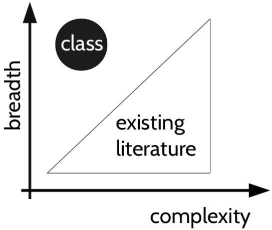
Bayesian Data Analysis in a Nutshell

An introduction to Geneticists and Breeders

Felipe Ferrão
Research Assistant Scientist
lferrao@ufl.edu
lfelipeferrao

June 15th, 2021

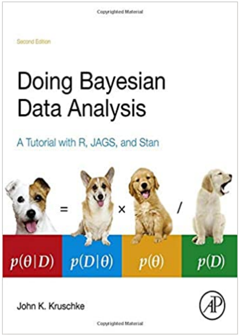
Introduction



- This class (should be) is a friendly introduction to Bayesian Data Analysis.
- My expectation:
  - Understand the general principles;
  - Keep mathematics to a minimum and focus instead on the intuition;
  - Implement some real analysis in R.

Introduction

This class was inspired by:



- There are many good books, videos and online Bayesian courses
- Chapters 1,2,3,4 and 5 from Kruschke (2014)
- A Student's Guide to Bayesian Statistics (Lambert, 2018) – also a very good material

Introduction

Bayesian Framework

- Bayesian involves reallocation of credibility across possibilities.
- All possibilities, over which we allocate credibility, are parameters in a mathematical model.

University of Florida

Introduction to Bayesian Statistics

https://lfelipe-ferrao.github.io/class/Bayesian/

Introduction

- You landed in a given airport and just left the airplane.
- Where is my luggage?
  - ▷ Prior knowledge: if you don't see any prohibitive signs, all directions are possible.
  - ▷ Data 1: you can read the signals.
  - ▷ Data 2: observe other passenger.
  - ▷ Data 3: you can ask for help !

- Credibility could be reallocated based on observation collected plus prior knowledge
- Reallocation of credibility across probabilities is the core of Bayesian inference.

Bayesian analysis will formalize this re-allocation in a logically, coherent and precise way!!

Introduction

- You landed in a given airport and just left the airplane.
- Where is my luggage?
  - ▷ Prior knowledge: if you don't see any prohibitive signs, all directions are possible.
  - ▷ Data 1: you can read the signals.
  - ▷ Data 2: observe other passenger.
  - ▷ Data 3: you can ask for help !

- Credibility could be reallocated based on observation collected plus prior knowledge
- Reallocation of credibility across probabilities is the core of Bayesian inference.

Bayesian analysis will formalize this re-allocation in a logically, coherent and precise way!!

Introduction

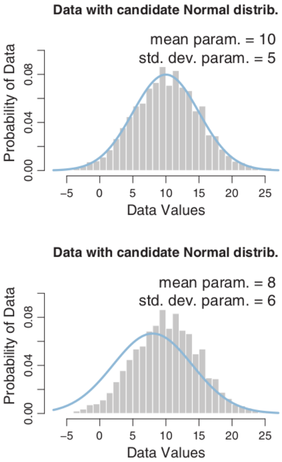
- You landed in a given airport and just left the airplane.
- Where is my luggage?
  - ▷ Prior knowledge: if you don't see any prohibitive signs, all directions are possible.
  - ▷ Data 1: you can read the signals.
  - ▷ Data 2: observe other passenger.
  - ▷ Data 3: you can ask for help !

- Credibility could be reallocated based on observation collected plus prior knowledge
- Reallocation of credibility across probabilities is the core of Bayesian inference.

Bayesian analysis will formalize this re-allocation in a logically, coherent and precise way!!

Introduction

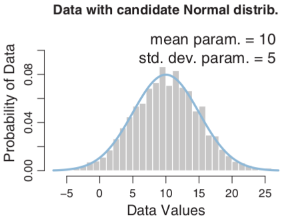
Possibilities are parameters values in a model!

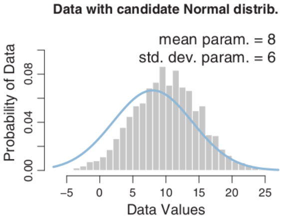


- Possibilities can be represented as mathematical functions.
- Probability distribution
- Same histogram and two candidates of parameters that create the curve.
- What group of parameters better describe the data?
- We use Bayesian inference to compute credibility of parameter values.

Introduction

Possibilities are parameters values in a model!

