50008

Probability and Statistics

Imperial College London

Contents

Chapter 1

introduction

Chapter 2

Elementary Probability Theory

Probability theory is a mathematical formalism to describe and quantify uncertainty.

Uses of probability include examples such as:

- Finding distribution of runtimes & memory usage for software.
- Response times for database queries.
- Failure rate of components in a datacenter.

2.1 Sample Spaces and Events

Sample Space **Definition 2.1.1**

The set of all possible outcomes of a random experiment. The set is usually denoted with set notation, and can be finite, countably or uncountably infinite.

For example:

Experiment	Sample Space
Coin Toss	$S = \{Heads, Tails\}$
6-Sided Dice Roll	$S = \{1, 2, 3, 4, 5, 6\}$
2 Coin Tosses	$S = \{(H, H), (H, T), (T, H), (T, T)\}$
Choice of Odd number	$S = \{x \in \mathbb{N} \mid \exists y \in \mathbb{N}.[2y + 1 = x]\}$

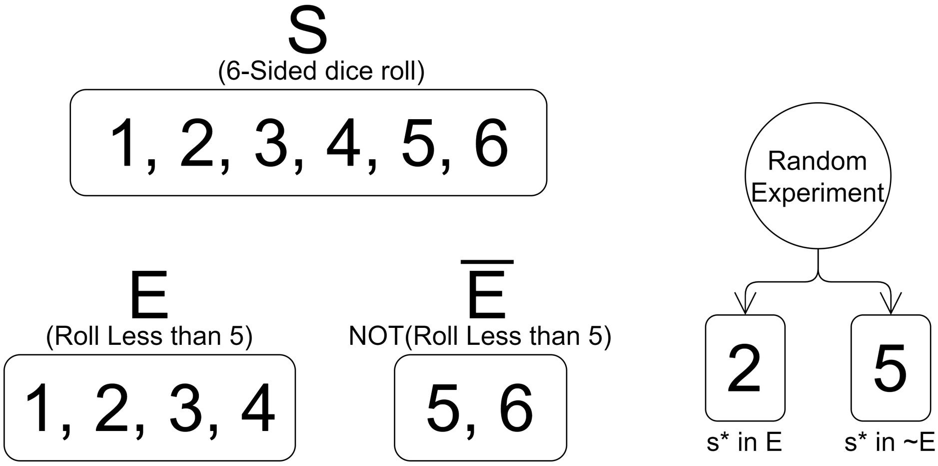
Event **Definition 2.1.2**

Any subset of the sample space $E \subseteq S$ (a set of possible outcomes).

- **null event**($\emptyset$) Empty event, can be used for impossible events.
- **universal event** (S) Event contains entire sample space and is therefore certain.
- **elementary events** Singleton subsets of the sample space (contain one element).

For example:

Event	Set of Event	Sample Space
6-Sided Dice Rolls 1	$E = \{1\}$	$S = \{1, 2, 3, 4, 5, 6\}$
6-Sided Dice Rolls Even	$E = \{2, 4, 6\}$	$S = \{1, 2, 3, 4, 5, 6\}$
6-Sided Dice Rolls 7	$E = \emptyset$	$S = \{1, 2, 3, 4, 5, 6\}$
2 Coin toss get 2 Tails	$E = \{(T, T)\}$	$S = \{(H, H), (H, T), (T, H), (T, T)\}$
Random Natural Number is 4	$E = \{4\}$	$S = \mathbb{N}$



- If we perform a random experiment with outcome $S* \in S$. If $s* \in E$, then event E has occurred.
- If E has not occurred ($s* \notin E$) then $s* \in \overline{E}$.
- The set $\{s*\}$ is an elementary event.
- Null event $\emptyset$ never occurs, the universal event S always occurs.

2.1.1 Set Operations on Events

- **Union / Or**

  $$\bigcup_i E_i = \{s \in S | \exists i.[s \in E_i]\}$$

  Occurs if at least one of the events E_i has occurred (has union of event sets).

  If 4 is rolled on a 6-sided dice, then union of (is 3) and (is 4) occurred.
- **Intersection / And**

  $$\bigcap_i E_i = \{s \in S | \forall i.[s \in E_i]\}$$

  Occurs if all the events E_i occur.

  If 4 is rolled on a 6-sided dice, the intersection of (is even) and (is 4) occurred.
- **Mutual Exlusion**

  $$E_1 \cap E_2 = \emptyset$$

  If sets are disjoint, then they are mutually exclusive (cannot occur simultaneously).

