An introduction to probabilistic programming with PyMC3

Sean Meling Murray and Solveig Masvie

10. februar 2020

Table of contents

Introduction

Bayesian statistics

Markov Chains

Road map

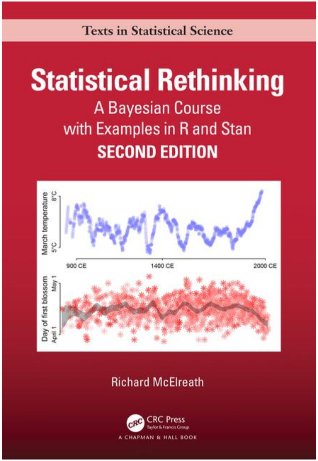
- Theory
  - The basics of Bayesianism
  - Markov chain Monte Carlo methods (MCMC)
- Practice
  - Probabilistic programming with PyMC3

What is Bayesian data analysis?

"A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule."

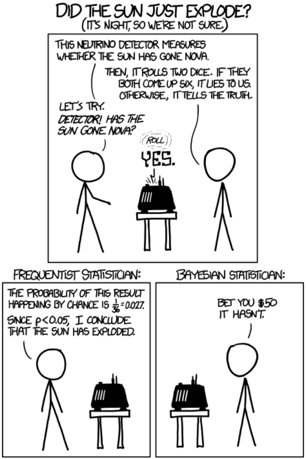
What is Bayesian data analysis?

- Richard McElreath: “Bayesian inference is just counting.”
- Count all the ways observed data could have arisen according to assumptions
- Assumptions that can arise in more ways are more consistent with the data, and therefore more plausible



The Frequentist vs. Bayesian debacle

- Frequentist statistics
  - Probability defined as the limiting frequency at which events occur
  - Uncertainty arises from sampling variation
- Bayesian statistics
  - Frequency and probability are different things
  - Uncertainty arises from our ignorance of the true state of the world



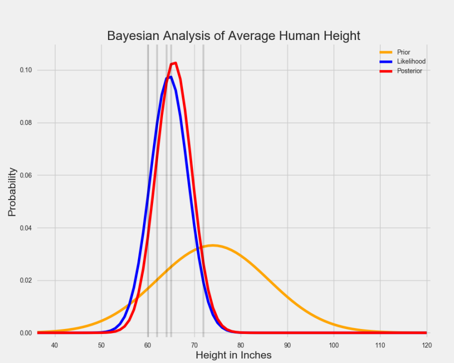
inmeta

Bayesian Analysis

The prior distribution combined with likelihood distribution (observed data) equals posterior distribution

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)\,d\theta} \tag{1}$$

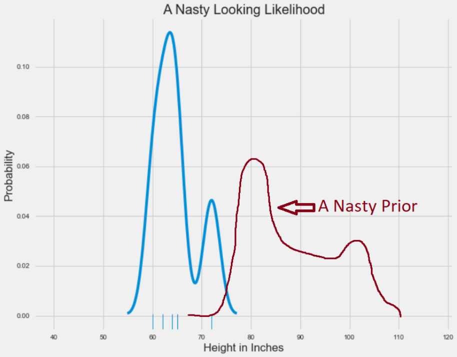
$$\mathscr{P} \propto \mathscr{L}\Pi \tag{2}$$



inmeta

Bayesian Analysis

How do we find the posterior when the prior and likelihood distribution are complicated



