Exploring Monte-Carlo-integration techniques in Bayesian model selection

Jakob Krause* and Dominic Schüchter†
(Dated: March 3, 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 [11, 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?

For this model-selection problem Bayesian inference provides quantitative measures, e.g. the Bayes-factor and the Bayes-complexity, these have among others been successfully used in astronomy as can be found in [9, 13, 14] 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 theorem and the underlying concepts of parameter estimation and model selection.

A. Bayes' Theorem

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

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

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 subsection (II C) [11, Chap. 2].

B. Parameter estimation

Assume we want to find the best parameters for a given dataset y and model M. Eq. (1) then can be written as

$$p(\boldsymbol{\theta}|y, M) = \frac{p(y|\boldsymbol{\theta}, M) \cdot p(\boldsymbol{\theta}|M)}{p(y|M)}. \tag{2}$$

Evaluating this equation will give probability density functions (PDFs)

$$p(\theta_i|y, M) = \int p(\boldsymbol{\theta}|y, M) \prod_{j \neq i} \mathrm{d}\theta_j \tag{3}$$

for each parameter θ_i. Because of equation (3) $p(\theta|D, M)$ is also called *marginal posterior*. From this one can find the best fit value either by finding the value of θ for which the marginal posterior is maximized or by calculating the mean with respect to the marginal posterior.

$$\begin{aligned} p(\theta|y, M) = \max &\Leftrightarrow \theta = \hat{\theta} \\ \langle\theta\rangle &= \int_{-\infty}^{\infty} \mathrm{d}\theta p(\theta|y, M) \cdot \theta \end{aligned} \tag{4}$$

Throughout this paper we will use $\langle\theta\rangle$ as our best fit estimate. [11]

C. Model comparison

To compare two (or more) competing models M_i that describe a dataset D let us write Bayes theorem once

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

again

$$p(M_i|y) = \frac{p(M_i) \cdot p(y|M_i)}{p(y)}. \tag{5}$$

We recognize $p(y|M_i)$ as the marginal likelihood of the model M_i. Let now the two competing models be those of Alice and Bob introduced in section (I).Thus,

$$p(y|M_i) = \int_{-\infty}^{\infty} \mathrm{d}\lambda p(y|\lambda, M) \cdot p(\lambda|M)$$

follows [11, Chap. 3].

1. *Bayes Factor*

One way of comparing two models is to compare their posterior odds, that is the ratio [14]

$$O_{ij} := \underbrace{\frac{p(M_i|y)}{p(M_j|y)}}_{\text{posterior odds}} = \underbrace{\frac{p(y|M_i)}{p(y|M_j)}}_{\text{BAYES Factor}} \cdot \underbrace{\frac{p(M_i)}{p(M_j)}}_{\text{prior odds}} = B_{ij} \cdot \frac{p(M_i)}{p(M_j)}. \tag{6}$$

Usually we can assume the prior odds to be ~ 1 because we do not favor one model over the other just on prior beliefs. With this equation (6) simply becomes

$$O_{ij} = B_{ij}.$$

Calculating B_{ij} will be referred to as calculating the BAYES factor *in favor of* model M_i. We can then evaluate the result of B_{ij} as follows

$$B_{ij} = B_{ji}^{-1} = \begin{cases} \geq 1 & \text{model } M_i \text{ favored} \\ < 1 & \text{model } M_j \text{ favored} \end{cases}$$

Nevertheless a BAYES Factor ≥ 1 does not mean that the favored model is far superior. There is an empirical scale which points to the strength of evidence for a given B_{ij}. This is displayed in table (I).

$\lvert \ln B_{ij} \rvert$	Odds	Probability	Strength of evidence
< 1.0	$\lesssim 3:1$	< 0.750	Inconclusive
1.0	$\sim 3:1$	0.750	Weak evidence
2.5	$\sim 12:1$	0.923	Moderate evidence
5.0	$\sim 150:1$	0.993	Strong evidence

TABLE I: Empirical scale for evaluating the strength of evidence when comparing two models M_i vs. M_j, adapted from [14]

