Chapter 1

Probability Spaces and Independence

This chapter treats two topics at the foundations of probability theory—*probability spaces* and *independence*. A probability space, in the abstract, serves as a universal template for modelling random phenomena. It consists just of a measure space together with a positive, additive measure, called a *probability measure*, having a total mass equal to one; the points of the measure space represent the different possible outcomes of the phenomenon, and the probability measure assigns probabilities to sets of outcomes. Obviously, you can't get more basic than this in setting out to formulate a mathematical theory of probability. In fact, the only non-trivial structure imposed in the definition of probability space is additivity of the probability measure, and this, too, is a very natural axiom. Nevertheless, the concept of a probability space is an important first step. It sets probability theory into the context of measure and integration theory, whose powerful tools thus become available, and it provides a unified framework for probabilistic analysis.

Independence is the mathematical formulation of what it means to have different random experiments whose outcomes do not affect one another statistically. It is the concept that most distinguishes probability from other branches of analysis. Mathematically, independence is just statement about integrals of products of functions and is closely related to products of measure spaces, but, in this purely analytic form, it is not a concept of central importance in analysis. In probability, however, independence enters from the very beginning, even in the interpretation of what a probability means. The basic probability model underlying statistical analysis is an infinite sequence of independent trials of a random experiment, because this is the (idealized) setting for discussing limits of empirical frequencies of the occurence of an event. Indeed, the frequentist interpretation of probability identifies probabilities with limit empirical frequencies. To make sense of this, one need first a definition of independence, then a probability space supporting an infinite sequence of independent trials, and then a theorem, called a large number law, identifying probabilities with long run averages. The probability space in this case can be constructed as an infinite product of probability spaces; the strong law of large numbers then says that the probability of an event is the limiting frequency of the occurence of the event in any random sequence of trials. In effect the strong law, which is treated in Chapter 3, is a theorem about infinite product measure spaces.

The aims of Chapter 1 are to present the definition of probability space, some general principles for constructing probability spaces, and the definition of independence, and then to put them together in the construction of infinite products of probability spaces. It is assumed that the reader has seen some measure theory and understands Lebesgue measure and integration, but we recall the definitions of algebra and σ-algebra in section 1, and carefully state or redefine measure theoretic concepts as they are first introduced. In section 2, we define probability spaces and outlines in theory the typical manner in which they are constructed. In Section 3, we develop in detail some important models, mainly for coin-tossing experiments. In particular, we construct explicitly a model for an infinite number of independent tosses of a fair coin. This example is a springboard into the topic of independence and product spaces, which we treat in section 5, after an important measure-theoretic interlude on monotone class theorems. In section 5.2, we arrive finally at one of the most important and distinctive probability spaces of probability theory—a product space that models an infinite sequence of independent, identical random trials. Along the way, both in the construction of example probability spaces and in the treatment of independence, we shall begin to see the importance of σ-algebras. In ordinary analysis, the only σ-algebras that typically arise are the Borel sets or the Lebesgue sets. In probability, the σ-algebra concept has a more central role.

Two caveats are in order. First, in this chapter, we motivate and apply the probability space as a working tool to model random phenomena. This is perhaps misleading; in reality, probabilists prefer to model using random variables, which we take up in the next chapter. However, pedagogically it is an acceptable fiction. Probability spaces and probability space techniques permeate the modern theory; in particular, random variables are measurable functions defined on a probability space. Second, while we have tried to spice up the exposition with concrete examples, chapter 1 is really a chapter on measure theory. You won't see much actual computation or approximation of probabilities! Please be patient! We'll get to the interesting stuff. But we first need to lay a good, measure-theoretic foundation.

1. Algebras and σ-algebras of collections of sets (Review)

Let Ω be a set and let $\mathcal{F}$ be a non-empty collection of subsets of Ω. We shall review standard terminology for describing closure properties of $\mathcal{F}$ under the set theoretic operations of union, intersection, and complementation.

Definition. $\mathcal{F}$ is called an algebra if

i) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$;   (A^c:=complement of A in Ω)
ii) $A, B \in \mathcal{F}$ implies $A \cup B \in \mathcal{F}$.

The following simple Lemma is left as an exercise.

Lemma 1.1. If $\mathcal{F}$ is an algebra, then

i) $\mathcal{F}$ contains Ω and $\emptyset$;
ii) $\mathcal{F}$ is closed under finite unions *and* finite intersections. (*Closure* under finite unions means that if $A_1, \ldots, A_n \in \mathcal{F}$ then $\cup_{i=1}^{n} A_i \in \mathcal{F}$ also; closure under finite intersections is defined similarly.)

Definition. The collection $\mathcal{F}$ is called a σ-algebra if $\mathcal{F}$ is an algebra and if $\mathcal{F}$ is closed under countable unions. Thus, if $\mathcal{F}$ is a σ-algebra and if $A_1, A_2, \ldots$ is a countably infinite sequence of sets in $\mathcal{F}$, then $\cup_1^\infty A_i \in \mathcal{F}$.

The following basic facts are also left as an exercise.

Lemma 1.2.

i) An algebra $\mathcal{F}$ is a σ-algebra if and only if it is closed under countable intersections.

ii) Let I be an index set and let $\{\mathcal{F}_\alpha\ ,\ \alpha \in I\}$ be a collection of σ-algebras of subsets of Ω indexed by I. Then $\cap_{\alpha \in I} \mathcal{F}_\alpha$ is also a σ-algebra.

In practice, σ-algebras are constructed indirectly from simpler collections of subsets of Ω, using the following general observation.

Lemma 1.3. Let $\mathcal{S}$ be a collection of subsets of Ω. Then there exists a smallest σ-algebra containing $\mathcal{S}$ and denoted $\sigma(\mathcal{S})$. That is, any σ-algebra which contains $\mathcal{S}$ contains $\sigma(\mathcal{S})$ and $\sigma(\mathcal{S})$ is itself a σ-algebra containing $\mathcal{S}$.

Proof: Let J denote the collection of all σ-algebras that contain $\mathcal{S}$. The collection 2^Ω of all subsets of Ω is certainly a σ-algebra containing $\mathcal{S}$, and so J is non-empty. By Lemma 1.2,

$$\bigcap_{\mathcal{G} \in J} \mathcal{G}$$

is a σ-algebra, and it is easy to see that it is the smallest σ-algebra containing $\mathcal{S}$. $\diamond$

Lemma 1.3 is useful in that it allows one to define a σ-algebra without giving an explicit construction of its elements. The most important examples are Borel σ-algebras. Let U be a topological space with the collection $\mathcal{T}$ of open sets. Then the Borel σ-algebra of U, denoted $\mathcal{B}(U)$, is defined by $\mathcal{B}(U) := \sigma(\mathcal{T})$. The elements of $\mathcal{B}(U)$ are called Borel sets. In analysis and probability theory we often use the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ of n-dimensional Euclidean space.

Exercise. (a) Let $\mathcal{C}$ be the collection of all closed sets of U. Show that $\mathcal{B}(U) = \sigma(\mathcal{C})$.
(b) Let $\mathcal{R}$ denote the collection of all finite disjoint unions of intervals of the form $(a, b]$ ($a = -\infty$ and/or $b = \infty$, included). Show that $\mathcal{R}$ is an algebra and that $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{R})$.

