7. Measure Theory

In this section we discuss probability spaces from a more advanced point of view. The section on Measure Theory in the chapter on Foundations is an essential prerequisite.

Sigma-algebras and Measurability

As usual, suppose that we have a random experiment with sample space S. It is sometimes impossible to include *all* subsets of S as events. Our ultimate goal is to assign probabilities to events in a random experiment. This cannot be done arbitrarily; the probabilities must be mathematically consistent in the sense of the Kolmogorov axioms. Roughly speaking, the more events that we include in the mathematical model of our random experiment, the harder it is to assign probabilities in a consistent way. However, we naturally want our collection of events to be **closed** under the set operations in a certain sense. Technically, the collection of events $\mathscr{S}$ is required to be a σ-algebra.

Formally, a **positive measure** μ on S is a nonnegative function defined on the σ-algebra $\mathscr{S}$ that satisfies the **countable additivity axiom**: If $\{A_i : i \in I\}$ is a countable, pairwise disjoint collection of sets in $\mathscr{S}$ then

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$$

Thus, a probability measure $\mathbb{P}$ on S is a positive measure on S with the additional requirement that $\mathbb{P}(S) = 1$. For a general measure μ, it's possible, of course, that $\mu(A) = \infty$ for some $A \in \mathscr{S}$. On the other hand, if $0 < \mu(S) < \infty$, then μ can be re-scaled into a probability measure..

Formally then, a **probability space** $(S, \mathscr{S}, \mathbb{P})$, the basic mathematical model of a random experiment, consists of three essential parts:

1. the sample space S
2. the σ-algebra of events $\mathscr{S}$
3. the probability measure $\mathbb{P}$.

Moreoever, σ-algebras are not just important for theoretical and foundational purposes, but are important for practical purposes as well. A σ-algebra can be used to specify **partial information** about an experiment--a concept of fundamental importance in probability, statistics, and especially random processes. Specifically, suppose that $\mathscr{A}$ is a collection of events in the experiment, and that we know whether or not A occurred for each $A \in \mathscr{A}$. Then in fact, we can determine whether or not A occurred for each $A \in \sigma(\mathscr{A})$, the σ-algebra generated by $\mathscr{A}$.

Suppose that X is a random variable for the experiment, taking values in a set T. Almost always, T will have a natural σ-algebra of admissable subsets $\mathscr{T}$. Technically, X is required to be measurable as a function from S into T. This ensures that $\{X \in B\}$ is a valid event (that is, a member of the σ-algebra $\mathscr{S}$) for each $B \in \mathscr{T}$. Therefore, the probability distribution of X, that is the mapping $B \mapsto \mathbb{P}(X \in B)$, really is a probability measure on the on the σ-algebra $\mathscr{T}$.

Also, $\{\{X \in B\} : B \in \mathscr{T}\}$ is a sub σ-algebra of $\mathscr{S}$, and in fact is the σ-algebra generated by X, denoted $\sigma(X)$. If we observe the value of X, then we know whether or not each event in $\sigma(X)$ has occurred. More generally, suppose X_i is a random variable for each i in an index set I (the random variables might take values in different spaces). If we observe the value of X_i for each $i \in I$ then we know whether or not each event in $\sigma(\{X_i : i \in I\})$ has occurred. This idea is very important in the study of random processes; see the chapter on Markov Chains for an example.

Null and Almost Sure Events

1. Show that the following collection of null and almost sure events (**essentially deterministic** events) forms a sub σ-algebra.

$$\mathscr{D} = \{A \in \mathscr{S} : (\mathbb{P}(A) = 0 \text{ or } \mathbb{P}(A) = 1)\}$$

Hint: Boole's inequality will be helpful.

Tail Events

Let $(X_1, X_2, \ldots)$ be a sequence of random variables for a random experiment. The **tail sigma algebra** of the sequence is

$$\mathscr{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots)$$

and an event $B \in \mathscr{T}$ is a **tail event** for the sequence. Thus, a tail event is an event that can be defined in terms of $\{X_n, X_{n+1}, \ldots\}$ for each $n \in \mathbb{N}_+$. The tail sigma algebra for a sequence of events $(A_1, A_2, \ldots)$ is defined analogously (let $X_k = \mathbf{1}(A_k)$, the indicator variable of A_k. for each k). The limit of a sequence of events that is either increasing or decreasing is a tail event of the sequence. More generally, the limit inferior and superior of a sequence of events are tail events of the sequence, and the event that a sequence of real-valued random variables converges is a tail event of the sequence. The concepts are studied in the section on Convergence.

2. Suppose that $(A_1, A_2, \ldots)$ is a sequence of events.

a. Show that if the events are increasing then $\bigcup_{n=1}^{\infty} A_n$ is a tail event of the sequence.
b. Show that if the events are decreasing then $\bigcap_{n=1}^{\infty} A_n$ is a tail event of the sequence.

3. Show that $\liminf_{n \to \infty} A_n$ and $\limsup_{n \to \infty} A_n$ are tail events for a sequence of events $(A_1, A_2, \ldots)$

4. Show that the event $\{X_n \text{ converges as } n \to \infty\}$ is a tail event for a sequence of real-valued random variables $(X_1, X_2, \ldots)$

The following exercise gives the **Kolmogorov zero-one law**, named for Andrey Kolmogorov. It states that the tail σ-algebra of an indpendent sequence is a sub σ-algebra of the σ-algebra of essentially deterministic events.

