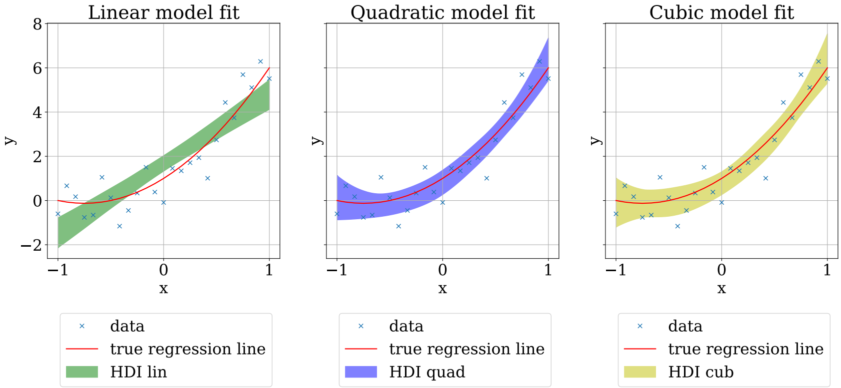
Bayesian model selection

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)
   - Error analysis and diagnostics

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

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

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

Monte-Carlo-Sampling

Benefits of Monte-Carlo-Sampling

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

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

Figure 1: `ArviZ` [1] and `PyMC3` [2]

Savage-Dickey-Density-Ratio (SDDR)

Error analysis and diagnostics

Inhalt...

Choosing priors and likelihoods

Computing Bayes-factor

Summary

-

References I

[1] *ArviZ: Exploratory analysis of Bayesian models — ArviZ dev documentation.* last visit: 12th March 2021. URL: https://arviz-devs.github.io/arviz/.

[2] *PyMC3 Documentation — PyMC3 3.10.0 documentation.* last visit: 12th March 2021. URL: https://docs.pymc.io/.