PROBABILITY THEORY

Introduction to Probability, Statistics and Random Processes

Hossein Pishro-Nik

Contents

1 Chapter 1

1.1 Set operations

A union of 2 sets is given by the combination of their elements:

$$A \cup B = \{1,2\} \cup \{2,3\} = \{1,2,3\}$$

The intersection of 2 sets is instead given by their shared elements:

$$A \cap B = \{1,2\} \cap \{2,3\} = \{2\}$$

Theorem 1 (De Morgan's law)**.** *For any sets $A_1, A_2, \ldots, A_n$ we have*

$$\begin{aligned}\overline{A_1 \cup A_2 \cup \ldots \cup A_n} &= \overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n} \\ \overline{A_1 \cap A_2 \cap \ldots \cap A_n} &= \overline{A_1} \cup \overline{A_2} \cup \ldots \cup \overline{A_n}\end{aligned}$$

Theorem 2 (Distributive law)**.** *For any sets A, B and C we have*

$$\begin{aligned}A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C)\end{aligned}$$

The complement of a set is given by all elements that are in the universal set, but not the set itself:

$$S = \{1,2,3,4,5\} \qquad \overline{A} = S \setminus A = \{1,2,3,4,5\} \setminus \{1,2\} = \{3,4,5\}$$

The difference between two sets is given by elements in the first but not the second:

$$A \setminus B = \{1,2\} - \{2,3\} = \{1\} \qquad A \setminus B = A \cap \overline{B}$$

Two sets are disjoint if their intersection is an empty set

$$A \cap B = \emptyset$$

Sets can be partitioned into smaller parts. The sets $A_1, A_2, \ldots, A_n$ are a partition of S if they're disjoint and:

$$\bigcup_{i=1}^{n} A_i = S$$

The cartesian product of two sets are given by the ordered pairs of both sets:

$$A \times B = \{1,2\} \times \{2,3\} = \{(1,2),(1,3),(2,2),(2,3)\}$$

Which can be expressed more generally as:

$$A \times B = \{(x, y) | x \in A \text{ and } y \in B\}$$

The number of elements contained in a (finite) sets is given by its cardinality:

$$|A| = |\{1, 2\}| = 2$$

For determining the cardinality of (finite) sets, the inclusion-exclusion principle is often used:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

This can once again be expanded to more sets:

$$|\bigcup_{i=1}^{n} A_i| = \sum_{i=1}^{n} |A_i| - \sum_{i<j} |A_i \cap A_j|$$
$$+ \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n+1} |A_1 \cap \cdots \cap A_n|$$

1.1.1 Cardinality and countable sets

Finite sets are obviously countable, however when we move onto infinite sets they are divided into countable **and** uncountable sets. A countable set is characterised by the ability to write it in one-to-one correspondance with the natural numbers, e.g.:

$$A = \{a_1, a_2, \ldots, a_n\}$$

Meaning you can list the elements, this is true for sets like the natural numbers, $\mathbb{N}$, and the integers, $\mathbb{Z}$, but also the rational numbers, $\mathbb{Q}$. Uncountable sets (such as the real- and complex numbers) on the other hand cannot be written as lists, but instead have to be denoted as intervals.

Definition 1 (Countability of a set)**.** A set, A, is called countable if one of the following is true:

- It is a finite set, $|A| < \infty$.
- The set can be written as a list with one-to-one correspondance with the natural numbers.

This means that any subset of $\mathbb{N}, \mathbb{Z}$ and $\mathbb{Q}$ are countable, whilst any set containing an interval on the real line is uncountable.

Theorem 3 (Countability of sub- and supersets)**.** *Any subset of a countable set is countable and any superset of an uncountable set is uncountable.*

Proof. Let A be a countable set and $B \subset A$. If A is finite, then it follows that $|B| \leq |A| < \infty$, thus B must be countable as its cardinality cannot exceed that of A, which must be smaller than ∞.

If A is instead countably infinite, then it follows that as B is a subset of A it must be possible to construct it by removing $\overline{B}$ from A, whereby it must also be countable, as it can be constructed as a list.

The opposite can be argued by assuming B is **not** countable, whereby a contradiction would occur in both proofs. □

Theorem 4 (Countability of union). *If $A_1, A_2, \ldots, A_n$ are countable sets, then the union of those must also be countable.*

Proof. As the sets are countable it must be possible to write them in the form

$$A_1 = \{a_{11}, a_{12}, \ldots, a_{1n}\}$$
$$A_2 = \{a_{21}, a_{22}, \ldots, a_{2n}\}$$
$$A_3 = \{a_{31}, a_{32}, \ldots, a_{3n}\}$$

As such the union of those sets must also be possible to construct as a list

$$\bigcup_{i=1}^{m} A_i = \{a_{11}, a_{12}, a_{21}, a_{22}, a_{31}, a_{32}, \ldots, a_{mn}\}$$

And as a result must be countable. □

Theorem 5 (Countability of carthesian product). *If A and B are countable, then $A \times B$ is also countable.*

Proof. As A and B are countable it must be possible to write them in the form

$$A = \{a_1, a_2, \ldots, a_n\}$$
$$B = \{b_1, b_2, \ldots, b_n\}$$

In accordance with the definition of the carthesian product, the two sets can be constructed as a list with the form

$$A \times B = \{(a_i, b_j) | i, j \in \mathbb{N}\}$$

Whereby it must be countable as it can be constructed as a list. □

As a result of this proof it also becomes clear that any set that can be written in the form

$$C = \bigcup_i \bigcup_j \{a_{ij}\} \text{ where } i, j \text{ belong to a countable set}$$

Must also be countable, the set of rational numbers is an example of this as it can be written as

$$\mathbb{Q} = \bigcup_{i \in \mathbb{Z}} \bigcup_{j \in \mathbb{N}} \left\{ \frac{i}{j} \right\}$$

1.1.2 Functions

Functions take an input from its domain, apply a rule to said input, whereby an output from the co-domain is produced.

$$f : A \to B \qquad f(x \in A) \in B$$

Definition 2 (Function)**.** A function maps elements from the domain set to elements in the co-domain with the property that each input is mapped to exactly one output.

In the same context the range operand is defined, as it is not necessary for a function to be able to output all elements of the codomain:

$$f : \mathbb{R} \xrightarrow{x^2} \mathbb{R}$$

Here both the domain- and co-domain are the real numbers, however it is clear that no value $x \in \mathbb{R}$ would ever produce a negative number, therefore:

$$\text{Range}(f) = \mathbb{R}^+$$

1.2 Random experiments

A random experiment will always have an **outcome** corresponding to an element from the **sample space**, S.

Definition 3 (Random experiment)**.** A random experiment is a process by which we observe something uncertain.

When a random experiment is repeated, each repetition is called a **trial**. The goal of analyzing a random experiment is to assign probabilities to **events**, which correspond to subsets of the sample space.

1.2.1 Probability

A probability is assigned to an event, $P(A) \in [0, 1]$. The derivation of probability theorem is based on 3 axioms:

- Axiom 1: For any event A, $1 \geq P(A) \geq 0$.
- Axiom 2: Probability of the sample space, S, is $P(S) = 1$.
- Axiom 3: If $A_1, A_2, \ldots, A_n$ are disjoint events, then $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$

Notationally unions and intersections can be read as:

$$\begin{aligned} P(A \cap B) &= P(A \text{ and } B) = P(A, B) \\ P(A \cup B) &= P(A \text{ or } B) \end{aligned}$$

Theorem 6 (Probability of complement)**.** *For any event A, $P(\overline{A}) = 1 - P(A)$.*

Proof. As the complement of a set contains all elements of the sample space that are not in the set

$$\overline{A} = S \setminus A$$

It is clear that their unions must be S and they must be disjoint whereby

$$P(A \cup \overline{A}) = P(S) = 1$$

As they are disjoint we can write the probability of their union as the sum of their probabilities

$$P(A) + P(\overline{A}) = 1 \Leftrightarrow P(A) = 1 - P(\overline{A})$$

□

Theorem 7 (Probability of empty set)**.** *The probability of the empty is zero, $P(\emptyset) = 0$.*

Proof. As the empty set must be the complement of the sample space we have that

$$P(\emptyset) = P(\overline{S}) = 1 - P(S) = 1 - 1 = 0$$

□

Theorem 8 (Probability must be equal to or less than 1)**.** *For any event A, $P(A) \leq 1$.*

Proof. By the first axiom we have that

$$P(\overline{A}) \geq 0$$

It becomes clear that

$$P(A) \leq 1$$

As $P(A) + P(\overline{A}) = 1$. □

Theorem 9 (Probability of a difference)**.** *The probability of a difference is given by $P(A \setminus B) = P(A) - P(A \cap B)$.*

Proof. As $A \cap B$ and $A \setminus B$ must be disjoint, whilst their union must be A

$$(A \cap B) \cup (A \setminus B) = A$$

We have by the third axiom that

$$P(A) = P((A \cap B) \cup (A \setminus B)) = P(A \cap B) + P(A \setminus B)$$

By rearranging it becomes clear that

$$P(A \setminus B) = P(A) - P(A \cap B)$$

□

Theorem 10 (Probability of a union)**.** *The probability of a union is given by $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

Proof. As A and $B \setminus A$ must be disjoint sets whilst their union must be $A \cup B$, it is clear that

$$P(A \cup B) = P(A \cup (B \setminus A))$$

As we know these are disjoint we write

$$P(A \cup B) = P(A) + P(B \setminus A)$$

Rewriting using the previous theorem we then have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

□

Theorem 11 (Probability of a subset must be less than or equal to its superset). *If $A \subset B$ then $P(A) \leq P(B)$.*

Proof. As $A \subset B$ it is clear that their union must be B

$$P(B) = P(A \cap B) + P(B \setminus A)$$

As their intersection is A we have that

$$P(B) = P(A) + P(B \setminus A)$$

As

$$P(B \setminus A) \geq 0$$

By the first axiom, we have that

$$P(B) \geq P(A)$$

$\square$

1.3 Conditional probability

Conditional probabilities are written as

$$P(A|B)$$

And are read as "*the probability of A, given that B has occurred*". The conditional probability will therefore be given by

$$\begin{aligned} P(A|B) &= \frac{|A \cap B|}{|B|} \\ &= \frac{\frac{|A \cap B|}{|S|}}{\frac{|B|}{|S|}} \text{ dividing by } |S| \\ &= \frac{P(A \cap B)}{P(B)} \end{aligned}$$

This is as when we know B has occured, the sample space of A is shrunk to B, whereby the cardinality of their intersection must be equal to the amount of favourable outcomes.

The earlier established probability axioms can also be formulated for conditional probabilities

- Axiom 1: For any event A, $P(A|B) \geq 0$.
- Axiom 2: Conditional probability of B given B is 1, i.e., $P(B|B) = 1$.
- Axiom 3: If $A_1, A_2, \ldots, A_n$ are disjoint events, then $P(\bigcup_{i=1}^{n} A_i|B) = \sum_{i=1}^{n} P(A_i|B)$.

And the same applies to the established fomulas

Theorem 12. *For any conditional event, $A|C$, $P(\overline{A}|C) = 1 - P(A|C)$.*

Proof. We know that

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

Assuming the theorem is correct we then have that

$$1 - P(\overline{A}|C) = 1 - \frac{P(\overline{A}|C)}{P(B)}$$

We wish to show that

$$\frac{P(A|B)}{P(B)} = 1 - \frac{P(\overline{A}|C)}{P(B)}$$

Multiplying by $P(B)$ gets us

$$P(A|B) = P(B) - P(\overline{A}|C)$$

Adding $P(\overline{A}|C)$ on the LHS we get

$$P(A|B) + P(\overline{A}|C) = P(B)$$

As the two are mutually exclusive by the definition of the complement

$$P((A|B) \cup (\overline{A}|C)) = P(B)$$

Which is true. □

Theorem 13. *The probability of the empty set is zero, $P(\emptyset|C) = 0$.*

Proof. As the empty is the complement of the sample set we have from the previous theorem that

$$P(S|C) = 1$$

By applying previous equation

$$P(\overline{S}|C) = 1 - 1 = 0$$

□

Theorem 14. *The probability of a conditional probability occuring must always be less than or equal to 1, $P(A|C) \leq 1$.*

Proof. From the definition of conditional probability

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

It is clear that the denominator and numerator can never be more than even as $A \cap B = B$ even if AB, resulting in the maximum value being 1.

$$P(A|B) \leq 1$$

□

Theorem 15. *The probability of a difference is given by $P(A \setminus B|C) = P(A|C) - P(A \cap B|C)$.*

Proof. As $A \cap B|C$ and $A \setminus B|C$ must be disjoint whilst their union must be $A|C$, we have that

$$(A \cap B|C)(A \setminus B|C) = A|C$$

By the third axiom we have that

$$P(A|C) = P(A \cap B|C) + P(A \setminus B|C)$$

Rearranging the terms we get

$$P(A \setminus B|C) = P(A|C) - P(A \cap B|C)$$

□

Theorem 16. *The probability of a union is given by $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.*

Proof. As $A|C$ and $B \setminus A|C$ must be disjoint and their union must be equal to $A \cup B|C$ we have that

$$P(A \cup B) = P(A|C(B \setminus A|C))$$

As these sets are disjoint we rewrite using the third axiom

$$P(A \cup B) = P(A|C) + P(B \setminus A|C)$$

By the previous theorem the difference is rewritten as

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

□

Theorem 17. *If $A \subset B$ then $P(A|C) \leq P(B|C)$.*

Proof. As $A \subset B$ it is clear that their union must be B

$$P(B|C) = P(A \cap B|C) + P(B \setminus A|C)$$

Since their intersection is A due to it being the subset

$$P(B|C) = P(A|C) + P(B \setminus A|C)$$

By the first axiom

$$P(B \setminus A|C) \geq 0$$

As such

$$P(B|C) \geq P(A|C)$$

□

This introduces some special cases

$$P(A|B) = \frac{P(\emptyset)}{P(B)} = 0, \text{ for } A \cap B = \emptyset$$
$$P(A|B) = \frac{P(B)}{P(B)} = 1, \text{ for } B \subset A$$
$$P(A|B) = \frac{P(A)}{P(B)}, \text{ for } A \subset B$$

Furthermore we also write the chain rule for conditional probability using the definition as a starting point

Theorem 18. *The extended chain rule is given by $P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1) \cdots P(A_n|A_{n-1}, A_{n-2}, \ldots, A_1)$*

Proof. From the definition of conditional probability we have that

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

By isolation for $P(A \cap B)$ we get

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(B)$$

Extending to 3 or more events we get that

$$P(A \cap B \cap C) = P(A \cap (B \cap C)) = P(A)P(B \cap C|A)$$

Applying the first equation

$$P(B \cap C) = P(B)P(C|B)$$

By conditioning both sides on A we get

$$P(B \cap C|A) = P(B|A)P(C|A, B)$$

Inserting in the original equation we then get

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A, B)$$

Which can be generalised to

$$\begin{gathered}P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1)\cdots \\ P(A_n|A_{n-1}, A_{n-2}, \ldots, A_1)\end{gathered}$$

□

1.3.1 Independence

Conditional probabilities are only relevant if two events are not independent.

Definition 4 (Independence)**.** Two events A, B are independent if $P(A \cap B) = P(A)P(B)$.

