# Foundations of probability theory

Joaquín Rapela

Gatsby Computational Neuroscience Unit
University College London

June 12, 2023

# Contents

# Main reference

I will mainly follow chapters seven *Foundations of probability theory* and eight *Conditional probability and Bayes* from Tijms (2012).

# Contents

# Historical notes

- explain the frequency-based interpretation of probability.
- constructing the mathematical foundations of probability theory has proven to be a long-lasting process of trial an error.
- the approach of defining probability as relative frequencies of repeatable experiments lead to unsatisfactory theory (why?) https://www.jstor.org/stable/pdf/20115155.pdf
- the frequency view of probability has a long history that goes back to Aristotle.
- in 1933 the Russian mathematician Andrej Kolmogrov (1903-1987) laid a satisfactory mathematical foundation of probability theory.

He created a set of axioms. Axioms state a number of minimal requirements that the probability objects should satisfy. From these few algorithms all claims of probability can be derived, as we will see.

# Probability model

**Definition 1 (Probability model)**

*A* **probability model** *is a matematical representation of a random experiment.*

# Probability model

**Definition 1 (Probability model)**

*A* **probability model** *is a matematical representation of a random experiment. It consists of a description of all possible outcomes of the experiment (i.e.,* **sample space***), a set of subsets of the sample space (i.e.,* **events***), and an assigment of probability to events (i.e.,* **probability measure***).*

# Sample space

**Definition 2 (Sample space)**

*The set of all samples in an experiment is called the* **sample space**. *It is denoted by* $\Omega$.

# Event



**Definition 3 (Event)**

*An* **event** *is a subset of the sample space. We denote the collection of all events by* $\mathcal{F}$.

# Event

**Definition 3 (Event)**

*An* **event** *is a subset of the sample space. We denote the collection of all events by $\mathcal{F}$.*

Notes:

1. We will only assign probabilities to events (i.e., to sets $A \in \mathcal{F}$).
2. For finite or countable sample spaces, we can assign probabilities to any subset of the sample space. Thus, any subset of a finite or countable sample space can be an event.
3. For uncountable sample spaces, we can only assign probabilities to well behaved subsets of the sample space (i.e., to elements in a $\sigma$ algebra of subsets of the sample space). Only well-behaved subsets of an uncountable sample space can be events.

# Probability measure



**Definition 4 (Probability measure)**

*A* **probability measure** *is a function that assigns numbers between zero and one to events (i.e.,* $P : \mathcal{F} \to [0, 1]$*).*

# Probability model: definitions

## Definition (Sample space)

*The set of all samples in an experiment is called the* **sample space**. *It is denoted by* $\Omega$.

## Definition (Event)

*An* **event** *is a subset of the sample space. We denote the collection of all events by* $\mathcal{F}$.

## Definition (Probability measure)

*A* **probability measure** *is a function that assigns numbers between zero and one to events (i.e.,* $P : \mathcal{F} \to [0, 1]$*).*

## Definition (Probability model)

*A* **probability model**, $\mathcal{M}$, *is a mathemtatical representation of a random experiment consiting of a sample space,* $\Omega$, *a set of events,* $\mathcal{F}$, *and a probability measure,* $P$ *(i.e.,* $\mathcal{M} = \{\Omega, \mathcal{F}, P\}$*).*

# Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once.

# Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H,\ T]$.

# Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, T]$. We assign a probability of 0.5 to each element of the sample space.

# Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair die and count the number of rolls until the first six shows up.

# Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair die and count the number of rolls until the first six shows up. The sample space is the set of positive integers.

# Examples of probability models



For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, \ T]$. We assign a probability of 0.5 to each element of the sample space.

2. The experiment is to repeately roll a fair die and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \ldots$ can be assigned to the outcomes $1, 2, 3, \ldots$.

# Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, \ T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair die and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \ldots$ can be assigned to the outcomes $1, 2, 3, \ldots$.
3. The experiment is to measure the time between the first and second spikes in an experimental trial.

# Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H,\ T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair die and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \ldots$ can be assigned to the outcomes $1, 2, 3, \ldots$.
3. The experiment is to measure the time between the first and second spikes in an experimental trial.