2. *Bayes Complexity*

When comparing models with a different amount of parameters, we obviously denote more parameters as more complex. A fundamental question of model-selection is then if the given data supports more or less parameters. A naïve measure of complexity would be the number of free parameters a model has, we denote this as $\mathcal{C}_0$. A more sophisticated measure of complexity is given by the so called BAYESIAN complexity, first introduced by Spiegelhalter et al [12], which can be written as [9]

$$\mathcal{C}_b = -2 \int \mathrm{d}\boldsymbol{\theta} p(\boldsymbol{\theta}|y, M) \log(\mathcal{L}(\boldsymbol{\theta})) + 2 \log(\mathcal{L}(\tilde{\boldsymbol{\theta}})), \tag{7}$$

with the likelihood $\mathcal{L}(\boldsymbol{\theta}) = p(D|\boldsymbol{\theta}, M)$ and $\tilde{\boldsymbol{\theta}} = \langle \boldsymbol{\theta} \rangle$. $\mathcal{C}_b$ describes how many model parameters the data is able to constrain [9] and is thus a useful tool for examining models with an increasing number of parameters. If we define a χ^2 as $\mathcal{L}(\boldsymbol{\theta}) \propto \exp(-\chi^2/2)$, we can write

$$\mathcal{C}_b = \overline{\chi^2(\boldsymbol{\theta})} - \chi^2(\tilde{\boldsymbol{\theta}}), \tag{8}$$

where $\overline{\chi}$ denotes the mean taken over the posterior PDF. The definition of $\mathcal{C}_b$ is chosen such that $\mathcal{C}_b \to \mathcal{C}_0$ for highly informative data [9].

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 [2]. `PyMC3` offers simple solutions to create models and a wide variety of sampling algorithms. Furthermore the `ArviZ`-Library provides methods for posterior analysis and visualization [1].

A. Monte-Carlo Sampling

First, the goal of MONTE-CARLO sampling is discussed. Often equations (3), (7) and (8) have no closed analytical solutions, so we are left with a numerical ansatz (or analytical approximations) [15]. One possibility is given by Markov-Chain-Monte-Carlo Sampling (MCMC) [**?**]. The main idea is to generate a chain of parameter samples $\boldsymbol{\theta}^{(t)}$ that follow the posterior PDF (2). For such a chain the mean with respect to the posterior is given by

$$\langle \boldsymbol{\theta} \rangle \approx \int p(\boldsymbol{\theta}|y) \boldsymbol{\theta} d\boldsymbol{\theta} = \frac{1}{N} \sum_{t=0}^{N-1} \boldsymbol{\theta}^{(t)}, \tag{9}$$

since the samples follow the PDF $p(\boldsymbol{\theta}|y)$. Analogously the expectation value of any function with respect to the posterior is

$$\langle f(\boldsymbol{\theta}) \rangle \approx \frac{1}{N} \sum_{t=0}^{N-1} f(\boldsymbol{\theta}^{(t)}). \tag{10}$$

Having established a MARKOV chain, one can compute the marginal posterior (3) by binning θ_i in the given parameter range and ignoring all other parameters [14].

B. Metropolis-Hastings Algorithm

So far we have discussed what we gain from sampling from the posterior distribution. We have not yet explained an algorithm how this is actually achieved. For our purposes we chose the *Sequential Monte Carlo* sampling algorithm provided by `PyMC3` [3], because it grants easy access to the marginal likelihood. This in turn is needed for model comparison via the Bayes-factor. Because the Metropolis-Hastings-algorithm is one of the fundamentals of understanding SMC, see subsection (III C), we briefly state it here. The main characterization of a Markov chain – which consists of random parameters $\boldsymbol{\theta}^{(t)}$ – is that each element is only determined by the previous element. In our case we want to sample in the parameter space according to the distribution $p(\boldsymbol{\theta}|y, M) := p(\boldsymbol{\theta})$ (2). We can achieve this by randomly proposing a new vector in parameter space $\boldsymbol{\theta'}$ according to a arbitrary proposal distribution $p_p(\boldsymbol{a}|\boldsymbol{b})$ and accept it with a probability

$$A(\boldsymbol{\theta'}, \boldsymbol{\theta}^{(t)}) = \min\left(1, \frac{p(\boldsymbol{\theta'})p_p(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta'})}{p(\boldsymbol{\theta}^{(t)})p_p(\boldsymbol{\theta'}|\boldsymbol{\theta}^{(t)})}\right).$$

