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	$S = \{x \in \mathbb{N} \mid \exists y \in \mathbb{N}.[2y + 1 = x]\}$

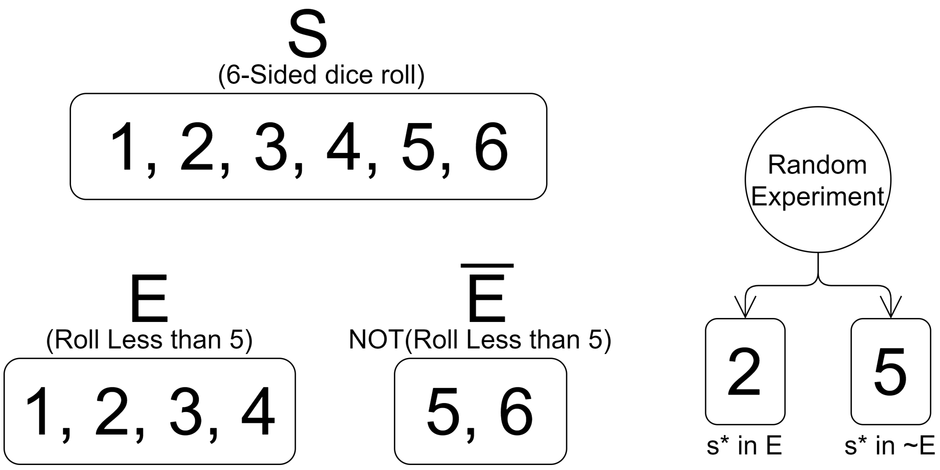
Event **Definition 2.1.2**

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: $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

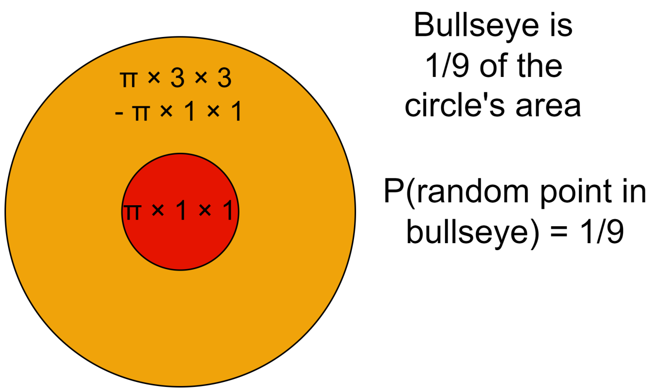
2.2 Interpretations of Probability

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: $\begin{matrix} \text{Option 1: } E \text{ occurs win } £1, \overline{E} \text{ occurs win } £0 \\ \text{Option 2: Regardless of outcome get } £P(E) \end{matrix}$.

Either outcome, the gambler receives $£P(E)$. 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:

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

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 1	2	3	4	5	6	Totals
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:

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

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.

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

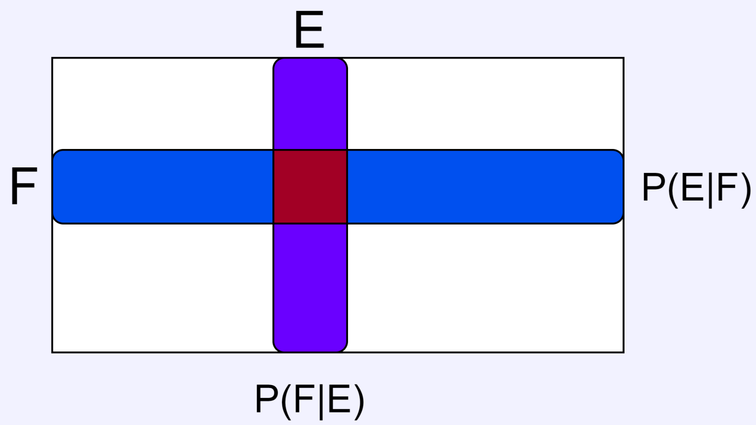
Bayes Theorem **Definition 2.4.2**

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:

