

1 Probability

Remark 1.1. To build a mathematical model of uncertainty and randomness we start by thinking what it is we will be measuring. We want to be able to measure the likelihood of some event happening. So we will measure events. What's an event though? It is an occurrence of something and this something is comprised of smaller events that the event we are concerned with is comprised of. These smaller events we will call outcomes, or singleton event. So for example, lets say I want to know the probability that my plane crashes. The event of a plane crash is composed of infinitely many outcomes... the event maybe consists of an outcome where a mechanic overlooked something, then something wobbled in just the right way, then the plane took a certain turn which caused some screw to loosen, ..., then the plane crashed. So the screw loosening is one outcome, the plane turning is another, etc. all these outcomes make up the event where a plane crashes. So we will consider sets of outcomes, we will call these events.

Remark 1.2. The below theorems will show us how to build these events, how to count the number of outcomes in the events.

Theorem 1.1. Let X_1, X_2, \dots, X_n be finite sets with cardinalities $|X_1|, |X_2|, \dots, |X_n|$. If a process consists of making sequential choices such that:

- The first choice is made from X_1 ,
- The second choice is made from X_2 ,
- ...,
- The n th choice is made from X_n ,

where the number of choices at each stage is independent of previous choices, then the total number of ways to complete the process is:

$$|X_1| \cdot |X_2| \cdots |X_n| = \prod_{i=1}^n |X_i|.$$

Theorem 1.2. Let n and k be nonnegative integers with $0 \leq k \leq n$. The number of distinct subsets of size k that a set of size n has is given by the binomial coefficient

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

Theorem 1.3. For any integer $n \geq 0$ and any real or complex numbers a, b ,

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

Theorem 1.4. The number of ways a set of n distinct objects can be partitioned into k subsets with n_k objects in the k th subset is

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

Theorem 1.5. The number of ways to arrange n distinct objects in a sequence is

$$P(n) = n! = n(n-1)(n-2) \cdots 2 \cdot 1$$

The number of ways to select and arrange k objects from n distinct objects is

$$P(n, k) = \frac{n!}{(n-k)!}.$$

Theorem 1.6. Let n, k , and j be nonnegative integers with $0 \leq j \leq k \leq n$. Then for a set with n distinct elements, all of the following hold

1.

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

2.

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

3.

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

4.

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

5.

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

6.

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1}$$

1.1 Axioms of Probability

Remark 1.3. So how will we define the abstract space we will be working in so that we can effectively measure the likelihood of events? So consider some event A that you want to know the likelihood of. If the event A is possible, then it should also be possible that A^c , meaning, we should be able to measure both of these. So let's call the space \mathcal{A} , we will add sets to \mathcal{A} that we think should be possible to measure if it is to be possible to measure A . So we have $A \in \mathcal{A}$ and $A^c \in \mathcal{A}$. If $A \in \mathcal{A}$ and $B \in \mathcal{A}$ then we should be able to measure $A \cup B$ and $A \cap B$, so we include all intersections and unions of possible events in \mathcal{A} .

Definition 1.1 (Algebra and σ -algebra). Let Ω be an abstract space. Let 2^Ω denote all subsets of Ω . With \mathcal{A} being a subset of 2^Ω . Then \mathcal{A} is an algebra if it satisfies (1), (2), and (3). \mathcal{A} is a σ -algebra if it satisfies (1), (2), and (4).

1. $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$ 2. If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.3. If the finite sequence of events $A_1, A_2, \dots, A_n \in \mathcal{A}$ then $\bigcup_{i=1}^n A_i \in \mathcal{A}$ and $\bigcap_{i=1}^n A_i \in \mathcal{A}$.4. If the countable sequence of events $A_1, A_2, \dots \in \mathcal{A}$ then $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$ and $\bigcap_{i=1}^\infty A_i \in \mathcal{A}$.

Remark 1.4. If $C \subset 2^\Omega$, then the σ -algebra generated by C , denoted $\sigma(C)$, is the smallest σ -algebra containing C .

