Probability and Measure - Patrick Billingsley

Chapter 1 - Probability

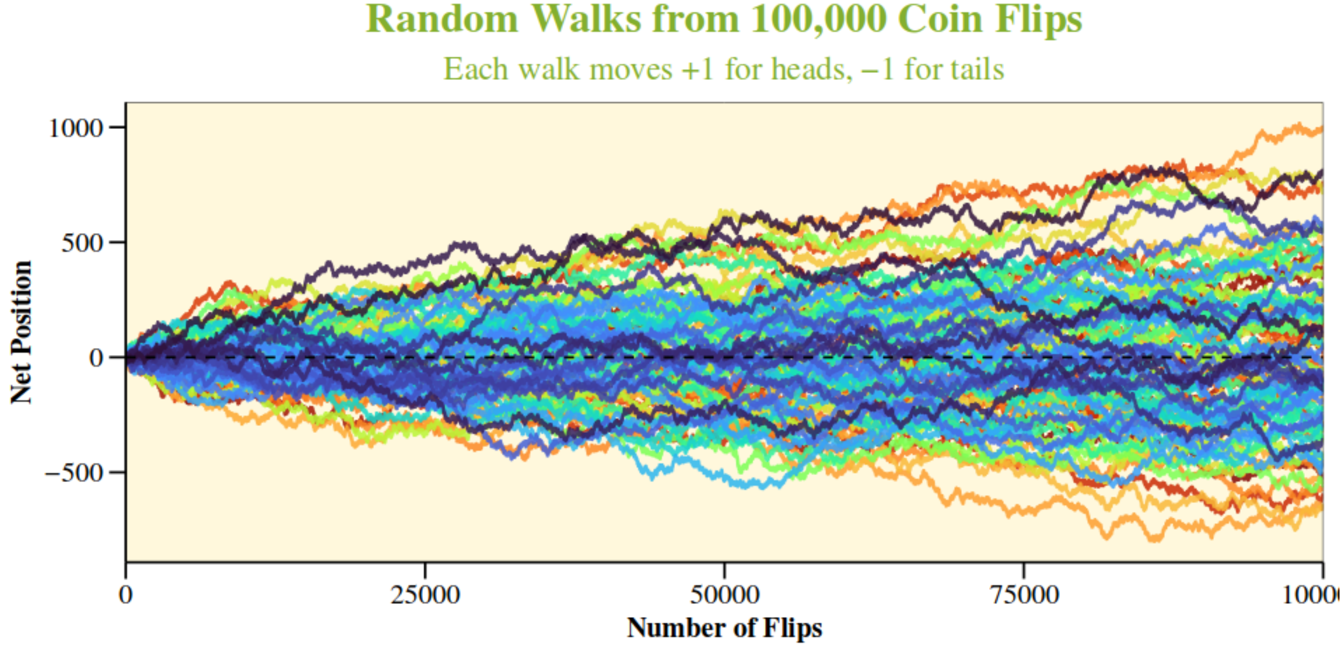
1.1 - Borel's Normal Number Theorem

Consider trying to analyze the frequency of heads and tails for an infinite number of coin tosses. Because academia must remain inaccessible to those without a lexile score in the 80th percentile, and arguably less for the sake of "*Mathematical Rigor*" Billingsley chooses to model this infinite sequence of coinflips as a so-called **nonterminating dyadic expansion** of the form:

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n} = .d_1(\omega)d_2(\omega)\cdots \tag{1.1}$$

That is basically just a really fancy way of saying it is an infinite sequence of ones and zeros —the finite case of this problem conforms with the familiar binomial distribution. However, this dyadic representation is cumbersome, it is often useful to use the **Rademacher Function** which is defined as:

$$r_n(\omega) = 2d_n(\omega) - 1 = \begin{cases} +1 & \text{if } d_n(\omega) = 1, \\ -1 & \text{if } d_n(\omega) = 0. \end{cases} \tag{1.2}$$