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

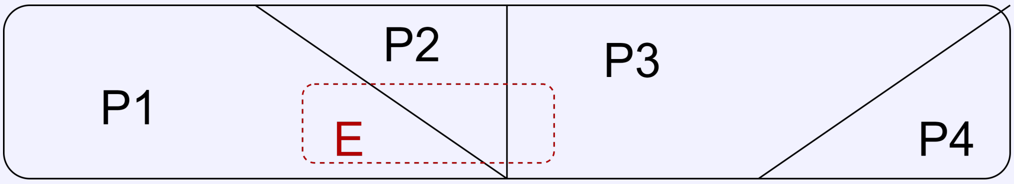


Partition Rule **Definition 2.4.3**

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

Random Variables

Probability Space **Definition 3.0.1**

$$(S, \mathcal{F}, P)$$

Models a random experiment where probability measure $P(E)$ is defined on subsets $E \subseteq S$ belonging to sigma algebra $\mathcal{F}$.

Within a sample space we can study quantities that are a function of randomly occurring events (e.g temperature, exchange rates, gambling scores).

Random Variable **Definition 3.0.2**

A *random variable* is a mapping from the sample space to the real numbers, for example *random variable* X:

$$X : S \to \mathbb{R}$$

Each element in the sample space $s \in S$ is assigned to a numerical value by $X(s)$.

When referring to the value of a random variable we use its name, e.g X in $P(5 < X \leq 30)$

- **Simple** Finite set of possible outcomes. (e.g dice faces)
- **Discrete** Countable outcomes/support/range. (e.g distance (m))
- **Continuous** Can be a continuous range (e.g temp)

Single Fair Dice Roll **Example Question 3.0.1**

$S = \{1, 2, 3, 4, 5, 6\}$, for any $s \in S.P(\{s\}) = \frac{1}{6}$.
We can define random variable X such that:

$$X(1) = 1, X(2) = 2, X(3) = 3, X(4) = 4, X(5) = 5, X(6) = 6$$

Then we can use X:

$$P_X(1 < X \leq 5) = P(\{2, 3, 4, 5\}) = {}^2\!/_3$$

$$P_X(X \in \{2, 3\}) = P(\{2, 3\}) = {}^1\!/_3$$

We can also define random variable Y such that:

$$Y(\epsilon) = \begin{cases} 0 & \epsilon \text{ is odd} \\ 1 & \epsilon \text{ is even} \end{cases}$$

And hence:

$$P_Y(Y = 0) = P(\{1, 3, 5\}) = {}^1\!/_2$$

3.1 Induced Probability

The probability measure P defined on a sample space S induces a probability distribution on the random variable in $\mathbb{R}$ (distribution of its outcomes).

$$S_X = \{s \in S | X(s) \leq x\}$$

Such that:

$$P_X(X \geq x) = P(S_X)$$

Note that unless there is ambiguity, $P_X(\dots)$ will often be written as $P(\dots)$.

Heads and Tails **Example Question 3.1.1**

We define random variable $X : \{H, T\} \to \mathbb{R}$ over the *continuum* $\mathbb{R}$ such that:

$$X(T) = 0 \text{ and } X(H) = 1$$

$$S_X = \begin{cases} \emptyset & \text{if } x < 0 \\ \{T\} & \text{if } 0 \leq x < 1 \\ \{H, T\} & \text{if } x \geq 1 \end{cases}$$

X represents the number of heads flipped.

$$P_X(X \leq x) = P(S_X) = \begin{cases} P(\emptyset) = 0 & \text{if } x < 0 \\ P(\{T\}) = 1/2 & \text{if } 0 \leq x < 1 \\ P(\{H, T\}) = 1 & \text{if } x \geq 1 \end{cases}$$

Now we can use X to compactly show probabilities.

