On Probability Axioms and Sigma Algebras

Abstract

These are supplementary notes that discuss the axioms of probability for systems with finite, countably infinite, and uncountably infinite sample spaces. Section I reviews basic material covered in class. Section II discusses a new and complex issue that arises in the uncountably infinite case. Specifically, if the sample space is uncountably infinite, then it is not possible to define probability measures for *all events*. Rather, probabilities are defined only for a large collection of events, called a *sigma algebra*. Fortunately, the standard sigma algebras that are used are so big that they encompass most events of practical interest. In fact, one can rigorously argue that they include *all* events of practical interest. Because of this, students can freely assume that all of the events considered in this course are *measurable* (so that they are in the sigma algebra and thus have well defined probabilities).

I. Axioms of Probability

Recall that a probabilistic system is defined by a *sample space* $\mathcal{S}$, which is a general set, and a *probability measure* $P[\mathcal{E}]$ defined on subsets $\mathcal{E} \subseteq \mathcal{S}$. Each subset $\mathcal{E}$ of the sample space is called an *event*. Let ϕ denote the *empty set*. The empty set is considered to be a subset of every set, and hence is also considered to be an event.

A. Simplified Axioms of Probability (without sigma algebras)

First assume that we want to define a probability measure $P[\mathcal{E}]$ for *all subsets* $\mathcal{E}$ of the sample space $\mathcal{S}$, including the empty set ϕ. The probability measure must satisfy the following three *axioms of probability*:

1) $P[\mathcal{S}] = 1$.
2) $P[\mathcal{E}] \geq 0$ for all events $\mathcal{E} \subseteq \mathcal{S}$.
3) If $\{\mathcal{E}_n\}_{n=1}^M$ is a finite sequence of mutually exclusive events (so that $\mathcal{E}_n \cap \mathcal{E}_m = \phi$ for all $n \neq m$), then:

$$P[\cup_{n=1}^M \mathcal{E}_n] = \sum_{n=1}^M P[\mathcal{E}_n] \quad \text{(finite additvity)}$$

Likewise, if $\{\mathcal{E}_n\}_{n=1}^\infty$ is a countably infinite sequence of mutually exclusive events, then:

$$P[\cup_{n=1}^\infty \mathcal{E}_n] = \sum_{n=1}^\infty P[\mathcal{E}_n] \quad \text{(countable additivity)}$$

B. Basic Consequences

Using these axioms, one can prove that $P[\phi] = 0$, and that $P[\mathcal{E}] \leq 1$ for all events $\mathcal{E}$, as shown below.

Lemma 1: $P[\phi] = 0$.

Proof: Note that $\mathcal{S}$ and ϕ are mutually exclusive because $\mathcal{S} \cap \phi = \phi$, and so $P[\mathcal{S} \cup \phi] = P[\mathcal{S}] + P[\phi]$ by axiom 3. Then:

$$P[\mathcal{S}] = P[\mathcal{S} \cup \phi] = P[\mathcal{S}] + P[\phi]$$

Because $P[\mathcal{S}] = 1$, the above implies that $1 = 1 + P[\phi]$, and so $P[\phi]$ must be 0. □

Lemma 2: For any event $\mathcal{E} \subseteq \mathcal{S}$, we have $P[\mathcal{E}] \leq 1$.

Proof: Let $\mathcal{E}$ be a subset of $\mathcal{S}$, and let $\mathcal{E}^c$ denote its complement (being all elements of $\mathcal{S}$ that are not in $\mathcal{E}$). Then $\mathcal{E}$ and $\mathcal{E}^c$ are mutually exclusive, and $\mathcal{E} \cup \mathcal{E}^c = \mathcal{S}$. Then:

$$\begin{aligned} P[\mathcal{E}] &\leq P[\mathcal{E}] + P[\mathcal{E}^c] && (1) \\ &= P[\mathcal{E} \cup \mathcal{E}^c] && (2) \\ &= P[\mathcal{S}] && (3) \\ &= 1 && (4) \end{aligned}$$

where (1) holds by axiom 2, (2) holds by axiom 3, and (4) holds by axiom 1. □

It is also easy to prove that if $\mathcal{E}_1$ and $\mathcal{E}_2$ are two events such that $\mathcal{E}_1 \subseteq \mathcal{E}_2$, then $P[\mathcal{E}_1] \leq P[\mathcal{E}_2]$ (this is left as an exercise).

