PROBABILITY THEORY

Introduction to Probability, Statistics and Random Processes

Hossein Pishro-Nik

Contents

PROBABILITY THEORY NOTES

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:

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

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. $\qquad\square$

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. 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:

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

1.1.3 Problems

Problem 3 a) Let $S = \{1, 2, 3\}$. Write all possible partitions of S.

As a partition is any collection of disjoint sets whos union makes up S we have that

1. $\{1\},\{2\},\{3\}$
2. $\{1,2\},\{3\}$
3. $\{1\},\{2,3\}$
4. $\{1,3\},\{2\}$
5. $\{1,2,3\}$

Problem 4 a) Determine whether each of the following sets are countable or countable:

- $A = \{x \in \mathbb{Q}| -100 \leq x \leq 100\}$
- $B = \{(x,y)|x \in \mathbb{N}, y \in \mathbb{Z}\}$
- $C =]0, 0.1]$
- $D = \left\{\frac{1}{n}|n \in \mathbb{N}\right\}$

As $A \subset \mathbb{Q}$ it is clear that it must be countable.

As B is the carthesian product of 2 countable sets it must be countable.

As C is a range it must be uncountable.

As D can be written in one-to-one correspondance with the naturals it must be countable.

Problem 5 a) Find the range of the function $f : \mathbb{R} \xrightarrow{\sin(x)} \mathbb{R}$.

As $\sin(x)$ has its extrema at $\sin\left(\frac{\pi}{2}\right) = 1$ and $\sin\left(\frac{3\pi}{2} = -1\right)$, it is clear that

$$\text{Range}(f) = [-1, 1]$$

1.2   Random experiments

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

Definition 3. 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:

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

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

7

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

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

$\square$

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

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

$\square$

1.2.2   Problems

Problem 2   Write the sample space, S, for the following random experiments:

a) We toss a coin until we see two consecutive tails. We record the number of coin tosses.

$$S = \{n \in \mathbb{N} | n \geq 2\}$$

b) A bag contains 4 balls: one is red, one is blue, one is white and one is green. We choose two distinct balls and record their color in order.

$$S = \{(R, B), (R, W), (R, G), (B, R), (B, W)(B, G), (W, R),$$
$$(W, B), (W, G), (G, R), (G, B), (G, W)\}$$

c) A customer arrives at a bank and waits in the line. We observe T, which is the total time (in hours) that the customer waits in the line. The bank has a strict policy that no customer waits more than 20 minutes under any circumstances.

$$S = \left[0; \frac{1}{3}\right]$$

Problem 3   Let A, B and C be three events in the sample space S. Suppose we know:

- $A \cup B \cup C = S$
- $P(A) = \frac{1}{2}$
- $P(B) = \frac{2}{3}$
- $P(A \cup B) = \frac{5}{6}$

Answer the following questions:

a) Find $P(A \cap B)$.

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

Inserting known information we get that

$$P(A \cap B) = \frac{1}{2} + \frac{2}{3} - \frac{5}{6} = \frac{1}{3}$$

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

No, as they aren't disjoint.

c) Find $P(C - (A \cup B))$.

We rewrite the event as

$$C \setminus (A \cup B) = (C \cup (A \cup B)) - (A \cup B)$$

As their union makes up the sample space

$$C \setminus (A \cup B) = S \setminus (A \cup B) = \overline{A \cup B}$$

Which we know is equal to

$$P(C \setminus (A \cup B)) = 1 - \frac{5}{6} = \frac{1}{6}$$

d) If $P(C \cap (A \cup B)) = \frac{5}{12}$, find $P(C)$.

As C must be possible to construct from the intersection of C with A and B and the difference between them we get that

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

By inserting we get

$$P(C) = \frac{5}{12} + \frac{1}{6} = \frac{7}{12}$$

Problem 4   I roll a fair die twice and obtain two numbers, $X_1 =$ result of the first roll, and $X_2 =$ result of the second roll. Find the probability of the following events.

a) A defined as $X_1 < X_2$.

We write the sample space

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

Is is clear that $|S| = 36$ and $|A| = 15$, as such

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

b) B defined $|\{6\}| \geq 1$.

From the sample space we get that $|B| = 11$, as such

$$P(B) = \frac{11}{36}$$

Problem 5 You purchase a certain product. The manual states that the lifetime, T, of the product defined as the amount of time (in years) the product works properly until it breaks down satisfies:

$$P(T \geq t) = e^{-\frac{t}{5}} \text{ for all } t \geq 0$$

For examples, the probability that the product lasts more than (or equal to) 2 years is

$$P(T \geq 2) = e^{-\frac{2}{5}} = 0.6703$$

a) This is an example of a continuous probability model. Write down the sample space, S.

