Review Probability Calculus
Conditional Probability Independence

MATH/STAT395: Probability II

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó
Section 1, 2

Vincent Roulet

Department of Statistics

UNIVERSITY *of* WASHINGTON

Outline

Axioms of probability

Definition

Let Ω be the sample space, and $\mathcal{F}$ be the set of all possible events[a].
The **probability measure** (also called **probability distribution** or simply **probabillity**) is a function from $\mathcal{F}$ into the real numbers such that

Axioms of probability

Definition

Let Ω be the sample space, and $\mathcal{F}$ be the set of all possible events[a]. The **probability measure** (also called **probability distribution** or simply **probabillity**) is a function from $\mathcal{F}$ into the real numbers such that

1. $0 \leq \mathbb{P}(A) \leq 1$ for any event A

Axioms of probability

Definition

Let Ω be the sample space, and $\mathcal{F}$ be the set of all possible events[a]. The **probability measure** (also called **probability distribution** or simply **probabillity**) is a function from $\mathcal{F}$ into the real numbers such that

1. $0 \leq \mathbb{P}(A) \leq 1$ for any event A
2. $\mathbb{P}(\Omega) = 1$

Axioms of probability

Definition

Let Ω be the sample space, and $\mathcal{F}$ be the set of all possible events[a].
The **probability measure** (also called **probability distribution** or simply **probabillity**) is a function from $\mathcal{F}$ into the real numbers such that

1. $0 \leq \mathbb{P}(A) \leq 1$ for any event A
2. $\mathbb{P}(\Omega) = 1$
3. For $A_1, A_2, \ldots$ any sequence of (pairwise) disjoint events ($A_i \cap A_j = \emptyset$ for $i \neq j$),

$$\boxed{\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).} \qquad \text{(Countable additivity)}$$

[a] If Ω is discrete, $\mathcal{F} = 2^{\Omega}$, if $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ (the Borel Algebra on $\mathbb{R}$)

Axioms of probability

Definition

Let Ω be the sample space, and $\mathcal{F}$ be the set of all possible events[a].
The **probability measure** (also called **probability distribution** or simply **probabillity**) is a function from $\mathcal{F}$ into the real numbers such that

1. $0 \leq \mathbb{P}(A) \leq 1$ for any event A
2. $\mathbb{P}(\Omega) = 1$
3. For $A_1, A_2, \ldots$ any sequence of (pairwise) disjoint events ($A_i \cap A_j = \emptyset$ for $i \neq j$),

$$\boxed{\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).} \qquad \text{(Countable additivity)}$$

[a] If Ω is discrete, $\mathcal{F} = 2^{\Omega}$, if $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ (the Borel Algebra on $\mathbb{R}$)

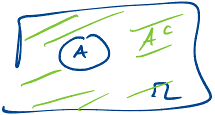
The triplet $(\underbrace{\Omega}_{\text{Sample space}}, \underbrace{\mathcal{F}}_{\text{Set of events}}, \underbrace{\mathbb{P}}_{\text{Probability measure}})$ is called a **probability space**.

Probability calculus

Probability calculus

- Finite additivity

Probability calculus



Probability calculus $A_i \cap A_j = \emptyset \quad \forall i, j \in \{1, \dots, n\}$

- Finite additivity
  - for $A_1, \dots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \dots \cup A_n) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)$

$> \mathbb{P}(A) = 1 - \mathbb{P}(A^c)$ $\quad A^c = \Omega \setminus A$

Probability calculus

Probability calculus

- Finite additivity
  - for $A_1, \ldots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$
- Compute via complement

Probability calculus

Probability calculus

- Finite additivity
  - for $A_1, \ldots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$
- Compute via complement
  - $P(A^c) = 1 - P(A)$ where $A^c = \Omega \setminus A$

> $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Probability calculus

Probability calculus

- Finite additivity
  - for $A_1, \ldots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$
- Compute via complement
  - $P(A^c) = 1 - P(A)$ where $A^c = \Omega \setminus A$
- Inclusion-Exclusion

Probability calculus

Probability calculus

- Finite additivity
  - for $A_1, \ldots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$
- Compute via complement
  - $P(A^c) = 1 - P(A)$ where $A^c = \Omega \setminus A$
- Inclusion-Exclusion
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probability calculus