This ensures the sampled values follow the desired PDF [15]. This is only a very brief overview of MCMC. More information on the underlying theory can be found e.g. in [10]

C. Sequential Monte Carlo (SMC)

Since the SMC algorithm joins several statistical concepts, including *importance sampling*, *tempering* and and MCMC kernel (Metropolis-Hastings) [3], it is a highly non-trivial algorithm. Thus, a detailed description is beyond the scope of this paper. We will now sketch the main idea of the algorithm. First let us introduce an auxiliary *parameter* β and write equation (2) as

$$p(\boldsymbol{\theta}|y)_\beta = \frac{p(y|\boldsymbol{\theta})^\beta \cdot p(\boldsymbol{\theta})}{Z_\beta},$$

with $Z_\beta = \int \mathrm{d}\boldsymbol{\theta} p(y|\boldsymbol{\theta})^\beta \cdot p(\boldsymbol{\theta})$. For $\beta = 1$ we get the same equation as before, that is $p(\boldsymbol{\theta}|y)_\beta = p(\boldsymbol{\theta}|y)$. The idea is to gradually sample from $\beta = 0$ to $\beta = 1$ using β to control the transition from an easy to sample distribution to a harder one. The final result is a collection of samples from the true posterior [3]. We can then estimate the marginal likelihood as [8]

$$\hat{p}(y) = \prod_i \widehat{\frac{Z_{\beta_i}}{Z_{\beta_{i-1}}}}. \tag{11}$$

The hat denotes, that this is a numerical estimation of $p(y)$ [3]. Using SMC we can in one run do parameter inference and also compare different models using the Bayes-factor (6).

D. Savage Dickey Density Ration (SDDR)

Now we introduce an alternative way of computing the Bayes-factor (6) for *nested* models; Consider model M_j with free parameters ω, ψ and a submodel M_i with one free parameter ψ and fixed $\omega = \omega_\star$. Let us further assume seperable priors (which is usually the case [13])

$$p(\omega, \psi|M_j) = p(\omega|M_j)p(\psi|M_i).$$

We can then write the Bayes-factor as [13]

$$B_{ij} = B_{ji}^{-1} = \left.\frac{p(\omega|y, M_j)}{p(\omega|M_j)}\right|_{\omega=\omega_\star} \text{(SDDR)}. \tag{12}$$

IV. EXAMPLES

We will now discuss two instructive examples applying the methods described in section (III).

A. Betabinomial example (coin flip)

Let us consider as a starting example, the flipping of a 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 problems with the option of either success or failure. Assume we throw a coin N times. Based on the outcome we wish to determine the probability for H and draw a conclusion whether the coin is fair or, in fact, biased. We simulated data for $N = 50$ and the biased coin with $p(H) = 0.25$.

1. *Analytical approach*

Since this example is relatively simple we can compute the Bayes-factor also analytically.

Let us remember Bayes theorem in the context of this example where we want to compare two models $M_i, i = 1, 2$ which assign the following posterior to the probability of heads $p(\theta|y, M_i)$ for a given dataset y

$$p(\theta|y, M_i) = \frac{p(y|\theta, M_i) \cdot p(\theta|M_i)}{p(y|M_i)}.$$

Let now M_1 be a model which assumes a fair coin, so that the *prior* is narrowly set around $\theta = 0.5$. M_2 assumes a biased coin, that is a *prior* centered around $\theta = 0.25$. If we want to obtain the Bayes-factor we have to calculate

$$p(y|M_i) = \int \mathrm{d}\theta p(y|\theta, M_i) \cdot p(\theta|M_i)$$

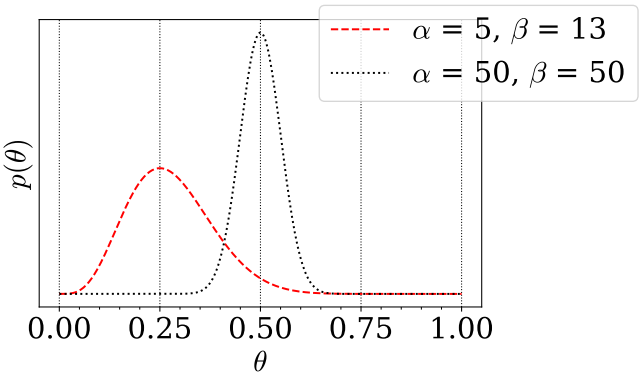

FIG. 1: Beta-distribution as *prior* $p(\theta|M_i)$ for the two different models

