CE 211
Mathematics for Engineers

Lecture 3

Independence and Introduction to Random Variables

Previously on Mathematics for Engineers

Suppose an unbiased coin is tossed thrice. Such an act is called an **experiment**. Each experiment results in an **outcome**, e.g., Heads, Tails, Heads.

The set of all possible outcomes is called the **sample space** which is usually denoted by Ω. 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

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

Note that not all subsets can be easily described in words. But we still treat them as events (barring a few) 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

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

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.

Previously on Mathematics for Engineers

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**.

Previously on Mathematics for Engineers

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

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

Previously on Mathematics for Engineers

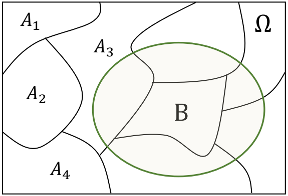
Definition (Conditional Probability)

If A and B are two events and if $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

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

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



Previously on Mathematics for Engineers

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

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

Lecture Outline

The Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.

He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

Lecture Outline

Independence

Independence

Intuition and Definition

Imagine a finite same space. If an event A can happen in $|A|$ ways and event B happens in $|B|$ ways. If both events are independent of each other, in how many ways can A and B occur? $|A||B|$.

The same logic can be extended to probability and other kinds of sample spaces.

Definition (Independence)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, two events A and B in $\mathcal{F}$ are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Events that are not independent are said to be dependent. The notation $A \perp\!\!\!\perp B$ is sometimes used to indicate that A and B are independent.

If two events A and B are independent of each other, what can you say about $\mathbb{P}(A|B)$? What about $\mathbb{P}(B|A)$? In other words, does the occurrence of B provide any information on the occurrence of A and vice versa?

Claim

If two events A and B are independent, $\mathbb{P}(A|B) = \mathbb{P}(A)$ and $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Is the converse true? We prefer using the earlier definition since it does not require the probabilities of the events to be strictly positive.

Suppose two events A and B are disjoint. Are they independent? Can you think of an example?

Independence

Quickercise

- Suppose a card is randomly picked from a deck of 52 cards. If A and B represent events that the selected card is an ace and spade respectively, are they independent?
- Suppose an unbiased coin is flipped twice. Let A denote the event that we see head on the first toss and B represent the event of seeing a tail on the second toss. Are these two independent?
- Suppose we roll two dice and A is the event that the sum is 6 and B is the event that the first dice lands on 4. Are A and B independent? What if A was the event that the sum is 7?

Independence

Caution

Note that in the coin toss example, it is okay to reason that the first toss has two outcomes H and T and hence the probability of A is $1/2$. The same argument applies for event B.

But to be very precise, we are mixing up two experiments. The probabilities of the events are defined using $(\Omega, \mathcal{F})$.

Hence, we must use the outcomes of the form (H, H) and (H, T) for the event A and (H, T) and (T, H) for event B.

What if we have more than three events? Consider the following three events in the context of rolling two dice:

- A: Sum of dice is 7
- B: First dice is 4
- C: Second dice is 3

Clearly, A and B, B and C, and C and A are independent of each other. However, if we knew that B and C occurred, we can definitely say something more about A.

Hence, we could decide to define independence as

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

.

Independence

Mutual Independence

However, $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ does not imply pairwise independence as seen in the following example. Suppose the events in the two dice case are:

- A: First dice is 1, 2, or 3
- B: First dice is 3, 4, or 5
- C: Sum of dice is 9

Definition (Independence of Three Events)

Three events A, B, and C are said to be independent if

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

Definition (Independence of n Events)

The $A_1, A_2, \ldots, A_n$ are said to be independent if

$$\mathbb{P}\left(\cap_{I \subset \{1,\ldots,n\}} A_i\right) = \Pi_{i \in I}\mathbb{P}(A_i)$$

Definition (Conditional Independence)

Given that an event C occurred, two events A and B are said to be conditionally independent if $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$

Show that if $\mathbb{P}(B \cap C) \neq 0$, the above definition is equivalent to

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

Here is an example where two events are independent but are conditionally dependent. Consider the following events in the context of tossing a coin twice

- A: First toss is a head
- B: Second toss is a head
- C: Both tosses have different sides

Clearly, A and B are independent but $\mathbb{P}(A \cap B|C) = 0$ and not equal to $\mathbb{P}(A|C)\mathbb{P}(B|C)$.

Here is another example where two events are conditionally independent but not independent.

Two biased coins 1 and 2 can be chosen at random for two tosses. The probability of observing heads on 1 is 0.99 and that on 2 is 0.01.