Independence for two or more events requires that all the individual events are independents, as well as all of them together, this means that for 3 events, all of the following must hold

$$\begin{aligned}P(A \cap B) &= P(A)P(B) \\ P(A \cap C) &= P(A)P(C) \\ P(B \cap C) &= P(B)P(C) \\ P(A \cap B \cap C) &= P(A)P(B)P(C)\end{aligned}$$

Theorem 19. *If A and B are independent, then A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$ are also independent.*

Proof. The first statement is proven as the others can be concluded from it. As the statement is equivalent to

$$P(A \cap \overline{B}) = P(A) - P(A \cap B)$$

As we know A and B are independent we have

$$P(A \cap \overline{B}) = P(A) - P(A)P(B)$$

We factor out $P(A)$ and get that

$$P(A \cap \overline{B}) = P(A)(1 - P(B))$$

And as

$$1 - P(B) = P(\overline{B})$$

It becomes clear that

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

□

To determine the probability of several unions of independent events, we make use of De Morgans law

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - P\left(\bigcap_{i=1}^{n} \overline{A_i}\right)$$

Which is equivalent to

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n}(1 - P(A_i))$$

1.3.2 Law of Total Probability

The law of total probability states that the probability of an event, A must be the sum of the probability of it occuring in every partition, $B_1, B_2, \ldots, B_n$ of the sample space

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

Proof. As $B_1, B_2, \dots, B_n$ are partitions of the sample space we write

$$S = \bigcup_{i=1}^{n} B_i$$

Now the event A occuring must be given by its intersection of the sample space

$$A = A \cap \left(\bigcup_{i=1}^{n} B_i \right)$$

By the distributive property it becomes clear that

$$A = \bigcup_{i=1}^{n} (A \cap B_i)$$

Now as the partitions by definition are disjoint we can determine the probability as the sum of probabilities

$$P(A) = P\left(\bigcup_{i=1}^{n} (A \cap B_i) \right) = \sum_{i=1}^{n} P(A \cap B_i)$$

Rewriting using the definition of conditional probability (as $A \in B_i$ can only occur if B_i has occured) we get

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

□

Theorem 20. *Bayes rule states that:* $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

Proof. From the definition of conditional probability

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

Multiplying by $P(B)$ on both sides

$$P(A|B)P(B) = P(A \cap B)$$

Dividing by $P(A)$ on both sides

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

□

1.3.3 Conditional Independence

Definition 5 (Conditional independence)**.** Two events A and B are conditionally independent given an event C if $P(A \cap B|C) = P(A|C)P(B|C)$.

Proof. From the definition of conditional probability we have

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

Conditioning both sides on C it becomes apparent that

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

Assuming that A and B are conditionally independent we have

$$P(A|B, C) = \frac{P(A|C)P(B|C)}{P(B|C)} = P(A|C)$$

□

2 Chapter 2

2.1 Counting

For a finite sample space with equal probabilities we recall that

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

As such determining the probability is a counting problem, determining the cardinality of A and S.

Definition 6 (Multiplication principle)**.** Suppose that we perform r experiments such that the k'th experiment has n_k possible outcomes for $k = \{1, 2, \dots, r\}$. Then there are a total of $n_1 \times n_2 \times \dots \times n_r$ possible outcomes for the sequence of r experiments.

For counting problems, some general terminology is relevant:

- Sampling: Choosing a random element from a set.
- With replacement, the sampled element is returned to the set and can therefore be drawn multiple times with repeated sampling.
- Without replacement, the sampled element is not returned to the set and can therefore not be drawn multiple times.
- Ordered means that the order at which elements are written matters, $\{a_1, a_2, a_3\} \neq \{a_3, a_1, a_2\}$.
- Unordered means that the order at which elements are written does not matter, $\{a_1, a_2, a_3\} = \{a_3, a_1, a_2\}$.

2.1.1 Ordered with replacement

Suppose we have a set consisting of n elements and we wish to draw k samples from the set, for example, say $A = \{1, 2, 3\}$ where we wish to sample $k = 2$, we then get 9 different possibilities

$$(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)$$

From this its clear that we create a list consisting of k-valued elements where each position has n options for values

$$\begin{array}{cccc} a_1 & a_2 & \dots & a_k \\ \uparrow & \uparrow & & \uparrow \\ n & n & & n \end{array}$$

Meaning that we can determine the total amount of possibilities as

$$n \times n \times \ldots \times n = n^k$$

2.1.2 Ordered without replacement (Permutations)

As opposed to the previous circumstance, a element is now removed every time we draw, resulting in there being one less options every time we move to the next position, as such, using the same example with A and $k = 2$ we get

$$(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)$$

Here we create a list of k-valued elements where each position has one less option than the previous

$$\begin{array}{cccc} (a_1 & a_2 & \ldots & a_k) \\ \uparrow & \uparrow & & \uparrow \\ n & n-1 & & n-k+1 \end{array}$$

Which is called a k-permutation of the elements in the set, the following notation is used to show the number of k-permutations of an n-element set

$$P^n_k = n \times (n-1) \times \ldots \times (n-k+1)$$

A special case can also occur here when $n < k$, as there then wont be enough options for every position and there therefore are no possible lists. Another special case is an n-permutation where $k = n$ which results in the sequence

$$\begin{aligned} P^n_n &= n \times (n-1) \times (n-2) \times \ldots \times (n-n+1) \\ &= n \times (n-1) \times (n-2) \times \ldots \times 1 \\ &= n! \end{aligned}$$

As such the factorial operator simply denotes the total number of permutations of an n element set, aka the total number of ways you can order n different objects. By definition $0! = 1$, using this we rewrite the formula for P^n_k

Theorem 21. *The amount of k-permutations of an n-element set is given by $P^n_k = \frac{n!}{(n-k)!}$.*

Proof. From the original expression for P^n_k we have that

$$P^n_k = n \times (n-1) \times (n-2) \times \ldots \times (n-k+1)$$

By multiplying by $\frac{(n-k)!}{(n-k)!}$ we get

$$
\begin{aligned}
&= n \times (n-1) \times (n-2) \times \ldots \times (n-k+1) \cdot \frac{(n-k)!}{(n-k)!} \\
&= \frac{n!}{(n-k)!}
\end{aligned}
$$

As multiplying by $(n-k)!$ in the numerator results in the sequence "finishing" and being equal to $n!$. □

2.1.3 Unordered without replacement (Combinations)

We now wish to determine the amount of possible lists when we sample k elements from an n-element set. This means that we want to determine the amount of possible k-element subsets of the n-element set. Using the same example as before with A and $k = 2$ we get 3 different combinations

$$(1, 2), (1, 3), (2, 3)$$

We show the number of k-element subsets of A as

$$\binom{n}{k}$$

Which is read as "n choose k", to determine the value of this we compare with P_k^n as the only difference between the two is ordering. This is because for any k-element subset of an n-element set, we can order the elements in $k!$ different ways, as such

$$P_k^n = \binom{n}{k} \times k!$$

Rewriting using our previously established formula for P_k^n and dividing by $k!$ we get

$$\binom{n}{k} = \frac{n!}{k!(n-k)}, \text{ for } 0 \leq k \leq n$$

This term is used extensively in the binomial theorem, which states that

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Theorem 22. *For any non-negative integers n, k it follows that* $\binom{n}{k} = \binom{n}{n-k}$.

Proof. Assume we wish to determine the amount of possible sequences consisting of k A's and j B's, as such we have $n = j + k$ positions to fill with either A or B, from these positions we need to choose j for A's and whatever is left is filled with B's, as such the amount of ways is

$$\binom{n}{j}$$

If we instead observe this from the point of B's, it is clear that the amount of ways would then be given by

$$\binom{n}{k}$$

As these must be equivalent we have that

$$\binom{n}{j} = \binom{n}{k}$$

From the initial determination of n we get that

$$n = j + k \implies j = n - k \vee k = n - j$$

As such

$$\binom{n}{j} = \binom{n}{n-k} = \binom{n}{k} = \binom{n}{n-j}$$

□

Theorem 23. *For any non-negative integers k, n it follows that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.*

Proof. From the Binomial theorem we know that

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

We let $a = b = 1$ and as such get

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

□

Theorem 24. *For non-negative integers $0 \leq k \leq n$ it follows that* $\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$.

Proof. We define an arbitrary set, A with $n+1$ elements

$$A = \{a_1, a_2, \ldots, a_n, a_{n+1}\}$$

From this set we wish to choose a $k+1$ element subset, call it B, by combinations we know this is equal to

$$\binom{n+1}{k+1}$$

B can also be constructed as the union of two subsets of B that are defined by either containing- or not containing a_{n+1}

$$B = B_1 \cup B_2, \text{ where } a_{n+1} \notin B_1, a_{n+1} \in B_2, B_1 \cap B_2 = \emptyset$$

To define B_1 we need to choose $k+1$ elements from the set $A \setminus a_{n+1}$ which is equal to

$$\binom{n}{k+1}$$

To complete the set we then need to choose k elements from A which can be done in

$$\binom{n+1}{k}$$

ways. As such the sum of the two must be equal to the initial expression resulting in

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k+1}$$

□

Theorem 25. *Vandermonde's identity states that* $\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i} \binom{n}{k-i}$

Proof. We construct a set A with $m+n$ elements, as such

$$A = \{a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_n\}$$

Determining the number of k-element subsets of A is equal to

$$\binom{m+n}{k}$$

This can also be done by choosing i elements from $\{a_1, a_2, \ldots, a_n$ first, and then $k-i$ elements from $\{b_1, b_2, \ldots, b_n\}$, which then can be done in

$$\binom{m}{i}\binom{n}{k-i}$$

ways. But as i can be any number from $0 \to k$ it is necessary to sum all the possible options whereby we write

$$\binom{m+n}{k} = \sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i}$$

□

An important class of random experiments are Bernoulli trials, a random experiemnt where there are two possible outcomes, succes and failure (which can be extended to any experiment with 2 outcomes as we can arbitrarily define success and failure).

In Bernoulli trials the probability of success if usually denoted by p and the probability of failure as its complement $q = 1 - p$. If we perform n independent Bernoulli trials and count the number of successes, it is called a binomial experiment, for example a coin toss where we define success as heads and count the number of heads.

Theorem 26. *The binomial formula is given by* $P(k) = \binom{n}{k} p^k (1-p)^{n-k}$

Proof. Imagine we toss a coin with $P(H) = p$ and $P(T) = 1 - p$ n times, we define C as the event of observing k heads (and $n-k$ tails), the probability of observing k heads will be given by

$$P(k) = |C| p^k (1-p)^{n-k}$$

To determine $|C|$, we realise that we can see the event as ordered sampling without replacement, as such we have that

$$|C| = \binom{n}{k}$$

Which we insert into the previous expression and get

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

□

Theorem 27. *If we have n people and wish to seperate them into r groups, the number of ways this is possible is given by the multinomial coefficients,* $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$.

Proof. Say we have n people we wish to divide into r groups, using combinations, this would be given as

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-\ldots-k_{r-1}}{k_r} = \frac{n!}{k_1!(n-k_1)!}\cdot\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}$$
$$\cdots\frac{(n-k_1-\ldots-k_{r-1})!}{k_r!(n-k_1-\ldots-k_{r-1}-k_r)!} = \frac{n!}{k_1 k_2! \ldots k_r!}$$

The simplification of this expression results from the following term always cancelling the "excess" in the previous, until we reach the final term where the numerator cancels the excess term from the previous fraction, whilst the denominator equals one as $n = \sum_{i=1}^{r} k_i \implies (n - k_1 - \ldots - k_r)! = (0)! = 1$ □

2.1.4 Unordered without replacement

As opposed to the previous section we're now working with replacement, as such we again use the example of A with $k = 2$ and get 6 possibilities given by

$$(1,1),(1,2),(1,3),(2,2),(2,3),(3,3)$$

One way to represent this sample is to list them as n-element vectors where each position corresponds to a number, eg.

$$\begin{aligned}
(a,b) &\to (x_1,x_2,x_3) = (n_1,n_2,n_3)\\
(1,1) &\to (x_1,x_2,x_3) = (2,0,0)\\
(1,2) &\to (x_1,x_2,x_3) = (1,1,0)\\
(1,3) &\to (x_1,x_2,x_3) = (1,0,1)
\end{aligned}$$

Constructing these vectors a pattern emerges, that is $\sum_{i=1}^{3} x_i = 2$, as such we can determine the amount of possibilities as the amount of integer solutions to $x_1 + x_2 + x_3 = 2$.

Theorem 28. *The number of distinct solutions to the equation* $x_1+x_2+\ldots+x_n = k$ *where* $x_n \in \{0, 1, 2, 3, \ldots\}$ *is equal to* $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$.

Proof. We first define a mapping where an integer $x_n \geq 0$ is replaced with x_n vertical lines. Suppose we now have a solution to the equation where $k = 6$, for example

$$3 + 0 + 2 + 1 \Leftrightarrow ||| + +|| + |$$

Here we realise that the equation can be represented by k vertical lines and $n - 1$ plus signs, as such we can use combinations to determine the amount of distinct sequences we can create using k vertical lines and $n-1$ plus signs, as such

$$n_{\text{solutions}} = \binom{k+n-1}{k} = \binom{k+n-1}{n-1}$$

□

3 Chapter 3

3.1 Random variables

Definition 7 (Random variable)**.** A random variable is a real-valued value determined by an underlying random experiment. Mathematically this can be expressed as a function that assigns a value to a possible outcome, $S \xrightarrow{X} \mathbb{R}$.

3.1.1 Discrete random variables

What defines whether at random variable is discrete, continuous or mixed is it's range. If the range of the random variable is strictly countable, the random variable is said to be discrete, this for example occurs if we wish to determine the amount of heads in a coin throw.

3.1.2 Probability mass function

Let X be a discrete random variable, as such

$$R_X = \{x_1, x_2, x_3, \dots\}$$

As the possible outcomes are discrete, they must each have their own probability of occuring, we define the event

$$A = \{s \in S | X(s) = x_i\}$$

Where the probabilities of each individual event $\{X = x_i\}$ is given by the probability mass function of X

$$P_X(x_i) = P(X = x_i), \text{for} i = 1, 2, 3, \dots$$

To ensure the function is well-defined, the definition of the mass probability function is frequently extended to all real numbers by a piecewise function

$$P_X(x) = \begin{cases} P(X = x) & x \in R_X \\ 0 & x \notin R_X \end{cases}$$

As this is a plottable discrete function, this is also called the probability mass function of the random variable X.

3.1.3 Independent random variables

Random variables can, like probabilities and conditional probabilities, be independent. Independence is defined in a similar way to in those examples, 2 random variables, X and Y, are independent if

$$P(X = x \cap Y = y) = P(X = x)P(Y = y), \text{ for all } x, y$$

This means that if they are independent we can write

$$P(X \in A \cap Y \in B) = P(X \in A)P(Y \in B), \text{ for all sets } A \text{ and } B$$

We can also condition random variables on eachother, for two independent random variables it then follows that

$$P(X = x|Y = y) = P(X = x), \text{ for all } x, y$$

As knowing the outcome of Y does not impact the probability of X.

3.2 Special distributions

3.2.1 Bernoulli distribution

We recall that a Bernoulli experiment is an experiment with only two possible outcomes, success or failure, defining $x = 1$ as success and $x = 0$ as failure, we get the Bernoulli PMF

$$X \sim Bernoulli(p) = P_X(k) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

3.2.2 Geometric distribution

For a geometric experiment, that being a Bernoulli experiment where we repeat the experiment until one of the two outcomes is observed has the probability $P(A) = p(1-p)^{k-1}$, where p is the probability of observing what we want, as such the PMF is given by

$$X \sim Geometric(p) = P_X(k) = \begin{cases} p(1-p)^{k-1} & k \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

3.2.3 Binomial distribution

We recall that the binomial theorem states that

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

As such the PMF is given by

$$X \sim Binomial(n,p) = P_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k \in \{0,1,2,\dots,n\} \\ 0 & \text{otherwise} \end{cases}$$

Simultaneously we can observe each cointoss that makes up the sequence as individual Bernoulli distributions, as such we can express a binomial as the sum of a sequence of Bernoulli experiments, i.e. we have the PMF X defined as the binomial giving the number of heads, which can be expressed as

$$X = \sum_{i=1}^{n} X_i$$

Where i is the ith cointoss in the sequence as given by the Bernoulli PMF.

