

## Foundations of probability theory

Historical notes

Axiomatic definition  
of probability

Some basic rules

## Conditional probability

Conditional  
probability

Assigning  
probabilities by  
conditioning

Independent events

Conditionally  
independent events

Law of conditional  
probability

## Statistical inference

Bayesian inference –  
discrete case

## References

# Foundations of probability theory

Joaquín Rapela

Gatsby Computational Neuroscience Unit  
University College London

June 26, 2023

# Contents

## Foundations of probability theory

Historical notes  
Axiomatic definition of probability  
Some basic rules

## Conditional probability

Conditional probability  
Assigning probabilities by conditioning  
Independent events  
Conditionally independent events  
Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 1 Foundations of probability theory
  - Historical notes
  - Axiomatic definition of probability
  - Some basic rules
- 2 Conditional probability
  - Conditional probability
  - Assigning probabilities by conditioning
  - Independent events
  - Conditionally independent events
  - Law of conditional probability
- 3 Statistical inference
  - Bayesian inference – discrete case

# Main reference

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

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

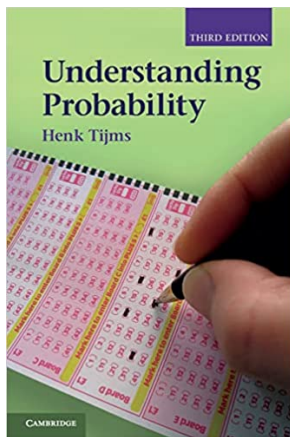
## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

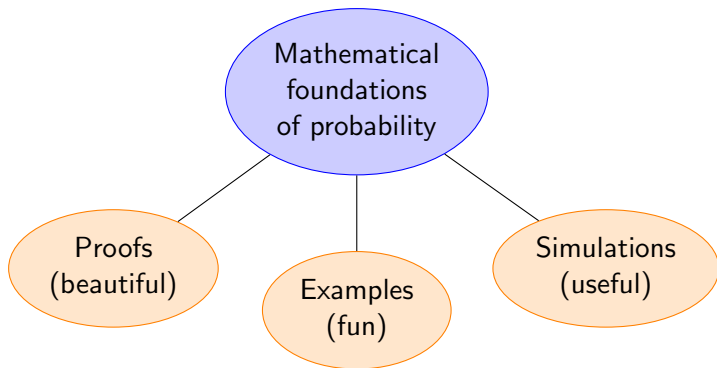
## Statistical inference

- Bayesian inference – discrete case

## References



# Main lecture goals



Foundations  
of probability  
theory

Historical notes

Axiomatic definition  
of probability

Some basic rules

Conditional  
probability

Conditional  
probability

Assigning  
probabilities by  
conditioning

Independent events

Conditionally  
independent events

Law of conditional  
probability

Statistical  
inference

Bayesian inference –  
discrete case

References

# Contents

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

- 1 Foundations of probability theory
- 2 Conditional probability
- 3 Statistical inference

# Historical notes

## Foundations of probability theory

### Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

**Gerolamo Cardano (1501-1576)** first mathematical analysis of games of chance, introduced probability as the ratio between favorable and the total number of possible outcomes

**Pierre Fermat and Blaise Pascal (1654)** correspondence about the solution of a problem posed by the Chevalier de Mere established the basic principles of probability.

**Christian Huygen (1629-1695)** laid the foundations of current probability theory. Introduced the concept of expected value.

**Pierre Simon Laplace (1749-1827)** published *Theorie Analytique des Probabilities*, the greatest contribution in the history of probability.

**Andrej Kolmogorov (1903-1987)** laid the current axiomatic foundation of probability.

# Probability model

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Definition 1 (Probability model)

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

# Probability model

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Inference

Bayesian inference – discrete case

## References

### Definition 1 (Probability model)

A **probability model** is a mathematical 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 assignment of probability to events (i.e., **probability measure**).



# Sample space

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Definition 2 (Sample space)

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

# Event

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Definition 3 (Event)

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

# Event

Foundations  
of probability  
theory

Historical notes

Axiomatic definition  
of probability

Some basic rules

Conditional  
probability

Conditional  
probability

Assigning  
probabilities by  
conditioning

Independent events

Conditionally  
independent events

Law of conditional  
probability

Statistical  
inference

Bayesian inference –  
discrete case

References

## 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Definition 4 (Probability measure)

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

# Probability model: definitions

## Foundations of probability theory

Historical notes  
Axiomatic definition of probability  
Some basic rules

## Conditional probability

Conditional probability  
Assigning probabilities by conditioning  
Independent events  
Conditionally independent events  
Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### 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.,  $P : \mathcal{F} \rightarrow [0, 1]$ ).*

### Definition (Probability model)

*A **probability model**,  $\mathcal{M}$ , is a mathematical representation of a random experiment consisting 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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 repeatedly roll a fair die and count the number of rolls until the first six shows up.

# Examples of probability models

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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 repeatedly 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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 repeatedly 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  $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \dots$  can be assigned to the outcomes  $1, 2, 3, \dots$

# Examples of probability models

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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 repeatedly 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  $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \dots$  can be assigned to the outcomes  $1, 2, 3, \dots$ .
- 3 The experiment is to measure the time between the first and second spikes in an experimental trial.

# Examples of probability models

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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 repeatedly 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  $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \dots$  can be assigned to the outcomes  $1, 2, 3, \dots$ .
- 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.

