Multi-messenger Astronomy: a Bayesian approach

G. Torralba Elipe[*]**, R. A. Vazquez and E. Zas**

Departamento de Física de Partículas & Instituto Galego de Física de Altas Enerxías,
Universidade de Santiago de Compostela, 15782, Santiago de Compostela, Spain
E-mail: guillermo.torralba@usc.es, vazquez@fpaxp1.usc.es *and*
zas@fpaxp1.usc.es

After the discovery of the gravitational waves and the observation of neutrinos of cosmic origin, we have entered a new and exciting era where cosmic rays, neutrinos, photons and gravitational waves will be used simultaneously to study the highest energy phenomena in the Universe. Here we present a fully Bayesian approach to the challenge of combining and comparing the wealth of measurements from existing and upcoming experimental facilities. We discuss the procedure from a theoretical point of view and using simulations, we also demonstrate the feasibility of the method by incorporating the use of information provided by different theoretical models and different experimental measurements.

35th International Cosmic Ray Conference — ICRC2017
10–20 July, 2017
Bexco, Busan, Korea

---

[*]Speaker.

1. Introduction

We are starting a new era for astroparticle physics. A lot of experiments are taking data simultaneously measuring different observables in a wide range of energies: Pierre Auger Observatory [1], Telescope Array [2], HAWC [3], IceCube [4], Magic [5], Antares [8]. New experiments will be developed such as Cherenkov Telescope Array [7] or KM3Net [9] and we are going to have unprecedented number of events to perform analyses that could answer the questions related with high-energy cosmic rays, neutrinos and photons from more than century ago such that: what are the cosmic rays?; where are they comming from?; how are they accelerated?. There are no doubts that for answer these questions the experiments must share their results and the answer will arrive by combining all the measurements.

In this work we present a brief review of Bayesian inference in Sec. 2, explaining the parameter estimation and hypothesis testing. The practice of these methods are shown using toy simulations in Sec. 3. First we consider that two experiments analyse different data without taking into account the results of the other experiment. After that, we consider that the experiments share their results and modify their prior information in their analyses.Finally the conclusions are reported in Sec. 4.

2. Review of Bayesian statistical inference

2.1 Parameter inference

Let $D = \{x_i\}_{i=1}^n$ be n realizations of a random variable X, *i.e,* n results of experiments consisting in measuring the variable X. Let θ be a parameter of interest. Notice that there are not restrictions on the dimensions of X and θ. The Bayesian inference consists of allocating probabilities to the possible values of θ according to the observed data set D by solving the equation

$$\pi(\theta|D,I) = \frac{f(D|\theta,I)\pi(\theta|I)}{f(D|I)} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}, \tag{2.1}$$

which is expressed in terms of probability density functions. Now we describe each term appearing in Eq. 2.1.

Likelihood function: $f(D|\theta,I)$

The likelihood function $f(D|\theta,I)$ is the conditional probability distribution of D given the unknown parameter θ and it is usually denoted as $\mathcal{L}(\theta|D)$. This function describes how the data set D is distributed assuming a given value of θ. The likelihood function expresses all information obtainable for the data satisfaying the *Likelihood principle*: All the information about θ that can be obtained from an experiment is contained in the likelihood function for θ given X. Two likelihood functions for θ (from the same or different experiments) contain the same information about θ if they are proportional to one another, see [10] and [11]. In [11] it is also shown that the likelihood principle is derived by the assumption of two principles: the *principle of sufficiency* and the *principle of conditionality*. These principles can be described informally as asserting the "irrelevance of observations independent of a sufficient statistic" (sufficiency) and the "irrelevance of experiments not actually performed" (conditionality).

The prior: $\pi(\theta|I)$

It describes all the information that we have about the parameter of interest before performing the experiment. A prior distribution can be created using information about past experiments, using theoretical knowledge or expressing our total ignorance about the problem. When we do not have information about the parameter of interest one should follow the *Laplace criterion rule* paraphrased as: "in the abscence of any further information (prior information) all possible results should be considered equally probable". This kind of prior is the so called "flat prior".

The posterior: $\pi(\theta|D,I)$

This function describes our knowledge about the θ parameter after the data analysis of the experimental results. Then one can read Eq. 2.1 as an update of the prior knowledge of θ, described by the prior, through the experiment described by the likelihood. For each event $x_i \in D$ of the data set, our knowledge about θ changes. Once the posterior distribution is known there are two standard estimators for the true value of θ: the mean of the posterior and the mode (the so called **M**aximum of **A P**osteriori distribution, MAP).