Probability calculus

- Finite additivity
  - for $A_1, \ldots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$
- Compute via complement
  - $P(A^c) = 1 - P(A)$ where $A^c = \Omega \setminus A$
- Inclusion-Exclusion
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(A \cup B \cup C) =$ $P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Probability calculus

Probability calculus

- Finite additivity
  - for $A_1, \ldots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$
- Compute via complement
  - $P(A^c) = 1 - P(A)$ where $A^c = \Omega \setminus A$
- Inclusion-Exclusion
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(A \cup B \cup C) =$
    $P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Probability calculus

Probability calculus

- Finite additivity
  - for $A_1, \ldots, A_n$ **disjoints**, $\mathbb{P}(A_1 \cup \ldots \cup A_n) = \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$
- Compute via complement
  - $P(A^c) = 1 - P(A)$ where $A^c = \Omega \setminus A$
- Inclusion-Exclusion
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(A \cup B \cup C) =$
    $P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Note: All these properties are direct consequences of finite additivity

Exercise

$$\left\{ \overset{2}{\underset{T}{H}} \ \overset{2}{\underset{T}{H}} \ \cdots \ \overset{2}{} \right\}$$

> **Exercise**
>
> I flip a fair coin 10 times, what is the probability that I get at least one heads?

$A = \{ \text{at least one heads} \}$ $A^c = \{ \text{no heads in 10 flips} \}$

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

$$= 1 - \frac{1}{2^{10}}$$

Exercise

Exercise

I flip a fair coin 10 times, what is the probability that I get at least one heads?

Solution Use the complement:

$$\mathbb{P}(\text{at least one heads}) = 1 - \mathbb{P}(\text{no heads}) = 1 - (1/2)^{10} \approx 0.999$$

Classic probability spaces

Equally likely outcomes $\rightarrow \mathbb{P}(\omega_i) = c$

$\sum_{i=1}^{N} \mathbb{P}(\omega_i) = 1$

$\rightarrow c = \frac{1}{N}$

- Consider Ω is finite and consists of N elements,

$$\Omega = \{\omega_1, \ldots, \omega_N\}$$

- A proba dist. $\mathbb{P}$ is equally likely if $\underline{\mathbb{P}(\omega_i) = 1/N}$ for all $i \in \{1, \ldots, N\}$
- As a consequence, for $E \subseteq \Omega$, $\mathbb{P}(E) = |E|/|\Omega|$ with $|E|$ the number of elements in E.

$E = \{\omega_3, \omega_5, \omega_7\}$

$\mathbb{P}(E) = \mathbb{P}(\{\omega_3\} \cup \{\omega_5\} \cup \{\omega_7\})$

$= \sum \mathbb{P}(\omega_i)$

$= \frac{|E|}{N} = \frac{|E|}{|\Omega|}$

Classic probability spaces

Equally likely outcomes

- Consider Ω is finite and consists of N elements,

$$\Omega = \{\omega_1, \dots, \omega_N\}$$

- A proba dist. $\mathbb{P}$ is equally likely if $\mathbb{P}(\omega_i) = 1/N$ for all $i \in \{1, \dots, N\}$
- As a consequence, for $E \subseteq \Omega$, $\mathbb{P}(E) = |E|/|\Omega|$ with $|E|$ the number of elements in E.

Examples:

- Toss of 2 *fair* coins, $\Omega = \{HH, TT, HT, TH\}$

Classic probability spaces

Equally likely outcomes

- Consider Ω is finite and consists of N elements,

$$\Omega = \{\omega_1, \dots, \omega_N\}$$

- A proba dist. $\mathbb{P}$ is equally likely if $\mathbb{P}(\omega_i) = 1/N$ for all $i \in \{1, \dots, N\}$
- As a consequence, for $E \subseteq \Omega$, $\mathbb{P}(E) = |E|/|\Omega|$ with $|E|$ the number of elements in E.

