1 Probability Spaces

Before diving in into the definition of a probability space, the main object of this course, we must note that this course is an introductory course in probability theory, which means we don't have the tools from measure theory to formalize probability. Thus, some proofs will be omitted, and we will also need to formalize discrete and continuous probability theory seperately.

First, let us introduce a paradox.

Paradox 1.1. (Bertrand's Paradox). *Consider an equilateral triangle inscribed in a circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?*

We can ponder about this paradox for a while, but Bertrand himself came up with three solutions, each with a different answer. The main difference in his methods lies in the way in which we choose the chords.

Definition 1.1. The sample space of an experiment, is a set Ω which contains all the possible outcomes of the experiment.

A good thing to note, is that we can choose different sample spaces for the same experiment. For example, if the experiment consists of rolling two dice, and we want to check for the sum of the results, we can set either $\Omega = \{1,2,3,4,5,6\}^2$, for the result of each dice, or $\Omega = \{1,2,\ldots,11,12\}$ for the sum of the results of the dice.

Definition 1.2 (Probability theory, intuitive definition)**.** A discrete probability space is a pair $(\Omega, \mathbf{P})$, where Ω is a countable sample set, and $\mathbf{P}\colon \Omega \to [0,1]$ is a function such that $\sum_{\omega\in\Omega} \mathbf{P}(\omega) = 1$. Intuitively, we say that $\mathbf{P}(\omega)$ represents the probability that ω will happen.

Definition 1.3. A subset of the sample space $A \subseteq \Omega$ is called an event. We also define:

$$\mathbf{P}(A) := \sum_{\omega\in\Omega} \mathbf{P}(\omega)$$

Here are a few properties of probability functions we can immediately verify:

1. $\mathbf{P}(\Omega) = 1$
2. $\mathbf{P}(\emptyset) = 0$
3. For $A \subset \Omega$ we have $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$
4. If $\{A_n\}_{n=1}^N$ are disjoint sets then
$$\mathbf{P}\left(\cup_{n=1}^n A_n\right) = \sum_{n=1}^N \mathbf{P}(A_n).$$
5. If $\{A_n\}_{n=1}^\infty$ is a sequence of pairwise disjoint sets then
$$\mathbf{P}\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mathbf{P}(A_n).$$

In a finite probability space we say that the probability function is continuous if for every $\omega \in \Omega$ we have $\mathbf{P}(\omega) = \frac{1}{|\Omega|}$.

We now proceed to consider an experiment in which we choose a direction in $\mathbb{R}^2$ at random, on S^1 and write it. The sample space is:

$$\Omega = S^1 = \left\{ e^{i\theta} \mid \theta \in [0, 2\pi) \right\}.$$

A natural question to ask, is if we can define a uniform probability function in the sense that for any arc $[a, b] \subset S^1$ we have $\mathbf{P}([a, b]) = b - a$. The answer is that with the definition we have worked with so far, we can't. We see that $\mathbf{P}(\{a\}) = 0$ for any $a \in S^1$, and thus we have that

$$\mathbf{P}(\Omega) = \sum_{\omega \in \Omega} \mathbf{P}(\omega) = 0.$$

To solve this problem, we may try to define a new function $\mathbf{P}\colon 2^\Omega \to [0, 1]$ that will directly assign each event its probability, but unfortunately for us, such a function, that satisfies the desired properties of a probability function, does not exist. The proof for this is in the course "real valued function", and will not be discussed here. However, we can give a proof, under the assumption of the following lemma.

Lemma 1.1. *Exists a set $E \subset S^1$ such that for any rational number $q \in (0, 2\pi) \cap \mathbb{Q}$ we have $e^{iq}E \cap E = \emptyset$.*

Indeed we see that

$$1 = \mathbf{P}(\Omega) = \mathbf{P}\left(\bigcup_{q \in [0,2\pi] \cap \mathbb{Q}} e^{iq}E\right) = \sum_{q \in [0,2\pi) \cap \mathbb{Q}} \mathbf{P}(e^{iq}E) = \sum_{q \in [0,2\pi) \cap \mathbb{Q}} \mathbf{P}(E)$$

And now we have a contradiction because if we set $\mathbf{P}(E) = a$ then we get

$$1 = \sum_{q \in [0,2\pi) \cap \mathbb{Q}} a$$

and this equation has no solution.

The classical solution to this problem, is to only define the probability function only on certain subsets of the sample space. Suppose we denote this new domain as $\mathcal{F} \in 2^\Omega$. In order for the desired properties to hold we must also accept that $\mathcal{F}$ holds certain conditions.

Definition 1.4 (σ-algebra)**.** Let Ω be a set. We say that $\subset 2^\Omega$ is a σ-algebra of sets (sometimes also called a σ-field), it satisfies the following properties:

1. $\Omega \in \mathcal{F}$.
2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
3. If $(A_n)_{n=1}^\infty \subset \mathcal{F}$, then $\cup_{n=1}^\infty A_n \in \mathcal{F}$.

