Handout 04: Formal Definition of Probability

The Axioms of Probability

We now come to the point where we will more formally define the central objects of study for this semester: probability spaces. We have already defined the concept of a sample space, and it would seem natural to define a probability space by specifying a function:

$$f : S \to [0, 1]$$

That maps each element of the sample space into a 'probability'. This is a reasonable thing to do, and we will often specify specific distribution in such a way. However, in the formal definition of probability we instead deal with the probability of *events*. An event E is a subset of the sample space S; we will actually define a probability function that maps events into probabilities rather than just the unique elements of S. That is:

$$f : \mathcal{P}(S) \to [0, 1]$$

Where $\mathcal{P}(S)$ is the power set, all possible subsets, of the sample space. Restrictions are then placed on these probabilities so they behave in natural ways.

Why the roundabout approach? We will eventually want to work with probability spaces defined on an uncountable set, such as the entire set of $\mathbb{R}$. Here, the probability of observing any particular element of S is exactly zero, and the only way to define probabilities is by describing the probabilities of events.

Definition 1 (Axioms of Probability) *For a sample space S and function $\mathbb{P} : \mathcal{P}(S) \to [0, 1]$, the following are nessisary and sufficent conditions for $(S, \mathbb{P}(\cdot))$ being a valid probability model:*

- *For any event $A \subseteq S$, $\mathbb{P}(A) \geq 0$.*
- *$\mathbb{P}(S) = 1$.*
- *For any collection of events $A_1, A_2, \ldots A_n, \ldots$ satisfying the conditions $A_i \cap A_j = \emptyset$ for all $i \neq j$ the probability that at least one of the events among the collection $\{A_i, i = 1, 2, 3, \ldots\}$ occurs may be computed as:*

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The first two of these axioms should cause no anxiety, since what they say is so simple and intuitive. Because we generally think of probabilities as long-run relative frequencies of occurrence, we would not want them to be negative. Further, we know when we define the sample space of a random experiment that it encompasses everything that can happen. It thus necessarily has probability one. The third statement may seem somewhat difficult to understand at first. Let's carefully examine what it says.

In the previous worksheet we referred to non-overlapping sets as "disjoint." In probability theory, it is customary to use the alternative phrase "mutually exclusive" to describe events which do not overlap. The message of the third axiom may be restated as: For any collection of mutually exclusive events in a discrete random experiment, the probability of their union is equal to the sum of their individual probabilities. The assertion made in Axiom 3 is intended to hold for a collection of arbitrary size, including collections that are countably infinite. Using the following theorems, we can easily show that it must also hold for any finite collection of sets.

Basic Probability Theorems

In this section we will state and prove several key results about general probability spaces.

Theorem 1 *For any event $A \subseteq S$, the probability of A^c, the complement of A, is equal to $1 - \mathbb{P}(A)$.*

Proof. Note that the sample space S may be represented as the union of two mutually exclusive events. Specifically, $S = A \cup A^c$, where $A \cap A^c = \emptyset$. It follows from the axioms that:

$$\begin{aligned} 1 &= \mathbb{P}(S) \\ &= \mathbb{P}(A \cup A^c) \\ &= \mathbb{P}(A) + \mathbb{P}(A^c) \end{aligned}$$

And the result follows by subtracting $\mathbb{P}(A^c)$ from both sides of the equation $\blacksquare$.

Theorem 2 *For any probability model, $\mathbb{P}(\emptyset) = 0$.*

Proof. Given the prior theorem, $\mathbb{P}(\emptyset)$ is equal to $1 - \mathbb{P}(S)$ which known to be 0 by the axioms $\blacksquare$.

Theorem 3 (The Monotonicity Property) *If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

Proof. Since $A \subseteq B$, we may write

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

Moreover, the events A and $B - A$ are mutually exclusive. Thus, we have,

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B - A).$$

Since, by the first axiom, $\mathbb{P}(B - A) \geq 0$, we see that $\mathbb{P}(B) \geq \mathbb{P}(A)$, as claimed ■.

So far we have several results for calculating the probability of the union of disjoint events. The next results compliments this by expressing the probability of the intersection of two events in terms of simplier probabilities.

Theorem 4 (The Addition Rule) *For arbitrary events A and B, we have:*

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

Proof. Note that $A \cup B$ may be written as $A \cup (B - A)$; since $A \cap (B - A) = \emptyset$, the third axiom implies that

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B - A).$$

Similarly, since $B = (A \cap B) \cup (B - A)$, and $(A \cap B) \cap (B - A) = \emptyset$, we also have that

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B - A).$$

We may rewrite this as

$$\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A \cup B).$$

Substituting into the first equation yields the desired result ■.

We now come to a somewhat more complex claim, a property of probabilities called "countable subadditivity." It is a result which is useful in providing a lower bound for probabilities arising in certain statistical applications (for example, when you want to be at least 95% sure of something). In addition to its utility, it provides us with a vehicle for garnering a little more experience in "arguing from first principles."

Theorem 5 (Countable subadditivity) *For any collection of events $A_1, A_2, \ldots, A_n, \ldots$ we have:*

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

Proof. Let's define a new collection of events, a collection that is mutually exclusive yet accounts for all the elements contained in the events A_1, A_2,.... Let

$$B_1 = A_1,$$

and let

$$B_2 = A_2 - A_1,$$

We may describe the event B_2 as "the new part of A_2," that is, the part of A_2 that has not been accounted for in A_1. Clearly, $B_1 \cup B_2 = A_1 \cup A_2$ and, in addition, $B_1 \cap B_2 = \emptyset$. Now, let

$$B_k = A_k - \left(\bigcup_{i=1}^{k-1} A_i \right) \quad \text{for all integers k} \geq 3$$

Each set B can be thought of as the "new part" of the corresponding A. In general, B_k is the set of all elements of A_k which are not contained in any of the preceding $(k-1)$ A's. Note that any two of the B events are mutually exclusive, that is, for any $i < j$,

$$B_i \cap B_j = \emptyset$$

This follows from the fact that, for arbitrary $i < j$, $B_i \subseteq A_i$ while $B_j \subset A_i^c$ since B_j only contains elements that were not members of any preceding event A, including of course, the event A_i. Moreover, for every positive k, a simple set theoretic argument shows that

$$\bigcup_{i=1}^{k} B_i = \bigcup_{i=1}^{k} A_i.$$

Using the same basic argument, we may conclude that

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

Now we are in a position to apply the third axiom, which together with the Monotonicity Property, yields

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i \right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i \right) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(B_i) \\ &\leq \sum_{i=1}^{\infty} \mathbb{P}(A_i) \end{aligned}$$

and completes the proof ■.