PROBABILISTIC MODELS AND PROBABILITY MEASURES

Contents

1. Probabilistic experiments
2. Sample space
3. Discrete probability spaces
4. σ-fields (σ-algebras)
5. Probability measures
6. Continuity of probabilities
7. Monotone class theorem

1 PROBABILISTIC EXPERIMENTS

Probability theory is a mathematical framework that allows us to reason about phenomena or experiments whose outcome is uncertain. A probabilistic model is a mathematical model of a probabilistic experiment that satisfies certain mathematical properties (the axioms of probability theory), and which allows us to calculate probabilities and to reason about the likely outcomes of the experiment.

A probabilistic model is defined formally by a triple $(\Omega, \mathcal{F}, \mathbb{P})$, called a **probability space**, comprised of the following three elements:

(a) Ω is the **sample space**, the set of possible outcomes of the experiment.

(b) $\mathcal{F}$ is a σ-**field**, a collection of subsets of Ω. (The term "σ-algebra" is also commonly used, as a synonym.)

(c) $\mathbb{P}$ is a **probability measure**, a function that assigns a nonnegative probability to every set in the σ-field $\mathcal{F}$.

Our objective is to describe the three elements of a probability space, and explore some of their properties.

2 SAMPLE SPACE

The sample space is a set Ω comprised of all the possible outcomes of the experiment. Typical elements of Ω are often denoted by ω, and are called **elementary outcomes**, or simply outcomes. The sample space can be finite, e.g., $\Omega = \{\omega_1, \ldots, \omega_n\}$, countable, e.g., $\Omega = \mathbb{N}$, or uncountable, e.g., $\Omega = \mathbb{R}$ or $\Omega = \{0,1\}^\infty$.

As a practical matter, the elements of Ω must be mutually exclusive and collectively exhaustive, in the sense that once the experiment is carried out, there is exactly one element of Ω that occurs.

Examples

(a) If the experiment consists of a single roll of an ordinary die, the natural sample space is the set $\Omega = \{1, 2, \ldots, 6\}$, consisting of 6 elements. The outcome $\omega = 2$ indicates that the result of the roll was 2.

(b) If the experiment consists of five consecutive rolls of an ordinary die, the natural sample space is the set $\Omega = \{1, 2, \ldots, 6\}^5$. The element $\omega = (3, 1, 1, 2, 5)$ is an example of a possible outcome.

(c) If the experiment consists of an infinite number of consecutive rolls of an ordinary die, the natural sample space is the set $\Omega = \{1, 2, \ldots, 6\}^\infty$. In this case, an elementary outcome is an infinite sequence, e.g., $\omega = (3, 1, 1, 5, \ldots)$. Such a sample space would be appropriate if we intend to roll a die indefinitely and we are interested in studying, say, the number of rolls until a 4 is obtained for the first time.

(d) If the experiment consists of measuring the velocity of a vehicle with infinite precision, a natural sample space is the set $\mathbb{R}$ of real numbers.

Note that there is no discussion of probabilities so far. The set Ω simply specifies the possible outcomes.

3 DISCRETE PROBABILITY SPACES

Before continuing with the discussion of σ-fields and probability measures in their full generality, it is helpful to consider the simpler case where the sample space Ω is finite or countable.

> **Definition 1.** *A **discrete probability space** is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ such that:*
>
> (a) *The sample space Ω is finite or countable: $\Omega = \{\omega_1, \omega_2, \ldots\}$.*
>
> (b) *The σ-field $\mathcal{F}$ is the set of all subsets of Ω.*
>
> (c) *The probability measure assigns a number in the set $[0, 1]$ to every subset of Ω. It is defined in terms of the probabilities $\mathbb{P}(\{\omega\})$ of the elementary outcomes, and satisfies*
>
> $$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}), \tag{1}$$
>
> *for every $A \subset \Omega$, and*
>
> $$\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1.$$

For simplicity, we will usually employ the notation $\mathbb{P}(\omega)$ instead of $\mathbb{P}(\{\omega\})$, and we will often denote $\mathbb{P}(\omega_i)$ by p_i.

The following are some examples of discrete probability spaces. Note that typically we do not provide an explicit expression for $\mathbb{P}(A)$ for every $A \subset \Omega$. It suffices to specify the probability of elementary outcomes, from which $\mathbb{P}(A)$ is readily obtained for any A.

Examples.

(a) Consider a single toss of a coin. If we believe that heads (H) and tails (T) are equally likely, the following is an appropriate model. We set $\Omega = \{\omega_1, \omega_2\}$, where $\omega_1 = H$ and $\omega_2 = T$, and let $p_1 = p_2 = 1/2$. Here, $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$, and $\mathbb{P}(\emptyset) = 0, \mathbb{P}(H) = \mathbb{P}(T) = 1/2, \mathbb{P}(\{H, T\}) = 1$.