We choose b_n to be strictly increasing so that the $[a_n, b_n]$ part of the interval $(a_n, b_n]$ converges to $[a, b]$. This is what allows us to have that $(a, b) = \bigcup_{n=1}^\infty (a_n, b_n]$.

Theorem 1.7 (Borel σ -algebra). If $\Omega = \mathbb{R}$, the Borel σ -algebra is the σ -algebra generated by open sets (or equivalently closed sets). Then the Borel σ -algebra can be generated by intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$.

Proof. Let C denote all open intervals. Since every open set in \mathbb{R} is the countable union of open intervals, we have $\sigma(C)$ = the Borel σ -algebra of \mathbb{R} . Let D denote all intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$. Let $(a, b) \in C$, and let $(a_n)_{n \geq 1}$ be a sequence of rationals decreasing to a and $(b_n)_{n \geq 1}$ be a sequence of rationals strictly increasing to b . Then

$$(a, b) = \bigcup_{n=1}^\infty (a_n, b_n] = \bigcup_{n=1}^\infty ((-\infty, b_n] \cap (a_n, \infty)) = \bigcup_{n=1}^\infty ((-\infty, b_n] \cap (-\infty, a_n]^c)$$

Since the right most expression is of the form of D and since we have that any element of C is equivalent to an element of D , we have $C \subset \sigma(D)$, hence $\sigma(C) \subset \sigma(D)$. However since $(-\infty, a]$ contains all its limit points, we know each element of D is a closed set, since closed sets are Borel sets, we have that $\sigma(D)$ is contained in the Borel sets \mathcal{B} . Thus we have

$$\mathcal{B} = \sigma(C) \subset \sigma(D) \subset \mathcal{B},$$

and hence $\sigma(D) = \mathcal{B}$. □

Remark 1.5. So the theorem above shows that when our sample space is the real numbers, or any space with the proper topology, we can generate the σ -algebra we define the probability measure on by using intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$. Since $a \in \mathbb{Q}$, we have made the σ -algebra from countable sets. Which is what we needed since before we knew that open sets \mathcal{C} could cover any set, we had to show that the countable collection of $(-\infty, a]$ could also cover (and thus measure) any set.

Remark 1.6. For our actual probability measure, we need an event that is guaranteed and an event that is impossible, so we include $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$ since we want $P(\emptyset) = 0$ and $P(\Omega) = 1$. We also would want that if we measure two events A and B where A and B share no outcomes $A \cap B = \emptyset$, then we want that $P(A \cup B) = P(A) + P(B)$.

Definition 1.2 (Probability Measure). A probability measure defined on a σ -algebra \mathcal{A} of Ω is a function $P : \mathcal{A} \rightarrow [0, 1]$ that satisfies

1. $P(\Omega) = 1$
2. For every pairwise disjoint $(A_n \cap A_m = \emptyset \text{ whenever } n \neq m)$ countable sequence $(A_n)_{n \geq 1}$ of elements of \mathcal{A} , we have

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Theorem 1.8. Let A_1, A_2, \dots, A_n be events, then

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 \leq i_2 \leq n} P(A_{i_1} \cap A_{i_2}) \\ &+ \sum_{1 \leq i_1 \leq i_2 \leq i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \sum_{1 \leq i_1 < i_2 \leq i_3 < i_4 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}) \\ &+ \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \end{aligned}$$

Proof. We prove the formula by induction on n . For $n = 2$ we already know it holds. Thus assume the statement holds for $n - 1$ events, i.e.,

$$P(A_1 \cup A_2 \cup \dots \cup A_{n-1}) = \sum P(A_i) - \sum P(A_i \cap A_j) + \sum P(A_i \cap A_j \cap A_k) - \dots + (-1)^n P(A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

We show it holds for n events. Observe that

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P((A_1 \cup \dots \cup A_{n-1}) \cup A_n).$$