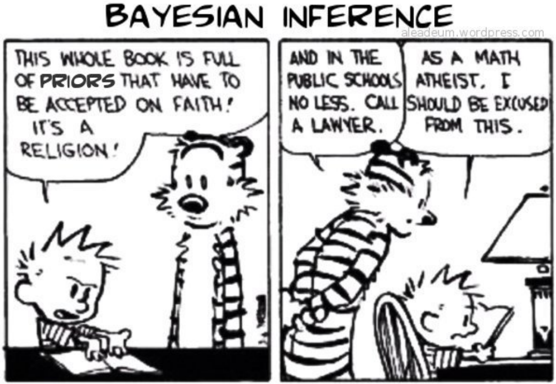




- Possibilities can be represented as mathematical functions.
- Probability distribution
- Same histogram and two candidates of parameters that create the curve.
- What group of parameters better describe the data?
- We use Bayesian inference to compute credibility of parameter values.

Probability

What is this stuff called probability?



Probabilities

- All uncertainty in terms of parameter estimation, in a Bayesian context, is probabilistic measured.
- What is this stuff called probability?
- Before define the Bayes'rule, we need a general understanding about probabilities

In the sequence ..

- Probability under different point of views
- Probability Distributions
- Discrete vs. Continuous
- Intercorrelations among common statistical distributions
- Kolmogorov's Three Axioms
- Deduce the Bayes'rule using simple probability rules

Probabilities

- All uncertainty in terms of parameter estimation, in a Bayesian context, is probabilistic measured.
- What is this stuff called probability?
- Before define the Bayes'rule, we need a general understanding about probabilities

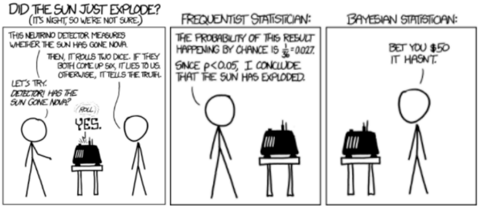
In the sequence ..

- Probability under different point of views
- Probability Distributions
- Discrete vs. Continuous
- Intercorrelations among common statistical distributions
- Kolmogorov's Three Axioms
- Deduce the Bayes'rule using simple probability rules

Probabilities

General Understanding

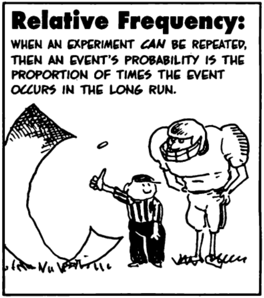
The general understanding about probabilities create two representative schools of thought: frequentist vs. Bayesian.



Probabilities

Outside the head

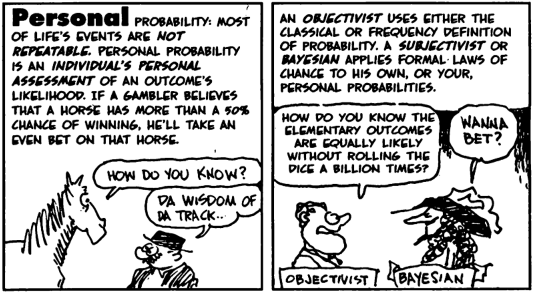
- Consider the idea to flip the coin.
- We can ask if the coin is fair, depending the number of come up heads.
- Ex: 50% of heads in a long run is a good indication that the coin is fair.
- Probability of head and the degree of belief is dictated as the limit of its relative frequency in many trials (long-run frequency).



Probabilities

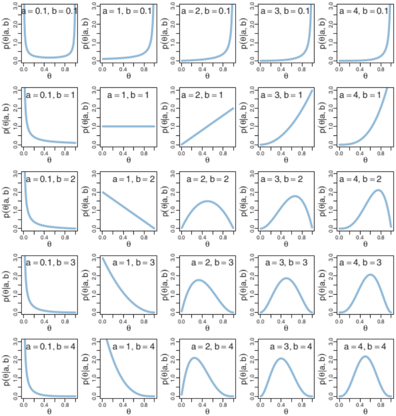
Inside the head

- What is the probability to rain tomorrow?
- We cannot use long-run relative frequency to answer this question.
- Specify subjective beliefs and how likely we think each possible outcome is.
- For example, we can check some historical data.
- We are calibrating a subjective belief.