(b) Consider a single roll of a die. if we believe that all six outcomes are equally likely, the following is an appropriate model. We set $\Omega = \{1, 2, \ldots, 6\}$ and $p_1 = \cdots = p_6 = 1/6$.

(c) This example is not necessarily motivated by a meaningful experiment, yet it is a legitimate discrete probability space. Let $\Omega = \{1, 2, 5, a, v, aaa, *\}$, and $\mathbb{P}(1) = .1$, $\mathbb{P}(2) = .1, \mathbb{P}(5) = .3, \mathbb{P}(a) = .15, \mathbb{P}(v) = .15, \mathbb{P}(aaa) = .2, \mathbb{P}(*) = 0$.

(d) Let $\Omega = \mathbb{N}$, and $p_k = (1/2)^k$, for $k = 1, 2, \ldots$. More generally, given a parameter $p \in [0, 1)$, we can define $p_k = (1 - p)p^{k-1}$, for $k = 1, 2, \ldots$. This results in a legitimate probability space because $\sum_{k=1}^{\infty} (1 - p)p^{k-1} = 1$.

(e) Let $\Omega = \mathbb{N}$. We fix a parameter $\lambda > 0$, and let $p_k = e^{-\lambda}\lambda^k/k!$, for $k = 0, 1, \ldots$. This results in a legitimate probability space because $\sum_{k=0}^{\infty} e^{-\lambda}\lambda^k/k! = 1$.

(f) We toss an unbiased coin n times. We let $\Omega = \{0, 1\}^n$, and if we believe that all sequences of heads and tails are equally likely, we let $\mathbb{P}(\omega) = 1/2^n$ for every $\omega \in \Omega$.

3

(g) We roll a die n times. We let $\Omega = \{1, 2, \ldots, 6\}^n$, and if we believe that all elementary outcomes (6-long sequences) are equally likely, we let $\mathbb{P}(\omega) = 1/6^n$ for every $\omega \in \Omega$.

Given the probabilities p_i, the problem of determining $\mathbb{P}(A)$ for some subset of Ω is conceptually straightforward. However, the calculations involved in determining the value of the sum $\sum_{\omega \in A} \mathbb{P}(\omega)$ can range from straightforward to daunting. Various methods that can simplify such calculations will be explored in future lectures.

4 σ-FIELDS

When the sample space Ω is uncountable, the idea of defining the probability of a general subset of Ω in terms of the probabilities of elementary outcomes runs into difficulties. Suppose, for example, that the experiment consists of drawing a number from the interval $[0, 1]$, and that we wish to model a situation where all elementary outcomes are "equally likely." If we were to assign a probability of zero to every ω, this alone would not be of much help in determining the probability of a subset such as $[1/2, 3/4]$. If we were to assign the same positive value to every ω, we would obtain $\mathbb{P}(\{1, 1/2, 1/3, \ldots\}) = \infty$, which is undesirable. A way out of this difficulty is to work directly with the probabilities of more general subsets of Ω (not just subsets consisting of a single element).

Ideally, we would like to specify the probability $\mathbb{P}(A)$ of every subset of Ω. However, if we wish our probabilities to have certain intuitive mathematical properties, we run into some insurmountable mathematical difficulties. A solution is provided by the following compromise: assign probabilities to only a partial collection of subsets of Ω. The sets in this collection are to be thought of as the "nice" subsets of Ω, or, alternatively, as the subsets of Ω of interest. Mathematically, we will require this collection to be a σ-field, a term that we define next.

Definition 2. *Given a sample space Ω, a σ-**field** is a collection $\mathcal{F}$ of subsets of Ω, with the following properties:*

(a) $\emptyset \in \mathcal{F}$.

(b) *If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.*

(c) *If $A_i \in \mathcal{F}$ for every $i \in \mathbb{N}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

A set A that belongs to $\mathcal{F}$ is called an **event**, *an* $\mathcal{F}$-**measurable** *set, or simply a* **measurable** *set. The pair $(\Omega, \mathcal{F})$ is called a* **measurable space**.

Remark. *A σ-field is often called a σ-**algebra**, and these terms will be used interchangeably. If we relax condition (c) and require only finite unions to be in $\mathcal{F}$, we get a definition of field (or algebra) of sets – see Def. 4 below.*

The term "event" is to be understood as follows. Once the experiment is concluded, the realized outcome ω either belongs to A, in which case we say that the event A has occurred, or it doesn't, in which case we say that the event did not occur.

It turns out that if $A_i \in \mathcal{F}$ for every $i \in \mathbb{N}$, then $\cap_{i=1}^{n} A_i \in \mathcal{F}$, i.e., a σ-field is closed under countable intersections as well.

