AARHUS UNIVERSITY

Methods 4 - 1

Chris Mathys

BSc Programme in Cognitive Science

Spring 2023

Methods 4 – Overview

- Wraps up the series of methods courses
- Modelling from a birds-eye view
- Advanced concepts
- Revisiting methods introduced in earlier methods courses
- Comprehensive Bayesian perspective

Methods 4 – Overview

Advanced concepts:

- Causal reasoning using directed acyclic graphs (DAGs)
- Mixture models
- Gaussian processes
- Measurement error
- Missing data
- [Stan programming]

Methods 4 – Overview

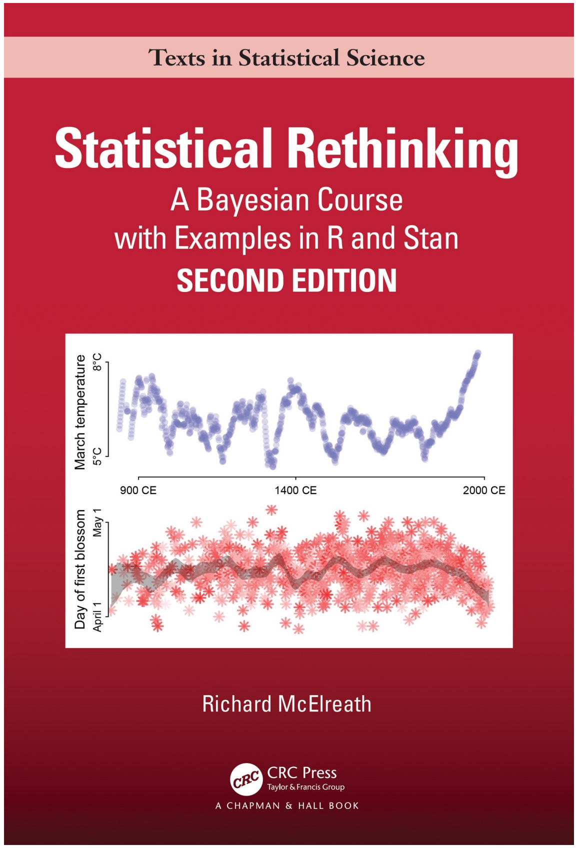
Revisit:

- Regression modelling
- Generalized linear models
- Multilevel modelling
- Markov chain Monte Carlo sampling, Hamiltonian Monte Carlo sampling
- Learning to fit models using probabilistic programming

Resources

Textbook:

[1]McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan (2nd ed.).* Chapman and Hall/CRC. doi:10.1201/9780429029608



Schedule

Course week	Week of year	Topics and readings
1	5	Statistical models (chapters 1,2)
2	6	Distributions and sampling (chapters 2,3)
3	7	Gaussian models and linear regression (chapter 4)
4	8	Several predictors, directed acyclic graphs (chapters 5)
5	9	Causal inference (chapter 6)
6	10	Model comparison (chapter 7)
7	12	Interactions (chapter 8)
8	13	Markov chain Monte Carlo, maximum entropy (chapters 9, 10)
9	14	Generalized linear models (chapters 11)
10	16	Mixture models, ordered categorical outcomes/predictors (chapter 12)
11	17	Multilevel models (chapter 13)
12	18	Varying slopes, Gaussian processes (chapter 14)
13	19	Measurement error, missing data, theory-driven models (chapters 15, 16)

Schedule

Course week	Week of year	Topics and readings
1	6	Statistical models (chapters 1,2)
2	7	Distributions and sampling (chapters 2,3)
3	8	Gaussian models and linear regression (chapter 4)
4	10	Several predictors, directed acyclic graphs (chapters 5)
5	11	Causal inference (chapter 6)
6	13	Model comparison (chapter 7)
7	15	Interactions (chapter 8)
8	16	Markov chain Monte Carlo, maximum entropy (chapters 9, 10)
9	17	Generalized linear models (chapters 11)
10	18	Mixture models, ordered categorical outcomes/predictors (chapter 12)
11a	19	Multilevel models (chapter 13) [only lecture]
11b	21	Multilevel models (chapter 13) [only class]

Resources

Author's current videos:

