CE 211
Mathematics for Engineers

Lecture 2
Axioms and Laws of Probability

Previously on Mathematics for Engineers

Definition (Permutations)

The number of ways of arranging n items is $n!$

Definition (Combinations)

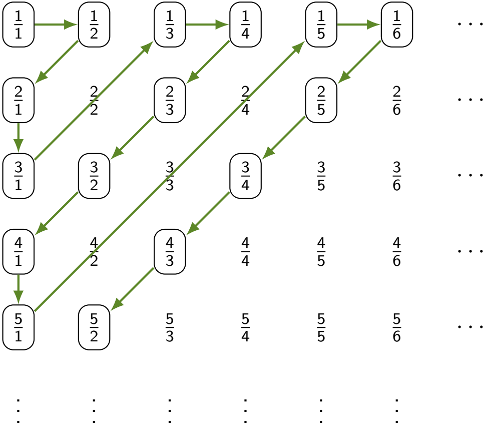
The number of ways of selecting k items from a set of n items is given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Definition (Countably Infinite)

A set A is countably infinite if there exists a bijective function f from A to the set of natural numbers $\mathbb{N} = \{1, 2, 3, \dots,\}$.

The set of rational numbers $\mathbb{Q}$ is countably infinite!



Previously on Mathematics for Engineers

Sets that are not countable are called **uncountable**. The set of real numbers $\mathbb{R}$ is uncountable since we cannot create a bijective function to $\mathbb{N}$. Cantor proved this using a diagonalization argument.

Let's proceed by contradiction. We will show that the unit interval $(0, 1)$ is uncountable. Extension to $\mathbb{R}$ is trivial. (Why?) Suppose we can list all reals in $(0, 1)$ and associate each number with a natural number

$$\begin{array}{cc}
1 & 0.\mathbf{3}895127\ldots \\
2 & 0.2\mathbf{5}00000\ldots \\
3 & 0.62\mathbf{4}6346\ldots \\
4 & 0.222\mathbf{2}222\ldots \\
5 & 0.1225\mathbf{7}43\ldots \\
6 & 0.58521\mathbf{5}8\ldots \\
\vdots & \vdots
\end{array}$$

Construct a number which differs from the diagonal elements shown in bold. E.g., 0.472561... This will never appear in the above list!

Previously on Mathematics for Engineers

We are usually interested in the probability that one or some of the outcomes the sample space occur.

These questions can be translated to a subset of outcomes that are called **events**. We say that an event happened if one of the outcomes in the event occurs during the experiment.

For example, in the previous experiments

- What is the event where we see exactly two heads

$$\{HHT, HTH, THH\}$$

- What is event where the sum of the dice is 7

$$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

Note that not all subsets can be easily described in words. But we still treat them as events and can ask the probability of its occurrence.

Previously on Mathematics for Engineers

An intuitive and familiar way for computing probabilities of events is to look at the number of elements in the event and divide it by the total number of outcomes.

Definition (Discrete Uniform Probability)

Suppose the sample space of an experiment consists of n outcomes which are equally likely, then the probability of an event A is

$$\mathbb{P}(A) = \frac{|A|}{n}$$

One must be careful in constructing the outcomes of the sample space. For example, when two dice are thrown, the sum has 11 outcomes: 2, 3, ..., 12. Using this argument, the probability is $1/11$. If you treat the dice to be indistinguishable, the answer would be $3/21$ (Why?) Both the answers are wrong because all outcomes are not equally likely.

Lecture Outline

The Problem

The Problem

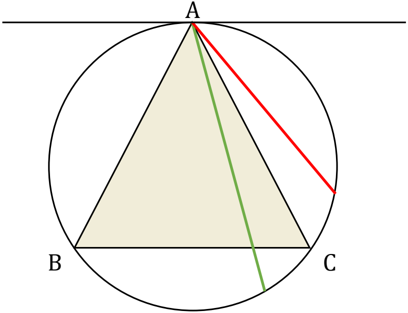
Bertrand Paradox

Imagine a circle of radius r and an equilateral triangle inscribed in it. Suppose a chord is randomly constructed. What is the probability that its length is greater than the side of the triangle $r\sqrt{3}$.