- A: First toss is a head
- B: Second toss is a head
- C: Coin 1 was selected

Here, $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C) = (0.99)(0.99)$. Calculate $\mathbb{P}(A), \mathbb{P}(B)$, and $\mathbb{P}(A \cap B)$. (Hint: Use law of total probability)

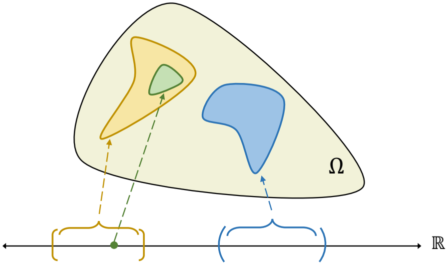
Lecture Outline

Random Variables

A random variable is an alternate way of constructing events. Defining random variables allows us to translate events of interest into probabilities more easily.

Definition (Random Variable)

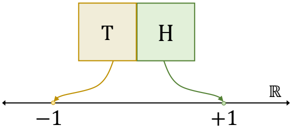
A real-valued random variable is a function or mapping $X : \Omega \to \mathbb{R}$ such that for all $S \subset \mathbb{R}$, $X^{-1}(S) \in \mathcal{F}$.



*Technically, there are some restrictions on S just like valid events, but we'll ignore those details.

For example, consider a coin toss. We could define a random variable as follows:

$$X(\omega) = \begin{cases} +1 & \text{if } \omega = H \\ -1 & \text{if } \omega = T \end{cases}$$



Imagine you get ₹1 if it lands on H and pay ₹1 if it lands on T, then it is natural to define a random variable this way.

We could ask what is the probability that you'll get between [2, 3], exactly -1, ≥ 0, etc. Each of these describes an event in $\mathcal{F}$.

- $X \in [2, 3] \equiv \emptyset$
- $X = -1 \equiv \{T\}$
- $X \geq 0 \equiv \{H\}$

There is nothing special about the $+1$ and -1 in the previous example. We could define multiple random variables for the same probability space.

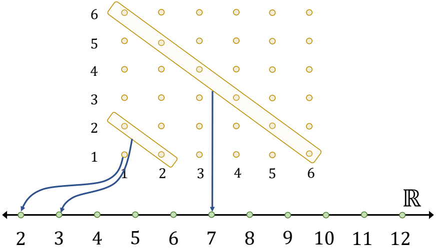
$$X(\omega) = \begin{cases} +2019 & \text{if } \omega = H \\ -560012 & \text{if } \omega = T \end{cases}$$

This is a valid random variable (perhaps not a useful one). The problem context will help in defining random variables.

We could ask again ask what is the probability that you'll get between [2, 3] $\cup$ [2010, 2020], exactly -1, $\leq$ 2019, etc. Each of these cases describes an event in $\mathcal{F}$.

- $X \in [2, 3] \cup [2010, 2020] \equiv \{H\}$
- $X = -1 \equiv \emptyset$
- $X \leq 2019 \equiv \{H, T\}$

Now consider the case where you roll two dice. Suppose, X is the random variable defined as the sum of the numbers.



Then, we could ask what is the probability that $X = 7$. This is same as the probability with which the following event will occur

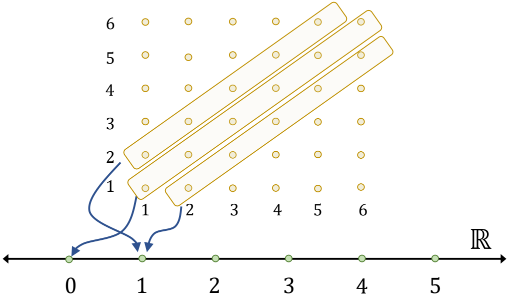
$$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

What is the event corresponding to $X \in [0,1]$?

Random Variables

Introduction

Instead, if we define X as the absolute value of difference in the numbers on the dice. What is the event corresponding to $X = 7$? $X = 1$?



Random Variables

Differences and Notation

Note the probability measure is a function $\mathbb{P} : \mathcal{F} \to [0,1]$ where you can think of $\mathcal{F}$ as 2^{Ω}, whereas the random variable X is another function $X : \Omega \to \mathbb{R}$.

As seen in the previous examples, for subsets $S \subset \mathbb{R}$, we can find an event $A \in \mathcal{F}$ such that $X^{-1}(S) = A = \{\omega \in \Omega | X(\omega) \in S\}$.