Which yields the following partial sum:
$$s_n(\omega) = \sum_{i=1}^{n} r_i(\omega). \tag{1.3}$$

If $f(\omega)$ is a step function taking value c_j in the interval $(x_{j-1}, x_j]$, where $0 = x_0 < x_1 < \cdots < x_k = 1$, then its integral in the sense of Riemann has the value
$$\int_0^1 f(\omega)\, d\omega = \sum_{j=1}^{k} c_j(x_j - x_{j-1}). \tag{1.4}$$

The Weak Law of Large Numbers

Theorem 1.1: For each $\varepsilon > 0$,
$$\lim_{n\to\infty} P\left[\omega : \left|\frac{1}{n}\sum_{i=1}^{n} d_i(\omega) - \frac{1}{2}\right| \geq \varepsilon\right] = 0 \tag{1.5}$$

Proof:

First things first, let us put this equation in a form that is easier to prove. Since $\sum_{i=1}^{n} d_i(\omega) = \frac{s_n(\omega)+n}{2}$, (1.4) with $\varepsilon/2$ in place of ε is the same thing as
$$\lim_{n\to\infty} P\left[\omega : \left|\frac{1}{n}s_n(\omega)\right| \geq \varepsilon\right] = 0. \tag{1.6}$$

Then, since this is a step function we know it must be of the following form:
$$\int_0^1 f(\omega)\, d\omega = \sum_{j=1}^{k} c_j(x_j - x_{j-1}). \tag{1.7}$$

Thus, the total length of the interval is $\frac{1}{2}$ meaning that:
$$\int_0^1 s_n(\omega) d\omega = 0 \tag{1.8}$$

Applying *Chebyshev's Inequality* to (1.5) yields:

$$P[\omega : |s_n(\omega)| \geq n\varepsilon] \leq \frac{1}{n^2\varepsilon^2} \int_0^1 s_n^2(\omega) d\omega = \frac{1}{n\varepsilon^2} \tag{1.9}$$

This inequality results from the finite union of intervals of a step function and an application of *Markov's Inequality*; this application is formally defined in the following lemma. **QED**

Let Ω denote the unit interval $(0, 1]$ if the intervals $I_i = (a_i, b_i]$ are disjoint and are contained in Ω, assign to A the probability

$$P(A) = \sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n} (b_i - a_i). \tag{1.10}$$

Lemma: If f is a nonnegative step function, then $[\omega : f(\omega) \geq \alpha]$ is for $\alpha > 0$ a finite union of intervals, and

$$P[\omega : f(\omega) \geq \alpha] \leq \frac{1}{\alpha} \int_0^1 f(\omega) d\omega \tag{1.11}$$

That is, the probability that the number of 1s in the sequence is greater than a number α is less than or equal to 1 over α times the expected number of 1s in the sequence.

Proof:

The set in question is the union of the intervals $(x_{i-1}, x_i]$ for which $c_i \geq a$. If $\sum'$ denotes summation over those i satisfying $c_i \geq a$, then

$$P[\omega : f(\omega) \geq a] = \sum{}' (x_i - x_{i-1}) \tag{1.12}$$

by the definition (1.9). On the other hand, since the c_i are all nonnegative by hypothesis, (1.10) gives

$$\begin{aligned} \int_0^1 f(\omega)\, d\omega = \sum_{i=1}^{k} c_i(x_i - x_{i-1}) &\geq \sum_{j=1}{}' c_i(x_i - x_{i-1}) \\ &\geq \sum{}' a(x_i - x_{i-1}). \end{aligned} \tag{1.13}$$

QED

The Strong Law of Large Numbers

is a special case of whats called **Borel's Normal Number Theorem**. Furthermore, the actual meat of this theorem cannot be adequately expressed in the familiar discrete probability theory. Consider the set:

$$N = \left[\omega : \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} d_i(\omega) = \frac{1}{2} \right] \tag{1.14}$$