We can now formally define a probability space.

Definition 1.5 (Probability Space)**.** A probability space is a triplet $(\Omega, \mathcal{F}, \mathbf{P})$ such that Ω is a set, $\mathcal{F}$ is a σ-algebra of Ω, and $\mathbf{P}\colon \to [0, 1]$ is a probability function that satisfies:

1. $\mathbf{P}(\Omega) = 1$
2. If $(A_n)_{n=1}^\infty \subset \mathcal{F}$ are disjoint, then $\mathbf{P}\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mathbf{P}(A_n)$.

In this case we shall call elements of $\mathcal{F}$ events.

Proposition 1.2. *Exists a σ-algebra $\mathfrak{B}$ of $\Omega = S^1$, and a unique function $\mathbf{P}\colon \to [0, 1]$ such that $(\Omega, \mathfrak{B}, \mathbf{P})$ is a probability space and $\mathbf{P}$ is invariant to spinning on the sphere.*

Definition 1.6 (Algebra of Sets)**.** A set $\mathcal{C} \subset 2^\Omega$ is called an algebra of sets if it satisfies the following properties:

1. $\Omega \in \mathcal{C}$.

2. If $A \in \mathcal{C}$, then $A^c \in \mathcal{C}$.

3. if $A, B \in \mathcal{C}$, then $A \cup B \in \mathcal{C}$.

We can immediately verify that any algebra $\mathcal{C}$ is closed under finite unions and finite intersections. We also notice that $\emptyset \in \mathcal{C}$, and that if $A, B \in \mathcal{C}$, then $A \setminus B \in \mathcal{C}$. We can also notice that any σ-algebra is closed under countable intersections, and that every σ-algebra is in particular also an algebra.

Example 1.1. If Ω is a set, and $A \subset \Omega$, then both 2^Ω and $\{\emptyset, A, A^c, \Omega\}$ are σ-algebras.

Example 1.2. Given a set Ω, the smallest σ-algebra of Ω is $\{\emptyset, \Omega\}$ which is called the trivial σ-algebra.

Proposition 1.3. *Let $(\mathcal{F}_\alpha)_{\alpha \in I}$ be a family of σ-algebras, then $\cap_{\alpha \in I} \mathcal{F}_\alpha$ is a σ-algebra.*

Proof. Obvious. □

Definition 1.7 (Minimal Sigma Algebra)**.** Let Ω be a set, and let $H \subset 2^\Omega$ be a family of its subsets. Then we define the minimal sigma algebra that contains H, denoted $\sigma(H)$, as the intersection of all the σ-algebras that contains all the elements in H. Notice that the intersection is never empty because 2^Ω is a σ-algebra that will always contain the elements of H.

Example 1.3. (Borel's σ-algebra). One of the most important minimal σ-algebras, is Borel's σ-algebra defined on $\mathbb{R}$. It is defined as such:

$$\mathfrak{B} = \mathfrak{B}(\mathbb{R}) := \sigma(\{(a, b) \mid a < b\}).$$

That is, the smallest σ-algebra that contains all the open intervals in $\mathbb{R}$. Similarly, we can define it on the space $\mathbb{R}^d$ as follows:

$$\mathfrak{B}_d = \mathfrak{B}(\mathbb{R}^d) := \sigma\left(\left\{\prod_{i=1}^{d}(a_i, b_i) \mid a_i < b_i\right\}\right).$$

Note that in general, Borel's σ-algebra is defined to be the smallest σ-algebra that contains all the open sets in a general topological space. It can be showen that this definition is equivalent to the definitions we just gave for $\mathfrak{B}$ and $\mathfrak{B}_d$.

Theorem 1.4. (Carathéodory). *Let Ω be a set, let $\mathcal{G}$ be an algebra of sets of Ω. If $\widehat{P}c\colon \mathcal{G} \to [0, 1]$ is a function that satisfies $f(\Omega) = 1$, and for each sequence of pairwise disjoint sets $\{A_n\}_{n=1}^\infty$ that*

$$\widehat{\mathbf{P}}\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^{\infty} \widehat{\mathbf{P}}(A_n),$$

then exists a single extension $\mathbf{P}\colon \sigma(\mathcal{G}) \to [0, 1]$ to $\widehat{\mathbf{P}}\colon \mathcal{G} \to [0, 1]$, such that the triplet $(\Omega, \sigma(\mathcal{G}), \mathbf{P})$ is a probability space.

