

Bayesian Inference

Probabilities:

Events: A and B

$P(A)$: (marginal) probability

$P(B)$: (marginal) probability

$P(A \cap B)$: joint probability

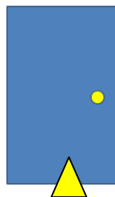
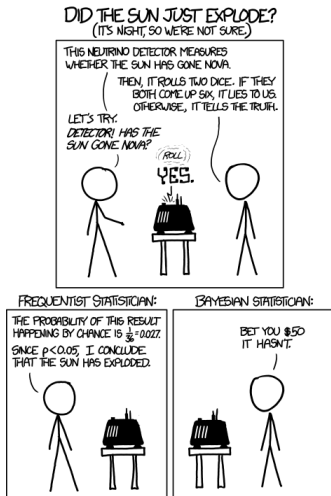
$P(A|B)$: conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



Door 1



Door 2



Door 3

What is 'Bayesian' analysis?

- ▶ Bayesian analysis is a different way to learn from your data
- ▶ It is based around the application of Bayes' theorem
- ▶ It is a type of statistical inference
- ▶ Bayesian analysis has two inputs:
 - What we knew before the we analyzed the data
 - The data itself
- ▶ Bayesian analysis tells us what the new state of knowledge is (including any uncertainty) after analyzing the data

Statistical inference

- ▶ Nearly all forms of statistical analysis follow these basic steps:
 - ▶ Construct a mathematical simplification of the study system that encapsulates the phenomena that you are interested in. This is referred to as a 'model'. Linear models (used in linear regression) are an example of a type of model.
 - ▶ Compare the model with the data
 - ▶ Make conclusions based on this comparison
- ▶ Up until now you have encountered a few different types of statistical inference:
 - ▶ null-hypothesis testing
 - ▶ (ordinary) least squares
 - ▶ maximum likelihood
- ▶ Bayesian inference is another type of statistical inference; it is not a model.

Example: coin flip

- ▶ We have a coin that may or may not be biased and we interested in finding out the degree of bias towards 'heads' (if one exists)
- ▶ the coin has been flipped 20 times and a total of 15 heads were observed
- ▶ First step: implement a model that describes how the data could be generated

Example: coin flip

- ▶ Binomial model is a good choice for this exercise.
- ▶ Binomial model has two parameters:
 - ▶ n = Number of trials and the probability of 'success' (p). n =
 - ▶ p = probability of 'success'

Coin flip: Null hypothesis testing

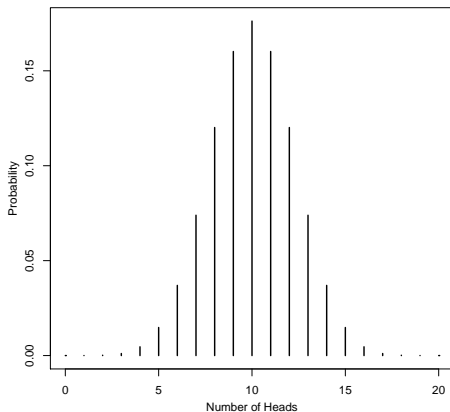
Coin flip: Null hypothesis testing

Probability of getting a value as or more extreme than the value observed given the Null hypothesis

Coin flip: Null hypothesis testing

- ▶ Null hypothesis: $p = 0.5$
- ▶ Calculate $p(H \geq h | p = 0.5)$
- ▶ see if this quantity falls below a certain threshold (often set to be 0.05)
- ▶ If it does, then reject the Null hypothesis

Coin flip: Null hypothesis testing



$$p(H \geq h | p = 0.5) = 0.02 \quad p \neq 0.5$$

Is $p = 0.51$?

