Exploring Monte-Carlo-integration techniques in Bayesian model selection

Jakob Krause* and Dominic Schüchter†

(Dated: March 1, 2021)

An article usually includes an abstract, a concise summary of the work covered at length in the main body of the article.

Usage: Secondary publications and information retrieval purposes.

Structure: You may use the `description` environment to structure your abstract; use the optional argument of the `\item` command to give the category of each item.

I. INTRODUCTION

In physics, one is often faced with the problem of *Model Selection* for a given data set. That means finding a mathematical description of the data, which on the one hand sufficiently characterizes the data structure and on the other hand satisfies the expected dependencies. This is by no means a trivial task; one has to understand the underlying physical model beforehand to not make the mistake of choosing a too complicated model although it may seemingly fit the data. At the same time too trivial assumptions can also lead in the wrong direction. Compactly this problem can be formulated in the following way (adapted from [3, Chap. 4]):

Alice has a theory; Bob also has a theory, but with an adjustable parameter λ. Whose theory should we prefer on the basis of data D?

Bayesian inference provides quantitative measures for this model-selection problem, e.g. the Bayes-factor and the Bayes-complexity, these have among others been successfully used in astronomy as can be found in [2, 4] respectively. In this paper we will investigate two simulated example problems and apply various measures of bayesian model selection as to find out the true underlying model which was used to generate the simulated data. We will focus on the numerical evaluation of such problems, especially Monte-Carlo techniques.

II. THEORY

In the following we will give a short introduction into Bayes Theory and the underlying concepts of model selection.

A. Bayes' Theorem

The fundamental equation of Bayesian statistics – for a dataset y and parameters $\boldsymbol{\theta}$ – is given by Bayes Theorem.

$$\mathrm{prob}(\boldsymbol{\theta}|y) = p(\boldsymbol{\theta}|y) = \frac{p(y|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{p(y)} \tag{1}$$

Where $p(\boldsymbol{\theta}|y)$ is the posterior probability for parameters $\boldsymbol{\theta}$ given the data y, $p(y|\boldsymbol{\theta})$ is the likelihood that the data fits a model with parameters $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ is the prior probability of $\boldsymbol{\theta}$ and $p(y) = \int_{-\infty}^{+\infty} \mathrm{d}\boldsymbol{\theta} p(y|\boldsymbol{\theta})p(\boldsymbol{\theta})$ is the marginal likelihood which acts as a normalization. In the case of parameter selection, the normalization can often be neglected since it is only a constant. In the case of the model comparison it is a crucial quantity, as we will disuss in section (II C) [3, Chap. 2].

B. Parameter estimation

C. Model comparison

1. Bayes Factor

2. Bayes Complexity

III. METHODS

We will now describe the functions of the used algorithms for the model selection. Since the implementation of the so called nested sampling is sufficiently dealt with by Gruppe - Bayesian parameter fitting our implementation uses the `PyMC3`-Python Library [1].

A. Monte-Carlo integration

Here we will explain Monte-Carlo sampling, that is *Sequential Monte Carlo* and therein Metropolis-Hastings. its probably better to put these two subsections in separate sections.

---

* http://www.github.com/krausejm; krause@hiskp.uni-bonn.de

† http://www.github.com/dschuechter; dschuechter@uni-bonn.de

B. Savage Dickey Density Ration (SDDR)

IV. EXAMPLES

A. Betabinomial example (coin flip)

Let us now consider as a starting example, the flipping of a two-sided coin, i.e. an experiment where we can measure either heads (H) or tails (T) with 50% probability, respectively. This, while simple, allows us an intuitive approach to Bayesian inference and model selection as well as to the MCMC techniques discussed before. Furthermore is this example easily altered to many real-life problems, such as birth rates, . . . , or anything with the option of either success or failure.

1. *Analytical approach?*

Assume we throw a coin 20 times. We observe 6 H and 14 T. "Is this a fair coin?" might be a question to ask yourself since the bias in outcome is quite large. Naively expecting a fair coin we could assign a *prior* to the probability of heads θ as centred around 0.5, so for example a gaussian with $\mu = 0.5, \sigma = 0.1$.

2. *Numerical approach*

B. Fitting a polynomial of unknown degree

1. *Analytical approach?*

2. *Numerical approach*

V. DISCUSSION

VI. SUMMARY

[1] Pymc3 documentation — pymc3 3.10.0 documentation. URL `https://docs.pymc.io/`.

[2] M. Kunz, R. Trotta, and D. R. Parkinson. Measuring the effective complexity of cosmological models. *Phys. Rev. D*, 74:023503, Jul 2006. doi: 10.1103/PhysRevD.74.023503. URL `https://link.aps.org/doi/10.1103/PhysRevD.74.023503`.

[3] D. Sivia and J. Skilling. *Data Analysis - A Bayesian tutorial*, volume 2. Oxford University Press, 2006.

[4] R. Trotta. Applications of Bayesian model selection to cosmological parameters. *Monthly Notices of the Royal Astronomical Society*, 378(1):72–82, 05 2007. ISSN 0035-8711. doi:10.1111/j.1365-2966.2007.11738.x. URL `https://doi.org/10.1111/j.1365-2966.2007.11738.x`.