Exercise 1.

(a) Let $\mathcal{F}$ be a σ-field. Prove that if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$. More generally, given a countably infinite sequence of events $A_i \in \mathcal{F}$, prove that $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$. *Hint:* Use De Morgan's law.

(b) Prove that property (a) of σ-fields (that is, $\emptyset \in \mathcal{F}$) can be derived from properties (b) and (c), assuming that the σ-field $\mathcal{F}$ is non-empty.

The following are some examples of σ-fields. (Check that this is indeed the case.)

Examples.

(a) The trivial σ-field, $\mathcal{F} = \{\emptyset, \Omega\}$.

(b) The collection $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$, where A is a fixed subset of .

(c) The set of all subsets of Ω: $\mathcal{F} = 2 = \{A \mid A \subset \Omega\}$.

(d) Let $\Omega = \{1, 2, \ldots, 6\}^n$, the sample space associated with n rolls of a die. Let $A = \{\omega = (\omega_1, \ldots \omega_n) \mid \omega_1 \leq 2\}$, $B = \{\omega = (\omega_1, \ldots, \omega_n) \mid 3 \leq \omega_1 \leq 4\}$, and $C = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega_1 \geq 5\}$, and $\mathcal{F} = \{\emptyset, A, B, C, A\cup B, A\cup C, B\cup C, \Omega\}$.

Example (d) above can be thought of as follows. We start with a number of subsets of Ω that we wish to have included in a σ-field (the sets A, B, and C, in this example). We then include more subsets, as needed, until a σ-field is constructed. More generally, given a collection of subsets of Ω, we can contemplate forming complements, countable unions, and countable intersections of these subsets, to form a new collection. We continue this process until no more sets are included in the collection, at which point we obtain a σ-field. This process is hard to formalize in a rigorous manner. An alternative way of defining this σ-field is provided below. We will need the following fact.

Proposition 1. *Let S be an index set (possibly infinite, or even uncountable), and suppose that for every s we have a σ-field $\mathcal{F}_s$ of subsets of the same sample space. Let $\mathcal{F} = \cap_{s \in S} \mathcal{F}_s$, i.e., a set A belongs to $\mathcal{F}$ if and only if $A \in \mathcal{F}_s$ for every $s \in S$. Then $\mathcal{F}$ is a σ-field.*

Proof. We need to verify that $\mathcal{F}$ has the three required properties. Since each $\mathcal{F}_s$ is a σ-field, we have $\varnothing \in \mathcal{F}_s$, for every s, which implies that $\varnothing \in \mathcal{F}$. To establish the second property, suppose that $A \in \mathcal{F}$. Then, $A \in \mathcal{F}_s$, for every s. Since each $\mathcal{F}_s$ is a σ-field, we have $A^c \in \mathcal{F}_s$, for every s. Therefore, $A^c \in \mathcal{F}$, as desired. Finally, to establish the third property, consider a sequence $\{A_i\}$ of elements of $\mathcal{F}$. In particular, for a given $s \in S$, every set A_i belongs to $\mathcal{F}_s$. Since $\mathcal{F}_s$ is a σ-field, it follows that $\cup_{i=1}^{\infty} A_i \in \mathcal{F}_s$. Since this is true for every $s \in S$, it follows that $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$. This verifies the third property and establishes that $\mathcal{F}$ is indeed a σ-field. $\qquad\square$

Suppose now that we start with a collection $\mathcal{C}$ of subsets of Ω, which is not necessarily a σ-field. We wish to form a σ-field that contains $\mathcal{C}$. This is always possible, a simple choice being to just let $\mathcal{F} = 2^\Omega$. However, for technical reasons, we may wish the σ-field to contain no more sets than necessary. This leads us to define $\mathcal{F}$ as the intersection of all σ-fields that contain $\mathcal{C}$. Note that if $\mathcal{H}$ is any other σ-field that contains $\mathcal{C}$, then $\mathcal{F} \subset \mathcal{H}$. (This is because $\mathcal{F}$ was defined as the intersection of various σ-fields, one of which is $\mathcal{H}$.) In this sense, $\mathcal{F}$ is the **smallest** σ-field containing $\mathcal{C}$. The σ-field $\mathcal{F}$ constructed in this manner is called the σ-field **generated** by $\mathcal{C}$, and is often denoted by $\sigma(\mathcal{C})$.

