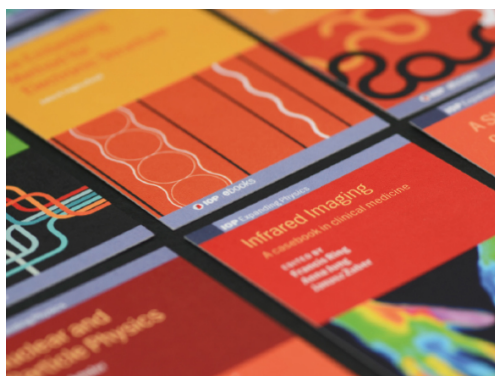
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PAPER • OPEN ACCESS

An introduction to Bayesian statistics for atomic physicists

To cite this article: M. Trassinelli 2020 *J. Phys.: Conf. Ser.* **1412** 062008

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

An introduction to Bayesian statistics for atomic physicists

M. Trassinelli

Institut des NanoSciences de Paris, INSP, CNRS, Sorbonne Université, F-75005 Paris, France

E-mail: martino.trassinelli@insp.jussieu.fr

Abstract. We present an introduction to some concepts of Bayesian data analysis in the context of atomic physics. Starting from basic rules of probability, we present the Bayes' theorem and its applications. In particular we discuss about the limits of classic statistics methods and in which cases a Bayesian analysis is mandatory. Moreover, we show how probability values can be assigned to different possible models (different numbers of peaks, type peak profile, etc.) from the analysis of experimental data.

1. Introduction

The standard analysis of atomic spectra consist on comparing n pairs $\{x_i, y_i\}$ of measured values (or $\{x_i, y_i, \sigma_i\}$ triplets if error bars are available) to the values of a function $F(x_i, \mathbf{a})$ used to model the data, where \mathbf{a} are the function parameters. The common practice to determine the set of parameters \mathbf{a}^{best} that describes at best the data is to find the maximum of the likelihood function

$$L(\mathbf{a}) = P(\{x_i, y_i, \sigma_i\}|\mathbf{a}) = \prod_{i=1}^n p(x_i, y_i, \sigma_i|\mathbf{a}), \quad (1)$$

where $p(x_i, y_i, \sigma_i|\mathbf{a})$ are the single conditional probability distribution for each channel to have x_i, y_i, σ_i for a given parameter values \mathbf{a} . If the data are normally distributed for each channel, we have that $p(x_i, y_i, \sigma_i|\mathbf{a}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i - F(x_i, \mathbf{a}))^2}{2\sigma_i^2}\right]$ and the equation above becomes

$$L(\mathbf{a}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{[y_i - F(x_i, \mathbf{a})]^2}{2\sigma_i^2}\right) \propto \exp\left(-\sum_{i=1}^n \frac{[y_i - F(x_i, \mathbf{a})]^2}{2\sigma_i^2}\right). \quad (2)$$

Its maximization corresponds then to minimization of the function $\chi^2 = \sum_{i=1}^n [y_i - F(x_i, \mathbf{a})]^2 / \sigma_i^2$ i.e. the chi-square function.

The maximum likelihood and chi-square minimization methods we just recalled are in mostly of the cases well adapted and sufficient for a correct data analysis. However they have some limitations that one should be aware:

- i) Cause-effect inversion:** The maximum likelihood method is based on the probability $L(\mathbf{a}) = P(\{x_i, y_i, \sigma_i\}|\mathbf{a})$ to obtain certain data values for a given set of parameters \mathbf{a} . But the final goal is just the opposite: for a given experimental data, one would like to evaluate the probability for having certain parameter values.



- ii) **Constraints on the parameter values:** Hard constraints on parameter values (eg. a mass that cannot be negative) or simply prior parameter distribution from previous experimental results cannot be taken into account in a coherent way.
- iii) **Assignment of probability to models:** If several models (hypotheses) can describe the experimental results, it is not possible assign probabilities to them from the analysis of the data set.