  For a 6-sided dice the events (is 4) and (is 6) are mutually exclusive.

2.1.2 Probability

When determining the probability of every subset $E \subseteq S$ occurring:

- S **is Finite** Can easily assign probabilites.
- S **is countable** Can assign probabilites.
- S **is uncountably infinite**
  Can initially assign some collection of subsets probabilities, but it then becomes impossible to define probabilities on reamining subsets.

  Cannot make probabilities sum to 1 with reasonably axioms.

For this reason when defining a probability function on sample space S, we must define the collection of subsets we will measure.

The subsets are referred to as $\mathcal{F}$ and must be:

1. nonempty ($S \in \mathcal{F}$)

2. closed under complements $E \in \mathcal{F} \Rightarrow \overline{E} \in \mathcal{F}$

3. closed under countable union $E_1, E_2, \dots \in \mathcal{F} \Rightarrow \bigcup_i E_i \in \mathcal{F}$

A collection of sets is known as σ-algebra.

Probability Measure **Definition 2.1.3**

A function $P : \mathcal{F} \to [0, 1]$ on the pair $(S, \mathcal{F})$ such that:

Axiom 1.	$\forall E \in \mathcal{F}.[0 \leq P(E) \leq 1]$
Axiom 2.	$P(S) = 1$
Axiom 3.	Countably additive, for **disjoint** sets $E_1, E_2, \dots \in \mathcal{F}$: $P(\bigcup_i E_i) = \sum_i P(E_i)$

$P(E)$ provides the probability (between 0 and 1 inclusive) that a given event occurs.

From the axioms satisfied by a *probability measure* we can derive that:

1. $P(\overline{E}) = 1 - P(E)$

2. $P(\emptyset) = 0$

3. For any events E_1 and E_2: $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

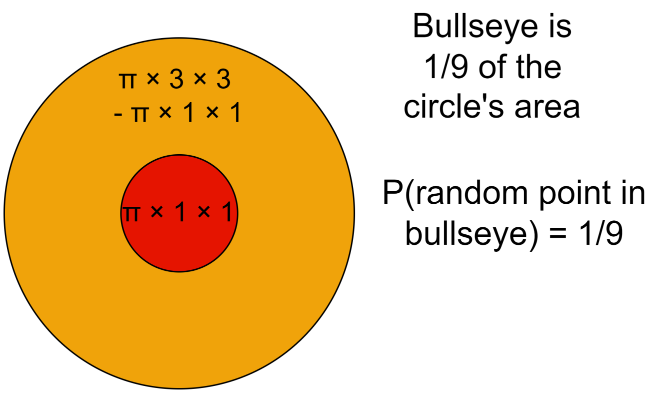
2.2 Interpretations of Probability

2.2.1 Classical Interpretation

Given S is finite and the *elementary events* are equally likely:

$$P(E) = \frac{|E|}{|S|}$$

We can also extend this *uniform probability distribution* to infinite spaces by considering measures such as area, mass or volume.



2.2.2 Frequentist Interpretation

Through repeated observations of identical random experiments in which E can occur, the proportion of experiments where E occurs tends towards the probability of E.

At an infinite number experiments, the proportion of occurrences of E is equal to $P(E)$.

Central Limit Theorem ***Extra Fun!* 2.2.1**

This can also be considered in terms of *central limit theorem*, where the greater the sample size taken from some distribution (with defined mean μ), the closer the mean of the sample to the distribution's mean. (more readings results in less variance in the sample means as they converge on the distribution's mean)

2.2.3 Subjective Interpretation

Probability is the degree of belief held by an individual.

For example if gambling: $\begin{array}{l} \text{Option 1: } E \text{ occurs win } \pounds 1,\ \overline{E} \text{ occurs win } \pounds 0 \\ \text{Option 2: Regardless of outcome get } \pounds P(E) \end{array}$.

Either outcome, the gambler receives $\pounds P(E)$. The value of $P(E)$ is the value for which the individual is indifferent about the choice between option 1 or 2. It is the *individuals probability* of event E occurring.

2.3 Joint Events and Conditional Probability

We commonly need to consider *Join Events* (where two events occur at the same time).