The points within (1.14) are called **Normal Numbers**. We want to show that a real number ω drawn from the unit interval at random is practically certain to be a normal number, which means that there is practical certainty that the asymptotic relative frequency of 1s in the sequence is equal to $\frac{1}{2}$

Before we can proceed, we need to formalize practical certainty through the idea of a negligible subset. A subset $A \subset \Omega$ is called **negligible** if for every $\varepsilon > 0$, there exists a finite or countable collection of intervals $I_1, I_2, \ldots$ such that:

$$A \subset \bigcup_k I_k, \quad \text{and} \quad \sum_k |I_k| < \varepsilon. \tag{1.15}$$

A negligible set is one that can be covered by intervals the total sum of which lengths can be made arbitrarily small. From this we can then say that it is practically impossible that the random ω will lie in A if A is negligible, and regard it as practically certain that ω will lie in A if its complement A^c is negligible.

An important consequence of this is that a set consisting of a single point is clearly negligible, and so every countable set is also negligible; this is why the cumulative density function of a single point of a continuous probability distribution is zero. In the case of our infinite sequence of coin tosses this can be interpreted as saying that it is practically impossible to toss a coin infinitely many times and correctly predict the *exact* sequence that the tosses will follow.

With these preliminaries in place, we can now move on to defining and proving our next theorem.

Borel's Normal Number Theorem

Theorem 1.2: The set of normal numbers has negligible complement.
Proof:

$$N = \left[\omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d_i(\omega) = \frac{1}{2}\right] = \left[\omega : \lim_{n\to\infty} \frac{1}{n} s_n(\omega) = 0\right] \tag{1.16}$$

To prove that N^c is negligible we need to show that $A = N^c$ satisfies the two conditions defined by (1.15). We begin by applying (1.9) where $f(\omega) = s_n^4(\omega)$ and $\alpha = n^4\varepsilon^4$

$$P\left[\omega : |s_n(\omega)| \geq n\varepsilon\right] \leq \frac{1}{n^4\varepsilon^4} \int_0^1 s_n^4(\omega) d\omega \tag{1.17}$$

Since $s_n^4(\omega)$ is simply our partial sum raised to the fourth power we can equivalently express it as the sum of four Rademacher functions multiplied together where each of the respective indices (α, β, γ, δ) range from 1 to n.

$$s_n^4(\omega) = \sum r_\alpha(\omega) r_\beta(\omega) r_\gamma(\omega) r_\delta(\omega) \tag{1.18}$$

This sum can be simplified by combining like terms to produce a new equation in which the indices are distinct. This yields an equation of one of the following forms depending on how the indices cancel with one another.

$$s_n^4(\omega) \begin{cases} r_i^4(\omega) = 1, \\ r_i^2(\omega)\, r_j^2(\omega) = 1, \\ r_i^2(\omega)\, r_j(\omega)\, r_k(\omega) = r_j(\omega)\, r_k(\omega), \\ r_i^3(\omega)\, r_j(\omega) = r_i(\omega)\, r_j(\omega), \\ r_i(\omega)\, r_j(\omega)\, r_k(\omega)\, r_l(\omega). \end{cases} \tag{1.19}$$

This is a linear system of equations, evidently the first two forms integrate to one. Each of the subsequent forms, however, becomes an odd function over sub-intervals due to alternation of the highest index function. These last three forms integrate to zero since integrating an odd function over symmetric intervals results in cancellation.