Examples:

- Toss of 2 *fair* coins, $\Omega = \{HH, TT, HT, TH\}$
- Roll of a *fair* die $\Omega = \{1, \dots, 6\}$

Classic probability spaces

$$\omega = (\underbrace{\underset{N}{3}, \underset{\times N}{2}, \underset{\times N}{1}, \underset{\times N}{3}}_{k})$$

Definition (Ordered sampling **with** replacement)

Consider an urn with $N \geq 2$ balls labeled as $\{1, 2, \ldots, N\}$. Pick one ball without looking, note its label, and put it back. Repeat this k times. The outcome is a k-tuple $(a_1, a_2, \ldots, a_k)$.

$$\Omega = \{\text{all } k\text{-tuples of } 1, \ldots, N\}, \quad |\Omega| = N^k.$$

Classic probability spaces

Definition (Ordered sampling **with** replacement)

Consider an urn with $N \geq 2$ balls labeled as $\{1, 2, \ldots, N\}$. Pick one ball without looking, note its label, and put it back. Repeat this k times. The outcome is a k-tuple $(a_1, a_2, \ldots, a_k)$.

$$\Omega = \{\text{all } k\text{-tuples of } 1, \ldots, N\}, \quad |\Omega| = N^k.$$

Example

Suppose our urn contains 5 balls labeled 1, 2, 3, 4, 5. Sample 3 balls with replacement and produce an ordered list of the numbers drawn.

$$\Omega = \{1, 2, 3, 4, 5\}^3 = \{(s_1, s_2, s_3) : \text{ each } s_i \in \{1, 2, 3, 4, 5\}\}, \qquad |\Omega| = 5^3 = 125$$

$$\mathbb{P}(\text{the sample is } (2, 1, 5)) = \mathbb{P}(\text{the sample is } (2, 2, 3)) = 1/125$$

Classic probability spaces

Definition (Ordered sampling **with** replacement)

Consider an urn with $N \geq 2$ balls labeled as $\{1, 2, \ldots, N\}$. Pick one ball without looking, note its label, and put it back. Repeat this k times. The outcome is a k-tuple $(a_1, a_2, \ldots, a_k)$.

$$\Omega = \{\text{all } k\text{-tuples of } 1, \ldots, N\}, \quad |\Omega| = N^k.$$

Example

Suppose our urn contains 5 balls labeled 1, 2, 3, 4, 5. Sample 3 balls with replacement and produce an ordered list of the numbers drawn.

$$\Omega = \{1, 2, 3, 4, 5\}^3 = \{(s_1, s_2, s_3) : \text{ each } s_i \in \{1, 2, 3, 4, 5\}\}, \qquad |\Omega| = 5^3 = 125$$

$$\mathbb{P}(\text{the sample is } (2, 1, 5)) = \mathbb{P}(\text{the sample is } (2, 2, 3)) = 1/125$$

Example

Repeated flips of a coin or rolls of a die are also sampling with replacements from the set $\{H, T\}$ or$\{1, 2, 3, 4, 5, 6\}$.

Classical probability spaces

Definition (Ordered sampling **without** replacement)

Pick one ball without looking, note its label, but don't put it back. Repeat this k times. The outcome is a k-tuple without repeats, i.e.,

$$\Omega = \{\text{all } k\text{-arrangements of } 1, \ldots, N\}, \quad \#\Omega = (N)_k.$$

Classical probability spaces

$(\underset{N}{\_},\ \underset{N-1}{\_},\ \underset{N-2}{\_},\ \_)$

Definition (Ordered sampling **without** replacement)

Pick one ball without looking, note its label, but don't put it back. Repeat this k times. The outcome is a k-tuple without repeats, i.e.,

$$\Omega = \{\text{all } k\text{-arrangements of } 1, \dots, N\}, \quad \#\Omega = (N)_k.$$

Example

$(2, 3, 5)$ $\quad$ $\cancel{(3, 2, 3)}$

Consider again the urn with 5 balls labeled 1, 2, 3, 4, 5. Sample 3 balls without replacement and produce an ordered list of the numbers drawn.

$$\Omega = \{(s_1, s_2, s_3) : \text{ each } s_i \in \{1,2,3,4,5\} \text{ and } s_1, s_2, s_3 \text{ are all distinct}\},$$

We have 5 choices for the first, 4 choices for the second, 3 choices for the third so $\#\Omega = 5 \cdot 4 \cdot 3 = (5)_3$ and

$$\mathbb{P}(\text{the sample is } (2,1,5)) = \frac{1}{5 \cdot 4 \cdot 3} = \frac{1}{60} \qquad \mathbb{P}(\text{the sample is } 2,2,3) = 0$$

Classical probability spaces

Definition (Unordered sampling **without** replacement)