Example. Let $\Omega = [0, 1]$. The smallest σ-field that includes every interval $[a, b] \subset [0, 1]$ is hard to describe explicitly (it includes fairly complicated sets), but is still well-defined, by the above discussion. It is called the **Borel** σ-field, and is denoted by $\mathcal{B}$. A set $A \subset [0, 1]$ that belongs to this σ-field is called a **Borel set**.[1]

---

[1]The Borel σ-field is usually defined as the σ-field generated by the collection of open sets of

4.1 Other reasons for "small" σ-fields.

As we discussed earlier, one reason for using a σ-field which does not include all subsets of Ω is in order to avoid insurmountable mathematical difficulties. However, there is also another reason: we may want to capture the perspective of an observer who receives only partial information about the outcome of the experiment. In that case, it is convenient (loosely speaking) to let $\mathcal{F}$ be just the set of events for which the observer will be able to tell whether they occurred or not.

With this perspective, a σ-field can be viewed as an abstract description of the information that an observer receives. In particular, if the information available to observers 1 and 2 is described by σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, and if $\mathcal{F}_2 \subset \mathcal{F}_1$, we have a situation in which observer 2 has less information.

Example. We flip a coin twice, and each flip results in Heads (H) or Tails (T). In this context, $\Omega = \{HH, HT, TH, TT\}$. The natural σ-field, $\mathcal{F}_1$, is the collection of all subsets of Ω. Consider now an observer who sees only the result of the first coin flip. In this case, we describe the information available to that observer in terms of the smaller σ-field

$$\mathcal{F}_2 = \Big\{ \varnothing, \ \Omega, \ \{HH, HT\}, \ \{TH, TT\} \Big\}.$$

In particular, this observer can tell whether the event $\{HH, HT\}$ has occurred or not, but cannot tell whether the event $\{HH\}$ has occurred.

We will turn to this association of σ-fields to observers much later, when we consider conditional expectations given partial information.

5 PROBABILITY MEASURES

We are now ready to discuss the assignment of probabilities to events. We have already seen that when the sample space Ω is countable, this can be accomplished by assigning probabilities to individual elements $\omega \in \Omega$. However, as discussed before, this does not work when Ω is uncountable. We are then led to assign probabilities to certain subsets of Ω, specifically to the elements of a σ-field $\mathcal{F}$, and require that these probabilities have certain "natural" properties.

Besides probability measures, it is also convenient to define the notion of a measure more generally.

We will be using the following terminology. We say that a collection of sets $A_\alpha \subset \Omega$, where α ranges over some index set is **mutually exclusive** or that the sets are **disjoint** if $A_\alpha \cap A_{\alpha'} = \varnothing$, whenever $\alpha \neq \alpha'$. Also, the sets $A_\alpha \subset \Omega$ are called **collectively exhaustive** if $\cup_\alpha A_\alpha = \Omega$.

a topological space. But for the case of the unit interval our definition is an equivalent one.

> **Definition 3.** *Let $(\Omega, \mathcal{F})$ be a measurable space. A **measure** is a function $\mu : \mathcal{F} \to [0, \infty]$, which assigns a nonnegative extended real number $\mu(A)$ to every set A in $\mathcal{F}$, and which satisfies the following two conditions:*
>
> *(a)* $\mu(\emptyset) = 0$;
>
> *(b)* *(**Countable additivity, or σ-additivity**) If $\{A_i\}$ is a sequence of disjoint sets that belong to $\mathcal{F}$, then $\mu(\cup_i A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.*
>
> *A **probability measure** is a measure $\mathbb{P}$ with the additional property $\mathbb{P}(\Omega) = 1$. In that case, the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space.***

In short, a measure is a nonnegative extended real valued σ-additive set function with domain $\mathcal{F}$.

For any $A \in \mathcal{F}$, $\mathbb{P}(A)$ is called the probability of the event A. The assignment of unit probability to the event Ω expresses our certainty that the outcome of the experiment, no matter what it is, will be an element of Ω. Similarly, the outcome cannot be an element of the empty set; thus, the empty set cannot occur and is assigned zero probability. If an event $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say that A occurs **almost surely**. Note, however, that A happening almost surely is not the same as the condition $A = \Omega$. For a trivial example, let $\Omega = \{1, 2, 3\}$, $p_1 = .5$, $p_2 = .5$, $p_3 = 0$. Then the event $A = \{1, 2\}$ occurs almost surely, since $\mathbb{P}(A) = .5 + .5 = 1$, but $A \neq \Omega$. The outcome 3 has zero probability, but is still possible. We will study more interesting examples of almost sure events later on when we give examples of non-discrete probability spaces.

The countable additivity property is very important. Its intuitive meaning is the following. If we have several events $A_1, A_2, \ldots$, out of which at most one can occur, then the probability that "one of them will occur" is equal to the sum of their individual probabilities. In this sense, probabilities (and more generally, measures) behave like the familiar notions of area or volume: the area or volume of a countable union of disjoint sets is the sum of their individual areas or volumes. Indeed, a measure is to be understood as some generalized notion of a volume. In this light, allowing the measure $\mu(A)$ of a set to be infinite is natural, since one can easily think of sets with infinite volume.