A remark on the last point, several classical statistics criteria are available to choose the most plausible model, like the χ^2 -test, the likelihood-ratio test, etc. [1–6]. However, in the unfortunate case where there is no a clear propensity to a specific model and we are interested on a parameter common to all models (as the position of the a peak with undefined shape), the value of the most probable of the parameter of interest (and its associated standard deviation) cannot be extracted with classical statistics criteria.

To overcome these problems, a different approach has to be implemented with a new and more general definition of probability based on propositional logic. The probability is here a positive number that indicates the propensity that a certain proposition X is true, where X can indicates a numerical value, eg. “the peak position is between x_1 and x_2 ”, or a more abstract concept, eg. “the model \mathcal{M} correctly describe the data”. Because of this general formulation, a probability function defined in this way can be applied to many different cases. In particular, as we will show in the next section, it can simply solve the three main problems of commonly used classic statistics methods (also called *frequentist* because of the definition of probability based on the repetitively of tries) mentioned above.

At present, Bayesian methods are routinely used in many fields: cosmology [7–9], particle physics [10], machine learning [11], . . . They start to be applied only in recent years to nuclear physics [12] and there are very few applications in atomic physics (see eg. Refs. [13–15]) with almost no use in atomic spectroscopy. In the present article we would like to fill this gap and provide a short introduction of Bayesian data analysis methods in the context of atomic physics. A general and more extended introduction and application to physics cases can be found elsewhere [16–19].

2. Probability definition and Bayes’ theorem

2.1. A general definition of probability

When basic logic and consistency are required, the probability function defined by propositions X, Y, \dots has to respect some basic properties [16, 20–22] (which are compatible to Kolmogorov probability axioms [23]):

$$0 \leq P(X|I) \leq 1, \quad (3a)$$

$$P(X|X, I) = 1, \quad (3b)$$

$$P(X + Y|I) = P(X|I) + P(Y|I) - P(X, Y|I), \quad (3c)$$

$$P(X, Y|I) = P(X|Y, I) \times P(Y|I) = P(Y|X, I) \times P(X|I). \quad (3d)$$

Here the probability $P(X|Y, I)$ means that X is true knowing that (“|”) that Y is true and were I represents the current state of knowledge (knowledge of well tested physics laws as eg.). The joint probability $P(X, Y|I)$ means that both “ X AND Y ” are true (equivalent to the logical conjunction ‘ \wedge ’). The symbol ‘+’ means the logical disjunction ($X + Y \equiv X \vee Y \equiv$ “ X OR Y is true”). When a set of mutual exclusive assertions are considered $\{Y_i\}$, with $P(Y_i|Y_{j \neq i}) = 0$, a consequence of Eqs. (3) is the so-called *marginalization rule*

$$P(X|I) = \sum_i P(X, Y_i|I) = P(X|I) = \int_{-\infty}^{\infty} P(X, Y|I) dY. \quad (4)$$

where the last term is valid in the limit of continuous case $Y_{i+1} - Y_i \rightarrow dY$.