Probabilities

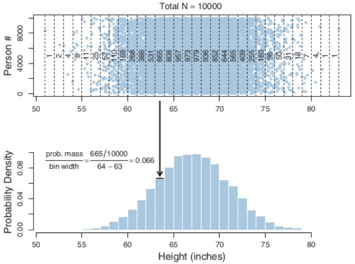
- When there are several possible outcomes, too much effort calibrate the subjective belief about every possible outcome.
- This is why we can use mathematical functions (**Probability Distribution**)



- Probability distribution is a mathematical function that gives the probabilities of occurrence of possible outcomes.
- Parameters can be interpreted as control knobs on mathematical devices that simulate the data generation.
- Beta distribution
- Two parameters (a,b)

Probabilities

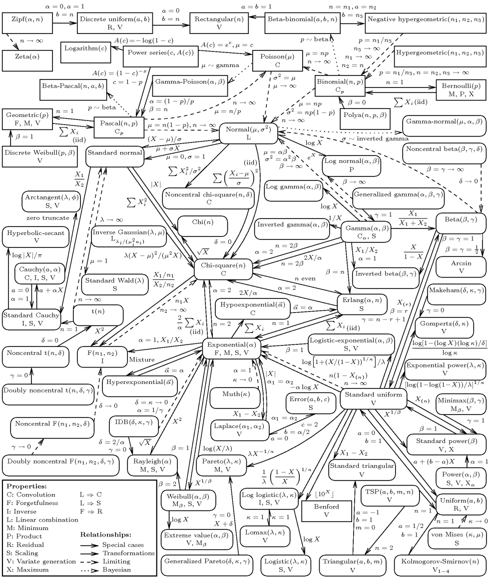
- Discrete: discrete outcomes, then we can talk about the probability of each distinct outcome. For example: Binomial, Poisson, Negative Bionomial
- Continues:continuous space and is problematic to talk about the probability of a specific value. Example: normal, t-distribution.



- Height (continuous outcome spaces)
- The probability that a randomly selected person has height of exactly 67.21413908 ... is essentially nil -> Continuous outcomes
- We can discretize the space into a finite number of intervals (bins).
- We can talk about the probability into intervals -> Discrete outcomes

Probabilities

Interrelations among Statistical Distributions

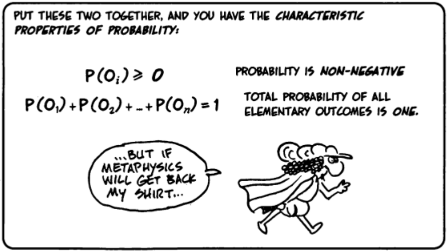


Probabilities

What all these distributions have in common...

Need to satisfy the Kolmogorov's Three Axioms

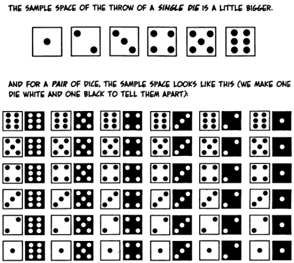
- Probability value must be non-negative;
- Sum of the probabilities across all events in the entire sample space must be one;
- For any two mutually exclusive events, the probability that one or the other occurs is the sum of their individual probability.



Probabilities

Two-way distributions

- In order to deduce the Bayes' rule, we need to understand some important properties of probability.
- A key example is the conjunction of two outcomes.
- Example: *What is the probability of meeting a person with both red hair and green eyes?*



Probabilities

Let's check the hair and eye color example

	Hair color				
Eye color	Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- In each of its main cells, the table indicates the **joint probability** of particular combinations of eye color and hair color. Notation: $p(e, h) = p(h, e)$
- We may be interested in the probabilities of the eye colors overall, collapsed across hair colors. These probabilities are indicated in the right margin, and they are called **marginal probabilities**. They are computed simply by summing the joint probabilities in each row, to produce the row sums. Notation: $p(h) = \sum_e p(e, h)$

Probabilities

Let's check the hair and eye color example

	Hair color				
Eye color	Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- In each of its main cells, the table indicates the **joint probability** of particular combinations of eye color and hair color. Notation: $p(e, h) = p(h, e)$
- We may be interested in the probabilities of the eye colors overall, collapsed across hair colors. These probabilities are indicated in the right margin, and they are called **marginal probabilities**. They are computed simply by summing the joint probabilities in each row, to produce the row sums. Notation: $p(h) = \sum_e p(e, h)$

Probabilities

Let's check the hair and eye color example

	Hair color				
Eye color	**Black**	**Brunette**	**Red**	**Blond**	**Marginal (eye color)**
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- In each of its main cells, the table indicates the **joint probability** of particular combinations of eye color and hair color. Notation: $p(e, h) = p(h, e)$
- We may be interested in the probabilities of the eye colors overall, collapsed across hair colors. These probabilities are indicated in the right margin, and they are called **marginal probabilities**. They are computed simply by summing the joint probabilities in each row, to produce the row sums. Notation: $p(h) = \sum_e p(e, h)$