- 2023 lectures: https://www.youtube.com/playlist?list=PLDcUM9US4XdPz-KxHM4XHt7uUVGWWVSus

Author's older videos:

- 2022 lectures: https://www.youtube.com/playlist?list=PLDcUM9US4XdMROZ57-OIRtIK0aOynbgZN
- 2019 lectures: https://www.youtube.com/playlist?list=PLDcUM9US4XdNM4Edgs7weiyIguLSToZRI

Resources

Code:

- This course's main repository:
  https://github.com/methods-4-f23/methods-4-resources

- R package (please install this!):
  https://github.com/rmcelreath/rethinking

Exam

https://kursuskatalog.au.dk/en/course/115683/Methods-4-Bayesian-Computational-Modeling

“Ordinary examination and re-examination:

The exam consists of a portfolio containing a number of assignments. The total length of the portfolio is: 3-7 assignments.

Their form and length will be announced on Brightspace by the teacher at the start of the semester. The portfolio may include products. Depending on their length, and subject to the teacher’s approval, these products can replace some of the standard pages in the portfolio.

It must be possible to carry out an individual assessment. So if some parts of the portfolio have been produced by a group, it must be stated clearly which parts each student is responsible for, and which parts the group as a whole is responsible for.

The complete portfolio must be submitted for assessment in the Digital Exam system. Each student submits a portfolio.”

Exam

- Portfolio consisting of 3 assignments
- Each assignment will require you to create an R Markdown notebook consisting of a mix of text and code.
- Due
  1. End of week 10 (Sunday 12 March, 23:59)
  2. End of week 15 (Sunday 16 April, 23:59)
  3. End of week 18 (Sunday 7 May, 23:59)

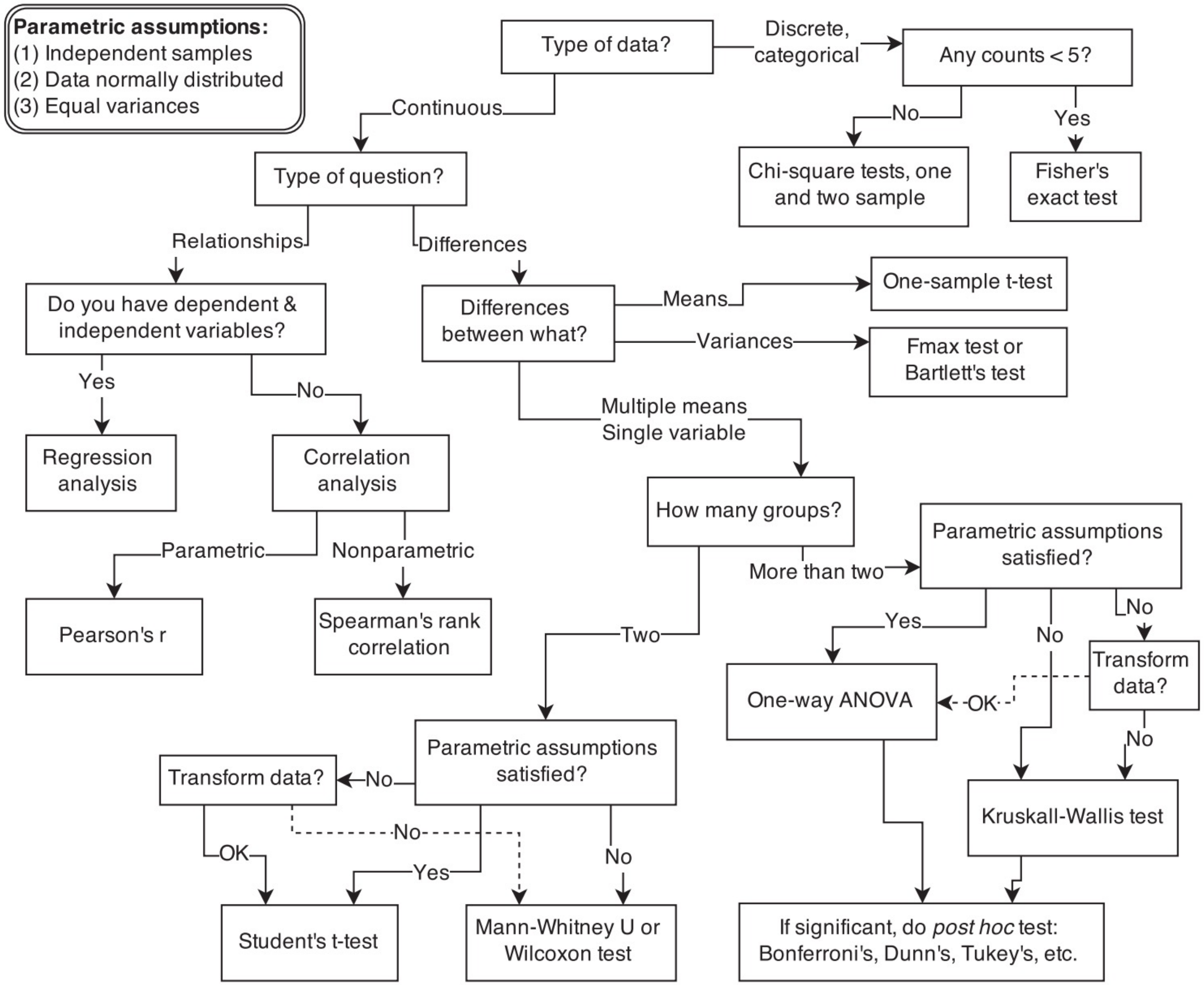
You will receive a (short) feedback message from us on your portfolio assignments that you can use for improvements before finalizing your hand-ins.