The evidence: $f(D|I)$

Also denoted as Z acts as a normalization constant in the parameter inference but takes an important role in the *Bayesian Model Selection* explained in Sec. 2.3. The evidence is given by:

$$Z = \int f(D|\theta,I)\pi(\theta|I)d\theta. \tag{2.2}$$

2.2 Confidence intervals

The confidence intervals or credible sets (here denoted as C.I) are easy to calculate in the Bayesian approach. Once the posterior distribution is known we want to find between which values $[\theta_1, \theta_2]$ the actual value of the parameter has been estimated. Usually this question is answered with an associated probability q which is typically 0.68, 0.9 and 0.95. The limits of the range are given by solving the equation

$$q = P(\theta_{low} \leq \theta \leq \theta_{up}) = \int_{\theta_{low}}^{\theta_{up}} \pi(\theta|D,I)d\theta. \tag{2.3}$$

When the inferred value of θ equal or near to one of the limits of the possible values of θ, one talk about upper or lower limits depending if $\theta \approx \theta_{min}$ or $\theta \approx \theta_{max}$.

2.3 Bayesian model selection

Consider now two hypotheses I_1 and I_2 that we want to constrast and we perform an experiment which gives us the data set $D = \{x_i\}_{i=1}^{n}$. We are going to consider that the likelihood functions are different for the different hypotheses, for I_1 we have $\mathscr{L}_1(\theta|D) = f_1(D|\theta)$ and for I_2 we have $\mathscr{L}_2(\omega|D) = f_2(D|\omega)$ where θ and ω could in principle have different dimensions (θ could be for

instance a shape of an exponential distribution and ω could be the mean and the variance of a normal distribution). The posterior distributions are given by

$$\pi(\theta|D,I_1) = \frac{\mathscr{L}_1(\theta|D)\pi(\theta|I_1)}{Z_1} \tag{2.4}$$

for the first hypothesis and

$$\pi(\omega|D,I_2) = \frac{\mathscr{L}_2(\omega|D)\pi(\omega|I_2)}{Z_2} \tag{2.5}$$

for the second hypothesis. Z_1 and Z_2 are the normalization factors for their respective equations:

$$Z_1 = \int f_1(D|\theta)\pi(\theta|I_1)d\theta = P(D|I_1), \tag{2.6}$$

which gives the probability of the data set D given the hypothesis I_1 (once $P(D|I_k)$ has been normalized to all the hypotheses). In the same way, Z_2 is the probability of D given the hypothesis I_2. The evidences have statistical meaning. Since we can calculate $P(D|I_1)$ and $P(D|I_2)$ we can also calculate $P(I_1|D)$ and $P(I_2|D)$ using the Bayes' theorem obtaining the probability of a given hypothesis given the data set D and independently of the parameters θ and ω:

$$P(I_m|D) = \frac{P(D|I_m)P(I_m)}{P(D)} = \frac{Z_m P(I_m)}{\sum_{l=1}^{M} Z_l P(I_l)}, \tag{2.7}$$

where here $M = 2$ and $m = 1,2$. The expression shown in Eq. 2.7 is the generalization for M possible hypotheses.

Once more the prior probabilities $P(I_1)$ and $P(I_2)$ must be chosen before the analysis. In this way, we obtain a probability mass function in which the variables are the different hypotheses. To compare which of the hypotheses is preferred by data, the ratio between the posterior probabilities is performed:

$$\frac{P(I_1|D)}{P(I_2|D)} = \frac{Z_1}{Z_2}\frac{P(I_1)}{P(I_2)}. \tag{2.8}$$

This ratio is called "posterior odds" and the ratio $P(I_1)/P(I_2)$ is called "prior odds". The ratio of the evidences Z_1/Z_2 is called the *Bayes' factor* of the hypothesis I_1 over I_2 ($B_{1,2}$) and represents the gain of probability of I_1 over the hypothesis I_2 after the data analysis:

$$\text{posterior odds } (I_1,I_2) = B_{1,2} \times \text{prior odds } (I_1,I_2). \tag{2.9}$$