Probabilities

Conditional Probabilities

Eye color	Hair color: Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- We often want to know the probability of one outcome, given that we know another outcome is true.
- Example: suppose I tell you that this person has blue eyes. Conditional on that information, what is the probability that the person has blond hair?

$$p(blue) = 0.36$$

$$p(blue, blond) = 0.16$$

$$p(blue|blond) = 0.16/0.36 = 0.45$$

Probabilities

Conditional Probabilities

Eye color	Hair color: Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- We often want to know the probability of one outcome, given that we know another outcome is true.
- Example: suppose I tell you that this person has blue eyes. Conditional on that information, what is the probability that the person has blond hair?

$$p(blue) = 0.36$$

$$p(blue, blond) = 0.16$$

$$p(blue|blond) = 0.16/0.36 = 0.45$$

Probabilities

When the Bayesian inference happens

Eye color	Hair color: Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- Probability of hair colors represent the credibility of each possible hair color.
- For this group of people, the general probability of having blond hair is 0.21.
- But when we learn that a person from this group has blue eyes, then the credibility of that person having blond hair increases to 0.45.
- This reallocation of credibility across the possible hair colors is Bayesian inference!

Probabilities

When the Bayesian inference happens

Eye color	Hair color: Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- Probability of hair colors represent the credibility of each possible hair color.
- For this group of people, the general probability of having blond hair is 0.21.
- But when we learn that a person from this group has blue eyes, then the credibility of that person having blond hair increases to 0.45.
- This reallocation of credibility across the possible hair colors is Bayesian inference!

Probabilities

When the Bayesian inference happens

Eye color	Hair color				Marginal (eye color)
	Black	Brunette	Red	Blond	
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

- Probability of hair colors represent the credibility of each possible hair color.
- For this group of people, the general probability of having blond hair is 0.21.
- But when we learn that a person from this group has blue eyes, then the credibility of that person having blond hair increases to 0.45.
- This reallocation of credibility across the possible hair colors is Bayesian inference!

Probabilities

Example of Conditional Probability

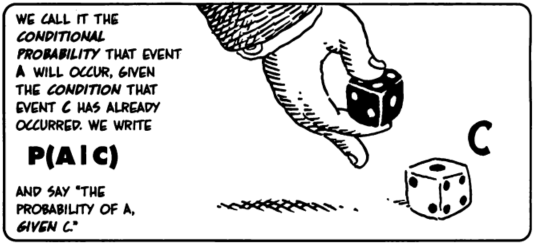
	Hair color				
Eye color	**Black**	**Brunette**	**Red**	**Blond**	**Marginal (eye color)**
Blue	$0.03/0.36 = 0.08$	$0.14/0.36 = 0.39$	$0.03/0.36 = 0.08$	$0.16/0.36 = 0.45$	$0.36/0.36 = 1.0$

Of the blue-eyed people in Table 4.1, what proportion have hair color h? Each cell shows $p(h|\text{blue}) = p(\text{blue}, h)/p(\text{blue})$ rounded to two decimal points.

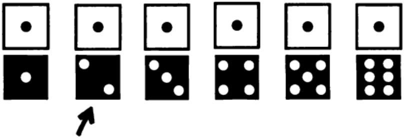
- Computations for conditional probability can be denoted by simple expression.
- We denote the conditional probability of hair color given eye color as $p(h|e)$, which is spoken "the probability of h given e."
- Calculation: $p(h|e) = \frac{p(e,h)}{p(e)}$

Probabilities

Example of Conditional Probability



BEFORE ANY DICE WERE THROWN, THE SAMPLE SPACE HAD 36 OUTCOMES, BUT NOW THAT THE EVENT C HAS OCCURRED, THE OUTCOME MUST BELONG TO THE *REDUCED SAMPLE SPACE C.*



IN THE REDUCED SAMPLE SPACE OF SIX ELEMENTARY OUTCOMES, ONLY ONE OUTCOME (1,2) SUMS TO 3. SO THE CONDITIONAL PROBABILITY IS 1/6.

Bayes' Rule

Now, we are ready to formally describe the Bayes' rule

Bayes' Rule

What we know so far

- We know that beliefs (or credibility) can be represented in the format of probability distribution;
- Probability, need to satisfy some rules (the Kolmogorov's Axioms)
- Conditional and marginal probabilities can be easily determined using two-way discrete tables.
- Most important: Reallocation of credibility (for example, across the possible hair colors) is Bayesian inference!