The properties of probability measures that are required by Definition 3 are often called the axioms of probability theory. Starting from these axioms, many other properties can be derived, as in the next proposition.

Proposition 2. *Probability measures have the following properties.*

(a) **(Finite additivity)** *If the events $A_1, \ldots, A_n$ are disjoint, then $\mathbb{P}(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i)$.*

(b) For any event A, we have $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(c) If the events A and B satisfy $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

(d) **(Union bound)** *For any sequence $\{A_i\}$ of events, we have*

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

(e) **(Inclusion-exclusion formula)** *For any collection of events $A_1, \ldots, A_n$,*

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i \Big) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{(i,j): \, i<j} \mathbb{P}(A_i \cap A_j)$$
$$+ \sum_{(i,j,k): \, i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Proof.

(a) This property is almost identical to condition (b) in the definition of a measure, except that it deals with a finite instead of a countably infinite collection of events. Given a finite collection of disjoint events $A_1, \ldots, A_n$, let us define $A_k = \emptyset$ for $k > n$, to obtain an infinite sequence of disjoint events. Then,

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

Countable additivity was used in the second equality, and the fact $\mathbb{P}(\emptyset) = 0$ was used in the last equality.

(b) The events A and A^c are disjoint. Using part (a), we have $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$. But $A \cup A^c = \Omega$, whose measure is equal to one, and the result follows.

(c) The events A and $B \setminus A$ are disjoint. Also, $A \cup (B \setminus A) = B$. Therefore, using also part (a), we obtain $\mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(B)$.

(d) Left as an exercise.

9

(e) Left as an exercise; a simple proof will be provided later, using random variables. □

For the special case where $n = 2$, part (e) of Proposition 2 simplifies to

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

Let us note that all properties (a), (c), and (d) in Proposition 2 are also valid for general measures (the proof is the same). Let us also note that for a probability measure, the property $\mathbb{P}(\emptyset) = 0$ need not be assumed, but can be derived from the other properties. Indeed, consider a sequence of sets A_i, each of which is equal to the empty set. These sets are disjoint, since $\emptyset \cap \emptyset = \emptyset$. Applying the countable additivity property, we obtain $\sum_{i=1}^{\infty} \mathbb{P}(\emptyset) = \mathbb{P}(\emptyset) \leq \mathbb{P}(\Omega) = 1$, which can only hold if $\mathbb{P}(\emptyset) = 0$.

Finite Additivity

Our definitions of σ-fields and of probability measures involve countable unions and a countable additivity property. A different mathematical structure is obtained if we replace countable unions and sums by finite ones. This leads us to the following definitions.

Definition 4. *Let Ω be a sample space.*

*(a) A **field** is a collection $\mathcal{F}_0$ of subsets of Ω, with the following properties:*

(i) $\emptyset \in \mathcal{F}$.
(ii) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
(iii) If $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$.

*(b) Let $\mathcal{F}_0$ be a field of subsets of Ω. A function $\mathbb{P} : \mathcal{F}_0 \to [0, 1]$ is said to be **finitely additive** if*

$$A, B \in \mathcal{F}_0, \ A \cap B = \emptyset \quad \Rightarrow \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Remark. *A field (of sets) is often called an **algebra** (of sets), and these terms will be used interchangeably.*

We note that finite additivity, for the two case of two events, easily implies finite additivity for a general finite number n of events, namely, the property in

part (a) of Proposition 2. To see this, note that finite additivity for $n = 2$ allows us to write, for the case of three disjoint events,

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3),$$

and we can proceed inductively to generalize to the case of n events.

Finite additivity is strictly weaker than the countable additivity property of probability measures. In particular, finite additivity on a field, or even for the special case of a σ-field, does not, in general, imply countable additivity. The reason for introducing the stronger countable additivity property is that without it, we are severely limited in the types of probability calculations that are possible. On the other hand, finite additivity is often easier to verify.

6 CONTINUITY OF PROBABILITIES

Consider a probability space in which $\Omega = \mathbb{R}$. The sequence of events $A_n = [1, n]$ converges to the event $A = [1, \infty)$, and it is reasonable to expect that the probability of $[1, n]$ converges to the probability of $[1, \infty)$. Such a property is established in greater generality in the result that follows. This result provides us with a few alternative versions of such a continuity property, together with a converse which states that finite additivity together with continuity implies countable additivity. This last result is a useful tool that often simplifies the verification of the countable additivity property.