because B_{12} is given by (6)

$$B_{12} = \frac{p(y|M_1)}{p(y|M_2)}.$$

To get an analytical solution we first have to assign *prior* and *likelihood* for each model. A natural ansatz for a prior $p(\theta|M_i)$ would be the Beta-distribution, since it is limited to the finite interval $[0, 1]$. Its PDF is given by [7]

$$\begin{aligned} f(\theta; \alpha, \beta) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\ &:= \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}. \end{aligned}$$

Where Γ is the *Gamma-function* [5] and β is the *Beta-function* [4]. To get a narrowly centered distribution around 0.5 we chose $\alpha = \beta = 50$. To get a distribution skewed around 0.25 we have to choose $\beta = 3\alpha - 2$ [6], in our case $\alpha = 5, \beta = 13$, see figure (1). Now we will assign the *likelihood* function $p(y|\theta, M)$. We can assume that each coin throw is independent of preceding coin throws and, naturally, only two outcomes are possible. Then a *Binomial Distribution* is a suitable choice for our likelihood $p(y|\theta, M_i)$. If we observe k heads out of N coin throws ($y = (N, k)$)

$$p(y|\theta, M_i) = \binom{N}{k}\theta^k(1-\theta)^{N-k}.$$

Since we only are only interested in the ratio B_{12} we can drop the binomial coefficient and write

$$p(y|\theta, M_i) \propto \theta^k(1-\theta)^{N-k}.$$

Then the integral becomes

$$\begin{aligned} p(y|M_i) &\propto \int_0^1 \mathrm{d}\theta \frac{1}{B(\alpha,\beta)} \cdot \theta^{\alpha+k-1} \cdot (1-\theta)^{N-k+\beta-1} \\ &= \frac{B(\alpha+k, \beta+N-k)}{B(\alpha,\beta)}. \end{aligned}$$

With this solution the Bayes-factor becomes

$$B_{12} = \frac{B(\alpha_1+k, \beta_1+N-k)\cdot B(\alpha_2,\beta_2)}{B(\alpha_2+k, \beta_2+N-k)\cdot B(\alpha_1,\beta_1)}. \tag{13}$$

For the biased coin M_2 should be superior. The analytical result of equation (13) is

$$B_{21} = B_{12}^{-1} = 9.5839.$$

2. *Numerical approach*

We now compute the Bayes-factor numerically using the SMC-algorithm provided by `PyMC3`. We assigned the same *priors* and *likelikhoods* as in the analytical approach with a sample size of 2000 and 20 Monte-Carlo runs. The sampling yielded the *marginal posteriors* depicted in figure (2). The Bayes-factor according to equation (11) evaluated to

$$B_{21} = B_{12}^{-1} = 9.5829 \pm 0.4719$$

in very good agreement with the analytical result.

B. Fitting a polynomial of unknown degree

A common problem in physics is to find an appropiate function to fit unknown data. In the following we will show how Bayesian model selection can help to answer this question. For that an example dataset – following a polynomial of second order with a known variation of $\sigma = [0.7, 0.2]$ – will be used and polynomials of first to third order will be considered as possible models. The true regression line is given by

$$f(x; \boldsymbol{\theta}) = a \cdot x^2 + b \cdot x + c = 2 \cdot x^2 + 3 \cdot x + 1.$$

To tackle this problem numerically via the SMC algorithm, we have to assign *priors* and *likelihoods*. The *priors* for the fit-parameters a, b and c are each described by a normal distribution with $\mu = 0$ and $\sigma = 2$, because we have some intuition what the parameters might be. That means $p(\boldsymbol{\theta}|M_i) = p(\theta_1|M_i) \cdot p(\theta_2|M_i) \cdots$. For the *likelihood* functions we used normal distributions because we considered the noise being gaussian. The *likelihood* can then be written as [11, Chap. 3]

$$p(y|\boldsymbol{\theta}, M_i) = \prod_{k=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{(-f(y_k;\boldsymbol{\theta}) - y_k)^2}{2\sigma^2}\right].$$

According to Bayes theorem we generated 2000 samples following the *posterior* PDF $p(\boldsymbol{\theta}|y, M_i) \propto p(y|\boldsymbol{\theta}, M_i) \cdot p(\boldsymbol{\theta}|M_i)$ using the SMC algorithm. The resulting functions with parameters a, b and c in the highest density intervals (HDI) containing 99% of the sampled values are displayed in figure (3).