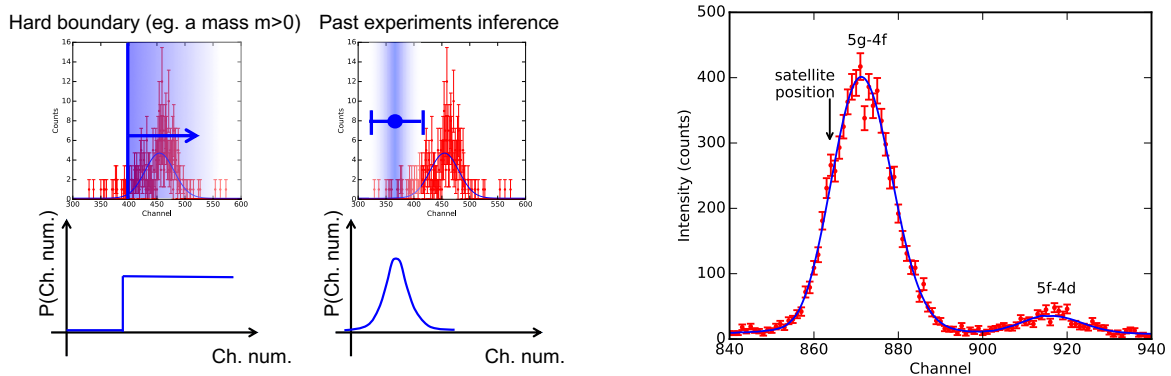


Figure 1. Left: Graphical representation of hard boundary from physical constraints (eg. a mass value that has to be positive and inference from past experiments in a given spectrum. In both cases, these constraints can be considered via the prior probability $P(a|I)$ of the different model parameter values (plots at the bottom). Right: Pionic nitrogen 5 – 4 transitions. Possible additional transitions from the presence of one remaining electron in the K shell are indicated.

2.2. Bayes' theorem and posterior probability

An important corollary of Eq. (3d) the so-called called the Bayes' theorem

$$P(X|Y, I) = \frac{P(Y|X, I) \times P(X|I)}{P(Y|I)}, \quad (5)$$

named after Rev. Thomas Bayes, who first formulated theorems of conditional probability before 1761 [24], and rediscovered in 1774 and further developed by Pierre-Simon Marquis de la Laplace [25].

For a better insight in the implication of this theorem, we consider the particular case where X represent the hypothesis that the parameter values set \mathbf{a} truly describes the data (via the function $F(x, \mathbf{a})$) and where Y correspond to the recorded data $\{x_i, y_i\}$. In this case, we have that the posterior probability of a certain set of parameter set is

$$P(\mathbf{a}|\{x_i, y_i\}, I) = \frac{P(\{x_i, y_i\}|\mathbf{a}, I) \times P(\mathbf{a}|I)}{P(\{x_i, y_i\}|I)} = \frac{L(\mathbf{a}) \times P(\mathbf{a}|I)}{P(\{x_i, y_i\}|I)}, \quad (6)$$

where I includes our available background information and where $P(\{x_i, y_i\}|\mathbf{a}, I)$ is by definition the likelihood function $L(\mathbf{a})$ for the given set of data. In addition to the likelihood function, we have here the term $P(\mathbf{a}|I)$ that includes the prior probability on the parameters \mathbf{a} , which can includes possible boundaries. With the above formula, we solve in fact at once the two problems i) and ii) discussed in the introduction. As suggested in Fig. 1 (left), if we have hard boundaries a_{min}, a_{max} , $P(a|I) = 1/(a_{max} - a_{min})$ is considered. In the case of inference of previous measurement with result $a' \pm \sigma_{a'}$, the prior parameter probability is
$$P(a|I) = \frac{1}{\sqrt{2\pi}\sigma_{a'}} \exp\left[-\frac{(a-a')^2}{2\sigma_{a'}^2}\right].$$

2.3. Model testing and Bayesian evidence

An important consequence of the Bayes' theorem is the possibility to assign probabilities to different hypotheses (models) with a simple and well-defined procedure. In this case, X in Eq. (5) represents the hypothesis that the model \mathcal{M} describes well the measured data and Y

represents the data themselves, as in the previous section. In particular case of atomic spectra, different hypotheses could correspond to different numbers of spectral component, presence of satellite lines or not, or different profiles modeling (Gaussian, Lorentzian, Voigt, ...). From Bayes' theorem, we have that the posterior probability of the model \mathcal{M} is [8, 16, 21]

$$P(\mathcal{M}|\{x_i, y_i\}, I) \propto P(\{x_i, y_i\}|\mathcal{M}, I) \times P(\mathcal{M}|I), \quad (7)$$