# Examples of probability models



For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H,\ T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair die and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \ldots$ can be assigned to the outcomes $1, 2, 3, \ldots$.
3. The experiment is to measure the time between the first and second spikes in an experimental trial. The sample space is the set $(0, \infty)$ of positive real numbers. We can assign a probability of $1 - \exp(-\lambda t)$ to the event that the second spike is fired less than $t$ seconds after the first spike.

# Axioms of probability theory

**Axiom 1**

$P(A) \geq 0, \quad \forall A \in \mathcal{F}$

**Axiom 2**

$P(\Omega) = 1$

**Axiom 3**

$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ *for every collection of pairwise disjoint events* $A_1, A_2, \ldots$

# Experiment with equally likely outcomes



An experiment with equally likely outcomes is one with a finite number of outcomes $\omega_1, \ldots, \omega_N$, where all outcomes have the same probability (i.e., $P(\omega_i) = \frac{1}{N}$).

# Experiment with equally likely outcomes



An experiment with equally likely outcomes is one with a finite number of outcomes $\omega_1, \ldots, \omega_N$, where all outcomes have the same probability (i.e., $P(\omega_i) = \frac{1}{N}$).

**Claim 1**

*For any event $A$, $P(A) = \frac{N(A)}{N}$, where $N(A)$ is the number of outcomes in the set $A$.*

# Example: equally likely outcomes

**Example 1**

*John, Pedro and Rosita each roll on fair die. How do we calculate the probability that the score of Rosita is equal to the sum of the scores of John and Pedro?*

# Some basic rules



**Rule 1**

*For any finite number of mutually exclusive events $A_1, \ldots, A_N$,*

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_N)$$

**Rule 2**

*For any event $A$,*

$$P(A) = 1 - P(A^c)$$

*where the event $A^c$ consists of all outcomes that are not in $A$.*

**Rule 3**

*Let $A, B$ be two events such that $A \subseteq B$. Then $P(A) \leq P(B)$.*

**Rule 4**

*For any two events $A$ and $B$,*

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

# Rule 1

**Proof.**

Denote by $\emptyset$ the empty set. We will firt prove that $P(\emptyset) = 0$. Take $A_i = \emptyset$ for $i = 1, 2, \ldots$. Then $\emptyset = \bigcup_{i=1}^{\infty} A_i$. Next, by Axiom 3, $P(\emptyset) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\emptyset)$. This implies that $P(\emptyset) = 0$. Define $A_{N+j} = \emptyset$ for $j = 1, 2, \ldots$. Then

$$
\begin{aligned}
P(A_1 \cup A_2 \cup \ldots \cup A_N) &= P(A_1 \cup A_2 \cup \ldots \cup A_N \cup A_{N+1} \cup A_{N+2} \cup \ldots) \\
&= P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \\
&= \sum_{i=1}^{N} P(A_i) + \sum_{j=1}^{\infty} P(A_{N+j}) = \sum_{i=1}^{N} P(A_i)
\end{aligned}
$$

$\square$

Notes:

1. the last equality in the second line holds by Axiom 1
2. the last equality in the third line holds because $P(A_{N+j}) = P(\emptyset) = 0$.

# Proof of rule 2

**Proof.**

$\Omega = A \cup A^c$. $A$ and $A^c$ are disjoint. Then, by Rule 1, $P(\Omega) = P(A) + P(A^c)$. From Axiom 2, $P(\Omega) = 1$. Thus, $1 = P(A) + P(A^c)$.

$\square$

# Exercise



Prove rule 3.

# Proof of rule 4

**Proof.**

$$\begin{aligned} A \cup B &= (A \setminus B) \cup (B \setminus A) \cup (A \cap B) \\ A &= (A \setminus B) \cup (A \cap B) \\ B &= (B \setminus A) \cup (A \cap B) \end{aligned}$$

Since the sets in the right-hand-side of the above equations are pairwise disjoint, by rule 1, we obtain

$$\begin{aligned} P(A \cup B) &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ P(A) &= P(A \setminus B) + P(A \cap B) \to P(A \setminus B) = P(A) - P(A \cap B) \\ P(B) &= P(B \setminus A) + P(A \cap B) \to P(B \setminus A) = P(B) - P(A \cap B) \end{aligned}$$

Replazing the equations on the right of the second and third line in the equation on the first line the rule is proved.

$$\begin{aligned} P(A \cup B) &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

$\square$

# Example 7.7: Chevalier de Mere to Blaise Pascal

The gambler Chevalier de Mere posed the following problem to the famous French mathematician Blaise Pascal in 1654. This problem marks the begining of probability theory.