Pick one ball without looking, note its label, but don't put it back. Repeat this k times. This time do not consider the order of the balls you drew but only their labels (which ball comes first does not matter).
Namely, consider the outcome to be a k-subset

$$\Omega = \{\text{all } k\text{-subset of } 1, \ldots, N\}, \quad \#\Omega = \binom{N}{k} = \frac{N!}{k!\,(N-k)!}$$

Classical probability spaces

Definition (Unordered sampling **without** replacement)

Pick one ball without looking, note its label, but don't put it back. Repeat this k times. This time do not consider the order of the balls you drew but only their labels (which ball comes first does not matter).
Namely, consider the outcome to be a k-subset

$$\Omega = \{\text{all } k\text{-subset of } 1, \dots, N\}, \quad \#\Omega = \binom{N}{k}.$$

Example

Same as before, an urn, 5 balls labeled 1, 2, 3, 4, 5. Sample 3 balls without replacement an produce a set of 3 balls (unordered)

$$\Omega = \{\omega : \omega \text{ is a subset of size 3 from } \{1,2,3,4,5\}\}$$

$$\mathbb{P}(\text{the sample is } \{1,2,5\}) = \frac{1}{\binom{5}{3}} = \frac{2!3!}{5!} = \frac{1}{10}$$

Exercises

Exercise

At a political meeting there are 6 liberals and 5 conservatives. We choose 5 people uniformly at random to form a committee. What is the probability that there are more conservatives than liberals in the committee?

A

- $B_i := \{$ there are i conservatives in the comitee $\}$

$$A = B_5 \cup B_4 \cup B_3$$

$$\mathbb{P}(A) = \mathbb{P}(B_5) + \mathbb{P}(B_4) + \mathbb{P}(B_3)$$

$$= \frac{\binom{5}{5}\binom{6}{0}}{\binom{11}{5}} + \frac{\binom{5}{4}\binom{6}{1}}{\binom{11}{5}} + \dots$$

$$= \dots$$

Exercises

Exercise

At a political meeting there are 6 liberals and 5 conservatives. We choose 5 people uniformly at random to form a committee. What is the probability that there are more conservatives than liberals in the committee?

Solution

- Denote $A_i = \{\text{there are i liberals in the comittee}\}$ so that $A = A_0 \cup A_1 \cup A_2$

Exercises

Exercise

At a political meeting there are 6 liberals and 5 conservatives. We choose 5 people uniformly at random to form a committee. What is the probability that there are more conservatives than liberals in the committee?

Solution

- Denote $A_i = \{\text{there are i liberals in the comittee}\}$ so that $A = A_0 \cup A_1 \cup A_2$
- We are doing sampling without replacement,

$$\mathbb{P}(A_i) = \frac{\binom{6}{i}\binom{5}{5-i}}{\binom{11}{5}}$$

Exercises

Exercise

At a political meeting there are 6 liberals and 5 conservatives. We choose 5 people uniformly at random to form a committee. What is the probability that there are more conservatives than liberals in the committee?

Solution

- Denote $A_i = \{\text{there are i liberals in the comittee}\}$ so that $A = A_0 \cup A_1 \cup A_2$
- We are doing sampling without replacement,

$$\mathbb{P}(A_i) = \frac{\binom{6}{i}\binom{5}{5-i}}{\binom{11}{5}}$$

- Since the events A_i are disjoint

$$\mathbb{P}(A) = \mathbb{P}(A_0) + \mathbb{P}(A_1) + \mathbb{P}(A_2) = \frac{1 \cdot 1 + 6 \cdot 5 + 15 \cdot 10}{462} \approx 0.39$$

Exercises

Exercise

Imagine a game of three players where exactly one player wins in the end and all players have equal chances of being the winner. The game is repeated four times. Find the probability that there is at least one person who wins no games.

- $A_i = \{\text{player } i \text{ wins no game}\}$
- $\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)$
$- \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_3)$
$+ \underbrace{\mathbb{P}(A_1 \cap A_2 \cap A_3)}_{=0}$

$\mathbb{P}(A_i) = \left(\frac{2}{3}\right)^4$

$\mathbb{P}(A_i \cap A_j) = \left(\frac{1}{3}\right)^4$

Exercises

Exercise

Imagine a game of three players where exactly one player wins in the end and all players have equal chances of being the winner. The game is repeated four times. Find the probability that there is at least one person who wins no games.