Theorem 1. (σ-additivity $\iff$ continuity) *Let $\mathcal{F}$ be a field of subsets of , and suppose that $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfies $\mathbb{P}(\Omega) = 1$ as well as the finite additivity property. Then, the following are equivalent:*

(a) *$\mathbb{P}$ is σ-additive on $\mathcal{F}$. In other words, if $\{A_j\}_{j=1}^{\infty}$–disjoint, $A_j \in \mathcal{F}$ and $A = \cup_{j=1}^{\infty} A_j \in \mathcal{F}$ then $\mathbb{P}(A) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$.*

(b) *If $\{A_i\}$ is an increasing sequence of sets in $\mathcal{F}$ (i.e., $A_i \subset A_{i+1}$, for all i), and $A = \cup_{i=1}^{\infty} A_i$ belongs to $\mathcal{F}$, then $\lim_{i \to \infty} \mathbb{P}(A_i) = \mathbb{P}(A)$.*

(c) *If $\{A_i\}$ is a decreasing sequence of sets in $\mathcal{F}$ (i.e., $A_i \supset A_{i+1}$, for all i), and $A = \cap_{i=1}^{\infty} A_i$ belongs to $\mathcal{F}$, then $\lim_{i \to \infty} \mathbb{P}(A_i) = \mathbb{P}(A)$.*

(d) *If $\{A_i\}$ is a decreasing sequence of sets in $\mathcal{F}$ (i.e., $A_i \supset A_{i+1}$, for all i) and $\cap_{i=1}^{\infty} A_i$ is empty, then $\lim_{i \to \infty} \mathbb{P}(A_i) = 0$.*

Notes:

- If $\mathcal{F}$ is also a σ-algebra then A in (a), (b) and (c) is automatically in $\mathcal{F}$.

- Theorem extends to general (non-probability) measures provided $\mathbb{P}(\Omega) < \infty$.

- **Notation:** If $\{A_i\}$ is a decreasing sequence of sets (i.e., $A_i \supset A_{i+1}$, for all i) and $\cap_{i=1}^{\infty} A_i = A$, we write $A_i \downarrow A$. Thus, in part (d) above, we are assuming that $A_i \downarrow \emptyset$.

Proof. We first assume that (a) holds and establish (b). Observe that $A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \ldots$, and that the events $A_1, (A_2 \setminus A_1), (A_3 \setminus A_2), \ldots$ are disjoint (check this). Therefore, using countable additivity,

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \sum_{i=2}^{\infty} \mathbb{P}(A_i \setminus A_{i-1})$$

$$= \mathbb{P}(A_1) + \lim_{n \to \infty} \sum_{i=2}^{n} \mathbb{P}(A_i \setminus A_{i-1})$$

$$= \mathbb{P}(A_1) + \lim_{n \to \infty} \sum_{i=2}^{n} \big(\mathbb{P}(A_i) - \mathbb{P}(A_{i-1})\big)$$

$$= \mathbb{P}(A_1) + \lim_{n \to \infty} \big(\mathbb{P}(A_n) - \mathbb{P}(A_1)\big)$$

$$= \lim_{n \to \infty} \mathbb{P}(A_n).$$

Suppose now that property (b) holds, let A_i be a decreasing sequence of sets, and let $A = \cap_{i=1}^{\infty} A_i \in \mathcal{F}$. Then, the sequence A_i^c is increasing, and De Morgan's law, together with property (b) imply that

$$\mathbb{P}(A^c) = \mathbb{P}\Big(\big(\cap_{i=1}^{\infty} A_i\big)^c\Big) = \mathbb{P}\big(\cup_{i=1}^{\infty} A_i^c\big) = \lim_{n \to \infty} \mathbb{P}(A_i^c),$$

and

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \lim_{n \to \infty} \mathbb{P}(A_i^c) = \lim_{n \to \infty} \big(1 - \mathbb{P}(A_i^c)\big) = \lim_{n \to \infty} \mathbb{P}(A_i).$$

Property (d) follows from property (c), because (d) is just the special case of (c) in which the set A is empty.

To complete the proof, we now assume that property (d) holds and establish that property (a) holds as well. Let $B_i \in \mathcal{F}$ be disjoint events. Let $A_n = \cup_{i=n}^{\infty} B_i$. Note that $\{A_n\}$ is a decreasing sequence of events. We claim that $\cap_{n=1}^{\infty} A_n = \emptyset$. Indeed, if $\omega \in A_1$, then $\omega \in B_n$ for some n, which implies that $\omega \notin \cup_{i=n+1}^{\infty} B_i = A_{n+1}$. Therefore, no element of A_1 can belong to all of the sets A_n, which means that that the intersection of the sets A_n is empty. Property (d) then implies that $\lim_{n \to \infty} \mathbb{P}(A_n) = 0$.