$$P_X(X = 1) = 1/2$$

Multiple Coin Flips **Example Question 3.1.2**

$$S = \{TTT, TTH, THT, HTT, THH, HHT, HTH, HHH\}$$

We can define X (number of heads):

$$X(s) = \begin{cases} 0 & s = TTT \\ 1 & s \in \{TTH, THT, HTT\} \\ 2 & s \in \{THH, HHT, HTH\} \\ 3 & s = HHH \end{cases}$$

Hence given 3 coin tosses:

$P_X(X > 1)$	More than one head
$P_X(X < 3)$	Not all heads
$P_X(X \leq 1)$	At least one head

Support/Range **Definition 3.1.1**

The set of all possible values of a random variable X:

$$\mathbb{X} \equiv supp(X) \equiv X(S) = \{x \in \mathbb{R} | \exists s \in S. X(s) = x\}$$

As S contains all possible experiment outcomes, $supp(X)$ contains all possible values/outcomes for the random variables X.

$$P_X(X \leq x) \text{ is defined for all } x \in supp(X)$$

3.2 Cumulative Distributions

Cumulative Distribution Function (F_X) **Definition 3.2.1**

The cumulative distribution function (cfd) of a random variable X is the probability where X takes some value less than or equal to some x:

$$F_X : \mathbb{R} \to [0,1] \text{ such that } F_X(x) = P_x(X \leq x)$$

To be a valid cfd, 3 criteria must be met:

1. **Probability between 0 and 1** $\forall x \in \mathbb{R}. 0 \leq F_X(x) \leq 1$
2. **Monotonicity** $\forall x_1, x_2 \in \mathbb{R} x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$
3. **Infinite Bounds** $F_X(-\infty) = 0, F_X(\infty) = 1$

For any random variable a *cfd* is right-continuous (a result of monotonicity).

$$x_1 > x_2 > x_3 ... > x \Rightarrow F_X(x_1) >= F_X(x_2) >= ... >= F_X(x)$$

We can determine the probability over finite intervals using the cumulative distribution:

$$\text{for } (a, b] \subseteq \mathbb{R} \; P_X(a < X \leq b) = F_X(b) - F_X(a)$$

Distributions

Probability Mass Function (p_X) **Definition 3.2.2**

Also called *probability function* gives the probability that a discrete random variable is exactly equal to a value.

The sample space S is mapped onto elements in the *support* of X (one-to-one).

We can then partition the sample space into a countable, disjoint collection od event subsets:

$$s \in E_i \Leftrightarrow X(s) = x_i, i = 1, 2 \ldots$$

A probability mass function is valid if and only if:

1. **No negative probabilities** $\forall x \in supp(X).\ p_X(x) \geq 0$
2. **Probabilities sum to 1** $\sum_{x \in supp(x)} p_X(x) = 1$

3.3 Discrete Random Variable

For a *discrete random variable* we define the probability mass function as:

$$p_X(x_i) = P(X = x_i) = P(E_i) \text{ where } x_i \in supp(X) \text{ and } x_i \text{ is the outcome of event } E_i$$

We can also define using *cfds*:

$$F_X(x_i) = \sum_{j=1}^{i} p_X(x_j) \Leftrightarrow p_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \quad \text{where } i = 2, 3 \ldots$$