C. Discussion on modifying the axioms

The third axiom can be modified by stating only the countable additivity case because the finite additivity statement follows as a consequence: First one must modify the above proof that $P[\phi] = 0$ to use only countably infinite unions via $\mathcal{S} = \mathcal{S} \cup \cup_{i=2}^{\infty}\phi$ (this is left as an exercise). Then, given a finite collection of disjoint events $\{\mathcal{E}_i\}_{i=1}^n$, we can write $\cup_{i=1}^n \mathcal{E}_i$ as a countably infinite union:

$$\cup_{i=1}^n \mathcal{E}_i = \cup_{i=1}^n \mathcal{E}_i \cup \cup_{i=n+1}^{\infty}\phi$$

and so

$$P[\cup_{i=1}^n \mathcal{E}_i] = P[\cup_{i=1}^n \mathcal{E}_i \cup \cup_{i=n+1}^{\infty}\phi] \overset{(a)}{=} \sum_{i=1}^n P[\mathcal{E}_i] + \sum_{i=n+1}^{\infty} P[\phi] = \sum_{i=1}^n P[\mathcal{E}_i] + \sum_{i=n+1}^{\infty} 0 = \sum_{i=1}^n P[\mathcal{E}_i]$$

where (a) holds by countable additivity.

In contrast, it can be shown that it is *impossible* to prove countable additivity only from finite additivity. This is because there are examples of systems that satisfy the first two axioms together with the finite additivity statement of Axiom 3, but do not satisfy the countable additivity statement. Such examples are surprisingly difficult to construct. A (partial) example is as follows: Define $\mathcal{S} = \mathbb{N} = \{1, 2, 3, ...\}$. For each finite subset $A \subseteq \mathbb{N}$ define $|A|$ as the number of elements in A. Now for each (possibly infinite) subset $A \subseteq \mathbb{N}$ define:

$$P[A] = \lim_{n\rightarrow\infty} \frac{|A \cap \{1, 2, ..., n\}|}{n} \tag{5}$$

whenever the limit exists. For example $P[\{evens\}] = P[\{2, 4, 6, ...\}] = 1/2$. Also, from (5) it follows that $P[A] = 0$ whenever A is a finite subset of $\mathbb{N}$. Unfortunately, there are some (infinite) sets $A \subseteq \mathbb{N}$ for which the above limit does not exist. However, values of $P[A]$ for such sets can be consistently defined using an advanced theory of *Banach limits*, so that $P[A]$ indeed exists for all sets $A \subseteq \mathbb{N}$ and satisfies (5) whenever the limit exists.[1] It can be shown that the resulting $P[A]$ satisfies the first two axioms, and satisfies the finite additivity statement of Axiom 3. It does not satisfy the countable additivity statement of Axiom 3 (and hence P not a valid probability measure) because:

$$\mathcal{S} = \cup_{i=1}^{\infty}\{i\}$$

but

$$1 = P[\mathcal{S}] \neq \sum_{i=1}^{\infty} P[\{i\}] = 0$$

where we have used $P[\{i\}] = 0$ for all natural numbers i because $\{i\}$ is a finite set.[2]

D. Probability Measures for Finite Sample Spaces

Suppose that the sample space $\mathcal{S}$ is finite with N possible outcomes, where N is a positive integer. Then $\mathcal{S}$ has the form:

$$\mathcal{S} = \{\omega_1, \omega_2, \dots, \omega_N\}$$

where ω_i represents the ith possible outcome, for $i \in \{1, \dots, N\}$. This sample space contains exactly 2^N possible events (including the empty set). We can easily construct a probability measure by defining probabilities p_i for each particular outcome $\omega_i \in \mathcal{S}$, noting that a particular outcome can be viewed as a set with one element:[3]

$$P[\omega_i] = p_i \ \ \forall i \in \{1, \dots, N\}$$

[1] A detailed development of Banach limits is beyond the scope of this course. That is why this is only a "partial" example.