**Example 7.7**

a. *How many rolls of a fair die are required to have at least a 50% chance of rolling at least one six?*
b. *How many rolls of two fair dice are required to have at least a 50% chance of rolling at least one double six?*

# Analytical solution to example 7.7a



(a)

- Let's fix the number of rolls $r$. The sample space is $\Omega = \{(i_1, \ldots, i_r) : 1 \leq i_k \leq 6\}$, where $i_k$ is the up face of the die on the kth roll. The outcomes in $\Omega$ are equiprobable.

# Analytical solution to example 7.7a

(a)

- Let's fix the number of rolls $r$. The sample space is $\Omega = \{(i_1, \ldots, i_r) : 1 \leq i_k \leq 6\}$, where $i_k$ is the up face of the die on the kth roll. The outcomes in $\Omega$ are equiprobable.
- We want to calculate the probability of the event $A$= "at least one six shows up in the $r$ rolls.". When you see the keyword "at least one" in an event, it is easier to calculate the probability of the complement $A^c$= "no six shows up in the $r$ rolls."

# Analytical solution to example 7.7a

(a)

- Let's fix the number of rolls $r$. The sample space is $\Omega = \{(i_1, \ldots, i_r) : 1 \leq i_k \leq 6\}$, where $i_k$ is the up face of the die on the kth roll. The outcomes in $\Omega$ are equiprobable.
- We want to calculate the probability of the event $A$= "at least one six shows up in the $r$ rolls.". When you see the keyword "at least one" in an event, it is easier to calculate the probability of the complement $A^c$= "no six shows up in the $r$ rolls."
- $P(A^c) = \frac{N(A^c)}{N} = \frac{5^r}{6^r} = \left(\frac{5}{6}\right)^r$.

# Analytical solution to example 7.7a

(a)

- Let's fix the number of rolls $r$. The sample space is $\Omega = \{(i_1, \ldots, i_r) : 1 \leq i_k \leq 6\}$, where $i_k$ is the up face of the die on the kth roll. The outcomes in $\Omega$ are equiprobable.
- We want to calculate the probability of the event $A$= "at least one six shows up in the $r$ rolls.". When you see the keyword "at least one" in an event, it is easier to calculate the probability of the complement $A^c$= "no six shows up in the $r$ rolls."
- $P(A^c) = \frac{N(A^c)}{N} = \frac{5^r}{6^r} = \left(\frac{5}{6}\right)^r$.
- $\frac{1}{2} < P(A) = 1 - P(A^c) = 1 - \left(\frac{5}{6}\right)^r$ iff $\left(\frac{5}{6}\right)^r < \frac{1}{2}$ iff $\log\left(\frac{5}{6}\right)^r < \log\frac{1}{2}$ iff $r\log\frac{5}{6} < \log\frac{1}{2}$ iff $r > \frac{\log\frac{1}{2}}{\log\frac{5}{6}} = 3.8$

# Simulated solution to example 7.7a

Please see code here.

# Exercise

Solve example 7.7b analytically and by simulation. Answer: you need at least 25 draws.

# Example 7.9: soccer teams in quarterfinal



## Example 7.9

*The eight soccer teams which have reached the quarterfinals of the Championship League are formed by two teams from of the the countries England, Germany, Italy and Spain. The four matches to be played in the quarterfinal are determined by drawing lots.*

a. *What is the probability that the two teams from the same country play against each other in each of the four matches?*
b. *What is the probability that there is a match between the two teams from England or between the two teams from Germany?*

# Analytical solution to example 7.9: setup



- Let $T$ be the set of eight teams
  $T = \{E_1, E_2, G_1, G_2, I_1, I_2, S_1, S_2\}$

# Analytical solution to example 7.9: setup

- Let $T$ be the set of eight teams $T = \{E_1, E_2, G_1, G_2, I_1, I_2, S_1, S_2\}$
- A match is a set of two different elements from $T$ (i.e., match $= \{T_i, T_j\}$, match $\subset T$, $T_i \neq Tj$).

