PSY4320 - Introduction to Bayesian statistics

Nikolai Czajkowski

Introduction

Question 1

When flipping a fair coin, we say that “the probability of flipping Heads is 0.5.” How do you interpret this probability?

1. If I flip this coin over and over, roughly 50% will be Heads.
2. Heads and Tails are equally plausible.
3. Both a and b make sense.

(Bayes rules!, 2022)

Question 2

- Zuofu claims that he can predict the outcome of a coin flip. To test his claim, you flip a fair coin 10 times and he correctly predicts all 10.
- Kavya claims that she can distinguish natural and artificial sweeteners. To test her claim, you give her 10 sweetener samples and she correctly identifies each.

In light of these experiments, what do you conclude?

1. You're more confident in Kavya's claim than Zuofu's claim.
2. The evidence supporting Zuofu's claim is just as strong as the evidence supporting Kavya's claim.

(Bayes rules!, 2022)

Question 3

Suppose that during a recent doctor's visit, you tested positive for a very rare disease. If you only get to ask the doctor one question, which would it be?

What's the chance that I actually have the disease? If in fact I don't have the disease, what's the chance that I would've gotten this positive test result?

(Bayes rules!, 2022)

- A different perspective on statistical inference.

Bayesian statistics is at the core of a longstanding dispute within the statistic community regarding the nature of probability.

A frequentist hypothesis test seeks to answer: If in fact the hypothesis is incorrect, what's the chance I'd have observed this, or even more extreme, data?

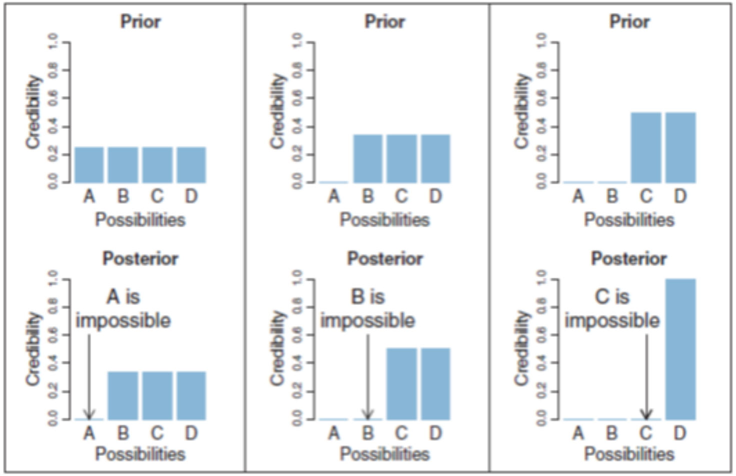
A Bayesian hypothesis test seeks to answer: In light of the observed data, what's the chance that the hypothesis is correct?

- *Bayesian* refers to English statistician, philosopher and Presbyterian minister, Thomas Bayes (1702–1761)
    - Formulated a theorem that is central to "Bayesian statistics"
- Bayesian statistics lost favor in the 1940's.
    - Ronald Fisher; *"the theory of inverse probability is founded upon an error, and must be wholly rejected"*
- Dramatic increase in interest after 1980's due to better algorithms (MCMC) and powerful computers.
- Currently there is a steep increase in papers citing bayesian methods.

B. Relative growth of Psychology papers mentioning "Bayesian"

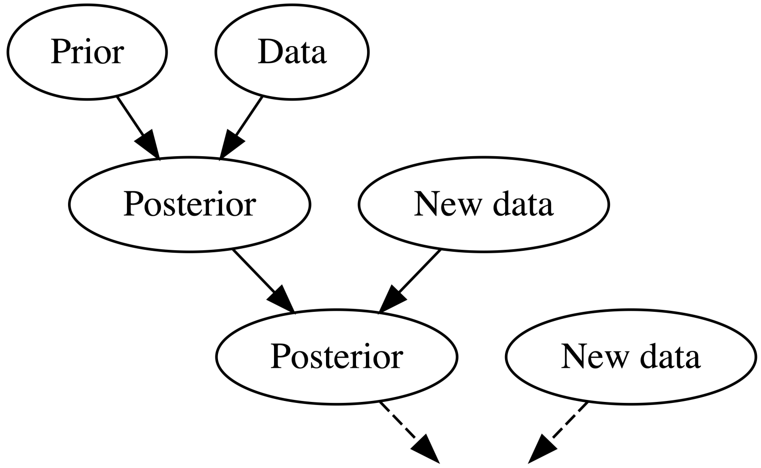
https://psycnet.apa.org/fulltext/2017-24635-002.html