Theorem 29. *Given two independent random variables described by binomial distributions, X, Y, with $k_1 = n$ and $k_2 = m$ the random variable defined as $Z = X + Y$ is given by*

$$P_Z(k) = \begin{cases} \binom{n+m}{k} p^k (1-p)^{n+m-k} & k = \{0,1,2,\dots,m+n\} \\ 0 & \textit{otherwise} \end{cases}$$

Proof. As the random variable can be described as the sum of the individual Bernoulli random variables we rewrite that

$$\begin{aligned} Z &= X + Y \\ &= X_1 + X_2 + \ldots + X_n + Y_1 + Y_2 + \ldots + Y_n \end{aligned}$$

A way to express this as probabilities is

$$
\begin{aligned}
P_Z(k) &= P(Z = k) \\
&= P(X + Y = k) \\
&= \sum_{i=0}^{n} P(X + Y = k | X = i) P(X = i) \\
&= \sum_{i=0}^{n} P(X = k - i) P(X = i) \\
&= \sum_{i=0}^{n} \binom{m}{k-i} p^{k-i} (1-p)^{m-k+i} \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} \binom{m}{k-i} \binom{n}{i} p^k (1-p)^{m-k+n} \\
&= p^k (1-p)^{m-k+n} \sum_{i=0}^{n} \binom{m}{k-i} \binom{n}{i} \\
&= \binom{m+n}{k} p^k (1-p)^{m-k+n} \\
&= \begin{cases} \binom{m+n}{k} p^k (1-p)^{m-k+n} & k \in \{0, 1, 2, \ldots, m+n\} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

$\square$

3.2.4 Negative binomial (Pascal) distribution

The Pascal distribution is a generalization of the geometric distribution where we wish to observe m successes instead of just 1, by this it follows that a Pascal distribution with $m = 1$ is simply a geometric distribution.

Suppose we toss a coin until m heads are observed, where X is the total number of tosses, in this context we define the event $A = \{X = k\}$ as the event that m heads are observed in k tosses. We rewrite this event as the intersection $A = B \cap C$ where

- B is the event that $m - 1$ heads are observed in the first $k - 1$ tosses.
- C is the event that heads is observed in the kth toss.

As these two events are independent (seperate coin tosses), we have that

$$P(A) = P(B \cap C) = P(B)P(C)$$

We can determine these probabilities. B is given by a binomial distribution whilst C is a constant value, as such we write that

$$\begin{aligned} P(B) &= \binom{k-1}{m-1} p^{m-1}(1-p)^{(k-1)-(m-1)} \\ &= \binom{k-1}{m-1} p^{m-1}(1-p)^{k-m} \\ P(C) &= p \end{aligned}$$

Since A is given by the product of these, we multiply them together whereby we get

$$\begin{aligned} P(A) &= P(B)P(C) \\ &= \binom{k-1}{m-1} p^{m-1}(1-p)^{k-m} p \\ &= \binom{k-1}{m-1} p^{m}(1-p)^{k-m} \end{aligned}$$

Whereby we can write the PMF for the Pascal distribution as

$$X \sim Pascal(m,p) = P_X(k) = \begin{cases} \binom{k-1}{m-1} p^m (1-p)^{k-m} & k \in \{m, m+1, m+2, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

3.2.5 Hypergeometric distribution

Imagine a bag containing b blue marbles and r red ones. Choosing $k \leq b+r$ marbles at random without replacements will at best give us $X \leq \min(k,b)$ blue marbles, whilst the number of red marbles in our sample must be $X \geq \max(0, k-r)$. As such we can define the range of X as

$$R_X = \{\max(0,k-r), \max(0,k-r)+1, \ldots, \min(k,b)\}$$

Simultaneously we can create an expression for determining the probability of getting x blue marbles (and $k-x$ red ones) by

$$\frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}}, \text{ for } x \in R_X$$

Using this expression we create the PMF as

$$X \sim Hypergeometric(b, r, k) = P_X(x) = \begin{cases} \dfrac{\dbinom{b}{x}\dbinom{r}{k-x}}{\dbinom{b+r}{k}} & x \in R_X \\ 0 & \text{otherwise} \end{cases}$$

3.2.6 Poisson distribution

The Poisson distribution is tupically used to model occurences in an interval using a variable, λ, as the known average.

Theorem 30. *Let X be a binomial PMF with $p = \frac{\lambda}{n}, \lambda > 0$. Then for any $k \in \{0, 1, 2, \dots\}$ we have that $\lim_{n\to\infty} P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$.*

Proof. The binomial will be given by

$$P_X(k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Taking the limit of this binomial we get

$$\begin{aligned} \lim_{n\to\infty} P_X(k) &= \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lambda^k \lim_{n\to\infty}\frac{n!}{k!(n-k)!}\left(\frac{1}{n^k}\right)\left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!}\lim_{n\to\infty}\frac{n!}{(n-k)!}\left(\frac{1}{n^k}\right)\left(1 - \frac{\lambda}{n}\right)^{n}\left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(\frac{n(n-1)\dots(n-k+1)}{n^k}\right)\left(1 - \frac{\lambda}{n}\right)^{n}\left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

We now take the limit of each of these expressions individually

$$\begin{aligned} \lim_{n\to\infty}\frac{n(n-1)\dots(n-k+1)}{n^k} &= 1 \\ \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{n} &= e^{-\lambda} \\ \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{-k} &= 1 \end{aligned}$$

Which gives us

$$\begin{aligned} P_X(k) &= \frac{\lambda^k}{k!} \cdot 1 \cdot 1 \cdot e^{-\lambda} \\ &= \frac{e^{-\lambda}\lambda^k}{k!} \end{aligned}$$

□

As such we can define the Poisson distributions PMF as

$$X \sim Poisson(\lambda) = P_X(k) = \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!} & k \in R_X \\ 0 & \text{otherwise} \end{cases}$$

3.3 Cumulative distribution function

As opposed to the discrete distribution function the cumulative distribution function describes the probability distribution on a continuous line.

Definition 8. The cumumlative distribution functions of a random variable X is defined as $F_X(x) = P(X \leq x)$, for all $x \in \mathbb{R}$.

A CDF can also be made for discrete random variables, this is done by extrapolating the range for which a probability is true to the real line, see for example a coin toss given by a binomial PMF with $p = \frac{1}{2}, n = 2$

$$\begin{aligned} P_X(k) &= \binom{n}{k}\left(\frac{1}{2}\right)^k\left(1-\frac{1}{2}\right)^{n-k} \\ P_X(0) &= \frac{2!}{0!(2-0)!} \cdot 1 \cdot \frac{1}{4} = \frac{1}{4} \\ P_X(1) &= \frac{2!}{1!(2-1)!} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} \\ P_X(2) &= \frac{2!}{2!(2-2)!} \cdot \frac{1}{4} \cdot 1 = \frac{1}{4} \end{aligned}$$

To convert this to a CDF we extrapolate the range for which the probabilities are valid

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4} & x \leq x < 1 \\ \frac{3}{4} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Here we see that the CDF is equal to summing up the probabilities gradually

$$F_X(x) = \sum_{x_k \leq x} P_X(x_k)$$

Theorem 31. *For all $a \leq b$ we have that $P(a < X \leq b) = F_X(b) - F_X(a)$.*

Proof. For $a \leq b$ we have that

$$P(X \leq b) = P(x \leq a) + P(a < X \leq b)$$

Rewriting this as CDFs we have that

$$F_X(b) = F_X(a) + P(a < X \leq b)$$

Using simple algebraic manipulation we then get

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

□

3.3.1 Expectation

A random variable has an expected value, also known as othe mean, which is the weighted average of all values in the range.

Definition 9. Let X be a discrete random variable with $R_X = \{x_1, x_2, \ldots\}$ (finite or countably infinite), the expected value of X denoted EX is then defined as

$$EX = \sum_{x_k \in R_X} x_k P(X = X_k) = \sum_{x_k \in R_X} x_k P_X(x_k)$$

The expected value of a random variable is the value which we expect the mean to take on as the number of trials approaches infinity, other ways to denote the expected value is

$$EX = E[X] = E(X) = \mu_X$$

This lets us establish models for the expected value of different known distribution functions

Bernoulli expected value The expected value of the Bernoulli is given by

$$\begin{aligned} EX &= 0P_X(0) + 1P_X(1) \\ &= 0 \cdot (1-p) + 1 \cdot (p) \\ &= p \end{aligned}$$

Geometric expected value The expected value of the geometric is given by

$$\begin{aligned} EX &= \sum_{x_k \in R_X} x_k P_X(x_k) \\ &= \sum_{k=1}^{\infty} k q^{k-1} p \\ &= p \sum_{k=1}^{\infty} k q^{k-1} \end{aligned}$$

We make use of the geometric sum formula to determine the value of this

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} \text{ for } |x| < 1$$

Our sum is however of a different form, however we realise we can take the derivative with respect to x on both sides and get

$$\begin{aligned} \frac{d}{dx} \sum_{k=0}^{\infty} x^k &= \frac{d}{dx} \frac{1}{1-x} \\ \sum_{k=0}^{\infty} k x^{k-1} &= \frac{1}{(1-x)^2} \end{aligned}$$

As such we can write that

$$\begin{aligned} EX &= p \sum_{k=1}^{\infty} k q^{k-1} \\ &= p \frac{1}{(1-q)^2} \\ &= p \frac{1}{p^2} \\ &= \frac{1}{p} \end{aligned}$$

Poisson expected value The expected value of the Poisson is given by

$$
\begin{aligned}
EX &= \sum_{x_k \in R_X} x_k P_X(x_k) \\
&= \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!}
\end{aligned}
$$

To allow a favourable rewrite we now define $j = k - 1$, as such

$$
\begin{aligned}
EX &= \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^{j+1}}{j!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}
\end{aligned}
$$

As this is the Taylor series for e^{λ} we get

$$
\begin{aligned}
EX &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}
$$

An important concept in expectation is linearity, which is useful for calculating the expected value of linear functions of random variables. The theorem states that

- $E[aX + b] = aEX + b$, for all $a, b \in \mathbb{R}$
- $E[X_1 + X_2 + \ldots + X_n] = EX_2 + EX_1 + \ldots + EX_n$, for any set of random variables.

Binomial expected value As demonstrated earlier, the binomial can be seen as a sum of Bernoulli random variables, as we know the expected value of the Bernoulli, we make use of linearity

$$\begin{aligned} EX &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= EX_1 + EX_2 + \ldots + EX_n \\ &= p + p + \ldots + p \\ &= np \end{aligned}$$

Pascal expected value As in the binomial we can see the pascal as the sum of a sequence of geometric random variables, as such we again have by linearity that

$$\begin{aligned} EX &= \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= EX_1 + EX_2 + \ldots + EX_n \\ &= \frac{1}{p} + \frac{1}{p} + \ldots + \frac{1}{p} \\ &= \frac{n}{p} \end{aligned}$$

3.4 Functions of random variables

If X is a random variable and $Y = g(X)$ then Y is a random variable. As such we can talk about its PMF, CDF and EX. The range of Y can be determined as

$$R_Y = \{g(x) | x \in R_X\}$$

This also means that we can determine the expected value of a function of a random variable. We let X with PMF $P_X(x)$ be a discrete random variable and $Y = g(X)$. One way to determine EY is to determine the PMF of Y and then applying the expectation formula, however the law of the unconscious statisticatian which states that

$$E[g(X)] = \sum_{x_k \in R_X} g(x_k) P_X(x_k)$$

Theorem 32. *The expected value of a function of a discrete random variable is given by the law of the unconscious statistician,* $E[g(X)] = \sum_{k_k \in R_X} g(x_k) P_X(x_k)$.

Proof. By definition we have that

$$EY = \sum_{y_k \in R_Y} y_k P_Y(y_k)$$

As $Y = g(X)$ we can rewrite this expression as

$$E[g(X)] = \sum_{x_k \in R_X} x_k P_X(x_k)$$

□

3.4.1 Variance

The expected value however has a problem, as it ignores the form of the distribution, two PMFs with the same expected value may have very different distributions. This is represented by the variance of the PMF, which denotes how spread out the distribution is.

Definition 10. The variance of a random variable X with $EX = \mu_X$ is defined as

$$\mathrm{Var}(X) = E[(X - \mu_X)^2]$$

This works, as the variability will be high if X is often far away from the mean, μ_X, whilst it'll be low if it is close to the mean, we can therefore use the variance as a measure of how spread the distribution is.

To compute the variance we simply use LOTUS with $g(X) = (X - \mu_X)^2$

$$\mathrm{Var}(X) = E[(X - \mu_X)^2] = \sum_{x_k \in R_X} (x_k - \mu_X)^2 P_X(x_k)$$

Theorem 33. *The variance of a random variable X can be computed as $Var(X) = E[X] - [EX]^2$.*

Proof. By the definition of variance we have that

$$\begin{aligned}\mathrm{Var}(X) &= E[(X - \mu_X)^2] \\ &= E[X^2 + \mu_X^2 - 2X\mu_X]\end{aligned}$$

By linearity we then have that

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2] + E[\mu_X] - E[2X\mu_X] \\ &= E[X^2] + \mu_X^2 - 2\mu_X E[X] \\ &= E[X^2] + \mu_X^2 - 2\mu_X^2 \\ &= E[X^2] - \mu_X^2 \\ &= \left[\sum_{x_k \in R_x} x_k^2 P_X(x_k)\right] - \mu_X^2\end{aligned}$$

□

One problem with the variance is that it ends up being the wrong unit as the value is squared, as such we define the standard deviation as the square root of the variance

$$\mathrm{SD}(X) = \sigma_X = \sqrt{\mathrm{Var}(X)}$$

Theorem 34. *For a random variable X and real numbers a, b we have that $Var(aX + b) = a^2 \, Var(X)$.*

Proof. If $Y = aX + b$ then

$$\mathrm{Var}(Y) = E[(Y - EY)^2]$$

By linearity we have that

$$\mathrm{Var}(Y) = E[(aX + b - aEX - b)^2]$$

Reducing and factoring a

$$\begin{aligned}\mathrm{Var}(Y) &= E[(aX - aEX)^2] \\ &= E[(a^2(X - EX)^2)]\end{aligned}$$

As a is a constant we have that

$$\begin{aligned}\mathrm{Var}(Y) &= a^2 E[(X - \mu_X)^2] \\ &= a^2 \mathrm{Var}(X)\end{aligned}$$

□

Theorem 35. *For a random variable X and real numbers a, b we have that $SD(aX + b) = |a|SD(X)$.*

Proof. As

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var(X)}}$$

We square both sides of the previous theorem

$$\begin{aligned} \sqrt{\mathrm{Var(Y)}} &= \sqrt{a^2 \mathrm{Var}(X)} \\ &= \sqrt{a^2}\sqrt{\mathrm{Var}(X)} \end{aligned}$$

This is equivalent to

$$\mathrm{SD}(Y) = |a|\mathrm{SD}(X)$$

□

4 Chapter 4

4.1 Continuous random variables and their distributions

Suppose we have an interval on the real line, $[a, b]$ with a uniformly distributed probability of a given value being picked, we know that the probability of each point $P(X = x) = 0$ as there is an infinite amount of points, as such, it only makes sense to look at subintervals of the interval when it comes to probability.