Solution

- By inclusion/exclusion formula, denoting $A_i = \{\text{player i wins no game}\}$

$$\mathbb{P}(A_1 \cup A_2 \cup_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_3 \cap A_1) + \mathbb{P}(A_1 \cap A_2 \cap A_3$$

Exercises

Exercise

Imagine a game of three players where exactly one player wins in the end and all players have equal chances of being the winner. The game is repeated four times. Find the probability that there is at least one person who wins no games.

Solution

- By inclusion/exclusion formula, denoting $A_i = \{\text{player i wins no game}\}$
  $\mathbb{P}(A_1 \cup A_2 \cup_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_3 \cap A_1) + \mathbb{P}(A_1 \cap A_2 \cap A_3$
- Clearly $\mathbb{P}(A_i) = (2/3)^4$

Exercises

Exercise

Imagine a game of three players where exactly one player wins in the end and all players have equal chances of being the winner. The game is repeated four times. Find the probability that there is at least one person who wins no games.

Solution

- By inclusion/exclusion formula, denoting $A_i = \{\text{player i wins no game}\}$

  $\mathbb{P}(A_1 \cup A_2 \cup_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_3 \cap A_1) + \mathbb{P}(A_1 \cap A_2 \cap A_3$
- Clearly $\mathbb{P}(A_i) = (2/3)^4$
- $A_1 \cap A_2$ means that player 3 won all games, so

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_3) = (1/3)^4$$

Exercises

Exercise

Imagine a game of three players where exactly one player wins in the end and all players have equal chances of being the winner. The game is repeated four times. Find the probability that there is at least one person who wins no games.

Solution

- By inclusion/exclusion formula, denoting $A_i = \{\text{player i wins no game}\}$
  $\mathbb{P}(A_1\cup A_2\cup_3) = \mathbb{P}(A_1)+\mathbb{P}(A_2)+\mathbb{P}(A_3)-\mathbb{P}(A_1\cap A_2)-\mathbb{P}(A_2\cap A_3)-\mathbb{P}(A_3\cap A_1)+\mathbb{P}(A_1\cap A_2\cap A_3$
- Clearly $\mathbb{P}(A_i) = (2/3)^4$
- $A_1 \cap A_2$ means that player 3 won all games, so
  $$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_3) = (1/3)^4$$
- Finally $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 0$

Exercises

Exercise

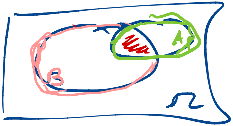
Imagine a game of three players where exactly one player wins in the end and all players have equal chances of being the winner. The game is repeated four times. Find the probability that there is at least one person who wins no games.

Solution

- By inclusion/exclusion formula, denoting $A_i = \{\text{player i wins no game}\}$
  $\mathbb{P}(A_1 \cup A_2 \cup_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_3 \cap A_1) + \mathbb{P}(A_1 \cap A_2 \cap A_3$
- Clearly $\mathbb{P}(A_i) = (2/3)^4$
- $A_1 \cap A_2$ means that player 3 won all games, so
  $$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_1 \cap A_3) = (1/3)^4$$
- Finally $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 0$
- So we get
  $$\mathbb{P}(A_1 \cup A_2 \cup_3) = 3 \cdot (2/3)^4 - 3 \cdot (1/3)^4 = 5/9$$

Outline

Conditional probability



Conditional probability

- Probability of A given B (with $\mathbb{P}(B) \neq 0$):

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

Conditional probability

Conditional probability

- Probability of A given B (with $\mathbb{P}(B) \neq 0$):

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

- **Multiplication rule**

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1, A_2) \cdots \mathbb{P}(A_n \mid A_1, \ldots, A_{n-1})$$

Conditional probability

Conditional probability

- Probability of A given B (with $\mathbb{P}(B) \neq 0$):

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

- **Multiplication rule**

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1, A_2) \cdots \mathbb{P}(A_n \mid A_1, \dots, A_{n-1})$$

- **Idea:**
  By conditioning on an event it is as if you were zooming into this world of possibility

Conditional Probability

Law of total Probability

- (Simple case)

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

Conditional Probability

Law of total Probability

- (Simple case)

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

- (General case)
  Let $B_1, \ldots, B_n$ be a partition of Ω with $\mathbb{P}(B_i) > 0$ for all i, for any event A,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

