BUILDING PROBABILITY SPACES

Recall that a *probability space* is a triple $(\Omega, \mathcal{F}, P)$, where:

- Ω is a set, called the *sample space*.
- $\mathcal{F}$ is a set of subsets of Ω, called the *events*. We require that it is a *σ-field*. This means that it satisfies the following properties:
  (1) $\emptyset \in \mathcal{F}$.
  (2) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
  (3) If $A_1, A_2, \ldots$ are all in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

  In other words, $\mathcal{F}$ contains the empty set and is closed under complementation and taking countable unions.
- $P : \mathcal{F} \to [0, 1]$ is called the *probability function*. It satisfies the following properties:
  (1) $P(\emptyset) = 0$.
  (2) If $A_1, A_2, \ldots$ are pairwise disjoint and in $\mathcal{F}$, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

  In other words, the probability of the empty set is zero, and P is *countably additive*.

We have seen several examples of *finite* probability spaces. For example, the uniform coin tossing space consists of $\Omega = \{H, T\}$, $\mathcal{F}$ consists of all subsets of Ω, and $P(A) = \frac{|A|}{2}$.

We now would like to give a probability space that corresponds to the experiment of picking a real number in $[0, 1]$ "uniformly at random". You should be able to convince yourself that, in such a space, we should have that for $a \leq b$, $P((a, b)) = b - a$, where (a, b) is the open interval $\{x \in [0, 1] \mid a < x < b\}$. Thus we want a probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = [0, 1]$, and for all $a, b \in [0, 1]$ with $a \leq b$, $(a, b) \in \mathcal{F}$ and $P((a, b)) = b - a$. How does one build such a space? We have seen (in assignment 1) that it can be tricky, and in fact it is impossible to have such a space where $\mathcal{F}$ contains *all* subsets of Ω. In fact we would be satisfied with having $\mathcal{F}$ being "as small as possible" but still containing the open intervals. Toward building such a small σ-field, we prove two lemmas:

Lemma 1. Fix a sample space Ω. Let $(\mathcal{F}_i)_{i \in I}$ be a non-empty (but not necessarily countable) collection of σ-fields on Ω. Then $\bigcap_{i \in I} \mathcal{F}_i$ is also a σ-field.

Proof. This is straightforward. For example, the empty set is in $\bigcap_{i \in I} \mathcal{F}_i$ since it is in each $\mathcal{F}_i$ (as they are σ-fields). Similarly, if A is in $\bigcap_{i \in I} \mathcal{F}_i$, then $A \in \mathcal{F}_i$ for each i so $A^c \in \mathcal{F}_i$ for each i (as they are σ-fields), so $A^c \in \bigcap_{i \in I} \mathcal{F}_i$. The proof that $\bigcap_{i \in I} \mathcal{F}_i$ is closed under countable unions is completely similar (try to write it down!). □

This lets us define a *smallest σ-field* containing a given collection.

Lemma 2. Fix a sample space Ω. Let $\mathcal{A}$ be any collection of subsets of Ω. Then there exists a smallest σ-field $\mathcal{F}$ containing $\mathcal{A}$.

Proof. Let $\mathcal{F}$ be the intersection of *all* σ-fields containing $\mathcal{A}$. Note that there is at least one such σ-field: the σ-field containing *all* subsets of Ω. Thus by Lemma 1,

$\mathcal{F}$ is a σ-field and the definition shows that it must be included in *all* other σ-fields containing $\mathcal{A}$, hence it is the smallest such σ-field. □

We will write $\sigma(\mathcal{A})$ for the smallest σ-field containing $\mathcal{A}$. We will apply Lemma 2 to the following setup:

Definition 3. An *interval* is a set A of reals such that for some reals $a \leq b$, A is either $[a, b]$, $[a, b)$, $(a, b]$, or (a, b).

