Sensitivity studies in the bayesian approach

Stefano Piacentini

7/12/2020

Introduction

- In the last days we worked on the development of the framework for the Bayesian analysis.
- We will see how it is possible to define a ‘3σ sensitivity’ and a ‘90 % credible interval’ in this approach, and we will show that the results are compatible with what has been done so far in the frequentist approach.
- As a computational tool we used the software **JAGS** (Just Another Gibbs Sampler), a software which implements the MCMC Gibbs Sampler algorithm in **R** and is interfaced with **Python** by the package **PyJags**.

Bayes theorem

[arXiv:2006.02453]

We recall that the posterior p.d.f for the parameters of the model can be computed by means of the Bayes theorem as:

$$p(r_S, \boldsymbol{\theta} \,|\, \{x_i\}, H_{r_S}) = \frac{p(\{x_i\} \,|\, r_S, \boldsymbol{\theta}, H_{r_S})\pi(r_S, \boldsymbol{\theta} \,|\, H_{r_S})}{\int_\Omega \int_0^\infty p(\{x_i\} \,|\, r_S, \boldsymbol{\theta}, H_1)\pi(r_S, \boldsymbol{\theta}|H_1) dr_S d\boldsymbol{\theta}}, \quad (13)$$

with

$$\mathcal{L}(r_s, \boldsymbol{\theta}) \equiv p(\{x_i\} \,|\, r_S, \boldsymbol{\theta}, H_{r_S}). \quad (14)$$

The marginal p.d.f. of the parameter of interest r_S is given by:

$$p(r_S \,|\, \{x_i\}, H_{r_S}) = \int_\Omega p(r_S, \boldsymbol{\theta} \,|\, \{x_i\}, H_{r_S}) d\boldsymbol{\theta}, \quad (15)$$

and similarly for any other parameter of the model.

Odds Ratio

[arXiv:2003.03340]

Suppose we want to compare two different models H_i and H_j, which for the moment we assume don't depend on any parameter. To compare these models we can compute the ratio of the posterior probabilities:

$$O_{ij} \equiv \frac{p\left(H_i|D,I\right)}{p\left(H_j|D,I\right)}, \tag{A.2}$$

this ratio gives the odds in favour of one model over the other.

Bayes Theorem

$$O_{ij} \equiv \frac{p\left(D|H_i,I\right)}{p\left(D|H_j,I\right)} \times \frac{\pi\left(H_i|I\right)}{\pi\left(H_j|I\right)}$$

Bayes Factor

[arXiv:2003.03340]

The data dependent part is the so called "Bayes factor" (BF), and it is defined as

$$BF_{ij} \equiv \frac{p\left(D|H_i, I\right)}{p\left(D|H_j, I\right)}. \tag{A.5}$$

In the hypothesis in which, due to our a priori ignorance, all the model's priors are equal, considering the odds ratio is equivalent to considering the Bayes factor.

How to compute number of σ from H_0

[arXiv:0803.4089]

To gain some intuition about how the Bayes factor works, consider two competing models: $\mathcal{M}_0$ predicting that a quantity $\theta = 0$ with no free parameters, and $\mathcal{M}_1$ which assigns θ a Gaussian prior distribution with 0 mean and variance Σ^2. Assume we perform a measurement of θ described by a normal likelihood of standard deviation σ, and with the maximum likelihood value lying λ standard deviations away from 0, i.e. $|\theta_{\max}/\sigma| = \lambda$. Then the Bayes factor between the two models is given by, from Eq. (20)

$$B_{01} = \sqrt{1 + (\sigma/\Sigma)^{-2}} \exp\left(-\frac{\lambda^2}{2(1 + (\sigma/\Sigma)^2)}\right). \tag{21}$$

Returning to the example of Eq. (21), if the data are informative with respect to the prior on the extra parameter (i.e., for $\sigma/\Sigma \ll 1$) the logarithm of the Bayes factor is given approximately by

$$\ln B_{01} \approx \ln(\Sigma/\sigma) - \lambda^2/2, \tag{26}$$

Then we can compute B_{01} and σ to get λ.