What is the next step

Formalize the Bayes'rule: mathematical relation between the prior allocation of credibility and the posterior reallocation of credibility conditional on data.

Bayes' Rule

What we know so far

- We know that beliefs (or credibility) can be represented in the format of probability distribution;
- Probability, need to satisfy some rules (the Kolmogorov's Axioms)
- Conditional and marginal probabilities can be easily determined using two-way discrete tables.
- Most important: Reallocation of credibility (for example, across the possible hair colors) is Bayesian inference!

What is the next step

Formalize the Bayes'rule: mathematical relation between the prior allocation of credibility and the posterior reallocation of credibility conditional on data.

Baye'rule

Below a simple derivation of the Bayes rule only using simple probability rules of conditional and marginal distributions.

$$p(r, c) = p(c, r)$$

$$p(r|c) = \frac{p(r, c)}{p(c)} \iff p(r, c) = p(r|c) \times p(c)$$

$$p(c|r) = \frac{p(c, r)}{p(r)} \iff p(c|r) = \frac{p(r|c) \times p(c)}{p(r)}$$

- What is key here is how we can move from $p(r|c) \iff p(c|r)$.
- It is very important, because $p(r|c) \neq p(c|r)$.
- It was coined as "inverse probability" in the past.

Baye'rule

Below a simple derivation of the Bayes rule only using simple probability rules of conditional and marginal distributions.

$$p(r,c) = p(c,r)$$

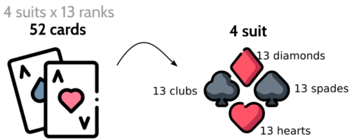
$$p(r|c) = \frac{p(r,c)}{p(c)} \iff p(r,c) = p(r|c) \times p(c)$$

$$p(c|r) = \frac{p(c,r)}{p(r)} \iff p(c|r) = \frac{p(r|c) \times p(c)}{p(r)}$$

- What is key here is how we can move from $p(r|c) \iff p(c|r)$.
- It is very important, because $p(r|c) \neq p(c|r)$.
- It was coined as "inverse probability" in the past.

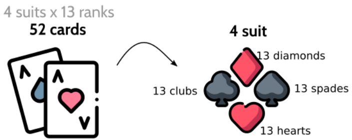
Baye'rule

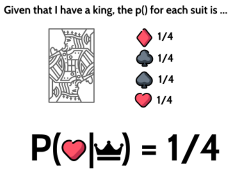
Simple example why $p(r|c) \neq p(c|r)$



Baye'rule

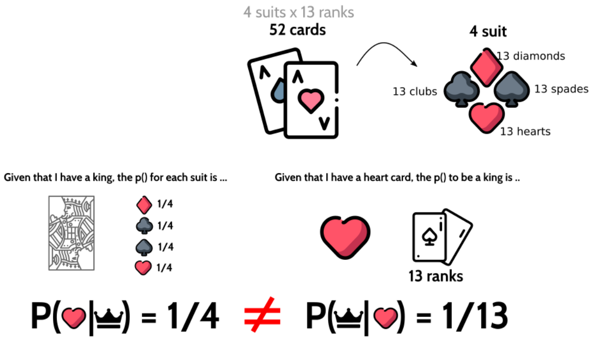
Simple example why $p(r|c) \neq p(c|r)$





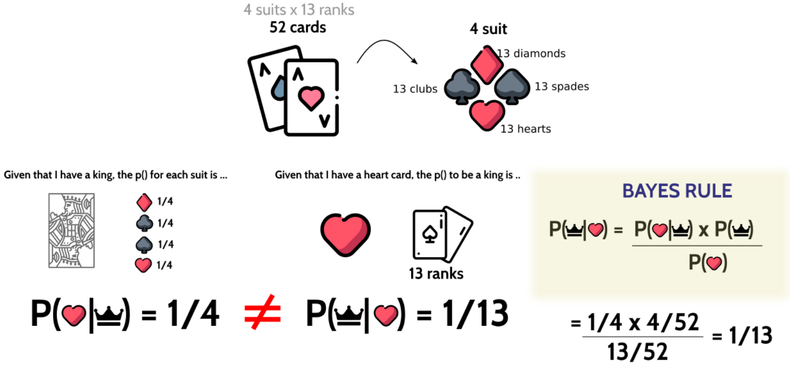
Baye'rule

Simple example why $p(r|c) \neq p(c|r)$



Baye'rule

Simple example why $p(r|c) \neq p(c|r)$



Baye'rule

Applied to parameters and data

- So far, we are using Bayes'rule in examples were the joint probabilities are directly provided as numerical values;
- In a real word we can use Bayes'rule applied to parameter and data;
- **p(data values|parameters)**: a model of data specifies the probability of particular data values given the model's structure and parameter values.
- **p(parameters|data values)**: we use Bayes' rule to convert that to what we really want to know, which is how strongly we should believe in the various parameter values