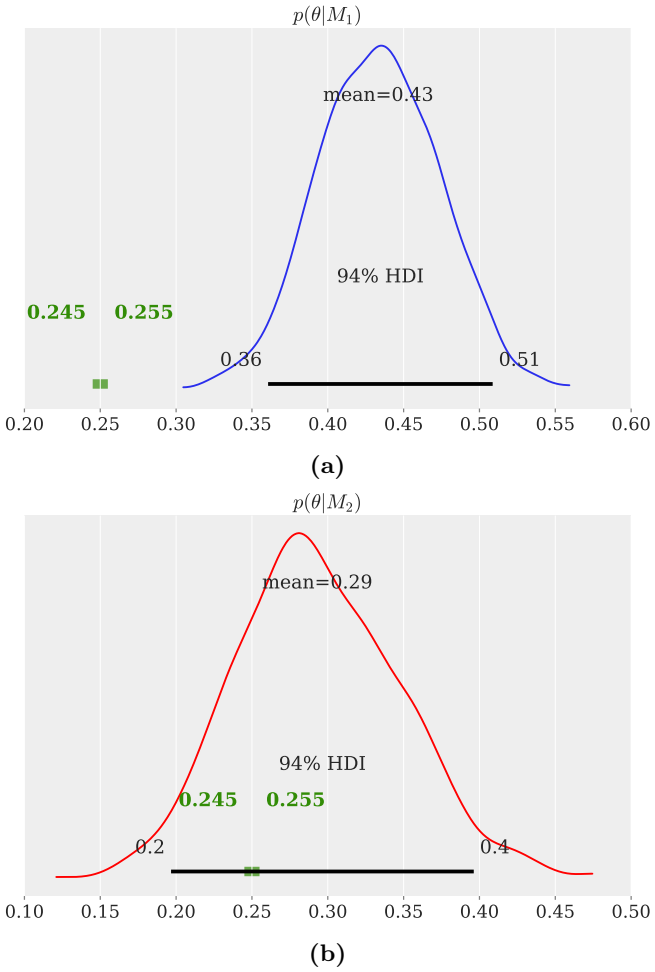

FIG. 2: The *marginal posterior* for $\alpha = \beta = 50$ (2a) and $\alpha = 5, \beta = 13$ (2b) of 2000 samples. HDI means highest density intervall. The highlighted green intervals denotes the expected value.

1. *Bayes-factor via SMC*

As before, it is now straightforward to compute the BAYES-factor comparing the models. Our results are in table (II).

Comparison M_1 vs. M_2	$\ln(B_{12})$
square vs. linear	8.5507 ± 0.053
cubic vs. linear	7.6225 ± 0.094
square vs. cubic	0.9371 ± 0.1093

TABLE II: Results of BAYES-factor via SMC

2. *Bayes-factor via SDDR*

In section (III D) we discussed an alternative way of computing the BAYES-factor. We notice this simplification applies to the case of polynomials of different degrees. Here each simpler model follows from the next more complex one by setting the coefficient of highest order to 0. Also the criterion of seperable priors applies. It is now a simple task to evaluate equation (12). We only need the normalized *marginal posterior* of the highest order coefficient and its corresponding prior at 0. The prior is known and the marginal posterior is directly obtained once we have sampled and can then be normalized. In figure (4) we show the priors and *marginal posteriors* of the highest order coefficient of the polynomial of second and third order respectively, which leads us to the following results, see table (III).