Conditional Probability

Law of total Probability

- (Simple case)

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$$

- (General case)
  Let $B_1, \ldots, B_n$ be a partition of Ω with $\mathbb{P}(B_i) > 0$ for all i, for any event A,

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

- **Idea:**
  Use law of total probability to decompose complex events (such as multiple stages experiments)

Bayes Formula

Bayes Formula

- (Simple case)
  For A, B, B^c such that $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(B^c) > 0$,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)}$$

Bayes Formula

Bayes Formula

- (Simple case)
  For A, B, B^c such that $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(B^c) > 0$,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)}$$

- (General case)
  Let $B_1, \dots, B_n$ be a partition of Ω with $\mathbb{P}(B_i) > 0$.
  For any event A, with $\mathbb{P}(A) > 0$, and any $k \in \{1, \dots, n\}$

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}$$

Bayes Formula

Bayes Formula

- (Simple case)
  For A, B, B^c such that $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(B^c) > 0$,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)}$$

- (General case)
  Let $B_1, \ldots, B_n$ be a partition of Ω with $\mathbb{P}(B_i) > 0$.
  For any event A, with $\mathbb{P}(A) > 0$, and any $k \in \{1, \ldots, n\}$

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}$$

- **Idea:**
  Get probability of one hypothesis given an observation

Exercises

Exercise

We pick one word uniformly at random in the following sentence and one letter uniformly at random from the word picked.

`MY PILLOW IS PURPLE`

What is the probability that the letter is P?

Exercises

Exercise

We pick one word uniformly at random in the following sentence and one letter uniformly at random from the word picked.

`MY PILLOW IS PURPLE`

What is the probability that the letter is P?

Solution

- Let $A = \{\text{the letter is P}\}$, $B_i = \{\text{the ith word is picked}\}$
- By the law of total probability,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A \mid B_1)\mathbb{P}(B_1) + \mathbb{P}(A \mid B_2)\mathbb{P}(B_2) + \mathbb{P}(A \mid B_3)\mathbb{P}(B_3) + \mathbb{P}(A \mid B_4)\mathbb{P}(B_4) \\ &= 0 \cdot \frac{1}{4} + \frac{1}{6} \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + \frac{2}{6} \cdot \frac{1}{4} = \frac{1}{8}\end{aligned}$$

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$

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

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

  which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
- Suppose A and B are independent. Then

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

  which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
- Suppose A and B are independent. Then
  1. A and B^c are independent

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

  which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$

- Suppose A and B are independent. Then
  1. A and B^c are independent
  2. A^c and B are independent

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

  which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
- Suppose A and B are independent. Then
  1. A and B^c are independent
  2. A^c and B are independent
  3. A^c and B^c are independent

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

  which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
- Suppose A and B are independent. Then
  1. A and B^c are independent
  2. A^c and B are independent
  3. A^c and B^c are independent
- **Idea:**
  Independence is used to simplify some computations

$A \cap B = \emptyset$

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

  which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
- Suppose A and B are independent. Then
  1. A and B^c are independent
  2. A^c and B are independent
  3. A^c and B^c are independent
- **Idea:**
  Independence is used to simplify some computations

Independence

Independence

- Two events A, B are independent if
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$
which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
- Suppose A and B are independent. Then
  1. A and B^c are independent
  2. A^c and B are independent
  3. A^c and B^c are independent
- **Idea:**
Independence is used to simplify some computations

Be careful
Do not confound **disjoint events** and **independent events**.

Independence

Independence

- Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

which implies that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ and $\mathbb{P}(B \mid A) = \mathbb{P}(B)$

- Suppose A and B are independent. Then
  1. A and B^c are independent
  2. A^c and B are independent
  3. A^c and B^c are independent

- **Idea:**
Independence is used to simplify some computations

Be careful
Do not confound **disjoint events** and **independent events**.

- For example A and A^c are disjoint,
but $0 = \mathbb{P}(A \cap A^c) \neq \mathbb{P}(A)\mathbb{P}(A^c) > 0$ for any A s.t. $0 < \mathbb{P}(A) < 1$.