Using the formula

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

with $B = A_1 \cup \dots \cup A_{n-1}$ and $C = A_n$, we obtain

$$P(A_1 \cup \dots \cup A_n) = P(A_1 \cup \dots \cup A_{n-1}) + P(A_n) - P((A_1 \cup \dots \cup A_{n-1}) \cap A_n).$$

By the induction hypothesis, the first term on the right-hand side is the sum of all inclusion-exclusion terms up to $(-1)^n P(A_1 \cap \dots \cap A_{n-1})$. The new term is

$$-P((A_1 \cup \dots \cup A_{n-1}) \cap A_n).$$

Expanding

$$(A_1 \cup \dots \cup A_{n-1}) \cap A_n = (A_1 \cap A_n) \cup \dots \cup (A_{n-1} \cap A_n),$$

and applying the inclusion-exclusion principle to these $n-1$ sets $A_1 \cap A_n, \dots, A_{n-1} \cap A_n$, we obtain an alternating sum of probabilities of intersections that include A_n . Careful bookkeeping of signs shows that each new term $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_m \cap A_n)$ appears with the correct factor $(-1)^m$.

Combining the induction hypothesis with these new terms yields

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum P(A_i) - \sum P(A_i \cap A_j) + \sum P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

Thus, the formula holds for all n . \square

Definition 1.3. (Indicator Function) If $A \in 2^\Omega$, then the indicator function $1_A(\omega)$ be given by

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

We say $A_n \in \mathcal{A}$ converges to A if $\lim_{n \rightarrow \infty} 1_{A_n}(\omega) = 1_A(\omega) \forall \omega \in \Omega$.

Remark 1.7. A few comments and clarifications about the definition above. Since 2^Ω is not necessarily all measurable, meaning, the σ -algebra \mathcal{A} may not include every subset of Ω . So for any singleton outcome $\omega \in \Omega$, we have if $\forall \omega \in A, \omega \in A_n$ as $n \rightarrow \infty$ then A_n converges to A . Note this is precisely what $\lim_{n \rightarrow \infty} 1_{A_n}(\omega) = 1_A(\omega) \forall \omega \in \Omega$ is stating.

Remark 1.8. So if we can define convergence of a sequence of sets, then we must have some conception of convergence of supremum and infimum. How will we define these? We want the supremum of a set to be the elements (outcomes) that are in **at least one** of the *infinite events*, that is, for all events events past some n th event, I want to know what is in **any** of the events that come after this one, then letting $n \rightarrow \infty$ we see that the elements remaining are in **at least one** of the *infinite events*. Whereas, we want the infimum to be the *smaller* set, when compared to the supremum. So instead of considering all elements that are in **any** event past the n th (we will again let $n \rightarrow \infty$) event, we will consider the elements that are in **every single** event past this n th one. From this, it is easy to see that the infimum is a subset of the supremum, which is what we wanted. We also want that when these are equivalent, the sequence of events converges.

Definition 1.4 (Supremum and Infimum of Sequence of Sets). Let A_n be a sequence of sets. If $A_n \in \mathcal{A} \forall n \in \mathbb{N}$ then define

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} A_m \\ \liminf_{n \rightarrow \infty} A_n &= \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} A_m. \end{aligned}$$

Lemma 1.1. Let \mathcal{A} be a σ -algebra and $(A_n)_{n \geq 1}^\infty$ be a sequence of sets in \mathcal{A} . Then,

$$\liminf_{n \rightarrow \infty} A_n \in \mathcal{A}, \quad \limsup_{n \rightarrow \infty} A_n \in \mathcal{A}, \quad \text{and} \quad \liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$$

Proof. By definition (1.1), the σ -algebra \mathcal{A} is closed under countable unions and intersections. since $A_n \in \mathcal{A}$, we have that for any fixed n , $\bigcap_{k \geq n} A_k \in \mathcal{A}$. Then countably infinite many unions of this is also in \mathcal{A} . That is,