As $t \geq 0$ we get that

$$S = [0; \infty[$$

b) Check that the statement in the manual makes sense by finding $P(T \geq 0)$ and $\lim_{t \to \infty} P(T \geq t)$

We insert that $t = 0$ and get

$$P(T \geq 0) = e^{-\frac{0}{5}} = e^0 = 1$$

As such it makes sense, we take the limit as $t \to \infty$

$$\lim_{t \to \infty} e^{-\frac{t}{5}} = 0$$

Which also makes sense.

c) Also check that if $t_1 < t_2$ then $P(T \geq t_1) \geq P(T \geq t_2)$. Why does this need to be true?

We assume $t_1 = 1, t_2 = 2$ and get

$$P(T \geq 1) = 0.8187 \qquad P(T \geq 2) = 0.6703$$

Whereby the statement is true. This has to be true as it is an exponential model, resulting in the monotony constantly being negative, as such the value must fall every time t is increased.

d) Find the probability that the product breaks down within three years of the purchase time.

We have that
$$P(T \geq 3) = e^{-\frac{3}{5}} = 0.5488$$

As this is the probability that it last longer than or equal to 3 years, the complement must be

$$P(T < 3) = 1 - P(T \geq 3) = 1 - 0.5488 = 0.4512$$

e) Find the probability that the product breaks down in the second year, i.e., find $P(1 \leq T < 2)$.

We determine the probability of it breaking down within 2 years and then subtract the probability of it breaking down within 1. Using results from c) we have
$$P(T < 1) = 0.1813 \qquad P(T < 2) = 0.3297$$

Subtracting the two we get that

$$P(1 \geq T < 2) = 0.3297 - 0.1813 = 0.1484$$

Problem 6   You get a stick and break it randomly into three pieces. What is the probability that you can make a triangle using the three pieces? You can assume the break points are chosen completely at random, i.e. if the length of the original stick is 1 unit, and x, y, z are the lengths of the three pieces, then (x, y, z) are uniformly chosen from the set

$$\{(x, y, z) \in \mathbb{R}^3 | x + y + z = 1, x, y, z \geq 0\}$$

By the triangle inequality we have that

$$x < y + z$$
$$y < x + z$$
$$z < x + y$$

Due to the sum requirement, the only time this is satisfied is when

$$x < \frac{1}{2}$$
$$y < \frac{1}{2}$$
$$z < \frac{1}{2}$$

Furthermore we have that the entire sample space must stretch a 3 dimensional plane from $P_1(1,0,0), P_2(0,1,0), P_3(0,0,1)$, while the possible values for forming a triangle must stretch one from $P_1(0.5,0,0), P_2(0,0.5,0), P_3(0,0,0.5)$. The area of this plane is $\frac{1}{4}$ of the sample spaces plane, and therefore

$$P(A) = \frac{1}{4}$$

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

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

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.

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

$\square$

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

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

$\square$

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

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

$\square$

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

$\square$

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

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

$\square$

1.3.1   Independence

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

Definition 4. 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

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

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

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

$\square$

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 $B_1, B_2, \ldots, B_n$ 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

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

$\square$

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

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

$\square$

1.3.3   Conditional Independence

Definition 5. 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

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

$$\square$$

1.3.4   Problems

Problem 1   You purchase a certain product. The manual states that the lifetime, T, of the product defined as the amount of time (in years) the product works properly untilk it breaks down satisfies:

$$P(T \geq t) = e^{-\frac{t}{5}} \text{ for all } t \geq 0$$

For example, the probability that the product lasts more than (or equal to) 2 years is

$$P(T \geq 2) = e^{-\frac{2}{5}} = 0.6703$$

a) I purchase the product and use it for two years without any problems. What is the probability that it breaks down in the 3rd year?

We define A as the probability of the product not breaking down in the first two years

$$P(A) = P(T \geq 2) = e^{-\frac{2}{5}} = 0.6703$$

Similarly we define B as the probability of it breaking down in the 3rd year

$$P(B) = P(2 \leq T \leq 3) = P(T \geq 2) - P(T \geq 3) = e^{-\frac{2}{5}} - e^{-\frac{3}{5}} = 0.1215$$

We now wish to determine the conditional probability $P(B|A)$

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

As $B \subset A$ it is clear that $A \cap B = B$

$$P(B|A) = \frac{P(B)}{P(A)} = \frac{0.1215}{0.6703} = 0.1812$$

Problem 2 You toss a fair coin three times

a) What is the probability of three heads, HHH?