# Analytical solution to example 7.9: setup

- Let $T$ be the set of eight teams $T = \{E_1, E_2, G_1, G_2, I_1, I_2, S_1, S_2\}$
- A match is a set of two different elements from $T$ (i.e., match $= \{T_i, T_j\}$, match $\subset T$, $T_i \neq Tj$).
- A quarterfinal is a set of four different matches.

# Analytical solution to example 7.9: setup



- Let $T$ be the set of eight teams $T = \{E_1, E_2, G_1, G_2, I_1, I_2, S_1, S_2\}$
- A match is a set of two different elements from $T$ (i.e., match $= \{T_i, T_j\}, \text{match} \subset T, T_i \neq Tj$).
- A quarterfinal is a set of four different matches.
- $\Omega$ is the set of all possible quarterfinals (e.g., $\{\{E1, G1\}, \{I1, G2\}, \{S2, G2\}, \{I2, G1\}\} \in \Omega$).

# Analytical solution to example 7.9: setup

- Let $T$ be the set of eight teams
  $T = \{E_1, E_2, G_1, G_2, I_1, I_2, S_1, S_2\}$
- A match is a set of two different elements from $T$ (i.e., match $= \{T_i, T_j\}$, match $\subset T$, $T_i \neq Tj$).
- A quarterfinal is a set of four different matches.
- $\Omega$ is the set of all possible quarterfinals (e.g., $\{\{E1, G1\}, \{I1, G2\}, \{S2, G2\}, \{I2, G1\}\} \in \Omega$).
- $\Omega$ contains equally-likely outcomes. Thus, for any event $A$, $P(A) = \frac{N(A)}{N}$.

# Analytical solution to example 7.9: setup

- Let $T$ be the set of eight teams $T = \{E_1, E_2, G_1, G_2, I_1, I_2, S_1, S_2\}$
- A match is a set of two different elements from $T$ (i.e., $\text{match} = \{T_i, T_j\}, \text{match} \subset T, T_i \neq Tj$).
- A quarterfinal is a set of four different matches.
- $\Omega$ is the set of all possible quarterfinals (e.g., $\{\{E1, G1\}, \{I1, G2\}, \{S2, G2\}, \{I2, G1\}\} \in \Omega$).
- $\Omega$ contains equally-likely outcomes. Thus, for any event $A$, $P(A) = \frac{N(A)}{N}$.
- $N = \frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}$. There are $\binom{8}{2}$, $\binom{6}{2}$, $\binom{4}{2}$ and 1 ways of selecting the first, second, third and fourth matches, respectively. We divide by 4! because the order between matches does not matter.

# Analytical solution to example 7.9a



- The event A="*two teams from the same country play against each other in each of the four matches*" contains only one outcome (i.e., $A = \{\{E_1, E_2\}, \{G_1, G_2\}, \{I_1, I_2\}, \{S_1, S_2\}\}$).

# Analytical solution to example 7.9a

- The event A="*two teams from the same country play against each other in each of the four matches*" contains only one outcome (i.e., $A = \{\{E_1, E_2\}, \{G_1, G_2\}, \{I_1, I_2\}, \{S_1, S_2\}\}$).
- $P(A) = \frac{N(A)}{N} = \frac{1}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.009524$

# Analytical solution to example 7.9b



- The event $B$=“*there is a match between the two teams from England or between the two teams from Germany*” is the union of the events $B_E$=“*there is a match between the two teams from England*” and $B_G$=“*there is a match between the two teams from Germany*” (i.e., $B = B_E \cup B_G$).

# Analytical solution to example 7.9b



- The event $B$=“*there is a match between the two teams from England or between the two teams from Germany*” is the union of the events $B_E$=“*there is a match between the two teams from England*” and $B_G$=“*there is a match between the two teams from Germany*” (i.e., $B = B_E \cup B_G$).
- Since $B_E$ and $B_G$ are not disjoint, we should use Rule 3 to compute $P(B_E \cup B_G)$ (i.e., $P(B_E \cup B_G) = P(B_E) + P(B_G) - P(B_E \cap B_G)$.