# Examples of probability models

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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 repeatedly 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  $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \dots$  can be assigned to the outcomes  $1, 2, 3, \dots$
- 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Axiom 1

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

### Axiom 2

$$P(\Omega) = 1$$

### Axiom 3

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

# Some basic rules

## Foundations of probability theory

Historical notes  
Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability  
Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Rule 1

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

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

### 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  $B$ ,*

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



# Experiment with equally likely outcomes

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

Foundations  
of probability  
theory

Historical notes  
Axiomatic definition  
of probability  
Some basic rules

Conditional  
probability

Conditional  
probability  
Assigning  
probabilities by  
conditioning  
Independent events  
Conditionally  
independent events  
Law of conditional  
probability

Statistical  
inference

Bayesian inference –  
discrete case

References

# Experiment with equally likely outcomes

An experiment with equally likely outcomes is one with a finite number of outcomes  $\omega_1, \dots, \omega_N$ , 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  $A$ .*

Foundations  
of probability  
theory

Historical notes  
Axiomatic definition  
of probability  
Some basic rules

Conditional  
probability

Conditional  
probability  
Assigning  
probabilities by  
conditioning  
Independent events  
Conditionally  
independent events  
Law of conditional  
probability

Statistical  
inference

Bayesian inference –  
discrete case

References

# Experiment with equally likely outcomes

## Foundations of probability theory

Historical notes  
Axiomatic definition of probability  
Some basic rules

## Conditional probability

Conditional probability  
Assigning probabilities by conditioning  
Independent events  
Conditionally independent events  
Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

An experiment with equally likely outcomes is one with a finite number of outcomes  $\omega_1, \dots, \omega_N$ , 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  $A$ .*

### Proof.

$$\begin{aligned} A &= \{a_1, \dots, a_{N(A)}\} = \{a_1\} \cup \dots \cup \{a_{N(A)}\} \\ P(A) &= P(\{a_1\}) + \dots + P(\{a_{N(A)}\}) = \sum_{i=1}^{N(A)} P(\{a_i\}) \\ &= \sum_{i=1}^{N(A)} \frac{1}{N} = \frac{N(A)}{N} \end{aligned}$$



# Example 7.1: John, Pedro and Rosita's dice game

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

### Example 7.1

*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?*

# Analytical solution of example 7.1

## Foundations of probability theory

Historical notes  
Axiomatic definition of probability  
Some basic rules

## Conditional probability

Conditional probability  
Assigning probabilities by conditioning  
Independent events  
Conditionally independent events  
Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

$$\Omega = \{(J, P, R) : 1 \leq J, P, R \leq 6\}$$

$$N = 6^3 = 216$$

$$\begin{aligned} A &= \{(J, P, R) \in \Omega : R = J + P\} \\ &= \{(1, 1, 2), (1, 2, 3), (1, 3, 4), (1, 4, 5), (1, 5, 6), \\ &\quad (2, 1, 3), (2, 2, 4), (2, 3, 5), (2, 4, 6), \\ &\quad (3, 1, 4), (3, 2, 5), (3, 3, 6), (4, 1, 5), (4, 2, 6), (5, 1, 6)\} \end{aligned}$$

$$N(A) = 15$$

$$P(A) = \frac{N(A)}{N} = \frac{15}{216} = 0.0694$$

# Simulated solution of example 7.1

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability

### Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

Please see code [here](#).

# Rule 1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Proof.

Denote by  $\emptyset$  the empty set. We will first prove that  $P(\emptyset) = 0$ . Take  $A_i = \emptyset$  for  $i = 1, 2, \dots$ . 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  $j = 1, 2, \dots$ . Then

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



### 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### 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,  $1 = P(A) + P(A^c)$ .





# Exercise

## Foundations of probability theory

Historical notes

Axiomatic definition  
of probability

Some basic rules

## Conditional probability

Conditional  
probability

Assigning  
probabilities by  
conditioning

Independent events

Conditionally  
independent events

Law of conditional  
probability

## Statistical inference

Bayesian inference –  
discrete case

## References

Prove rule 3.

# Proof of rule 4

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Proof.

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

$$A = (A \setminus B) \cup (A \cap B)$$

$$B = (B \setminus A) \cup (A \cap B)$$

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

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

$$P(A) = P(A \setminus B) + P(A \cap B) \rightarrow P(A \setminus B) = P(A) - P(A \cap B)$$

$$P(B) = P(B \setminus A) + P(A \cap B) \rightarrow P(B \setminus A) = P(B) - P(A \cap B)$$

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

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



# Example 7.7: Chevalier de Mere to Blaise Pascal

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

The gambler Chevalier de Mere posed the following problem to the famous French mathematician Blaise Pascal in 1654. This problem marks the beginning 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 of example 7.7a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

(a)

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

# Analytical solution of example 7.7a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

(a)

- Let's fix the number of rolls  $r$ . The sample space is  $\Omega = \{(i_1, \dots, i_r) : 1 \leq i_k \leq 6\}$ , where  $i_k$  is the up face of the die on the  $k$ th 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 of example 7.7a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

(a)

- Let's fix the number of rolls  $r$ . The sample space is  $\Omega = \{(i_1, \dots, i_r) : 1 \leq i_k \leq 6\}$ , where  $i_k$  is the up face of the die on the  $k$ th 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 of example 7.7a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

(a)

- Let's fix the number of rolls  $r$ . The sample space is  $\Omega = \{(i_1, \dots, i_r) : 1 \leq i_k \leq 6\}$ , where  $i_k$  is the up face of the die on the  $k$ th 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 of example 7.7a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Please see code [here](#).