As the coin tosses are independent we have that

$$P(H \cap H \cap H) = P(H)P(H)P(H) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

b) What is the probability that you observe exactly one heads?

The probability of observing exactly one heads must be the sum of the probabilities of

$$A = \{H, T, T\}$$
$$B = \{T, H, T\}$$
$$C = \{T, T, H\}$$

As the likelihood of any sequence is equal we have that

$$P(HHH) = P(A) = P(B) = P(C)$$

As such
$$P(1H) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$

c) Given that you have observed at least one heads, what is the probability that you observe at least two heads?

We wish to determine the probability $P(A|B)$ where A is observing ≥ 2 heads and B is knowing there has already been observed at least one. The favourable events for B are

$$P(B) = P(S \setminus \{T, T, T\}) = 1 - \frac{1}{8} = \frac{7}{8}$$

Whilst the favourable events for A are

$$P(A) = P(\{(H,H,T),(H,T,H),(T,H,H),(H,H,H)\}) = \frac{4}{8} = \frac{1}{2}$$

By the definition of conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{2}}{\frac{7}{8}} = \frac{4}{7}$$

Problem 3   For three events, A, B and C, we know that

- A and C are independent.

- B and C are independent.

- A and B are disjoint.

- $P(A \cup C) = \frac{2}{3}, P(B \cup C) = \frac{3}{4}$ and $P(A \cup B \cup C) = \frac{11}{12}$.

a) Find $P(A), P(B)$ and $P(C)$.

As these are all probabilities of unions we have that

$$P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{2}{3}$$
$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = \frac{3}{4}$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap C) - P(B \cap C) = \frac{11}{12}$$

We add the first 2 equations together and get

$$P(A \cup C) + P(B \cup C) = P(A) + P(B) + 2P(C) - P(A \cap C) - P(B \cap C)$$

Subtracting the third equation from these we then get

$$[P(A) + P(B) + 2P(C) - P(A \cap C) - P(B \cap C)] - [P(A) + P(B)+$$
$$P(C) - P(A \cap C) - P(B \cap C)] = P(C) = \frac{2}{3} + \frac{3}{4} - \frac{11}{12} = \frac{6}{12} = \frac{1}{2}$$

Which we then substitute into the other equations

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

As such $P(A) = \frac{1}{3}, P(B) = \frac{1}{2}$ and $P(C) = \frac{1}{2}$.

Problem 4   Let $C_1, C_2, \ldots, C_M$ be a partition of the sample space, S, and A, B be two events. Suppose we know that

- A and B are conditionally independent given C_i for all $i \in \{1, 2, \ldots, M\}$.

- B is independent of all C_i's.

a) Prove that A and B are independent.

We have that

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

By the law of total probability

$$
\begin{aligned}
P(A \cap B) &= \sum_{i=1}^{M} P(A \cap B|C_i)P(C_i) \\
&= \sum_{i=1}^{M} P(A|C_i)P(B|C_i)P(C_i) \ A, B \text{ are conditionally independent} \\
&= \sum_{i=1}^{M} P(A|C_i)P(B)P(C_i) \ B \text{ is indepedet of } C \\
&= P(B) \sum_{i=1}^{M} P(A|C_i)P(C_i) \\
&= P(B)P(A)
\end{aligned}
$$

Problem 5   In my town, it's rainy one third of the days. Given that it is rainy, there will be heavy traffic with probability $\frac{1}{2}$, and given that it is not rainy, there will be heavy traffic with probability $\frac{1}{4}$. If it's rainy and there is heavy traffic, I arrive late for work with probability $\frac{1}{2}$. On the other hand, the probability of being late is reduced to $\frac{1}{8}$ if it is not rainy and there is no heavy traffic. In other situations (rainy and no traffic, not rainy and traffic) the probability of being late is $\frac{1}{4}$. You pick a random day:

We define the events, R, T, L corresponding to it being rainy, traffic being heavy and the person being late, based on this a probability tree is made.

a) What is the probability that it's not raining and there is heavy traffic and I am not late?

We read from the probability tree that

$$P(\overline{L}|\overline{R}, T) = \frac{1}{8}$$

b) What is the probability that I am late?

We sum the probabilities that result in being late

$$P(L) = \frac{1}{12} + \frac{1}{24} + \frac{1}{24} + \frac{1}{16} = \frac{11}{48}$$

c) Given that I arrived late at work, what is the probability that it rained that day?

$$P(R|L) = \frac{P(R \cap L)}{P(L)} = \frac{\frac{1}{12} + \frac{1}{24}}{\frac{11}{48}} = \frac{6}{11}$$

Problem 6   A box contains three coins: two regular coins and one fake two-headed coin.

a) You pick a coin at random and toss it. What is the probability that it lands heads up?

