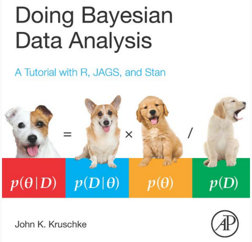
# Introduction to Bayesian Inference and PyMC3

Fraternali group retreat

Marius Kaušas

2018-09-02

# Introduction




**(a)** The puppy book.


**(b)** Superb book on ML.

**Figure 1:** Material is based on the following books.

# Introduction

Exercises from *Doing Bayesian Data Analysis* 1E by John K. Kruschke have been ported to PyMC3.

**GitHub link:** `https://github.com/aloctavodia/Doing_bayesian_data_analysis`

# What is Bayesian Inference?

- Bayesian inference is reallocation of credibility accross possibilities.

- The possibilities, over which we allocate credibility, are parameter values in meaningful mathematic models.

# Steps of doing Bayesian data analysis

1. Identify the data relevant to the research questions. Which data variables are to be predicted and which data variables are supposed to act as predictors?
2. Define a descriptive model for the relevant data.
3. Specify a prior distribution on the parameters.
4. Use Bayesian inference to re-allocate credibility across parameter values. Interpret the posterior distribution.
5. Conduct a posterior predictive check.

# Bayes' Rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{1}$$

# Bayes' Rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{1}$$

The "prior", $p(\theta)$, is the credibility of $\theta$ values without the data $D$.

# Bayes' Rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{1}$$

The "likelihood", $p(D|\theta)$, is the probability that the data coule be generated by the model with parameter value $\theta$.

# Bayes' Rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{1}$$

The "marginal likelihood", $p(D)$, is the of the overall probability of the data according to the model, determined by averaging across all possible parameter values weighted by the strength of belief in those parameter values.

# Bayes' Rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{1}$$

The "posterior", $p(\theta|D)$, is the credibility of $\theta$ values with the data $D$ taken into the account.

# Inferring a Binomial probability via analytical solution

We start with our favourite coin flipping example.

# Inferring a Binomial probability via analytical solution

**The following "simple" example will provide with:**

- Understanding of underlying concepts of Bayesian inference on a continuous parameter.

- A clear (hopefully!) sense of what is being approximated.

# Inferring a Binomial probability via analytical solution

**Context:**

Let's estimate the underlying probability of the two possible outcomes:

$p(x = 1|\theta) = \theta$ $\qquad\qquad \theta \in [0, 1]$

$p(x = 0|\theta) = 1 - \theta$, where *heads* $= 1$ and *tails* $= 0$

$$Bernoulli(x|\theta) = \theta^x (1 - \theta)^{1-x} \tag{2}$$

# Inferring a Binomial probability via analytical solution

**Dataset:**

$D = x_1, x_2, ..., x_N$, where $N$ is the total number of flips and $z$ is the number of heads.

**Likelihood:**

$$
\begin{aligned}
p(D|\theta) &= \prod_{n=1}^{N} p(x_n|\theta) && (3)\\
&= \prod_{n=1}^{N} \theta^{x_n}(1-\theta)^{1-x_n} && (4)\\
&= \theta^{z}(1-\theta)^{N-z} && (5)
\end{aligned}
$$

# Inferring a Binomial probability via analytical solution

**Prior distribution:**

$$Beta(\theta|\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \textit{ where } \alpha,\ \beta > 0 \tag{6}$$

$$B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \tag{7}$$

# Inferring a Binomial probability via analytical solution

## Probability density functions (PDFs) of Beta distribution:

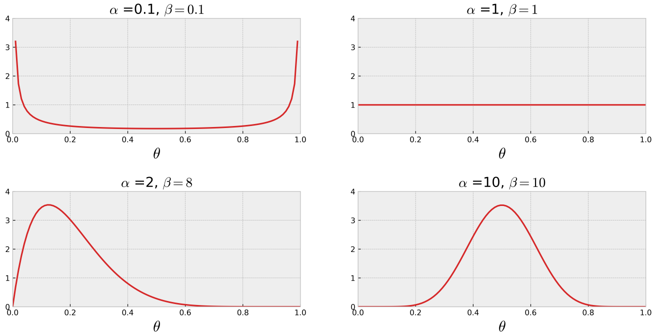



**Figure 2:** PDFs of *Beta distribution* for various $\alpha$ and $\beta$ parameters.

# Inferring a Binomial probability via analytical solution

**Conjugate prior:**

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{1}$$

When the forms of $P(D|\theta)$ and $p(\theta)$ combine so that the posterior distribution has the same form as the prior distribution, then $p(\theta)$ is called *conjugate prior* for $p(D|\theta)$.

# Inferring a Binomial probability via analytical solution

**Posterior is a compromise of prior and likelihood:**

$$p(\theta|z, N) = \frac{p(z, N|\theta)p(\theta)}{p(z, N)} \tag{8}$$

$$= \theta^{z}(1-\theta)^{N-z}\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}/p(z, N) \tag{9}$$

# Inferring a Binomial probability via analytical solution

**Posterior is a compromise of prior and likelihood:**

$$
\begin{aligned}
p(\theta|z,N) &= \frac{p(z,N|\theta)p(\theta)}{p(z,N)} && (8)\\
&= \theta^{z}(1-\theta)^{N-z}\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}/p(z,N) && (9)\\
&= \theta^{z}(1-\theta)^{N-z}\theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha,\beta)p(z,N) && (10)
\end{aligned}
$$

The denominator must be a normalizer for a posterior distribution to be a probability distribution. We have already seen one like that in the *Beta distribution*.