Maximum Likelihood: Coin flip

- ▶ Coin: estimate the probability of getting head
- ▶ Flip the coin 20 times: 15 heads
- ▶ Construct a mathematical simplification of the study system that encapsulates the phenomena that you are interested in.
 - ▶ binomial distribution
 - ▶ n: number of trials
 - ▶ p: probability of success

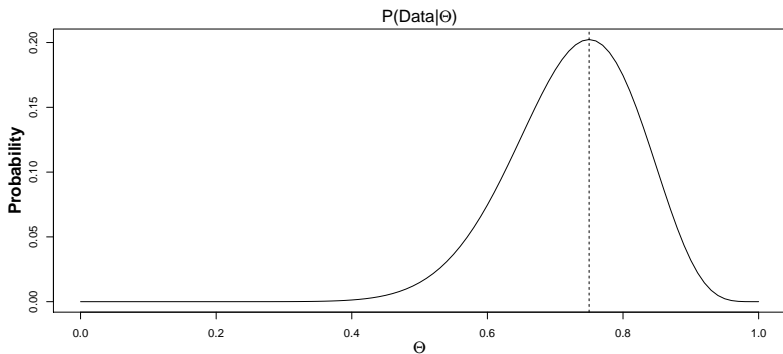
$$P(H = h|p) = \frac{n!p^h(1-p)^{n-h}}{h!(n-h)!}$$

h: number of heads

Likelihood Function

$$L(p) = P(H = h|p)$$

Probability of obtaining data as function of parameter



Maximum Likelihood: Coin flip

$$L(p) = \frac{n! p^h (1-p)^{n-h}}{h! (n-h)!}$$

$$L(p) = \frac{n!}{h! (n-h)!} p^h (1-p)^{n-h}$$

$$l(p) = \ln\left(\frac{n!}{h! (n-h)!}\right) + \ln(p)h + \ln(1-p)(n-h)$$

n: number of trials

h: number of heads

Likelihood

$$l(p) = \ln\left(\frac{n!}{h!(n-h)!}\right) + \ln(p)h + \ln(1-p)(n-h)$$

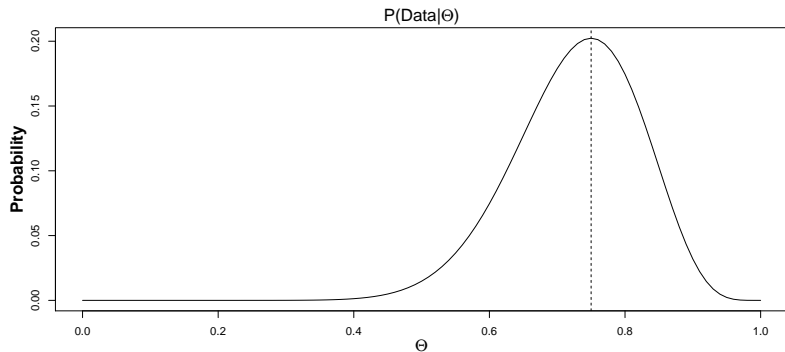
Derivative and set 0

$$\frac{dl(p)}{dp} = \frac{h}{p} - \frac{n-h}{1-p} = 0$$

$$h - hp = np - hp$$

$$p = \frac{h}{n} \quad q.e.d.$$

Likelihood function: Graphic



- Flip the coin 20 times: 15 heads

Maximum likelihood (classical inference): $p = 0.75$

Bayesian inference vs maximum likelihood

Maximum Likelihood:

$$\hat{\theta} = \arg \max \{P(Data|\Theta)\}$$

Probability of getting a data set given a parameter value

Bayesian inference

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$

Probability of a parameter value given a data set

Bayesian inference:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$

Posterior : Probability of parameter value given data

Likelihood : Probability of data given parameter value

Prior : Probability of parameter value

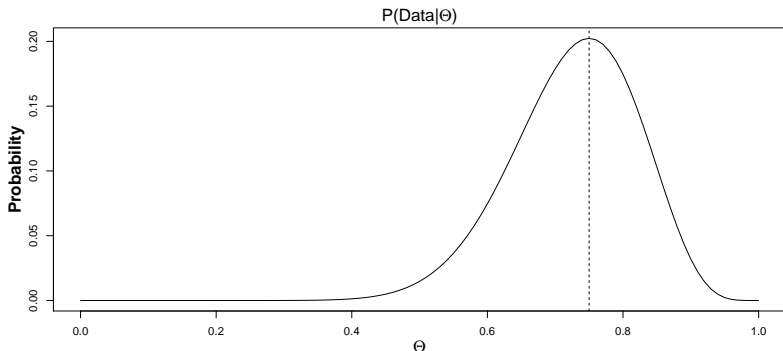
Normalizing constant

$$P(\Theta|Data) \propto P(Data|\Theta)P(\Theta)$$

Likelihood

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$

Likelihood : Probability of data given parameter value



Prior: Knowledge before looking at the data

- ▶ The prior takes into account all the information you know about the system before looking at the data
- ▶ The prior is usually expressed as a probability density/mass function
- ▶ If something is known about the system then probability functions with more weight on expected parameter values can be given. These are often called informative or sometimes subjective priors
- ▶ If nothing is known about the system then the investigator can specify flat looking priors. Flat-shaped priors are often called vague, uninformative, objective or minimally informative priors