12

Applying finite additivity to the n disjoint sets $B_1, B_2, \ldots, B_{n-1}, \cup_{i=n}^{\infty} B_i$, we have

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i \right) = \sum_{i=1}^{n-1} \mathbb{P}(B_i) + \mathbb{P}\left(\bigcup_{i=n}^{\infty} B_i \right).$$

This equality holds for any n, and we can take the limit as $n \to \infty$. The first term on the right-hand side converges to $\sum_{i=1}^{\infty} \mathbb{P}(B_i)$. The second term is $\mathbb{P}(A_n)$, and as observed before, converges to zero. We conclude that

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i \right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i),$$

and property (a) holds. $\qquad\square$

6.1 Discrete probability spaces revisited

In Section 3, we defined $\mathbb{P}(A)$ for every $A \subset \Omega$ in terms of the probabilities of individual outcomes. We actually need to verify that this formula results in probabilities that satisfy countable additivity.

To this effect, we can use Theorem 1. We only need to verify (i) finite additivity and (ii) the continuity property in part (d).

Regarding finite additivity, it suffices to consider the case of two sets; the general case is obtained by induction on the number of sets. Suppose that the sets $A = \{\omega_1, \omega_2, \ldots\}$ and $B = \{\omega_1', \omega_2', \ldots\}$ are disjoint. Let $a_i = \mathbb{P}(\omega_i)$ and $b = \mathbb{P}(\omega_i')$. We then have $A \cup B = \{\omega_1, \omega_1', \omega_2, \omega_2', \ldots\}$ and

$$\mathbb{P}(A \cup B) = a_1 + b_1 + a_2 + b_2 + \cdots = \sum_{i=1}^{\infty} (a_i + b_i) = \sum_{i=1}^{\infty} a_i + \sum_{i=1}^{\infty} b_i = \mathbb{P}(A) + \mathbb{P}(B).$$

The second and third equalities above are elementary properties of infinite series involving nonnegative numbers (more generally of absolutely convergent infinite series); namely, the order of summation or the grouping of the summands does not matter.

Regarding continuity, we need to show that

$$A_n \downarrow \varnothing \quad \Rightarrow \quad \mathbb{P}(A_n) \to 0.$$

Indeed, without loss of generality, we may assume $\Omega = \{1, 2, \ldots\}$ is the set of natural numbers (to be denoted in this course by either $\mathbb{N}$ or $\mathbb{Z}_+$). Fix some $\epsilon > 0$. Since $\sum_{i=1}^{\infty} \mathbb{P}(i) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$ is a convergent series, it follows that there exists some $m \in \mathbb{N}$ for which

$$\sum_{i \geq m} \mathbb{P}(i) \leq \epsilon.$$

13

On the other hand, since $A_n \downarrow \emptyset$, it follows that for every i, there exists some n_i such that $i \notin A_n$, for $n \geq n_i$. By using this property for $i = 1, \ldots, m-1$, we see that

$$A_n \subseteq \{m, m+1, \ldots\},$$

when n is large enough. Thus, for all large enough n,

$$\mathbb{P}(A_n) \leq \mathbb{P}\big(\{m, m+1, \ldots\}\big) = \sum_{i \geq m} \mathbb{P}(i) \leq \epsilon \,.$$

It follows that $\lim_{n \to \infty} \mathbb{P}(A_n) \leq \epsilon$. Since ϵ can be an arbitrarily small positive number, we conclude that $\lim_{n \to \infty} \mathbb{P}(A_n) = 0$.

7 MONOTONE CLASS THEOREM

We will soon find that one often needs to prove that a certain collection of sets is a σ-algebra. Such verifications are facilitated by the following theorem.

Definition 5. *A collection of sets $\mathcal{M}$ is a **monotone class** if all increasing and decreasing sequences of sets from $\mathcal{M}$ have limits belonging to $\mathcal{M}$. Formally, let $A_n \in \mathcal{M}$ for all n*

$$A_n \nearrow A \quad \Rightarrow \quad A \in \mathcal{M}$$
$$A_n \searrow A \quad \Rightarrow \quad A \in \mathcal{M}.$$

The minimal monotone class containing a collection $\mathcal{C}$ is denoted $\mu(\mathcal{C})$.

Note that $\mu(\mathcal{C})$ is well-defined by an analog of Proposition 1 for intersections of monotone classes.

Theorem 2. *If $\mathcal{A}$ is an algebra (field) of sets, then*

$$\mu(\mathcal{A}) = \sigma(\mathcal{A}) \,.$$

Proof. First, note that any σ-algebra is necessarily a monotone class. Thus

$$\mu(\mathcal{A}) \subseteq \sigma(\mathcal{A}).$$

Second, any collection $\mathcal{F}$ of sets which is simultaneously a monotone class and

an algebra is a σ-algebra. To see this, suppose $A_k \in \mathcal{F}$ and notice that

$$\bigcup_{k=1}^{\infty} A_k = \lim_{n \to \infty} B_n, \qquad B_n \triangleq \bigcup_{k=1}^{n} A_k.$$

Then, $B_n \in \mathcal{F}$ because $\mathcal{F}$ is an algebra and $\lim_n B_n \in \mathcal{F}$ because $\mathcal{F}$ is a monotone class.