Independent Events **Definition 2.3.1**

Two events are independent if the occurence of one does not affect the other. Given E_1 and E_2 are independent:

$$E_1 \textit{ and } E_2 \textit{ independent} \Leftrightarrow P(E_1 \textit{ occurrs and } E_2 \textit{ occurs}) = P(E_1) \times P(E_2)$$

More generally, the set of events $\{E_1, E_2, \dots\}$ are independent if for any finite subset $\{E_{i_1}, E_{i_2}, \dots, E_{i_n}\}$:

$$p(\bigcap_{j=1}^{n} E_{i_j}) = \prod_{j=1}^{n} P(E_{i_j})$$

If E_1 and E_2 are independent, then so are $\overline{E_1}$ and E_2.

For example with a coin toss, subsequent coin tosses do not effect the next coin toss's probability of heads.

We can show that if E_1 and E_2 are independent, so are $\overline{E_1}$ and E_2:

(1)	$F = (E_1 \cap E_2) \cup (\overline{E_1} \cap E_2)$	By set operations
(2)	$P(E_2) = P(E_1 \cap E_2) + p(\overline{E_1} \cap E_2)$	As **1** was a disjoint union, Axiom 3
(3)	$P(\overline{E_1} \cap E_2) = P(E_2) - P(E_1 \cap E_2)$	
(4)	$P(\overline{E_1} \cap E_2) = P(E_2) - P(E_1) \times P(E_2)$	
(5)	$P(\overline{E_1} \cap E_2) = P(E_2) \times (1 - P(E_1)$	
(6)	$P(\overline{E_1} \cap E_2) = P(E_2) \times P(\overline{E_1})$	By $P(\overline{E}) = 1 - P(E)$

We can show that $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$:

(1)	$E_1 \cup E_2 = E_1 \cup (E_2 \cap \overline{E_1})$	From set theory
(2)	$P(E_1 \cup E_2) = P(E_1 \cup (E_2 \cap \overline{E_1}))$	By Axiom 3
(3)	$P(E_1 \cup E_2) = P(E_1) + P(E_2 \cap \overline{E_1})$	
(4)	$P(E_2 \cap \overline{E_1}) = P(E_2) - P(E_1 \cap E_2)$	By **3** of the previous proof and as E_1 and E_2 are independent

Dice for Money **Example Question 2.3.1**

We can construct a *Probability Table*:

		Dice						Totals
		1	2	3	4	5	6	
Coin	H	1/12	1/12	1/12	1/12	1/12	1/12	1/2
	T	1/12	1/12	1/12	1/12	1/12	1/12	1/2
Totals		1/6	1/6	1/6	1/6	1/6	1/6	

We can determine the probability of any event by summing the probabilities of elementary events represented by cells in the table.

$P(H)$ is called a *marginal probability*, as it the probability of one event occurring irrespective of the other (the dice in this case).

$P((H, 3))$ is called a *joint probability* as it involves both events (dice roll and the coin toss).

Roll of the Die — **Example Question 2.3.2**

A crooked die (called a top) has the same faces on either side.

We flip the coin, then if it is heads we use the normal die, else we use the top.

		Dice						Totals
		1	2	3	4	5	6	
Coin	H	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/2$
	T	$1/6$	0	$1/6$	0	$1/6$	0	$1/2$
Totals		$1/4$	$1/12$	$1/4$	$1/12$	$1/4$	$1/12$	

We can now see that $P(\{(H,3)\}) \neq P(\{H\}) \times P(\{3\})$ and hence they are dependent, as the dice roll depends on the coin toss.

2.4 Conditional Probability

For two events E and F in *sample space* S, where $P(F) \neq 0$:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Probability of E given F is the probability of both occurring over the probability of F.

Independence — ***Extra Fun!* 2.4.1**

If E and F are independent:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E) \times P(F)}{P(F)} = P(E)$$

Conditional Independence — **Definition 2.4.1**

$P(\bullet|F)$ defines a probability measure obeying the axioms of probability on set F (When have just reduced S to F).

Three events E_1, E_2, F are conditionally independent if and only if:

$$P(E_1 \cap E_2|F) = P(E_1|F) \times P(E_2|F)$$

W — **Example Question 2.4.1**

hat is the probability the dice rolls a 3 given the dice rolls an odd number?

$$P(\{3\}|\{1,3,5\}) = \frac{P(\{3\} \cap \{1,3,5\})}{P(\{1,3,5\})} = \frac{P(\{3\})}{P(\{1,3,5\})} = \frac{1/6}{1/2} = \frac{1}{3}$$

Go big or go home! — **Example Question 2.4.2**

Throw a die from each hand. What is the probability the die thrown from the left is larger than the die thrown from the right.