# Analytical solution to example 7.9b

- The event $B$=“*there is a match between the two teams from England or between the two teams from Germany*” is the union of the events $B_E$=“*there is a match between the two teams from England*” and $B_G$=“*there is a match between the two teams from Germany*” (i.e., $B = B_E \cup B_G$).
- Since $B_E$ and $B_G$ are not disjoint, we should use Rule 3 to compute $P(B_E \cup B_G)$ (i.e., $P(B_E \cup B_G) = P(B_E) + P(B_G) - P(B_E \cap B_G)$.
- $P(B_E) = P(B_G)$. Since the match $\{E_1, E_2\}$ is in all quarterfinals in $B_E$, to calculate $N(B_E)$ we need to computer the number of matches between six teams of three countries, as done in part (a). This gives $N(B_E) = \frac{\binom{6}{2}\binom{4}{2}}{3!}$. Then $P(B_E) = P(B_G) = \frac{\frac{\binom{6}{2}\binom{4}{2}}{3!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.142857$.

# Analytical solution to example 7.9b

- The event $B$="*there is a match between the two teams from England or between the two teams from Germany*" is the union of the events $B_E$="*there is a match between the two teams from England*" and $B_G$="*there is a match between the two teams from Germany*" (i.e., $B = B_E \cup B_G$).
- Since $B_E$ and $B_G$ are not disjoint, we should use Rule 3 to compute $P(B_E \cup B_G)$ (i.e., $P(B_E \cup B_G) = P(B_E) + P(B_G) - P(B_E \cap B_G)$.
- $P(B_E) = P(B_G)$. Since the match $\{E_1, E_2\}$ is in all quarterfinals in $B_E$, to calculate $N(B_E)$ we need to computer the number of matches between six teams of three countries, as done in part (a). This gives $N(B_E) = \frac{\binom{6}{2}\binom{4}{2}}{3!}$.
  Then $P(B_E) = P(B_G) = \dfrac{\frac{\binom{6}{2}\binom{4}{2}}{3!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.142857$.
- Since the match $\{E_1, E_2\}$ and $\{G_1, G_2\}$ are in all quarterfinals in $B_E \cap B_G$, as in part (a), $N(B_E \cap B_G) = \frac{\binom{4}{2}}{2!}$. Then
  $$P(B_E \cap B_G) = \frac{\frac{\binom{4}{2}}{2!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.028571.$$

# Analytical solution to example 7.9b

- The event $B$=“*there is a match between the two teams from England or between the two teams from Germany*” is the union of the events $B_E$=“*there is a match between the two teams from England*” and $B_G$=“*there is a match between the two teams from Germany*” (i.e., $B = B_E \cup B_G$).
- Since $B_E$ and $B_G$ are not disjoint, we should use Rule 3 to compute $P(B_E \cup B_G)$ (i.e., $P(B_E \cup B_G) = P(B_E) + P(B_G) - P(B_E \cap B_G)$.
- $P(B_E) = P(B_G)$. Since the match $\{E_1, E_2\}$ is in all quarterfinals in $B_E$, to calculate $N(B_E)$ we need to computer the number of matches between six teams of three countries, as done in part (a). This gives $N(B_E) = \frac{\binom{6}{2}\binom{4}{2}}{3!}$. Then $P(B_E) = P(B_G) = \frac{\frac{\binom{6}{2}\binom{4}{2}}{3!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.142857$.
- Since the match $\{E_1, E_2\}$ and $\{G_1, G_2\}$ are in all quarterfinals in $B_E \cap B_G$, as in part (a), $N(B_E \cap B_G) = \frac{\binom{4}{2}}{2!}$. Then $P(B_E \cap B_G) = \frac{\frac{\binom{4}{2}}{2!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.028571$.
- Thus $P(B_E \cup B_G) = P(B_E) + P(B_G) - P(B_E \cap B_G) = 2 \times 0.142857 - 0.028571 = 0.257143$.