The Problem

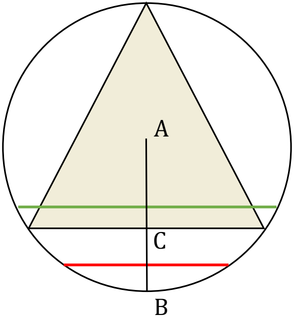
Bertrand Paradox

Solution 1: Randomly construct an equilateral triangle ABC. Draw a chord hinged at point A. If the chord, falls within the angle made by AB and AC, its length is greater than the side.



This can happen with a probability $60/180 = 1/3$.

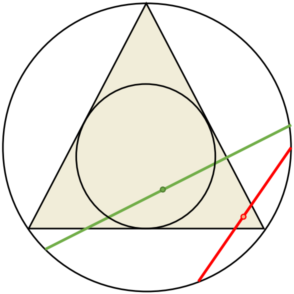
Solution 2: Pick a radius AB and randomly choose a point C on it. Draw a chord orthogonal to C. Draw an equilateral triangle such that AB is perpendicular to one of the sides. Using geometry, it can be shown that this side of the triangle intersects AB at the midway point.



Thus, the constructed chord is longer than the side of the equilateral triangle with a probability 1/2.

The Problem
Bertrand Paradox

Solution 3: Draw a chord inside the circle at random. If the center of the chord falls inside a circle inscribed in the equilateral triangle, then the chord length is greater than the side of the triangle.



The mid-point can fall inside the inner circle with a probability $\frac{\pi r^2/4}{\pi r^2} = 1/4$.

Which of these three answers is correct?

Lecture Outline

Axioms of Probability

Axioms of Probability

Intuition

The finite case definition cannot clearly be applied to situations where the state space is countably infinite or uncountable.

It does, however, provide the intuition for an axiomatic approach to probability which assumes, without proof, that some statements are true.

Axioms of Probability

Intuition

Suppose Ω is the sample space of an experiment (can be finite, countably infinite, or uncountable)

Axioms

1. For every event $A \subset \Omega$, $\mathbb{P}(A) \geq 0$
2. $\mathbb{P}(\Omega) = 1$
3. If $A_1, A_2, \ldots$ are disjoint events, i.e., $A_i \cap A_j = \emptyset \, \forall i, j$, then

$$\mathbb{P}\big(\cup_{i=1}^{\infty} A_i \big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

When Ω is finite, the third axiom only involves a finite number of events.

Axioms of Probability

Propositions

The three axioms of probability can be used to prove standard results which we saw in the last class without using the cardinality of the events.

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}((A \cup B)^c) = \mathbb{P}(A^c \cap B^c)$

Axioms of Probability

Countably Infinite Sample Spaces

Let us revisit the earlier question on picking an even number from a countably infinite sample space such as $\mathbb{N}$.

If it is assumed that every number can be picked uniformly with equal probability, we have a problem since the third axiom is violated.

To see why, let event A_i represent selecting the number i. If all such events are equally likely and $\mathbb{P}(A_i) = \epsilon$, $\mathbb{P}\big(\cup_{i=1}^{\infty} A_i\big) = 1$ but $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$.

Axioms of Probability

Countably Infinite Sample Spaces

What went wrong in the previous example? Should the sample space not be countably infinite?

It is perfectly okay to have countably infinite state spaces. For instance, our experiment could involve counting the number of calls made over a cellular network.

The problem arose since we presupposed that a probability function which places an equal weight on all outcomes exists.

Such a function does not exist for a countably infinite spaces. In other words, it is wrong to assume that we can pick a number from $\mathbb{N}$ randomly with uniform probability.

Axioms of Probability

Uncountable Sample Spaces

Now consider the problem of throwing a dart randomly on the real line between 0 and 1. What is the probability of hitting a number, say $1/4$?

We could define a probability function using the length of intervals as a proxy for the 'number of elements in an event'. The the length of the set $\{1/4\}$ is zero while that of the interval $[0, 1]$ is 1, and hence the required probability is 0.

What is the probability that the dart lands in the sub-interval $[0, 1/4]$? The length of the interval is $1/4$ and hence the probability is $\frac{1/4}{1}$.

Note that the third axiom only allows unions of countably infinite events and hence we cannot take unions of all elements and get $0 = \infty$ from the third axiom.

What is the probability of hitting a

- Rational number? Zero
- Irrational number? One

This notion of length can be used to define a probability function and can be extended to areas and volumes in higher dimensions and is called the Lebesgue measure.

But there is a catch!