Prior: Knowledge before looking at the data

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$

Prior : Probability of parameter value

Monty Hall problem: $\frac{1}{3}$ probability of having car/chocolate behind/under each door/cup

Prior:

- ▶ Don't use a probability distribution for the prior that gives zero weight to a parameter value that is possible
 - ▶ **Cromwell's rule** "I beseech you, in the bowels of Christ, think it possible that you may be mistaken" – Oliver Cromwell in a letter to the Synod of the Church of Scotland
- ▶ Do use an informative prior if there is information available
 - ▶ Previous research:
 - ▶ Field data
 - ▶ Meta-analysis of published literature
 - ▶ Known climatic, physical, or biological tolerances
- ▶ Do use a probability distribution for the prior that gives zero weight to a parameter value that is impossible such as:
 - ▶ Values outside of the range 0-1 for parameters that describe a probability
 - ▶ Values less than zero for measures of weight, length etc.
 - ▶ Values that represent some form of physical or biological impossibility

Prior: coin flip

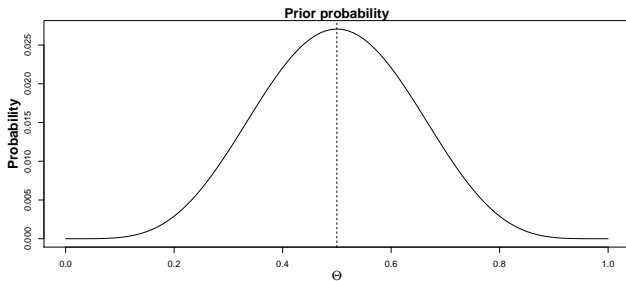
Probability of heads: p

Any prior knowledge?

Prior:

Coin flip: reasonable to set a prior centered on $p = 0.5$
(i.e. probability of heads and tails is equal)

20 flips with another coin, 10 heads



Beta distribution: two parameters equivalent to number of heads and number of tails in a coin flipping experiment

Normalizing constant:

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$

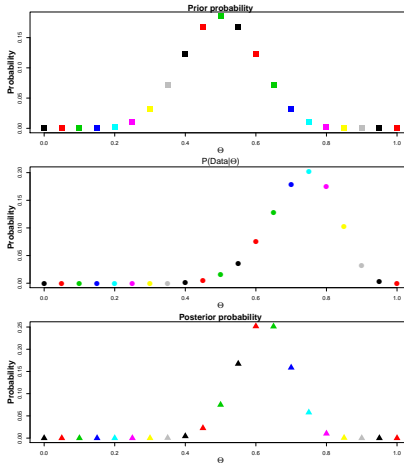
- ▶ The normalizing constant exists to ensure that the resulting posterior distribution is proper (sums to 1)
- ▶ the normalizing constant does not depend on Θ

$$P(\Theta|Data) \propto P(Data|\Theta)P(\Theta)$$

Posterior: Normalized product of Likelihood and Prior

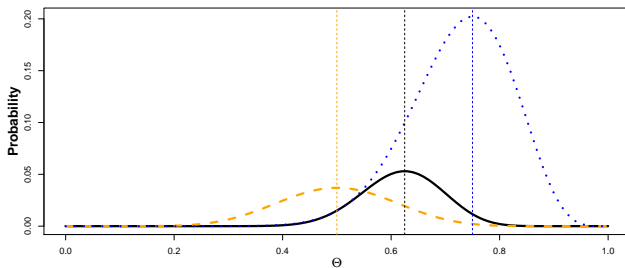
$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$