# Simulated solution to example 7.9a

Please see code here.

# Exercise

Solve example 7.7b by simulation. Answer: you should obtain a solution close to the analytical one.

# Contents

# Definition of conditional probability: motivation



- We are given a probability model $(\Omega, \mathcal{F}, P)$.

# Definition of conditional probability: motivation



- We are given a probability model $(\Omega, \mathcal{F}, P)$.
- We are interested in the probability of event $A \in \mathcal{F}$. This model provides us the unconditioned probability of $A$, $P(A)$.

# Definition of conditional probability: motivation

- We are given a probability model $(\Omega, \mathcal{F}, P)$.
- We are interested in the probability of event $A \in \mathcal{F}$. This model provides us the unconditioned probability of $A$, $P(A)$.
- Our colleague performs an experiment and tells us that event $B$ ocurred. How does the fact that $B$ occurred changes the probability of $A$?

$$P(A) \xrightarrow{\text{B occurred}} P(A|B)$$

# Definition of conditional probability: motivation

- We are given a probability model $(\Omega, \mathcal{F}, P)$.
- We are interested in the probability of event $A \in \mathcal{F}$. This model provides us the unconditioned probability of $A$, $P(A)$.
- Our colleague performs an experiment and tells us that event $B$ ocurred. How does the fact that $B$ occurred changes the probability of $A$?

$$P(A) \xrightarrow{\text{B occurred}} P(A|B)$$

- Definition 1: $P(\cdot|B) = P(\cdot \cap B)$, where $\cdot$ can be any event.

# Definition of conditional probability: motivation

- We are given a probability model $(\Omega, \mathcal{F}, P)$.
- We are interested in the probability of event $A \in \mathcal{F}$. This model provides us the unconditioned probability of $A$, $P(A)$.
- Our colleague performs an experiment and tells us that event $B$ ocurred. How does the fact that $B$ occurred changes the probability of $A$?

$$P(A) \xrightarrow{\text{B occurred}} P(A|B)$$

- Definition 1: $P(\cdot|B) = P(\cdot \cap B)$, where $\cdot$ can be any event.
- Problem with Definition 1: we want $P(B|B) = 1$, but from Definition 1, $P(B|B) = P(B \cap B) = P(B) \leq 1$.

# Definition of conditional probability: motivation

- We are given a probability model $(\Omega, \mathcal{F}, P)$.
- We are interested in the probability of event $A \in \mathcal{F}$. This model provides us the unconditioned probability of $A$, $P(A)$.
- Our colleague performs an experiment and tells us that event $B$ ocurred. How does the fact that $B$ occurred changes the probability of $A$?

$$P(A) \xrightarrow{\text{B occurred}} P(A|B)$$

- Definition 1: $P(\cdot|B) = P(\cdot \cap B)$, where $\cdot$ can be any event.
- Problem with Definition 1: we want $P(B|B) = 1$, but from Definition 1, $P(B|B) = P(B \cap B) = P(B) \leq 1$.
- Definition 2: $P(\cdot|B) = \frac{P(\cdot \cap B)}{P(B)}$.

# Definition of conditional probability: motivation

- We are given a probability model $(\Omega, \mathcal{F}, P)$.
- We are interested in the probability of event $A \in \mathcal{F}$. This model provides us the unconditioned probability of $A$, $P(A)$.
- Our colleague performs an experiment and tells us that event $B$ ocurred. How does the fact that $B$ occurred changes the probability of $A$?

$$P(A) \xrightarrow{\text{B occurred}} P(A|B)$$

- Definition 1: $P(\cdot|B) = P(\cdot \cap B)$, where $\cdot$ can be any event.
- Problem with Definition 1: we want $P(B|B) = 1$, but from Definition 1, $P(B|B) = P(B \cap B) = P(B) \leq 1$.
- Definition 2: $P(\cdot|B) = \frac{P(\cdot \cap B)}{P(B)}$.
    - $P(B|B) = 1$ ✓

# Definition of conditional probability: motivation

- We are given a probability model $(\Omega, \mathcal{F}, P)$.
- We are interested in the probability of event $A \in \mathcal{F}$. This model provides us the unconditioned probability of $A$, $P(A)$.
- Our colleague performs an experiment and tells us that event $B$ ocurred. How does the fact that $B$ occurred changes the probability of $A$?