How to compute number of σ from H_0

[arXiv:0803.4089]

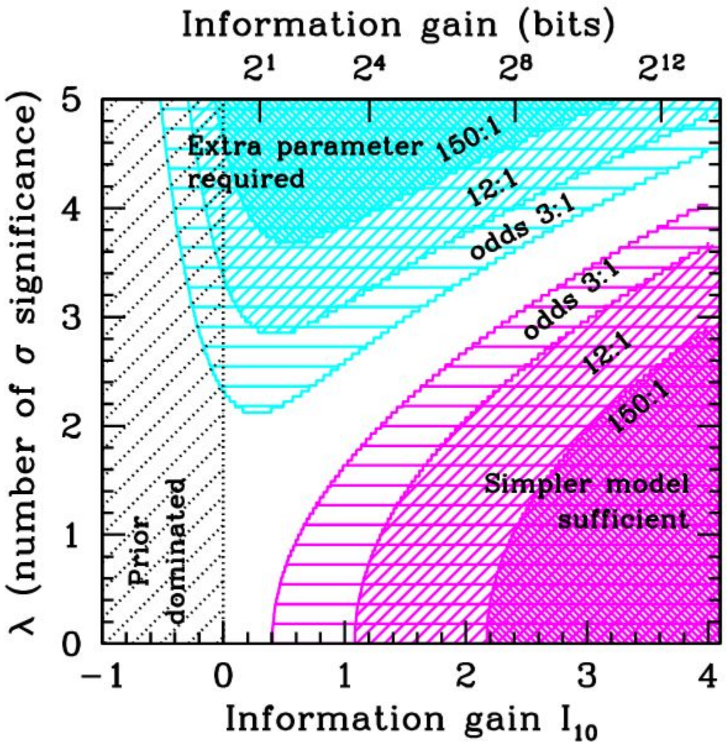


Table 2. Translation table (using Eq. (27)) between frequentist significance values (p–values) and the upper bounds on the odds ($\overline{B}_{10}$) in favour of the more complex model. No other choice of prior (within the family considered in the text) will give higher evidence in favour of the extra parameters. The "sigma" column is the corresponding number of standard deviations away from the mean for a normal distribution. The "category" column gives the Jeffreys' scale of Table 1 (from [65]).

p–value	$\overline{B}_{10}$	$\ln \overline{B}_{10}$	sigma	category
0.05	2.5	0.9	2.0	
0.04	2.9	1.0	2.1	'weak' at best
0.01	8.0	2.1	2.6	
0.006	12	2.5	2.7	'moderate' at best
0.003	21	3.0	3.0	
0.001	53	4.0	3.3	
0.0003	150	5.0	3.6	'strong' at best
6×10^{-7}	43000	11	5.0	

$$I_{10} \equiv \log_{10}(\Sigma/\sigma).$$

$$B_{01} = \sqrt{1 + (\sigma/\Sigma)^{-2}} \exp\left(-\frac{\lambda^2}{2(1 + (\sigma/\Sigma)^2)}\right)$$

How to compute limits [90% C.I.]

[arXiv:2006.02453]

The marginal p.d.f. of the parameter of interest r_S is given by:

$$p(r_S \,|\, \{x_i\},\, H_{r_S}) = \int_\Omega p(r_S,\, \boldsymbol{\theta} \,|\, \{x_i\},\, H_{r_S}) d\boldsymbol{\theta}, \tag{15}$$

We compute the upper bound for the DM signal as the 90% Credible Interval (C.I.). This is defined as the value of $\sigma_{SI}(m_\chi)$ corresponding to the 90% quantile of the posterior p.d.f. for r_S:

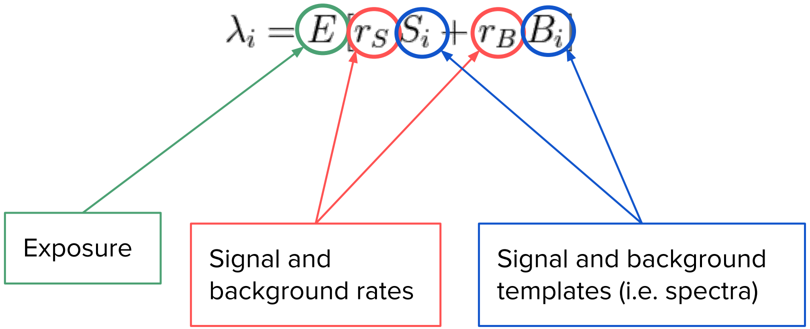
$$r_S(90\%\,\text{C.I.}) : \int_0^{r_S(90\%\,\text{C.I.})} p(r_S \,|\, \{x_i\},\, H_{r_S})\, dr_S = 0.9. \tag{16}$$

In the Bayesian approach the upper bound is a statement on the true value of the parameter of interest. The quantity $r_S(90\%\,\text{C.I.})$ has to be interpreted as the value below which we believe at 90% probability level the true value of r_S lies, given the present experimental information.