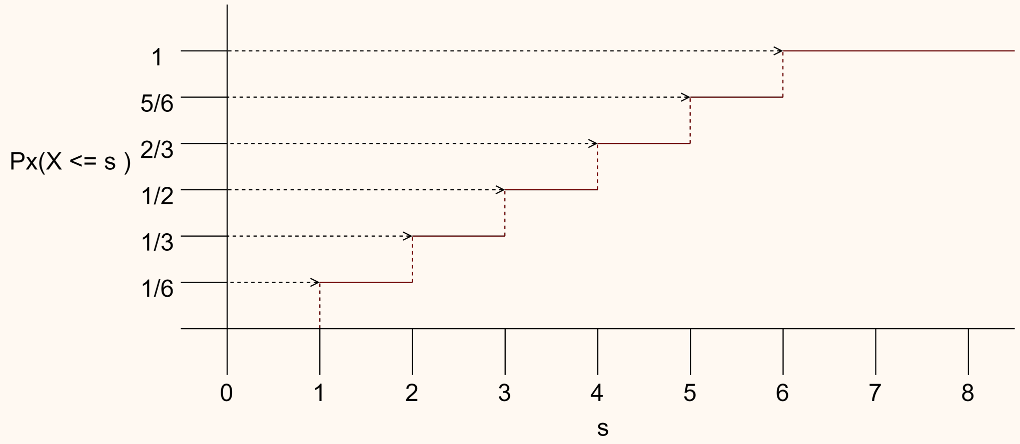
Or more simply:

$$p_X(x_i) = P_X(X = x_i) = P(X \leq x_i) - P(X \leq x_{i-1}) = F_X(x_i) - F_X(x_{i-1})$$

When graphed, F_X is a monotonically increasing, stepped function with jumps at points in $S(X)$.

Six Sided Dice — **Example Question 3.3.1**

Here we have X representing the value of the dice roll. We can plot the cumulative distribution (showing probability a dice roll is less than or equal to a given value).



Discrete CFDs have several properties:

- **Limiting Cases**
$$\lim_{x \to -\infty} F_X(x) = 0 \quad \lim_{x \to \infty} F_X(x) = 1$$
At ∞ the whole set of outcomes is covered, probabilities sum to 1. At $-\infty$ none are covered.
- **Continuous from the right**
$$\text{For } x \in \mathbb{R} \lim_{h \to 0^+} F_X(x+h) = F_X(x)$$
Moving from the right to the left the probability will reduce and tend towards the value.
- **Non-Decreasing**
$$a < b \Rightarrow F_X(a) \leq F_X(b)$$
As it is cumulative, the value can only grow larger moving right.
- **Can cover a range**
$$\text{For } a < b.\ P(a < X \leq b) = F_X(b) - F_X(a)$$

Poisson Distribution **Definition 3.3.1**

A discrete probability distribution expressing the probability of a given number of events occuring in a fixed time interval, given a constant mean.

$$Pois(\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ where } k \text{ is the number of occurrences}$$

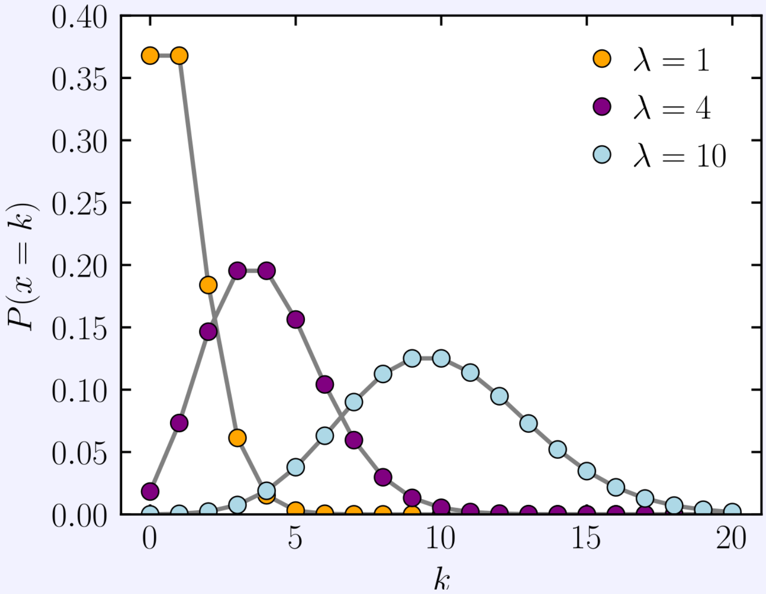
e.g What is the probability exactly 7 people buy pizzas at a stall in one hour, given on average is 4 people per hour?

