Signal discovery in sparse spectra: a Bayesian analysis

F. Beaujean[1], *A. Caldwell*[1*], *D. Kollár*[2], *K. Kröninger*[3], *S. Pashapour*[3]
1 Max-Planck-Institut für Physik, München, Germany
2 CERN, Geneva, Switzerland
3 II Physikalisches Institut, Universität Göttingen, Germany
* Corresponding author

Abstract
A Bayesian analysis of the probability of a signal in the presence of background is described. As an example, the method was used to calculate the sensitivity of the GERDA experiment to neutrinoless double beta decay. In addition, we discuss the use of consensus priors, the look-elsewhere-effect in Bayesian analysis and other topics.

1 Introduction

Scientific knowledge, i.e., justified belief, comes from inductive reasoning. Experimental tests allow us to build our justification for believing in particular models. In the context of the models, frequency distributions can be produced and probabilities of different outcomes calculated. However, it is impossible to make a statement on the truth of the model without considering all possible models which could give similar results, and assigning prior beliefs to the models. Frequentist approaches avoid using priors and therefore in principle do not allow statements on how strongly we should believe in a particular model. Statements of belief in a model become maximally subjective - each interpreter of the data is advised to reach their own conclusions on what to believe [1]. In contrast, in the Bayesian approach the prior beliefs are explicitly stated so that posterior beliefs can be evaluated. While the posterior beliefs are also subjective, the reasoning which led to the conclusion is made clear. Given that the goal is to make a statement on how strongly we believe our models, the Bayesian approach seems to us appropriate.

2 Signal discovery in an event counting setting

Imagine we have a collections of events where we have measured some physical quantity x which can take on a continuous range of values. We assume that we have a background model, with background contribution B, for the distribution of the values of x, possibly with nuisance parameters involved, and we can predict the distribution of x values for some new physics, which could depend on parameters of interest (e.g., for a Gaussian distribution for signal events, we have some position μ, width parameter σ, and amplitude S). To proceed, we need our prior belief that the background model accounts completely for the observations, $P_0(H_1)$, and the prior belief that there could be new physics contributing to the observations, $P_0(H_2) = 1 - P_0(H_1)$. For the models, we also need the prior beliefs in the possible values of the parameters: e.g., for the 'new physics' model $P_0(\mu, S, \sigma|H_2)$. We then group the observations $\{x\}$ in intervals Δx_i and compare the predictions with the observations. Using D to represent the data, we have for the posterior belief in H_2:

$$P(H_2|D) = \frac{P(D|H_2)P_0(H_2)}{P(D|H_2)P_0(H_2) + P(D|H_1)P_0(H_1)} \tag{1}$$

where

$$\begin{aligned} P(D|H_2) &= \int P(D|\mu, S, \sigma, B)P_0(\mu, S, \sigma|H_2)P_0(B)d\mu dS d\sigma dB \\ P(D|H_1) &= \int P(D|B)P_0(B)dB \end{aligned}$$

and

$$P(D|\mu, S, \sigma, B) = \prod_i \frac{e^{-\nu_i} \nu_i^{n_i}}{n_i!} \tag{2}$$

$$P(D|B) = \prod_i \frac{e^{-\lambda_i} \lambda_i^{n_i}}{n_i!} \tag{3}$$

and

$$\lambda_i(B) = \int_{\Delta x_i} f_B(x|B) dx \tag{4}$$

$$\nu_i(\mu, A, \sigma, B) = \lambda_i(B) + \int_{\Delta x_i} f_S(x|\mu, A, \sigma) dx \tag{5}$$

with n_i the observed number of events in bin i, λ_i the expectation for bin i for the background model, and ν_i the expectation including the new physics signal given the parameter values.

3 Sensitivity analysis for GERDA

This analysis method was used to estimate the sensitivity of the GERDA experiment to neutrinoless double beta decay [2]. In the GERDA case, the location and shape of the signal are known (i.e., μ and σ above are fixed), so that the only physics parameter is the expectation for the number of signal events.