How do science and statistics relate to each other?

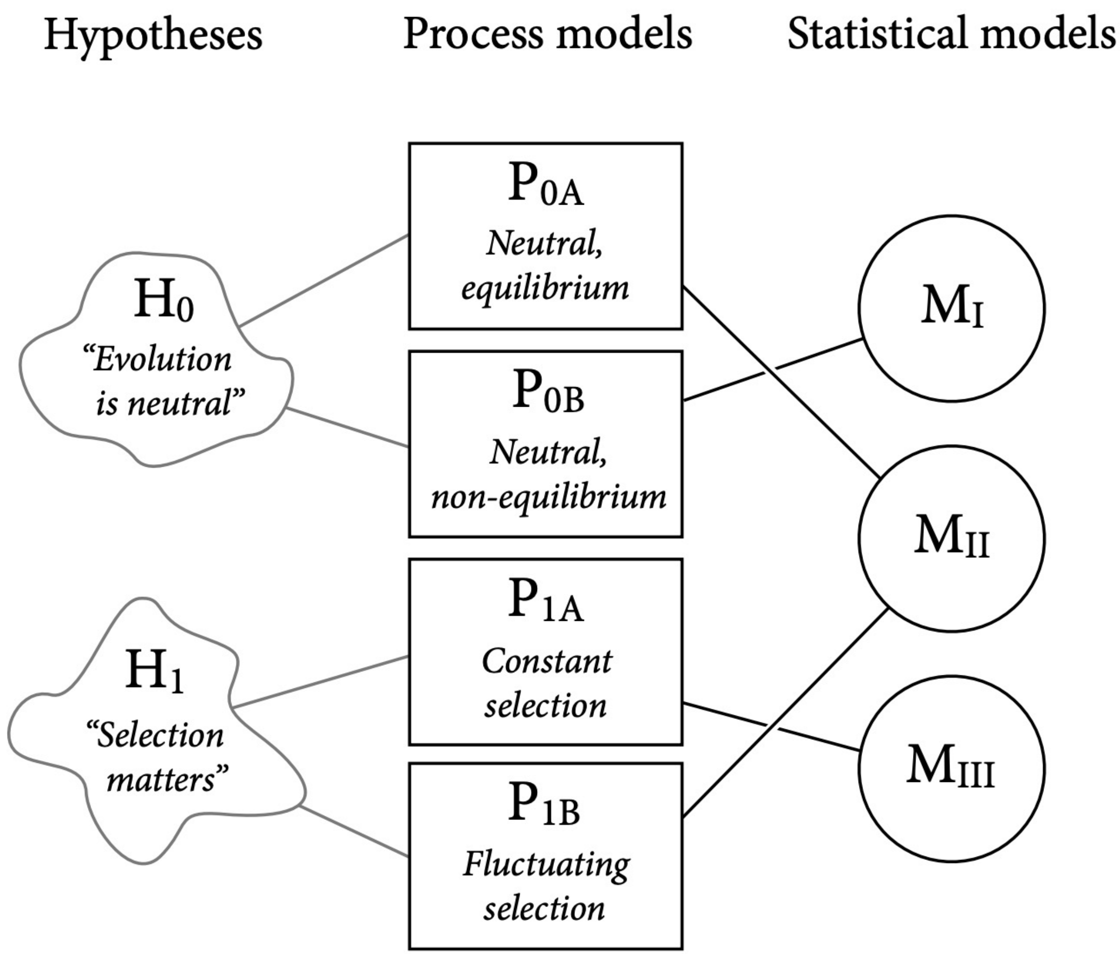
- Your take on this?
- What's the difference between a **statistical** model and a **mechanistic**/process model?
- Example: nutrition science based on
  - Statistical analysis of the outcomes associated with observed/manipulated eating habits (what could go wrong?)
  - Interventions based on knowledge of human physiology
- Karl Popper: science advances by **falsification**.
- Falsification of what?
- Not falsification of null hypotheses. What can go wrong when we try to falsify **null hypotheses**?
- Further: what's the additional problem with **point** null hypotheses?
- Answer: they **can't** be true, so you can always falsify them. You just have to collect enough data.