# Exercise

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

# Example 7.9: soccer teams in quarterfinal

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Example 7.9

*The eight soccer teams which have reached the quarterfinals of the Championship League are formed by two teams from each of 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 of example 7.9: setup

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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 of example 7.9: setup

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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.,  $\text{match} = \{T_i, T_j\}, \text{match} \subset T, T_i \neq T_j$ ).

# Analytical solution of example 7.9: setup

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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.,  $\text{match} = \{T_i, T_j\}, \text{match} \subset T, T_i \neq T_j$ ).
- A quarterfinal is a set of four different matches.

# Analytical solution of example 7.9: setup

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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.,  $\text{match} = \{T_i, T_j\}, \text{match} \subset T, T_i \neq T_j$ ).
- A quarterfinal is a set of four different matches.
- $\Omega$  is the set of all possible quarterfinals (e.g.,  $\{\{E_1, G_1\}, \{I_1, G_2\}, \{S_2, G_2\}, \{I_2, G_1\}\} \in \Omega$ ).

# Analytical solution of example 7.9: setup

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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.,  $\text{match} = \{T_i, T_j\}, \text{match} \subset T, T_i \neq T_j$ ).
- A quarterfinal is a set of four different matches.
- $\Omega$  is the set of all possible quarterfinals (e.g.,  $\{\{E_1, G_1\}, \{I_1, G_2\}, \{S_2, G_2\}, \{I_2, G_1\}\} \in \Omega$ ).
- $\Omega$  contains equally-likely outcomes. Thus, for any event  $A$ ,  
$$P(A) = \frac{N(A)}{N}.$$

# Analytical solution of example 7.9: setup

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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.,  $\text{match} = \{T_i, T_j\}, \text{match} \subset T, T_i \neq T_j$ ).
- A quarterfinal is a set of four different matches.
- $\Omega$  is the set of all possible quarterfinals (e.g.,  $\{\{E_1, G_1\}, \{I_1, G_2\}, \{S_2, G_2\}, \{I_2, G_1\}\} \in \Omega$ ).
- $\Omega$  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 of example 7.9a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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 of example 7.9a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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 of example 7.9b

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

# Analytical solution of example 7.9b

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- The event  $B = \text{"there is a match between the two teams from England or between the two teams from Germany"}$  is the union of the events  $B_E = \text{"there is a match between the two teams from England"}$  and  $B_G = \text{"there is a match between the two teams from Germany"}$  (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 of example 7.9b

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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 compute the number of matches between six teams of three countries, as done in part (a). This gives  $N(B_E) = \frac{\binom{6}{2}\binom{4}{2}}{3!}$ .

$$\text{Then } P(B_E) = P(B_G) = \frac{\frac{\binom{6}{2}\binom{4}{2}}{3!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.142857.$$

# Analytical solution of example 7.9b

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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 compute 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  $P(B_E) = P(B_G) = \frac{\frac{\binom{4}{2}\binom{2}{2}}{2!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.142857$ .
- 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 of example 7.9b

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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 compute 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  $P(B_E) = P(B_G) = \frac{\frac{\binom{4}{2}\binom{2}{2}}{2!}}{\frac{\binom{8}{2}\binom{6}{2}\binom{4}{2}}{4!}} = 0.142857$ .
- 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 of example 7.9a

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Please see code [here](#).



# Exercise

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

# Contents

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

- 1 Foundations of probability theory
- 2 Conditional probability
- 3 Statistical inference

# Definition of conditional probability: motivation

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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

**Conditional probability**

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

# 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  $A$ ,  $P(A)$ .

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

# Definition of conditional probability: motivation

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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  $B$  occurred. 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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  $B$  occurred. 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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  $B$  occurred. How does the fact that  $B$  occurred changes the probability of  $A$ ?

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

- 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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  $B$  occurred. How does the fact that  $B$  occurred changes the probability of  $A$ ?

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

- 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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  $B$  occurred. 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:  $P(\cdot|B) = \frac{P(\cdot \cap B)}{P(B)}$ .
  - $P(B|B) = 1 \checkmark$

# Definition of conditional probability: motivation

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- 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  $B$  occurred. 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Definition 5 (Conditional probability)

*For any two 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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### 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 of example 8.1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

# Analytical solution of example 8.1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

# Analytical solution of example 8.1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$ .  $N = 36$ . 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 of example 8.1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$ .  $N = 36$ . 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 of example 8.1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$ .  $N = 36$ . 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 of example 8.1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- $\Omega = \{(i_1, i_2) : 1 \leq i_1, i_2 \leq 6\}$ .  $N = 36$ . 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 of example 8.1

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Please see code [here](#).

# Exercise

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

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

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Rule 5

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

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

# Proof of rule 5

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Proof.

By induction.

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

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

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



Notes:

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

# Example 8.3: allocating tourists to hotels

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Example 8.3

*A group of 15 tourists is stranded in a city with four hotels of the same class. Each of the hotels has enough room available to accommodate the 15 tourists. The group's guide, who has a good working relationship with each of the four hotels, assigns the tourists to the hotels as follows. First, he randomly determines how many tourists will go to hotel A, then how many of the remaining tourists will go to hotel B, and next how many of the still remaining tourists will go to hotel C. All remaining tourists are sent to hotel D. At each stage of the assignment the guide draws a random number between zero and the number of tourists left.*

- a** Calculate the probability of any assignment of tourists to hotels.
- b** Check that the probability of all possible assignments equals one.
- c** What is the probability that all four hotels receive guests?
- d** Is the selected rule fair to the four hotels?