Simply from the definition of a CDF we have that $F_X(x < a) = 0$, whilst $F_X(x \geq b) = 1$, simultaneously we can establish a general equation for the probability of an interval as the proportion it constitutes of the total interval

$$\begin{aligned} F_X(a \leq x_1 \leq x_2 \leq b) &= P(X \in [x_1, x_2]) \\ &= \frac{x_2 - x_1}{b - a} \end{aligned}$$

Using this definition we can create a CDF as

$$F_X(x) \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$$

Whether we use $<$ or $\leq$ (or the reverse), doesn't matter as the probability of each individual point is equal to 0, as such $P(X < 2) = P(X \leq 2)$.

Definition 11. A random variable X with CDF $F_X(x)$ is said to be continuous if $F_X(x)$ is a continuous function for all $x \in \mathbb{R}$.

4.2 Probability density function

As its impossible to define a PMF for a continuous function (as $P(X = x) = 0$ for all $x \in \mathbb{R}$), as instead define the probability density function as

$$f_X(x) = \lim_{\Delta \to 0^+} \frac{P(x < X \leq x + \Delta)}{\Delta}$$

We recall that $P(a < X \leq b) = F_X(b) - F_X(a)$, as such we rewrite the limit as

$$f_X(x) = \lim_{\Delta \to 0^+} \frac{F_X(x + \Delta) - F_X(x)}{\Delta}$$

We then recognize this as the definition of the derivate, as such we can write that

$$f_X(x) = \lim_{\Delta \to 0^+} \frac{F_X(x+\Delta) - F_X(x)}{\Delta} = F_X'(x)$$

Definition 12. Consider a continuous random variable X with an absolutely continuous CDF $F_X(x)$. The function $f_X(x)$ defined by

$$f_X(x) = \frac{d}{dx} F_X(x) = F_X'(x)$$

if $F_X(x)$ is differentiable at x is called the probability density function of X.

Using the general example from the previous section we determine the PDF as

$$\begin{aligned} F_X'(x) &= \frac{d}{dx} \frac{x-a}{b-a} \\ &= \frac{1}{b-a} \left(\frac{d}{dx} x + \frac{d}{dx} - a \right) \\ &= \frac{1}{b-a} \end{aligned}$$

As such we can define the probability density function as

$$f_X(X) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x \notin [a, b] \end{cases}$$

As the PDF is the derivative of the CDF, we can integrate a segment of the interval and get the probability thereof assuming it is absolutely continuous in that interval

$$F_X(x) = \int_{-\infty}^{x} f_X(x) dx$$

At the same time the integral over the entire real line must be equal to 1 in accordance with the axioms of probability

$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

Whilst we can determine the probability of an interval as

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx$$

Similarly we can define the range of a random variable X as the possible values of the random varibles where the PDF is larger than 0, as such

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

4.3 Expected value and variance

In discrete random variables we frequently used sums to determine different values, in the case of continuous random variables we replace the sum with an integral sign and the PMF with the PDF, using LOTUS as an example we for example get that

$$
\begin{aligned}
EX_{discrete} &= \sum_{x_k \in R_X} x_k P_X(x_k) \\
EX_{continuous} &= \int_{-\infty}^{\infty} x f_X(x) dx
\end{aligned}
$$

For variance we can do the same as

$$
\begin{aligned}
\mathrm{Var}(X) &= E[(X - \mu_X)^2] \\
&= EX^2 - (EX)^2 \\
&= \int_{-\infty}^{\infty} x^2 f_X(x) dx - \mu_X^2
\end{aligned}
$$

4.4 Functions of continuous random variables

If X is a continuous random variable and $Y = g(X)$ is a function of X, then Y is a random variable, as such finding the CDF and PDF of Y should be possible, either directly or by determining CDF and taking the derivative, to do so we however have to make sure the function is continuous over the real line.

4.4.1 Uniform distribution

A uniform distribution is a distribution with a PDF given by

$$
f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x \notin [a, b] \end{cases}
$$

The expected value of a uniform distribution is given by

$$\begin{aligned}
\text{Var}(x) &= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_{-\infty}^{\infty} x \times \left(\frac{1}{b-a}\right) dx \\
&= \frac{1}{b-a} \times \left[\frac{x^2}{2}\right]_a^b \\
&= \frac{1}{b-a} \times \left(\frac{b^2}{2} - \frac{a^2}{2}\right) \\
&= \frac{b^2 - a^2}{2(b-a)} \\
&= \frac{b-a}{2}
\end{aligned}$$

To determine the variance we need EX^2

$$\begin{aligned}
EX^2 &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\
&= \int_{-\infty}^{\infty} x^2 \times \left(\frac{1}{b-a}\right) dx \\
&= \frac{1}{b-a} \times \left[\frac{x^3}{3}\right]_a^b \\
&= \frac{1}{b-a} \times \left(\frac{b^3}{3} - \frac{a^3}{3}\right) \\
&= \frac{b^3 - a^3}{3(b-a)} \\
&= \frac{(b-a)(b^2 + a^2 + ab)}{3(b-a)} \\
&= \frac{b^2 + a^2 + ab}{3}
\end{aligned}$$

As such we can determine the variance as

$$
\begin{aligned}
Var(X) &= EX^2 - (EX)^2 \\
&= \frac{b^2 + a^2 + ab}{3} - \left(\frac{b-a}{2}\right)^2 \\
&= \frac{b^2 + a^2 + ab}{3} - \frac{(b-a)^2}{4} \\
&= \frac{4(b^2 + a^2 + ab)}{12} - \frac{3(b^2 + a^2 - 2ab)}{12} \\
&= \frac{4b^2 + 4a^2 + 4ab - 3b^2 - 3a^2 + 6ab}{12} \\
&= \frac{b^2 + a^2 - 2ab}{12}
\end{aligned}
$$

4.4.2 Exponential distribution

An exponential distribution is a distribution with a PDF given by

$$
f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}
$$

The expected value of a uniform distribution is given by

$$
\begin{aligned}
EX &= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_{-\infty}^{\infty} x \times \lambda e^{-\lambda x} dx \\
&= \lambda \int_{-\infty}^{\infty} x e^{-\lambda x} dx
\end{aligned}
$$

We now make use of integration by parts with $f = x$ and $g' = e^{-\lambda x}$

$$
\int x e^{-\lambda x} = -\frac{x e^{-\lambda x}}{\lambda} - \int -\frac{e^{-\lambda x}}{\lambda} dx
$$

And solve the integral using substitution with $u = -\lambda x \implies dx = \frac{1}{-\lambda} du$

$$
\begin{aligned}
\int -\frac{e^u}{\lambda}\frac{1}{-\lambda} du &= \frac{1}{\lambda^2} \int e^u du \\
&= \frac{e^u}{\lambda^2} \\
&= \frac{e^{-\lambda x}}{\lambda^2}
\end{aligned}
$$

As such we can compute the expected value as

$$\begin{aligned}
EX &= \int_{-\infty}^{\infty} \lambda \times \left(-\frac{xe^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2} \right) dx \\
&= \int_{-\infty}^{\infty} -xe^{-\lambda x} - \frac{e^{-\lambda x}}{\lambda} dx \\
&= \int_{-\infty}^{\infty} -\frac{\lambda x e^{-\lambda x} - e^{-\lambda x}}{\lambda} dx \\
&= \left[\frac{(\lambda x + 1)e^{-\lambda x}}{\lambda} \right]_0^{\infty} \\
&= \frac{(\lambda \times 0 + 1)e^0}{\lambda} - \lim_{x\to\infty} \frac{(\lambda x + 1)e^{-\lambda x}}{\lambda} \\
&= \frac{1}{\lambda}
\end{aligned}$$

To determine the variance we need EX^2

$$\begin{aligned}
EX^2 &= \int_0^{\infty} x^2 \times \lambda e^{-\lambda x} dx \\
&= \lambda \int_0^{\infty} x^2 \times e^{-\lambda x} dx \\
&= \lambda \int_0^{\infty} \left(-\frac{x^2 e^{-\lambda x}}{\lambda} - \int -\frac{2xe^{-\lambda x}}{\lambda} \right) \\
&= \lambda \int_0^{\infty} \left(-\frac{x^2 e^{-\lambda x}}{\lambda} + \frac{2}{\lambda} \int xe^{-\lambda x} \right) \\
&= \lambda \int_0^{\infty} \left(-\frac{x^2 e^{-\lambda x}}{\lambda} + \frac{2}{\lambda} \times \left(\frac{e^{-\lambda x}}{\lambda^2} \right) \right) \\
&= \int_0^{\infty} -x^2 e^{-\lambda x} + \frac{2e^{-\lambda x}}{\lambda^2} \\
&= \int_0^{\infty} \frac{-\lambda^2 x^2 e^{-\lambda x} + 2e^{-\lambda x}}{\lambda^2} \\
&= \left[\frac{-\lambda^2 x^2 e^{-\lambda x} + 2e^{-\lambda x}}{\lambda^2} \right]_0^{\infty} \\
&= \frac{-\lambda^2 0^2 e^{-\lambda 0} + 2e^0}{\lambda^2} - \lim_{x\to\infty} \frac{-\lambda^2 x^2 e^{-\lambda x} + 2e^{-\lambda x}}{\lambda^2} \\
&= \frac{2}{\lambda^2}
\end{aligned}$$

As such we can determine the variance as

$$
\begin{aligned}
\mathrm{Var}(X) &= EX^2 - (EX)^2 \\
&= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2}
\end{aligned}
$$

4.4.3 Normal (Gaussian) distribution

A standard normal random variable is a random variable whos PDF follows

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Where the first term ensures that the area under the curve is equal to one. The expected value of such a distribution is given by

$$
\begin{aligned}
EZ &= \int_{-\infty}^{\infty} z f_Z(z) dz \\
&= \int_{-\infty}^{\infty} z \times \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \times e^{-\frac{z^2}{2}} dz
\end{aligned}
$$

Making use of integration by substitution with $u = -\frac{z^2}{2} \implies dz = \frac{1}{-z} du$

$$
\begin{aligned}
EZ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \times e^u \times \frac{1}{-z} du \\
&= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^u du \\
&= -\frac{1}{\sqrt{2\pi}} \left[e^u\right]_{-\infty}^{\infty} \\
&= -\frac{1}{\sqrt{2\pi}} \left[e^{-\frac{z^2}{2}}\right]_{-\infty}^{\infty} \\
&= -\frac{1}{\sqrt{2\pi}} \times \left(\lim_{z\to\infty} e^{-\frac{z^2}{2}} - \lim_{z\to-\infty} e^{-\frac{z^2}{2}}\right) \\
&= 0
\end{aligned}
$$

This makes sense as the distribution is symmetrical around 0. We now determine EZ^2 to enable determining the variance.

$$\begin{aligned}
EZ^2 &= \int_{-\infty}^{\infty} z^2 f_Z(z)dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} z \left(ze^{-\frac{z^2}{2}} dz \right) + \int_{0}^{\infty} z \left(ze^{-\frac{z^2}{2}} \right) dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\left[-ze^{-\frac{z^2}{2}} \right]_{-\infty}^{0} + \int_{-\infty}^{0} e^{-\frac{z^2}{2}} dz - \left[ze^{-\frac{z^2}{2}} \right]_{0}^{\infty} + \int_{0}^{\infty} e^{-\frac{z^2}{2}} dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} e^{-\frac{z^2}{2}} dz + \int_{0}^{\infty} e^{-\frac{z^2}{2}} dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \int_{-\infty}^{\infty} f_Z(z)dz \\
&= 1
\end{aligned}$$

As such we can determine the variance of the Gaussian distribution as

$$\begin{aligned}
\mathrm{Var}(Z) &= EZ^2 - (EZ)^2 \\
&= 1 - 0 \\
&= 1
\end{aligned}$$

The CDF of the normal distribution which we would typically find by integrating the PDF, this integral does however not have a closed form solution, and as such we typically denote the CDF of a standard normal as

$$\phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dz$$

Furthermore we define a normal random variable, a random variable which can be transformed into any other random variable, for this we have that

$$\begin{aligned}
X &= \sigma Z + \mu, \text{ where } \sigma > 0 \\
EX &= \sigma EZ + \mu = \mu \\
\mathrm{Var}(X) &= \sigma^2 \mathrm{Var}(Z) = \sigma^2
\end{aligned}$$

By this definition we call X a normal random variable with mean μ and variance σ^2, notationally this can be written

$$X \sim N(\mu, \sigma^2)$$

A random variable with such distribution has that

$$\begin{aligned} P(X < z) &= F_X(z) \\ &= \phi\left(\frac{z-\mu}{\sqrt{\sigma}}\right) \end{aligned}$$

Theorem 36. *A linear transformation of a normal random variable is itself a normal random variable.*

Proof. Let

$$X \sim N(\mu_X, \sigma_X^2) \text{ and } Y = aX + b \text{ where } a, b \in \mathbb{R}$$

Expanding the shorthand we have that

$$X = \sigma_X Z + \mu_X$$

As such we write that

$$\begin{aligned} Y &= aX + b \\ &= a(\sigma_X Z + \mu_X) + b \\ &= a\sigma_X Z + a\mu_X + b \end{aligned}$$

Which can also be written as

$$\begin{aligned} Y &= (a\sigma_X)Z + (a\mu_X + b) \\ &= N(a\mu_X + b, a^2\sigma_X^2) \end{aligned}$$

□

4.4.4 Gamma distribution

The gamma distribution makes use of the gamma function, denoted by $\Gamma(x)$, which is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx,\ \alpha > 0$$

In practice, the gamma function is an extension of the factorial operation to the real (and complex) numbers. This function has some notable properties

- $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$
- $\int_0^\infty x^{\alpha-1}e^{-\lambda x}dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}$ for $\lambda > 0$
- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
- $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

The gamme function has a PDF given by

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

If we let $\alpha = 1$ we get

$$f_X(x) \begin{cases} \lambda e^{-\lambda} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

We recognize that this is the PDF of an exponential distribution, and as such the gamma distribution can be seen as the sum of n independent exponential distributions. We again determine the expected value and variance of a gamma distribution, using LOTUS we have that

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x) \\ &= \int_{-\infty}^{\infty} x \frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x} \\ &= \frac{\lambda^a}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \\ &= \frac{\lambda^a}{\Gamma(\alpha)} \times \frac{\alpha\Gamma(\alpha)}{\lambda^{\alpha+1}} \\ &= \frac{\alpha}{\lambda} \end{aligned}$$

The same method is employed to determine EX^2, as such

$$\begin{aligned}
EX^2 &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_{-\infty}^{\infty} x^2 \frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+1}e^{-\lambda x}dx \\
&= \frac{\lambda^a}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} \\
&= \frac{\lambda^a}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} \\
&= \frac{\lambda^a}{\Gamma(\alpha)} \times \frac{(\alpha+1)\Gamma(\alpha+1)}{\lambda^{\alpha+2}} \\
&= \frac{\lambda^a}{\Gamma(\alpha)} \times \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\lambda^{\alpha+2}} \\
&= \frac{(\alpha+1)\alpha}{\lambda^2} \\
&= \frac{a^2+\alpha}{\lambda^2}
\end{aligned}$$

We now determine the variance and get that

$$\begin{aligned}
\mathrm{Var}(X) &= EX^2 - (EX)^2 \\
&= \frac{a^2+a}{\lambda^2} - \frac{a^2}{\lambda^2} \\
&= \frac{a}{\lambda^2}
\end{aligned}$$

4.5 Mixed random variables

As opposed to the previous exclusively discrete or continuous random variables, mixed random variables feature both a discrete and continuous part.