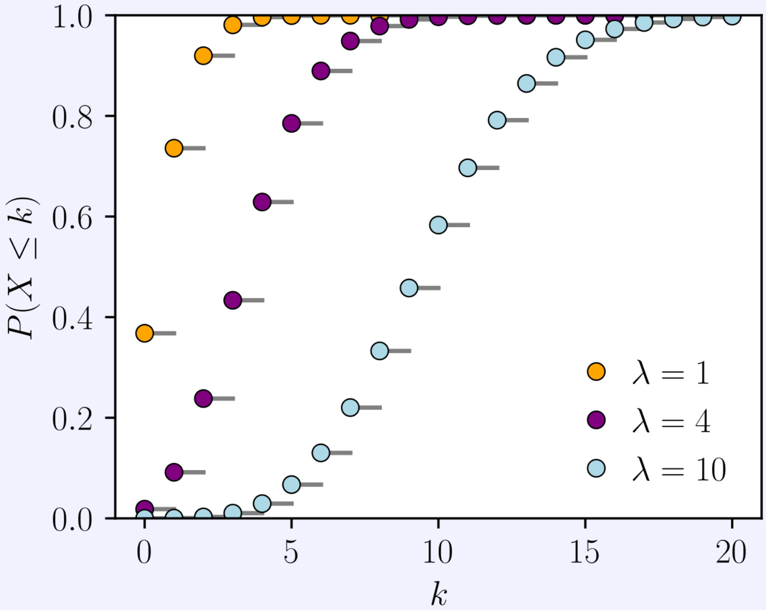
$$X \approx Poisson(4)$$

For a poisson distribution the mean (expected) and variance are equal.

$$E(X) = Var(X)$$

$$P(X = 7) = \frac{4^7 e^{-4}}{7!}$$





3.4 Link with Statistics

We can consider a set of data as realisations of a random variable defined on some underlying population of the data.

- Frequency histogram is an empirical estimate for the *pmf*.
- Cumulative histogram is an empirical estimate of the *cdf*.

3.5 Expectation

Expected Value **Definition 3.5.1**

The expectation of a *discrete random variable* X is:

$$E_X(X) = \sum_x xp(x)$$

Also referred to as μ_X it is the mean value of the distribution.

$$E(g(X)) = \sum_x g(x)p_X(x)$$

$$E(a \times X + b) = a \times E(X) + b$$

$$E(a \times g(X) + b \times f(X)) = a \times E(g(X)) + b \times E(f(X))$$

Given another distribution Y:

$$E(X + Y) = E(X) + E(Y)$$

Dice Rolls **Example Question 3.5.1**

Given random variable X representing the value of a dice roll:

$$X(n) = n \text{ where } 1 \geq n \geq 6$$

$$P(X = x) = \begin{cases} 1/6 & 1 \geq n \geq 6 \\ 0 & otherwise \end{cases}$$

We can get the expected as:

$$E(X) = 1/6 \times 1 + 1/6 \times 2 + 1/6 \times 3 + 1/6 \times 4 + 1/6 \times 5 + 1/6 \times 6 = 21/6 = 3.5$$

We can base scoring on the dice roll:

$$score(x) = 4 \times x + 2$$

Hence we can calculate that the expected score is $E(score(X)) = 4 \times 3.5 + 2 = 16$.

Dice and Coins **Example Question 3.5.2**

Given random variable D of a fair dice, and fair coin C:

$$P(D = x) = \begin{cases} 1/6 & 1 \geq n \geq 6 \\ 0 & otherwise \end{cases} \text{ and } P(C = x) = \begin{cases} 1/2 & x \in \{H, T\} \\ 0 & otherwise \end{cases}$$

Given $score = dice\ roll + 1$ if coin flip is heads what is the expected score?

$$E(D) = 3.5 \quad E(C) = 0.5 \quad E(score) = 3.5 + 2 * 0.5 = 4.5$$

3.6 Variance

Moment **Definition 3.6.1**

A function which measures the shape of a function's graph.