# Analytical solution of example 8.3

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Let  $i, j$  and  $k$  be the number of tourists assigned to hotels A, B and C respectively. Then

$$\begin{aligned}\Omega = \{(i, j, k, 15 - (i + j + k)) : 0 \leq i \leq 15, \\ 0 \leq j \leq 15 - i; \\ 0 \leq k \leq 15 - (i + j)\}\end{aligned}$$

- a Define the events  $E_j(k)$  = "hotel  $W$  receives  $k$  tourists," with  $W \in \{A, B, C, D\}$ . The assignment  $(i, j, k, 15 - (i + j + k))$  is the only member of the event  $E_A(i) \cap E_B(j) \cap E_C(k) \cap E_D(15 - (i + j + k))$ . Then

$$\begin{aligned}P(\{(i, j, k, 15 - (i + j + k))\}) &= P(E_A(i) \cap E_B(j) \cap E_C(k)) \\ &= P(E_C(k) | E_A(i) \cap E_B(j)) P(E_B(j) | E_A(i)) P(E_A(i))\end{aligned}$$

$$P(E_A(i)) = \frac{1}{16}$$

$$P(E_B(j) | E_A(i)) = \frac{1}{16 - i}$$

$$P(E_C(k) | E_B(j) \cap E_A(i)) = \frac{1}{16 - (i + j)}$$

Therefore

$$P(\{(i, j, k, 15 - (i + j + k))\}) = \frac{1}{16 - (i + j)} \frac{1}{16 - i} \frac{1}{16}$$



# Analytical solution of example 8.3

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Let  $i, j$  and  $k$  be the number of tourists assigned to hotels A, B and C respectively. Then

$$\Omega = \{(i, j, k, 15 - (i + j + k)) : 0 \leq i \leq 15, \\ 0 \leq j \leq 15 - i, \\ 0 \leq k \leq 15 - (i + j)\}$$

b

$$\begin{aligned} P(\Omega) &= \sum_{\omega \in \Omega} P(\omega) = \sum_{i=0}^{15} \sum_{j=0}^{15-i} \sum_{k=0}^{15-(i+j)} P(\{(i, j, k, 15 - (i + j + k))\}) \\ &= \sum_{i=0}^{15} \sum_{j=0}^{15-i} \sum_{k=0}^{15-(i+j)} \frac{1}{16 - (i + j)} \frac{1}{16 - i} \frac{1}{16} \\ &= \frac{1}{16} \sum_{i=0}^{15} \frac{1}{16 - i} \sum_{j=0}^{15-i} \frac{1}{16 - (i + j)} \sum_{k=0}^{15-(i+j)} 1 \\ &= \frac{1}{16} \sum_{i=0}^{15} \frac{1}{16 - i} \sum_{j=0}^{15-i} \frac{1}{16 - (i + j)} (16 - (i + j)) \\ &= \frac{1}{16} \sum_{i=0}^{15} \frac{1}{16 - i} \sum_{j=0}^{15-i} 1 = \frac{1}{16} \sum_{i=0}^{15} \frac{1}{16 - i} (16 - i) = \frac{1}{16} \sum_{i=0}^{15} 1 = \frac{1}{16} 16 = 1 \end{aligned}$$

# Analytical solution of example 8.3

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References



Define the event  $E = \text{"all hotels receive guests."}$  Then

$$E = \{(i, j, k, 15 - (i + j + k)) : 1 \leq i \leq 12, \\ 1 \leq j \leq 13 - i; \\ 1 \leq k \leq 14 - (i + j)\}$$

$$P(E) = \sum_{i=1}^{12} \sum_{j=1}^{13-i} \sum_{k=1}^{14-(i+j)} P(\{(i, j, k, 15 - (i + j + k))\}) \\ = \sum_{i=1}^{12} \sum_{j=1}^{13-i} \sum_{k=1}^{14-(i+j)} \frac{1}{16 - (i + j)} \frac{1}{16 - i} \frac{1}{16} \\ = 0.2856$$



$$P(\{(15, 0, 0, 0)\}) = \frac{1}{16} \\ P(\{(0, 15, 0, 0)\}) = \frac{1}{16^2}$$

Thus, the selected rule is unfair to the four hotels.

# Simulated solution of example 8.3c

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Please see code [here](#).

# Independent events

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

## Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

If event  $A$  is independent from event  $B$ , then knowing that  $B$  has occurred does not change the probability that event  $A$  occurs.

### Definition 6

*Two events  $A$  and  $B$  are independent if*

$$P(A|B) = P(A)$$

# Independent events

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

## Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

If event  $A$  is independent from event  $B$ , then knowing that  $B$  has occurred does not change the probability that event  $A$  occurs.

### Definition 6

*Two events  $A$  and  $B$  are independent if*

$$P(A|B) = P(A)$$

### Claim

*If events  $A$  and  $B$  are independent if and only if*

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

# Independent events

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

## Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

If event  $A$  is independent from event  $B$ , then knowing that  $B$  has occurred does not change the probability that event  $A$  occurs.