The sample space is:

$$S = \left\{ \begin{matrix} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{matrix} \right\}$$

We want the event such that the left value of the pair is larger.

For value 1 there are 0 possible, for 2 there is 1 and so on.

$$(1:0),(2:1),(3:2),(4:3),(5:4),(6:5)$$

Hence there are $0+1+2+3+4+5=15$ possible pairs with the left larger than the right.

$$P(E)=\frac{15}{36}=\frac{5}{12}$$

However if we know the left or right die, we can determine a new probability. For example if we know the left die is 4 then we know there are 6 pairs with the left as 4, and 3 of those pairs have a smaller right.

$$P(E|4)=\frac{3}{6}=\frac{1}{2}$$

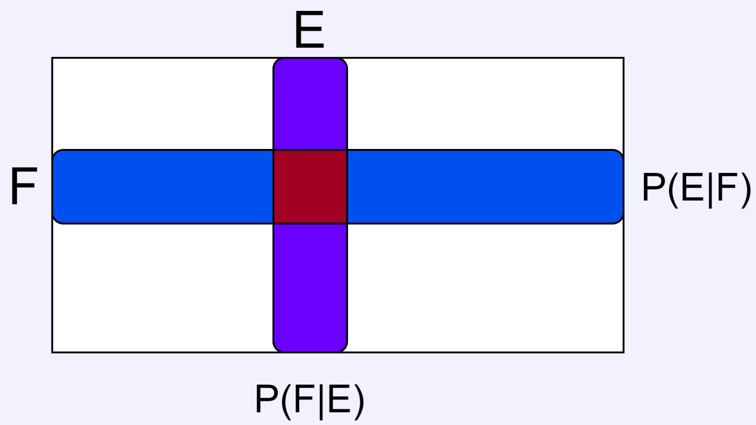
Bayes Theorem **Definition 2.4.2**

For two events E and F we have:

$$P(E\cap F)=P(F)\times P(E|F)=P(F)\times\frac{P(E\cap F)}{P(F)}=P(E)\times P(F|E)=P(E)\times\frac{P(E\cap F)}{P(E)}$$

Hence we can deduce:

$$P(E|F)=\frac{P(E)\times P(F|E)}{P(F)}$$

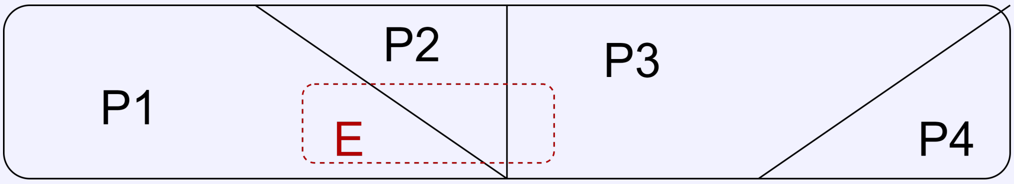


Partition Rule **Definition 2.4.3**

Given a set of events $\{F_1, F_2, \dots\}$ which forms a partition of S (disjoint sets that contain all of F).

For any event $E \subseteq S$:

$$P(E) = \sum_i P(E|F_i) \times P(F_i)$$



Proof:

(1)	$E = E \cap S = E \cap \bigcup_i F_i = \bigcup_i (E \cap F_i)$	By set theory and disjointness of partitions.
(2)	$P(E) = P(\bigcup_i (E \cap F_i))$	
(3)	$P(E) = \sum_i P(E \cap F_i)$	By axiom 3 and disjointness of partitions.
(4)	$P(E) = \sum_i P(E|F_i) \times P(F_i)$	

Law of Total Probability **Definition 2.4.4**

Given some event E and events $\{F_1, F_2, \dots\}$:

$$P(E) = \sum_i P(E \cap F_i)$$

For example the 6-Sided dice, $E = H$ and $F = [\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}]$, the marginal probability is the same as the sum of all cells in row H.

Using complement as a partition we can deduce that:

$$P(E) = P(E \cap F) + P(E \cap \overline{F})$$

$$P(E) = P(E|F) \times P(F) + P(E|\overline{F}) \times P(\overline{F})$$

2.4.1 Terminology Recap

- **Conditional Probabilities** Of the form $P(E|F)$.
- **Joint Probabilities** Of the form $P(E \cap F)$.
- **Marginal Probabilities** Of the form $P(E)$.

Chapter 3

credit