The first form corresponds to the case where all 4 indices are equal, meaning there is effectively a single index that runs from 1 to n giving us n possible terms. The second form corresponds to the case where there is effectively 2 distinct indices. However, since there are three different ways to produce each pair of distinct indices the number of possible terms contributed by the second form is equal to:

$$3 \cdot \binom{n}{2} = 3 \cdot \frac{\cancel{2}n(n-1)}{\cancel{2}} = 3n(n-1) \tag{1.20}$$

We can then integrate $s_n^4(\omega)$ term by term, which yields:

$$\int_0^1 s_n^4(\omega)d\omega = n + 3n(n-1) \leq 3n^2 \tag{1.21}$$

Substituting this simplified expression into (1.17) gives the following:

$$P\left[\omega : \left|\frac{1}{n}s_n(\omega)\right| \geq \varepsilon\right] \leq \frac{3}{n^2\varepsilon^4} \tag{1.22}$$

Fix a positive sequence $\{\varepsilon_n\}$ tending to 0 slowly enough so that the following series converges.

$$\sum_n \varepsilon_n^{-4} n^{-2} \tag{1.23}$$

Now define the sets:

$$A_n = \left\{\omega : \left|\tfrac{1}{n}s_n(\omega)\right| \geq \varepsilon_n\right\} \quad \& \quad \begin{array}{l} P(A_n) \leq 3\varepsilon_n^{-4}n^{-2}, \\ \sum_n P(A_n) < \infty \end{array} \tag{1.24}$$

If for some m,
$\omega \in A_n^c$ for all $n \geq m$, then

$$\left|\frac{1}{n}s_n(\omega)\right| < \varepsilon_n \quad \text{for all } n \geq m \tag{1.25}$$

and since $\varepsilon_n \to 0$, it follows that ω is a normal number.

Thus, for each m,

$$\bigcap_{n=m}^{\infty} A_n^c \subseteq N \quad \text{or equivalently} \quad N^c \subseteq \bigcup_{n=m}^{\infty} A_n. \tag{1.26}$$

This last inclusion leads directly to the required covering.
Given any $\varepsilon > 0$, choose m so that

$$\sum_{n=m}^{\infty} P(A_n) < \varepsilon. \tag{1.27}$$

Each set A_n is a finite disjoint union of dyadic intervals:

$$A_n = \bigcup_k I_{nk}, \quad \sum_k |I_{nk}| = P(A_n). \tag{1.28}$$

Therefore, the union

$$\bigcup_{n=m}^{\infty} A_n = \bigcup_{n=m}^{\infty} \bigcup_k I_{nk} \tag{1.29}$$

is a countable union of intervals (not necessarily disjoint), with

$$\sum_{n=m}^{\infty} \sum_k |I_{nk}| = \sum_{n=m}^{\infty} P(A_n) < \varepsilon. \tag{1.30}$$

Hence, the collection of intervals $\{I_{nk} : n \geq m, k \geq 1\}$ provides a covering of N^c satisfying

$$N^c \subseteq \bigcup I_{nk}, \quad \sum |I_{nk}| < \varepsilon, \tag{1.31}$$

which is exactly the form required by the definition of negligibility (as given in equation (1.11)). **QED**

Length and Negligibility

The next theorem will act as our entry point for defining measure. These three statements taken together ensure the additivity of disjoint events that we expect within a sample space. In discrete probability this is a result of the properties of the Riemann integral. However, theorem 1.3 provides us with a way to guarantee the additivity of disjoint events independent of the Riemann integral, allowing us to generalize beyond the restrictive assumptions of Riemann integration.

Theorem 1.3: Let $I = (a, b]$ be an interval of length $|I| = b - a$ and, let $I_k = (a_k, b_k]$ be a finite or infinite sequence of bounded intervals (they need not be subintervals of $(0, 1]$). Then:

i. If $\bigcup_k I_k \subset I$, and the I_k are disjoint, then $\sum_k |I_k| \leq |I|$.

ii. If $I \subset \bigcup_k I_k$ (the I_k need not be disjoint), then $|I| \leq \sum_k |I_k|$.

iii. If $I = \bigcup_k I_k$, and the I_k are disjoint, then $|I| = \sum_k |I_k|$.