where the first term of the right part is the so-called *Bayesian evidence* E of the model (also called *marginal likelihood* or *model likelihood*) and the second term is the prior probability assigned to the model from our background knowledge. Using the marginalization rule (Eq. (4)) and the probability properties (Eqs. (3a–3d)), we have

$$E(\mathcal{M}) \equiv P(\{x_i, y_i\}|\mathcal{M}, I) = \int P(\{x_i, y_i\}|\mathbf{a}, \mathcal{M}, I)P(\mathbf{a}|\mathcal{M}, I)d^N \mathbf{a} = \int L^{\mathcal{M}}(\mathbf{a})P(\mathbf{a}|\mathcal{M}, I)d^N \mathbf{a}, \quad (8)$$

where N is the number of the parameters of the considered model, and where we explicitly indicate the dependency of the likelihood function $L^{\mathcal{M}}(\mathbf{a})$ on the related model \mathcal{M} .

Considering equal priors, the probability of a model is higher if the evidence $E(\mathcal{M})$ is higher, which means that the likelihood function integral over the model parameter space is higher. This does not imply that the maximum of the likelihood function is larger. Considering maximal values of $L^{\mathcal{M}}(\mathbf{a})$ only can in fact introduce errors in the analysis, as recently discussed in the particular case of nuclear physics in Ref. [12]. Model selection criteria based on it, like the likelihood-ratio test and the Akaike information criterion, should be used then with precautions.

Probabilities evaluated from Eq. (7) are generally lower for models with higher number of parameters because of the higher dimensionality of the integral that corresponds to a larger parameter volume (and then to a lower average value of the likelihood function). This means that when the values of the likelihood function are similar, simple models (with a small number of parameters) are favored, accordingly to Ockham's razor principle.

The possibility to assigning probabilities to models has another important advantage. In the case we are interested to determine the probability distribution of a common parameter a_j without the need to identify the most probable model among the available choices \mathcal{M}_ℓ , we can obtain the probability distribution $P(a_j|\{x_i, y_i\}, I)$ from the weighted sum

$$P(a_j|\{x_i, y_i\}, I) = \sum_{\ell} P(a_j|\{x_i, y_i\}, \mathcal{M}_\ell, I) \times P(\mathcal{M}_\ell|I), \quad (9)$$

where $P(a_j|\{x_i, y_i\}, \mathcal{M}_\ell, I)$ are the probability distributions of a_j for each model and $P(\mathcal{M}_\ell|I)$ are the probabilities of the different models. This is particularly useful when model probabilities are similar and classic statistics criteria are powerless.

3. A simple example: a possible satellite line presence

We consider a common case in spectroscopy where we would like to test the presence or not of a weak and unresolved spectral line close to an intense line, an example already discussed in details in Ref. [19]. In this specific example, we consider the $5g - 4f$ transition in pionic nitrogen, an hydrogen-like atom formed by a nitrogen nucleus and a negatively charged pion. During the formation of the pionic atoms, all electrons are expected to be ejected. The presence of a remaining electron in the K shell cannot however completely be excluded. Its presence can cause a shift of the main transition energy due to the Coulomb screening and then an appearance of a new component in the spectrum. To determine the probability of such a scenario, we have to calculate the evidence for the two possible models: Model 1 corresponding to the situation

without remaining electrons (a pure hydrogen-like pionic atom) and Model 2 with the possible presence of one remaining electron. More details on the physics case can be found in Refs. [26,27].

The examined data consists in seven distinct spectra similar to the one represented in Fig. 1 (right) with a total of about 60000 recorded counts. Compared to model 1, model 2 has only one additional free parameter of the satellite line intensity, whose relative position with respect to the main line is fixed by theory.

The probabilities of the two hypotheses are calculated from the corresponding evidence E (Eq. (8)) of the models (calculated by `Nested_fit` program [19,28]) and they are resumed in Tab. 1 together with some parameter model values.