Now, if we consider again our previous problem, and let $\Omega = S^1$, in order to find a uniform probabiliy function on it we can define the set $\mathcal{G}$ to be the set of all finite unions of intervals on S^1. As it is closed under union of pairs, and complements, it is an algebra. Now define $\widehat{\mathbf{P}}\colon \mathcal{G} \to [0, 1]$ as such:

$$\widehat{\mathbf{P}}\left(\biguplus_{i=1}^{N}(a_i, b_i)\right) = \sum_{i=1}^{N} \frac{b_i - a_i}{2\pi},$$

We can see that $\widehat{\mathbf{P}}$ satisfies the conditions in Theorem 1.4 and thus exists an extension $\mathbf{P}$ defined on the sigma algebra $\mathcal{B} = \sigma(\mathcal{G})$ which is also called the Borel σ-algebra of S^1. We have that $(\Omega, \mathcal{B}, \mathbf{P})$ is a probability space and we call $\mathbf{P}$ the uniform probability function on S^1.

Now we can more formally consider the properties of probability functions.

Proposition 1.5. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.*

1. $\mathbf{P}(\emptyset) = 0$.
2. *If $\{A_n\}_{n=1}^N \subset \mathcal{F}$ are disjoint sets then $\cup_{n=1}^n A_n \in \mathcal{F}$ and*
$$\mathbf{P}\left(\cup_{n=1}^n A_n\right) = \sum_{n=1}^N \mathbf{P}(A_n).$$
3. *For every $A \in \mathcal{F}$ we have $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.*
4. *If $A, B \in \mathcal{F}$ and $A \subset B$, then $(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A)$ and thus $\mathbf{P}(A) \leq \mathbf{P}(B)$.*
5. *If $A, B \in \mathcal{F}$, then*
$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

Proposition 1.6 (Continuity of the Probability Function)**.** *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.*

1. *If $(A_n)_{n=1}^\infty \subset \mathcal{F}$ is an increasing sequence of events, that is $A_1 \subset A_2 \subset A_3, \ldots$, then*
$$\mathbf{P}\left(\bigcup_{n=1}^\infty A_n)\right) = \lim_{n\to\infty}(A_n).$$
2. *If $(A_n)_{n=1}^\infty \subset \mathcal{F}$ is a decreasing sequence of events, that is $A_1 \supset A_2 \supset A_3, \ldots$, then*
$$\mathbf{P}\left(\bigcap_{n=1}^\infty A_n)\right) = \lim_{n\to\infty}(A_n).$$

In fact the last proposition is a not more than a case of the following proposition.

Proposition 1.7. *Let $(A_n)_{n=1}^\infty$ be a sequence of events in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. If the limit $\lim_{n\to\infty} A_n$ exists, then $\lim_{n\to\infty} A_n \in \mathcal{F}$, and*
$$\mathbf{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n)$$

Let us prove this theorem for the case $(A_n)_{n=1}^\infty$ is increasing. Define the following sequence:
$$B_1 = A_1$$
$$B_n = A_n \setminus A_{n-1}$$

It is clear that:

1. The sets $(B_n)_{n=1}^\infty$ are disjoint.
2. For every $N \in \mathbb{N}$ we have:
$$\bigcup_{n=1}^N B_n = \bigcup_{n=1}^N A_n = A_N.$$
3. $\cup_{n=1}^\infty B_n = \cup_{n=1}^\infty A_n$.

We now have:
$$\mathbf{P}\left(\bigcup_{n=1}^\infty A_n\right) = \mathbf{P}\left(\bigcup_{n=1}^\infty B_n\right) = \sum_{n=1}^\infty (B_n) = \lim_{N\to\infty}\sum_{n=1}^N \mathbf{P}(B_n) = \lim_{N\to\infty} \mathbf{P}\left(\bigcup_{n=1}^N B_n\right)$$
$$= \lim_{N\to\infty} \mathbf{P}(A_N).$$

2 Conditional Probability

Definition 2.1 (Conditional Probability)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and let $A, B \in \mathcal{F}$, such that $\mathbf{P}(B) > 0$. We define the probability of A given that B already happened as:

$$\mathbf{P}(A \mid B) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

The intuition behind this definition should be clear. We calculate the probability of event A "inside" event B.

Notice that we can also use conditional probability to calculate the the probability of an intersection of two events.

Proposition 2.1. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, let $B \in \mathcal{F}$ be an event such that $\mathbf{P}(B) > 0$. Then, the map $A \mapsto \mathbf{P}(A \mid B)$ is a probability function.*

The proof that the range of the function is $[0, 1]$ and that $(\Omega \mid B) = 0$ is clear from expanding the definitions, so we will only prove sigma additivity.