Examples

Examples

- Successive flips of a coin or rolls of a die are usually assumed to be independent

Examples

Examples

- Successive flips of a coin or rolls of a die are usually assumed to be independent
- Sampling with replacement in a an urn with n red balls, p blue balls and considering

$$A = \{\text{first ball is red}\} \quad B = \{\text{second ball is blue}\}$$

then A, B are independent

Examples

Examples

- Successive flips of a coin or rolls of a die are usually assumed to be independent
- Sampling with replacement in a an urn with n red balls, p blue balls and considering

$$A = \{\text{first ball is red}\} \quad B = \{\text{second ball is blue}\}$$

  then A, B are independent
- Same setting but using sampling without replacement, we get that A, B are not independent

Examples

Examples

- Successive flips of a coin or rolls of a die are usually assumed to be independent
- Sampling with replacement in a an urn with n red balls, p blue balls and considering

$$A = \{\text{first ball is red}\} \quad B = \{\text{second ball is blue}\}$$

  then A, B are independent
- Same setting but using sampling without replacement, we get that A, B are not independent
- See Example 2.19 in the book for example

Independence

Mutual Independence

Events $A_1, \ldots, A_n$ are **independent** (or **mutually independent)** if for any collection $A_{i_1}, \ldots, A_{i_k}$ with $2 \leq k \leq n$, $1 \leq i_1 < \ldots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k})$$

Independence

Mutual Independence
Events $A_1, \ldots, A_n$ are **independent** (or **mutually independent)** if for any collection $A_{i_1}, \ldots, A_{i_k}$ with $2 \leq k \leq n$, $1 \leq i_1 < \ldots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k})$$

Conditional Independence
Let $B \subseteq \Omega$ s.t. $\mathbb{P}(B) > 0$, events A_1, A_2 are **conditionally independent given** B if

$$\mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

Quiz next lecture

Quiz

Two factories I and II produce phones for brand ABC. Factory I produces 60% of all ABC phones, and factory II produces 40%. 10% of phones produced by factory I are defective, and 20% of those produced by factory II are defective. You know that the store where you buy your phones is supplied by one of the factories, but you do not know which one. You buy two phones, and both are defective. What is the probability that the store is supplied by factory II?

Hint: Make an appropriate assumption of conditional independence.

Outline

Finer points

- In some lectures, there will be additional slides for some technical details if you are interested,
- The slides will be denoted by asterisks
- These slides won't be covered during class and won't be used for assignments/exams,
- However if you are interested in pursuing some math/probability courses, these are good concepts/proofs to know

σ-algebra*

Motivation

- The definition of probabilities require a special handling of the possible events that we consider.
- Formally the set of events must be a σ-algebra as defined below

Definition

A σ-algebra $\mathcal{F}$ on a set Ω is a collection of subsets of Ω satisfying

1. $\Omega \in \mathcal{F}, \quad \emptyset \in \mathcal{F}$
2. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
3. if $A_1, A_2, \ldots, \in \mathcal{F}$, then $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{F}$

Borel Algebra

- To manipulate probabilities on $\mathbb{R}$ we use the Borel algebra (special case of a σ-algebra defined through intervals)

Definition (Borel algebra*)

The Borel algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest[a] set of subset of $\mathbb{R}$ such that

1. $(-\infty, b]$ belongs to $\mathcal{B}$ for any $b \in \mathbb{R}$

[a] smallest in the sense of inclusion. Namely a set A is smaller than a set B if $A \subseteq B$

Note:

Borel Algebra

- To manipulate probabilities on $\mathbb{R}$ we use the Borel algebra (special case of a σ-algebra defined through intervals)

Definition (Borel algebra*)

The Borel algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest[a] set of subset of $\mathbb{R}$ such that

1. $(-\infty, b]$ belongs to $\mathcal{B}$ for any $b \in \mathbb{R}$
2. if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$

[a] smallest in the sense of inclusion. Namely a set A is smaller than a set B if $A \subseteq B$

Note:

Borel Algebra

- To manipulate probabilities on $\mathbb{R}$ we use the Borel algebra (special case of a σ-algebra defined through intervals)

Definition (Borel algebra*)

The Borel algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest[a] set of subset of $\mathbb{R}$ such that

1. $(-\infty, b]$ belongs to $\mathcal{B}$ for any $b \in \mathbb{R}$
2. if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$
3. if $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{B}$

[a] smallest in the sense of inclusion. Namely a set A is smaller than a set B if $A \subseteq B$

Note:

Borel Algebra

- To manipulate probabilities on $\mathbb{R}$ we use the Borel algebra (special case of a σ-algebra defined through intervals)