Given the lack of theoretical consensus on the Majorana nature of neutrinos and the cloudy experimental picture, the prior probabilities for H_1 and H_2 were chosen to be equal, i.e.

$$P_0(H_1) = 0.5, \tag{6}$$

$$P_0(H_2) = 0.5\,. \tag{7}$$

The prior probability for the number of expected signal events, assuming H_2, was taken flat up to a maximum value, S_{max}, consistent with existing limits[1]. It should be noted that the prior probability for H_1 depends on the maximum allowed signal rate. $S_{\max}$ was chosen so that the probability for the hypothesis H_1 is 50 %, which is a reasonable assumption. The effect of choosing a different prior for the number of signal events was studied in Ref. [2].

The overall background contribution B was chosen to be Gaussian with mean value $\mu_{\mathrm{B}} = B_0$ and width $\sigma_{\mathrm{B}} = B_0/2$. The prior probabilities for the expected signal and background contributions were taken as

$$P_0(\mu, S, \sigma|H_2) = P_0(S|H_2) = \frac{1}{S_{\max}},\ 0 \le S \le S_{\max},\ p_0(S) = 0 \text{ otherwise}, \tag{8}$$

$$P_0(B) = \frac{e^{-\frac{(B-\mu_{\mathrm{B}})^2}{2\sigma_{\mathrm{B}}^2}}}{\int_0^\infty e^{-\frac{(B-\mu_{\mathrm{B}})^2}{2\sigma_{\mathrm{B}}^2}} dB},\ B \ge 0,\ P_0(B) = 0 \text{ otherwise}\,. \tag{9}$$

Ensemble tests were then used to evaluate the sensitivity of the experiment to both signal discovery and probability limits on the half-life $T_{1/2}$ for neutrinoless double beta decay. An example data set as well as the resulting discovery sensitivity are given in Fig. 1.

[1] $S_{\max}$ was calculated assuming a half-life of $T_{1/2} = 0.5 \cdot 10^{25}$ years.

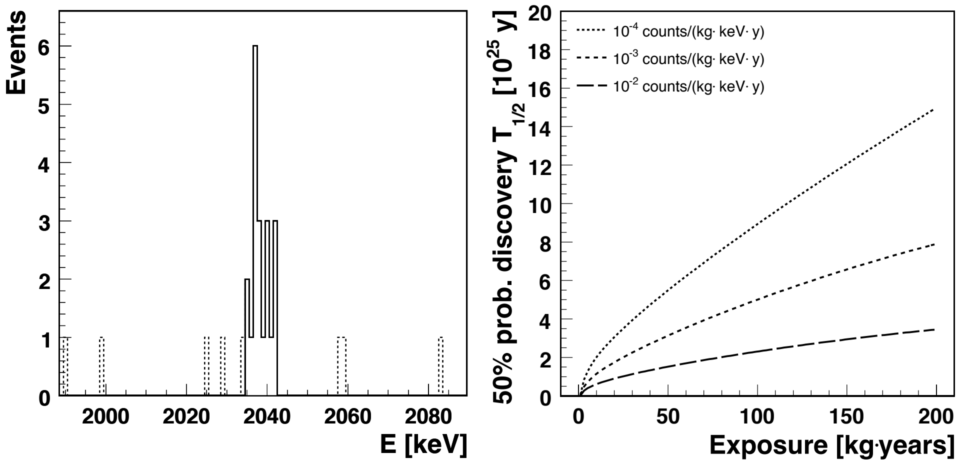

Fig. 1: Left: an example data set generated with $T_{1/2} = 2 \cdot 10^{25}$ yr, a background index of $1 \cdot 10^{-3}$/(keV·kg·yr) and an exposure of 100 kg·yr. Right: the curves indicate the half-life where an experiment would have a 50 % chance of claiming a discovery as a function of exposure, and for different background indices. Discovery was defined in [2] as $P(H_1|D) < 0.0001$.