Statistical analysis by flowchart

Why not?



Hypotheses and models



Bayesian Inference

A surprising piece of information

BBC Sign in News Sport Weather iPlayer TV Ra

NEWS MAGAZINE

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/E

Video & Audio | Magazine | Editors' Blog | In Pictures | Also in the News | Have Your Say | Special Repor

19 November 2012 Last updated at 18:19

44K Share

Does chocolate make you clever?

By Charlotte Pritchard
BBC News

Eating more chocolate improves a nation's chances of producing Nobel Prize winners - or at least that's what a recent study appears to suggest. But how much chocolate do Nobel laureates eat, and how could any such link be explained?

A surprising piece of information

Messerli, F. H. (2012). Chocolate Consumption, Cognitive Function, and Nobel Laureates. *New England Journal of Medicine, 367*(16), 1562–1564.

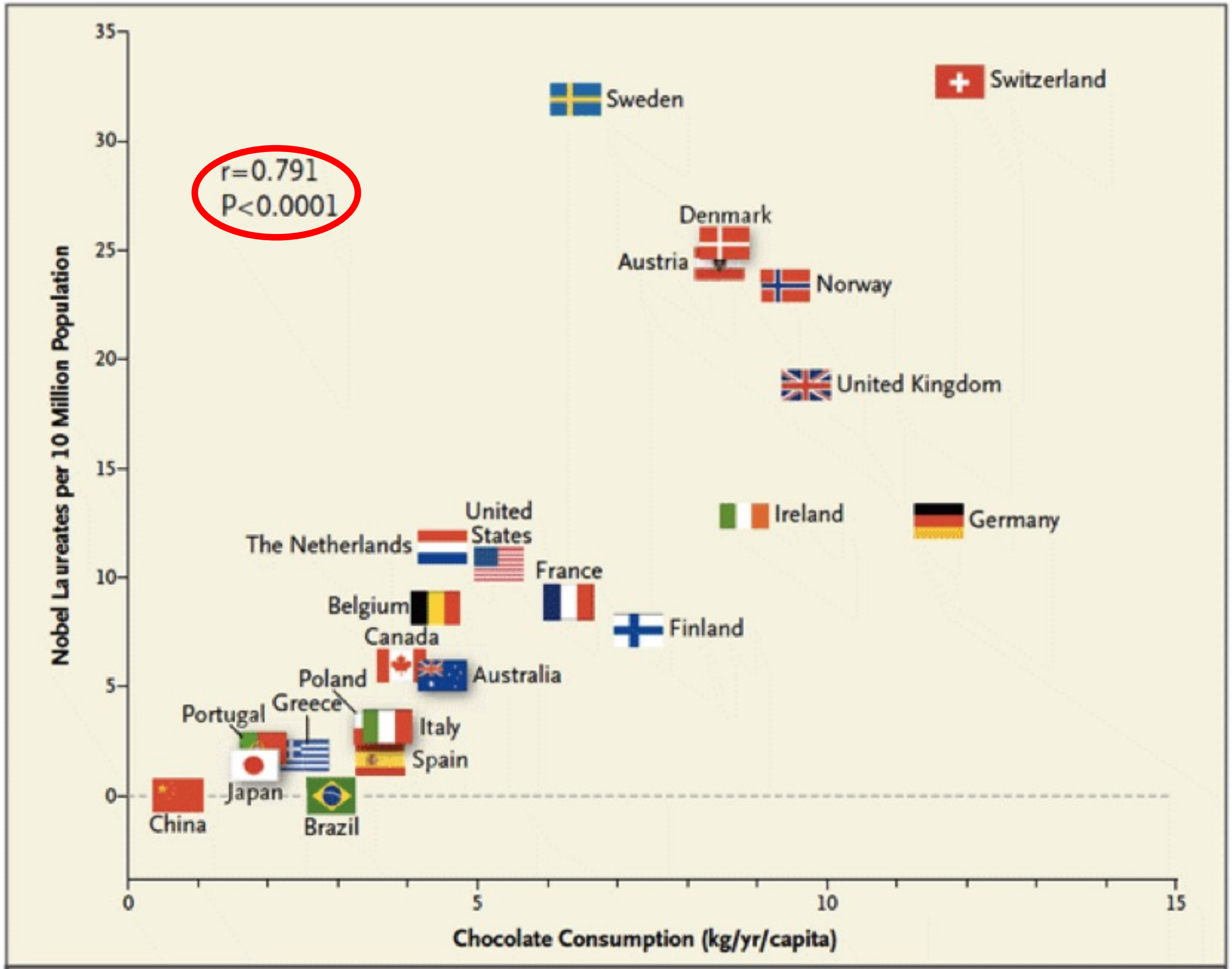


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

NEW ENGLAND JOURNAL OF MEDICINE

So will I win the Nobel prize if I eat lots of chocolate?

This is a question referring to **uncertain quantities**. Like almost all scientific questions, it cannot be answered by deductive logic. *Nonetheless, quantitative answers can be given – **but they can only be given in terms of probabilities***.

Our question here can be rephrased in terms of a conditional probability:

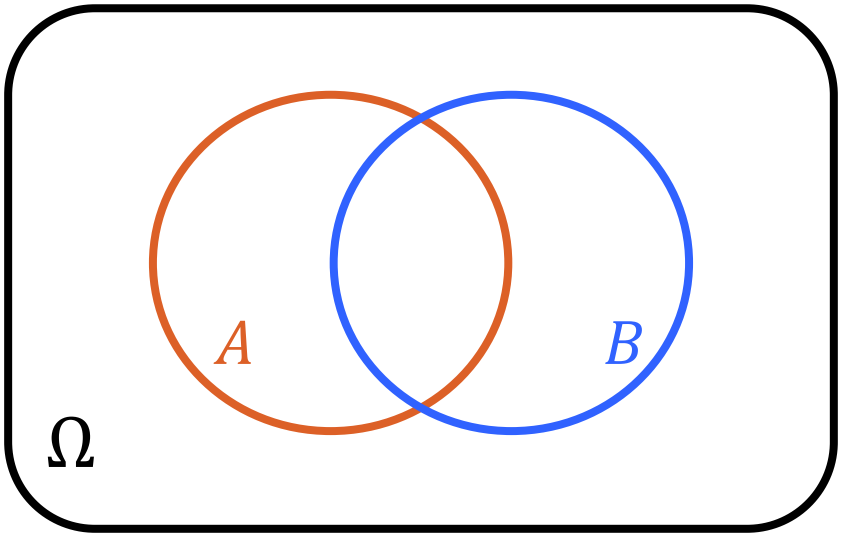
$$p(Nobel \mid amount\ of\ chocolate) = ?$$

To answer it, we have to learn to calculate such quantities. The tool for this is **Bayesian inference**.