Table 1. Evidence, probability and parameters values of the models 1 and 2 without and with a satellite line, respectively. The analysis is obtained with `Nested_fit` program [19,28] and partially reported in previous publications [26,27].

	Model 1	Model 2	Average value
log(Evidence)	-5031.00 ± 0.23	-5039.41 ± 1.82	
Probability	$99.98^{+0.02\%}_{-0.12\%}$	$0.02^{+0.12\%}_{-0.02\%}$	
Relative satellite amplitude	0.	$(9.8 \pm 2.2) \times 10^{-3}$	2.2×10^{-6}
Main line shift (ch.)	0.	0.079 ± 0.048	1.8×10^{-5}

As it can be observed, in the considered case the probability for having a satellite line is very low (only 0.02%). In addition using Eq. (9), the average of common parameter values can be calculated from the single most probable values relative to each model and the probabilities of the models themselves $P(\mathcal{M}_{1,2}|I)$. This feature of Bayesian methods is, to the opinion of the author, one of the most useful tools for the analysis of complex or statistically-poor atomic spectra. Even if the probability of the two models was similar, a correct evaluation of the main peak position would have been possible.

4. Discussions and conclusions

In the previous sections we present the limitation of classic (frequentist) statistics analysis with respect to Bayesian statistics methods. Classic analysis can be safely used in many cases but one should be aware on its limitation. Classic methods cannot correctly take into account prior probabilities of parameters of the considered fit function (model/profile). Moreover, the study of maximal (minimal) value of likelihood (chi-square) function, instead of its behavior over the range of parameter values, can induce errors in the analysis as non-existing correlations and miss presence of multiple maxima (minima). These problems are avoided when the prior probabilities is included via Bayes' theorem (Eq. 6) and when the entire likelihood function $P(\mathbf{a}|\{x_i, y_i\}, I) = L(\mathbf{a})$ is considered.

For model selection, classic statistics methods provide only criteria to choose one model with respect to others, without the possibility to assign a probability to the different models. This is particularly annoying when these criteria fail to clearly indicate a preference for one particular model. With the Bayesian approach, this difficulty becomes trivial. The probability distribution (and mean, standard deviations, confidence intervals, etc.) of the different parameters that are common to the different models can be obtained from the probability distribution relative each model and the model probability.

Even if tremendously advantageous, Bayesian statistics methods have a drawback, partially responsible to the limited use among the scientific community. The evaluation of the Bayesian

evidence and, more generally, the evaluation of the likelihood function $L(\mathbf{a})$ in the N -dimensional parameter space (where N is the number of the model parameter) can be very computationally costly. $L(\mathbf{a})$ is generally non analytic and its values and integral can be studied with Monte Carlo methods. In present days, when performing multicore computers are more and more available, the use of Bayesian data analysis libraries (eg. BAT [29], PyMC3 [30]) and programs (eg. CosmoMC [7], Multinest [9], Polychord [31], Nested_fit [19, 28]) is more and more easy to implement and it should be used as much as possible. This message is addressed in particular to the atomic physics community, where Bayesian methods are still rarely used.