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k \in \mathcal{A}.$$

Similarly,

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k \in \mathcal{A}.$$

Now suppose $x \in \liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k$, then for some $N \in \mathbb{N}$, $x \in A_k$, $\forall k \geq N$. Thus $x \in \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$ since x is in all such A_k where $k \geq N$ and $\limsup A_n$ only requires that x be in at least one. Therefore,

$$\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n.$$

□

Lemma 1.2. Let \mathcal{A} be a σ -algebra and $(A_n)_{n \geq 1}^{\infty}$ be a sequence of sets in \mathcal{A} . Then,

$$\lim_{n \rightarrow \infty} A_n = A \iff \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n = A$$

Proof. Suppose $x \in A$ and $\lim A_n = A$. Then $\exists N \in \mathbb{N}$ such that $\forall k \geq N$ we have $x \in A_k$. Thus,

$$x \in \bigcap_{k \geq N} A_k \implies x \in \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k = \liminf_{n \rightarrow \infty} A_n$$

Hence, $A \subseteq \liminf A_n$ since we showed that $x \in A \implies x \in \liminf A_n$. Suppose $x \notin A$. Then $\exists N \in \mathbb{N}$ such that $\forall k \geq N$ we have $x \notin A_k$. Thus,

$$x \notin \bigcup_{k \geq N} A_k \implies x \notin \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k = \limsup_{n \rightarrow \infty} A_n$$

Lets summarize what we have showed here. We have all of the below conditions,

$$(1.1) \quad \liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n, \quad A \subseteq \liminf_{n \rightarrow \infty} A_n, \quad \text{and} \quad \limsup_{n \rightarrow \infty} A_n \subseteq A$$

Thus

$$A = \liminf_{n \rightarrow \infty} A_n = \limsup_{n \rightarrow \infty} A_n$$

□

Theorem 1.9 (Continuity of Probability Measure). Let P be a probability measure, and let A_n be a sequence of events in the σ -algebra \mathcal{A} which converges to A . Then $A \in \mathcal{A}$ and $\lim_{n \rightarrow \infty} P(A_n) = P(A)$.

Proof. Define $\limsup A_n$ and $\liminf A_n$ as definition (1.4). By lemma 1.1, we have $\limsup_{n \rightarrow \infty} A_n \in \mathcal{A}$ and $\liminf_{n \rightarrow \infty} A_n \in \mathcal{A}$. So by hypothesis, A_n converges to A , then from lemma 1.2,

$$\lim_{n \rightarrow \infty} 1_{A_n} = 1_A, \quad \forall \omega \iff A = \limsup_{n \rightarrow \infty} A_n = \liminf_{n \rightarrow \infty} A_n$$

□

Definition 1.5 (Monotone Sequence of Sets). A sequence of events $(A_n)_{n \geq 1}^{\infty}$ is said to be an *monotone increasing* sequence of sets if

$$A_1 \subseteq A_2 \subseteq \dots \subseteq A_k \subseteq A_{k+1} \subseteq \dots$$

Similarly, a sequence of sets $(A_n)_{n \geq 1}^{\infty}$ is said to be a *monotone decreasing* sequence if

$$A_1 \supseteq A_2 \supseteq \dots \supseteq A_k \supseteq A_{k+1} \supseteq \dots$$

Further, if an increasing sequence $(A_n)_{n \geq 1}^{\infty}$ converges to some event A , then we write $A_n \uparrow A$ and we have $A = \bigcup_{n \geq 1}^{\infty} A_n$. Similarly, if $(A_n)_{n \geq 1}^{\infty}$ decreases to A then we write $A_n \downarrow A$, with $A = \bigcap_{n \geq 1}^{\infty} A_n$.

Theorem 1.10. Let \mathcal{A} be a σ -algebra and let $(A_n)_{n \geq 1}^{\infty} \in \mathcal{A}$ be a sequence of sets. Suppose $P : \mathcal{A} \rightarrow [0, 1]$ is a probability measure. Then the following are equivalent,

1. Axiom (2) of definition (1.2)
2. $A_n \downarrow A \implies P(A_n) \downarrow P(A)$.
3. $A_n \uparrow A \implies P(A_n) \uparrow P(A)$

Remark 1.9. In the (3) \implies (1) proof, note that we assume A_n is pairwise disjoint because that is what needs to be satisfied, by the definition of the probability measure.