However: *note that no amount of statistical analysis will tell you anything about the causal mechanism behind this if you don't have **a hypothesis about that mechanism and a causal scientific model of it**!*

Calculating with probabilities: the setup

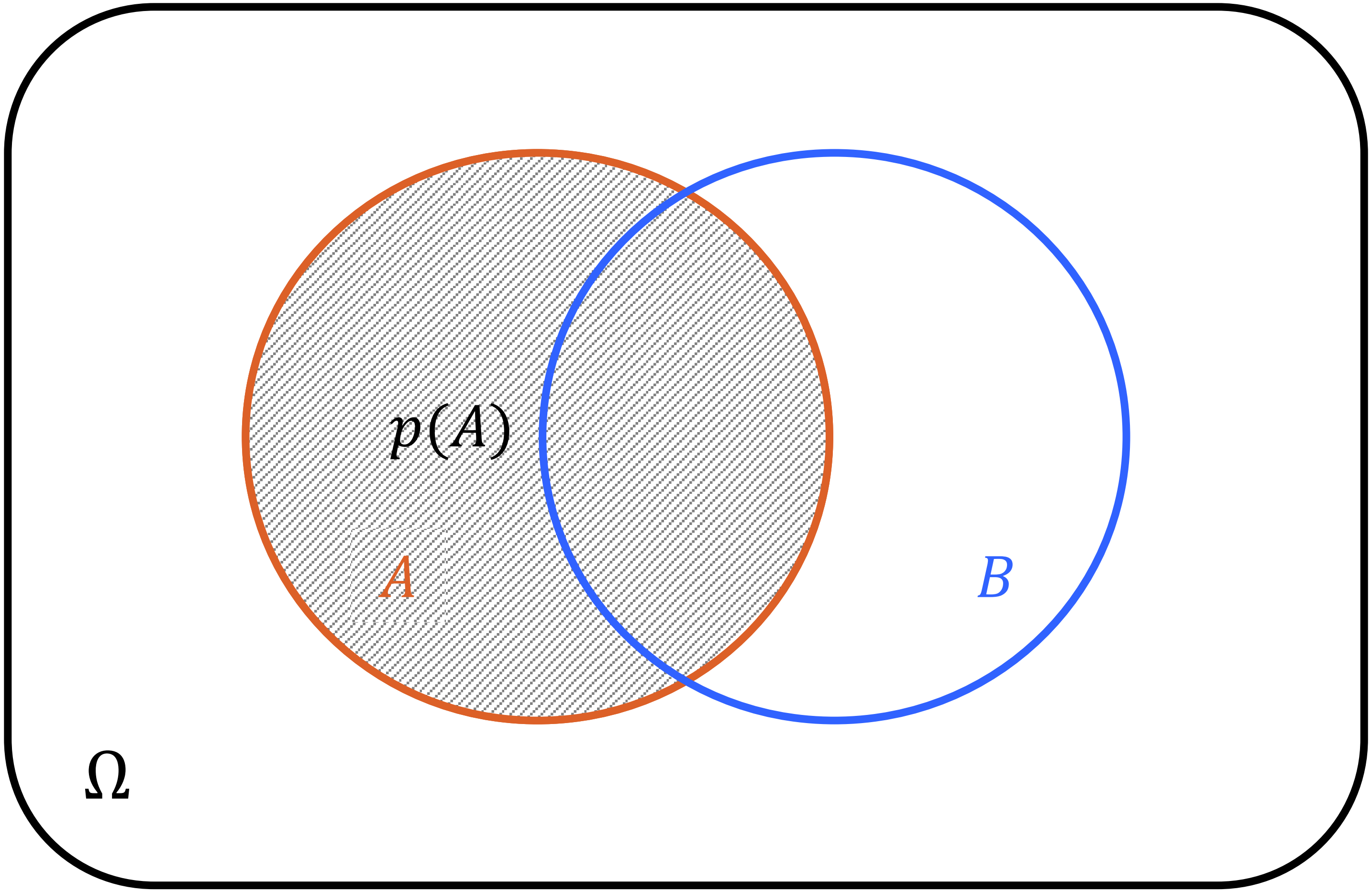
We assume a probability space Ω with subsets A and B