2.4 Predictive distributions

Suppose that an observer wants to prepare an experiment to infer certain parameter θ which can take values in the Θ space with prior probabilities $\pi(\theta,I)$. The distribution of the random variable X is given by the likelihood function $f(x|\theta,I)$. The data distribution before the experiment is

$$f(\tilde{x}|I) = \int_{\Theta} f(\tilde{x}|\theta,I)\pi(\theta|I)d\theta \tag{2.10}$$

where $\tilde{x}$ denotes unobserved data. $f(\tilde{x}|I)$ is called the prior predictive distribution. After the experiment has been built and the data D analysed, the knowledge about θ has changed: $\pi(\theta,I) \to \pi(\theta|D,I)$. Now the expected data distribution has also changed:

$$f(\tilde{x},I) \to f(\tilde{x}|D,I) = \int_{\Theta} f(\tilde{x}|\theta,I)\pi(\theta|D,I)d\theta \tag{2.11}$$

where $f(\tilde{x}|D,I)$ is called the posterior predictive distribution. This distribution can be used to compare with the observed data distribution to get a feeling of how well the estimation of θ fits the measured data or for future experiments.

3. Simulations

Let E_X and E_Y be two experiments measuring different observables X and Y. The experiments are interested in to measure the fraction of certain distribution (*signal*) that there is in their data. As an example, X can be the proton fraction of cosmic rays at ultra-high energies while Y can be the astrophysical photon or neutrino fractions at energies in the PeV region. Let M_1 and M_2 two models predicting different signals both in X and Y and predicting different relations between the signals as it is illustrated in Fig. 1. In our example $\alpha_y = \alpha_x^{1.2}$ for M_1 and $\alpha_y = \alpha_x^2$ for M_2.
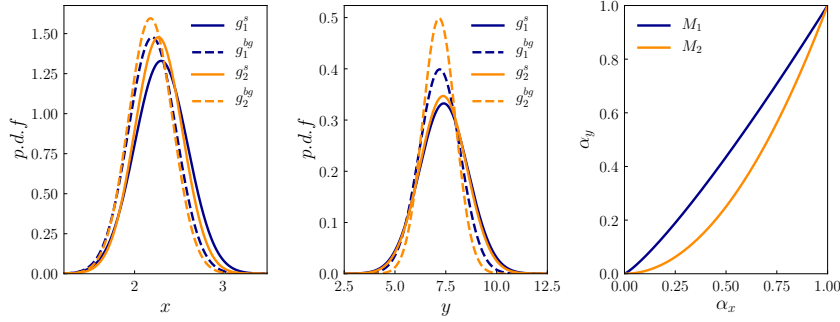

Figure 1: Signal and background distributions (continuous and dashed lines) predicted from M_1 (blue) and M_2 orange for the two experiments: E_X in the left panel and E_Y in the center. The fraction of the signal in E_Y as a function of the signal in E_X is shown in the right panel for the two models.

The signal and background are normal distributions (denoted by g^s and g^{bg} respectively) for the two models with the following parameters:

$$
M_1 \begin{cases} g_1^s(x): \mu = 2.3 \text{ and } \sigma = 0.3 \\ g_1^{bg}(x): \mu = 2.2 \text{ and } \sigma = 0.27 \\ g_1^s(y): \mu = 7.4 \text{ and } \sigma = 1.2 \\ g_1^{bg}(y): \mu = 7.2 \text{ and } \sigma = 1 \end{cases} \quad M_2 \begin{cases} g_2^s(x): \mu = 2.27 \text{ and } \sigma = 0.27 \\ g_2^{bg}(x): \mu = 2.18 \text{ and } \sigma = 0.25 \\ g_2^s(y): \mu = 7.37 \text{ and } \sigma = 1.15 \\ g_2^{bg}(y): \mu = 7.17 \text{ and } \sigma = 0.8 \end{cases} \tag{3.1}
$$

We simulate two data samples (one for each experiment) following the model M_1 with $\alpha_x^{true} = 0.3$. E_X measures 300 events and E_Y measures 200 events. For these simulations we have $\langle x \rangle = 2.23$ and $\sigma_x = 0.3$; $\langle y \rangle = 7.27$ and $\sigma_y = 1.07$. The likelihood function for the model M_i ($i = 1,2$) and variable z ($z = x, y$) is given by:

$$
f(z|\alpha_z, M_i) = \alpha_z g_i^s(z) + (1 - \alpha_z) g_i^{bg}(z). \tag{3.2}
$$

Now we perform two analyses: one where each experiment analyse the data without any kind of information (Sec. 3.1) and another one where the experiments use the information obtained from the other (Sec. 3.2).