2. Probability spaces

Here is a statement you would certainly believe: the probability of getting an even number in one roll of a die is 1/2. Here is another you could verify by a simple combinatorial argument: the probability of getting exactly two heads in 3 tosses of a penny is 3/8. We understand these statements intuitively, but what do they mean precisely? More generally, what does it mean to assign a probability to an uncertain event? Before beginning the mathematical theory, it is useful to make a few naive remarks about the interpretation of probability, as a guide to intuition. We avoid going deeply into the matter. It is a philosophical problem, that is to say, one that admits contending approaches and has no, single, scientifically verifiable resolution. Fortunately, simple, if imprecise, intuition is enough to motivate a mathematical framework for probability.

There are two main approaches to the interpretation of probability. The first is called the frequentist approach, and it applies most convincingly to random experiments which can, in principle, be repeated as often as desired under identical circumstances. A coin toss, die roll, or, more generally, any game of chance provide typical examples. Another example that is useful to think about is the motion due to random molecular collisions of a microscopic particle suspended in a fluid; this is called Brownian motion, after the botanist Robert Brown who observed the phenomenon for pollen grains in 1826. In each case, we can play the game of chance or observe the Brownian motion a large number of times and record the results. For any given event A, for example, the event of getting an even number in the die rolling example, we can then calculate the empirical frequency of A, that is, the frequency of occurence of A in our sequence of trials. As the number of trials becomes larger and larger, this empirical frequency appears always to settle down to some limit. We assume that this limit always exists and we identify it with the probability of A. Thus, for example, saying that the probability of an even roll is 1/2 means that in a long run of die rolls, approximately 1/2 are even and that the approximation ultimately improves as the number of trials increases. Notice that in the frequentist view, a probability is an objective, physical property of the random system under study.

The frequentist approach does not however apply to situations, such as the next presidential election, or tomorrow's horse race, that are unique and not repeatable. Yet the outcomes are still uncertain and people still assign them probabilities, if only to make book, whether on a horse race or an election. In such cases, people form judgements about the relative likelihood of different outcomes using knowledge of previous performance (whether of politicians or horses), polls, recent history, the alignment of the planets, or whatever else they deem relevant. Necessarily, these judgements will be subjective and vary from individual to individual; an expert in presidential politics will have a much more precise and informed assessment about the presidential election than a casual observer, but even experts will disagree among themselves. In these cases, probabilities are subjective assessments of relative likelihood, subject to change as new data are acquired. The Bayesian philosophy insists that this is the only philosophically consistent way to think of probablity in all cases, even those, such as games of chance, where the frequentist approach seems natural. The term *Bayesian* derives from Bayes rule, which is a formula for updating probabilities after new information is acquired. I believe that Mr. Bayes (1702-61), the inventor of this rule, was not himself a Bayesian; indeed, he predates the formulation of the Bayesian philosophy.

We come now to the definition of probability space. For motivation think about what, in the abstract, would constitute a full description of a random phenomenon. The first thing we need to do is identify all possible outcomes. We gather these into a set Ω, called the *outcome space*. For instance, $\{1, 2, 3, 4, 5, 6\}$ is the outcome space for one roll of a die; a list of all possible sequences of length 10 of the numbers 1 through 6 is the outcome space for the experiment of 10 die rolls. In more complicated examples we may want an outcome space that includes at least all physically possible outcomes and may contain more. Consider a Brownian motion over a one second period. When we observe it, we record a continuous path in space as a function of time. Thus, a suitable outcome space might be the set of all continuous, $\mathbb{R}^3$-valued functions on $[0,1]$. A priori, this space might be too large and include paths that a Brownian motion could not actually follow. However, we cannot decide which paths could be excluded until we have a mathematicial model for the Brownian motion. The space of continuous functions is thus a simple and convenient choice for the outcome space.

The second element in any full description of a random phenomena is a class of *events*. By *event* we mean just a subset of possible outcomes. In the simple die rolling example the subset $\{1\}$ is the event of rolling a 1, while the subset $\{2,4,6\}$ is the event of rolling an even number. In the abstract setting, we gather all the events, whose probabilities we can measure, in an ensemble $\mathcal{F}$.

The final and most important element is a measure $\mathbb{P}$ that assigns each event A in the ensemble $\mathcal{F}$ a number $\mathbb{P}(A)$ between 0 and 1 representing the *probability* that A occurs. $\mathbb{P}(A) = 1$ means that A is certain to occur, while $\mathbb{P}(A) = 0$ means that A is certain not to occur, and numbers in between measure the relative likelihood of events that may or may not occur.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$, consisting of outcome space, event ensemble, and probability assignment, provides the basic constituents of a probability space. To complete the definition, we impose on the triple some mimimal structure that is consistent with our intuitive notion of what probability means, irrespective of the particular random phenomenon under study. This structure may be summed up in one word: *additivity*. We introduce additivity in in two steps, the first being finite additivity.

Definition. $(\Omega, \mathcal{F}, \mathbb{P})$ is called a finitely additive probability space if $\mathcal{F}$ is an algebra and $\mathbb{P}$ is a positive measure satisfying $\mathbb{P}(\Omega) = 1$ and

$$\mathbb{P}\Big(\bigcup_1^n A_i\Big) = \sum_1^n \mathbb{P}(A_i) \quad \text{for disjoint } A_1, \ldots, A_n \text{ in } \mathcal{F}. \tag{1}$$

This definition is very reasonable. If $\mathbb{P}$ is to represent a relative likelihood, it should be additive; in fact, the frequentist interpretation demands this. The assumption that $\mathcal{F}$ is an algebra is also very natural; it says that if we can measure the probabilities of events A and B, then we can measure those of $A \cup B$, A^c, and B^c as well.

Here are some elementary properties of finitely additive probability spaces; the easy proofs are left as an exercise.

$$\text{if } A \in \mathcal{F}, \text{ then } \mathbb{P}(A^c) = 1 - \mathbb{P}(A), \tag{2}$$

$$0 \leq \mathbb{P}(A) \leq 1, \text{ and } \mathbb{P}(\emptyset) = 0 \tag{3}$$

$$\text{if } A, B \in \mathcal{F} \text{ and } A \subset B, \text{ then } P(A) \leq \mathbb{P}(B). \tag{4}$$

Example 2.1 *Formal probability space for a die roll.* Let $\Omega := \{1,2,3,4,5,6\}$, $\mathcal{F} := 2^\Omega$ (=the ensemble of all subsets of Ω), and let $\mathbb{P}(A) := |A|/6$, for every

$A \subset \Omega$, where $|A|$ is the cardinality of A. This is a probability model for one roll of a fair die. It is immediate from the definition that $I\!P$ is finitely additive. $\diamond$

Finitely additive probability measures are all that is needed so long as one works with an outcome space Ω of finite cardinality. However, they come up short when the outcome space is infinite. In particular, finite additivity is not strong enough to handle the important limit questions of probability that require countable operations on events. Therefore, it is useful strengthen finite additivity to countable additivity.

Definition. A *probability space* $(\Omega, \mathcal{F}, I\!P)$ is a triple consisting of a space Ω, a σ-algebra $\mathcal{F}$ of subsets of Ω, and a positive, countably additive measure $I\!P$ on $(\Omega, \mathcal{F})$ satisfying $I\!P(\Omega) = 1$. In this case, $I\!P$ is called a probability measure.