Proof. (2) \iff (3): Suppose $A_n \uparrow A$ and $P(A_n) \uparrow P(A)$. Then $A_n^c \downarrow A^c$ and $P(A_n^c) \downarrow P(A^c)$. But since $P(A_n^c) = 1 - P(A_n)$, proving (3) \iff (2) suffices.

(3) \implies (1): Suppose $A_n \uparrow A$ and $P(A_n) \uparrow P(A)$. Also, assume A_n is *pairwise disjoint*, meaning $\forall i, j \in [n]$, where $i \neq j$, we have $A_i \cap A_j = \emptyset$. Let $B_n = \cup_{p=1}^n A_p$ and let $B = \cup_{n \geq 1} A_n$. Then by axiom (2) of the probability measure (1.2), we have $P(B_n) = \sum_{p=1}^n P(A_p)$. Then as $n \rightarrow \infty$, we have that $P(B_n) \uparrow P(B)$, so $P(B_n)$ is increasing sequence, increasing to $P(B)$ since $P(A_n) \uparrow P(A)$ so $P(B_n) = P(\cup_{n \geq 1} A_n)$.

(1) \iff (3): Suppose A_n is a sequence increasing to A . Define the sequence $(B_n)_{n \geq 1}$

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_2 \setminus A_1 \\ &\vdots \\ B_k &= A_k \setminus A_{k-1} \\ &\vdots \end{aligned}$$

Remark 1.10. Since A_n is increasing, we have that the A_{k-1} set contains every set before it, so letting $B_k = A_k \setminus A_{k-1}$ for every k ensures each B_k contains only the elements that A_k provided. Then since probabilities are nonnegative, we see that B_n is monotone increasing.

Then we have $A = \cup_{i=1}^{\infty} B_i$ and $B_n \cap B_m = \emptyset$ whenever $m \neq n$, meaning B_n is pairwise disjoint. Thus from (1),

$$P(A) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i)$$

But since we also have

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

thus we have $P(A_n) \uparrow P(A)$. □

Proposition 1.1. *Let $A_i \in \mathcal{A}$ be a sequence of events. Then,*

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^{\infty} P(A_i).$$

Proof. We proceed by induction. With $n = 2$ we have

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \implies P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$$

Assume for some n the below holds,

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

Then consider

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^n A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^n A_i \cap A_{n+1}\right) \quad (1.8) \\ &\leq \sum_{i=1}^n P(A_i) + P(A_{n+1}) = \sum_{i=1}^{n+1} P(A_i) \\ n \rightarrow \infty &\implies P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) \\ &\implies P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^{\infty} P(A_i) \end{aligned}$$

□

1.2 Conditional Probability and Independence

Remark 1.11. Suppose we wanted to determine the probability of some event but we want to update this probability given we observed that the event B occurred, or, we want to see how the likelihood changes given B happened. Then we want the probability *density* associated with the portion of A that is in B , then we want to normalize this figure to represent that B is being assumed. Another way of seeing the above is, given that B occurred, we would never consider accounting for the event $A \cap B^c$.

Definition 1.6. Let B be an event in the sample space Ω such that $P(B) > 0$. Then for all events A the *conditional probability* of A given B is defined as

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

Proposition 1.2 (Conditional Probability Measure). *The conditional probability is a probability measure (1.2).*

Proof. Define $Q(A) = P(A | B)$. Then

$$Q(\Omega) = P(\Omega | B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Now suppose A_n is a sequence of pairwise disjoint sets. Then

$$Q(\cup_{n=1}^{\infty} A_n) = \frac{P(\cup_{n=1}^{\infty} A_n \cap B)}{P(B)} = \frac{P(\cup_{n=1}^{\infty} (A_n \cap B))}{P(B)}$$