Likelihood

$$\mathcal{L}\ (r_S, r_B, \boldsymbol{\theta}; \{x_i\}) \equiv p(\{x_i\} \,|\, r_S, r_B, \boldsymbol{\theta}, H) = \prod_{i=1}^{N_{bin}} \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i},$$

$$\lambda_i = E[r_S S_i + r_B B_i]$$




```
model {

   for (i in 1:n) {


      x[i] ~ dpois(lambda[i])

      lambda[i] <- t * mub * (bkg_f[i]) +
                   t * mus * (sig[i])

   }

   # prior on overall background
   mub     ~ dpois(bkg_mean)

   # prior on signal is quite vague
    mus     ~ dnorm(0.0, 1/(3.0 * sqrt(0.5 * mubmax))^2);
}
```

Differences between the two approaches

- In the **frequentist approach μ_S and μ_B are not random variables**, and you have to work on toy data samples assuming a certain value of μ_S and μ_B and computing the distribution of your test statistics (i.e. the log likelihood ratio) on your dataset {x} :

$$L(\{x\} \mid \mu_S , \mu_B)/L(\{x\} \mid \mu_S = 0, \mu_B).$$

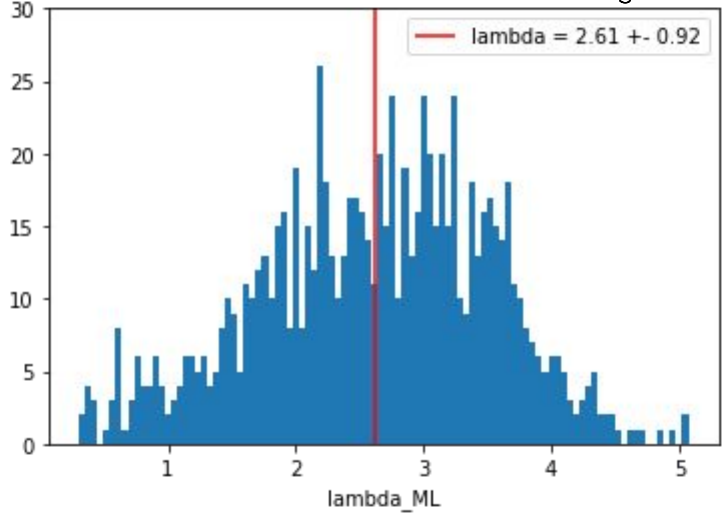
- In the Bayesian approach once you have a data sample, you want to infer μ_S and μ_B and get their posterior p.d.f. in the two models (i.e. signal + bkg and bkg only), from which you can compute mean, std, quantiles, Bayes factor, …
- The Bayes factor is proportional to the marginal likelihood ratio and it could be "compared" with the frequentist LR, although they are different by definition:
  - μ_S and μ_B are extracted from each data sample
  - the likelihood is marginalized on all the possible values of the parameters and not maximized, profiled, …
  - since marginalizing means taking the average of the likelihood in the parameter space, the result will be worse than the maximum likelihood approach by definition

Comparison with the previous analysis - I

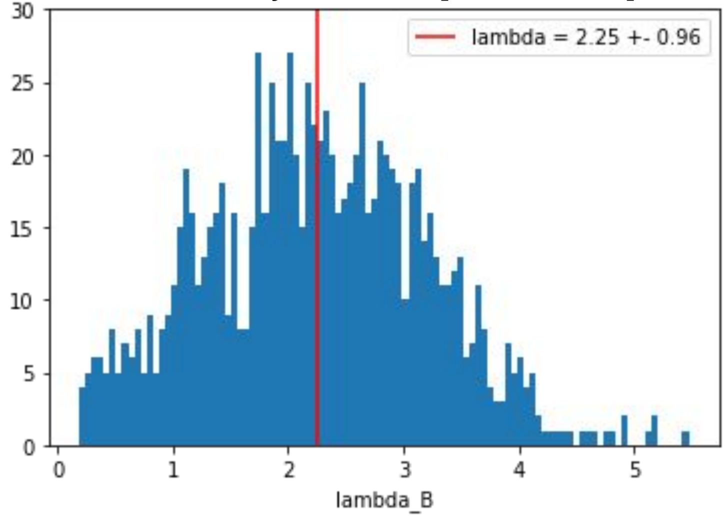
```
mass       = 0.85 GeV
N_sample = 1000
```