Recall that countable additivity means that if $A_1, A_2, \ldots$ is a countable sequence of disjoint sets in $\mathcal{F}$ then

$$I\!P(\bigcup_1^\infty A_i) = \sum_1^\infty I\!P(A_i).$$

Example 2.2 *Uniform probability measure on* $[a, b)$, $b > a$. Let $\Omega := [a, b)$, let $\mathcal{F}$ be the Borel subsets, $\mathcal{B}([a, b))$, of $[a, b)$, and let $I\!P(A) := \lambda(A)/(b - a)$, where $\lambda(A)$ denotes the Lebesgue measure of A, for a Borel set A. $I\!P$ is countable additive by construction and $I\!P([a, b)) = 1$; hence $([a, b), \mathcal{B}([a, b)), I\!P)$ is a probability space. Consider an experiment whose outcome is modelled by this space. Then $I\!P(A)$, the probability that the outcome falls in A, depends only on the length of A and is invariant under translation (modulo $b - a$) of A. For this reason we call it the uniform distribution on $b - a$.

Here is a simple random trial for which the uniform distribution is a reasonable model. Fix a vector in a planar surface and drop a pin onto the surface from a decent height. Let θ denote the angle between the pin, thought of as a vector pointing from the head of the pin to its tip, and the fixed vector. Since there should be no prefered direction for θ, the probability measure for θ should be invariant with respect to translations modulo 2π. Show that any such translation invariant, probability measure on the Borel sets of $0 \leq \theta < 2\pi$ is the uniform measure. $\diamond$

Notice that in the last example the probability of any single point is 0. But since the outcome of any trial is fact a single point, every outcome that we actually observe had zero probability! This is a somewhat annoying feature of uncountably infinite probability spaces. I think the best way to understand it is to realize that in such models the probability measure is really a convenient idealization. In the pin drop experiment we cannot actually measure θ to arbitrary accuracy. This has no physical meaning. Instead we can measure θ only to a certain decimal accuracy, say of 10^{-N}. Let $1 \leq z < 2\pi$ be a number represented by an N place decimal expansion. Then the event that, in a trial pin drop, we measure the angle z is really the event that the ideal angle falls in an interval of length 10^{-N} centered at

z, and this event has probability $10^{-N}/2\pi > 0$. In other words, the true probability distribution is really an atomic measure on the discrete set of points in $[0, 2\pi)$ with decimal expansions out to the 10^{-N}th place, and the uniform measure should be considered as a mathematically convenient approximation. The tools of calculus apply directly to the uniform measure and make it much easier to work with than the discrete measure.

Remark. Notice that examples 1.1 and 1.2 each present a probability space model of one random trial. Of themselves, they contain no reference to a frequentist interpretation of probability. The first major task of probability theory that we take up is to construct probability spaces modelling ideal, infinite sequences of identical trials of a random experiment and to verify the connection between probabilities and long-time statistical averages. We will show later in this chapter how the infinite trial probability spaces are constructed. The study of statistical averages on these spaces is part of the subject of large number laws, which is treated in chapter 3.

Countably additivity is not so intuitively obvious a primitive axiom as finite additivity. However, finite additivity, plus an appealing continuity property implies countable additivity. A finitely additive probability measure P on an algebra $\mathcal{R}$ is said to be continuous from above at $\emptyset$ if

(5) for any decreasing sequence $\{A_n\}_{n\geq 1} \subset \mathcal{R}$ with $\cap_n A_n = \emptyset, \quad P(A_n) \downarrow 0.$

Notice that any countably additive probability measure is necessarily continuous from above; this is a straightforward consequence of countable additivity that we leave to the reader.

We state the following basic result without proof. Although it is stated only for probability measures it is true for σ-finite, finitely additive measures. Recall that a measure $I\!\!P$ on $\sigma(\mathcal{R})$ *extends* P on $\mathcal{R}$ if $I\!\!P(A) = P(A)$ for every $A \in \mathcal{R}$.

Theorem 2.1. (Carathéodory's Extension Theorem) Let $I\!\!P$ be a finitely-additive probability measure on an algebra $\mathcal{R}$. Then P admits an extension to a probability measure $I\!\!P$ on $\sigma(\mathcal{R})$ if and only if P is continuous from above at $\emptyset$. The extension $I\!\!P$ is unique.

Carathéodory's Theorem will almost always allow one to extend a finitely additive probability measure to a probability measure. In fact, it is the tool that is generally applied to construct countably additive measure spaces, in particular, probability spaces. Typically, to define a probability space, we first construct a finitely additive probability measure on an algebra by an explicit formula; generally, it is easy to do this if we choose an algebra whose elements can be described simply. Then we check continuity from above at $\emptyset$ and invoke Carathéodory's theorem. This same procedure is at the heart of textbook constructions of Lebesgue measure. As an example in probability theory, we shall work out in detail, in section 3, the construction of a probability space for an infinite sequence of tosses of a fair coin.

The necessity of condition (5) in Carathéodory's theorem is a consequence of a more general continuity property of probability measures stated in Theorem 2.2 below. We shall derive the uniqueness of the extension in section 4, as a consequence of the Monotone Class theorem. For the (harder) proof of sufficiency, consult any standard text on measure theory, such as Halmos, *Measure Theory*, chapter 2, or Folland, *Real Analysis*, chapter 3.

Properties (2)-(3) stated for a finitely additive probability measure are of course still true in the countably additive case. In addition, countably additive probability measures satisfy the important continuity properties stated in the next result.

Proposition 2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

$\mathbb{P}$ is continuous from above; that is, if A_n is a decreasing sequence of measurable sets,

$$\mathbb{P}(\bigcap_1^\infty A_n) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Also, $\mathbb{P}$ is continuous from below; that is, if A_n is an increasing sequence of measurable sets,

$$\mathbb{P}(\bigcup_1^\infty A_n) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Proof: We prove continuity from above. Let $\{A_n\}$ be a decreasing sequence of events. and set $B = \cap_{n\geq 1} A_n$ For every n, let $V_n := A_n - A_{n+1}$. Then, the V_n are disjoint and are disjoint from B, and $A_n = B \cup [\cup_n^\infty V_i]$ for every n. By countable additivity,

$$\mathbb{P}(A_n) = \mathbb{P}(B) + \sum_n^\infty \mathbb{P}(V_i) \tag{6}$$

for each n. In particular,

$$\infty > \mathbb{P}(A_1) - \mathbb{P}(B) = \sum_1^\infty \mathbb{P}(V_i).$$

It follows that

$$\sum_n^\infty \mathbb{P}(V_i) \to 0 \qquad \text{as } n \to \infty,$$

Thus, from (6), $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(B)$, proving continuity from above. $\diamond$

Exercise. Prove the continuity from below in Proposition 2.2.

3. Probability spaces; elementary examples, coin tossing.

In this section we construct probability spaces, mostly for coin tossing examples, and discuss discrete spaces in general.

3.1 Discrete probability spaces

We call a probability space $(\Omega, \mathcal{F}, I\!\!P)$ discrete if Ω is finite or countably infinite. In this case the task of specifying a probability measure is relatively simple and nothing fancy is needed for satisfying countable additivity. Let $\Omega = \{\omega^i\ ,\ 1 \leq i < \infty\}$ be an enumeration of the sample space. The smallest σ-algebra which includes all the singleton sets is the power set 2^Ω of all subsets, since any subset is a countable union of singletons; thus we take 2^Ω as the σ-algebra $\mathcal{F}$ of events. A probability measure on $(\Omega, \mathcal{F})$, is then completely specified by the probabilities $p_i = I\!\!P(\{w^i\})$ of the individual elements of Ω, because if $E \subset \Omega$, countable additivity of $I\!\!P$ implies that

$$I\!\!P(E) = \sum_{i;\omega^i \in E} p_i.$$

The only restrictions on the p_i are that they be non-negative and that they sum to 1: $\sum_1^\infty p_i = 1$.