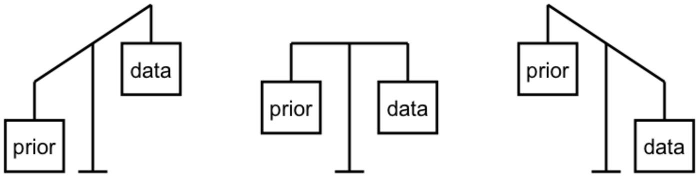
- Bayesian statistics is an approach based on the Bayesian interpretation of probability where probability expresses a degree of belief in an event.
- Bayesian methods use data to *reallocate credibility* to outcomes.
- Bayes rule will give us the tools we need to perform this reallocation.



(Kruschke, 2014)

Prior, data, posterior





Benefits of a bayesian approach

1. A unified approach to inference
2. Avoid null-hypothesis significance testing
3. Naturally incorporate known information into the model
4. Bayesian Statistics correspond to intuition
5. Richer results than p-values
6. Flexibility
7. Results are valid for any sample size

Why Bayesian statistics in PSY4320?

- **Typically only classical (frequentistic) statistical methods are taught to psychology students**
  - Rely on assumptions that make the maths tractable.
  - Involve concepts (sampling distributions, p-values) that even experienced researchers misinterpret.
- **Neuroscience is an early adopter of novel methods**
  - You will see these methods in research papers.
- **Computational and machine learning approaches increasingly important**
  - Solve problems by simultaion rather than exact analysis.

Seminar 1: Probability

- Probability, probability rules, Bayes theorem.
- *Use Bayes theorem to solve simple probability problems*

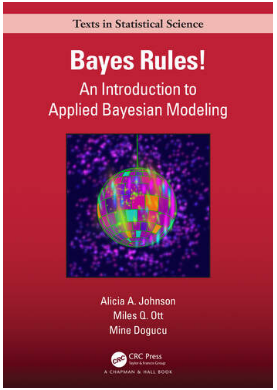
Seminar 2: How to get and use the posterior

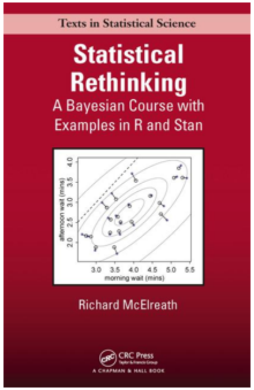
- Central concepts in bayesian analysis
    - Likelihood, Prior and posterior distributions, credibility intervals, bayes factor.
- *Estimate parameter in distribution by grid approximation*

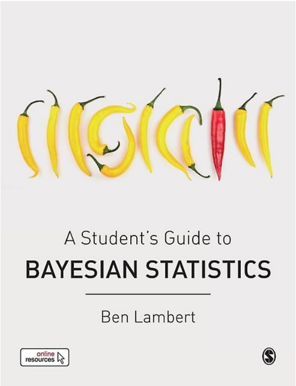
Seminar 3: Bayesian estimation using MCMC

- Markov-Chain Monte Carlo (MCMC) methods
    - interpretation and diagnostics
- Stan - Bayesian analysis in R
- *Run and interpret a bayesian linear regression model*

Do you want to learn bayesian statistics?







- Bayes Rules! is a new book freely avaliable here:*
  - https://www.bayesrulesbook.com
- *Statistical rethinking* by Richard McElreath is extremely pedagogical, and you can see video lectures for a full bayesian course
  - http://xcelab.net/rm/statistical-rethinking/
- Lambert has a book with detailed solutions, and complete set of videos
  - https://study.sagepub.com/lambert

Bayes Rules! Chapter 2

In mathematics, a **set** is a well-defined collection of distincit elements or members.

$$A = \{1, 2, 3\} \quad B = \{3, 4, 5\}$$

The *union* of A and B, denoted by $A \cup B$ is the set of all things that are members of either A or B.

$$A \cup B = \{1, 2, 3, 4, 5\}$$

The *intersection* of A and B, denoted by $A \cap B$ is the set of all things that are members of both A and B. If $A \cap B = \emptyset$, then A and B are said to be disjoint.

$$A \cap B = \{3\}$$

The complement of an event A, denoted by A', is the set of all outcomes in S that are not contained in A. If S contain all the integers from 1 to 10, then

$$A^C = \{6, 7, 8, 9, 10\}$$

Sets

We are going to concider a single dice throw.

- A is the event of getting an even number.
- B is the event of getting a number <=3.

