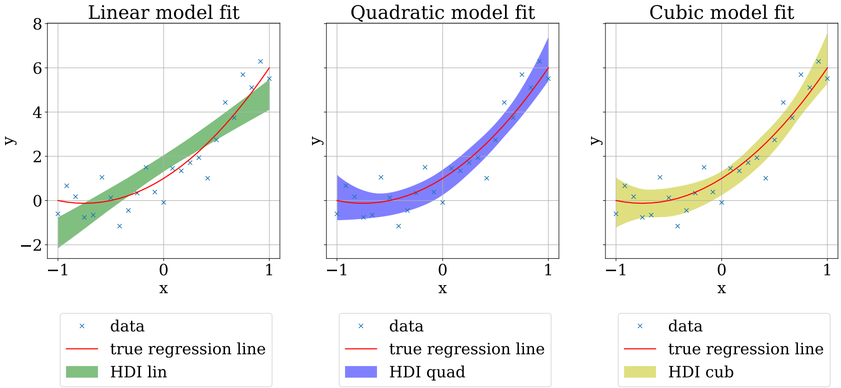
Bayesian model selection

Seminar physics760 – Computational Physics



Dominic Schüchter
dschuechter@uni-bonn.de | dschuechter

Jakob Krause
krause@hiskp.uni-bonn.de | krausejm

Tutor: Andreas Wirzba
a.wirzba@fz-juelich.de

19.03.2021

Introduction

Bayes' Theorem

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

with

- *posterior* $p(\boldsymbol{\theta}|y)$
- *likelihood* $p(y|\boldsymbol{\theta})$
- *prior* $p(\boldsymbol{\theta})$
- *marginal likelihood* $p(y) = \int_{-\infty}^{+\infty} \mathrm{d}\boldsymbol{\theta} p(y|\boldsymbol{\theta})p(\boldsymbol{\theta})$

This can be used for *model selection* (?)

Table of Contents

1. Theory
   - Parameter estimation
   - Model comparison
2. Methods
   - Monte-Carlo-Sampling
   - Savage-Dickey-Density-Ratio (SDDR)
3. Examples
   - Coin Flip
   - Fitting a polynomial of unknown degree
4. Summary

Parameter estimation

JAN and MARIUS already talked about this, so here we only sketch the basics again

$$p(\theta_i|y,M) = \int p(\boldsymbol{\theta}|y,M) \prod_{j\neq i} \mathrm{d}\theta_j \tag{1}$$

$$\begin{aligned} p(\theta|y,M) &= \max \Leftrightarrow \theta = \hat{\theta} \\ \langle\theta\rangle &= \int_{-\infty}^{\infty} \mathrm{d}\theta p(\theta|y,M)\cdot\theta \end{aligned} \tag{2}$$

Model comparison

How do we turn BAYES' theorem into a tool for model comparison?

BAYES factor

$$p(M_i|y) = \frac{p(M_i) \cdot p(y|M_i)}{p(y)}. \tag{3}$$

$$O_{ij} := \underbrace{\frac{p(M_i|y)}{p(M_j|y)}}_{\text{posterior odds}} = \underbrace{\frac{p(y|M_i)}{p(y|M_j)}}_{\text{BAYES Factor}} \cdot \underbrace{\frac{p(M_i)}{p(M_j)}}_{\text{prior odds}} = B_{ij} \cdot \frac{p(M_i)}{p(M_j)}. \tag{4}$$

Model comparison

Bayesian complexity

$$\mathcal{C}_b = -2 \int \mathrm{d}\boldsymbol{\theta} p(\boldsymbol{\theta}|y, M) \log(\mathcal{L}(\boldsymbol{\theta})) + 2 \log(\mathcal{L}(\tilde{\boldsymbol{\theta}})), \tag{5}$$

$$\mathcal{C}_b = \overline{\chi^2(\boldsymbol{\theta})} - \chi^2(\tilde{\boldsymbol{\theta}}), \tag{6}$$

1. Theory
   - Parameter estimation
   - Model comparison

2. Methods
   - Monte-Carlo-Sampling
   - SAVAGE-DICKEY-Density-Ratio (SDDR)

3. Examples
   - Coin Flip
   - Fitting a polynomial of unknown degree

4. Summary

Monte-Carlo-Sampling

Benefits of Monte-Carlo-Sampling

$$\langle \boldsymbol{\theta} \rangle \approx \int p(\boldsymbol{\theta}|y)\boldsymbol{\theta} d\boldsymbol{\theta} = \frac{1}{N}\sum_{t=0}^{N-1} \boldsymbol{\theta}^{(t)}, \tag{7}$$

$$\langle f(\boldsymbol{\theta}) \rangle \approx \frac{1}{N}\sum_{t=0}^{N-1} f(\boldsymbol{\theta}^{(t)}). \tag{8}$$

also, *marginal posterior* distributions are obtained trivially by binning values of θ_i ignoring $\theta_{j\neq i}$

Figure 1: `ArviZ` [1] and `PyMC3` [3]

Monte-Carlo-Sampling

But how exactly do we get samples $\boldsymbol{\theta}^{(t)}$?