Comparison M_1 vs. M_2	$\ln(B_{12})$
square vs. linear	> 2.4301
square vs. cubic	0.8091
cubic vs. linear	> 0.3329

TABLE III: Results of BAYES-factor via SDDR. Note that for the comparison square vs. linear we could only provide a lower bound because our sample size did not provide us with values around $\omega_\star$

3. *Bayes-Complexity*

As another measure of model selection, the BAYESIAN complexity $\mathcal{C}_b$ was introduced. Having sampled in parameter space $\boldsymbol{\theta}^{(t)}$, equation (8) is evaluated instantly as [16]

$$\mathcal{C}_b = \frac{1}{N_S}\sum_{t=0}^{N_S-1}\left(\sum_{k=0}^{N_y-1}\frac{[y_k - f(y_k;\boldsymbol{\theta}^{(t)})]^2}{\sigma^2}\right) - \sum_{k=0}^{N_y-1}\frac{[y_k - f(y_k;\frac{1}{N_S}\sum_{t=0}^{N_S-1}\boldsymbol{\theta}^{(t)})]^2}{\sigma^2}.$$

In figure (5) we plotted the computed complexity $\mathcal{C}_b$ as a function of the number of parameters and also the input complexity $\mathcal{C}_0$.

V. DISCUSSION

VI. SUMMARY

VII. APPENDIX

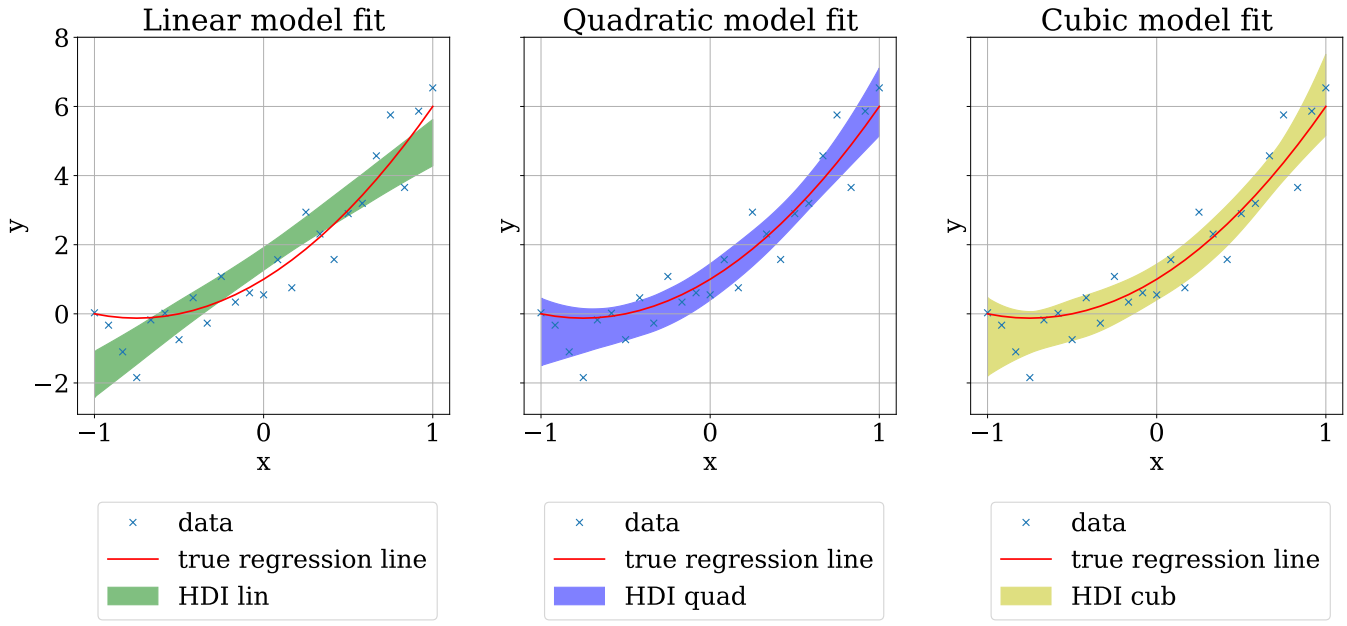

FIG. 3: Result of parameter estimation with SMC. The data was generated with $\sigma = 0.7$