What is the set that corresponds to:

1. B^C
2. $A \cap B$
3. $A^C \cup B$

A phenomenon is **random** if there is the apparent or actual lack of pattern or predictability in events.

Individual random events are, by definition, unpredictable, but if the probability distribution is known, the frequency of different outcomes over repeated events (or "trials") is predictable.

The set of all possible outcomes of a random phenomenon is called the **sample space (S)**.

- What is the sample space for the outcomes of a coin toss?
    - What is the sample space for the outcomes of throwing a dice?
    - What is the sample space for the outcomes of throwing three coins?

An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space.

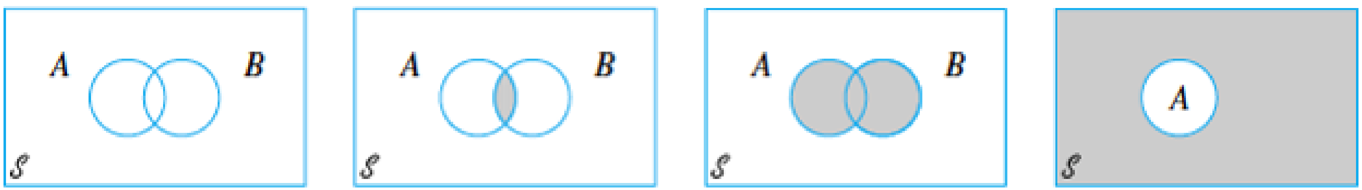
$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Which subset of S corresponds to:

- the event A that all the results are tails.
- the event B that you get exactly two heads on three throws.

A **Venn diagram** is a diagram that shows all possible logical relations between a finite collection of different sets.

Usually it is intended as an informal representation to aide reasoning, and exact sizes/shapes of regions are not important.



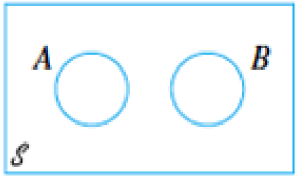
(a) Venn diagram of events A and B (b) Shaded region is $A \cap B$ (c) Shaded region is $A \cup B$ (d) Shaded region is A'

(Devore, 2012)

Sketch the venn diagrams

- Throwing a dice and getting one of the outcomes {1,2,3,4,5,6}
- The three events; i) It is sunny, ii) It is raining, iii) you are carrying an umbrella.

When A and B have no outcomes in common, they are said to be **disjoint** or **mutually exclusive events**. Mathematicians write this compactly as $A \cap B = \emptyset$, where $\emptyset$ denotes the event consisting of no outcomes whatsoever



(e) Mutually exclusive events

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Are these events mutually exclusive?

- A: Getting only heads, B: Getting only tails
- A: Getting exactly two heads, B: Getting exactly two tails
- A: Getting at least one head, B: Getting at least one tail.

If a random phenomenon has outcomes that are equally likely, then the **probability** of an event A is:

$$P(A) = \frac{Number\ of\ outcomes\ in\ A}{Number\ of\ outcomes\ in\ S}$$

What is the probability of throwing three tails in three coins?

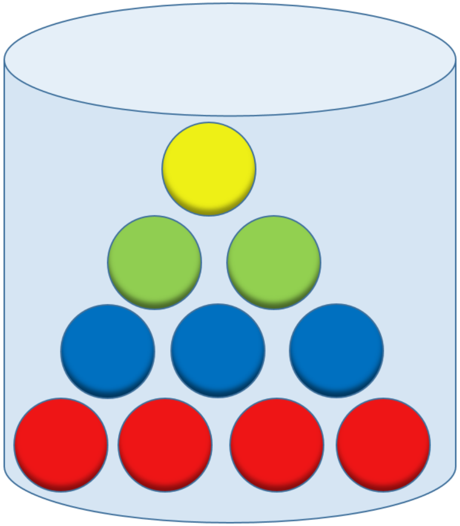
What is the probability of getting exactly two heads on three throws?