### Definition 6

*Two events  $A$  and  $B$  are independent if*

$$P(A|B) = P(A)$$

### Claim

*If events  $A$  and  $B$  are independent if and only if*

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

### Proof.

$$\rightarrow: P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

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



# Example 8.5: independent dice events?

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

## Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Example 8.5

*Two fair dice are thrown. Let  $A$  be the event that the number shown in die one is even, and  $B$  the event that the sum of the two dice is odd. Are  $A$  and  $B$  independent?*

# Analytical solution of example 8.5

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

## Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

The sample space is

$$\begin{aligned}\Omega &= \{(d_1, s) : d_1 \text{ number shown in die one, } s \text{ sum of the two dice}\} \\ &= \{(1, 2), (1, 3), \dots, (1, 7), \\ &\quad (2, 3), (2, 4), \dots, (2, 8), \\ &\quad \dots, \\ &\quad (6, 7), (6, 8), \dots, (6, 12)\}\end{aligned}$$

$$N(\Omega) = 36, N(A) = N(B) = 18, N(A \cap B) = 9$$

$$P(B) = \frac{N(B)}{N(\Omega)} = \frac{18}{36} = \frac{1}{2}$$

$$P(B|A) = \frac{N(A \cap B)}{N(A)} = \frac{9}{18} = \frac{1}{2}$$

Since  $P(B) = P(B|A)$ , events  $A$  and  $B$  are independent.



# Exercise: independence-related proofs

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

## Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Prove the following claims:

**Claim (disjoint events are generally not independent)**

*Let  $A$  and  $B$  be two events and  $P$  a probability measure such that  $P(A) \neq 0$  and  $P(B) \neq 0$ . Then, it is always true that the fact that  $A$  and  $B$  are disjoint implies that  $A$  and  $B$  are not independent.*

**Claim (symmetry of independence)**

*$P(A|B) = P(A)$  if and only if  $P(B|A) = P(B)$ .*

# Conditionally independent events

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Definition 7

*Events  $A$  and  $B$  are conditionally independent given event  $C$  if and only if  $P(A|B \cap C) = P(A|C)$ .*

# Example on conditional independence

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

### Example tossing a possibly unfair coin

*Suppose both Martin and Norman toss the same coin. Let  $A$  be the event “Norman’s toss resulted in heads”, and  $B$  the event “Martin’s toss resulted in heads.” Assume also that there is a possibility that the coin is biased towards heads but we do not know this for certain. In this case events  $A$  and  $B$  are not independent. For example, observing that  $B$  occurred causes us to increase our belief that  $A$  occurred (i.e.,  $P(A|B) > P(A)$ ).*

*In the above, events  $A$  and  $B$  are dependent on event  $C$ , “the coin is biased towards heads.” Although events  $A$  and  $B$  are not independent, it turns out that they are conditionally independent given event  $C$ , i.e., once we know for certain that event  $C$  occurred then any evidence about  $B$  cannot change our belief about  $A$  (i.e.,  $P(A|C) = P(A|B \cap C)$ ).*

# Example on conditional independence

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Example tossing a possibly unfair coin

*Let's assign probabilities to the events above and verify the claims that (a) events  $A$  and  $B$  are dependent (i.e.,  $P(A|B) > P(A)$ ) and that (b) events  $A$  and  $B$  are conditionally independent given event  $C$  (i.e.,  $P(A|B \cap C) = P(A|B)$ ).*

- a** *list the elements of the sample space  $\Omega$ , and of events  $A$ ,  $B$  and  $C$ .*
- b** *assign probabilities to events containing only one element of  $\Omega$ , so that
  - *events  $A$  and  $B$  and events  $A$  and  $B^c$  are conditionally independent given that the coin is biased or given that the coin is unbiased,*
  - *the probability of heads given that the coin is biased is 0.7*
  - *the coin is equally likely biased or unbiased.**
- c** *with the previous assigned probabilities, check that events  $A$  and  $B$  are not independent but that they are conditionally independent given  $C$ .*

# Analytical solution of example on conditional independence

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References



$$\Omega = \{HHU, HTU, THU, TTU, HHB, HTB, THB, TTB\}$$

$$A = \{HHU, HTU, HHB, HHH\}$$

$$B = \{HHU, THU, HHB, HHH\}$$

$$C = \{HHB, HTB, THB, TTB\}$$

In the above, for example, HTB means that Norman tossed a heads, Martin tossed a tail and the coin was biased towards heads.

# Analytical solution of example on conditional independence

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

**b** Given the example setup we know that

$$P(A|C^c) = P(B|C^c) = \frac{1}{2}$$

$$P(A|C) = P(B|C) = \frac{1}{7}$$

$$P(C) = P(C^c) = \frac{1}{2}$$

$$P(A|B, C^c) = P(A|B^c, C^c) = P(A|C^c)$$

$$P(A|B, C) = P(A|B^c, C) = P(A|C)$$

# Analytical solution of example on conditional independence

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

**b** Next we will assign probabilities using Rule 5,

$$\begin{aligned}P(\{HHU\}) &= P(A \cap B \cap C^c) = P(A|B \cap C^c)P(B|C^c)P(C^c) \\&= P(A|C^c)P(B|C^c)P(C^c) = \frac{1}{2} \frac{1}{2} \frac{1}{2} = \frac{1}{8}\end{aligned}$$

The second equality above uses Rule 5, and the third equality uses the fact that  $A$  is conditionally independent from  $B$  given  $C^c$ .