A Chapter problems

A.1 Chapter 1

Problem 1 Suppose the universal set is defined as $S = \{x \in \mathbb{N} | 1 \geq x \geq 10\}$, $A = \{1, 2, 3\}$, $B = \{x \in S | 2 \geq x \geq 7\}$ and $C = \{7, 8, 9, 10\}$.

a) Find $A \cup B$

$$A \cup B = \{1, 2, 3\} \cup \{2, 3, 4, 5, 6, 7\} = \{1, 2, 3, 4, 5, 6, 7\}$$

b) Find $(A \cup C) \setminus B$

$$(A \cup C) \setminus B = \{1, 2, 3, 7, 8, 9, 10\} \setminus \{2, 3, 4, 5, 6, 7\} = \{1, 8, 9, 10\}$$

c) Find $\overline{A} \cup (B \setminus C)$

$$\overline{A} \cup (B \setminus C) = \{4, 5, 6, 7, 8, 9, 10\} \cup \{2, 3, 4, 5, 6\} = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

d) Do A, B and C form a partition of S?

No as $A \cap B \neq \emptyset$ and $B \cap C \neq \emptyset$.

Problem 2 When working with real numbers, our universal set is $\mathbb{R}$. Find each of the following sets a) $[6, 8] \cup [2, 7[$

$$[2, 8]$$

b) $[6, 8] \cap [2, 7[$

$$[6, 7[$$

c) $\overline{[0, 1]}$

$$]-\infty, 0[\cup]1, \infty[$$

d) $[6, 8] \setminus]2, 7[$

$$[7, 8]$$

Problem 3 For each of the follopwing Venn diagrams, write the sets denoted by the shaded area

a)

$$A \cup B \setminus A \cap B$$

b)

$$B \setminus C$$

c)

$$A \cap B + A \cap C$$

d)

$$C \cup (A \cap B) \setminus ((C \cap A) \cup (C \cap B))$$

Problem 4 A coin is tossed twice. Let $S = \{H, T\} \times \{H, T\}$. Write the following sets by listing their elements:

a) First toss is heads

$$A = \{(H, T), (H, H)\}$$

b) At least one tails

$$B = \{(T, T), (T, H), (H, T)\}$$

c) Two tosses are different

$$C = \{(H, T), (T, H)\}$$

Problem 5 Let $A = \{1, 2, \ldots, 100\}$. For any $i \in \mathbb{N}$, define A_i as the set of numbers in A that are divisible by i.

a)

$$|A_1| = \left\lfloor \frac{|S|}{1} \right\rfloor = \left\lfloor \frac{100}{1} \right\rfloor = 100$$
$$|A_2| = \left\lfloor \frac{|S|}{2} \right\rfloor = \left\lfloor \frac{100}{2} \right\rfloor = 50$$
$$|A_3| = \left\lfloor \frac{|S|}{3} \right\rfloor = \left\lfloor \frac{100}{3} \right\rfloor = 33$$
$$|A_4| = \left\lfloor \frac{|S|}{4} \right\rfloor = \left\lfloor \frac{100}{4} \right\rfloor = 25$$
$$|A_5| = \left\lfloor \frac{|S|}{5} \right\rfloor = \left\lfloor \frac{100}{5} \right\rfloor = 20$$

b)

By the distributive property we have that

$$A_2 \cap A_3 \cap A_5 = (A_2 \cap A_3) \cap A_5$$

As the intersection of A_2 and A_3 must be the even factors of 3 we get

$$A_2 \cap A_3 = \{x \in A | x = 6n, n \in \mathbb{N}\}$$

The cardinality of this must be half of the original as every other value is valid

$$|A_2 \cap A_3| = \left\lfloor \frac{33}{2} \right\rfloor = 16$$

As only every fifth factor of 6 is a factor of 5 we get that

$$A_2 \cap A_3 \cap A_5 = \{x \in A | x = 30n, n \in \mathbb{N}\}$$

As this is every fifth value of the previous intersection, the cardinality must be one fifth of the previous

$$|A_2 \cap A_3 \cap A_5| = \left\lfloor \frac{16}{5} \right\rfloor = 3$$

Problem 6 As A_1, A_2, A_3 form a partition of the universal set, the cardinality of B must be equal to the sum of the cardinalities in the individual partitions.

$$|B| = \sum_{i=1}^{3} |B \cap A_i| = 10 + 20 + 15 = 45$$

Problem 7 a) As the numbers can be listed in one-to-one correspondance with the natural numbers the set is countable.

b) As the set is made up of the union of 2 countable sets we have that

$$B = \bigcup_{i=\mathbb{Q}} \bigcup_{j=\mathbb{Q}} \{a_i + b_j\sqrt{2}\}$$

As such it must be countable as its constituents are.

c) As the set is a subset of an uncountable set (real numbers), it is not countable.

Problem 8 We take the limit of the upper bound of the interval as $n \to \infty$

$$\lim_{n\to\infty} \frac{n-1}{n} = 1$$

As this limit is the upper bound of the interval, corresponding to the maximal interval included in the union of sets we get that

$$A = \bigcup_{n=1}^{\infty} A_n = [0; 1[$$

Problem 9 Opposed to the previous problem the smallest set will here define the set as the intersection is limited to the smallest component. We take the limit as $n \to \infty$ as the value is inverse proportional to n

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

As such

$$A = \bigcap_{n=1}^{\infty} A_n = \{0\}$$

Problem 10 a)

b)

Problem 11 As the set is given as an interval it is clear that

$$[0, 1[\subset \mathbb{R}$$

As subsets of uncountable sets are uncountable, it becomes clear that the range is uncountable.

Problem 12 a)

Reading the function it becomes clear that the domain of the function is given by

$$\{H, T\}^3$$

While the co-domain is given by

$$\mathbb{N} \cup \{0\}$$

b)

As the function is limited by the amount of heads that can appear in the sequence it is clear that

$$\text{Range}(f) = \{0, 1, 2, 3\}$$

c)

Knowing that $x = 2$ tells us that 2 heads are present in the sequence, and as such 1 tails must also be present, as such the possible events are

$$\{(H, H, T), (H, T, H), (T, H, H)\}$$

Problem 13 a)

We logically assume that the events are disjoint as two teams cant win, as such it is clear that

$$0.5 + P(b) + 0.25 = 1 \Leftrightarrow P(b) = 1 - 0.5 - 0.25 = 0.25$$

b) As the events are disjoint we determine the probability as

$$P(b \cup d) = 0.25 + 0.25 = 0.5$$

Problem 14 a)

By inclusion-exclusion principle we have that

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.7 - 0.9 = 0.2$$

b)

We have that

$$P(\overline{A} \cap B) = P(B \setminus A)$$

Expanding this expression we get that

$$P(\overline{A} \cap B) = P(B) - P(A \cap B) = 0.7 - 0.2 = 0.5$$

c)

Expanding the expression we again get

$$P(A \setminus B) = P(A) - P(A \cap B) = 0.4 - 0.2 = 0.2$$

d)

Expanding the expression we get

$$P(\overline{A} \setminus B) = P(\overline{A}) - P(\overline{A} \cap B) = (1 - 0.4) - 0.5 = 0.1$$

e)

By the inclusion-exclusion principle we have

$$P(\overline{A} \cup B) = P(\overline{A}) + P(B) - P(\overline{A} \cap B) = (1 - 0.4) + 0.7 - 0.5 = 0.8$$

f) By the distributive law

$$A \cap (B \cup \overline{A}) = (A \cap B) \cup (A \cap \overline{A})$$

As

$$A \cap \overline{A} = \emptyset$$

by definition. We get that

$$P(A \cap (B \cup \overline{A})) = P((A \cap B) \cup \emptyset) = 0.2$$

Problem 15 a)

We assume the rolls are independent, as such

$$P(X_2 = 4) = \frac{1}{6}$$

b)

The sample space is given by

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

The outcomes that satisfy the event are

$$\{X_1 + X_2 = 7\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

As such

$$P(X_1 + X_2 = 7) = \frac{6}{36} = \frac{1}{6}$$

c)

As they are independent we determine the probability as

$$P(X_1 \neq 2 \cap X_2 \geq 4) = P(X_1 \neq 2)P(X_2 \geq 4) = \frac{5}{6} \cdot \frac{3}{6} = \frac{15}{36} = \frac{5}{12}$$

Problem 16 a)

As the sum of the individual probabilities must equate to 1 we wish to determine a c that satisfies the equation

$$\sum_{k=1}^{\infty} P(k) = \sum_{k=1}^{\infty} \frac{c}{3^k} = 1$$

We apply the formula for a geometric series

$$\sum_{k=0}^{\infty} cr^k = \frac{a}{1-r} \text{ for } |r| < 1$$

As this starts from 0 we subtract c as $P(\{0\}]) = \frac{c}{k^0} = c$

$$\begin{aligned} 1 &= -c + \sum_{k=0}^{\infty} c \left(\frac{1}{3}\right)^k \\ &= -c + \frac{c}{1 - \frac{1}{3}} \end{aligned}$$

By isolation for c we then get that

$$\frac{2}{3} = -\frac{2}{3}c + c \Leftrightarrow \frac{2}{3} = \frac{1}{3}c \Leftrightarrow 2 = c$$

b)

The set is given by the union of the 3, as such

$$\begin{aligned} P(\{2,4,6\}) &= P(\{2\} \cup \{4\} \cup \{6\}) \\ &= P(\{2\}) + P(\{4\}) + P(\{6\}) \\ &= \frac{2}{9} + \frac{2}{81} + \frac{2}{729} \\ &= \frac{182}{729} \end{aligned}$$

c) This must equate to the complement of $P(\{1,2\})$, as such

$$\begin{aligned} P(\{3,4,5,\ldots\}) &= 1 - P(\{1,2\}) \\ &= 1 - \left(\frac{2}{3} + \frac{2}{9}\right) \\ &= 1 - \frac{8}{9} \\ &= \frac{1}{9} \end{aligned}$$

Problem 17 We have that

$$\begin{aligned} P(A) &= P(B) \\ P(C) &= 2P(D) \\ P(A \cup C) &= P(A) + P(C) = 0.6 \end{aligned}$$

We rewrite all terms as functions of $P(A)$, these must equate to 1 as a team has to win, making it the sample space

$$\begin{aligned} 1 &= P(A) + P(A) + (0.6 - P(A)) + \frac{0.6 - P(A)}{2} \\ &= 0.5P(A) + 0.9 \end{aligned}$$

Isolating for $P(A)$ we get

$$P(A) = \frac{1 - 0.9}{0.5} = 0.2$$

From this we can determine the rest of the probabilities using the requirements set earlier

$$\begin{aligned} P(A) &= P(B) = 0.2 \\ P(C) &= 0.6 - P(A) = 0.4 \\ P(D) &= 0.5P(C) = 0.2 \\ \sum P &= 0.2 + 0.2 + 0.4 + 0.2 = 1 \end{aligned}$$

Problem 18 a)

We insert $t = 1$

$$P(T \leq 1) = \frac{1}{16}1^2 = \frac{1}{16}$$

b)

This must equate to the complement of it taking 2 hours, as such

$$P(2 > t) = 1 - \left(\frac{1}{16}2^2\right) = \frac{3}{4}$$

c)

This equates to the difference between the probability of the job taking more than 1 hour and more than 3 hours

$$P(1 \leq T \leq 3) = P(T \leq 1) - P(T \leq 3) = \frac{1}{16}3^2 - \frac{1}{16}1^2 = \frac{1}{2}$$

Problem 19 The problem has the form of the quadratic equation

$$ax^2 + bx + c = 0 \implies x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

For this to have real solutions it is necessary that

$$b^2 - 4ac \geq 0 \vee a \neq 0$$

Translating this to the given equation we wish to determine values $A, B \in [0, 1]$ that satisfy

$$1^2 - 4AB \geq 0 \vee A \neq 0$$

Seeing this as a function based on the given figure, we write that

$$1^2 - 4xy \geq 0$$

Isolating for y it becomes apparent that

$$\begin{aligned} 1^2 - 4xy &\geq 0 \\ 1^2 &\geq 4xy \\ \frac{1}{4x} &\geq y \end{aligned}$$

As the maximal y-value is 1 we determine the intersection

$$1 = \frac{1}{4x} \Leftrightarrow x = \frac{1}{4}$$

Now we can determine the probability as the proportion of the area that fits under this curve in the interval for which it is inside the unit square

$$\begin{aligned} P(1 - 4AB \geq 0) &= \frac{1}{4} + \frac{1}{4}\int_{\frac{1}{4}}^{1} \frac{1}{4x} \\ &= \frac{1}{4} + \frac{1}{4}\left[\ln x\right]_{\frac{1}{4}}^{1} \\ &= \frac{1}{4} + \frac{1}{4}\left(\ln 1 - \ln\frac{1}{4}\right) \\ &\approx \frac{3}{5} \end{aligned}$$

Problem 20 a)

As every set is a subset of the next, it is clear that A_n will be the largest or equal to the largest set in the sequence, as such

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n)$$

As all the sets are subsets of A_n and therefore their union will create A_n.

b)

Here it is the opposite where A_n will be the smallest or equal to the smallest set in the sequence as such

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n)$$

As the intersection will always be limited by the smallest set in the sequence.

Problem 21 DO LATER

Problem 22 Let A be the event that the customer purchased a cup of coffee and B the event they purchased a piece of cake, then

$$\begin{aligned} P(A) &= 0.7 \\ P(B) &= 0.4 \\ P(A \cap B) &= 0.2 \end{aligned}$$

By conditional probability we have that

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.2}{0.4} \\ &= 0.5 \end{aligned}$$

Problem 23 a)

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.2}{0.35} \\ &\approx 0.57 \end{aligned}$$

b)

$$\begin{aligned} P(C|B) &= \frac{P(B \cap C)}{P(B)} \\ &= \frac{0.15}{0.35} \\ &\approx 0.43 \end{aligned}$$

c)

$$\begin{aligned} P(B|A \cup C) &= \frac{P(B \cap (A \cup C))}{P(A \cup C)} \\ &= \frac{0.25}{0.7} \\ &\approx 0.36 \end{aligned}$$

d)

$$
\begin{aligned}
P(B|A\cap C) &= \frac{P(B\cap(A\cap C))}{P(A\cap C)} \\
&= \frac{0.1}{0.2} \\
&= 0.5
\end{aligned}
$$

Problem 24 As this is an interval with infinite points, we determine the proportion of the interval that is spanned by the interval were observing, as such

$$P(A \leq X \leq B) = \frac{B-A}{10}$$

a)

$$P(2 \leq X \leq 5) = \frac{5-2}{10} = 0.3$$

b)

$$P(X \leq 2|X \leq 5) = \frac{2}{5} = 0.4$$

c)

$$
\begin{aligned}
P(3 \leq X \leq 8|X \geq 4) &= \frac{P(3 \leq X \leq 8 \cap X \geq 4)}{X \geq 4} \\
&= \frac{P(4 \leq X \leq 8)}{0.6} \\
&= \frac{\frac{8-4}{10}}{0.6} \\
&= \frac{2}{3}
\end{aligned}
$$

Problem 25 We define the event A as getting an A in the course and B as living on campus

$$
\begin{aligned}
P(A) &= \frac{120}{600} = 0.2 \\
P(B) &= \frac{200}{600} = \frac{1}{3} \\
P(A|\overline{B}) &= \frac{80}{400} = 0.2 P(A|B) & &= \frac{40}{200} = 0.2
\end{aligned}
$$

For A and B to be independent we would expect

$$P(A|B) = P(A)$$

Which is true, as such, the 2 events are independent.