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

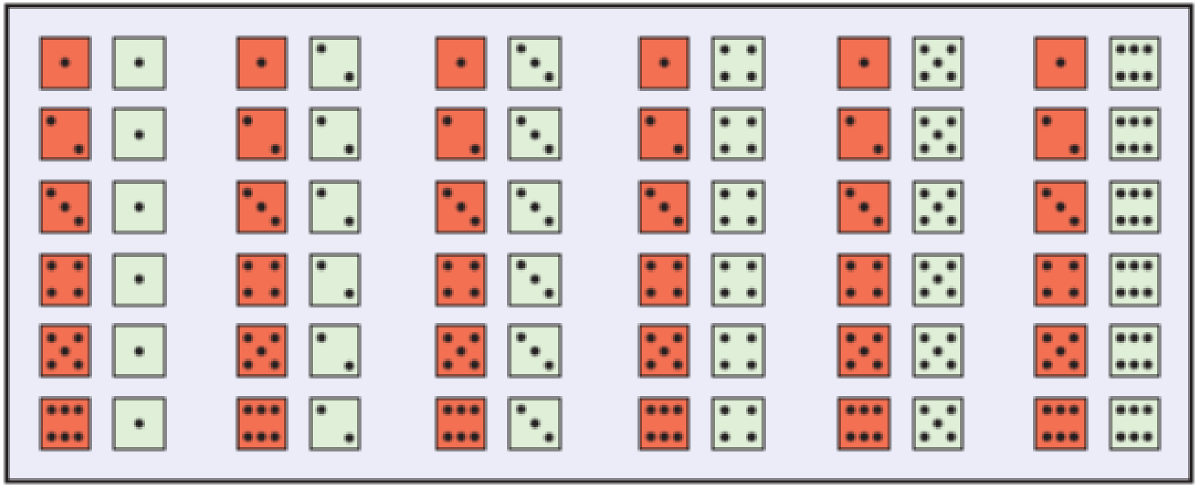
Probability is a numerical description of how likely an event is to occur or how likely it is that a proposition is true.

1. The probability $P(A)$ of any event A satisfies $0 \leq P(A) \leq 1$.
2. If S is the sample space in a probability model, then $P(S) = 1$.
3. P(A does not occur) = $P(A^C) = 1 - P(A)$.

Exercise



A is the event that the ball you pick is green. What is the probability of A?

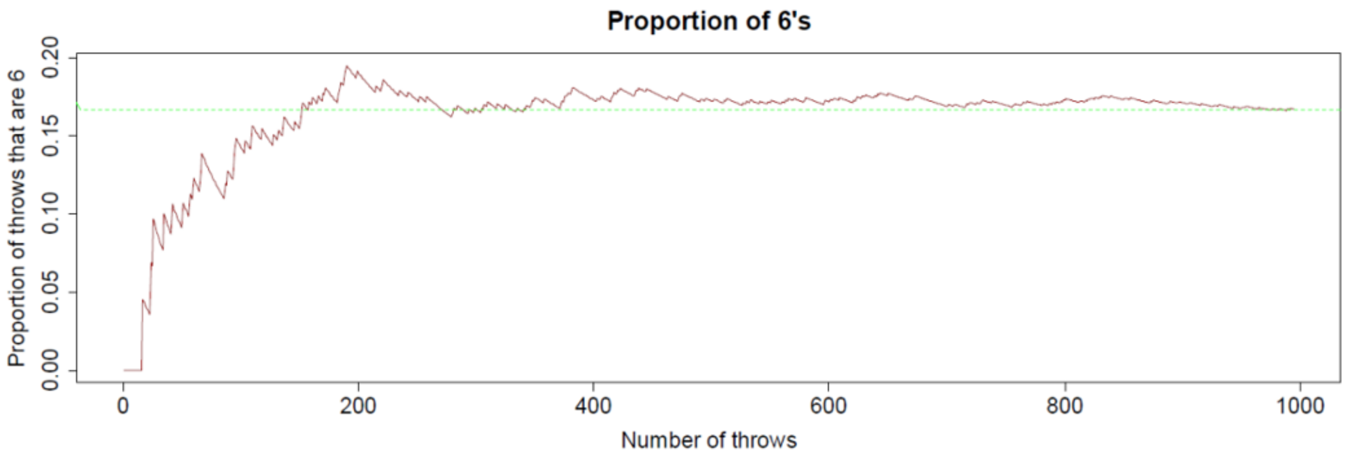


What is the probability that the sum of the eyes is 5 when you roll two dice?

(Moore, 2015)

Imagine that you throw a fair dice.

- What is the probability that you get the value 6?
- How do you know that this is the probability?