In the same way we can see that

$$P(\{HTU\}) = P(\{THU\}) = P(\{TTU\}) = \frac{1}{8}.$$

# Analytical solution of example on conditional independence

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

**b** For a biased coin we have

$$\begin{aligned}P(\{HHB\}) &= P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C) \\&= P(A|C)P(B|C)P(C) = \frac{7}{10} \frac{7}{10} \frac{1}{2} = \frac{49}{200}\end{aligned}$$

$$\begin{aligned}P(\{HTB\}) &= P(A \cap B^c \cap C) = P(A|B^c \cap C)P(B^c|C)P(C) \\&= P(A|C)P(B^c|C)P(C) = P(A|C)(1 - P(B|C))P(C) \\&= \frac{7}{10} \frac{3}{10} \frac{1}{2} = \frac{21}{200}\end{aligned}$$

$$\begin{aligned}P(\{THB\}) &= P(A^c \cap B \cap C) = P(A^c|B \cap C)P(B|C)P(C) \\&= P(A^c|C)P(B|C)P(C) = (1 - P(A|C))P(B|C)P(C) \\&= \frac{3}{10} \frac{7}{10} \frac{1}{2} = \frac{21}{200}\end{aligned}$$

$$\begin{aligned}P(\{TTB\}) &= P(A^c \cap B^c \cap C) = P(A^c|B^c \cap C)P(B^c|C)P(C) \\&= P(A^c|C)P(B^c|C)P(C) = (1 - P(A|C))(1 - P(B|C))P(C) \\&= \frac{3}{10} \frac{3}{10} \frac{1}{2} = \frac{9}{200}\end{aligned}$$



# Analytical solution of example on conditional independence

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Bayesian inference

Bayesian inference – discrete case

## References

Thus, the probabilities of single-outcome events are:

$\{HHU\}$	$\{HTU\}$	$\{THU\}$	$\{TTU\}$	$\{HHB\}$	$\{HTB\}$	$\{THB\}$	$\{TTB\}$
$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{49}{200}$	$\frac{21}{200}$	$\frac{21}{200}$	$\frac{9}{200}$

$$\begin{aligned}
 P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P((A \cap B) \cap (C \cup C^c))}{P(B \cap ((A \cap C) \cup (A \cap C^c) \cup (A^c \cap C) \cup (A^c \cap C^c)))} \\
 &= \frac{P((A \cap B \cap C) \cup (A \cap B \cap C^c))}{P((B \cap A \cap C) \cup (B \cap A \cap C^c) \cup (B \cap A^c \cap C) \cup (B \cap A^c \cap C^c))} \\
 &= \frac{P(A \cap B \cap C) + P(A \cap B \cap C^c)}{P(B \cap A \cap C) + P(B \cap A \cap C^c) + P(B \cap A^c \cap C) + P(B \cap A^c \cap C^c)} \\
 &= \frac{P(\{HHB\}) + P(\{HHU\})}{P(\{HHB\}) + P(\{HHU\}) + P(\{THB\}) + P(\{THU\})} \\
 &= \frac{\frac{49}{200} + \frac{1}{8}}{\frac{49}{200} + \frac{1}{8} + \frac{21}{200} + \frac{1}{8}} = \frac{\frac{74}{200}}{\frac{120}{200}} = \frac{74}{120} = 0.62
 \end{aligned}$$

$$P(A) = P(B) = \frac{120}{200} = 0.6 < 0.62 = P(A|B) \quad \text{Therefore, A and B are not independent.}$$

# Analytical solution of example on conditional independence



$$\begin{aligned}
 P(A|B \cap C) &= \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(A \cap B \cap C)}{P((B \cap C) \cap (A \cup A^c))} = \frac{P(A \cap B \cap C)}{P((A \cap B \cap C) \cup (A^c \cap B \cap C))} \\
 &= \frac{P(A \cap B \cap C)}{P(A \cap B \cap C) + P(A^c \cap B \cap C)} = \frac{P(\{HHB\})}{P(\{HHB\}) + P(\{THB\})} = \frac{\frac{49}{200}}{\frac{49}{200} + \frac{21}{200}} = \frac{7}{10} \\
 P(A|C) &= \frac{P(A \cap C)}{P(C)} = \frac{P((A \cap C) \cap (B \cup B^c))}{P(C \cap ((A \cap B) \cup (A \cap B^c) \cup (A^c \cap B) \cup (A^c \cap B^c)))} \\
 &= \frac{P((A \cap B \cap C) \cup (A \cap B^c \cap C))}{P((A \cap B \cap C) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C) \cup (A^c \cap B^c \cap C))} \\
 &= \frac{P(A \cap B \cap C) + P(A \cap B^c \cap C)}{P(A \cap B \cap C) + P(A \cap B^c \cap C) + P(A^c \cap B \cap C) + P(A^c \cap B^c \cap C)} \\
 &= \frac{P(\{HHB\}) + P(\{HTB\})}{P(\{HHB\}) + P(\{HTB\}) + P(\{THB\}) + P(\{TTB\})} \\
 &= \frac{\frac{49}{200} + \frac{21}{200}}{\frac{49}{200} + \frac{21}{200} + \frac{21}{200} + \frac{9}{200}} \\
 &= \frac{\frac{70}{200}}{\frac{100}{200}} = \frac{7}{10} = P(A|B \cap C) \quad \text{Therefore, A and B are conditionally independent given C.}
 \end{aligned}$$