Definition (Borel algebra*)

The Borel algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest[a] set of subset of $\mathbb{R}$ such that

1. $(-\infty, b]$ belongs to $\mathcal{B}$ for any $b \in \mathbb{R}$
2. if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$
3. if $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{B}$

[a] smallest in the sense of inclusion. Namely a set A is smaller than a set B if $A \subseteq B$

Note:

- By 2., $(a, +\infty) \in \mathcal{B}$ for any $a \in \mathbb{R}$

Borel Algebra

- To manipulate probabilities on $\mathbb{R}$ we use the Borel algebra (special case of a σ-algebra defined through intervals)

Definition (Borel algebra*)

The Borel algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest[a] set of subset of $\mathbb{R}$ such that

1. $(-\infty, b]$ belongs to $\mathcal{B}$ for any $b \in \mathbb{R}$
2. if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$
3. if $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{B}$

[a] smallest in the sense of inclusion. Namely a set A is smaller than a set B if $A \subseteq B$

Note:

- By 2., $(a, +\infty) \in \mathcal{B}$ for any $a \in \mathbb{R}$
- By 3. and 2. using $(A \cup B)^c = A^c \cap B^c$ for any $A_1, A_2, \ldots \in \mathcal{B}$, $\bigcap_{i=1}^{+\infty} A_i \in \mathcal{B}$

Borel Algebra

- To manipulate probabilities on $\mathbb{R}$ we use the Borel algebra (special case of a σ-algebra defined through intervals)

Definition (Borel algebra*)

The Borel algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest[a] set of subset of $\mathbb{R}$ such that

1. $(-\infty, b]$ belongs to $\mathcal{B}$ for any $b \in \mathbb{R}$
2. if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$
3. if $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{B}$

[a] smallest in the sense of inclusion. Namely a set A is smaller than a set B if $A \subseteq B$

Note:

- By 2., $(a, +\infty) \in \mathcal{B}$ for any $a \in \mathbb{R}$
- By 3. and 2. using $(A \cup B)^c = A^c \cap B^c$ for any $A_1, A_2, \ldots \in \mathcal{B}$, $\bigcap_{i=1}^{+\infty} A_i \in \mathcal{B}$
- Therefore

$$(a, b] = (a, +\infty) \cap (-\infty, b] \in \mathcal{B} \text{ for any } a < b$$

Borel Algebra

- To manipulate probabilities on $\mathbb{R}$ we use the Borel algebra (special case of a σ-algebra defined through intervals)

Definition (Borel algebra*)

The Borel algebra $\mathcal{B}$ on $\mathbb{R}$ is the smallest[a] set of subset of $\mathbb{R}$ such that

1. $(-\infty, b]$ belongs to $\mathcal{B}$ for any $b \in \mathbb{R}$
2. if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$
3. if $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{B}$

[a] smallest in the sense of inclusion. Namely a set A is smaller than a set B if $A \subseteq B$

Note:

- By 2., $(a, +\infty) \in \mathcal{B}$ for any $a \in \mathbb{R}$
- By 3. and 2. using $(A \cup B)^c = A^c \cap B^c$ for any $A_1, A_2, \ldots \in \mathcal{B}$, $\bigcap_{i=1}^{+\infty} A_i \in \mathcal{B}$
- Therefore

$$(a, b] = (a, +\infty) \cap (-\infty, b] \in \mathcal{B} \text{ for any } a < b$$

- Since $\{a\} = \cap_{n=1}^{+\infty}(a - 1/n, a]$, $\{a\} \in \mathcal{B}$ and $[a, b] = (a, b] \cup \{a\} \in \mathcal{B}$

Probability space, rigorous definition

Definition

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three parts:

- A set Ω called the *sample space*, the set of all possible *outcomes* of a random action,
- A σ-algebra $\mathcal{F}$ called the *set of events*, where each *event* $E \in \mathcal{F}$ is a subset of Ω,
- A *probability measure* $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ that assigns probabilities to events and satisfies the **axioms of probability**
  1. For all $A \in \mathcal{F}$, $0 \leq \mathbb{P}(A) \leq 1$,
  2. $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$
  3. For any sequence $A_1, A_2, \ldots \in \mathcal{F}$ of *disjoint* sets,

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} \mathbb{P}(A_i)$$