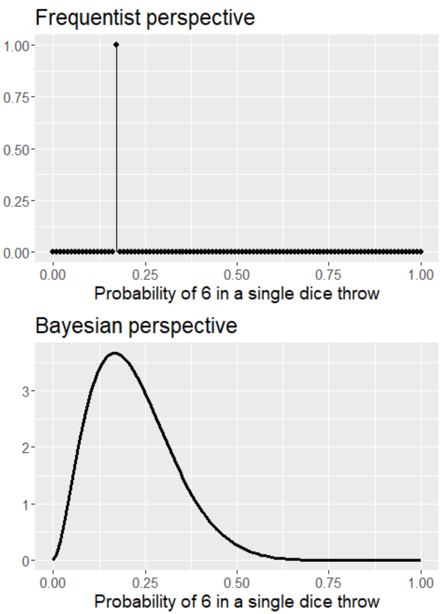


- Frequentist approaches require all probabilities be defined by connection to countable events, and their frequencies in very large samples.
    - premised on actual (or imagined) resampling of data, resulting in sampling distributions of statistics.

- In frequentist statistics, parameters are regarded as *fixed* and cannot have probability distributions, only measurements (statistics) can.
    - Ex. Probability of observing dice value 6 is θ is fixed, only the estimate $\hat{\theta}$ has random (sampling) variance.

Bayesian notion of probability

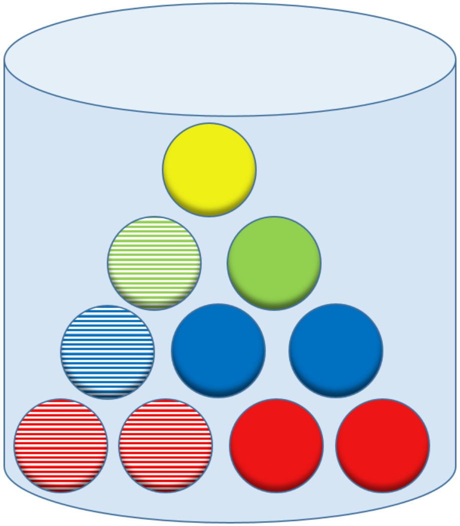
- In the Bayesian philosophy, a probability measures the relative plausibility of an event.
- Randomness is treated as a property of *information*, not of the world.
- Given our lacking information about the exact initial state of the dice toss, the probability of six is 1/6.
- Somtimes dismissively referred to as "subjectivist".



Saturn as Galileo saw it

Joint probability: The probability of *co-occurrence* of two or more events.

Given two events A and B, the **joint probability** is typically denoted as $P(A \cap B)$ or $P(A, B)$.



A is the event that the ball is red. B is the event that the ball is striped What is $P(A \cap B)$?

Table 4.1 Proportions of combinations of hair color and eye color

	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

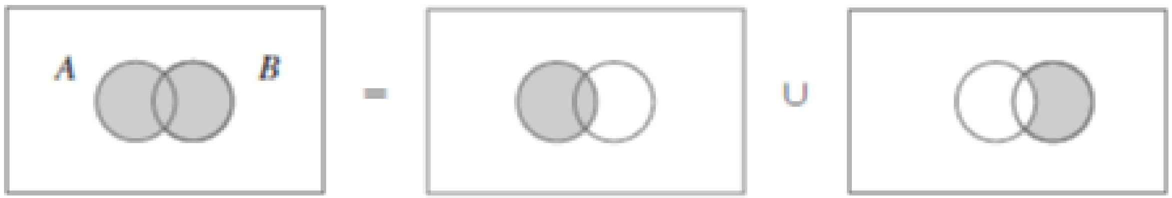
Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974).

What is the probability that a randomly chosen individual has:

a. Red hair?

b. Brown eyes?

c. Brown eyes and blond hair?

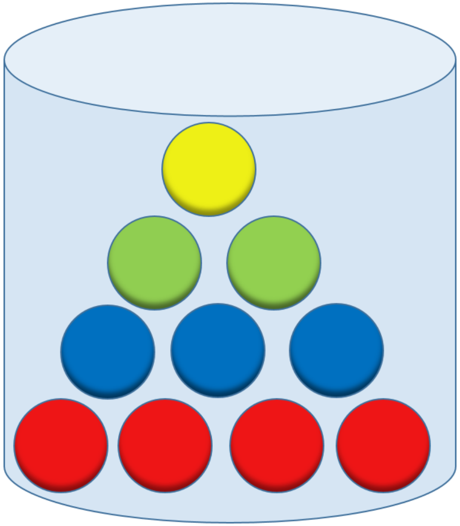


If A and B are two events, then the probability.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