Definition 4. Let $\Omega = [0, 1]$ and let $\mathcal{A}$ be the set of all intervals that are subsets of $[0, 1]$. We call a set in $\sigma(\mathcal{A})$ a *Borel* subset of $[0, 1]$ and call $\sigma(\mathcal{A})$ the *Borel* σ-field of $[0, 1]$.

Intuitively, a Borel set is one that can be obtained from intervals by taking unions or complementing (maybe doing these operations many times). Any set that we are likely to encounter is Borel, and in fact it is quite hard to exhibit a set that is *not* Borel.

The set $\mathcal{A}$ of subintervals of $[0, 1]$ is not a σ-field or even a field (it is not closed under complements). However, it is closed under intersections, and complements are disjoint unions of members of $\mathcal{A}$. In other words, it is a semi-algebra:

Definition 5. Fix a sample space Ω. A *semi-algebra* is a collection $\mathcal{A}$ of subsets of Ω such that:

(1) $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$.
(2) If A_1 and A_2 are in $\mathcal{A}$, then $A_1 \cap A_2$ are in $\mathcal{A}$.
(3) If $A \in \mathcal{A}$, then there exists pairwise disjoint $B_1, B_2, \ldots, B_k \in \mathcal{A}$ such that $A^c = B_1 \cup B_2 \cup \ldots \cup B_k$.

You should check for yourself that the collection $\mathcal{A}$ of subintervals of $[0, 1]$ is a semi-algebra. To see one idea for the proof, observe for example that $(0, \frac{1}{2})^c$ can be written as $\{0\} \cup [\frac{1}{2}, 1] = [0, 0] \cup [\frac{1}{2}, 1]$, a disjoint union of intervals.

When A is a subinterval of $[0, 1]$, it is clear how one should define the probability of A (when each number in $[0, 1]$ should be equally likely): we must have $P(A) = b - a$, when for example $A = [a, b]$, and similarly for the other types of intervals (one point does not matter). Further, it turns out that every Borel set can in some sense be approximated by intervals, hence the property that $P(A) = b - a$ uniquely specifies the probability function P. In general, once a probability function is specified on a semi-algebra $\mathcal{A}$ it can (under some mild conditions) be extended to a unique probability function on $\sigma(\mathcal{A})$. This is the content of the *extension theorem*:

Theorem 6 (The extension theorem)**.** Fix a sample space Ω and let $\mathcal{A}$ be a semi-algebra. Let $P_0 : \mathcal{A} \to [0, 1]$ satisfy the following three conditions:

(1) $P_0(\emptyset) = 0$ and $P_0(\Omega) = 1$.
(2) Finite additivity: $P_0(\bigcup_{i=1}^k A_i) = \sum_{i=1}^k P_0(A_i)$ whenever $A_1, A_2, \ldots, A_k$ are in $\mathcal{A}$, $\bigcup_{i=1}^k A_i \in \mathcal{A}$, and the A_i's are pairwise disjoint.
(3) Countable monotonicity: $P_0(A) \leq \sum_{i=1}^\infty P_0(A_i)$ whenever $A, A_1, A_2, \ldots$ are all in $\mathcal{A}$ and $A \subseteq \bigcup_{i=1}^\infty A_i$.

Then there exists a unique probability function $P : \sigma(\mathcal{A}) \to [0, 1]$ which extends P_0 (that is, $P(A) = P_0(A)$ for any $A \in \mathcal{A}$).

We take the extension theorem as a black box and will *not* prove it (the proof is not that difficult but quite long and technical; you can find it for example in Rosenthal's *a first look at rigorous probability* - see the syllabus for a precise reference).

The extension theorem implies we can build the desired uniform probability space on $[0, 1]$:

Corollary 7. There is a unique probability space $(\Omega, \mathcal{F}, P)$ such that $\Omega = [0, 1]$, $\mathcal{F}$ is the σ-field of Borel sets, and $P((a, b)) = P([a, b)) = P((a, b]) = P([a, b]) = b - a$ for all $a \leq b$ in $[0, 1]$.