3.1 Independent analyses

In this approach the experiments have no any prior information but they are interested in the fraction of the signal, then the fraction of signal plus the fraction of background must be one. For this reason, each experiment choose a uniform distribution between 0 and 1 as its prior.
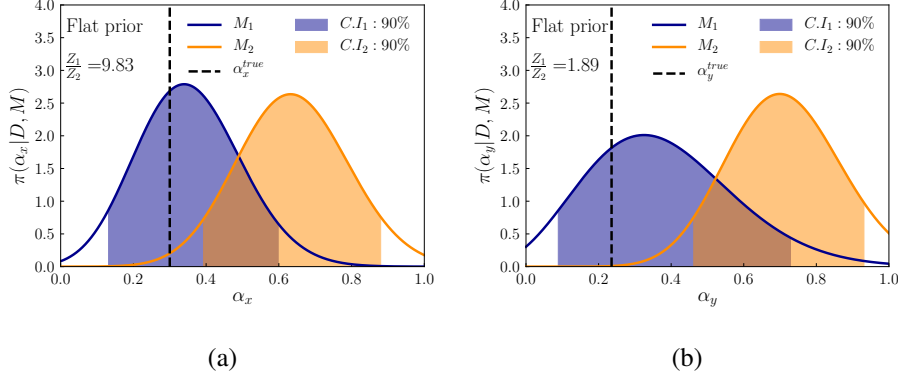

Figure 2: Posterior probability distributions of E_X (a) and E_Y (b) for the fraction of the interesting signal assuming the different models: M_1 (blue) and M_2 (orange).

In Fig. 2 the posterior probabiltiy distributions of the signals for each experiment under the assumption of the different models are displayed. We obtain numerically that E_X obtains $\langle \alpha \rangle_x = 0.33$ and a C.I at 90% $[0.13, 0.6]$ assuming M_1 while assuming M_1 E_X obtains $\langle \alpha \rangle_x = 0.6$ and $[0.4, 0.9]$ as the posterior mean value of the fraction of the signal and C.I respectively. E_Y obtains $\langle \alpha \rangle_y = 0.38$ and C.I $0.09, 0.74$ assuming M_1 and $\langle \alpha \rangle_y = 0.62$ as fraction of the interesting signal with $[0.46, 0.93]$ as a C.I assuming M_2. Since each experiment assumes $P(M_1) = P(M_2) = 0.5$, before the analysis, E_X arrives to the conclusion that M_1 is almost ten times most probable than M_2 while the resolution to discriminate between the models in E_Y is smaller and for this experiment $P(M_1|D)/P(M_1|D) \sim 2$.

3.2 Combined analyses

When one experiment has analysed some data, its prior knowledge change, and these change can be use for the same experiment to analyse new data or for another experiment. In this example we show how the results of each experiment is used by the other. Assuming the results of E_X in the previous section E_Y can modify the prior of α_y for each theoretical model or scenario. In the same way, E_X can do the same in sight of the analysis done by E_Y. These new priors are shown together with the new results in Fig. 3.
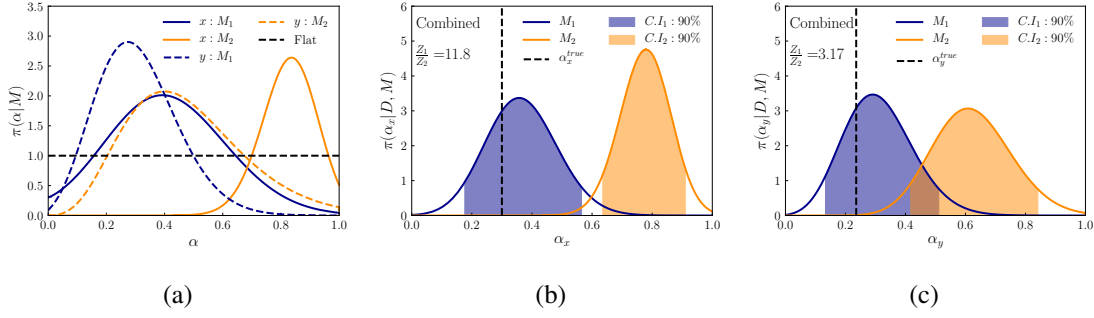
Figure 3: Prior probabilities for the signals given the independent analysis in panel (a): prior for α_x given the results of E$_Y$ assuming M_1 (M_2) is shown as continuous blue (orange) line; prior for α_y given the results of E$_X$ assuming M_1 (M_2) are shown as dashed blue (orange) line. They are compared with the uniform prior taken in the independent analysis (black dashed line). The posterior distributions for each experiment are shown in panels (b) and (c) for E$_X$ and E$_Y$ respectively.