Axioms of Probability

Vitali Sets

When defining events in the finite case, every subset of the sample space could be treated as an event and hence it was okay to ask the probability of occurrence of any set in 2^{Ω}.

It turns out that that's no longer true for uncountable cases. There are strange subsets where the axioms of probability break down!

An example is the **Vitali Set** which can be used to construct countably infinite subsets of $[0, 1]$, one for every rational number A_q such that $\cup_q A_q = [0, 1]$ and all A_q's have equal probability. This violates the third axiom.

From an engineering standpoint, we will just assume that the events of interest are always 'nice' subsets of the sample space and we will not encounter such pathological instances.

As a teaser, let's briefly discuss how these cases are handled but it is not of much importance for the rest of the course.

Axioms of Probability

Every event is a subset of the sample space but not all subsets are valid events. The set of all valid events is called a σ-**algebra** and is denoted by $\mathcal{F}$.

The tuple $(\Omega, \mathcal{F})$ is said to be a measurable space and given such a space, we can define a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$ which satisfies the three axioms. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called the **probability space**.

Axioms of Probability

σ-algebra

A σ-algebra is a collection of subsets which is closed under countable unions and complements. Closed here implies that these operations will produce sets which also belong to the collection.

For example, let $\Omega = \{1, 2, 3, 4\}$. The following collection of sets form a σ-algebra and are all treated as valid events when studying probability

- 2^{Ω}
- $\{\emptyset, \{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}\}$

For uncountable sets such as the real line, we usually take all open intervals and add to it all possible countable unions and complements and continue this process. This resulting set is also called the **Borel σ-algebra**.

Events that are of usual interest will always lie in this set.

Axioms of Probability

Summary

To summarize, a probability space consists of three components

- A sample space Ω which is the set of all outcomes
- A set of events $\mathcal{F}$
- A probability measure or a function $\mathbb{P} : \mathcal{F} \to [0, 1]$

The probability measure must satisfy the following three axioms.

Axioms

1. For every event $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$
2. $\mathbb{P}(\Omega) = 1$
3. If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint events, i.e., $A_i \cap A_j = \emptyset \, \forall i, j$, then

$$\mathbb{P}\big(\cup_{i=1}^{\infty} A_i \big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Axioms of Probability

Summary

Although probability was widely studied in the 18th and 19th century, it was only in 1930s that Andrey Kolmogorov laid out the foundations of the axiomatic approach to probability.

Conditional Probability

Conditional Probability

Consider the problem of two dice. If we are told that sum equals 7, what is the probability that the first dice was less than or equal to 4.

Recall that the set of outcomes associated with the event – Observing a sum of 7 is

$$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

Since we know that the number on the first dice is ≤ 4, of the above 7 outcomes, the outcomes of interest are

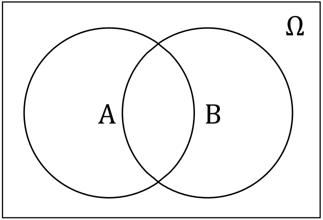
$$\{(1,6),(2,5),(3,4),(4,3)\}$$

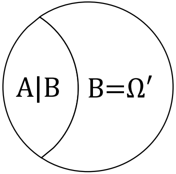
Therefore, the required probability is $4/6$.

Conditional Probability

Introduction

Such problems can be modelled using the concept of conditional probability. We write $\mathbb{P}(A|B)$ to indicate the probability of observing event A given that B occurred. Pictorially,





In the first case, we know that Ω always occurs. This is replaced with B in the second case since B is known to always occur.

Conditional Probability

Introduction

Definition (Conditional Probability)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If A and B are two events in $\mathcal{F}$ and if $\mathbb{P}(B) > 0$, the probability of A given B is

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

Many a times, we will skip writing the probability space but it is implicitly assumed that the events belong to $\mathcal{F}$.

Solving the previous problem using this definition, let events A and B be observing a sum of 7 and observing at least 4 on the first die respectively.

$\mathbb{P}(A \cap B)$ is given by $4/36$ and $\mathbb{P}(B) = 6/36$. Therefore, $\mathbb{P}(A|B) = 4/6$.