The n^{th} moment of a random variable is the expected value of its n^{th} power:

$$n^{th} \text{ moment of } X = \mu_X(n) = E(X^n) = \sum_x x^n p(x)$$

- **First Moment** The expected value.
- **Central Moment** The variance ($E[(X - E(X))^2]$)
- **Standardized Moment** The skew ($\frac{E(X - E(X))^3}{sd(X)^3}$)

Variance **Definition 3.6.2**

The expectation of the deviation from the expected/mean value squared.

$$Var(X) = Var_X(X) = \sigma_X^2 = E[(X - E(X))^2] = E(X^2) - (E(X))^2$$

Note that:

$$Var(a \times X + b) = a^2 Var(X)$$

Standard Deviation **Definition 3.6.3**

The square root of the variance.

$$\sigma_X = sd_X(X) = \sqrt{Var_X(X)}$$

Dice Roll **Example Question 3.6.1**

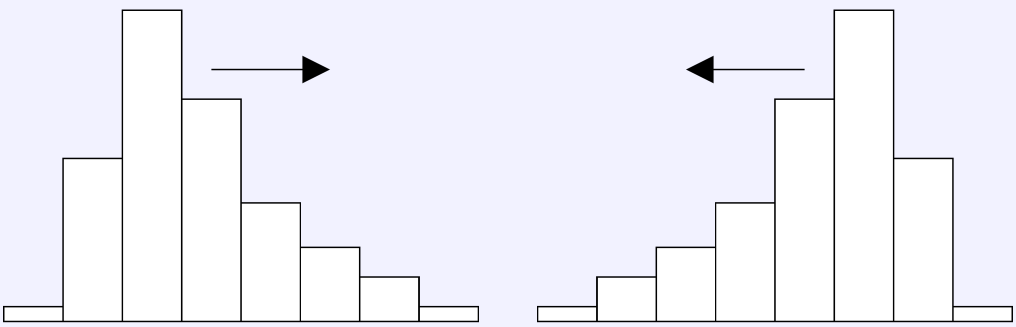
For a random variable representing a dice X:

$$Var(X) = E(X^2) - (E(X^2)) = \sum_x x^2 p(x) - (\sum_x xp(x))^2 = {}^{91}/_{6} - {}^{49}/_{4} = {}^{35}/_{12}$$

Skewness **Definition 3.6.4**

A measure of asymmetry (the standardized moment):

$$\gamma_1 = \frac{E(X - E(X))^3}{sd(X)^3} = \frac{E(X - \mu)}{\sigma^3} \text{ where } \mu = E(X), \sigma = Sd(X)$$



Positive Skew Negative Skew

3.7 Sum of Random Variables

Given random variables $X_1, X_2, \ldots, X_n$ (not necessarily independent, and potentially from different distributions), the sum is:

$$\text{The sum } S_n = \sum_{i=1}^{n} X_i \text{ and the average is } \frac{S_n}{n}$$

(The sum of the outcomes from all random variables)

The expected/mean value of S_n (expected value of the sum of all the random variables) is:

$$E(S_n) = \sum_{i=1}^{n} E(X_i) \text{ and } E(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} E(X_i)}{n}$$

- **All independent**

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) \text{ and } Var(\frac{S_n}{n}) = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2}$$

- **All independent and Identically Distributed**
  Given that for all i, $E(X_i) = \mu_X$ and $Var(X_i) = \sigma_X^2$:

$$E(\frac{S_n}{n}) = \mu_X \text{ and } Var(\frac{S_n}{n}) = \frac{\sigma_X^2}{n}$$

Important Discrete Random Variables

Bernouli Distribution **Definition 3.7.1**

For an experiment with only two outcomes, encoded as 1 and 0.