Foundations  
of probability  
theory

Historical notes

Axiomatic definition  
of probability

Some basic rules

Conditional  
probability

Conditional  
probability

Assigning  
probabilities by  
conditioning

Independent events

Conditionally  
independent events

Law of conditional  
probability

Statistical  
inference

Bayesian inference –  
discrete case

References

# Example p86: dice followed by coin tosses

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Example page 266

*A fair die is rolled to yield a number between one and six, and then a coin is tossed that many times. What is the probability that heads will not appear?*

# Analytical solution of example p86

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Bayesian inference

Bayesian inference – discrete case

## References

$$\begin{aligned}P(\text{no\_heads}) &= P(\text{no\_heads} \cap \Omega) \\&= P(\text{no\_heads} \cap (\text{die\_equals\_1} \cup \text{die\_equals\_2} \cup \dots \cup \text{die\_equals\_6})) \\&= P((\text{no\_heads} \cap \text{die\_equals\_1}) \cup (\text{no\_heads} \cap \text{die\_equals\_2}) \cup \dots \cup (\text{no\_heads} \cap \text{die\_equals\_6})) \\&= \sum_{i=1}^6 P(\text{no\_heads} \cap \text{die\_equals\_i}) = \sum_{i=1}^6 P(\text{no\_heads} | \text{die\_equals\_i}) P(\text{die\_equals\_i}) \\&= P(T)P(\text{die\_equals\_1}) + P(TT)P(\text{die\_equals\_2}) + \dots + P(TTTTTT)P(\text{die\_equals\_6}) \\&= \frac{1}{2} \frac{1}{6} + \left(\frac{1}{2}\right)^2 \frac{1}{6} + \dots + \left(\frac{1}{2}\right)^6 \frac{1}{6} = 0.1640625\end{aligned}$$

# Simulated solution of example p86

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

**Law of conditional probability**

## Statistical inference

Bayesian inference – discrete case

## References

Please see code [here](#).

# Law of conditional probability

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

### Rule 6

**Law of conditional probability** *Let  $A$  be an event that can only occur if one of the mutually exclusive events  $B_1, \dots, B_n$  occurs. Then*

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

# Exercise: proof of law of conditional probability

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

**Law of conditional probability**

## Statistical inference

Bayesian inference – discrete case

## References

Prove the law of conditional probability. Hint: use induction as in the proof of rule 5.

# Exercise: tour de France

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

### Example 8.6

*The upcoming Tour de France bicycle tournament will take place from July 1 through July 23. One hundred and eighty cyclists will participate in event. What is the probability that two or more participating cyclists will have birthdays on the same day during the tournament?*



# Contents

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

1 Foundations of probability theory

2 Conditional probability

3 Statistical inference

# Bayesian inference

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- In Bayesian statistics parameters of a model,  $\theta$  are considered random quantities. They are assigned a (subjective) probability  $P(\theta)$ .
- The probability of parameters is subject to change as new data arrives,  $P(\theta|D)$ .

$$P(\theta) \xrightarrow{\text{data observed}} P(\theta|D)$$

- How to calculate the posterior probability  $P(\theta|D)$ ?

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

$$P(\theta \cap D) = P(D|\theta)P(\theta)$$

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

# Bayesian inference

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### • Sequential update

$$\begin{aligned}P(\theta|D_1 \cap D_2) &= \frac{P(\theta \cap D_1 \cap D_2)}{P(D_1 \cap D_2)} \\&= \frac{P(D_2|D_1 \cap \theta)P(D_1 \cap \theta)}{P(D_1 \cap D_2)} \\&= \frac{P(D_2|\theta)P(D_1|\theta)P(\theta)}{P(D_1 \cap D_2)} \\&= K P(D_2|\theta)P(\theta|D_1) \propto P(D_2|\theta)P(\theta|D_1)\end{aligned}$$

Notes:

- 1 the first and second equalities use the **definition of conditional probability**.
- 2 the third equality assumes that the datasets  $D_1$  and  $D_2$  are **conditionally independent** given  $\theta$ .
- 3 the last equality groups all factors that don't depend on  $\theta$  in the proportionality constant  $K$ .

# Exercise: sequential posterior probability for arbitrary number of observations

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

For any  $n \in \mathbb{N}$ , prove that

$$P(\theta|D_1 \cap \dots \cap D_{n+1}) = P(D_{n+1}|\theta)P(\theta|D_1 \cap D_2 \cap \dots \cap D_n)$$

# Example 8.13: disease treatment effect

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

### Example 8.13

*A new treatment is tried out for a disease. A standard treatment has a success probability of 35%. The discrete uniform distribution on  $0.00, 0.01, \dots, 0.99, 1.00$  is taken as prior for the success probability of the new treatment. The experiment design is to make exactly ten observations by treating ten patients. The experimental study yield seven successes and three failures (Berry, 2006).*

- a** *Plot the prior of the success probability of the new treatment.*
- b** *Plot the likelihood of the success probability of the new treatment.*
- c** *Plot the posterior of the success probability of the new treatment.*
- d** *What is the posterior probability that the new treatment is more effective than the standard one?*

# Analytical solution of example 8.13

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

 prior:

$$p(\theta) = \frac{1}{101}, \quad \text{for } \theta = 0.00, 0.01, \dots, 0.99, 1.00$$

# Analytical solution of example 8.13

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

**b** likelihood:

$$\Omega = \{(r_1, \dots, r_{10}) : r_i \in \{S, F\}\}$$

Define the events  $k\_successes$  = “experiments where  $k$  out of the 10 observations were successes”, with  $k \in \{0, 1, \dots, 10\}$ .