A particularly important case is the finite outcome space with *equally likely probabilities*; this means

$$I\!\!P(\{w^i\}) = \frac{1}{|\Omega|} \qquad \text{for each } i,$$

where $|\Omega|$ denotes the cardinality of Ω. In this case,

$$I\!\!P(A) = \frac{|A|}{|\Omega|} \qquad \text{for any } A \subset \Omega.$$

A basic coin tossing example is defined next.

Example 3.1. *Sequence of N tosses of a fair coin.* Imagine flipping a fair coin N times, denoting each occurence of a head by 1 and each of a tail by 0. The oucome space may then be represented as $\{0,1\}^N$, the set of all N-vectors of 0's and 1's. If the coin is fair, all sequences should have equally likely outcomes; hence we define our probability measure by requiring that $I\!\!P^{(N)}(\{\omega\}) = 2^{-N}$ for any element $\omega = (\omega_1, \ldots, \omega_N)$ of $\{0,1\}^N$, because the cardinality of $\{0,1\}^N$ is 2^N. Then, for any $E \subset \Omega$,

$$I\!\!P^{(N)}(E) = \sum_{\omega \in E} \frac{1}{2^N} = \frac{|E|}{2^N}.$$

For instance, if A is the event that exactly m heads (1's) occur in the sequence of N tosses,

$$I\!\!P^{(N)}(A) = \binom{N}{m} \frac{1}{2^N},$$

since the cardinality of A is the number of ways to choose m from N.

Let $M > N$. An event $A \subset \{0,1\}^N$ concerning N consecutive tosses may be lifted to an equivalent event in $\{0,1\}^M$, namely

$$\{\omega \in \{0,1\}^M \mid (\omega_1, \ldots, \omega_N) \in A\} = A \times \{0,1\}^{M-N}.$$

The family of probability measures $\{P^{(k)}\}_{k \geq 1}$ are consistent one to another in the sense that

$$\mathbb{P}^{(M)}\left(A \times \{0,1\}^{M-N}\right) = \mathbb{P}^{(N)}(A), \tag{1}$$

for M, N, and A as above. To check this, note that

$$\mathbb{P}^{(M)}\left(A \times \{0,1\}^{M-N}\right) = \frac{|A \times \{0,1\}^{M-n}|}{2^M} = \frac{|A| 2^{M-N}}{2^M} = \frac{|A|}{2^N}.$$

Example 3.2. In this example we build a countably infinite example on the idea of coin tossing. Consider a game in which a fair coin is flipped until ten heads appear. The sample space Ω for this experiment is the set of all finite sequences of 0's and 1's containing exactly ten 1's and ending with a 1, plus the set of all infinite sequences containing strictly less than ten 1's. If $\alpha = (\alpha_1, \ldots, \alpha_N)$ is a finite sequence in the sample space of length N, then, in accordance with Example 3.1, it makes sense to let $\mathbb{P}(\{\alpha\}) = 2^{-N}$. What about the probability of an infinite sequence in the sample space? Let E_N be the set of all sequences of length N having ten 1's and ending in a 1. Let $E = \cup_{10}^{\infty} E_N$, which is the event consisting of all finite length sequences for our game. Noting that the cardinality of E_N is the same as the cardinality of all choices of 9 from $N-1$, we have $P(E_N) = \binom{N-1}{9}\frac{1}{2^N}$. Hence

$$\begin{aligned}
\mathbb{P}(E) = \sum_{10}^{\infty} \mathbb{P}(E_N) &= \sum_{10}^{\infty} \binom{N-1}{9} \frac{1}{2^N} \\
&= \frac{1}{9! 2^{10}} \frac{d^9}{dt^9} \left(\frac{1}{1-t}\right)\Big|_{t=1/2} \\
&= 1.
\end{aligned}$$

Thus $\mathbb{P}(\text{an infinite sequence occurs}) = \mathbb{P}(E^c) = 0$; in other words, with probability 1 the game ends in a finite number of tosses. We may therefore simplify the probability space by replacing Ω by the subset E without any loss to the validity of the model.

Exercise. In this example, note that $E^c = \cap_{N \geq 1} F_N$, where F_N is the event that the game last longer than N plays. Show directly that $\lim_{N \to \infty} \mathbb{P}(F_N) = 0$ to give an alternate proof of $\mathbb{P}(E^c) = 0$.

3.2. Tossing a fair coin infinitely often.

The statistical behavior of the outcomes of a large number of repetitions of a random experiment is a major issue of probability theory. One object is to establish large limit laws, stating that the frequency of occurence of an event tends to its probability as the number of trial increases to infinity, so long as individual trials do not affect each other 'too much.' Such theorems validate the frequency interpretation of a probability. To discuss large number laws, it is useful to construct a probability space in which the outcome is an infinite sequence of trials. In the next example, we do this for infinite tosses of a fair coin. The construction is actually just a particular case of an infinite product of probability spaces, which we develop in greater generality in a later section in connection with independence.

The result of tossing a coin infinitely often may be recorded as an infinite sequences of 0's and 1's, where a 1 in the n^{th} spot signifies that the n^{th} toss came up heads. The appropriate outcome space is thus the product space $\{0,1\}^\infty$ of countably infinite sequences of 0's and 1's. Notice that $\{0,1\}^\infty$ has the cardinality of the real line and hence is uncountably infinite. We shall use ω to denote a point in $\{0,1\}^\infty$ and ω_n to denote the n^{th} element of ω; thus, $\omega = (\omega_1, \omega_2, \ldots)$. We wish to determine an appropriate σ-algebra , $\mathcal{B}$, of subsets of $\{0,1\}^\infty$, and a probability measure $\mathbb{P}$, for modelling a fair coin. As a starting point, it is thus reasonable to impose the two conditions:

(a) $\mathcal{B}$ contains any event depending only on the outcomes of a finite number of the tosses; and

(b) the probability that $\mathbb{P}$ assigns to any event depending only on the first N tosses, should be consistent with the probability assigned by the measure $P^{(N)}$ defined in Example 3.1.

We show that condition (b) determines a unique probability measure on the σ-algebra generated by $\mathcal{B}$.

An event $A \subset \{0,1\}^\infty$ depending only on the first N tosses is an event that can be written in the form

$$A = C_A^N \times \{0,1\}^\infty := \{\omega \ ; \ (\omega_1, \ldots, \omega_N) \in C_A^N\} \tag{2}$$

where C_A^N is a subset of $\{0,1\}^N$. Let $\dot{\mathcal{B}}$ denote the collection of all such subsets, where N ranges over the positive integers and C_A^N over the subsets of $\{0,1\}^N$. It is easy to check that $\dot{\mathcal{B}}$ is an algebra, but *not* a σ-algebra (exercise!). Let A be an element of $\dot{\mathcal{B}}$ as in (2). The appropriate probability to assign to A is

$$\dot{P}(A) = \dot{P}(C_A^N \times \{0,1\}^\infty) = \mathbb{P}^{(N)}(C_A^N). \tag{3}$$

Note that if $A \in \dot{\mathcal{B}}$ and N is given, then C_A^N is uniquely determined. However, it is possible to represent the same A in the form (2) using many different N. Luckily, the consistency among the measures $\{P(k)\}$ proved in (1) of Example 3.1, shows that the definition of $\dot{P}(A)$ in (3) does not depend on N and hence is meaningful.