For$X \sim Bernoulli(p)$ where $x \in S(X) = \{0, 1\}$ and $0 \leq p \leq 1$:

PMF	*Expected*	*Variance*
$p_X(x) = p^x(1-p)^{1-x}$	$\mu = E(X) = p$	$\sigma^2 = Var(X) = p(1-p)$

Binomial Distribution **Definition 3.7.2**

Given n trials with two options, binomial models the number of outcomes. (e.g 3 coin tosses, number of ways to get 2 heads out of total outcomes).

For $X \sim Bionomial(n, p)$ where X takes values $0, 1, 2, \ldots, n$ and $0 \leq p \leq 1$:

PMF	*Expected*	*Variance*	*Skewness*
$p_X(x) = \binom{n}{x} p^x(1-p)^{n-x}$	$\mu = E(X) = np$	$\sigma^2 = Var(X) = np(1-p)$	$\gamma_1 = \frac{1-2p}{\sqrt{np(1-p)}}$

Note that choice is: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

Poisson Distribution **Definition 3.7.3**

Given a constant mean number of events per fixed itme interval, provides probabilities of different numbers of events occuring. (e.g sell on average 6 cookies an hour, what is the probability 10 cookies are sold in a given hour).

For $X \sim Poisson(\lambda)$ where λ is the mean number of events and $\lambda > 0$:

PMF	*Expected*	*Variance*	*Skewness*
$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$	$\mu = E(X) = \lambda$	$\sigma^2 = Var(X) = \lambda$	$\gamma_1 = \frac{1}{\sqrt{\lambda}}$

Note that for poisson the skew is always positive (but decreases as λ increases), and $E(X) \equiv Var(X)$.

Geometric Distribution **Definition 3.7.4**

A potentially infinite number of trials to get an outcome (e.g attempts required to shoot a target, given probability of hit).

We can consider it infinite Bernoulli trials $X_1, X_2, \ldots,$ where $X = \{i | X_i = 1\}$ (X is number of attempts to get outcome 1).

For $X \sim Geometric(p)$ where X takes all values in $\mathbb{Z}^+ = \{1, 2, \ldots\}$ and $0 \le p \le 1$:

PMF	*Expected*	*Variance*	*Skewness*
$p_X(x) = p(1-p)^{x-1}$	$\mu = E(X) = \frac{1}{p}$	$\sigma^2 = Var(X) = \frac{1-p}{p^2}$	$\gamma_1 = \frac{2-p}{\sqrt{1-p}}$

Alternatively we can consider the number of trials *before* getting an outcome:
If $X \sim Geometric(P)$ consider $Y = X - 1$ where Y takes values $\mathbb{N} = \{0, 1, 2, \ldots\}$:

PMF	*Expected*	*Variance*	*Skewness*
$p_Y(x) = p(1-p)^y$	$\mu = E(Y) = \frac{1-p}{p}$	Unchanged	Unchanged

Discrete Uniform Distribution **Definition 3.7.5**

Where a discrete number of outcomes are equally likely (e.g fair dice, colour wheel).

For $X \sim U(\{1, 2, \ldots, n\})$:

PMF	*Expected*	*Variance*	*Skewness*
$p_X(x) = \frac{1}{n}$	$\mu = E(X) = \frac{n+1}{2}$	$\sigma^2 = Var(X) = \frac{n^2-1}{12}$	$\gamma_1 = 0$

3.8 Poisson Limit Theorem

We can use the *Binomial Distribution* to approximate the *Poisson Distribution*:

$$Poisson(\lambda) \approx Binomial(n, p) \text{ when } \lambda = np \text{ and } n \text{ is very large, } p \text{ is very small}$$

This is as for a *Poisson distribution* mean and variance are equal and for binomial, mean is np and variance $np(1-p)$ so as p gets smaller (and n larger) $np \approx np(1-p)$.

Chapter 4

credit