Problem 26 Define N_1, N_6 as the number of times out of n that a 1 or 6 is rolled, and let X_i be the i'th roll, then

$$P(N_1 \geq 1 \cap N_6 \geq 1) = 1 - P(\overline{N_1 \geq 1 \cap N_6 \geq 1})$$

By De Morgans Law we get

$$\begin{aligned}
&= 1 - P(N_1 = 0 \cap N_6 = 0) \\
&= 1 - (P(X_1 \neq 1, X_2 \neq 1, \ldots, X_n \neq 1) + P(X_1 \neq 6, X_2 \neq 6, \ldots, X_n \neq 6) \\
&- P((X_1 \neq 1, X_2 \neq 1, \ldots, X_n \neq 1) \cap (X_1 \neq 1, X_2 \neq 6, \ldots, X_n \neq 6)))
\end{aligned}$$

As the final term is equivalent to

$$\bigcap_{i=1}^{n} P(X_i \neq 1 \cap X_i \neq 6)$$

We get that

$$\begin{aligned}
&= 1 - \left(2\left(\frac{5}{6}\right)^n - \left(\frac{4}{6}\right)^n\right) \\
&= 1 - \frac{2 \cdot 5^n - 4^n}{6^n}
\end{aligned}$$

Problem 27 a)

Tree is finished on paper.

b)

We sum the probabilities that result in an error

$$\begin{aligned}
P(E) &= P(E|G) + P(E\overline{G}) \\
&= 0.08 + 0.06 \\
&= 0.14
\end{aligned}$$

c)

From the definition of conditional probability we have

$$\begin{aligned}
P(G|\overline{E}) &= \frac{P(G \cap \overline{E})}{P(\overline{E})} \\
&= \frac{0.72}{1 - 0.14} \\
&= \frac{0.72}{0.86} \\
&\approx 0.84
\end{aligned}$$

Problem 28 We define D as the unit being defective, the probability of picking a defective unit on the first draw is

$$P(D_1) = \frac{5}{100}$$

Assuming that one wasnt defective the next draw has

$$P(D_2|\overline{D_1}) = \frac{5}{99}$$

And the 3rd

$$P(D_3|\overline{D_2}, \overline{D_1}) = \frac{5}{98}$$

We wish to determine the probaility of the event

$$P(D = 1) = P(\{(D, \overline{D}, \overline{D}), (\overline{D}, D, \overline{D}), (\overline{D}, \overline{D}, D)\})$$

As these are disjoint we determine them as the sum of their individual probailities

$$\begin{aligned}
P(D = 1) &= P(\{(D, \overline{D}, \overline{D})\}) + P(\{\overline{D}, D, \overline{D}\}) + P(\{\overline{D}, \overline{D}, D)\}) \\
&= \frac{5}{100} \cdot \frac{95}{99} \cdot \frac{94}{98} + \frac{95}{100} \cdot \frac{5}{99} \cdot \frac{94}{98} + \frac{95}{100} \cdot \frac{94}{99} \cdot \frac{5}{98} \\
&\approx 0.14
\end{aligned}$$

Problem 29 a)

As the components are connected in series, the probability of the system being functional is equal to the complement of one or more of the components not working. Let F be the event that the system is functional and $P(C_i)$ be the event that the i'th component is functional.

$$P(F) = P(C_1 \cap C_2 \cap C_3) = P(C_1)P(C_2)P(C_3)$$

b)

As they're connected in parallel they must all fail for the circuit to be non-functional, as such the probability must be given be the union of them

$$P(F) = P(C_1 \cup C_2 \cup C_3)$$

By the inclusion-exclusion principle

$$\begin{gathered}
P(F) = P(C_1) + P(C_2) + P(C_3) - P(C_1)P(C_2) - P(C_1)P(C_3) - \\
P(C_2)P(C_3) + P(C_1)P(C_2)P(C_3)
\end{gathered}$$

c)

The probability that this circuit works must be given by the probability that either C_1 and C_3 or C_2 and C_3 are functional.

$$\begin{aligned} P(F) &= P((C_1 \cap C_3) \cup (C_2 \cap C_3)) \\ &= P(C_1 \cap C_3) + P(C_2 \cap C_3) - P((C_1 \cap C_3) \cap (C_2 \cap C_3)) \\ &= P(C_1)P(C_3) + P(C_2)P(C_3) - P(C_1 \cap C_2 \cap C_3) \end{aligned}$$

d)

The probability that this circuit works is given by the probability that either C_3 or both C_1 and C_2 work

$$\begin{aligned} P(F) &= P(C_3 \cup (C_1 \cap C_2)) \\ &= P(C_3) + P(C_1 \cap C_2) \\ &= P(C_3) + P(C_1)P(C_2) - P(C_1)P(C_2)P(C_3) \end{aligned}$$

e)

The probability that this circuit works is determined by whether C_1, C_2 and C_5 or C_3, C_4 and C_5 work, therefore

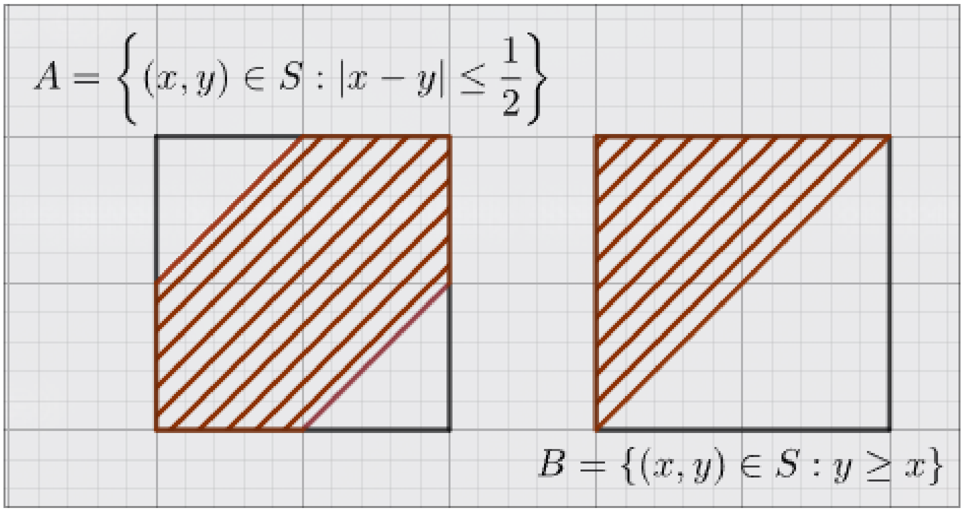
$$\begin{aligned} P(F) &= P(C_1 \cap C_2 \cap C_5) + P(C_3 \cap C_4 \cap C_5) - P(C_1 \cap C_2 \cap C_3 \cap C_4 \cap C_5) \\ &= P(C_1)P(C_2)P(C_5) + P(C_3)P(C_4)P(C_5) - \prod_{i=1}^{5} P(C_i) \end{aligned}$$

Problem 30 a)

Rewriting the set as a piecewise function we get

$$A = \begin{cases} x - y \leq \frac{1}{2} & x \geq y \\ y - x \leq \frac{1}{2} & y \geq x \end{cases} \implies \begin{cases} y \leq x - \frac{1}{2} & x \geq y \\ y \leq \frac{1}{2} + x & y \geq x \end{cases}$$

At the same time B is equal to the area above the $45°$ line in the unit square. Drawing this in Geogebra gives us



b)

Using geometry to calculate the areas we get that

$$P(A) = 1 - 2 \cdot 0.5 \cdot 0.5 \cdot 0.5 = \frac{3}{4}$$
$$P(B) = 1 - 0.5 \cdot 1 \cdot 1 = \frac{1}{2}$$

c)

We determine the probability of the intersection using geometry

$$P(A \cap B) = 1 - 0.5 - 0.5 \cdot 0.5 \cdot 0.5 = \frac{3}{8}$$

For them to be independent

$$P(A \cap B) = P(A)P(B) = \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8}$$

As such they are independent.

Problem 31 Let A be the event that an email is spam, B the event that it contains the word refinance, we then have that

$$P(A) = 0.5$$
$$P(B|A) = 0.01$$
$$P(B|\overline{A}) = 0.00001$$

We wish to determine the probability

$$P(A|B)$$

By Bayes rule we have that

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{0.01 \cdot 0.5}{0.01 \cdot 0.5 + 0.00001 \cdot 0.5} \\ &\approx 0.99 \end{aligned}$$

Problem 32 a)

For the path to be open it is clear that B_1, B_4 or B_2, B_5 or B_1, B_3, B_5 or B_2, B_3, B_4 must be open

$P(A) = P(\{B_1 \cap B_3 \cap B_4\} \cup \{B_1 \cap B_3 \cap B_5\} \cup \{B_2 \cap B_3 \cap B_5\} \cup \{B_2 \cap B_3 \cap B_4\})$

We rename each path to P_i where i is the arbitrary number assigned to the path. By the inclusion exclusion principle we have that

$$
\begin{aligned}
P(A) &= P(P_1) + P(P_2)P(P_3) + P(P_4) - P(P_1)P(P_2) - P(P_1)P(P_3) \\
&- P(P_1)P(P_4) - P(P_2)P(P_3) - P(P_2)P(P_4) - P(P_3)P(P_4) \\
&+ P(P_1)P(P_2)P(P_3) + P(P_1)P(P_2)P(P_4) + P(P_1)P(P_3)P(P_4) \\
&+ P(P_2)P(P_3)P(P_4) - P(P_1)P(P_2)P(P_3)P(P_4) \\
&= P(B_1 \cap B_4) + P(B_2 \cap B_5) + P(B_1 \cap B_3 \cap B_5) + P(B_2 \cap B_3 \cap B_4) \\
&- P((B_1 \cap B_4) \cap (B_2 \cap B_5)) - P((B_1 \cap B_4) \cap (B_1 \cap B_3 \cap B_5)) \\
&- P((B_1 \cap B_4) \cap (B_2 \cap B_3 \cap B_4)) - P((B_2 \cap B_5) \cap (B_1 \cap B_3 \cap B_5)) \\
&- P((B_2 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) - P((B_1 \cap B_3 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&+ P((B_1 \cap B_4) \cap (B_2 \cap B_5) \cap (B_2 \cap B_3 \cap B_5)) \\
&+ P((B_1 \cap B_4) \cap (B_2 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&+ P((B_1 \cap B_4) \cap (B_1 \cap B_3 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&+ P((B_2 \cap B_5) \cap (B_2 \cap B_3 \cap B_4) \cap (B_2 \cap B_3 \cap B_4)) \\
&- P((B_1 \cap B_4) \cap (B_2 \cap B_5) \cap (B_1 \cap B_3 \cap B_5) \cap (B_2 \cap B_3 \cap B_4)) \\
&= P(B_1 \cap B_4) + P(B_2 \cap B_5) + P(B_1 \cap B_3 \cap B_5) + P(B_2 \cap B_3 \cap B_4) \\
&- P(B_1 \cap B_2 \cap B_4 \cap B_5) - P(B_1 \cap B_3 \cap B_4 \cap B_5) \\
&- P(B_1 \cap B_2 \cap B_3 \cap B_4) - P(B_1 \cap B_2 \cap B_3 \cap B_5) \\
&- P(B_2 \cap B_3 \cap B_4 \cap B_5) - P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\
&+ P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) + P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\
&+ P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) + P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\
&- P(B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_5) \\
&= P_1P_4 + P_2P_5 + P_1P_3P_5 + P_2P_3P_4 - P_1P_2P_4P_5 - P_1P_3P_4P_5 - P_1P_2P_3P_4 \\
&- P_1P_2P_3P_5 - P_2P_3P_4P_5 - P_1P_2P_3P_4P_5 + 4P_1P_2P_3P_4P_5 - P_1P_2P_3P_4P_5 \\
&= P_1P_4 + P_2P_5 + P_1P_3P_5 + P_2P_3P_4 - P_1P_2P_4P_5 - P_1P_3P_4P_5 - P_1P_2P_3P_4 \\
&- P_1P_2P_3P_5 - P_2P_3P_4P_5 + 2P_1P_2P_3P_4P_5 \\
&= P_1P_4(1 - P_2P_5 - P_3P_5 - P_2P_3 + 2P_2P_3P_5) + P_2P_5 \\
&+ P_1P_3P_5 + P_2P_3(P_4 - P_1P_5 - P_4P_5)
\end{aligned}
$$

b) We use Bayes rule to determine the probability, conditioning A on B_3 we effectively cancel out the P_3 terms as $P(P_3|B_3) = 1$

$$
\begin{aligned}
P(A|B_3) &= P_1P_4(1 - P_2P_5 - P_5 - P_2 + 2P_2P_5) + P_2P_5 + P_1P_5 \\
&+ P_2(P_4 - P_1P_5 - P_4P_5)
\end{aligned}
$$

By Bayes rule we have that

$$P(B_3|A) = \frac{P(A|B_3)P_3}{P(A)}$$

Inserting known information

$$P(B_3|A) = \frac{P_1P_3P_4(1 - P_2P_5 - P_5 - P_2 + 2P_2P_5) + P_2P_3P_5 + P_1P_3P_5 + P_2P_3(P_4 - P_1P_5 - P_4P_5)}{P_1P_4(1 - P_2P_5 - P_3P_5 - P_2P_3 + 2P_2P_3P_5) + P_2P_5 + P_1P_3P_5 + P_2P_3(P_4 - P_1P_5 - P_4P_5)}$$

Problem 33 Before any extra information is given we have that

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}$$

We let C_i be the event that the host opens the i'th door and H_i be the event that the car is behind the i'th door, assuming we opened the 1st door we get the following probabilities

$$\begin{aligned} P(C_1|H_1) &= 0 \\ P(C_2|H_1) &= \frac{1}{2} \\ P(C_3|H_1) &= \frac{1}{2} \\ P(C_1|H_2) &= 0 \\ P(C_2|H_2) &= 0 \\ P(C_3|H_2) &= 1 \\ P(C_1|H_3) &= 0 \\ P(C_2|H_3) &= 1 \\ P(C_3|H_3) &= 0 \end{aligned}$$

If the host choses to open the 3rd door, we wish to determine whether

$$P(H_2|C_3) > P(H_1)$$

As that would put us in an advantageous situation. We apply Bayes rule and get

$$\begin{aligned} P(H_2|C_3) &= \frac{P(C_3|H_2)P(H_2)}{P(C_3|H_1)P(H_1) + P(C_3|H_2)P(H_2) + P(C_3|H_3)P(H_3)} \\ &= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} \\ &= \frac{2}{3} \end{aligned}$$

As

$$P(H_2|C_3) > P(H_1)$$

It would be advantageous to switch our guess.