Observe that $\cup_{n=1}^{\infty} (A_n \cap B)$ is a pairwise disjoint partition (1.8) of B , thus applying (1.2)

$$\sum_{n=1}^{\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=1}^{\infty} P(A_n | B) = \sum_{n=1}^{\infty} Q(A_n).$$

□

Definition 1.7. A collection of events $(A_i)_{i \in I}$ is an independent collection if for every finite subset J of I , one has

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

If the above condition is satisfied for the whole collection, we say the collection $(A_i)_{i \in I}$ is mutually independent. Also, if A_i and A_j are independent $\forall i, j$ with $i \neq j$, that is if any two events you pick from the collection $(A_i)_{i \in I}$ are independent, then the collection is pairwise independent.

Exercise 1.1. If A and B are independent, so also are A and B^c , A^c and B , and A^c and B^c .

Proof.

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

Suppose A and B are independent events, so that

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

We show that $P(A^c \cap B^c) = P(A^c)P(B^c)$. Notice that

$$P(A^c \cap B^c) = 1 - P(A \cup B).$$

By the inclusion-exclusion formula,

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

Hence,

$$P(A^c \cap B^c) = 1 - [P(A) + P(B) - P(A \cap B)] = 1 - P(A) - P(B) + P(A \cap B).$$

On the other hand,

$$P(A^c)P(B^c) = (1 - P(A))(1 - P(B)) = 1 - P(A) - P(B) + P(A)P(B).$$

Since $P(A \cap B) = P(A)P(B)$, we see that the two expressions match:

$$P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A \cap B) = 1 - P(A) - P(B) + P(A)P(B) = P(A^c)P(B^c).$$

Therefore, A^c and B^c are independent. \square

Proposition 1.3 (Partition Equation). *If $A_1, A_2, \dots, A_n \in \mathcal{A}$ and if $P(A_1 \cap \dots \cap A_{n-1}) > 0$, then*

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

Proof. We use induction. For $n = 2$, the theorem is simply Definition 1.6. Suppose the theorem holds for $n - 1$ events. Let $B = A_1 \cap \dots \cap A_{n-1}$. Then by Definition 1.6,

$$P(B \cap A_n) = P(A_n | B)P(B);$$

next, we replace $P(B)$ by its value given in the inductive hypothesis:

$$P(B) = P(A_1)P(A_2 | A_1) \dots P(A_{n-1} | A_1 \cap \dots \cap A_{n-2}),$$

and we get the result \square

Definition 1.8 (Partition). A countable collection of events B_1, \dots, B_n are a *partition* of Ω if the sets B_i are pairwise disjoint and together they make up Ω . That is, for all i and j , $B_i \cap B_j = \emptyset$ whenever $i \neq j$ and $\bigcup_{i=1}^n B_i = \Omega$

Proposition 1.4. *Suppose that B_1, \dots, B_n is a partition of Ω with $P(B_i) > 0$ for $i = 1, \dots, n$. Then for any event A we have*

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

Theorem 1.11 (Bayes Theorem). *Let B_1, B_2, \dots, B_n be a partition of the sample space Ω such that each $P(B_i) > 0$. Then for any event A with $P(A) > 0$, and for any $k = 1, \dots, n$, we have:*

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

Proof. By Proposition 1.3, we have that the denominator

$$\sum_n P(A | B_n)P(B_n) = P(A).$$

Therefore, the formula becomes

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

\square

Definition 1.9. Let A_1, A_2, \dots, A_n and B be events with $P(B) > 0$. Then A_1, A_2, \dots, A_n are *conditionally independent, given B* , if the following condition holds:

For any $k \in \{2, \dots, n\}$ and indices $1 \leq i_1 < i_2 < \dots < i_k \leq n$,

$$P(A_{i_1}A_{i_2} \dots A_{i_k} | B) = P(A_{i_1} | B)P(A_{i_2} | B) \dots P(A_{i_k} | B).$$

1.3 Random Variables

A random variable X is a single-valued real function that assigns a value of X to each set in \mathcal{A} . So X is just a regular function the only difference is that the domain is comprised of sets. For each such set, X will give some value, then we can assign probabilities to these values. We will see that we are often not concerned with the individual values of X , instead we investigate the range of X and the *probability distribution* associated with it.