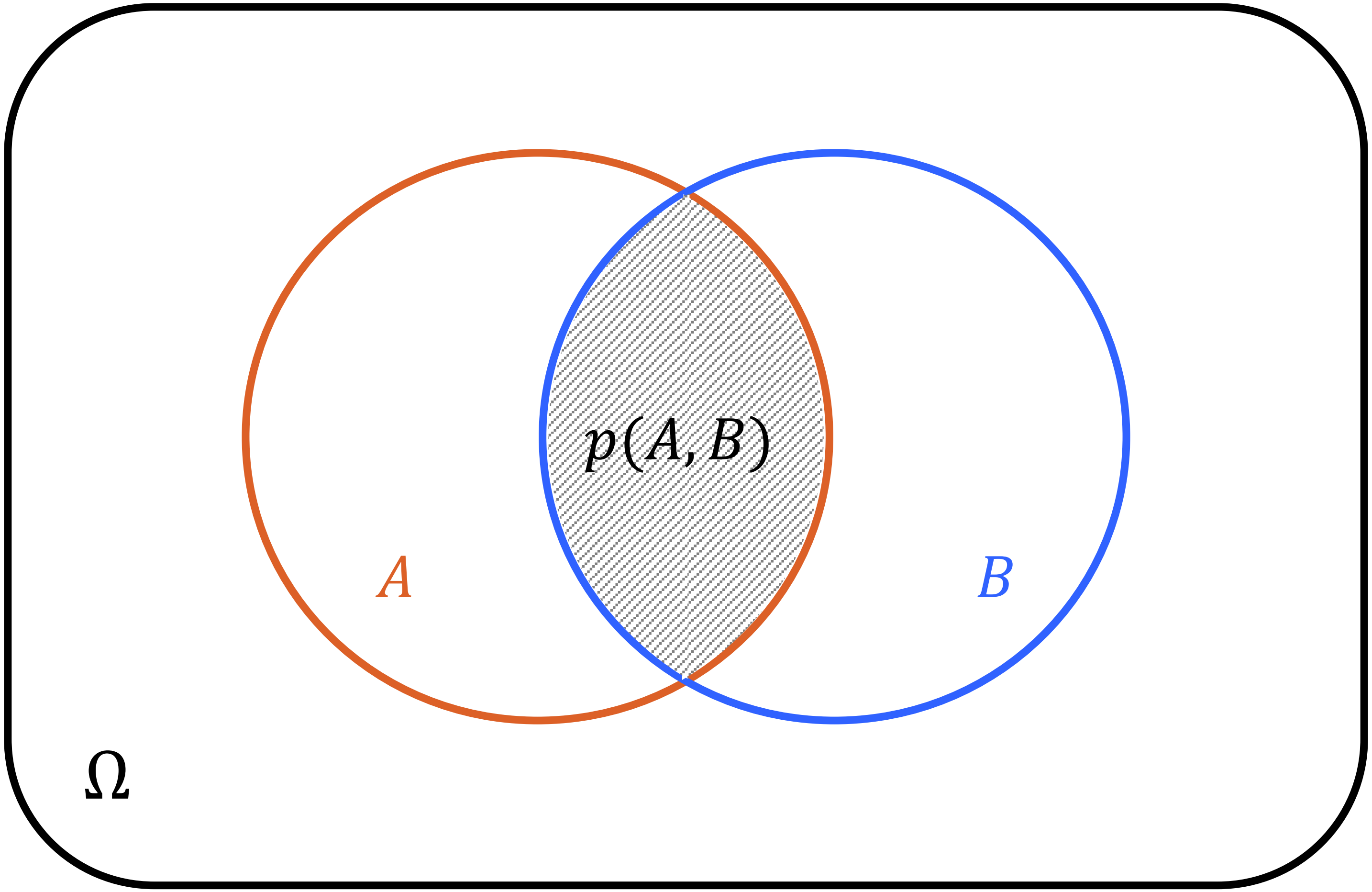
In order to understand *the rules of probability*, we need to understand **three kinds of probabilities**

- *Marginal* probabilities like $p(A)$
- *Joint* probabilities like $p(A, B)$
- *Conditional* probabilities like $p(B|A)$

Marginal probabilities



Joint probabilities



What is 'marginal' about marginal probabilities?