Conditional Probability

Introduction

It can be easily shown that conditional probability also satisfies all three axioms of probability. That is,

- $\mathbb{P}(A|B) \geq 0$
- $\mathbb{P}(\Omega|B) = 1$
- If $A_1, A_2, \ldots$ are disjoint events, i.e., $A_i \cap A_j = 0 \, \forall i, j$, then

$$\mathbb{P}\big(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$$

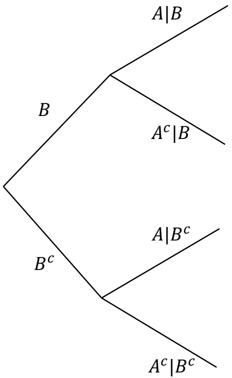
Note that $\mathbb{P}$ function is still defined from $2^{\Omega} \rightarrow [0,1]$ or $\mathcal{F} \rightarrow [0,1]$ or write a new probability function $\mathbb{P}(.|B)$

Hence, we can write other identities such as

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

Suppose a medical test for a rare disease successfully detects the disease in individuals having it 90% of the time. It also known to result in false positives 5% of the time when individuals are known to not have the disease.

If 10% of the population have the disease, what is the probability that an individual has the disease and the test fails? What is the probability that the individual does not have the disease and the test is positive?



Define events A and B as $B = \{\text{The individual has the disease}\}$ and $A = \{\text{The medical test is positive}\}$

Conditional Probability

Introduction

The expression for conditional probability $\mathbb{P}(B \cap A) = \mathbb{P}(B)\mathbb{P}(A|B)$ can be extended to multiple events as follows

Definition (Multiplication Rule)

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$$

Conditional Probability

Quickercise

Solve with and without using conditional probability.

- A box contains three white balls and two red balls. Two balls are removed without replacement. What is the probability that the first is white and the second is red?
- A deck of 52 cards is randomly divided into 4 piles of 13 cards each. What is the probability that each pile has exactly one Ace?
- Suppose you throw a dart uniformly on a line between 0 and 1 but you don't get to observe the result. If you were told that it was not in the interval $[0, 1/4]$, what is the probability that it landed in the interval $[7/8, 1]$?

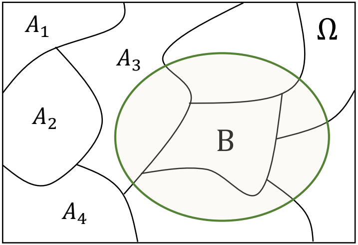
Conditional Probability

Law of Total Probability

Definition (Law of Total Probability)

Suppose $A_1, \ldots, A_n$ represents a partition of the sample space Ω and $\mathbb{P}(A_i) > 0 \,\forall\, i = 1, \ldots, n$. Then, for any event B

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B) + \ldots + \mathbb{P}(A_n \cap B)\\ &= \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \ldots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)\end{aligned}$$



Theorem (Bayes' Theorem)

Suppose $A_1, \ldots, A_n$ *represents a partition of the sample space* Ω *and* $\mathbb{P}(A_i) > 0 \, \forall \, i = 1, \ldots, n$*. Then, for any event* B *with* $\mathbb{P}(B) > 0$

$$\begin{aligned}\mathbb{P}(A_i|B) &= \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(A_1)\mathbb{P}(B|A_1) + \ldots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)}\end{aligned}$$

For two events A and A^c, Bayes' theorem can be rewritten as

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

Conditional Probability

Quickercise

Suppose a medical test for a rare disease successfully detects the disease in individuals having it 90% of the time. It also known to result in false positives 5% of the time when individuals are known to have the disease.

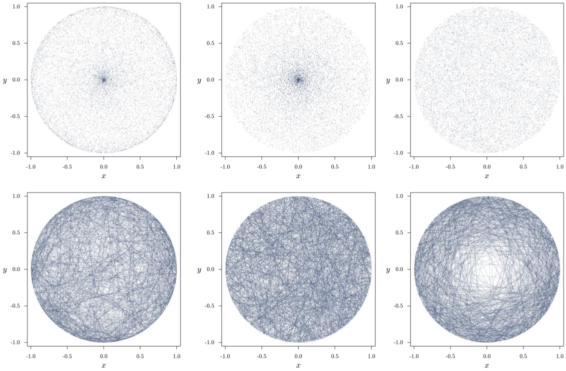
If 10% of the population have the disease and the test result is positive. What is the probability that the individual actually has the disease?

A Solution

A Solution

Bertrand Paradox

We get different answers in Bertrand's paradox since they are 'different problems'.



The word 'random' is not specific enough and all three answers are correct for a certain version of randomness.

Your Moment of Zen