Baye'rule

Applied to parameters and data

- So far, we are using Bayes'rule in examples were the joint probabilities are directly provided as numerical values;
- In a real word we can use Bayes'rule applied to parameter and data;
- **p(data values|parameters)**: a model of data specifies the probability of particular data values given the model's structure and parameter values.
- **p(parameters|data values)**: we use Bayes' rule to convert that to what we really want to know, which is how strongly we should believe in the various parameter values

Baye'rule

Factors of Bayes' rule have specific names

Likelihood: the probability that the data could be generated by the model with parameter value (**θ**)

Prior: the prior is the credibility value of **θ** without the data.

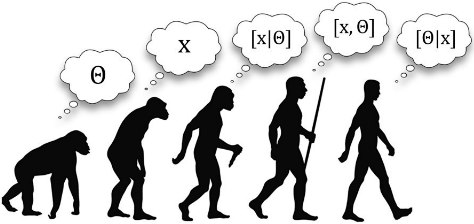
$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)}$$

Posteriori: is the credibility of **θ** with the data (D) taken into account

Evidence: is 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.

Estimating Bias in a Coin

Let's consider a real example



Estimating Bias in a Coin

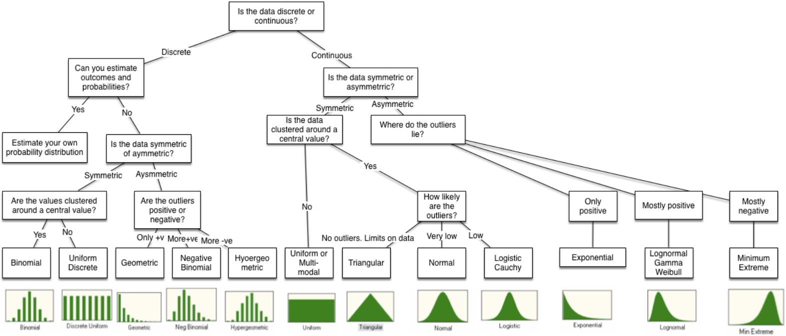
Objectives

- Pipeline for Bayesian data Analysis;
- Influence of the sample size on the posterior;
- Influence of the prior on the posterior;
- Why Bayesian inference can be difficult;

Estimating Bias in a Coin

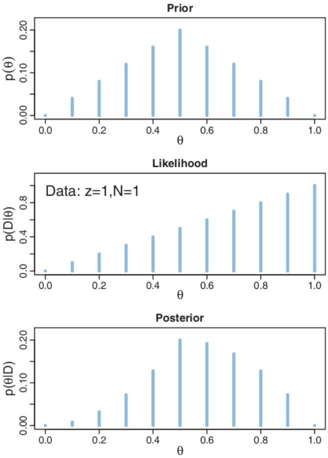
Pipeline of Bayesian analyses

1. Identifying the type of data being described.
2. Creating a descriptive model with meaningful parameters. For that, we need a mathematical expression of the likelihood function in Bayes' rule.



Estimating Bias in a Coin

Pipeline of Bayesian analyses



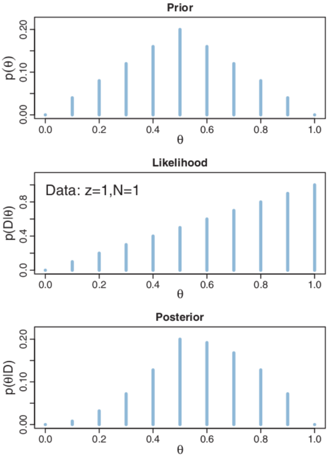
3 Establishing a prior distribution over the parameter values. This could be trick. Here, we believe that the factory tends to produce fair coins ($\theta = 0.5$);

4 Collecting the data and applying Bayes' rule to re-allocate credibility across the possible parameter values. Suppose that we flip the coin once and observe heads.