Hence, the following probabilities are the same

$$\mathbb{P}(X \in S) = \mathbb{P}(X^{-1}(S)) = \mathbb{P}(A) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\})$$

Be careful to not write $X(A)$ and $\mathbb{P}(S)$, where $A \in \mathcal{F}$ and $S \subset \mathbb{R}$ (unless of course $\Omega = \mathbb{R}$).

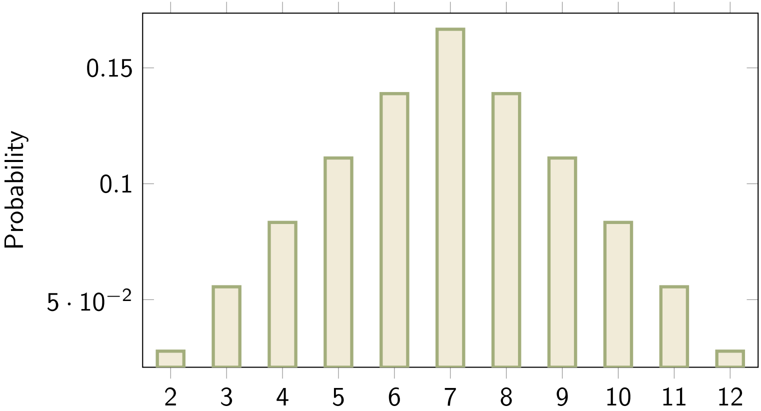
Discrete random variables are random variables that have a finite or countably infinite images. (They can still be defined on an uncountable Ω.)

In the previous examples, we could list the probability of each outcome. For instance, when two dice are thrown and the random variable is the sum of numbers, the probabilities of the outcomes are

ω	$\mathbb{P}(\{\omega\})$	ω	$\mathbb{P}(\{\omega\})$
2	$1/36$	8	$5/36$
3	$2/36$	9	$4/36$
4	$3/36$	10	$3/36$
5	$4/36$	11	$2/36$
6	$5/36$	12	$1/36$
7	$6/36$		

Random Variables

Discrete Random Variables



*The bar widths have been magnified only for better readability. They are actually point masses concentrated at each outcome.

Definition (Probability Mass Function)

The probability mass function (PMF) of a random variable X represents the probability of each outcome. It is denoted as $p_X(x)$ and is defined as

$$p_X(x) = \mathbb{P}(X = x)$$

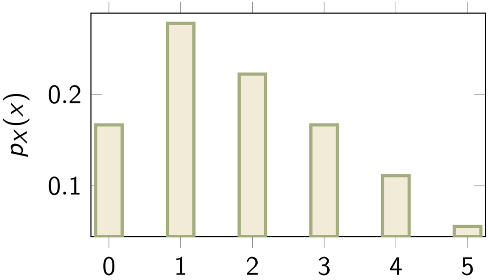
If it is clear from the context, the subscript X can be ignored and the pmf is simply written as $p(x)$.

Given the PMF of X, we can find $\mathbb{P}(X \in S)$ using $\sum_{x \in S} p_X(x)$.

Note that capital letters X, Y are reserved for the random variables and x, y are used to indicate its realizations or outcomes. So as pointed out earlier, $\mathbb{P}(x)$ is bad notation.

In the problem of throwing two dice, what is the PMF of the random variable defined as the absolute value of the difference of the numbers on the faces.

x	$p_X(x)$
0	6/36
1	10/36
2	8/36
3	6/36
4	4/36
5	2/36



The probability with which random variables lie in the interval $(-\infty, x]$ is of special interest.

Definition (Cumulative Distribution Function)

The cumulative distribution function (CDF) of a random variable X is denoted by $F_X(x)$ and is defined as

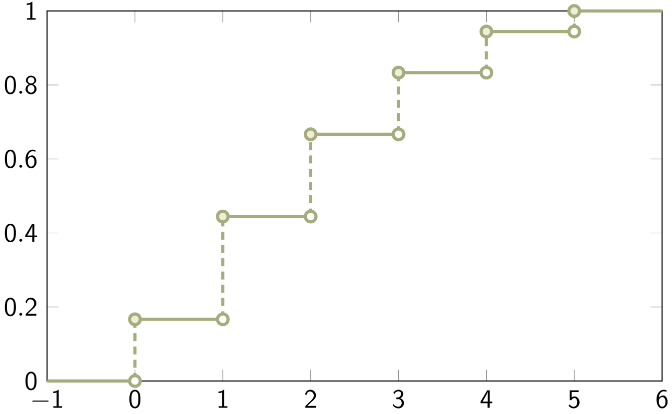
$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x' \leq x} p_X(x')$$

Construct the CDF function for the coin toss experiment and the two random variables defined for the dice experiment.