Proof. Let $\mathcal{A}$ be the semi-algebra of subintervals of $[0, 1]$. Define P_0 on $\mathcal{A}$ by $P_0((a, b)) = b - a$, and similarly for the other variants of intervals. Using some hard result of analysis (you will do this in assignment 6), one can show that the three conditions of the extension theorem are satisfied (the third, countable monotonicity, is the hard one). Thus there is a unique probability function $P : \mathcal{F} \to [0, 1]$ which extends P_0, as desired. □

Remark 8. The probability function derived in the corollary is often called the *Lebesgue measure* on $[0, 1]$. Intuitively, it gives the "length" of any nice (i.e. Borel) subset of $[0, 1]$.

The extension theorem has other applications. You will see in assignment 6 how to build the probability space modeling the experiment of tossing a coin infinitely-many times. Here, we show how to model the *product* of two experiments. This is the experiment which consists in running the first experiment, getting a result x, then independently running the second experiment, getting a result y and having the result in the product experiment be (x, y). The probability of getting an event of the form $A \times B$ should then be $P(A)P(B)$ (as the events A and B come from different experiments). Formally:

Definition 9. Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be probability spaces. We define their *product* $(\Omega, \mathcal{F}, P)$ as follows:

- $\Omega = \Omega_1 \times \Omega_2$ (recall that this is the set of pairs (a, b) such that $a \in \Omega_1$ and $b \in \Omega_2$).
- $\mathcal{F} = \sigma(\mathcal{A})$, where:
$$\mathcal{A} = \{A_1 \times A_2 \mid A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$$
is the collection of "rectangles" in Ω. As an exercise in assignment 6, you will check that $\mathcal{A}$ is a semi-algebra.
- First define $P_0 : \mathcal{A} \to [0, 1]$ by $P_0(A_1 \times A_2) = P_1(A_1)P_2(A_2)$. Again, it is possible (but not trivial; you can take it as a blackbox in this class) to show that P_0 satisfies the conditions of the extension theorem. Thus there is a unique probability function $P : \mathcal{F} \to [0, 1]$ that extends P_0.

Applying this to the uniform probability space on $[0, 1]$ (see Corollary 7), we obtain:

Corollary 10. There is a probability space $(\Omega, \mathcal{F}, P)$ with Ω the unit rectangle $[0, 1] \times [0, 1]$ such that (for $a \leq b, c \leq d$ all in $[0, 1]$) $P([a, b] \times [c, d]) = (b - a)(d - c)$.

Proof. Take the product of $([0, 1], \mathcal{F}, P)$ as described in Corollary 7 with itself. □

The space described by this corollary is essentially the uniform space on $[0, 1]^2$: it is the probability space one obtains when throwing a dart uniformly at random at the unit square and recording the result. The probability of hitting a given set is given by its *area*. When the set is a rectangle, its area is clear. When it is not (for example when it is a disk), we can approximate the disk by rectangles and compute the area of the disk via the approximation by rectangles (in much the same way that the integral is defined).

What if we want instead to throw darts at only the disk of radius 1? We can then consider the *subspace* of the uniform probability space $[0,1]^2$ which consists of only the disk. The probability is obtained by conditioning on the event "the dart hits the disk". Formally:

Definition 11. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A \in \mathcal{F}$ be such that $P(A) \neq 0$. The *subspace* of this probability space induced by A is the space $(\Omega_0, \mathcal{F}_0, P_0)$ defined as follows:

(1) $\Omega_0 = A$.
(2) $\mathcal{F}_0 = \{B \in \mathcal{F} \mid B \subseteq A\}$.
(3) $P_0(B) = P(B|A)$.

You will show in assignment 6 that this is indeed a probability space. In the discussion above, the starting space would be the uniform space on $[0,1]^2$ and A would be the disk of radius 1.