5 Compute the posteriori distribution for the parameters.

Estimating Bias in a Coin

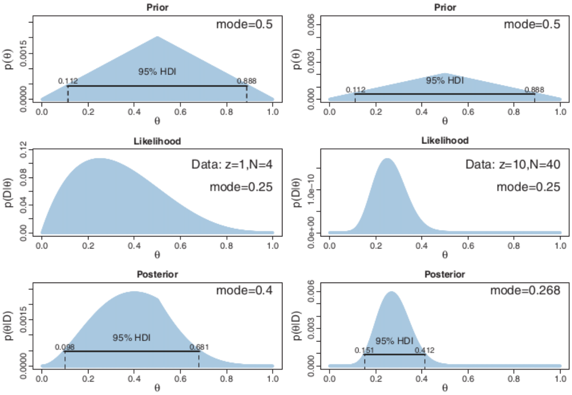
General conclusions



- The overall contour of the posterior distribution is different from the prior distribution.
- Despite the data showing 100% heads (in a sample consisting of a single flip), the posterior of large θ values is low.
- This illustrates a general phenomenon in Bayesian inference: The posterior is a compromise between the prior distribution and the likelihood function.

Estimating Bias in a Coin

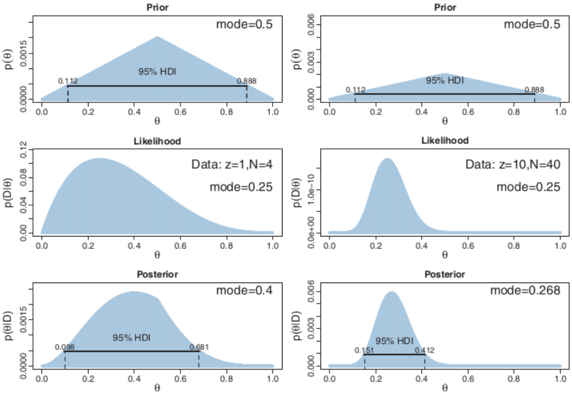
What about the influence of sample size?



- More data, the more precise is the estimate of the parameter(s) in the model.
- Larger sample sizes yield greater precision or certainty of estimation.
- Less influence of the prior !

Estimating Bias in a Coin

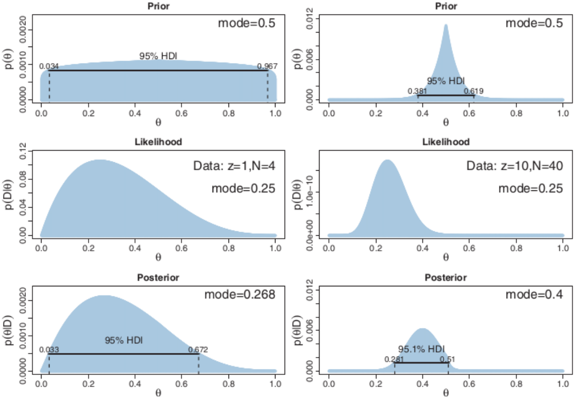
What about the influence of sample size?



- More data, the more precise is the estimate of the parameter(s) in the model.
- Larger sample sizes yield greater precision or certainty of estimation.
- Less influence of the prior !

Estimating Bias in a Coin

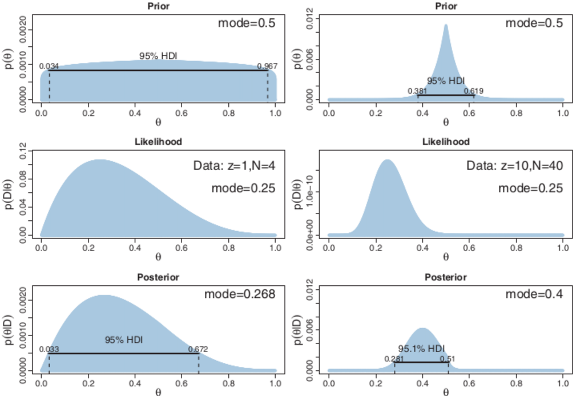
What about the influence of the prior?



- Posterior is very close to the likelihood function. If prior distribution is relatively broad compared with the likelihood, prior has little influence on the posterior.
- Prior is so sharp that the posterior distribution is noticeably influenced by the prior.

Estimating Bias in a Coin

What about the influence of the prior?



- Posterior is very close to the likelihood function. If prior distribution is relatively broad compared with the likelihood, prior has little influence on the posterior.
- Prior is so sharp that the posterior distribution is noticeably influenced by the prior.