Definition 1.10 (Random Variable). A random variable is a measurable function $X : \Omega \rightarrow \mathbb{R}$ such that for all Borel measurable sets $B \subseteq \mathbb{R}$, the preimage of B is an event in \mathcal{A} , that is

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}.$$

This means that X is \mathcal{A} -measurable, ensuring that we can compute probabilities of the form $P(X \in B)$

Remark 1.12. So a random variable inputs events or outcomes and outputs a real number, then the probability measure will assign probabilities in $[0, 1]$ to the values of X . We can then define the distribution of X by

$$P^X(A) = P(\omega \mid X(\omega) \in A) = P(X^{-1}(A)) = P(X \in A)$$

When Ω is finite or countable, this is completely determined by the following

$$p_j^X = P(X = j) = \sum_{\{\omega \mid X(\omega) = j\}} p_\omega \text{ and } P_X(A) = \sum_{j \in A} p_j^X$$

So basically, the probability distribution of the random variable X is the collection of probabilities $P(X \in B)$.

Further, a random variable is a discrete random variable if there exists a finite or countably infinite set k_1, k_2, \dots of reals such that

$$\sum_i P(X = k_i) = 1$$

where the sum ranges over the entire set of points $\{k_1, k_2, \dots\}$.

Remark 1.13. We are now going to go off on a bit of a tangent in order to construct the probability measure over uncountable Ω , which we have not touched on yet. In the above remark, we concluded that countable probabilities are completely determined by

$$P(A) = \sum_{\{\omega \mid \forall \omega \in A\}} p_\omega$$

When we consider uncountable sets, this no longer holds. For example, if $\Omega = [0, 1] \subset \mathbb{R}$, then we need to assign a probability measure to the collection of all subsets of $[0, 1]$.

Definition 1.11. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X : \Omega \rightarrow \mathbb{R}$ be a real-valued random variable. The cumulative distribution function (CDF) of X , denoted $F_X : \mathbb{R} \rightarrow [0, 1]$, is defined by

$$F_X(x) = P(X \leq x), \quad \forall x \in \mathbb{R}.$$

The function $F_X(x)$ satisfies the following properties

1. (Monotonicity) $F_X(x)$ is monotone increasing
2. (Right Continuity) $F_X(x)$ is right continuous

$$\lim_{h \rightarrow 0^+} F_X(x + h) = F_X(x)$$

3. (Limits at Infinity)

$$\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

4. (Jumps) If $F_X(x)$ has a jump at x , say

$$P(X = x) = F_X(x) - \lim_{y \rightarrow x^-} F_X(y),$$

then X has positive probability at x

5. (Absolute Continuity) If $F_X(x)$ is absolutely continuous, then there exists a probability density function (PDF) $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

Definition 1.12 (Probability Mass Function (PMF)). The probability mass function of a discrete random variable X is the function p defined by

$$p(k) = P(X = k)$$

for possible values k of X .

We can then find the probabilities of other events of X . That is, for some event $X \in B$, we sum each $p(k)$, $\forall k \in B$.

Definition 1.13. [Expected Value] Let X be a real-valued random variable on a countable space Ω . The expectation of X , denoted $E(X)$, is defined to be

$$E(X) = \sum_{\omega} X(\omega)p_{\omega} \quad \text{or} \quad \int_{-\infty}^{\infty} X(\omega)p_{\omega}d\omega??$$

provided this sum converges. Notice that if the random variable is discrete we use the finite sum, if it is continuous, we use the continuous sum.