The CDF function is always non-decreasing and right continuous. It also satisfies

$$\lim_{x\to-\infty} F_X(x) = 0 \qquad \lim_{x\to+\infty} F_X(x) = 1$$

Further, for discrete random variables the CDF is a step function.

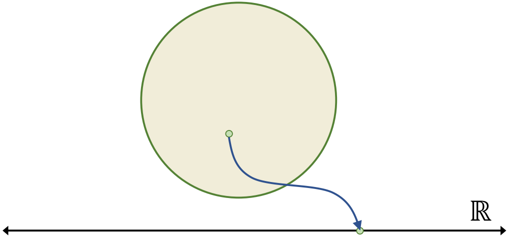


Random Variables

Continuous Random Variables

Continuous random variables are ones which have uncountable images. (Their domain will hence be an uncountable sample space.)

For example, X could represent the location of a randomly thrown dart on the interval $[0, 1]$ in which case it can be written as $X : [0, 1] \to \mathbb{R}$ or on a two-dimensional circle of some radius, i.e., $X : C \to \mathbb{R}$, where $C = \{(x, y) | x^2 + y^2 \leq r\}$ etc.



Just like the discrete case, we define a probability density function but with a small twist since the probability of observing a singleton event is 0.

Definition (Probability Density Function)

The probability density function (PDF) of a continuous random variable is denoted as $f_X(x)$ and is defined as

$$f_X(x)dx = \mathbb{P}\big(X \in [x, x + dx]\big)$$

Thus, the probability that the random variable lies in a subset S is given by

$$\mathbb{P}(X \in S) = \int_{x \in S} f_X(x)dx$$

Since the probability that X equals any value is 0, the above definition could have been written using $(x, x + dx]$, $[x, x + dx)$, or $(x, x + dx)$.

The definition of CDF for continuous random variables remains unchanged

Definition (Cumulative Distribution Function)

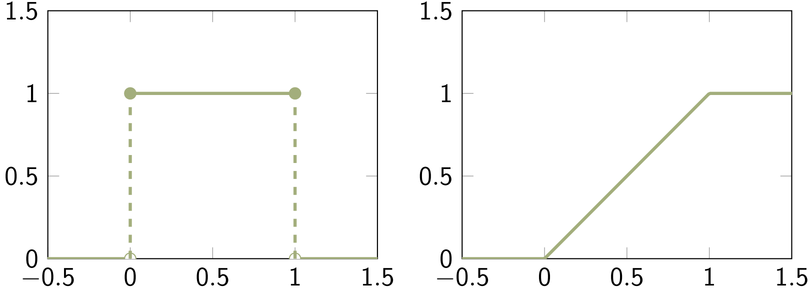
The cumulative distribution function (CDF) of a random variable X is denoted by $F_X(x)$ and is defined as

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)dx$$

Random Variables

Continuous Random Variables

The PDF and CDF of a uniformly distributed random variable is shown below.



Note that just as the sum of the lengths of the vertical bars in the discrete case had to equal 1, the area under the PDF must be 1.

Random Variables

Support

Often we use the term **support of a random variable** to indicate the points where the PMF or PDF is positive. Specifically,

For discrete random variables,

$$R_X = \{x \in \mathbb{R} | p_X(x) > 0\}$$

For continuous random variables,

$$R_X = \{x \in \mathbb{R} | f_X(x) > 0\}$$

What is the support of the random variable representing the sum of numbers on two dice? What is the support of the uniformly distributed random variable that we saw earlier?

A Solution

A Solution

Monty Hall Problem

Here are the equiprobable outcomes of the Monty Hall problem. You could also name the goats and write six outcomes.

Door 1	Door 2	Door 3	Stick	Switch
Car	Goat	Goat	W	L
Goat	Car	Goat	L	W
Goat	Goat	Car	L	W

If you picked door 1 and the host opened door 3, it is tempting to think that the odds of finding the goat or the car in the second door is $1/2$.

The host does not pick any door. He chooses one of the two doors without the car and always shows a goat. If A is the event of winning from your initial choice (say door 1) and B is the event that the host shows you a goat, then Bayes' theorem gives $\mathbb{P}(A|B) = 1/3$ which hasn't improved from before.

However, the complement of A, i.e., winning from switching (or losing from sticking to door 1) has a probability $2/3$.

Your Moment of Zen