4 Error bars

No error bars are shown in Fig. 1, since error bars on distributions of observed numbers of events are at best misleading. There is certainly no uncertainty on the number of observed events. The only uncertainty comes when the observed number of events is used to estimate the mean of the underlying Poisson distribution. There are different ways in which this mean can be extracted, and placing the estimate for the mean at the number of observed events is in any case not always the best choice. The second problem arises with the size of the error bar. This is routinely plotted as the square root of the number of events, taking the Poisson result that the variance is equal to the mean. However, this definition does not lead to an error bar which contains 68 % probability. The probability range covered varies dramatically for small numbers of events and is asymmetric around the point. This leads to great confusion when non-experts analyze data/model agreement 'by eye'. We would strongly favor ending the practice of putting error bars on the number of observed events. It is better to give no extra information than to give misleading information.

5 The Look-Elsewhere Effect (LEE) in Bayesian Analysis

There is no look-elsewhere effect in the GERDA example since the location of the signal is known. In general, the LEE is suppressed in Bayesian analysis, since a penalty is built into the prior for allowing a signal to appear in different places during a search. This is demonstrated here for a simple example of searching for a signal in a 1-D distribution. Assume that the resolution (width of the peak, σ) for the potential signal is known as well as the amplitude, but we allow a search with the location of the signal free. Define H_1 as the null hypothesis - only known backgrounds are present. H_2 is the hypothesis that in addition to the known backgrounds, there is also a signal. In this case, using μ as the location of the new physics signal, we have

$$P(H_2|D) = \frac{\int P(D|H_2,\mu)P_0(H_2,\mu)d\mu}{\int P(D|H_2,\mu)P_0(H_2,\mu)d\mu + P(D|H_1)P_0(H_1)} \quad . \tag{10}$$

where D represents the data and we assume that the null hypothesis has no free parameters.

Taking a simple example,

$$P_0(H_2,\mu) \quad = \quad P_0(H_2)P_0(\mu|H_2)$$

$$P_0(H_2) = P_0(H_1) = 1/2$$

our equation (10) becomes

$$P(H_2|D) = \frac{\int P(D|H_2,\mu)P_0(\mu)d\mu}{\int P(D|H_2,\mu)P_0(\mu)d\mu + P(D|H_1)} \quad . \tag{11}$$

Now assume we can use a flat prior for μ, given by

$$P_0(\mu) = \frac{1}{L_\mu}$$

where L_μ is the range over which the parameter can vary. Our equation further simplifies to

$$P(H_2|D) = \frac{\int P(D|H_2,\mu)d\mu}{\int P(D|H_2,\mu)d\mu + L_\mu P(D|H_1)} \tag{12}$$

The integral can be written as

$$\int P(D|H_2,\mu)d\mu = P(D|H_2,\mu^*)\delta_\mu$$

where μ^* is the parameter value which maximizes the probability of the data, and δ_μ is an effective width of the distribution $P(D|H_2,\mu)$. We expect $\delta_\mu \approx \sqrt{2\pi}\sigma$. Using these results, we find

$$P(H_2|D) = \frac{P(D|H_2,\mu^*)\delta\mu}{P(D|H_2,\mu^*)\delta\mu + L_\mu P(D|H_1)} \quad .$$

The probability $P(D|H_2,\mu^*)$ tends to grow relative to $P(D|H_1)$ as we allow searches over bigger ranges (new data sets). However, there is a penalty δ_μ/L_μ for allowing the signal to appear anywhere in the spectrum, and this will shrink as L_μ is expanded, compensating for the larger $P(D|H_2,\mu^*)$. Since the search range L_μ is presumably much greater than the resolution σ, the penalty factor can be quite small. Every additional parameter (dimension in which we search) will bring such a reduction factor.