Definition 1.14. The n th moment of the random variable X is the expectation $E(X^n)$.

$$E(X^n) = \sum_{\omega} X^n(\omega)p_{\omega} \quad \text{or} \quad \int_{-\infty}^{\infty} X^n(\omega)P(X(\omega))d\omega$$

Theorem 1.12. Let $h : \mathbb{R} \rightarrow [0, \infty)$ be a nonnegative function and let X be a real valued random variable. Then

$$P(\{\omega \mid h(X(\omega)) \geq a\}) \leq \frac{E(h(X))}{a}, \quad \forall a > 0.$$

Corollary 1.1 (Markov's Inequality).

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

Definition 1.15. Let X be a real valued random variable with $X^2 \in \mathcal{L}^1$ where \mathcal{L}^1 is the space of real valued random variables on (Ω, \mathcal{A}, P) . The variance of X is defined to be

$$\sigma^2 = \sigma_X^2 = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

The standard deviation of X , σ_X , is the nonnegative square root of the variance.

Corollary 1.2 (Chebyshev's Inequality). If X^2 is in \mathcal{L}^1 , then for $a > 0$ we have

1. $P(\{|X| \geq a\}) \leq \frac{EX^2}{a^2}$
2. $P(\{|X - E(X)| \geq a\}) \leq \frac{\sigma_X^2}{a^2}$

1.3.1 Distributions

We will now look at the different distributions associated with a random variable and we will discuss the motivations for using each one.

Repeated Independent Trials

If we are sampling with replacement, that is, whatever *data* we are looking at is being drawn from equivalent sets. This means the information of each given trial (or occurrence/value/outcome of the random variable) does not provide any information for the next trial.

The simplest trial is one which has two outcomes, success or failure, for each trial. If we let the probability of a success be p , then the probability of fail is $1 - p$. So suppose we do n trials and we have $k = 1$ success. Then the probability of that one success is $p(1 - p)^{n-1}$, but the number of ways to have this success is $\binom{n}{1} = n$ since for each of the n spots, only one of them was a success. Extending this, we get the definition below.

Definition 1.16 (Binomial Distribution). Let n be a positive integer and $0 \leq p \leq 1$. A random variable X has the *binomial distribution* with parameters n and p if the possible values of X are $\{0, 1, \dots, n\}$ and the probabilities are

$$P(\{X = k\}) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k = 0, 1, \dots, n.$$

This is denoted $X \sim \text{Bin}(n, p)$.

Now consider you are trying to figure out how many flips of a coin till you get a head. Let the probability of a head be p . Then the probability of obtaining heads on the 10th flip is $(1 - p)^9 p$. This motivates the below definition

Definition 1.17 (Geometric Distribution). A random variable X follows a Geometric distribution with parameter p (success probability per trial) if the probability of k independent trials till a success on the k th trial is given by,

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \dots$$

Now imagine you have a finite population with N objects. Then suppose K objects are of type 1 and $N - K$ objects are of type 2. We draw a sample of n without replacement from the N total and want to know the probability of getting exactly k objects of type 1. This is the exact same as the binomial distribution only now we are not replacing.

So for some $P(X = k)$ we have $\binom{K}{k}$ ways of choosing the k type 1 objects from the total K amount of them. Then we have $\binom{N-K}{n-k}$ ways to select the remaining $n - k$ objects of type 2 from the total $N - K$ of type 2. Then $\binom{N}{n}$ is the total number of ways to select the n objects.

Definition 1.18 (Hypergeometric Distribution). A hypergeometric random variable represents the number of successes of size n , drawn without replacement from a population of size N that contains K successes. The PMF is given by

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}, \quad \max(0, n - (N - K)) \leq k \leq \min(n, K).$$

Definition 1.19 (Poisson Distribution). A Poisson random variable models the number of events occurring in a fixed interval of time or space, under the assumption that events occur independently and at a constant average rate λ . A random variable X follows a Poisson distribution with rate parameter $\lambda > 0$ if

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

Definition 1.20 (Normal Distribution). A random variable X follows a Normal distribution with mean μ and variance σ^2 , written as $X \sim \mathcal{N}(\mu, \sigma^2)$, if its probability density function (PDF) is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$