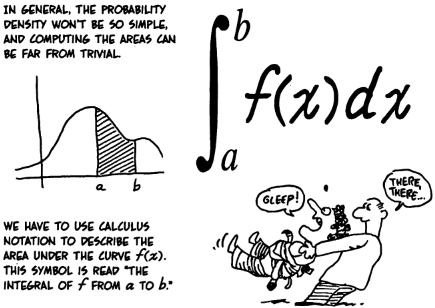
Conclusions

- This is why Bayesian inference is intuitively rational
- A strong prior that uses a lot of previous data to put high credibility over a narrow range of parameter values has a large influence.
- With a weakly prior that spreads credibility over a wide range of parameter values, it takes relatively little data to shift the peak of the posterior distribution toward the data.
- Posteriori is a compromise between prior elicitation and data observations

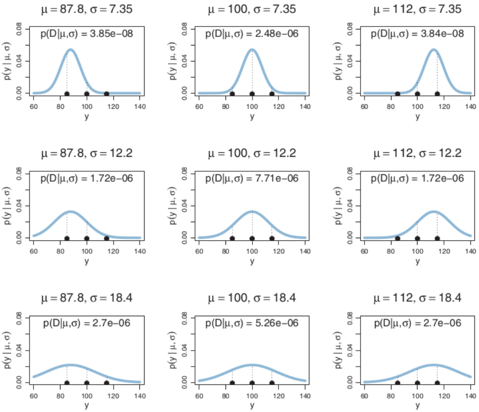
Final Considerations

Why Bayesian inference can be difficult?

- To be able to use Bayesian analysis we must ensure that the posterior is a probability distribution.
- The denominator is a number that ensure it by normalising the numerator.
- It requires complex integrals, it can become intractable in higher dimensions !



Final Considerations



Why Bayesian inference can be difficult?

- Let's look first to a normal probability density function.
- $p(y \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Three data values y1 = 85, y2 = 100, and y3 = 115 were considered
- Given the data, how should we allocate credibility for μ and σ?

Final Considerations

Why Bayesian inference can be difficult?

- The answer this question we need to use the Bayes' rule and solve the follow equation:

$$p(\mu, \sigma \mid D) = \frac{p(D \mid \mu, \sigma)\ p(\mu, \sigma)}{\iint d\mu\ d\sigma\ p(D \mid \mu, \sigma)\ p(\mu, \sigma)}$$

- Difficulty in computing the denominator of the Bayes' rule application.
- More complex problems with multiple parameters we need have multiple-dimensional integrals to be solved

Final Considerations

Why Bayesian inference can be difficult?

- The answer this question we need to use the Bayes' rule and solve the follow equation:

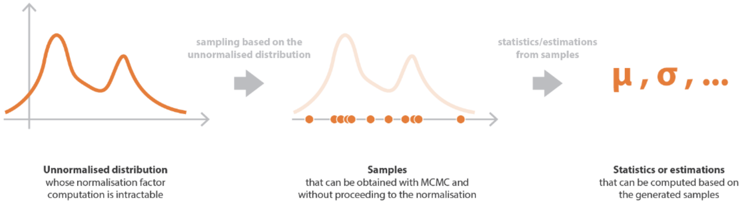
$$p(\mu,\sigma \mid D) = \frac{p(D \mid \mu,\sigma)\ p(\mu,\sigma)}{\iint d\mu\ d\sigma\ p(D \mid \mu,\sigma)\ p(\mu,\sigma)}$$

- Difficulty in computing the denominator of the Bayes' rule application.
- More complex problems with multiple parameters we need have multiple-dimensional integrals to be solved

Final Considerations

Why Bayesian inference can be difficult?

- Simplest solution: restricting models to simple likelihood with corresponding prior distributions (conjugate priors) to produce a tractable integral.
- MCMC: randomly sampling a large number of representative combinations of parameter values from the posterior distribution. They can generate representative parameter-value combinations of complex models without computing integrals.



Final Considerations

More details about MCMC

- To explain MCMC we would need more time
- Chapter 7 from Doing Bayesian Data Analysis (Kruschke, 2014)
- Chapter 12,13,14 and 15 from A Student's Guide to Bayesian Statistics (Lambert, 2018)
- More details in the Hands-on

Final Considerations

Final Considerations

- Thank you !!
- Questions ??
- Hands-on
  - Introduction to Probabilistic Language (rstan)
  - Use of Bayesian Statistics for genomic prediction (BGLR package)

University *of* Florida

Introduction to
Bayesian
Statistics

https://lfelipe-ferrao.github.io/class/Bayesian/