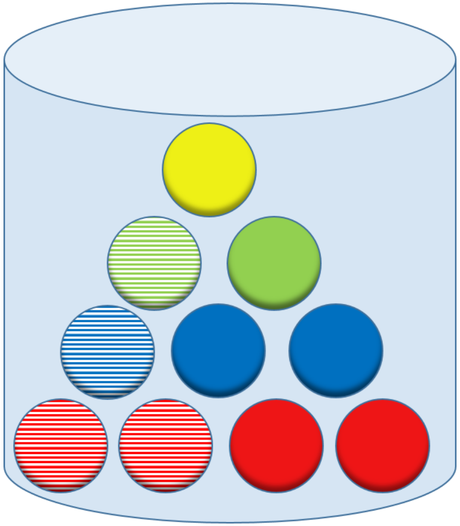
For mutually exclusive events (disjoint sets) this simplifies to

$$P(A \cup B) = P(A) + P(B)$$



A is the event that the ball is red. B is the event that the ball is green. What is $P(A \cup B)$? (Either red or green)

- Make use of a venn diagram.



A is the event that the ball is red. B is the event that the ball is striped What is $P(A \cup B)$?

- Make use of a venn diagram.

Consider randomly selecting a student at a certain university, and let A denote the event that the selected individual has a Visa credit card and B be

the analogous event for a MasterCard. Suppose that $P(A) = .5$, $P(B) = .4$, and $P(A \cap B) = .25$.

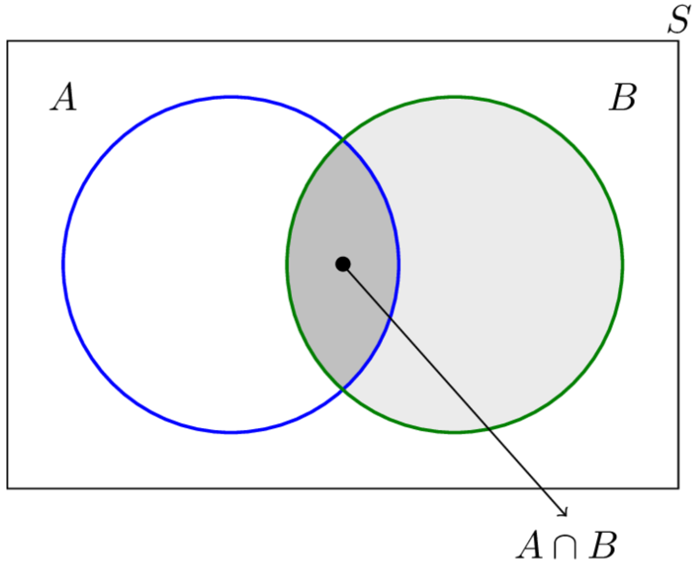
a. Compute the probability that the selected individual has at least one of the two types of cards (i.e., the probability of the event $A \cup B$).

b. What is the probability that the selected individual has neither type of card?

c. Describe, in terms of A and B, the event that the selected student has a Visa card but not a MasterCard, and then calculate the probability of this event.

For any two events A and B with $P(B) \neq 0$, the conditional probability of A *given* that B has occurred is defined by:

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



Which of the following statements corresponds to the expression:
$P(Monday|rain)$?

1. The probability of rain on Monday.
2. The probability of rain, given that it is Monday.
3. The probability that it is Monday, given that it is raining.
4. The probability that it is Monday and that it is raining

Which of the expressions below correspond to the statement: the probability of rain on Monday?

1. $P(rain)$
2. $P(rain|Monday)$
3. $P(Monday|rain)$
4. $P(rain, Monday)$
5. $P(rain, Monday)/P(Monday)$

Exercise

	Blood Group			
Ethnic Group	**O**	**A**	**B**	**AB**
1	.082	.106	.008	.004
2	.135	.141	.018	.006
3	.215	.200	.065	.020

Suppose that an individual is randomly selected from the population, and define events by $A = \{\text{type A selected}\}$, $B = \{\text{type B selected}\}$, and $C = \{\text{ethnic group 3 selected}\}$.

a. Calculate $P(A)$, $P(C)$, and $P(A \cap C)$.

b. Calculate both $P(A|C)$ and $P(C|A)$, and explain in context what each of these probabilities represents.

c. If the selected individual does not have type B blood, what is the probability that he or she is from ethnic group 1?

Three questions:

1. Is $P(X) < P(X|Y)$? or is $P(X) > P(X|Y)$?
2. Can $P(X|Y)$ be eual to $P(X)$?
3. Is $P(X|Y) = P(Y|X)$?

Can you think of an example or counterexample for each of these questions?