Proof. Let $(A_n)_{n=1}^{\infty} \subset \mathcal{F}$ be disjoint sets, then $(A_n \cap B)_{n=1}^{\infty} \subset \mathcal{F}$ are also disjoint sets and we have:

$$\begin{aligned}
\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n \mid B\right) &= \frac{\mathbf{P}\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right)}{\mathbf{P}(B)} \\
&= \frac{\mathbf{P}\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right)}{\mathbf{P}(B)} \\
&= \sum_{n=1}^{\infty} \frac{\mathbf{P}(A_n \cap B)}{\mathbf{P}(B)} \\
&= \sum_{n=1}^{\infty} \mathbf{P}(A_n \mid B)
\end{aligned}$$

□

Proposition 2.2 (Law of Total Probability)**.** *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let $N \in \mathbb{N} \cup \{\infty\}$, and $(A_n)_{n=1}^{N}$ be disjoint events such that $\cup_{n=1}^{N} A_n = \Omega$. Then,*

$$\mathbf{P}(B) = \sum_{n=1}^{N} \mathbf{P}(A_n)\mathbf{P}(B|A_n).$$

Proof.

$$\begin{aligned}
\mathbf{P}(B) &= \mathbf{P}(B \cap \Omega) \\
&= \mathbf{P}\left(B \cap \bigcup_{n=1}^{N} A_n\right) \\
&= \mathbf{P}\left(\bigcup_{n=1}^{N} (A_n \cap B)\right) \\
&= \sum_{n=1}^{N} \mathbf{P}(A_n \cap B) \\
&= \sum_{n=1}^{N} \mathbf{P}(A_n)\mathbf{P}(B \mid A_n).
\end{aligned}$$

□

Example 2.1 (Pólya's urn, simplified)**.** Let there be 1 white and 1 black ball in an urn. At each step, one ball is drawn uniformly at random from the urn, and its color observed; it is then returned in the urn, and an additional ball of the same color is added to the urn. What is the probability that there are k black balls in the urn after the n-th step?

First denote:

$$\begin{aligned} A_{n,k} &= \{\text{there are } k \text{ black balls after the } n\text{-th step.}\} \\ p_{n,k} &= \mathbf{P}(A_{n,k}). \end{aligned}$$

In order for there to be k black balls after the n-th step, there must either have been $k-1$ or k black balls in the $n-1$-th step. Thus,

$$\begin{aligned} \mathbf{P}(A_{n,k}) &= \mathbf{P}(A_{n,k} \cap (A_{n-1,k-1} \cup A_{n-1,k})) \\ &= \mathbf{P}(A_{n-1,k-1})\mathbf{P}(A_{n,k} \mid A_{n-1,k-1}) + \mathbf{P}(A_{n-1,k})\mathbf{P}(A_{n,k} \mid A_{n-1,k}). \end{aligned}$$

This implies that

$$p_{n,k} = \frac{k-1}{n+1} p_{n-1,k-1} + \frac{n+1-k}{n+1} p_{n-1,k}.$$

Coupled with the fact that $p_{0,1} = 1$ we can verify that the only solution under these conditions is $p_{n,k} = \frac{1}{n+1}$. In general, these problems are very hard to solve.

Another useful trick is Bayes' theorem. In its simplified version it states that,

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(B \mid A)\mathbf{P}(A)}{\mathbf{P}(B)},$$

and can be solved without much thought. Here's the general theoerm.

Theorem 2.3. (Bayes' Theorem). *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Let $N \in \mathbb{N} \cup \{\infty\}$, and $(A_n)_{n=1}^N$ be disjoint events such that $\cup_{n=1}^N A_n = \Omega$. Then,*

$$\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(B \mid A_i)\mathbf{P}(A_i)}{\sum_{n=1}^N \mathbf{P}(A_n)\mathbf{P}(B \mid A_n)}.$$

Proof. Left as an exercise to the reader. □

Example 2.2. Suppose we have a test for checking whether a person has the terrible the terrible "cooties". It has a true positive rate of 0.98, and a false positive rate of 0.01. Assume that 0.1% of the population has the cooties, what is the probability that a person who got a positive result has the cooties?

Denote,

$$\begin{aligned} A &= \{\text{the person is healthy}\} \\ B &= \{\text{the answer is positive}\}. \end{aligned}$$

From Bayes' theorem we have:

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(B \mid A)\mathbf{P}(A)}{\mathbf{P}(B)} = \frac{0.01 \cdot 0.999}{\mathrm{P}(B)}.$$

From the law of total probability we have

$$\begin{aligned} \mathbf{P}(B) &= \mathbf{P}(A)\mathbf{P}(B \mid A) + \mathbf{P}(A^c)\mathbf{P}(B \mid A^c) \\ &= 0.01 \cdot 0.999 + 0.98 \cdot 0.001 = 0.01097. \end{aligned}$$

And thus,

$$\mathbf{P}(A \mid B) = \frac{0.01 \cdot 0.999}{0.01097} \approx 0.91$$

3 Independance and Repeating Experiments