Problem 34 a)

We have that

$$P(A) = \frac{1}{6}$$
$$P(B) = \frac{1}{6}$$
$$P(C) = \frac{1}{6}$$

For A and B to be independent it must be true that

$$P(A \cap B) = P(A)P(B)$$

The intersection of the 2 is only satisfied by the pair $(2, 5)$, as such

$$P(A \cap B) = \frac{1}{36} \stackrel{?}{=} \frac{1}{6} \cdot \frac{1}{6}$$

As the sides are equivalent the two are independent.

b)

Again we determine $P(A \cap C)$, this is again a single pair, $(2, 3)$ as such the rest of the problem is equivalent to the previous. The two are independent.

c)

Again $P(B \cap C)$ is a single pair $(4, 3)$, as such they are once again independent.

d)

As we already know we satisfy 3 of the requirements we must at last determine whether

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

This however is not true as the intersection is empty due to 2 and 3 never adding up to 7, as such

$$P(A \cap B \cap C) = P(\emptyset) = 0$$

Which is not equal to the product of the individual probabilities as

$$0 \neq \left(\frac{1}{6}\right)^3$$

Problem 35

Problem 36 We define $H \dots H$ as the event that n coin tosses resulted in heads, and H_{n+1} as the event that the next coin toss results in heads and F be that we picked the fair coin. We wish to determine the probability

$$P(H_{n+1}|H \dots H)$$

From the law of total probability it becomes clear that

$$P(H_{n+1}) = P(H_{n+1}|F)P(F) + P(n_{+1}|\overline{F})P(\overline{F})$$

Conditioning both sides on $H \dots H$ we then get

$$P(H_{n+1}|H \dots H) = P(H_{n+1}|F, H \dots H)P(F|H \dots H) + P(n_{+1}|\overline{F}, H \dots H)P(\overline{F}|H \dots H)$$

The terms with two conditions are conditionally independent on $H \dots H$ as only the selected coin matters, as such

$$P(H_{n+1}|F, H \dots H) = \frac{1}{2}$$
$$P(H_{n+1}|\overline{F}, H \dots H) = 1$$

By using Bayes rule we can determine $P(F|H \dots H|)$ as

$$\begin{aligned}
P(F|H \dots H) &= \frac{P(H \dots H|F)P(F)}{P(H \dots H|F)P(F) + P(H \dots H|\overline{F})P(\overline{F})} \\
&= \frac{\frac{1}{2}^n \cdot \frac{1}{2}}{\frac{1}{2}^n \cdot \frac{1}{2} + 1 \cdot \frac{1}{2}} \\
&= \frac{\frac{1}{2}^n}{\frac{1}{2}^n + 1} \\
&= \frac{1}{1 + \frac{1}{\frac{1}{2}^n}} \\
&= \frac{1}{2^n + 1}
\end{aligned}$$

And $P(\overline{F}|H \dots H|)$ as its complement

$$P(\overline{F}|H \dots H) = 1 - \frac{1}{2^n + 1}$$

The probability of the n'th toss being must therefore be the sum of these probabitilies

$$\begin{aligned}
P(H_{n+1}|H\dots H) &= P(F|H\dots H)P(F) + P(\overline{F}|H\dots H) \\
&= \frac{1}{2}\cdot\frac{1}{2^n+1} + \left(1 - \frac{1}{2^n+1}\right) \\
&= \frac{1}{2(2^n+1)} + \left(1 - \frac{1}{2^n+1}\right) \\
&= 1 + \frac{-1}{2(2^n+1)} \\
&= 1 - \frac{1}{2(2^n+1)}
\end{aligned}$$

Problem 37

Problem 38

Problem 39

A.2 Chapter 2

Problem 1

Problem 2 We make use of combinations with $n = 12, k = 8$

$$\binom{12}{8} = \frac{12!}{8!(12-8)!} = 495$$

Problem 3 a)

The probability of picking exactly 4 black phones will be equal to

$$P(B = 4) = \frac{|B = 4|}{|S|}$$

As they are not put back we determine $|S|$ as the amount of 10-element subsets of a 50 element set using combinations

$$|S| = \binom{50}{10} = \frac{50!}{10!(50-10)!} = 10272278170$$

Simultaneously we can determine the amount of sets containing 4 black phones and 6 white phones as

$$|B = 4| = \binom{20}{4}\binom{30}{10-4} = \frac{20!}{4!(20-4)!} \cdot \frac{30!}{6!(30-6)!} = 2876839875$$

As such

$$P(B = 4) = \frac{2876839875}{10272278170} \approx 0.28$$

b) We wish to determine $P(B \in \{0, 1, 2\})$, we use the same method as the previous section and get that

$$\begin{aligned}
|B \in \{0, 1, 2\}| &= \sum_{i=0}^{2} \binom{20}{i}\binom{30}{10-i} \\
&= \binom{20}{0}\binom{30}{10} + \binom{20}{1}\binom{30}{9} + \binom{20}{2}\binom{30}{8} \\
&= 30045015 + 286143000 + 1112055750 \\
&= 1428243765
\end{aligned}$$

And again we determine the probability as

$$P(B \in \{0, 1, 2\}) = \frac{1428243765}{10272278170} \approx 0.14$$

Problem 4 a)

We once again make use of combinations, the sample space is given by

$$|S| = \binom{52}{5} = \frac{52!}{5!(52-5)!} = 2598960$$

The amount of 5 element sets consisting of 1 ace must be given by

$$|A=1| = \binom{4}{1}\binom{48}{4} = \frac{4!}{1!(4-1)!} \cdot \frac{48!}{4!(48-4)!} = 778320$$

Determining the probability is then as simple as

$$P(A=1) = \frac{778320}{2598960} \approx 0.30$$

b) The phrasing of the question suggests that its easier to determine the complement, as such we find the probability of getting no aces using the same method as in the previous section

$$|A=0| = \binom{4}{0}\binom{48}{5} = \frac{4!}{0!(4)!} \cdot \frac{48!}{5!(48-5)!} = 1712304$$

The probability is then

$$P(A=0) = \frac{1712304}{2598960} \approx 0.66$$

And we then determine the complement as this is equal to the probability of getting 1 or more aces

$$P(A \in \{1,2,3,4\}) = P(\overline{A=0}) = 1 - P(A=0) \approx 0.34$$

Problem 5 We wish to determine the probability

$$P(A=2|A \geq 1)$$

As the reverse conditional probability is independent of the condition we know that

$$P(A \geq 1|A=2) = 1$$

Using Bayes law we then find that

$$P(A=2|A \geq 1) = \frac{P(A \geq 1|A=2)P(A=2)}{P(A \geq 1)}$$

We determine $|A = 2|$ using the same method as in previous section

$$|A = 2| = \binom{4}{2}\binom{48}{3} = \frac{4!}{2!(4-2)!} \cdot \frac{48!}{3!(48-3)!} = 103776$$

And then make use of this to determine the probability of getting 2 aces

$$P(A = 2) = \frac{103776}{2598960} \approx 0.04$$

As such we have that

$$P(A = 2|A \geq 1) = \frac{1 \cdot 0.04}{0.34} \approx 0.12$$

Problem 6 Assuming the cards are dealt in bulk we can determine that

$$\begin{aligned} n_{\text{cards}} &= 52 - 26 = 26 \\ n_{\text{spades}} &= 13 - 7 = 6 \end{aligned}$$

As such 26 cards containing 6 spades will be dealt to player C, giving us that the sample space is

$$|S| = \binom{26}{13} = \frac{26!}{13!(26-13)!}$$

Whilst the amount of decks containing exactly 4 spades is

$$|C = 4| = \binom{6}{4}\binom{20}{13-4} = \frac{6!}{4!(6-4)!} \cdot \frac{20!}{9!(20-9)!}$$

As such the probability is

$$P(C = 4) = \frac{\binom{6}{4}\binom{20}{9}}{\binom{26}{13}} \approx 0.24$$

Problem 7 The sample space is given by

$$|S| = \binom{50}{15} = \frac{50!}{15!(50-15)!}$$

By inclusion exclusion we have that

$$P(Y \cup J) = P(Y) + P(J) - P(Y \cap J)$$

The amount of sets containing either you or Joe is

$$|Y| = |J| = \binom{1}{1}\binom{49}{14} = \frac{1!}{1!(1-1)!} \cdot \frac{49!}{14!(49-14)!}$$

The amount of sets containing both you and your friend Joe is

$$|Y, J| = \binom{2}{2}\binom{48}{13} = \frac{2!}{2!(2-2)!} \cdot \frac{48!}{13!(48-13)!}$$

As such the probability of both you and Joe being in group is

$$P(Y \cup J) = \frac{\binom{2}{2}\binom{48}{13}}{\binom{50}{15}} \approx 0.09$$

Inserting in the inclusion exclusion principle we then get

$$P(Y \cup J) = 2\frac{\binom{1}{1}\binom{49}{14}}{\binom{50}{15}} - \frac{\binom{2}{2}\binom{48}{13}}{\binom{50}{15}} \approx 0.51$$

Problem 8 For a set with n elements of which r are unique we have that there are

$$N = \frac{n!}{n_1!n_2!\dots n_r!}$$

Where n_i is the amount of repetitions of the given element, in Massachusetts we observe that

$$\begin{aligned} n_m &= 1 \\ n_a &= 2 \\ n_s &= 4 \\ n_c &= 1 \\ n_h &= 1 \\ n_u &= 1 \\ n_e &= 1 \\ n_t &= 2 \end{aligned}$$

As such

$$N = \frac{13!}{2!4!2!} = 64864800$$

Problem 9 a)

We use the binomial theorem and as such have that

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The probability of it containing 8 heads and 12 tails must simply be given by inserting $k = 8$ and determining the amount of possible sets using combinations

$$P(8) = \binom{20}{8} p^8 (1-p)^{20-8}$$

b)

Observing more than 8 heads and more than 8 tails is only possible when observing either 9, 10 or 11 heads, as such the probability must be given by

$$P(H \geq 8, T \geq 8) = \sum_{k=9}^{11} \binom{20}{k} p^k (1-p)^{20-k}$$

Problem 10 As every path will be described by some scrambled sequence of 20 movements to the right and 10 movements up, as such the problem boils down to finding out how many ways you can move right 20 times and up 10 times, which is given by

$$\frac{30!}{20!10!} = 30045015$$

Problem 11 We have the sample size from the previous problem, we determine the amount of paths that end up at $(10, 5)$ by determining the amount of paths that end at that point, this is given by

$$\frac{15!}{10!5!} = 3003$$

Afterwards there is once again 10 rights and 5 ups to reach the final point, giving us the same amount of paths once again, as such

$$P((10,5) \in \text{Path}) = \frac{\left(\frac{15!}{10!5!}\right)^2}{\frac{30!}{20!10!}} \approx 0.30$$

Problem 12 Using the binomial theorem we have that

$$P(k) = \binom{n}{k} p_a^k (1-p_a)^{n-k}$$

As p_a denotes the probability of moving up for it to reach the wanted point we must have that $k = 5$ and $n = 15$, therefore

$$P(5) = \binom{15}{5} p_a^5 (1-p_a)^{15-5}$$

Problem 13 a)

We make use of the law of total probability as well as the binomial theorem to write that

$$\begin{aligned} P(H \geq 3) &= \sum_{i=1}^{2} P(H=3|C_i)P(C_i) + P(H=4|C_i)P(C_i) + P(H=5|C_i)P(C_i) \\ &= 0.5 \sum_{i=1}^{2} \binom{5}{3} p_i^3 (1-p_i)^2 + \binom{5}{4} p_i^4 (1-p_i)^1 + \binom{5}{5} p_i^5 \\ &\approx 0.35 \end{aligned}$$

b)

We wish to determine the probability

$$P(C_2|H \geq 3)$$

We can easily determine the reverse conditional probability using the binomial theorem

$$\begin{aligned} P(H \geq 3|C_2) &= \binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p)^1 + \binom{5}{5} p^5 \\ &\approx 0.21 \end{aligned}$$

As we know $P(H \geq 3)$ from the previous question, we apply Bayes rule and get that

$$\begin{aligned} P(C_2|H \geq 3) &= \frac{P(H \geq 3|C_2)P(C_2)}{P(H \geq 3)} \\ &= \frac{0.21 \cdot 0.5}{0.35} \\ &= 0.3 \end{aligned}$$

Problem 14

Problem 15

Problem 16

Problem 17

Problem 18

Problem 19 We apply that the number of possible solutions for an equation equating to k with n distinct elements is given by

$$\binom{k+n-1}{k}$$

As the x_i's are limited to the natural numbers we rewrite such that we are working in the domain $\{0, 1, 2, \ldots\}$ by stating that

$$y_i = x_i - 1 \Leftrightarrow x_i = y_i + 1$$

Rewriting the equation we then get

$$y_1 + y_2 + y_3 + y_4 + y_5 = 95$$

As such the number of solutions is given by

$$\binom{95+5-1}{k} = \frac{99!}{95!} = 3764376$$

Problem 20 We can determine that as the sum of solutions of the equation

$$x_2 + x_3 + x_4 = 100 - x_1$$

We rewrite x_1 as i and thus this can be expressed as

$$x_2 + x_3 + x_4 = 100 - i$$

By sum we then have

$$\begin{aligned} n_{\text{solutions}} &= \sum_{i=0}^{10} \binom{100-i+3-1}{100-i} \\ &= \binom{102}{100} + \binom{101}{99} + \ldots + \binom{92}{90} \\ &= 51271 \end{aligned}$$

Problem 21

A.3 Chapter 3

Problem 1 a)

$$R_X = \{0, 1, 2\}$$

b)

$$P(X \geq 1.5) = P(x = 2) = \frac{1}{6}$$

c)

$$P(0 < X < 2) = P(x = 1) = \frac{1}{3}$$

d)

$$P(X = 0|X < 2) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3}} = 0.6$$

Problem 2

$$\begin{aligned} P(X = 2) &= P(X = 1) = \frac{P(X = 3)}{2} \\ P(X = 0) &= P(X = 3) \end{aligned}$$

As the total probability must be 3 its clear that

$$\begin{aligned} 1 &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= P(X = 3) + \frac{P(X = 3)}{2} + \frac{P(X = 3)}{2} + P(X = 3) \\ &= 3P(X) \end{aligned}$$

Solving for $P(X = 3)$ we get

$$P(X = 3) = \frac{1}{3}$$

As such we get that the PMF for the function is given by

$$P_X(x) = \begin{cases} \frac{1}{3} & x = 0 \\ \frac{1}{6} & x = 1 \\ \frac{1}{6} & x = 2 \\ \frac{1}{3} & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Problem 3

Problem 4 As they are independent we determine that

a)

$$\begin{aligned} P(X \le 2 \cap Y \le 2) &= P(X \le 2)P(Y \le 2) \\ &= (\frac{1}{4}+\frac{1}{8})(\frac{1}{6}+\frac{1}{6}) \qquad = \frac{1}{8} \end{aligned}$$

b) By inclusion exclusion

$$\begin{aligned} P(X > 2 \cup Y > 2) &= P(X > 2) + P(Y > 2) - P(X > 2 \cap Y > 2) \\ &= \frac{1}{8}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3} - P(X > 2)P(Y > 2) \\ &= \frac{1}{8}+\frac{1}{2}+\frac{1}{3}+\frac{1}{3} - (\frac{1}{8}+\frac{1}{2})(\frac{1}{3}+\frac{1}{3}) \\ &= \frac{7}{8} \end{aligned}$$

c) As they are independent we have that

$$\begin{aligned} P(X > 2 | Y > 2) &= P(X > 2) \\ &= \frac{1}{8}+\frac{1}{2} = \frac{5}{8} \end{aligned}$$

d) We make use of the law of total probability and condition and get that

$$\begin{aligned} P(X < Y) &= \sum_{i=1}^{4} P(X < Y | Y = i)P(Y = i) \\ &= \sum_{i=1}^{4} P(X < i)P(Y = i) \\ &= \frac{1}{4}\cdot\frac{1}{6} + \left(\frac{1}{4}+\frac{1}{8}\right)\cdot\frac{1}{3} + \left(\frac{1}{4}+\frac{1}{8}+\frac{1}{8}\right)\cdot\frac{1}{3} \\ &= \frac{1}{3} \end{aligned}$$

Problem 5 We make use of the binomial theorem and write that

$$P(k) = \binom{50}{k}\left(\frac{1}{2}\right)^{k}\left(1-\frac{1}{2}\right)^{n-k}$$