Proof:

To begin, we can observe that **iii.** follows from **i.** & **ii.** so we just need to prove the first two claims:

Proof of i:

Finite Case: Assume that there are n intervals. The result clearly holds if $n = 1$, we assume that it holds for $n - 1$. If a_n is the largest among $a_1, ..., a_n$, then $\bigcup_{k=1} I_{n-1}(a_k, b_k] \subset (a, a_n)$, so that $\sum_{k=1}^{n-1}(b_k - a_k) \leq a_n - a$ by induction, and thus $\sum_{k=1}^{n}(b_k - a_k) \leq a_n - a + (b_n - a_n) \leq b - a$.

Infinite Case: If there are an infinite amount of intervals, we just showed that any finite subcollection must satisfy **i.**, as a result $\sum_{k=1}^{n}(b_k - a_k) \leq b - a$, and since n is arbitrary this holds in the infinite case.

Proof of ii:

Finite Case: Assume that **ii.** holds for the case of $n - 1$ intervals and that $(a, b] \subset \bigcup_{k=1}^{n}(a_k, b_k]$ where $a_n < b \leq b_n$. If $a_n \leq a$, then the result clearly holds. Otherwise, $(a, a_n] \subset \bigcup_{k=1}^{n-1}(a_k, b_k]$, so that $\sum_{k=1}^{n-1}(b_k - a_k) \geq a_n - a$ by induction. Thus, $\sum_{k=1}^{n}(b_k - a_k) \geq (a_k - a) + (b_n - a_n) \geq b - a$. The finite case follows from this by induction.

Infinite Case: Assume that $(a, b] \subset \bigcup_{k=1}^{\infty}(a_k, b_k]$. If $0 < \varepsilon < b - a$, the open intervals $(a_k, b_k + \varepsilon 2^{-k})$ cover the closed interval $[a + \varepsilon, b]$, and it follows by the *Heine-Borel theorem* that $(a+\varepsilon, b] \subset \bigcup_{k=1}^{n}(a_k, b_k + \varepsilon 2^{-k}]$, meaning that by the finite case we have $b - (a+\varepsilon) \leq \sum_{k=1}^{n}(b_k + \varepsilon 2^{-k} - a_k) \leq \sum_{k=1}^{\infty}(b_k - a_k) + \varepsilon$. Thus, since ε is arbitrarily small the infinite case follows.

QED

Part **ii.** of theorem 1.3 proves the fact that an interval which has positive length cannot be negligible. The intuition behind this fact follows directly from our definition of a negligible interval being one that can be made arbitrarily small. Any interval of positive length cannot be made smaller than its length, and therefore cannot be negligible. However, in order to rigorously prove this fact we can defer to the intrinsic properties of the real numbers.

1.2 - Probability Measures

Spaces, Semi-Rings, Rings, and Fields

Let Ω be an arbitrary **space** or set of points ω. As it pertains to probability, Ω is the set of all possible outcomes ω; a subset of Ω is an **event** and an element ω of Ω is a **sample point**.

This definition is just too broad to do anything with in it's current form. Instead, we are interested in certain **classes** of sets that have special properties that allow us to actually accomplish something. A set that is part of a given class $\mathcal{F}$ is said to be **measurable** over $\mathcal{F}$. Specifically, if we want to systematize the kinds of problems presented in 1.1 we are interested in the class of sets that contains the intervals and is closed under the formation of countable unions and intersections.

A collection of subsets $\mathcal{F}$ on Ω is called a **Semi-ring** if it satisfies the following three conditions:

i. $\emptyset \in \mathcal{F}$
ii. for any $A, B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$
iii. for any $A, B \in \mathcal{F}$, $B \setminus A = \bigcup c_i$ where $c_i \in \mathcal{F}$

A collection of subsets $\mathcal{F}$ on Ω is called a **Ring** if it satisfies the following three conditions:

$$\left.\begin{array}{rl} \textbf{i.} & \emptyset \in \mathcal{F} \\ \textbf{ii.} & \text{for any } A, B \in \mathcal{F},\ A \cup B \in \mathcal{F} \\ \textbf{iii.} & \text{for any } A, B \in \mathcal{F},\ B \setminus A \in \mathcal{F} \end{array}\right\} \text{ A Ring that also contains } \Omega \text{ is called a } \textbf{Field}$$

Out interest will primarily lie in considering fields, however it is important to clarify the subtle differences between semiring, rings, and fields.