To complete our discussion, we shall show that $\dot{P}$ is finitely additive on $\dot{\mathcal{B}}$ and that it is countably additive from above at $\emptyset$. Clearly, $\dot{P}(\{0,1\}^\infty) = 1$. By

Carathéodory's extension theorem, there is then a unique countably additive probability measure $I\!P$ on $\sigma(\dot{\mathcal{B}})$. Define $\mathcal{B} = \sigma(\dot{\mathcal{B}})$; then $(\{0,1\}^\infty, \mathcal{B}, I\!P)$ is our probability model for an infinite sequence of tosses of a fair coin.

To prove finite additivity of $\dot{P}$ let $A_1, \ldots, A_k$ be disjoint subsets in $\dot{\mathcal{B}}$. Then there is some N large enough so that each A_i, $1 \le i \le k$, may be written in the form $A_i = C_i^N \times \{0,1\}^\infty$. Necessarily, $C_{A_i}^N$, $1 \le i \le k$, are also disjoint. Thus, from finite additivity of $I\!P^{(N)}$,

$$\dot{P}\left(\bigcup_1^k A_i\right) = I\!P^{(N)}\left(\bigcup_1^k C_i^N\right) = \sum_1^k I\!P^{(N)}(C_i^N) = \sum_1^k \dot{P}(A_i).$$

To prove continuity of $\dot{P}$ from above at $\emptyset$, we show the contrapositive: if $\{A_n\}$ is a decreasing sequence of sets in $\dot{\mathcal{B}}$ and $\lim_{n\to\infty} P(A_n) > 0$, then $\cap_1^\infty A_n$ is nonempty. The basic identity that we use is

$$\dot{P}\left(\{i\}\times C\times\{0,1\}^\infty\right) = \frac{1}{2}\dot{P}\left(C\times\{0,1\}^\infty\right), \tag{4}$$

for $i = 0$ or $i = 1$, and $C \subset \{0,1\}^N$ for some finite N. This is a simple consequence of the definition of $\dot{P}$ and the easily checked identity $I\!P^{(N+1)}(\{i\}\times C) = \frac{1}{2} I\!P^{(N)}(C)$, $i = 0$ or $i = 1$.

Now assume $\{A_n\}$ is a decreasing and $\lim_{n\to\infty} P(A_n) > 0$. For $i = 0$ or $i = 1$, let $A_n(i) := \{\omega\ ;\ (i,\omega) \in A_n\}$, and observe that $\{A_n(i)\}_{n\ge1}$ is a decreasing sequence for each i. Since $A_n = [\{0\}\times A_N(0)] \cup [\{1\}\times A_N(1)]$, finite additivity and (4) imply

$$\dot{P}(A_n) = \frac{1}{2}\dot{P}(A_n(0)) + \frac{1}{2}\dot{P}(A_n(1)).$$

Hence, there must be an $i_1 \in \{0,1\}$ such that

$$\lim_{n\to\infty} P(A_n(i_1)) > 0;$$

otherwise $\lim_{n\to\infty} P(A_n) > 0$ could not hold. Now apply the preceding argument to $\{A_n(i_1)\}$ itself to obtain the existence of i_2 such that if

$$A_n(i_1, i_2) = \{\omega \mid (i_1, i_2, \omega) \in A_n\},$$

then, again,

$$\lim_{n\to\infty} P(A_n(i_1, i_2)) > 0.$$

Continuing in this manner, one constructs by induction a sequence $(i_1, i_2, \ldots)$ such that for each positive integer r,

$$\lim_{n\to\infty} P(A_n(i_1, \ldots, i_r)) > 0,$$

where $A_n(i_1, \ldots, i_r) = \{\omega \mid (i_1, \ldots, i_r, \omega) \in A_n\}$.

Clearly $A_n(i_1, \ldots, i_r)$ is non-empty for every n and r. But if $A_n = C_n^N \times \{0,1\}^\infty$, with $C_n^N \subset \{0,1\}^N$, then $A_N(i_1, \ldots, i_N)$ can be non-empty only if $(i_1, \ldots, i_N) \in C_n^N$. Hence, it follows that $(i_1, i_2, \ldots) \in A_n$. Since this is true for any n, $(i_1, i_2, \ldots) \in \cap_1^\infty A_n$, proving that the intersection is non-empty. $\diamond$

The probability space $(\{0,1\}^\infty, \mathcal{B}, I\!\!P)$ constructed in the last example is rich enough to state the strong law of large numbers for independent tosses of a fair coin. Of course, the strong law applies in much greater generality, but as a preview, and as an example in which we make essential use of the countably additive extension $I\!\!P$, we shall state it for the specialized case here.

Suppose we perform the (ideal) experiment of tossing our fair coin infinitely often and obtain the outcome $\omega \in \{0,1\}^\infty$. The frequency of heads (1's) in the first n tosses of ω is

$$S_n(\omega) = \frac{1}{n} \sum_1^n \omega_i, \tag{5}$$

and it is called the *empirical* probability of heads in N tosses, because it is the frequency that is experimentally observed. Our measure $I\!\!P$ will be consistent with the frequency interpretation of probability if $S_N(\omega)$ tends to $1/2$ as $n \to \infty$ with $I\!\!P$ probability 1. The strong law says that this is the case.

Theorem 3.1 (Strong law of large numbers for Example 3.3)

$$I\!\!P\big(\lim_{n\to\infty} \frac{1}{n} \sum_1^n \omega_i = 1/2\big) = 1.$$

Exercise. Show that the event that the average frequency converges to $1/2$ is in $\mathcal{B}$, but not in $\dot{\mathcal{B}}$.

4. The monotone class theorem

The monotone class theorem is a result from measure theory that finds frequent application to questions of measurabilty in probability theory. It gives conditions allowing one to conclude properties of a σ-algebra of sets from study of its generating algebra.

Definition: Let $\mathcal{G}$ be a non-empty collection of subsets of Ω. $\mathcal{G}$ is called a monotone class if it is closed under unions of increasing sequences and intersections of decreasing sequences; that is, if

(i) given $\{A_n\}_1^\infty \subset \mathcal{G}$ such that $A_n \subset A_{n+1}$ for every n, it follows that $\cup_1^\infty A_n \in \mathcal{G}$, and

(ii) given $\{A_n\}_1^\infty \subset \mathcal{G}$ such that $A_n \supset A_{n+1}$ for every n, it follows that $\cap_1^\infty A_n \in \mathcal{G}$.

Given a collection $\mathcal{S}$ of sets, let $M(\mathcal{S})$ denote the smallest monotone class containing $\mathcal{S}$. It is easily checked that $M(\mathcal{S})$ exists for any class $\mathcal{S}$, because arbitrary intersections of monotone classes are monotone classes. (Exercise)

Theorem 4.1. (Monotone Class Theorem). Let $\mathcal{R}$ be an algebra. Then $M(\mathcal{R}) = \sigma(\mathcal{R})$.

Proof: Since any σ-algebra is clearly a monotone class, it follows immediately that $\sigma(\mathcal{R})$ is contained in $M(\mathcal{R})$. Thus we need only prove the reverse inclusion.

First note that any monotone class $\mathcal{G}$ which is also an algebra is in fact a σ-algebra. Indeed, let $\mathcal{G}$ be both a monotone class and an algebra. Let $A_1, A_2, \ldots$ be a countable sequence of events in $\mathcal{G}$. Then, for every n, $B_n := \cup_1^n A_j$ is an element of $\mathcal{G}$, by the definition of algebra. Since $\{B_n\}$ is an increasing sequence, $\cup_1^\infty B_n$ is also in $\mathcal{G}$ by the monotone class property of $\mathcal{G}$. Hence it follows that $\cup_1^\infty A_n = \cup_1^\infty B_n$ is in $\mathcal{G}$. Thus $\mathcal{G}$ is closed under countable unions, making $\mathcal{G}$ a σ-algebra.