Two questions:

1. Is $P(X) < P(X|Y)$?
    - $P(orchestra|practice) > P(orchestra)$
    - $P(flue|washhands) > P(flu)$

2. Is $P(X|Y) = P(Y|X)$?
    - $P(pregnant|female) \approx \mathbf{0.03?}$
    - $P(female|pregnant) >> \mathbf{0.03}$

Table 4.1 Proportions of combinations of hair color and eye color

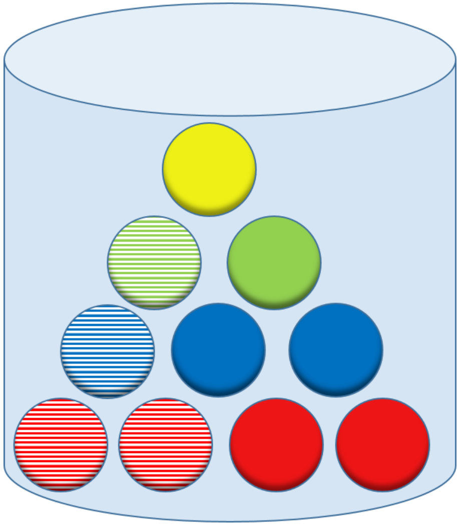
	Hair color				
Eye color	**Black**	**Brunette**	**Red**	**Blond**	**Marginal (eye color)**
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.0

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).

a. What is the probability of having both black hair and brown eyes?
b. What is the probability of having brown eyes given that the hair color is black?
c. What has the highest probability, green eyes given that the hair is brown, or green eyes give that the hair is red?

Multiplicative rule

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



A is the event that the ball is red. B is the event that the ball is striped.Using the multiplicative rule, what is $P(A \cap B)$?

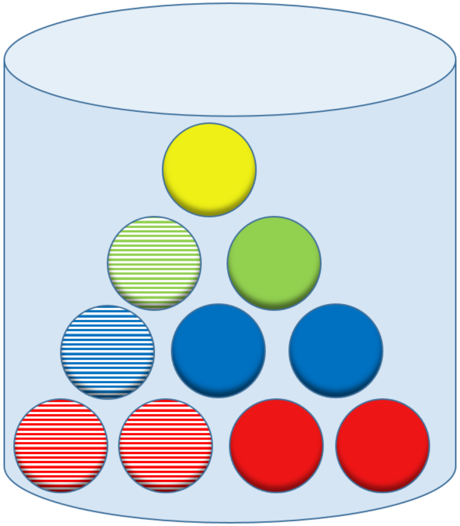
$$P(A \cap B) = P(A|B) \cdot P(B)$$

Two events A and B are independent if knowing that one occurs does not change the probability that the other occurs.

A and B are independent if $P(A|B) = P(A)$, and dependent otherwise.

So, A and B are independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$



A is the event that the ball is red. B is the event that the ball is striped Is A independent of B?

???

No, since the result using the multiplicative rule is: 0.4*0.5=0.02

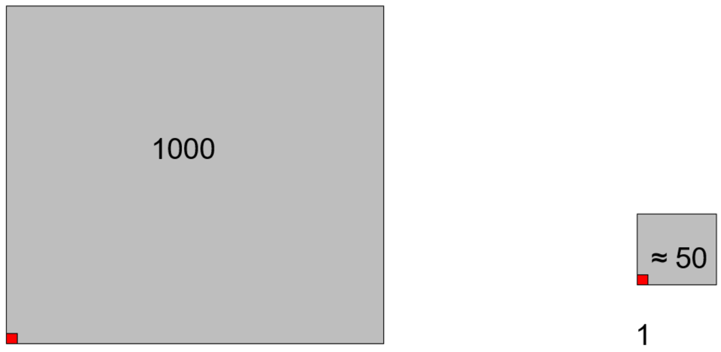
- Assume that 1 out of every 1000 people are vampires.
- A test to detect vampires has a sensitivity of 100% (will always identify someone as vampire, if they are one), and a false positive rate of 5%.
- This can be stated as:

$$P(positive|vampire) = 1.0$$

$$P(positive|not\ vampire) = 0.05$$

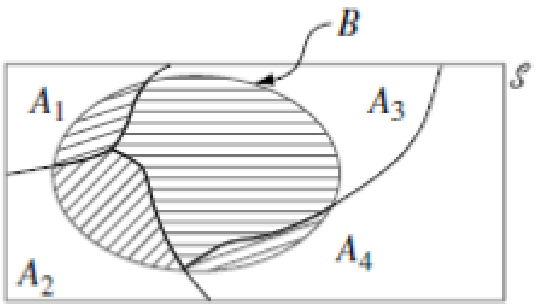
- If a randomly selected individual is found to be positive, what is the probability that he/she really is a vampire? What is:

$$P(vampire|positive) = ?$$



- Testing the 1000 people (left) will give (approximately) 51 positive results, only one of which is really a vampire.
- The probability of being a vampire given a positive test is therefore only 1/51 = 0.0196.