References

- [1] Neyman J and Pearson E 1933 *Phil. Trans. R. Soc.* **231** 289–337
- [2] Yates F 1934 *Sup. J. R. Stat. Soc.* **1** 217–235 ISSN 14666162
- [3] Bevington P and Robinson D 2003 *Data reduction and error analysis for the physical sciences* (McGraw-Hill) ISBN 9780072472271 URL <https://books.google.fr/books?id=0poQAQAIAAJ>
- [4] Massey F J 1951 *J. Am. Stat. Assoc.* **46** 68–78 ISSN 0162-1459
- [5] Akaike H 1974 *IEEE Trans. Automat. Contr.* **19** 716–723 ISSN 0018-9286
- [6] Spiegelhalter D J, Best N G, Carlin B P and Van Der Linde A 2002 *J. R. Stat. Soc. B* **64** 583–639 ISSN 1467-9868 URL <http://dx.doi.org/10.1111/1467-9868.00353>
- [7] Lewis A and Bridle S 2002 *Phys. Rev. D* **66** 103511
- [8] Trotta R 2008 *Contemp. Phys.* **49** 71–104
- [9] Feroz F, Hobson M P and Bridges M 2009 *Mon. Not. R. Astron. Soc.* **398** 1601–1614 ISSN 1365-2966
- [10] Particle Data Group 2018 *Phys. Rev. D* **98** 030001
- [11] Barber D 2012 *Bayesian Reasoning and Machine Learning* (Cambridge University Press) ISBN 9780521518147
- [12] King G, Lovell A, Neufcourt L and Nunes F 2019 *Phys. Rev. Lett.* **122** 232502
- [13] Stockton J K, Wu X and Kasevich M A 2007 *Phys. Rev. A* **76** 033613
- [14] Calonico D, Levi F, Lorini L and Mana G 2009 *Metrologia* **46** 267 ISSN 0026-1394
- [15] Mooser A, Kracke H, Blaum K, Bräuninger S A, Franke K, Leiteritz C, Quint W, Rodegheri C C, Ulmer S and Walz J 2013 *Phys. Rev. Lett.* **110** 140405
- [16] Sivia D S and Skilling J 2006 *Data analysis: a Bayesian tutorial* 2nd ed (Oxford University Press)
- [17] von der Linden W, Dose V and von Toussaint U 2014 *Bayesian Probability Theory: Applications in the Physical Sciences* (Cambridge University Press) ISBN 9781139952460
- [18] von Toussaint U 2011 *Rev. Mod. Phys.* **83** 943–999
- [19] Trassinelli M 2017 *Nucl. Instrum. Methods B* **408** 301–312 ISSN 0168-583X
- [20] Cox R T 1946 *Am. J. Phys.* **14** 1–13
- [21] Jaynes E and Bretthorst G 2003 *Probability Theory: The Logic of Science* (Cambridge University Press) ISBN 9781139435161
- [22] Ballentine L E 1986 *Am. J. Phys.* **54** 883–889 URL <http://scitation.aip.org/content/aapt/journal/ajp/54/10/10.1119/1.14783>
- [23] Kolmogorov A 1956 *Foundations of the theory of probability* (Chelsea Pub. Co.)
- [24] Bayes M and Price M 1763 *Philosophical Transactions* **53** 370–418 URL <http://rstl.royalsocietypublishing.org/content/53/370.short>
- [25] Laplace P 1825 *Essai philosophique sur les probabilités* (Bachelier) URL <https://books.google.fr/books?id=pY9aAAAAcAAJ>
- [26] Trassinelli M, Anagnostopoulos D F, Borchert G, Dax A, Egger J P, Gotta D, Hennebach M, Indelicato P, Liu Y W, Manil B, Nelms N, Simons L M and Wells A 2016 *Phys. Lett. B* **759** 583–588 ISSN 0370-2693 URL <http://www.sciencedirect.com/science/article/pii/S0370269316302787>
- [27] Trassinelli M, Anagnostopoulos D, Borchert G, Dax A, Egger J P, Gotta D, Hennebach M, Indelicato P, Liu Y W, Manil B, Nelms N, Simons L and Wells A 2016 *EPJ web conf.* **130** 01022
- [28] Trassinelli M 2019 *submitted to MDPI Proceedings*
- [29] Caldwell A, Kollár D and Kröninger K 2009 *Comp. Phys. Comm.* **180** 2197–2209 ISSN 0010-4655
- [30] Salvatier J, Wiecki T V and Fonnesbeck C 2016 *PeerJ Comput. Sci.* **2** e55 ISSN 2376-5992
- [31] Handley W J, Hobson M P and Lasenby A N 2015 *Mon. Not. R. Astron. Soc.* **453** 4384–4398 ISSN 0035-8711