In order to complete the proof of the theorem, it will thus suffice to prove that $M(\mathcal{R})$ is an algebra. To this end, fix $A \in \mathcal{R}$ and consider the class of sets

$$\mathcal{C}_A := \{B \;\big|\; B \subset \Omega,\ A \cup B \in M(\mathcal{R})\ \}.$$

Clearly, $\mathcal{R} \subset \mathcal{C}_A$ since $\mathcal{R}$ is an algebra. Moreover, it is easy to check (do so!) that $\mathcal{C}_A$ is a monotone class. Hence, $M(\mathcal{R}) \subset \mathcal{C}_A$ by definition of $M(\mathcal{R})$ as minimal. It follows that

$$A \cup B \in M(\mathcal{R}) \quad \text{for every } A \in \mathcal{R} \text{ and every } B \in M(\mathcal{R}). \tag{1}$$

Now let $E \in M(\mathcal{R})$ and consider $\mathcal{C}_E$. Observe that $\mathcal{R} \subset \mathcal{C}_E$ because of (1), and that, again, $\mathcal{C}_E$ is a monotone class. Hence, $M(\mathcal{R}) \subset \mathcal{C}_E$, which says precisely that

$$E \cup B \in M(\mathcal{R}) \quad \text{for every } E \in M(\mathcal{R}) \text{ and every } B \in M(\mathcal{R}). \tag{2}$$

This proves that $M(\mathcal{R})$ is closed under finite unions. To show $M(\mathcal{R})$ is closed under complements we show that $\mathcal{C} := \{A \;\big|\; A \subset \Omega,\ A^c \in M(\mathcal{R})\ \}$ is a monotone class containing $\mathcal{R}$, and hence that $M(\mathcal{R}) \subset \mathcal{C}$. $\diamond$

As a first application of the monotone class theorem we shall prove the following proposition, which implies the uniqueness of the extension $I\!\!P$ in the statement of Carathéodory's extension theorem, stated as Theorem 2.1.

Proposition 4.2. Let $\mathcal{R}$ be an algebra, and let $I\!\!P$ and $I\!\!K$ be two countably additive measures on $\sigma(\mathcal{R})$. If $I\!\!P(A) = I\!\!K(A)$ for every $A \in \mathcal{R}$ then $I\!\!P = I\!\!K$.

Proof: Let $\mathcal{C} := \{\ A \;\big|\; A \in \sigma(\mathcal{R}),\ I\!\!P(A) = I\!\!K(A)\ \}$. Then $\mathcal{R} \subset \mathcal{C}$ by assumption. Moreover, $\mathcal{C}$ is a monotone class because $I\!\!P$ and $I\!\!K$ are countably additive. For example if $A_1, A_2, \ldots$ is a decreasing sequence of sets in $\mathcal{C}$,

$$I\!\!P(\bigcap_1^\infty A_n) = \lim_{n\to\infty} I\!\!P(A_n) = \lim_{n\to\infty} I\!\!K(a_n) = I\!\!K(\bigcap_1^\infty A_n),$$

and so $\mathcal{C}$ is closed under decreasing limits. Similarly, $\mathcal{C}$ is closed under increasing limits. Since $\mathcal{C}$ is a monotone class containing $\mathcal{R}$, $\sigma(\mathcal{R}) = M(\mathcal{R}) \subset \mathcal{C}$. Thus $I\!P(A) = I\!K(A)$ for every $A \in \sigma(\mathcal{R})$. $\diamond$

Here is another result, on a similar theme. It can also be proved using the monotone class theorem (or even proved directly), and it will occasionally be useful. We state it for probability measures, although it is true for more general measures. We leave the proof as an exercise.

Proposition 4.3. Let $I\!P$ be a probability measure on $\sigma(\mathcal{R})$ where $\mathcal{R}$ is an algebra. Let $\epsilon > 0$ be arbitrary. Then for every set $A \in \sigma(\mathcal{R})$ there is a set $A_0 \in \mathcal{R}$ such that $I\!P(A \triangle A_0) < \epsilon$. (Here, $A \triangle A_0$ denotes the symmetric difference $(A - A_0) \cup (A_0 - A)$.)

5. Independence and product spaces.

5.1. Independence The concept of independence is the key feature that distinguishes the study of probability spaces from general measure theory. This section presents the definition of independence, some of its basic properties, and its intimate connection to product spaces.

Definition. Events A and B in the probability space $(\Omega, \mathcal{F}, I\!P)$ are independent if

$$I\!P(A \cap B) = I\!P(A) I\!P(B). \tag{1}$$

The events $\{A_\alpha | \alpha \in \mathcal{I}\}$, where $\mathcal{I}$ is some index set, are (mutually) independent if

$$I\!P(A_{\alpha_1} \cap \cdots \cap A_{\alpha_n}) = I\!P(A_{\alpha_1}) \cdots I\!P(A_{\alpha_n}), \tag{2}$$

for any finite set $\{\alpha_1, \ldots, \alpha_n\} \subset \mathcal{I}$.

Exercise. Pairwise independence does not impy mutual independence! Construct an example of a probability space with three events A, B, and C which are pairwise but not mutually independent.

The terminology, "A and B are independent," conveys the idea that knowing B has occured does not change the probability we assign to A, nor does knowledge of A affect the probability of B. To explain this idea, assume $I\!P(B) > 0$ and consider the ratio

$$\frac{I\!P(A \cap B)}{I\!P(B)}, \tag{3}$$

If B is fixed, the assignment $A \to I\!P(A \cap B)/I\!P(B)$ defines a new probability measure supported on B, and, and so measures the probabilities of events given that B has occured. The ratio (3) is thus called the conditional probability of A

given B, and denoted $I\!P(A/B)$. If A and B are independent, we see easily from (1) that $I\!P(A/B) = I\!P(A)$; i.e., prior knowledge of B does not affect the probability assigned to A. (The use of conditional probabilities and their generalization is a major topic we shall treat later.)

In practice it is useful to extend the definition of independence between events to independence between families of events.

Definition. For each α in the index set $\mathcal{I}$, let $\mathcal{G}_\alpha$ be a family of events. The families $\{\mathcal{G}_\alpha | \alpha \in \mathcal{I}\}$ are said to be (mutually) independent if $A_{\alpha_1}, \ldots, A_{\alpha_n}$ are independent whenever $\{\alpha_1, \ldots, \alpha_n\}$ is a finite subset of $\mathcal{I}$ and $A_{\alpha_i} \in \mathcal{G}_{\alpha_i}$ for each i.

Consideration of independence between families of events is natural even at the most elementary level. For example, let A and B be independent events. Then

$$
\begin{aligned}
I\!P(A \cap B^c) &= I\!P(A) - I\!P(A \cap B) = I\!P(A)[1 - I\!P(B)] \\
&= I\!P(A) I\!P(B^c).
\end{aligned} \tag{1}
$$

Hence, A is independent of B^c, and, by interchanging the roles of A and B in (1), A^c is independent of B. Repeating (1) with A replaced by A^c, we find A^c and B^c are independent as well. The reader can easily check that $\emptyset$ and Ω are independent of all other events. Hence we have shown that the independence of A and B automatically entails that of the respective σ-algebras $\{\emptyset, \Omega, A, A^c\}$ and $\{\emptyset, \Omega, B, B^c\}$. In fact, a more general statement is true.