$$0\_successes = \{(F, F, F, F, F, F, F, F, F, F)\}$$

$$1\_successes = \{(S, F, F, F, F, F, F, F, F, F), (F, S, F, F, F, F, F, F, F, F), \dots, (F, F, F, F, F, F, F, F, F, S)\}$$

$$2\_successes = \{(S, S, F, F, F, F, F, F, F, F), (S, F, S, F, F, F, F, F, F, F), \dots, (F, F, F, F, F, F, F, F, S, S)\}$$

Calculate the events probabilities

$$P(0\_successes|\theta) = P(\{(F, F, F, F, F, F, F, F, F, F)\}) = (1 - \theta)^{10}$$

$$\begin{aligned} P(1\_successes|\theta) &= P(\{(S, F, F, F, F, F, F, F, F, F), (F, S, F, F, F, F, F, F, F, F), \dots, \\ &\quad (F, F, F, F, F, F, F, F, F, S)\}) \\ &= P(\{(S, F, F, F, F, F, F, F, F, F)\}) + P(\{(F, S, F, F, F, F, F, F, F, F)\}) + \dots + \\ &\quad P(\{(F, F, F, F, F, F, F, F, F, S)\}) = 10P(\{(S, F, F, F, F, F, F, F, F, F)\}) \\ &= 10 \theta (1 - \theta)^9 \end{aligned}$$

# Analytical solution of example 8.13

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

**b** likelihood:

$$\begin{aligned}P(2\_successes|\theta) &= P(\{ (S, S, F, F, F, F, F, F, F, F), (S, F, S, F, F, F, F, F, F, F), \dots, \\&\quad (F, F, F, F, F, F, F, F, S, S) \}) \\&= P(\{ (S, S, F, F, F, F, F, F, F, F) \}) + \\&\quad P(\{ (S, F, S, F, F, F, F, F, F, F) \}) + \dots + \\&\quad P(\{ (F, F, F, F, F, F, F, F, S, S) \}) \\&= \binom{10}{2} P(\{(S, F, F, F, F, F, F, F, F, F)\}) \\&= \binom{10}{2} \theta^2 (1 - \theta)^8 \\P(k\_successes|\theta) &= \binom{10}{k} \theta^k (1 - \theta)^{10-k}, \quad k \in \{0, 1, \dots, 10\}\end{aligned}$$



# Analytical solution of example 8.13

## Foundations of probability theory

Historical notes  
Axiomatic definition of probability  
Some basic rules

## Conditional probability

Conditional probability  
Assigning probabilities by conditioning  
Independent events  
Conditionally independent events  
Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

Ⓒ posterior:

$$P(\theta|7\_successes) = K_1 P(7\_successes|\theta)P(\theta) = K_1 \binom{10}{7} \theta^7 (1-\theta)^3 \frac{1}{101} = K_2 \theta^7 (1-\theta)^3$$

$$1 = \sum_{k=0}^{100} P(\theta_k|7\_successes) = K_2 \sum_{k=0}^{100} \theta_k^7 (1-\theta_k)^3; \quad \theta_k = \frac{k}{100}$$

$$K_2 = \frac{1}{\sum_{k=0}^{100} \theta_k^7 (1-\theta_k)^3}$$

$$P(\theta|7\_successes) = \frac{\theta^7 (1-\theta)^3}{\sum_{k=0}^{100} \theta_k^7 (1-\theta_k)^3}$$

Ⓓ

$$P(\text{new treatment more effective than standard one}) = \sum_{k=36}^{100} P(\theta_k|7\_successes)$$

# Exercise: sequential posterior probability

## Foundations of probability theory

- Historical notes
- Axiomatic definition of probability
- Some basic rules

## Conditional probability

- Conditional probability
- Assigning probabilities by conditioning
- Independent events
- Conditionally independent events
- Law of conditional probability

## Statistical inference

- Bayesian inference – discrete case

## References

Continuing example 8.13, calculate the posterior probability of the new treatment after a second observation with six successes and a first observation with seven successes.

# July 19: online Bayesian linear regression

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

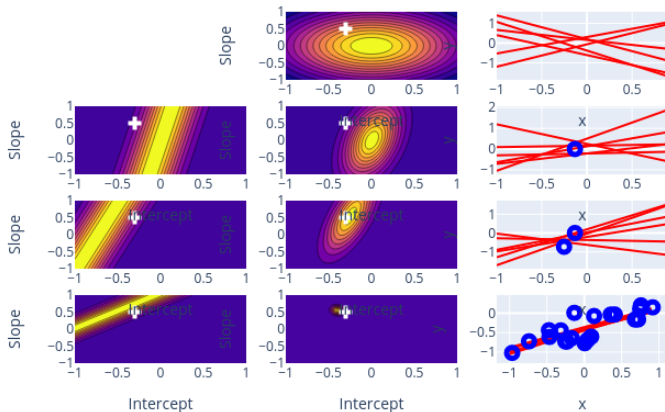
Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References



# References

## Foundations of probability theory

Historical notes

Axiomatic definition of probability

Some basic rules

## Conditional probability

Conditional probability

Assigning probabilities by conditioning

Independent events

Conditionally independent events

Law of conditional probability

## Statistical inference

Bayesian inference – discrete case

## References

- Berry, D. A. (2006). Bayesian clinical trials. *Nature reviews Drug discovery*, 5(1):27–36.
- Tijms, H. (2012). *Understanding probability*. Cambridge University Press.