[2] The *union bound* $P[\cup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} P[A_i]$ also fails for this example. The statement of the union bound with countably infinite unions can only be proven from countable additivity.

[3] Recall that the probability measure $P[\mathcal{E}]$ is defined on sets $\mathcal{E}$ that are subsets of $\mathcal{S}$. Thus, one might prefer to use the notation $P[\{\omega_i\}]$ to emphasize that $\{\omega_i\}$ is the *set* consisting of the single element ω_i. However, we use the notation $P[\omega_i]$ because it is simpler and it causes no confusion.

We can then extend the probability measure to *all events* $\mathcal{E} \subseteq \mathcal{S}$ by defining $P[\phi] = 0$, and by defining $P[\mathcal{E}]$ for each non-empty subset $\mathcal{E} \subseteq \mathcal{S}$ as follows:

$$P[\mathcal{E}] = \sum_{\omega_i \in \mathcal{E}} P[\omega_i]$$

where the summation on the right-hand-side represents a sum over each of the elements ω_i in the set $\mathcal{E}$. It is easy to show that this definition of $P[\mathcal{E}]$ satisfies the three axioms of probability if and only if $p_i \geq 0$ for all $i \in \{1, 2, \ldots, N\}$, and $\sum_{i=1}^{N} p_i = 1$ (this is left as an exercise).

E. Probability Measures for Countably Infinite Sample Spaces

Now suppose the sample space $\mathcal{S}$ has a countably infinite number of outcomes, so that it can be written:

$$\mathcal{S} = \{\omega_1, \omega_2, \omega_3, \ldots\}$$

We can easily construct a probability measure $P[\mathcal{E}]$ for all possible events $\mathcal{E} \subseteq \mathcal{S}$ in the same way as before: Define $P[\phi] = 0$, and define p_i as the probability of outcome ω_i:

$$P[\omega_i] = p_i \ \ \forall i \in \{1, 2, 3, \ldots\}$$

Because we can always sum non-negative numbers over a finite or countably infinite number of terms, we can define $P[\mathcal{E}]$ for each non-empty subset $\mathcal{E} \subseteq \mathcal{S}$ by:

$$P[\mathcal{E}] = \sum_{\omega_i \in \mathcal{E}} P[\omega_i]$$

Again, it is easy to show that the above definition satisfies the three axioms of probability if and only if $p_i \geq 0$ for all $i \in \{1, 2, 3, \ldots\}$ and $\sum_{i=1}^{\infty} p_i = 1$ (this is left as an exercise).

II. A Can of Worms: Uncountably Infinite Sample Spaces

The previous section defined the probability measure $P[\mathcal{E}]$ for all events $\mathcal{E}$ of a finite or countably infinite sample space. This was done by summing the probabilities of individual outcomes in $\mathcal{E}$. This is no longer possible when the sample space is uncountably infinite, because summation over an uncountable number of terms is not defined.

Suppose the sample space $\mathcal{S}$ contains an uncountably infinite number of outcomes. For example, $\mathcal{S}$ might represent all real numbers in the unit interval $[0, 1]$. Rather than building the probability measure $P[\mathcal{E}]$ by attempting to sum individual probabilities, one typically constructs $P[\mathcal{E}]$ by first defining it for classes of simple events, and then extending this definition to more complex events in such a way that the axioms of probability are satisfied. An example of this for the sample space consisting of the unit interval is given in the next subsection.