Lemma 5.1. Consider a family of events $\{A_\alpha | \alpha \in \mathcal{I}\}$, for some index set $\mathcal{I}$, and, for each α, let $\mathcal{A}_\alpha = \{\emptyset, \Omega, A_\alpha, A_\alpha^c\}$ be the smallest σ-algebra that contains A_α. If the sets A_α, $\alpha \in \mathcal{I}$ are independent, then the σ-algebras $\mathcal{A}_\alpha$, $\alpha \in \mathcal{I}$ are independent as well.

Proof: It suffices to consider the case when $\mathcal{I}$ is finite, since, by definition, the independence of the members of any family requires only the independence of any finite subset. We showed above that the proposition is true for $n = 2$. We leave the case $|\mathcal{I}| = n > 2$ as an exercise. $\diamond$

Example 5.1: Independence of coin tosses. In section 3 we constructed a probability space to model repeated trials of a fair coin toss, using the assumption that all outcomes of a fixed number of trials were equally likely. We shall show that, the outcomes of different trials are independent in this probability space. Therefore , let $(\{0,1\}^\infty, \mathcal{B}, P^\infty)$ denote the space constructed in Example 3.3. We have changed notation for the probability measure; P^∞ is identical to $I\!P$ of Example 3.3. For each k, let $\mathcal{A}_k$ be the σ-algebra of events describing all possible outcomes of toss k; thus $\mathcal{A}_k$ consists of all events of the form

$$\{\omega \in \{0,1\}^\infty \ : \omega_k \in B\}$$

where $B \subset \{0,1\}$. ($\mathcal{A}_k$ consist only of 4 events–the event that toss k comes up heads, the event it comes up tails, the empty set, and the whole probability space.)

By Lemma 5.1, it suffices to show that for any n and for any $A_1 \in \mathcal{A}_1$, …, $A_n \in \mathcal{A}_n$, the events $A_1, \ldots, A_n$ are mutually independent. It may be assumed that each A_i is non-empty; otherwise, condition (2) is automatically satisfied. Write each A_i as

$$A_i = \{\omega \in \{0,1\}^\infty \ : \omega_i \in B_i\}.$$

Then $P^\infty(A_{k_i}) = |B_i|/2$. Also

$$A_{k_1} \cap \cdots \cap A_{k_r} = \{\omega \in \{0,1\}^\infty \ : \omega_{k_i} \in B_i,\ 1 \le i \le r\}.$$

A simple calculation shows

$$P^\infty(A_{k_1} \cap \cdots \cap A_{k_r}) = \frac{\prod_1^r |B_i|}{2^r} = \prod_1^r P^\infty(A_{k_i}). \quad \diamond$$

which verifies (2) and hence the independence of $A_1, \ldots, A_n$.

The independence of coin tosses under the measure P^∞ is really just a consequence of the fact that P^∞ is a product measure. We elaborate on this in great generality later in the section.

Example 5.2 (Independent tosses of a biased coin) We will construct the probability space for N independent tosses of a coin, with sides labelled 1 and 0, such that, in each toss, $P(\{1\}) = p$ and $P(\{0\}) = 1 - p$. We may write $I\!P(\{i\}) = p^i(1-p)^{1-i}$. As in Example 3.1, the outcome space is $\{0,1\}^N$ and the σ-algebra is the collection of all subsets of $\{0,1\}^N$. Given an outcome $\omega = (\omega_1, \ldots, \omega_N)$, independence dictates that the appropriate probability measure $I\!P_p$ on $\{0,1\}^N$ is

$$I\!P(\{(\omega_1, \ldots, \omega_N)\}) = \prod_1^N P(\{\omega_i\}) = p^{(\omega_1 + \cdots + \omega_N)}(1-p)^{N-(\omega_1+\cdots+\omega_N)}.$$

Note that $\sum_1^N \omega_i$ is the number of 1's observed in the N tosses.

Exercise. Construct a probability space for infinite independent sequences of the toss of a biased coin.

In example 5.1, we have shown that the tosses are mutually independent. It is natural to suspect that any group of tosses is independent from any other group. More precisely, if $\mathcal{A}$ is the σ-algebra generated by all tosses in an index set I and $\mathcal{B}$ is the σ-algebra generated by all tosses in an index set J disjoint from I then $\mathcal{A}$ and $\mathcal{B}$ should be independent. This is true,but it requires a proof. Rather than work only within the context of the coin tossing example we state and prove a general theorem. Let $(\Omega, \mathcal{F}, I\!P)$ be any probability space, and let $\{\mathcal{F}_i,\ 1 \le i < \infty\}$ be a family of σ-algebras which are sub-σ-algebras of $\mathcal{F}$. If I and J are two disjoint subsets of the positive integers, define

$$\mathcal{F}^I := \sigma(\cup_{i \in I} \mathcal{F}_i) \qquad \mathcal{F}^J := \sigma(\cup_{j \in J} \mathcal{F}_j).$$

Proposition 5.2. If $\mathcal{F}_i$, $1 \leq i < \infty$ are mutually pbindependent and if I and J are disjoint, then $\mathcal{F}^J$ and $\mathcal{F}^J$ are independent as well.

The hypothesis of the Proposition implies that a generating collection of events of $\mathcal{F}^J$ is independent of a generating collection of $\mathcal{F}^I$. The problem is to pass from independence of these generating sets to independence of the σ-algebra . The monotone class theorem is the natural tool. We develop the proof in a series of preparatory results.

Lemma 5.3. Let A be an event and let

$$\mathcal{G} := \{B \in \mathcal{F};\ B \text{ is independent of } A\}.$$

Then $\mathcal{G}$ is a monotone class.

The proof is left as an exercise.

Lemma 5.4 Let $\mathcal{R}$ and $\mathcal{S}$ be independent algebras of events. Then the σ-algebras $\sigma(\mathcal{R})$ and $\sigma(\mathcal{S})$ are independent as well.

Proof: Let A be an event in $\mathcal{R}$. Then, by Lemma 5.3, the class $\mathcal{G}_A$ of events independent of A is a monotone class. Since $\mathcal{G}_A$ includes $\mathcal{S}$ by assumption, the Monotone Class Theorem says that it must also include $\sigma(\mathcal{S})$. Now take any B in $\sigma(\mathcal{S})$. The set $\mathcal{G}_B$ of events independent of B is a monotone class that includes $\mathcal{R}$ and so includes $\sigma(\mathcal{R})$. We have thus shown that any B in $\sigma(\mathcal{S})$ is independent of any A in $\sigma(\mathcal{R})$. $\diamond$

Lemma 5.5. Let I be any subset of the positive integers, and define $\dot{\mathcal{F}}^I$ to be the class of all finite, disjoint unions of sets of the form

$$\cap_{k=1}^N A_k, \qquad A_k \in \cup_{i \in I} \mathcal{F}_i, \quad 1 \leq k \leq N. \tag{2}$$

Then $\dot{\mathcal{F}}^I$ is an algebra and $\mathcal{F}^I = \sigma(\dot{\mathcal{F}})$.

We leave the proof as an exercise.