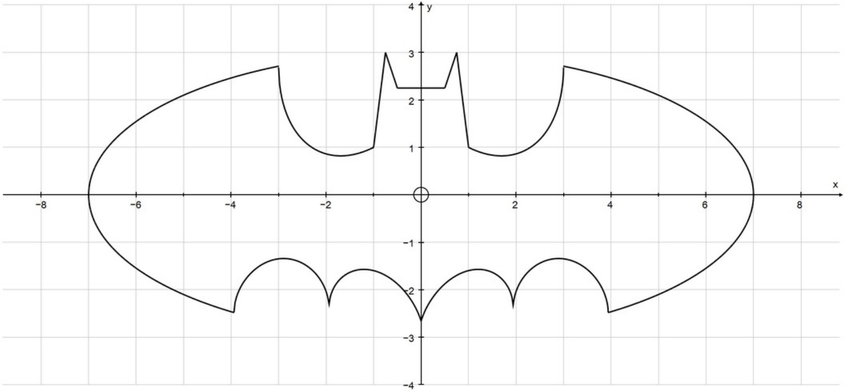
inmeta

Monte Carlo simulations

- Monte Carlo (MC) simulations are just a way to approximate numerical results using repeated random sampling
- Main idea:
  - Generate N samples $x_1, \ldots, x_N$ from $p(x)$, approximate f using the empirical distribution of $\{f(x_n)\}_{n=1}^{N}$
  - $\mathbb{E}[f] = \int f(x)p(x)\,dx \approx \frac{1}{N}\sum_{n=1}^{N} f(x_n) = \hat{f}$
- MC estimates converge thanks to Law of Large Numbers (LLN)
  - $(\hat{f} - f) \to \mathcal{N}\left(0, \frac{\sigma^2}{N}\right)$ as $N \to \infty$

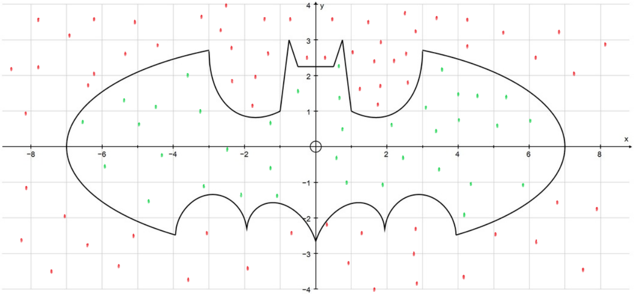
Practical example

- Calculate difficult integrals, such as the area of the Batman sign



Monte Carlo Simulations - Example

- Repeatedly sample $(u_1, u_2) \sim \text{Uniform}(-\frac{1}{2}, \frac{1}{2})$
- Calculate area A of Batman sign as $A = \text{area of rectangle} \times \frac{\text{green dots}}{\text{all dots}}$



inmeta

Markov Chain

- Sequences of events that have a probabilistic relation to one another
- Markov chains are memoryless. All we need to calculate the next event are available in the current state

Bayesian MCMC

Steps in MCMC

- Define function for $\mathscr{L}$, Π and thus $\mathscr{P}$
- Define initial guess for θ (based on the prior)
- Try a jump in θ
- Accept/reject based on chosen method/sampler (Metropolis)
- Keep jumping
- After doing many steps remove burn-in steps

Sampler

Metropolis rule

- If $\mathscr{P}_{new} > \mathscr{P}_i$ accept the jump so $\theta_{i+1} = \theta_i$
- If $\mathscr{P}_{new} < \mathscr{P}_i$ accept the jump with probability $\frac{\mathscr{P}_{new}}{\mathscr{P}_i}$

inmeta

Jump in θ

How to make a jump in θ

$$\theta_{new} = \theta_i + \mathcal{N}(0, \Delta\theta) \tag{3}$$

We call $\Delta\theta$ for the jump scale. Normally this must be tuned manually for every dimension. A rule of thumb is that we want a jump scale that gives a reasonable acceptance rate.

A slide with a theorem and a proof.

Theorem (Integral)

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

Bevis.

Here's the proof. □

inmeta

A slide with blocks

title of the bloc

bloc text

title of the bloc

bloc text

A slide using pause

- Represent Abelian groups on the computer

A slide using pause

- Represent Abelian groups on the computer
- Compute on Abelian groups

A slide using pause

- Represent Abelian groups on the computer
- Compute on Abelian groups
- Solve equations, factor group homomorphisms