$$P(A) \xrightarrow{\text{B occurred}} P(A|B)$$

- Definition 1: $P(\cdot|B) = P(\cdot \cap B)$, where $\cdot$ can be any event.
- Problem with Definition 1: we want $P(B|B) = 1$, but from Definition 1, $P(B|B) = P(B \cap B) = P(B) \leq 1$.
- Definition 2: $P(\cdot|B) = \frac{P(\cdot \cap B)}{P(B)}$.
  - $P(B|B) = 1$ ✓
  - $P(A) = P(A|\Omega) = \frac{P(A \cap \Omega)}{P(\Omega)} = \frac{P(A)}{1} = P(A)$ ✓

# Definition of conditional probability



**Definition 5 (Conditional probability)**

*For any to events $A$ and $B$, with $P(B) > 0$, the conditional probability of $A$ given $B$, $P(A|B)$, is defined as*

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

# Example 8.1: conditional probability for two dice



**Example 8.1**

*Someone has rolled two dice. You know that one of the dice turned up a face value of six. What is the probability that the other die turned up a six as well?*

# Analytical solution to example 8.1

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$. $N = 36$. Equally probable outcomes.

# Analytical solution to example 8.1

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$. $N = 36$. Equally probable outcomes.
- $B$= "*one die turned up a face value of six*" $= \{(6,i), (j,6), (6,6) : 1 \leq i,j \leq 5\}$. $N(B) = 11$. $P(B) = \frac{N(B)}{N} = \frac{11}{36}$.

# Analytical solution to example 8.1

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$. $N = 36$. Equally probable outcomes.
- $B$=*"one die turned up a face value of six"* $= \{(6, i), (j, 6), (6, 6) : 1 \leq i, j \leq 5\}$. $N(B) = 11$. $P(B) = \frac{N(B)}{N} = \frac{11}{36}$.
- $A$=*"the other die turned up a face value of six"*

# Analytical solution to example 8.1

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$. $N = 36$. Equally probable outcomes.
- $B$="*one die turned up a face value of six*"$=$ $\{(6,i),(j,6),(6,6) : 1 \leq i,j \leq 5\}$. $N(B) = 11$. $P(B) = \frac{N(B)}{N} = \frac{11}{36}$.
- $A$="*the other die turned up a face value of six*"
- $A \cap B$="*the two dice turned up a face value of six*"$=$ $\{(6,6)\}$. $N(A \cap B) = 1$. $P(A \cap B) = \frac{N(A \cap B)}{N} = \frac{1}{36}$.

# Analytical solution to example 8.1



- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$. $N = 36$. Equally probable outcomes.
- $B$="*one die turned up a face value of six*"$=\{(6,i),(j,6),(6,6) : 1 \leq i,j \leq 5\}$. $N(B) = 11$. $P(B) = \frac{N(B)}{N} = \frac{11}{36}$.
- $A$="*the other die turned up a face value of six*"
- $A \cap B$="*the two dice turned up a face value of six*"$=\{(6,6)\}$. $N(A \cap B) = 1$. $P(A \cap B) = \frac{N(A \cap B)}{N} = \frac{1}{36}$.
- Approach 1 [definition]: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{11/36} = \frac{1}{11}$.

# Analytical solution to example 8.1

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$. $N = 36$. Equally probable outcomes.
- $B$=*"one die turned up a face value of six"*$=$ $\{(6,i),(j,6),(6,6) : 1 \leq i,j \leq 5\}$. $N(B) = 11$. $P(B) = \frac{N(B)}{N} = \frac{11}{36}$.
- $A$=*"the other die turned up a face value of six"*
- $A \cap B$=*"the two dice turned up a face value of six"*$=$ $\{(6,6)\}$. $N(A \cap B) = 1$. $P(A \cap B) = \frac{N(A \cap B)}{N} = \frac{1}{36}$.
- Approach 1 [definition]: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{11/36} = \frac{1}{11}$.
- Approach 2 [$B$ as $\Omega$]: $P(A|B) = \frac{N(A \cap B)}{N(B)} = \frac{1}{11}$.