The toy data samples are generated from the likelihood with fixed μ_S and μ_B, with μ_S and μ_B chosen such to give 3σ significance

From Maximum likelihood [assuming μ_S and μ_B]

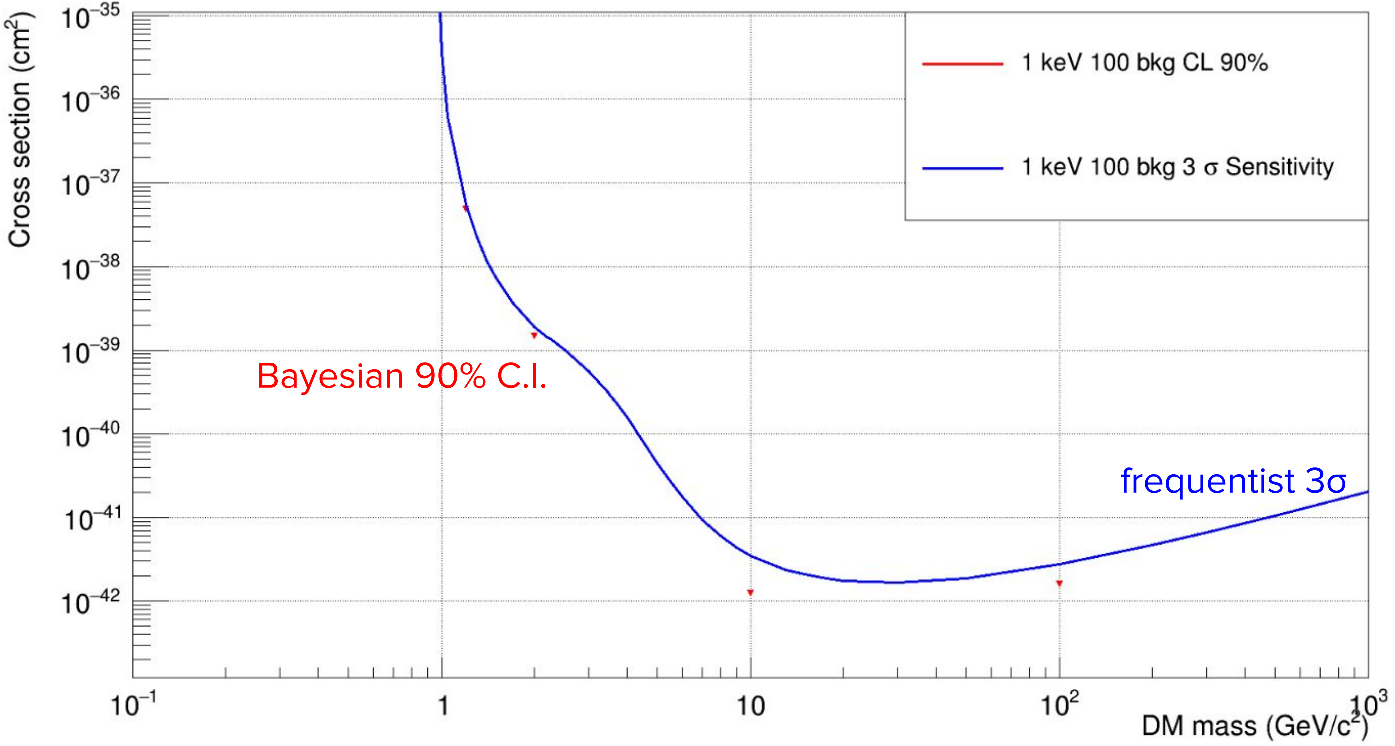


From Bayes factor [see slide 6]



[lambda = # of σs]

Comparison with the previous analysis - II



Conclusions

- We developed the framework for the Bayesian analysis.
- We have shown how it is possible to define a ‘3σ sensitivity’ and a ‘90 % credible interval’ in this approach
- We compared the results form two different codes and approaches and we are confident there the two codes are sound
- The code is ready to be adapted to a more complex likelihood if needed.
- **We studied in detail the differences between the two approaches and we concluded that they deliver the same message but the bayesian approach is simpler and can model more complex scenarios without much additional effort.**

We then propose to move to a Bayesian approach