It remains to prove that $\mathcal{M} \triangleq \mu(\mathcal{A})$ is an algebra. To that end, define another collection of sets

$$\mathcal{L}_1 \triangleq \{A \in \mathcal{M} \mid A^c \in \mathcal{M}\}.$$

Note that clearly $\mathcal{L}_1$ contains $\mathcal{A}$. Furthermore, for any increasing sequence $E_n \nearrow E$ of subsets of $\mathcal{L}_1$ we have $E \in \mathcal{L}_1$ since

$$(E_n)^c \searrow E^c$$

and $\mathcal{M}$ is a monotone class. Similarly, $\mathcal{L}_1$ is closed under decreasing limits. Thus $\mathcal{L}_1$ is a monotone class. By minimality of $\mathcal{M} = \mu(\mathcal{A})$ we conclude

$$\mathcal{L}_1 = \mathcal{M},$$

and hence $\mathcal{M}$ is closed under taking complements.

Proceeding in the same way, fix a set $B \in \mathcal{A}$ and define

$$\mathcal{L}_B \triangleq \{A \in \mathcal{M} \mid A \cap B \in \mathcal{M}\}.$$

Clearly, $\mathcal{L}_B$ contains $\mathcal{A}$ and is a monotone class (since $A_n \nearrow A \Rightarrow A_n \cap B \nearrow A \cap B$). Again, $\mathcal{L}_B = \mathcal{M}$. Hence, as B was arbitrary, $\mathcal{M}$ is closed under taking intersections and (by taking complements) unions with sets from $\mathcal{A}$.

Finally, let

$$\mathcal{L}_2 \triangleq \{A \in \mathcal{M} \mid A \cap M \in \mathcal{M} \text{ and } A \cup M \in \mathcal{M} \text{ for all sets } M \in \mathcal{M}\}.$$

As we have shown above, $\mathcal{L}_2$ contains $\mathcal{A}$. If $A_n \in \mathcal{L}_2$ and $A_n \nearrow A$, then for any M we have

$$A_n \cap M \in \mathcal{M}$$

by the definition of $\mathcal{L}_2$ and, on the other hand,

$$A_n \cap M \nearrow A \cap M \in \mathcal{M}$$

since $\mathcal{M}$ is a monotone class. Thus $A \cap M \in \mathcal{M}$. Applying this argument to $(A_n^c \cap M^c) \searrow A^c \cap M^c$ and noticing that $\mathcal{M}$ is closed under complements we obtain $A \cup M \in \mathcal{M}$. Hence $A \in \mathcal{L}_2$ and $\mathcal{L}_2$ is a monotone class implying

$$\mathcal{L}_2 = \mathcal{M}$$

or that $\mathcal{M}$ is an algebra. $\qquad\square$

Remark (Caution: real analysis). *The importance of the monotone class theorem is that it allows one to avoid the use of transfinite induction when proving properties of σ-algebras. However, if you understand transfinite induction many of the tricky constructions involving monotone classes become much less mysterious. For example, constructing $\mu(\mathcal{A})$ involves taking $\mathcal{A}$, then adding all the limits of increasing and decreasing sets (thus forming new sets "tier 2"), then adding the limits of increasing and decreasing sets in tier 2 (forming "tier 3"), etc. Transfinite induction gives a rigorous sense to the definition, "let $\mu(\mathcal{A})$ be the first tier at which this procedure stabilizes". Intuitively, then, $\mu(\mathcal{A})$ is closed under the operation of taking limits. Now if E is a set in any tier then $E \cap A$ is also a set in the same tier (assuming $A \in \mathcal{A}$). Consequently, $\mu(\mathcal{A})$ is automatically closed under intersections with sets from $\mathcal{A}$. Similarly, one may replace $A \in \mathcal{A}$ with any A in tier 2, 3, etc – eventually proving $\mu(\mathcal{A})$ is closed under intersections.*

References

(a) E. Cinlar, Chapter I, Sections 1-4.

(b) Grimmett and Stirzaker, Chapter 1, Sections 1.1-1.3.

(c) Williams, Chapter 1, Sections 1.0-1.5, 1.9-1.10.

(d) Florescu, Tudor: Chapter 1, Sections 2.1–2.3.5.

6.436J / 15.085J Fundamentals of Probability
Fall 2018