We sum up the final 20 terms as this will be the probability of there being more than 30 cars

$$\begin{aligned} P(C > 30) &= \sum_{i=31}^{50} \binom{50}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k} \\ &\approx 0.06 \end{aligned}$$

Problem 6

Problem 7 We note that the probabilities we wish to determine can be written as

$$\begin{aligned} P(X > 5) &= 1 - \sum_{k=\min(R_X)}^{5} P_X(k) \\ P(2 < X \leq 6) &= \sum_{k=3}^{6} P_X(k) \\ P(X > 5 | X < 8) &= \frac{\sum_{k=6}^{7} P_X(k)}{\sum_{k=\min(R_X)}^{7} P_X(k)} \end{aligned}$$

a)

$$P_X(k) = \begin{cases} \frac{1}{5}\left(1 - \frac{1}{5}\right)^{k-1} & k \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} P(X > 5) &= 1 - \sum_{k=1}^{5} P_X(k) \\ P(2 < X \leq 6) &= \sum_{k=3}^{6} P_X(k) \\ P(X > 5 | X < 8) &= \frac{\sum_{k=6}^{7} P_X(k)}{\sum_{k=1}^{7} P_X(k)} \end{aligned}$$

b)

$$P_X(k) = \begin{cases} \begin{pmatrix} 10 \\ k \end{pmatrix} \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)^{n-k} & k \in \{0, 1, 2, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

$$P(X > 5) = 1 - \sum_{k=0}^{5} P_X(k)$$

$$P(2 < X \leq 6) = \sum_{k=3}^{6} P_X(k)$$

$$P(X > 5 | X < 8) = \frac{\sum_{k=6}^{7} P_X(k)}{\sum_{k=0}^{7} P_X(k)}$$

c)

$$P_X(k) = \begin{cases} \binom{k-1}{2} \left(\frac{1}{2}\right)^m \left(1 - \frac{1}{2}\right)^{k-m} & k \in \{3, 4, 5, \ldots\} \\ 0 & \text{otherwise} \end{cases}$$

$$P(X > 5) = 1 - \sum_{k=3}^{5} P_X(k)$$

$$P(2 < X \leq 6) = \sum_{k=3}^{6} P_X(k)$$

$$P(X > 5 | X < 8) = \frac{\sum_{k=6}^{7} P_X(k)}{\sum_{k=3}^{7} P_X(k)}$$

d)

$$P_X(k) = \begin{cases} \dfrac{\binom{10}{x}\binom{10}{12-x}}{\binom{10+10}{12}} & x \in \{2, 3, \ldots, 10\} \\ 0 & \text{otherwise} \end{cases}$$

$$P(X > 5) = 1 - \sum_{k=2}^{5} P_X(x)$$

$$P(2 < X \leq 6) = \sum_{k=3}^{6} P_X(x)$$

$$P(X > 5 | X < 8) = \frac{\sum_{x=6}^{7} P_X(x)}{\sum_{x=2}^{7} P_X(x)}$$

e) We assume its range is given by $R_X = \{0, 1, 2, \dots\}$

$$P_X(k) = \begin{cases} \frac{e^{-5}5^k}{k!} & k \in R_X \\ 0 & \text{otherwise} \end{cases}$$

$$P(X > 5) = 1 - \sum_{k=0}^{5} P_X(k)$$
$$P(2 < X \leq 6) = \sum_{k=3}^{6} P_X(k)$$
$$P(X > 5 | X < 8) = \frac{\sum_{k=6}^{7} P_X(k)}{\sum_{k=0}^{7} P_X(k)}$$

Problem 8 As the different exams are independent the probability of a given sequence will be given by

$$\begin{aligned} P(X = k) &= P(F_1)P(F_2)\dots P(F_{k-1})P(S_k) \\ &= \left[\prod_{j=0}^{k-1} P(F_j)\right] P(S_k) \\ &= \left[\prod_{j=0}^{k-1} \left(\frac{1}{2}\right)^j\right] \left(1 - \left(\frac{1}{2}\right)^k\right) \end{aligned}$$

a)

$$P(X = 1) = \left[\prod_{j=0}^{0} \left(\frac{1}{2}\right)^j\right] \left(1 - \left(\frac{1}{2}\right)^1\right) = \frac{1}{2}$$
$$P(X = 2) = \left[\prod_{j=0}^{1} \left(\frac{1}{2}\right)^j\right] \left(1 - \left(\frac{1}{2}\right)^2\right) = \frac{3}{8}$$
$$P(X = 3) = \left[\prod_{j=0}^{2} \left(\frac{1}{2}\right)^j\right] \left(1 - \left(\frac{1}{2}\right)^3\right) = \frac{7}{64}$$

b) Done in opening of question c) This must be the complement of finishing the quest in the first 2 attempts, as such

$$\begin{aligned} P(X > 2) &= 1 - \left(\frac{1}{2} + \frac{3}{8}\right) \\ &= 1 - \frac{7}{8} \\ &= \frac{1}{8} \end{aligned}$$

d) This is the probability

$$P(X = 2 | X > 1)$$

By the definition of conditional probability this is

$$\begin{aligned} P(X = 2 | X > 1) &= \frac{P(X = 2)}{P(X > 1)} \\ &= \frac{P(X = 2)}{P(\overline{X = 1})} \\ &= \frac{\frac{3}{8}}{1 - \frac{1}{2}} \\ &= \frac{3}{4} \end{aligned}$$

Problem 9 We wish to show that

$$P_X(k) = p(1-p)^{k-1} \implies P(X > m + l | X > m) = P(X > l)$$

From the condition we already know that $X > m$, as such, we can interpret this as the first m experiments having already been conducted resulting in $P(X \leq m) = 1$, this makes the first m experiments redundant, as we are only interested in the experiments following the first m experiments, resulting in only the values of X that are larger than l being relevant.

Problem 10 a) We make use of combinations and get that

$$\begin{aligned} P(R = 4) &= \frac{|R = 4||G = 10 - 4|}{|S|} \\ &= \frac{\binom{20}{4}\binom{30}{10-4}}{\binom{50}{10}} \\ &\approx 0.28 \end{aligned}$$

b) We wish to determine the probability

$$P(R = 4|R \geq 3)$$

As the opposite conditional is conditionally independent we make use of Bayes rule

$$\begin{aligned}
P(R = 4|R \geq 3) &= \frac{P(R \geq 3|R = 4)P(R = 4)}{P(R \geq 3)} \\
&= \frac{1 \times 0.28}{1 - \sum_{i=0}^{2} \binom{20}{i} \binom{30}{10-i}} \\
&\approx 0.33
\end{aligned}$$

Problem 11 We have that

$$\begin{aligned}
X_{weekday} &\sim Poisson\left(\frac{1}{6}\right) \\
X_{weekend} &\sim Poisson\left(\frac{1}{30}\right)
\end{aligned}$$

a) We make use of the Poisson distribution, multiplying the argument by 4×60 and setting $k = 0$ after which we find the probability

$$P_X(k) = \frac{e^{-4\times 60\times \frac{1}{30}} \times \left(4 \times 60 \times \frac{1}{30}\right)^k}{k!}$$

We set $k = 0$ to determine the probability during the 4 hour interval and get

$$\begin{aligned}
P_X(0) &= \frac{e^{-8} \times 8^0}{0!} \\
&= e^{-8} \\
&\approx 3.4 \times 10^{-4}
\end{aligned}$$

b) We make use of the law of total probability as well as the definition of conditional probability

$$\begin{aligned}
P(\text{weekday}|k = 0) &= \frac{P(\text{weekday} \cap k = 0)}{\frac{5}{7}P_{X_{\text{weekday}}}(k) + \frac{2}{7}P_{X_{\text{weekend}}}(k)} \\
&= \frac{\frac{5}{7}e^{-10}}{\frac{5}{7}e^{-10} + \frac{2}{7}e^{-2}} \\
&\approx 8.4 \times 10^{-4}
\end{aligned}$$

Problem 12

$$F_X(x) = \begin{cases} 0 & x < -2 \\ 0.2 & -2 \leq x < -1 \\ 0.5 & -1 \leq x < -0 \\ 0.7 & 0 \leq x < 1 \\ 0.9 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Problem 13 a)

$$R_X = \{0, 1, 2, 3\}$$

b)

$$P_X(x) = \begin{cases} \frac{1}{6} & x = 0 \\ \frac{1}{3} & x = 1 \\ \frac{1}{4} & x = 2 \\ \frac{1}{4} & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

Problem 14 a) Making use of LOTUS we get that

$$\begin{aligned} EX &= \sum_{x=1}^{3} x P_X(x) \\ &= 1 \times 0.5 + 2 \times 0.3 + 3 \times 0.2 \\ &= 0.5 + 0.6 + 0.6 \\ &= 1.7 \end{aligned}$$

b) We compute the variance using $\text{Var}(X) = E[X^2] - EX^2$, determining the first term using LOTUS we get

$$\begin{aligned} E[X^2] = \sum_{x=1}^{3} x^2 P_X(x) & \qquad = 1^2 \times 0.5 + 2^2 \times 0.3 + 3^2 \times 0.2 \\ = 3.5 & \end{aligned}$$

To determine the standard deviation we square the variance

$$\text{SD}(X) = \sqrt{3.5}$$

c) Making use of LOTUS we have that

$$\begin{aligned} EY &= \sum_{x\in R_X} \frac{2}{x} P_X(x) \\ &= \frac{2}{1}\times 0.5 + \frac{2}{2}\times 0.3 + \frac{2}{3}\times 0.2 \\ &\approx 1.43 \end{aligned}$$

Problem 15

Problem 16

Problem 17 By LOTUS we have that

$$\begin{aligned} EX &= \sum_{k\in R_X} kp(1-p)^{k-1} \\ &= \sum_{k=1}^{\infty} k(1-p)^{k-1} \\ &= p\sum_{k=1}^{\infty}(1-p)^{k-1} \end{aligned}$$

We wish to make use of the geometric sum and do so by taking the first derivative with respect to x on both sides

$$\begin{aligned} \frac{d}{dx}\sum_{k=0}^{\infty} x^k &= \frac{d}{dx}\frac{1}{1-x} \\ \sum_{k=0}^{\infty} kx^{k-1} &= \frac{1}{(1-x)^2} \end{aligned}$$

Which we make use of using the form written above

$$\begin{aligned} EX &= p\sum_{k=1}^{\infty} k(1-p)^{k-1} \\ &= p\frac{1}{(1-(1-p))^2} \\ &= p\frac{1}{p^2} \\ &= \frac{1}{p} \end{aligned}$$

Problem 18 As the Pascal is the sum of a sequence of geometric random variables we can write that

$$\begin{aligned} EX &= \binom{n-1}{k-1} p^{k-1})(1-p)^{n-k} \\ &= EX_1 + EX_2 + \ldots + EX_n \\ &= \frac{1}{p} + \frac{1}{p} + \ldots + \frac{1}{p} \\ &= \frac{n}{p} \end{aligned}$$

Problem 19

Problem 20 We define the value k as the amount of people in a household, as such the PMF is given by

$$P_X(k) = \begin{cases} 0.1 & k = 1 \\ 0.2 & k = 2 \\ 0.3 & k = 3 \\ 0.2 & k = 4 \\ 0.1 & k = 5 \\ 0.1 & k = 6 \\ 0 & \text{otherwise} \end{cases}$$

We now determine the expected value using LOTUS

$$\begin{aligned} EX &= \sum_{k=1}^{6} x_k P_X(k) \\ &= 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.2 + 5 \times 0.1 + 6 \times 0.1 \\ &= 3.3 \end{aligned}$$

b) We use the same method as before, except we have to weigh the amount of total people in a house with k people proportional to the total amount of people meaning that the probability of each $k \in R_X$ is given by $\frac{k \times N_k}{3300}$

$$P_X(k) = \begin{cases} \frac{1}{33} & k = 1 \\ \frac{4}{33} & k = 2 \\ \frac{3}{11} & k = 3 \\ \frac{8}{33} & k = 4 \\ \frac{5}{33} & k = 5 \\ \frac{2}{11} & k = 6 \\ 0 & \text{otherwise} \end{cases}$$

We again determine the expected value using LOTUS

$$\begin{aligned} EX &= \sum_{k=1}^{6} x_k P_X(k) \\ &= 1 \times \frac{1}{33} + 2 \times \frac{4}{33} + 3 \times \frac{3}{11} + 4 \times \frac{8}{33} + 5 \times \frac{5}{33} + 6 \times \frac{2}{11} \\ &= \frac{43}{11} \end{aligned}$$

Problem 21

Problem 22 The probability of getting a sequence of k heads followed by tails on a fair coin is given by the geometric distribution

$$X \sim Geometric\left(\frac{1}{2}\right)$$

a)

$$\begin{aligned} EX &= \sum_{k=1}^{\infty} 2^{k-1} P_X(k) \\ &= \infty \end{aligned}$$

b) For the player to win more than 65 dollars k has to satisfy the inequality

$$2^{k-1} > 65 \implies k > 7$$

As such the probability of winning more than 65 dollars must be the complement of getting $k \leq 7$, meaning that

$$
\begin{aligned}
P(W > 65) &= P(\overline{k \leq 7}) \\
&= 1 - \sum_{k=1}^{7} p(1-p)^{k-1} \\
&= 1 - \left(0.5^1 + 0.5^2 + \ldots + 0.5^7\right) \\
&= \frac{1}{128}
\end{aligned}
$$

c) This places a limit onto the values that k can have, as such we rewrite the LOTUS from previous

$$
\begin{aligned}
EX &= \sum_{k=1}^{31} 2^{k-1} P_X(k) + \sum_{k=32}^{\infty} 2^{30} P_X(k) \\
&= \frac{1}{2} \sum_{k=1}^{31} 2^k \left(\frac{1}{2}\right)^k + 2^{30} \sum_{k=32}^{\infty} \left(\frac{1}{2}\right)^k \\
&= \frac{1}{2} \sum_{k=1}^{31} 1 + 2^{30} \times \left(1 - \sum_{k=1}^{31} \left(\frac{1}{2}\right)^k\right) \\
&= \frac{31}{2} + 2^{30} \times (1 - 0.9999999995343387) \\
&= 16
\end{aligned}
$$

Problem 23 By linearity we get that

$$
\begin{aligned}
E[(X-\alpha)^2] &= E[X^2 + \alpha^2 - 2X\alpha] \\
&= E[X^2] + \alpha^2 - 2X\alpha
\end{aligned}
$$

As only the 2nd and 3rd terms are dependent on α we wish to minimize this, we do so by taking the first derivative and solving when its equal 0

$$
\frac{d}{d\alpha}(\alpha^2 - 2X\alpha) = 2\alpha - 2X
$$

We set this equal zero to identify its extremum and get

$$
\begin{aligned}
2\alpha - 2X &= 0 \\
2\alpha &= 2X \\
\alpha &= X
\end{aligned}
$$

And as $X = \mu$ we get that the function is minimized when

$$\alpha = \mu$$

Problem 24

Problem 25 a) We write the PDF as a CDF as this lets us identify the interval containing the midpoint of probability summation

$$F_X(k) = \begin{cases} 0.4 & k \leq 1 \\ 0.7 & 1 < k \leq 2 \\ 1 & 2 < k \leq 3 \end{cases}$$

Here we see that the interval containing the midpoint of probabilities is the one with $k = 2$, as such the median is 2.

b) As this is a distribution with an even amount of outcomes with equal probabilities we list them numerically

$$S = \{1, 2, 3, 4, 5, 6\}$$

And identify the 2 middle numbers as $3, 4$, we take the average of this as the amount of elements in the sample space is even

$$\frac{3+4}{2} = 3.5$$

As such the median is 3.5.

c)