A. The Uniform Distribution over $[0, 1]$

The *uniform probability distribution* over the sample space $\mathcal{S} = [0, 1]$ is defined by first considering events $\mathcal{E}$ that are *open intervals* of the type (a, b) for real numbers a, b that satisfy $0 \leq a < b \leq 1$. Probabilities $P[(a, b)]$ for such intervals are defined:

$$P[(a, b)] = b - a$$

This means that the probability of the open interval $(0, 1)$ is equal to 1. It follows that the probability of the outcome 0 is 0, as is the probability of the outcome 1. In fact, if the probability of open intervals is defined this way, one can use the axioms of probability to show that the probability of any individual outcome is 0. It follows that the probability of any open or closed interval in $[0, 1]$ is simply the size of that interval. Again using the axioms of probability, it follows that the probability of any event that is the (finite or countable) union of non-overlapping intervals in $[0, 1]$ is just the sum of the interval sizes. We can use further limiting arguments to find what the probabilities must be for more complicated sets that can be written as countable unions of intervals, and complements of these, and so forth. This results in well defined probabilities for most sets of practical interest. The collection of these sets is called the *Borel sigma algebra*.

However, a surprising result is that this procedure *does not allow probabilities to be defined for all subsets of* $[0, 1]$. That is, it can be shown that *there exist non-measurable sets*, being sets for which it is impossible to define probabilities. This strange situation arises only for sample spaces $\mathcal{S}$ that are uncountably infinite, and is discussed in more detail in advanced probability texts that use measure theory, such as [1]. For our purposes, this creates some confusion regarding the axioms of probability: How can these axioms hold if some events $\mathcal{E}$ do not even have probabilities? This is solved by restricting the axioms to apply only to certain collections of events in $\mathcal{S}$. These collections are called *sigma algebras* and are defined in the next subsection.

B. Sigma Algebras

Let $\mathcal{S}$ be any (non-empty) set that represents a sample space. Here we allow the set $\mathcal{S}$ to be arbitrary (either finite, countably infinite, or uncountably infinite). Let $\mathcal{F}$ denote a *collection of subsets of* $\mathcal{S}$. We say a collection $\mathcal{F}$ is a *sigma algebra* (also called σ*-algebra* or σ*-field*) if it satisfies the following three properties:[4]

- The full sample space $\mathcal{S}$ is in $\mathcal{F}$, and the empty set ϕ is in $\mathcal{F}$.
- If a set $\mathcal{E}$ is in $\mathcal{F}$, then its complement $\mathcal{E}^c$ is also in $\mathcal{F}$.
- If $\{\mathcal{E}_n\}_{n=1}^M$ is a finite sequence of subsets, each of which is in $\mathcal{F}$, then the union $\cup_{n=1}^M \mathcal{E}_n$ is also in $\mathcal{F}$. Likewise, if $\{\mathcal{E}_n\}_{n=1}^\infty$ is a countable sequence of subsets, each of which is in $\mathcal{F}$, then the countable union $\cup_{n=1}^\infty \mathcal{E}_n$ is also in $\mathcal{F}$.

We will use $\mathcal{F}$ to represent the collection of events $\mathcal{E} \subseteq \mathcal{S}$ for which we can define probabilities $P[\mathcal{E}]$. The set $\mathcal{F}$ consisting of *all subsets* of the sample space $\mathcal{S}$ is (trivially) a sigma algebra. However, it is possible to have sigma algebras that do *not* contain all subsets of $\mathcal{S}$. For example, given any non-empty sample space $\mathcal{S}$, the 2-element collection of subsets consisting of only ϕ and $\mathcal{S}$ is trivially a sigma algebra. For another example, if we have a finite sample space $\mathcal{S} = \{1, 2, 3, 4\}$, it is easy to show that the following collection of sets $\mathcal{F}$ satisfies the defining properties of a sigma algebra:

$$\mathcal{F} = \{\phi, \{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}\} \tag{6}$$

This particular sigma algebra $\mathcal{F}$ does not include the set $\{1, 2, 3\}$.