Sequential Monte Carlo (SMC)

First let us introduce an auxiliary *temperature parameter* $\beta \in [0, 1]$ and write

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

with $Z_\beta = \int \mathrm{d}\boldsymbol{\theta} p(y|\boldsymbol{\theta})^\beta \cdot p(\boldsymbol{\theta})$.
Main idea: gradually sample from simple distribution ($\beta = 0$) to complex/true distribution ($\beta = 1$) using Metropolis-Hastings
SMC then allows us to estimate the *marginal likelihood* as

$$\hat{p}(y) = \prod_i \widehat{\frac{Z_{\beta_i}}{Z_{\beta_{i-1}}}}. \tag{9}$$

Savage-Dickey-Density-Ratio (SDDR)

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 separable priors (which is usually the case [5])

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

We can then write the Bayes factor as [5]

$$B_{ij} = B_{ji}^{-1} = \left. \frac{p(\omega | y, M_j)}{p(\omega | M_j)} \right|_{\omega = \omega_\star} \text{(SDDR)}. \tag{10}$$

Coin Flip

- remember the flipping of a biased coin from the lecture
- we want to verify that the coin is biased by using the Bayes factor
- two models: M_1 – assumes a fair coin, M_2 – assumes biased coin

Posterior of the coin flip problem

$$p(\theta|y, M_i) = \frac{p(y|\theta, M_i) \cdot p(\theta|M_i)}{p(y|M_i)} \tag{11}$$

to get $p(y|M_i) = \int_{-\infty}^{+\infty} \mathrm{d}\theta p(y|\theta, M_i) p(\theta|M_i)$ we need to specify a *prior* $p(\theta|M_i)$ and a *likelihood* $p(y|\theta, M_i)$

Coin Flip

Choosing a prior

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

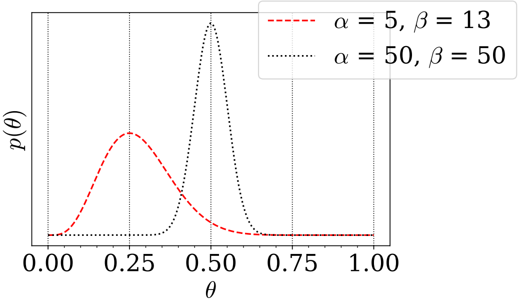


Figure 2: Beta-distribution as *prior* $p(\theta|M_i)$ for the two different models

Coin Flip

Choosing a likelihood

Since we can assume i.i.d. outcomes of the coin flip a natural choice is a *Binomial distribution*. If we observe k heads out of N coin throws ($y = (N, k)$)

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

We simulated data for $N = 50$ and the biased coin with $p(H) = 0.25$.

Coin Flip

Finally computing the BAYES factor

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

$$\Rightarrow B_{21} = B_{12}^{-1} = \frac{B(\alpha_2+k,\beta_2+N-k)\cdot B(\alpha_1,\beta_1)}{B(\alpha_1+k,\beta_1+N-k)\cdot B(\alpha_2,\beta_2)} = 9.5839$$

Can we reproduce this numerically?

Coin Flip

let us obtain 2000 samples from the posterior distribution using the same *likelihood* and *priors* via the SMC algorithm provided by `PyMC3`

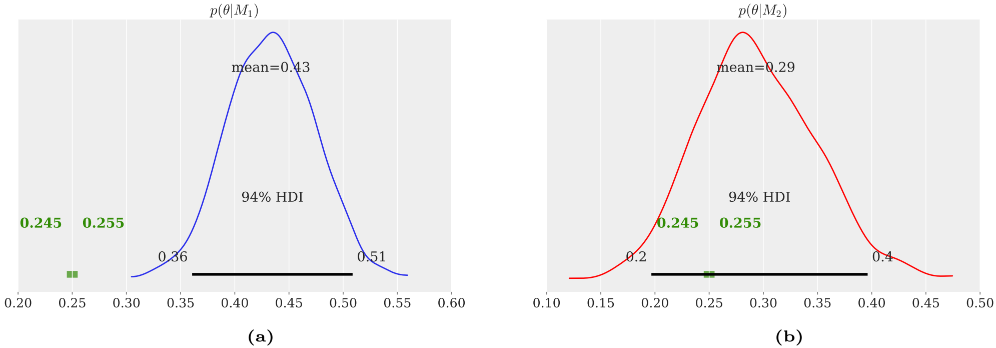


Figure 3: The *marginal posterior* for $\alpha = \beta = 50$ (3a) and $\alpha = 5, \beta = 13$ (3b) of 2000 samples. HDI means highest density interval. The highlighted green intervals denote the expected value.

$$\text{We find } B_{21} = B_{12}^{-1} = 9.5829 \pm 0.4719 \text{ (noice! } \checkmark\text{)}$$

Summary

References I

[1] *ArviZ: Exploratory analysis of Bayesian models — ArviZ dev documentation.* last visit: 13th March 2021. URL: https://arviz-devs.github.io/arviz/.

[2] 'Beta distribution'. In: *Wikipedia* (). last visit: 13th March 2021. URL: https://en.wikipedia.org/wiki/Beta_distribution.

[3] *PyMC3 Documentation — PyMC3 3.10.0 documentation.* last visit: 13th March 2021. URL: https://docs.pymc.io/.

[4] *PyMC3 Documentation — Sequential Monte Carlo.* last visit: 13th March 2021. URL: https://docs.pymc.io/notebooks/SMC2_gaussians.html.

[5] Roberto Trotta. 'Applications of Bayesian model selection to cosmological parameters'. In: *Monthly Notices of the Royal Astronomical Society* 378.1 (May 2007), pp. 72–82. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2007.11738.x. eprint: https://academic.oup.com/mnras/article-pdf/378/1/72/3961005/mnras0378-0072.pdf. URL: https://doi.org/10.1111/j.1365-2966.2007.11738.x.