Posterior : Probability of parameter value given data



Posterior: Normalized product of Likelihood and Prior

$$P(\Theta|Data) = \frac{P(Data|\Theta)P(\Theta)}{P(Data)}$$



Exercise 1

- ▶ Change number of coin flips used to build the prior
 - ▶ no coin flips
 - ▶ half the the data
 - ▶ twice the data
- ▶ Change number of coin flips used as data
- ▶ only four coin flips
- ▶ twice the number of prior flips
- ▶ five times the number of prior flips
- ▶ divide your coin flips into two parts:
 - ▶ 20 flips then use the new posterior as prior and add an additional 20 coin flips
 - ▶ compare this to 40 coin flips

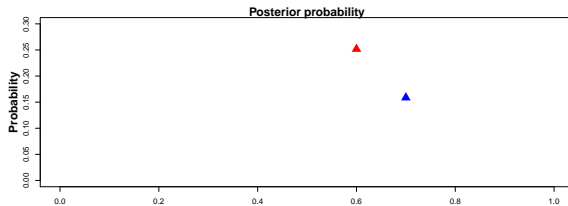
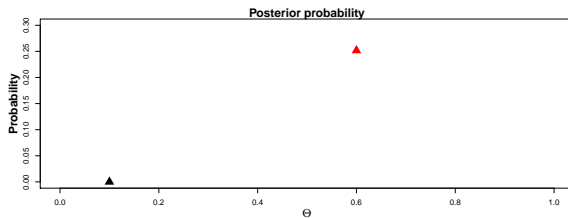
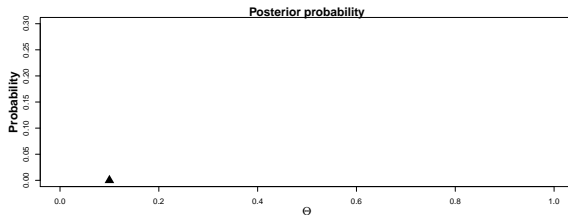
Posterior:

- ▶ more observations increase importance of likelihood compared to prior
- ▶ add data step wise and obtain the same posterior
- ▶

Sampling: Markov Chain Monte Carlo

- ▶ Draw parameter Θ at random
- ▶ Compute posterior probability for this parameter value
- ▶ Select a new parameter value (many different selection options)
- ▶ Compute posterior probability of proposed value
- ▶ ratio $\frac{P_{prop}}{P_{old}}$
- ▶ if ratio > 1 move else ratio is probability of moving
- ▶

Sampling: Markov Chain Monte Carlo



Sampling: MCMC

- ▶ Person wants to visit all states in the US repeatedly, number of visits depends on population size
- ▶ Start with a randomly chosen state
- ▶ Draw a new state at random
- ▶ Compare populations of the two states
- ▶ ratio $\frac{P_{prop}}{P_{old}}$
- ▶ if ratio > 1 move else ratio is probability of moving

Repeat many times

Proportion of visits will converge to proportion of population

Linear regression

- ▶ Nearly all forms of statistical analysis follow these basic steps:
 - ▶ Construct a mathematical simplification of the study system that encapsulates the phenomena that you are interested in.
 - ▶ linear regression:

$$y_i = a + bx_i + e_i \quad e \sim N(0, s^2)$$

$$y_i \sim N(a + bx_i, s^2)$$

- ▶ Compare the model with the data

$$P(y|a, b, s^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{1}{2} \left(\frac{a + bx_i - y_i}{s} \right)^2}$$

- ▶ **Maximum Likelihood** Choose a , b , and s^2 so that probability is maximized

Linear regression

- Compare the model with the data

$$P(y|a, b, s^2) = \prod_{i=1}^n C \frac{1}{s} e^{-\frac{1}{2} \left(\frac{a + bx_i - y_i}{s} \right)^2}$$

$$P(y|a, b, s^2) \propto \prod_{i=1}^n \frac{1}{s} e^{-\frac{1}{2} \left(\frac{a + bx_i - y_i}{s} \right)^2}$$

$$P(a, b, s^2|y) \propto P(y|a, b, s^2)P(a)P(b)P(s^2)$$

Prior for a ?

Prior for b ?

Prior for s^2 ?

Linear regression: result

With MCMC and 10000 samples from the posterior

10000 values of a , b , and s^2

10000 possible regression models