- Let A be the statement 'the sun is shining'
- Let B be the statement 'it is raining'
- $\bar{A}$ negates A, $\bar{B}$ negates B

Consider the following table of joint probabilities:

	B	$\bar{B}$	Marginal probabilities
A	$p(A,B) = 0.1$	$p(A,\bar{B}) = 0.5$	$p(A) = 0.6$
$\bar{A}$	$p(\bar{A},B) = 0.2$	$p(\bar{A},\bar{B}) = 0.2$	$p(\bar{A}) = 0.4$
Marginal probabilities	$p(B) = 0.3$	$p(\bar{B}) = 0.7$	Sum of all probabilities $\sum p(\cdot,\cdot) = 1$

Marginal probabilities get their name from being at the margins of tables such as this one.

Conditional probabilities

- In the previous example, what is the probability that the sun is shining given that it is not raining?
- This question refers to a conditional probability: $p(A|\bar{B})$
- You can find the answer by asking yourself: out of all times where it is not raining, which proportion of times will the sun be shining?

	B	$\bar{B}$	Marginal probabilities
A	$p(A,B) = 0.1$	$p(A,\bar{B}) = 0.5$	$p(A) = 0.6$
$\bar{A}$	$p(\bar{A},B) = 0.2$	$p(\bar{A},\bar{B}) = 0.2$	$p(\bar{A}\,) = 0.4$
Marginal probabilities	$p(B) = 0.3$	$p(\bar{B}) = 0.7$	Sum of all probabilities $\sum p(\cdot,\cdot) = 1$

- This means we have to divide the joint probability of 'sun shining, not raining' by the sum of all joint probabilities where it is not raining:

$$p(A|\bar{B}) = \frac{p(A,\bar{B})}{p(A,\bar{B}) + p(\bar{A},\bar{B})} = \frac{p(A,\bar{B})}{p(\bar{B})} = \frac{0.5}{0.7} \approx 0.71$$

The rules of probability

Considerations like the ones above led to the following definition of the **rules of probability:**

1. $\sum_a p(a) = 1$ (*Normalization*)

2. $p(B) = \sum_a p(a, B)$ (*Marginalization* – the **sum rule**)

3. $p(A, B) = p(A|B)p(B) = p(B|A)p(A)$ (*Conditioning* – the **product rule**)

These are **axioms**, ie they are assumed to be true. Therefore, we cannot test them the way we could test a theory. However, we can see if they turn out to be useful.

The rules of probability

R. T. Cox showed in 1946 that the rules of probability theory can be derived from *three basic desiderata*:

1. Representation of degrees of plausibility by real numbers
2. Qualitative correspondence with common sense (in a well-defined sense)
3. Consistency

By mathematical proof (i.e., by *deductive* reasoning) the three desiderata as set out by Cox imply the rules of probability (i.e., the rules of *inductive* reasoning).

This means that anyone who accepts the desiderata must accept the rules of probability.

«Probability theory is nothing but common sense reduced to calculation.»

— Pierre-Simon Laplace, 1819

Bayes' rule

- The product rule of probability states that

$$p(A|B)p(B) = p(B|A)p(A)$$

- If we divide by $p(B)$, we get **Bayes' rule:**

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)p(A)}{\sum_a p(B|a)p(a)}$$

- The last equality comes from unpacking $p(B)$ according to the product and sum rules:

$$p(B) = \sum_a p(B,a) = \sum_a p(B|a)p(a)$$

Bayes' rule: what problem does it solve?

- Why is Bayes' rule important?
- It allows us to invert conditional probabilities, ie to pass from $p(B|A)$ to $p(A|B)$:

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

- In other words, it allows us to update our belief about A in light of observation B

Bayes' rule: the chocolate example

In our example, it is immediately clear that $P(Nobel|chocolate)$ is very different from $P(chocolate|Nobel)$. While the first is hopeless to determine directly, the second is much easier to find out: ask Nobel laureates how much chocolate they eat. Once we know that, we can use Bayes' rule:

$$\underbrace{p(Nobel|chocolate)}_{\text{posterior}} = \frac{\overbrace{\overbrace{p(chocolate|Nobel)}^{\text{likelihood}}\,\overbrace{P(Nobel)}^{\text{prior}}}^{\text{model}}}{\underbrace{p(chocolate)}_{\text{evidence}}}$$

However: *note that no amount of statistical analysis will tell you anything about the causal mechanism behind this if you don't have* ***a hypothesis about that mechanism and a causal scientific model of it****!*