One can observe that by including the results of one experiment in the other the results change. Now E$_X$ obtains that the posterior odds in favour of M_1 are: $P(M_1|D)/P(M_2|D) = 11.8 \times 1.89 \approx 22$, increasing the evidence in favour of the model M_1. When E$_Y$ analyse its data taking into account the results of E$_X$ the posterior odds also increase being now $P(M_1|D)/P(M_2|D) \approx 31$. Therefore both experiments have reasons to beleave that the true model is M_1 and the joined results will be $\langle \alpha_x \rangle = 0.36$ and $\langle \alpha_y \rangle = 0.31$ with C.I $[0.18, 0.56]$ and $[0.13, 0.51]$ respectively being M_1 at least 22 times more probable than M_2.

Finally, the posterior predictive distributions taking the results of the combined analysis are shown in Fig. 4. Even though the data can be well described by the two models, the Bayesian combined analysis permits us distinguish numerically between the two models.
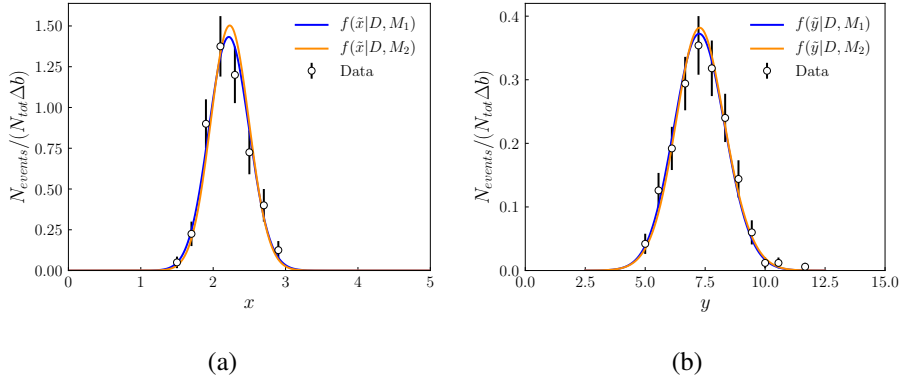

Figure 4: Posterior predictive distributions for E$_X$ (a) and E$_Y$ (b) compared with the observed data.

4. Conclusions

The Bayesian approach for the combination of different measurements and detectors has been presented and tested using simulations. With these methods the estimation of the parameters of interest and the discrimination among different theoretical models or scenarios can be improved using past or present experimental results from different experiments.

In this work we show how the combination of the information obtained with two different detectors can improve the parameter estimation, reduce the uncertainty and distinguish between theoretical models that can explain the same data.

References

[1] Pierre Auger Collaboration, https://www.auger.org/

[2] Telescope Array Project, www.telescopearray.org/

[3] The High-Altitude Water Cherenkov Gamma-Ray Observatory, http://www.hawc-observatory.org/

[4] IceCube Neutrino Observatory, https://icecube.wisc.edu/

[5] MAGIC Collaboration, https://magic.mpp.mpg.de/

[6] ANTARES Collaboration, http://antares.in2p3.fr/

[7] Cherenkov Telescope Array, https://www.cta-observatory.org/

[8] ANTARES Collaboration, http://antares.in2p3.fr/

[9] KM3Net Collaboration, https://www.km3net.org/

[10] J. O. Berger, R. L. Wolpert, M. J. Bayarri, M. H. DeGroot, B. M. Hill, D. A. Lane and L. LeCam *The Likelihood Principle*, *IMS Lecture Notes–Monograph Series* (1988) Institute of Mathematical Statistics

[11] A. Brihnbaum, *On the foundations of statistical inference*, *Journal of the American Statistical Association*, **57** (1962)