# Simulated solution to example 8.1



Please see code here.

# Exercise



Someone has rolled two dice. You know that the first die turned up a face value of six. What is the probability that the second die turned up a six as well?

# Assigning probabilities by conditioning



**Rule 5**

*For any $n \in \mathbb{N}$, $n \geq 2$, and any sequence of events $A_1, \ldots, A_n$,*

$$P(A_1 \cap \ldots \cap A_n) = P(A_n | A_{n-1} \cap \ldots \cap A_1) P(A_{n-1} | A_{n-2} \cap \ldots \cap A_1) \ldots P(A_1)$$

# Proof of rule 5

**Proof.**

By induction.

$$P_2: \quad P(A_1 \cap A_2) = P(A_2|A_1)P(A_1)$$

$P_n \to P_{n+1}$:

$$\begin{aligned} P(A_1 \cap A_2 \cap \ldots \cap A_n \cap A_{n+1}) &= P(B \cap A_{n+1}) \\ &= P(A_{n+1}|B)P(B) \\ &= P(A_{n+1}|A_n \cap A_{n-1} \cap \ldots A_1) \\ &\quad P(A_n \cap A_{n-1} \cap \ldots \cap A_1) \\ &= P(A_{n+1}|A_n \cap A_{n-1} \cap \ldots A_1) \\ &\quad P(A_n|A_{n-1} \cap \ldots \cap A_1) \\ &\quad P(A_{n-1}|A_{n-2} \cap \ldots \cap A_1) \ldots P(A_1) \end{aligned}$$

□

Notes:

1. $P_2$ follows from the definition of conditional probability.
2. in the first equality in $P_n \to P_{n+1}$ we defined the event $B = A_1 \cap \ldots \cap A_n$.
3. in the second equality in $P_n \to P_{n+1}$ we used the definition of conditional proability.
4. in the third equality in $P_n \to P_{n+1}$ we replaced $B$ by its definition.
5. in the fourth equality in $P_n \to P_{n+1}$ we used the inductive hypothesis $P(A_1 \cap \ldots \cap A_n) = P(A_n|A_{n-1} \cap \ldots \cap A_1)P(A_{n-1}|A_{n-2} \cap \ldots \cap A_1) \ldots P(A_1)$.

# Example:



- redo Example 7.9 (solution following Rule 4)

# Example:



- probability that it takes 10 or more cards before the first ace appears

# Independent events



- motivation of independence definition with conditional probabilities
- Definition 8.2

# Example

- Example 8.5

# Example



- Example 8.6 (uses birthday problem, example 7.10)

# Law of conditional probability



- example of die followed by coin tosses

Rule 5 law of conditional probability. Let $A$ be an event that can only occur if one of the mutually exclusive events $B_1, \ldots, B_n$ occurs. Then

$$P(A) = P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)$$

# Example

- example 8.6: tour the France (difficult!)

# Bayes rule in odds form

- true/false hypothesis

Rule 6 The posterior probability $P(H|E)$ satisfies

$$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(H)}{P(\bar{H})}\frac{P(E|H)}{P(E|\bar{H})}$$

- interpretation of rule 6
- avoid need of P(E)
- prior odds + likelihood ratio or Bayes factor
- prior odds update with new evidence
- sequential update (mention Bayesian linear regression)

# Example

- example 8.8

# Example



- example 8.11
- add to the problem statement:
- in 1992, 4936 women were murdered in the US, of which roughly 1430 were murdered by their (ex)husbands or boyfriends
- 5% of the married women in the US have at some point been physically abused by their husbands.
- assume that a woman who has been murdered by some other than her husband had the same same chance of being abused by her husband as a randomly selected woman
- Alan Dershowitz admitted that a substantial percentage of the husbands who murder their wives, previous to the murder, also physically abuse their wives. Given this statement, we assume that the proability that a husband physically abused his wife, given that he killed her, is 50 percent.

# Bayesian inference – discrete case



- explain posterior sequential update

# Example



- example 8.13 (solve it analytically and by sampling)

# References

Tijms, H. (2012). *Understanding probability*. Cambridge University Press.