6 p-values and incomplete sets of models

A full Bayesian analysis is only possible if we have a complete set of models. In the GERDA example, we performed a kind of either/or (background model or background+specific signal). We are often in a situation where we are not sure if we have found a complete (enough) set of models. What do we do if we want to include also other possibilities (other types of signals could be present, the background estimate could be faulty) ? We may not even know whether we should include other possibilities. A hierarchical structure can be set up as was done for the BAT solution to the BANFF challenge [3]. The logic for searching for new physics in this case is shown diagrammatically in Fig. 2. The logic is based on using p-values, and is implicitly a Bayesian argument (see [4]). It is assumed that incorrect models have p-value distributions sharply peaked at 0, so that a small p-value gives reason to believe that we have found an incorrect model. Without specifying prior beliefs, the argumentation remains vague.

7 Consensus priors

Our degree-of-belief that we have found new physics depends on both the data and the prior belief. The discussion in the physics community on how many sigmas are needed to define a discovery clearly reflects the need for the definition of consensus priors. Different signals will clearly have different priors. E.g., it would come as no great surprise to find the Higgs particle with a mass around 120 GeV. A search for the Higgs in this mass range would start with a sizeable prior belief. On the other hand, signals for

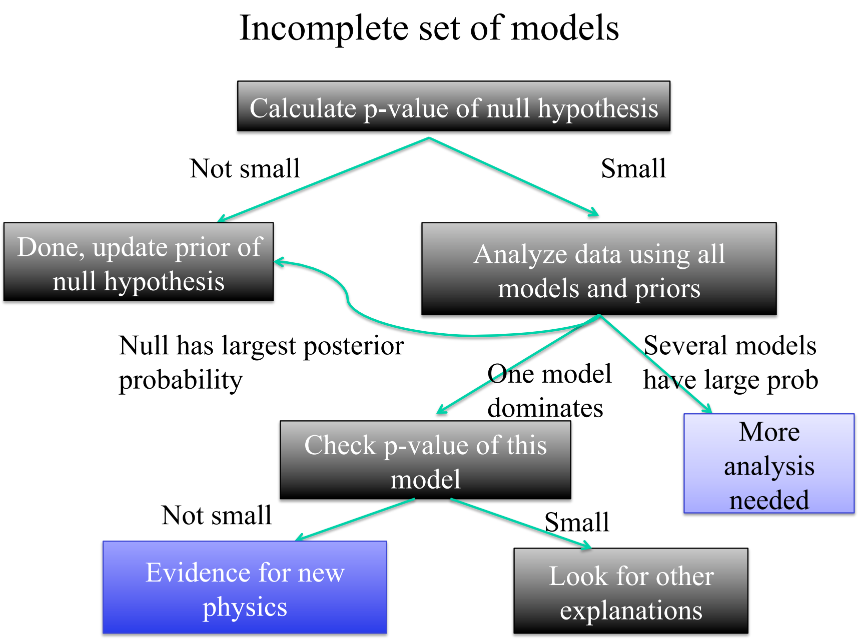

Fig. 2: The logical flow for claiming evidence for new physics (or not) in situations where we are reluctant to define a set of models summing to prior probability of one.

large extra dimensions are a priori considered much more unlikely in this mass range, and would come with a much smaller prior belief. The PHYSTAT community could be a good place to start a discussion towards consensus priors for new physics, at the LHC and also for other experimental searches (direct dark matter detection, neutrinoless double beta decay, ...). For each type of new physics searched for, both 'conservative' and 'optimistic' priors could be defined. Basing analyses on these consensus priors would allow for a transparent means of drawing conclusions on the belief in the new physics. The consensus priors would be updated as the new data came along by a representative body of the community; e.g., represented by a subcommittee of the PDG.

References

[1] See, e.g, sec 33.2.2 of K. Nakamura et al. (Particle Data Group), J. Phys. **G 37**, 075021 (2010).

[2] A. Caldwell, K. Kröninger, Phys. Rev. **D 74** 092003 (2006).

[3] S. Pashapour, 'Bayesian Analysis Toolkit for searches', these Proceedings.

[4] F. Beaujean, A. Caldwell, D. Kollár, and K. Kröninger, Phys. Rev. **D 83**, 012004 (2011).