The example (6) illustrates a sigma algebra for a finite sample space $\mathcal{S}$. Of course, for the context of probability theory, there is no reason to talk about sigma algebras at all unless we consider sample spaces $\mathcal{S}$ that are uncountably infinite. In the case when $\mathcal{S}$ is the unit interval $[0, 1]$, the *Borel sigma algebra* $\mathcal{F}$ is the collection of all open intervals (a, b) such that $0 \leq a < b \leq 1$, together with all complements of these, all finite or countably infinite unions of these, all complements of these unions, and so on. It is not obvious that this procedure does not include all possible subsets of $[0, 1]$, but it can be shown that it does not.

C. Probability Axioms with Sigma Algebras

We can now present a rigorous statement of the probability axioms that hold for any sample space $\mathcal{S}$ (finite, countably infinite, or uncountably infinite). These axioms assume a-priori that we are given a particular sigma algebra $\mathcal{F}$ for $\mathcal{S}$. Recall that $\mathcal{F}$ necessarily includes both $\mathcal{S}$ and the empty set ϕ. The sigma algebra $\mathcal{F}$ represents the collection of all events for which we have well defined probabilities. Assume we are given a function $P[\mathcal{E}]$ that assigns a real number to each event $\mathcal{E}$ in $\mathcal{F}$ (the function does not need to be defined for events $\mathcal{E}$ that are not in $\mathcal{F}$). Then $P[\mathcal{E}]$ is a probability measure if it satisfies the following three axioms:

1) $P[\mathcal{S}] = 1$.
2) $P[\mathcal{E}] \geq 0$ for all events $\mathcal{E}$ in the collection $\mathcal{F}$.
3) If $\{\mathcal{E}_n\}_{n=1}^M$ is a finite sequence of mutually exclusive events, each of which is in the collection $\mathcal{F}$, then:

$$P[\cup_{n=1}^M \mathcal{E}_n] = \sum_{n=1}^M P[\mathcal{E}_n] \quad \text{(finite additivity)}$$

Likewise, if $\{\mathcal{E}_n\}_{n=1}^\infty$ is a countably infinite sequence of mutually exclusive events, each of which is in the collection $\mathcal{F}$, then:

$$P[\cup_{n=1}^\infty \mathcal{E}_n] = \sum_{n=1}^\infty P[\mathcal{E}_n] \quad \text{(countable additivity)}$$

[4]We say that $\mathcal{F}$ is an *algebra* if the same three properties hold, with the exception that the third property is modified by removing the second sentence of that property. That is, we remove the requirement about countably infinite unions, so that we only require finite unions of elements in $\mathcal{F}$ to also be in $\mathcal{F}$. The "sigma" part of "sigma algebra" emphasizes that we enforce the stronger requirement about countable unions. Thus, all sigma algebras are algebras, but the reverse is not true.

Thus, the only difference between these axioms and those stated in Section I is that here the axioms only involve events $\mathcal{E}$ that are in the sigma algebra $\mathcal{F}$. Because any event $\mathcal{E}$ in $\mathcal{F}$ also has its complement $\mathcal{E}^c$ in $\mathcal{F}$, it is easy to see that the complement of any event in $\mathcal{F}$ has a well defined probability that is equal to 1 minus the probability of the original event. That is, for any event $\mathcal{E}$ in $\mathcal{F}$, we have that $\mathcal{E}^c$ is also in $\mathcal{F}$, and:

$$P[\mathcal{E}^c] = 1 - P[\mathcal{E}]$$

The same consequences discussed in Section I also apply here, namely, that $P[\phi] = 0$, that $P[\mathcal{E}] \leq 1$ for any event $\mathcal{E}$ in $\mathcal{F}$, and that $P[\mathcal{E}_1] \leq P[\mathcal{E}_2]$ whenever $\mathcal{E}_1$ and $\mathcal{E}_2$ are events in $\mathcal{F}$ such that $\mathcal{E}_1 \subseteq \mathcal{E}_2$.