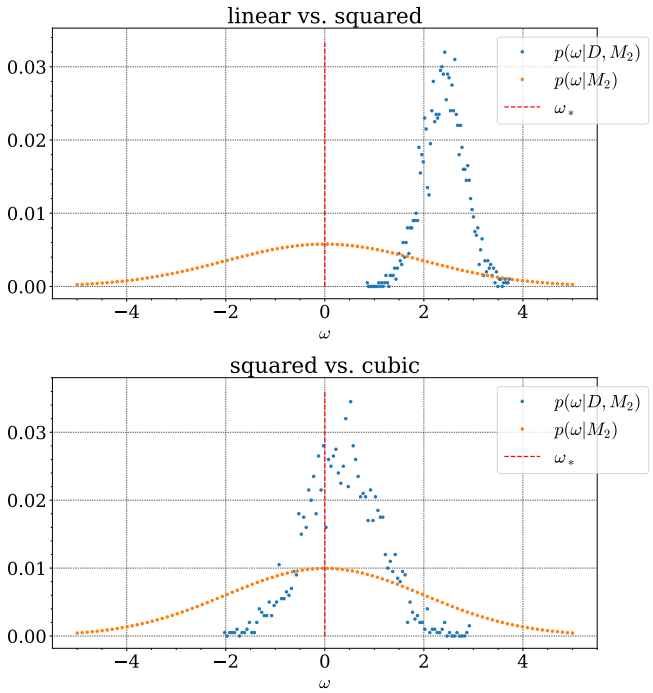

FIG. 4: Computation of the SDDR

[1] Arviz: Exploratory analysis of bayesian models — arviz dev documentation. URL `https://arviz-devs.github.io/arviz/`. last visit: March 3, 2021.

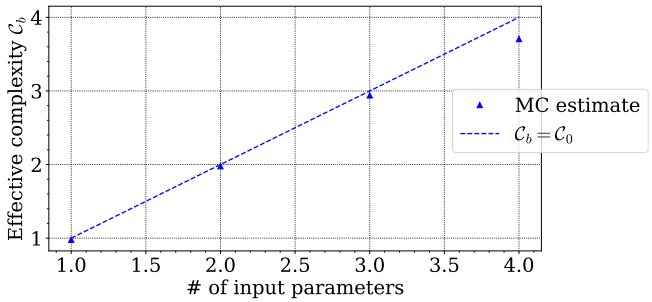

FIG. 5: BAYESIAN complexity as a function of number of input parameters

[2] Pymc3 documentation — pymc3 3.10.0 documentation, . URL https://docs.pymc.io/. last visit: March 3, 2021.

[3] Pymc3 documentation — sequential monte carlo, . URL https://docs.pymc.io/notebooks/SMC2_gaussians.html. last visit: March 3, 2021.

[4] Beta function. *Wikipedia*. URL https://en.wikipedia.org/wiki/Beta_function. last visit: March 3, 2021.

[5] Gamma function. *Wikipedia*. URL https://en.wikipedia.org/wiki/Gamma_function. last visit: March 3, 2021.

[6] Beta distribution. *Wikipedia*. URL https://en.wikipedia.org/wiki/Beta_distribution. last visit: March 3, 2021.

[7] S. Celik and M. Korkmaz. Beta distribution and inferences about the beta functions. *Asian Journal of Science and Technology*, 7:2960–2970, 05 2016.

[8] D. Gunawan, K.-D. Dang, M. Quiroz, R. Kohn, and M.-N. Tran. Subsampling sequential monte carlo for static bayesian models. 2020. URL https://arxiv.org/pdf/1805.03317.pdf. last visit: March 3, 2021.

[9] 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.

[10] R. M. Neal. Probabilistic inference using markov chain monte carlo methods. *Technical Report CRG-TR-93-1*.

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

[12] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(4):583–639, 2002. ISSN 13697412, 14679868. URL http://www.jstor.org/stable/3088806.

[13] 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.

[14] R. Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104, Mar 2008. ISSN 1366-5812. doi: 10.1080/00107510802066753. URL http://dx.doi.org/10.1080/00107510802066753.

[15] U. von Toussaint. Bayesian inference in physics. *Rev. Mod. Phys.*, 83:943–999, Sep 2011. doi: 10.1103/RevModPhys.83.943. URL https://link.aps.org/doi/10.1103/RevModPhys.83.943.

[16] A. Wirzba. private communication, Feb 2021.