The probability of getting heads must be the union of the probabilities of getting heads with the individual coins.

$$P(H) = P(H|C_1)P(C_1) + P(H|C_2)P(C_2) + P(H|C_3)P(C_3)$$
$$= \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}$$
$$= \frac{2}{3}$$

b) You pick a coin at random and toss it, and get heads, what is the probability that it is the two headed coin?

We apply Bayes rule and have that

$$P(C_3|H) = \frac{P(H|C_3)P(C_3)}{P(H)}$$
$$= \frac{1 \cdot \frac{1}{3}}{\frac{2}{3}}$$
$$= \frac{1}{2}$$

Problem 7   A family has two children. We ask the father, "Do you have at least one daughter named Lilia?". He replies, "yes", what is the probability that both children are girls? In other words, we want to find the probability that both children are girls, given that the family has at least one daughter named Lilia. Here you can assume that if a child is a girl, her name will be Lilia with a probability $\alpha \ll 1$ independently from other children's names. If the child is a boy, his name will not be Lilia

The sample space is

$$P(S) = \{(B, B), (B, G), (G, G), (G, B)$$

Given that we know they have at least one daughter named Lilia we have

$$P(L|BB) = 0$$
$$P(L|GB) = P(L|BG) = \alpha$$
$$P(L|GG) = \alpha(1 - \alpha) + (1 - \alpha)\alpha + \alpha^2 = 2\alpha - 2\alpha^2$$

Applying Bayes rule we get that

$$
\begin{aligned}
P(GG|L) &= \frac{P(L|GG)P(GG)}{P(L)} \\
&= \frac{(2\alpha - \alpha^2)\frac{1}{4}}{(2\alpha - \alpha^2)\frac{1}{4} + \alpha\frac{1}{4} + \alpha\frac{1}{4} + 0\frac{1}{4}} \\
&= \frac{0.5a - 0.25a^2}{(2\alpha - \alpha^2)\frac{1}{4} + \alpha\frac{1}{2}} \\
&= \frac{\frac{1}{4}\alpha(-\alpha + 2)}{\frac{1}{4}\alpha(-\alpha + 4)} \\
&= \frac{2 - \alpha}{4 - \alpha} \\
&\approx \frac{1}{2} \text{ as } \alpha \ll 1
\end{aligned}
$$

Problem 8   A family has two children. We choose one at random and find out that she is a girl. What is the probability that both children are girls?

$$P(GG) = P(GB) = P(BG) = P(BB)$$
$$P(G_r|GG) = 1$$

$$P(G_r|BG) = P(G_r|GB) = \frac{1}{2}P(G_r|BB) = 0$$

Using Bayes rule we determine $P(GG|G_r)$

$$
\begin{aligned}
P(GG|G_r) &= \frac{P(G_r|GG)P(GG)}{P(G_r)} \\
&= \frac{1 \cdot \frac{1}{4}}{\frac{1}{4}(1 + \frac{1}{2} + \frac{1}{2})} \\
&= \frac{1}{2}
\end{aligned}
$$

Problem 9 (NOT SOLVED AS I DONT GET THE EXPLANATION)
I toss an unfair coin where $P(H) = p$ repeatedly. The game ends the first time two consecutive heads or tails is observed. I win if two heads are observed and lose if two tails are observed.

The probability of tails must be the complement of heads

$$
P(T) = 1 - p
$$

The probability of winning is given by

$$
\begin{aligned}
P(W) = P(\{(HH),(HTHH),(HTHTHH),\ldots\}) + \\
P(\{(THH),(THTHH),(THTHTHH,\ldots)\})
\end{aligned}
$$

In regards to the above statement we then have that

$$
\begin{aligned}
P(W) = P(H)^2 + P(H)^3 P(T) + P(H)^4 P(T)^2 + \ldots + \\
P(T)^1 P(H)^2 + P(T)^2 P(H)^3 + P(T)^3 + P(H)^4
\end{aligned}
$$

1.4 Chapter problems

1.4.1 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)

Problem 6

Problem 7

Problem 8

Problem 9

Problem 10

Problem 11

Problem 12

Problem 13

Problem 14

Problem 15

Problem 16

Problem 17

Problem 18

Problem 19

Problem 20

Problem 21

Problem 22

Problem 23

Problem 24

Problem 25

Problem 26

Problem 27

Problem 28

Problem 29

Problem 30

Problem 31

Problem 32

Problem 33

Problem 34

Problem 35

Problem 36

Problem 37

Problem 38

Problem 39