It follows that a probabilistic system can be repressed by a *triple* $(\mathcal{S}, \mathcal{F}, P)$, where $\mathcal{S}$ is a given (non-empty) sample space, $\mathcal{F}$ is a sigma algebra of events on that sample space, and "P" is shorthand representation of the probability measure $P[\mathcal{E}]$ defined over all events $\mathcal{E}$ that are in $\mathcal{F}$.

D. *Typical Sigma Algebras*

For finite or countably infinite sample spaces $\mathcal{S}$, the set $\mathcal{F}$ is typically defined as the set of all subsets of $\mathcal{S}$. For sample spaces consisting of the real numbers or intervals of the real numbers, the set $\mathcal{F}$ is typically defined as the Borel sigma algebra (extended in the obvious way outside of the unit interval). Sample spaces consisting of vectors of real numbers are treated with sigma algebras that are multi-dimensional versions of the Borel sigma algebra.

Random Variables are useful because they map outcomes of a sample space to real numbers, so that probabilities on this sample space can be defined in terms of the Borel sigma algebra. Specifically, a random variable $X(\omega)$ is a mapping that assigns a real number to each outcome ω of the sample space $\mathcal{S}$. Then we can talk about the probability of events of the form $\{\omega | X(\omega) \leq x\}$, where x is a given real number, by using a probability measure defined over the interval $(-\infty, x]$, an interval that falls within the Borel sigma algebra. This probability value, viewed as a function of x, is called the *cumulative distribution function for* X, and is usually denoted $F_X(x)$.

Random variables are usually written without the "(ω)" notation, so that $X(\omega)$ is written more simply as X, with the understanding that X is a real number with a value that depends on the outcome ω. While ω is simply an element of the sample space $\mathcal{S}$, if one intuitively views ω as a random outcome of a probability experiment, then X inherits this randomness and can be intuitively viewed as a random real number whose value depends on the random outcome ω. Now let Y be a new random variable defined by a continuous function of the random variable X. Specifically, let $Y(\omega) = g(X(\omega))$, where $g(\cdot)$ is a continuous function from the real numbers to the real numbers. For a given real number y, the event $\{\omega | Y(\omega) \leq y\}$ is equivalent to the event $\{\omega | X(\omega) \in g^{-1}((-\infty, y])\}$, where $g^{-1}((-\infty, y])$ is the *inverse image* of the set $(-\infty, y]$ through the function $g(\cdot)$, being all real numbers x such that $g(x) \in (-\infty, y]$. It can be shown that the inverse image of a set in the Borel sigma algebra through a continuous function is again in the Borel sigma algebra, and so probabilities for events specified by the random variable Y are also well defined.

E. *An algebra that is not a sigma algebra*

Define the sample space as the natural numbers:

$$\mathcal{S} = \mathbb{N} = \{1, 2, 3, ...\}$$

Define $\mathcal{F}$ as the collection of all subsets $A \subseteq \mathbb{N}$ such that either A or A^c is finite. It can be shown that $\mathcal{F}$ is an algebra (exercise). However, it is not a sigma algebra because the set $\{evens\} = \{2, 4, 6, 8, ...\}$ not an element of $\mathcal{F}$, but it is a countable union of sets in $\mathcal{F}$

$$\{evens\} = \{2\} \cup \{4\} \cup \{6\} \cup \{8\} \cup ...$$

On this algebra, it is easy to define $P : \mathcal{F} \rightarrow \mathbb{R}$ by $P[A] = 0$ if A is a finite set, and $P[A] = 1$ if A is an infinite set. In this case it is easy to see that P satisfies Axioms 1-2 of probability. It also satisfies the finite additivity statement of Axiom 3. However, it does not satisfy countable additivity because

$$\mathbb{N} = \cup_{i=1}^{\infty} \{i\}$$

but

$$P[\mathbb{N}] = 1 \neq 0 = \sum_{i=1}^{\infty} P[\{i\}] = 0$$

This function P is not a valid probability measure because it is not defined on a sigma algebra, and it does not satisfy countable additivity.

REFERENCES

[1] P. Billingsley. *Probability and Measure, 3rd edition*. New York: John Wiley & Sons, 1995.