5. Suppose that B is a tail event for a sequence of independent random variables $(X_1, X_2, \ldots)$ Show that $\mathbb{P}(B) = 0$ or $\mathbb{P}(B) = 1$.

a. Argue that for each n, $\{X_1, X_2, \ldots, X_n, \mathbf{1}(B)\}$ is an independent set of random variables.
b. From (a) argue that $\{X_1, X_2, \ldots, \mathbf{1}(B)\}$ is an independent set of random variables.
c. From (b) argue that the event B is independent of itself.

d. From (c) show that $\mathbb{P}(B) = 0$ or $\mathbb{P}(B) = 1$.

From Exercise 3 and Exercise 5, note that if $(A_1, A_2, \ldots)$ is a sequence of independent events, then $\limsup_{n\to\infty} A_n$ must have probability 0 or 1. The second Borel-Cantelli lemma gives a condition under which the probability is in fact 1.

Uniqueness and Extension of Probability Measures

In some cases, it is impossible to define a probability measure $\mathbb{P}$ on a σ-algebra $\mathscr{S}$ explicitly, by giving a "formula" for computing $\mathbb{P}(A)$ for each $A \in \mathscr{S}$. Rather, we usually know how the probability measure $\mathbb{P}$ should work on some class of events $\mathscr{B}$. We would then like to know that $\mathbb{P}$ can be **extended** to a probability measure on the σ-algebra generated by $\mathscr{B}$, and that this extension is **unique**.

We will now give a basic existence and uniqueness theorem. For a proof, see for example the book, Probability and Measure. Recall first that an algebra $\mathscr{A}$ of subsets of S is a collection of subsets that contains S and is closed under complements and finite unions (and hence also finite intersections). A **probability measure** $\mathbb{P}$ on $\mathscr{A}$ is a nonnegative function with $\mathbb{P}(S) = 1$ that satisfies countable additivity axiom whenever the countable union happens to be in $\mathscr{A}$. Thus, $\mathbb{P}$ is finitely additive and *partially* countably additive. The basic **extension and uniqueness theorem** states that a probability measure on an algebra $\mathscr{A}$ can be uniquely extended to a probability measure on $\sigma(\mathscr{A})$.

Next, a collection $\mathscr{B}$ of subsets of S is a **π-system** if $\mathscr{B}$ is closed under finite intersections: if $B \in \mathscr{B}$ and $C \in \mathscr{B}$ then $B \cap C \in \mathscr{B}$ The basic **uniqueness theorem** states that if $\mathbb{P}_1$ and $\mathbb{P}_2$ are probability measures on $\mathscr{S}$ and $\mathbb{P}_1(B) = \mathbb{P}_2(B)$ for all $B \in \mathscr{B}$ where $\mathscr{B}$ is a π-system with $\sigma\left(\mathscr{B}\right) = \mathscr{S}$ then $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ for any $A \in \mathscr{S}$.

The Real Numbers

For example, the standard (Borel) σ-algebra on $\mathbb{R}$ is generated by the collection of all open intervals of finite length, which is clearly closed under intersection. Thus, a probability measure $\mathbb{P}$ on $\mathbb{R}$ is completely determined by its values on the finite open intervals. In addition, the σ-algebra on $\mathbb{R}$ is generated by the collection of closed, infinite intervals of the form $(-\infty, x]$. Thus, a probability measure on $\mathbb{R}$ is completely determined by its values on these intervals. This is important in the study of distribution functions.

Finite Product Sets

Next, suppose that we have a sequence of n sets $(S_1, S_2, \ldots, S_n)$, with σ-algebras $(\mathscr{S}_1, \mathscr{S}_2, \ldots, \mathscr{S}_n)$, respectively. Recall that the product set

$$T = S_1 \times S_2 \times \cdots \times S_n$$

is a natural sample space for an experiment that consists of multiple measurements, or for a compound experiment that consists of performing n basic experiments in sequence. Usually, we give T the σ-algebra $\mathscr{T}$ generated by the collection of product sets of the form

$$A = A_1 \times A_2 \times \cdots \times A_n \quad \text{where} \quad A_i \in \mathscr{S}_i \quad \text{for each} \quad i \in \{1, 2, \ldots, n\}$$

This collection of product sets is closed under intersection, and hence a probability measure on T is completely determined by its values on these product sets. An important special case occurs when $S_i = S$ and $\mathscr{S}_i = \mathscr{S}$ for each $i \in \{1, 2, \ldots, n\}$. In this case, T is the natural sample space for the experiment that consists of n repetitions of a basic experiment.

Infinite Product Sets

Generalizing, suppose that we have an infinite sequence of sets $(S_1, S_2, \ldots)$ with σ-algebras $(\mathscr{S}_1, \mathscr{S}_2, \ldots)$ respectively. The product set

$$T = S_1 \times S_2 \times \cdots$$

is a natural sample space for an experiment that consists of infinitely many measurements, or for a compound experiment that consists of combining an infinite sequence of basic experiments. Usually, we give T the σ-algebra $\mathscr{T}$ generated by the collection of **cylinder sets** of the form

$$A = A_1 \times A_2 \times \cdots \times A_n \times S_{n+1} \times S_{n+2} \times \cdots \quad \text{where} \quad n \in \mathbb{N}_+ \quad \text{and} \quad A_i \in \mathscr{S}_i \quad \text{for each} \quad i \in \{1, 2, \ldots, n\}$$

This collection of product sets is closed under intersection, and hence a probability measure on S is completely determined by its values on these product sets. Again, an important special case occurs when $S_i = S$ and $\mathscr{S}_i = \mathscr{S}$ for each $i \in \mathbb{N}_+$. In this case, T is the natural sample space for the experiment that consists of infinite repetitions of a basic experiment.