# Inferring a Binomial probability via analytical solution

**Posterior is a compromise of prior and likelihood:**

$$
\begin{aligned}
p(\theta|z,N) &= \frac{p(z,N|\theta)p(\theta)}{p(z,N)} && (8)\\
&= \theta^z(1-\theta)^{N-z}\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}/p(z,N) && (9)\\
&= \theta^z(1-\theta)^{N-z}\theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha,\beta)p(z,N) && (10)\\
&= \theta^{((z+\alpha)-1)}(1-\theta)^{((N-z+\beta)-1)}/B(z+\alpha,\ N-z+\beta) && (11)
\end{aligned}
$$

# Inferring a Binomial probability via analytical solution

**Posterior is a compromise of prior and likelihood:**

$$p(\theta|z, N) = Beta(\theta|\alpha_N, \beta_N) \tag{12}$$

Where,

$$\alpha_N = \alpha_0 + z \tag{13}$$
$$\beta_N = \beta_0 + (N - z) \tag{14}$$

... and we end up with a *Beta distribution* (6) .That we can easily update.

# Inferring a Binomial probability via analytical solution

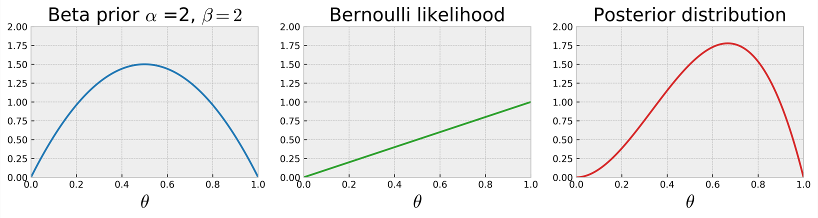



**Figure 3:** Inference of a single flip.

# Markov chain Monte Carlo (MCMC)

- Sometimes our priors must be expressed by other distributions, which do not yield analytically solvable posterior distributions.

- However, one can approximate a posterior distribution.

- A way to do that is to sample a number of representative points from the posterior using MCMC method.

# Metropolis algorithm

- Generate a random value from the proposal distribution, to create a $\theta_{proposed}$.

- Evaluate the target distribution at any proposed position to compute:

$$p_{move} = min\left(1, \frac{P(\theta_{proposed})}{P(\theta_{current})}\right) \tag{15}$$

- Generate a random value from *Uniform distribution* $\in [0, 1]$.

- Accept the move, if the $U(0, 1) \in [0, p_{move}]$.

# Inferring a Binomial probability via MCMC

**Back to our coin flipping:**

- Propose a jump: $\theta_{pro} = \theta_{cur} + \Delta\theta$, where $\Delta\theta \sim Normal(\mu, \sigma)$
- Compute the probability of moving to the proposed value:

$$p_{move} = min\left(1, \frac{P(\theta_{pro})}{P(\theta_{cur})}\right) \tag{16}$$

$$= min\left(1, \frac{p(D|\theta_{pro})p(\theta_{pro})}{p(D|\theta_{cur})p(\theta_{cur})}\right) \tag{17}$$

$$= min\left(1, \frac{Bernoulli(z, N|\theta_{pro})Beta(\theta_{pro}|\alpha, \beta)}{Bernoulli(z, N|\theta_{cur})Beta(\theta_{cur}|\alpha, \beta)}\right) \tag{18}$$

# Inferring a Binomial probability via MCMC

- Compute the probability of moving to the proposed value:

$$p_{move} = min\left(1, \frac{Bernoulli(z, N|\theta_{pro})Beta(\theta_{pro}|\alpha, \beta)}{Bernoulli(z, N|\theta_{cur})Beta(\theta_{cur}|\alpha, \beta)}\right) \tag{18}$$

$$= min\left(1, \frac{\theta_{pro}^{z}(1-\theta_{pro})^{(N-z)}\theta_{pro}^{(\alpha-1)}(1-\theta_{pro})^{(\beta-1)}/B(\alpha,\ \beta)}{\theta_{cur}^{z}(1-\theta_{cur})^{(N-z)}\theta_{cur}^{(\alpha-1)}(1-\theta_{cur})^{(\beta-1)}/B(\alpha,\ \beta)}\right) \tag{19}$$

- Accept the move if:

$$U(0,1) \in [0, p_{move}]$$

# Posterior Predictive Checks

- Posterior predictive checks (PPCs) are a way to validate a model by generating data from the model parameters sampled from the posterior.

- PPCs are made for revising, simplifying or expanding the model as one examines how well it fits the data.

# Lights, camera, action!

To the PyMC3 mobile!

# References I


C. M. Bishop.
**Pattern recognition and machine learning.**
Springer, 2006.

J. K. Kruschke and T. M. Liddell.
**Bayesian data analysis for newcomers.**
*Psychonomic Bulletin and Review*, 25:155–177, 2018.

J. K. Kruschke and T. M. Liddell.
**The Bayesian New Statistics.**
*Psychonomic Bulletin and Review*, 25:178–206, 2018.

J. L. Kruschke.
**Doing Baeysian Data Analysis.**
Elsevier, 2015.

# References II

D. van Ravenzwaaij, P. Cassey, and S. D. Brown.
**A simple introduction to Markov Chain Monte Carlo sampling.**
*Psychonomic Bulletin and Review*, 25:143–154, 2018.