Bayes' theorem is a direct application of conditional probabilities.



Let $A_1, A_2, \ldots, A_k$ be a collection of k mutually exclusive events, with $P(A_i) > 0$ for $i = 1, \ldots, k$. Then for any other event B for wich $P(B) > 0$,

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

$$P(V|+) = \frac{P(V \cap +)}{P(+)}$$

$$
\begin{aligned}
P(V|+) &= \frac{P(V \cap +)}{P(+)} \\
&= \frac{P(+|V)P(V)}{P(+|V)P(V) + P(+|\bar{V})P(\bar{V})} \\
&= \frac{1 \cdot 0.001}{1 \cdot 0.001 + 0.05 \cdot 0.999} \\
&= 0.0196
\end{aligned}
$$

Bayes theorem exercise 1

Assume that the word ‘offer’ occurs in 80% of the spam messages in my account. Also, let’s assume ‘offer’ occurs in 10% of my desired e-mails. If 30% of the received e-mails are considered as a scam, and I will receive a new message which contains ‘offer’, what is the probability that it is spam?

https://towardsdatascience.com/bayes-theorem-101-example-solution-ff54147d6c7f

In a casino in Blackpool there are two slot machines: one that pays out 10 % of the time, and one that pays out 20 % of the time. Obviously, you would like to play on the machine that pays out 20 % of the time but you do not know which of the two machines is the more generous. You thus adopt the following strategy: you assume initially that the two machines are equally likely to be the generous machine. You then select one of the two machines at random and put a coin into it. Given that you loose that first bet estimate the probability that the machine you selected is the more generous of the two machines.

https://towardsdatascience.com/bayes-theorem-101-example-solution-ff54147d6c7f

Bayes rule with parameters and data

Bayes rule tells us how to update our prior beliefs in order to derive better, more informed, beliefs about a situation in light of new data.

$$\underbrace{\mathrm{P}(\text{parameters} \mid \text{data})}_{\text{posterior}} = \frac{\overbrace{\mathrm{P}(\text{data} \mid \text{parameters})}^{\text{likelihood}} \overbrace{\mathrm{P}(\text{parameters})}^{\text{prior}}}{\underbrace{\mathrm{P}(\text{data})}_{evidence}}$$

With data set $\boldsymbol{y}$ and model parameters $\boldsymbol{\theta}$, Bayes' theorem can be written as:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Terms you should be familiar with

- A set
- Random phenomenon
- Event, sample space
- Venn diagram
- Mutually exclusive events
- Probability of an event
    - Frequentistic notion of probability
    - Bayesian notion of probability
- Joint probability
- Additive rule of probability
- Conditional probability
- Multiplicative rule
- Independent events
- Bayes theorem/rule

The R Project for Statistical Computing

- R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation.
  - https://www.r-project.org/

- The R language is widely used platform for data analysis and data science across all sciences.

- Comes with a small core of functions, but can be extended from a vast library of Several thousands *packages*.

https://en.wikipedia.org/wiki/R_(programming_language))

- RStudio is an integrated development environment (IDE) for R, a programming language for statistical computing and graphics.
  - https://rstudio.com/
  - Greatly simplifies work in large projects.
- RStudio, Inc. is a commercial enterprise, and has no formal connection to the R Foundation.

Some R resources

Official (brief) R documentation:

- http://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

If you plan to use R in your work, read this book!:

- https://r4ds.had.co.nz/

R-bloggers is a popular portal for R users:

- https://www.r-bloggers.com/2017/03/the-5-most-effective-ways-to-learn-r/

Statistical Inference via Data Science

- https://moderndive.com/

we'll examine a sample of 150 articles which were posted on Facebook and fact checked by five BuzzFeed journalists (Shu et al. 2017).

```
# Load packages
library(bayesrules)
library(tidyverse)
library(janitor)

# Import article data
data(fake_news)
```

You see the person point to a fizzy cola drink and say “please pass my pop.” Though the country is united in its love of fizzy drinks, it’s divided in what they’re called, with common regional terms including “pop,” “soda,” and “coke.” This data, i.e., the person’s use of “pop,” provides further information about where they might live.

We will look at da data consisting of 374250 responses to a volunteer survey conducted at popvssoda.com