Proof of Proposition 5.2. Let I and J be disjoint. Let F be a finite intersection of sets in $\cup_{i \in I} \mathcal{F}_i$, as in (2), and let G be a finite intersection of sets in $\cup_{j \in J} \mathcal{F}_j$. The mutual independence of the σ-algebras $\mathcal{F}_k$, $1 \leq k < \infty$, implies that F and G are independent. Now let A be in $\dot{\mathcal{F}}^I$, so that $A = \cup_1^n F_k$, where each F_k is a finite intersection of sets in $\cup_{i \in I} \mathcal{F}_i$ and the F_k are disjoint, and let B be in $\dot{\mathcal{F}}^J$, so that $B = \cup_1^m G_l$, where each G_l is a finite intersection of sets in $\cup_{j \in J} \mathcal{F}_j$. Then, the sets $F_k \cap G_l$ are disjoint for different values of k and l, and so

$$\begin{aligned} P^\infty(A \cap B) &= \sum P^\infty(F_k \cap G_l) = \sum P^\infty(F_k) P^\infty(G_l) \\ &= \sum_k P^\infty(F_k) \sum_l P^\infty(G_l) = P^\infty(A) P^\infty(B). \end{aligned} \tag{3}$$

This calculation shows that $\dot{\mathcal{F}}^I$ and $\dot{\mathcal{F}}^J$ are independent. Lemma 5.3 then implies that $\mathcal{F}^I$ and $\mathcal{F}^J$ are independent. $\diamond$

Returning now to the coin toss example, we see that if I and J are disjoint sets of positive integers, and $\mathcal{A}^I = \sigma(\cup_{i\in I}\mathcal{A}_i)$ denotes the σ-algebra of all events concerning the outcomes of tosses i, $i \in I$, and similarly for $\mathcal{A}^J$, then $\mathcal{A}^I$ and $\mathcal{A}^J$ are independent.

5.2 Products of probability spaces. The coin toss probability spaces, both finite and infinite, that we have constructed are all examples of product probability spaces. We shall present the general construction of a product space and state its connection to independence.

Let $\{(\Omega_i\, \mathcal{F}_i, I\!P_i)\}$ be a family of probability spaces. We wish to construct a product probability measure on the product space

$$\otimes_1^\infty \Omega_i := \Omega_1 \times \Omega_2 \times \cdots.$$

A subset A of $\otimes_1^\infty \Omega_i$ is called a *rectangle* if there exists some positive integer n and sets $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}$ such that

$$A = A_1 \times \cdots \times A_n \times [\otimes_{n+1}^\infty \Omega_i] \tag{6}$$

The product σ-algebra $\otimes_1^\infty \mathcal{F}_i$ is defined to be the smallest σ-algebra containing all rectangles.

Theorem 5.6 There is a unique probability measure on $(\otimes_1^\infty \Omega_i, \otimes_1^\infty \mathcal{F}_i)$, denoted by $\otimes_1^\infty I\!P_i$ and called the product measure, with the property that for any rectangle of the form (6)

$$[\otimes_1^\infty I\!P_i]\,(A_1 \times \cdots \times A_n \times [\otimes_{n+1}^\infty \Omega_i]) = \prod_1^n I\!P_i(A_i) \tag{7}$$

Proof: The proof is similar to the construction of the space for infinite tosses of a fair coin. However it shall use a bit of integration theory for general measures. If you need review of this, you may wish to wait until you study expectations before reading the proof. Let $\mathcal{R}_1$ be the collection of all finite disjoint unions of rectangles of $\otimes_1^\infty \Omega_i$. Clearly $\mathcal{R}_1$ generates $\otimes_1^\infty \mathcal{F}_i$. It is easy to check that $\mathcal{R}_1$ is an algebra; hence it includes all finite unions of rectangles. Define a measure Q_1 on $\mathcal{R}_1$ by

$$Q_1(A_1 \times \cdots \times A_n \times [\otimes_{n+1}^\infty \Omega_i]) = \prod_1^n I\!P_i(A_i).$$

It is not hard to show that this measure is consistently defined and that it is finitely additive. To complete the proof it is only necessary to show that Q_1 is continuous from above at $\emptyset$, because then Carathéodory's theorem implies it has a unique, countably additive extension to $\otimes_1^\infty \mathcal{F}_i$. The technique for demonstrating continuity from above at $\emptyset$ generalizes the method used in the construction of the product measure in the coin tossing example. Again, we prove the contrapositive of continuity from above at the empty set by the following method. Let $\omega = (\omega_1, \omega_2, \ldots)$ denote a typical point of the infinite product space $\otimes_1^\infty \Omega_i$. If $A \subset \otimes_1^\infty \Omega_i$, define the section

$$A(\omega_1, \ldots, \omega_n) = \left\{ \omega' \in \otimes_{n+1}^\infty \Omega_i \;\middle|\; (\omega, \omega') \in A \right\}.$$

Suppose now that A_i, $1 \le i \le \infty$, is a decreasing sequence of sets. We shall show that if

$$\lim_{i \to \infty} Q_1(A_i) > 0, \tag{8}$$

then there exists a $\omega = (\omega_1, \omega_2, \ldots)$ such that for each positive integer N

$$A_i(\omega_1, \ldots, \omega_N) \text{ is non-empty for all positive integers } i. \tag{9}$$

It will follow then that $\omega \in \cap_i A_i$, thereby proving the contrapositive.

Some additional notation is needed to proceed. Let $\mathcal{R}_k$ denote the algebra of finite disjoint unions rectangles of $\otimes_k^\infty \Omega_i$, and define the finitely additive measure Q_i on $([\otimes_k^\infty \Omega_i], \mathcal{R}_k)$ in the same way we defined Q_1. Observe that if $A \in \mathcal{R}_1$, then the section $A(\omega_1, \ldots, \omega_k) \in \mathcal{R}_k$. For $k = 1$ and $A \in \mathcal{R}_1$, one can verify the formula,

$$Q_1(A) = \int Q_2\left(A(\omega_1)\right)\, dP_1(\omega_1). \tag{10}$$

(As part of this verification one shows that the map $\omega_1 \to Q_2\left(A(\omega_1)\right)$ is $\mathcal{F}_1$-measurable.)

We will verify (8) and (9) for $N = 1$. Thus consider a decreasing sequence A_i of subsets of $\mathcal{R}_1$ satisfying (8). Since the sequence of sections $A_i(\omega_1)$ is decreasing for each ω_1, $\lim_{i \to \infty} Q_2\left(A_i(\omega_1)\right)$ exists for every ω_1 and, by the dominated convergence theorem, assumption (8), and formula (10),

$$0 < \lim_i Q_1(A_i) = \int \lim_{i \to \infty} Q_2\left(A_i(\omega_1)\right)\, dP(\omega_1).$$

Therefore, there exists some $\omega_1 \in \Omega_1$ such that $Q_2\left(A_i(\omega_1)\right) > 0$ for all i, and hence that $A_i(\omega_1)$ is non-empty for all i. One now proves the general case by induction. The proof of each induction step is the same except that Q_1 is replaced by Q_N and A_i by $A_i(\omega_1, \ldots, \omega_N)$. $\diamond$.

The connection between product spaces and independence found in the coin tossing example generalizes completely. By the definition of the product measure,

$$\otimes_1^\infty I\!P_i\left(\{\omega\,;\, \omega_i \in A_i\}\right) = I\!P_i(A_i).$$

Thus (7) says precisely that the events $\{\omega\,;\, \omega_i \in A_i\}$, $1 \le i \le n$ are independent with respect to measure $\otimes_1^\infty I\!P_i$. The following proposition restates this fact in fancier language.

Proposition 5.7 Let $\mathcal{A}_n$ be the sub-σ-algebra of $\otimes_1^\infty \mathcal{F}_i$ consisting of sets of the form $\{\omega\ ;\ \omega_n \in A\}$, where $A \in \mathcal{F}_n$. The σ-algebras $\{\mathcal{A}_i;\ 1 \le i < \infty\}$ are independent in the probability space $(\otimes_1^\infty \Omega_i, \otimes_1^\infty \mathcal{F}_i, \otimes_1^\infty I\!P_i)$.