Field Axioms:

A class $\mathcal{F}$ of an arbitrary nonempty space Ω is called a **Field** if it contains Ω itself and is closed under the formation of complements and finite unions. In otherwords, if it satisfies the first three of the following axioms:

i. $\Omega \in \mathcal{F}$ (Non-Empty)
ii. $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$ (Closure Under Complementation)
iii. $A, B \in \mathcal{F}$ implies $A \cup B \in \mathcal{F}$ (Closure Under **Finite** Union)
↳ implies $A \cap B \in \mathcal{F}$ (Closure Under **Finite** Intersection)

iv. $A_1, A_2, \cdots \in \mathcal{F}$ implies $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$ (Closure Under **Countable** Union)
↳ implies $A_1 \cap A_2 \cap \cdots \in \mathcal{F}$ (Closure Under **Countable** Intersection)

↳ A field $\mathcal{F}$ that also satisfies **iv.** is called a ***σ*-Field**.

If $\mathcal{F}$ is closed under Complementation, then it is closed under the formation of finite unions if and only if it is closed under the formation of finite intersections. Furthermore, since all finite values are by definition countable, the definition of a σ-field is more restrictive than that of a field.

The largest σ-field in Ω is the **Power Class** 2^Ω, which consists of *all* possible subsets of Ω; conversely, the smallest σ-field consists only of the empty set and Ω itself. Let $\mathcal{A}$ be a class.

Consider the class of sets that...

i. Contains $\mathcal{A}$. ii. Is a σ-field. iii. Is as small as possible.	$\Longrightarrow$	Such a class is the ***σ*-field generated by $\boldsymbol{\mathcal{A}}$**; it is the intersection of all the σ-fields containing $\mathcal{A}$. Denoted by $\sigma(\mathcal{A})$

When we say that $\sigma(\mathcal{A})$ is as small as possible we are saying that every σ-field that contains $\mathcal{A}$ also contains $\sigma(\mathcal{A})$.

Stated more formally, $\sigma(\mathcal{A})$ has the following properties:

i. $\mathcal{A} \subset \sigma(\mathcal{A})$
ii. $\sigma(\mathcal{A})$ is a σ-field
iii. If $\mathcal{A} \subset \mathcal{G}$ and $\mathcal{G}$ is a σ-field, then $\sigma(\mathcal{A}) \subset \mathcal{G}$

Probability Measures

A **set function** is a real-valued function that is defined on some class of subsets of Ω. A set function P on a field $\mathcal{F}$ is a **Probability Measure** if it satisfies the following three conditions (Kolmogorov Axioms):

i. $0 \leq P(A) \leq 1$ for $A \in \mathcal{F}$; **Nonegativity of Probability**
ii. $P(\emptyset) = 0$, $P(\Omega) = 1$; **Law of Total Probability**

iii. If $A_1, A_2, \ldots$ is a disjoint sequence of $\mathcal{F}$-sets and if $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k). \quad \textbf{Finite Additivity of Probability} \tag{2.1}$$

If $\mathcal{F}$ is a σ-field in Ω and P is a probability measure on $\mathcal{F}$, the triple $(\Omega, \mathcal{F}, P)$ is called a **Probability Measure Space** (or simply a probability space), and any $\mathcal{F}$-set A for which $P(A) = 1$ is called a **support** of P.

It is usually relatively easy to tell whether or not a given measure satisfies the first two conditions. In general, the hardest part of showing that a given $(\Omega, \mathcal{F}, P)$ is a probability measure space is showing that it satisfies finite additivity.

Example 2.1:

Space: Let Ω be a countable space.
σ-field: Let $\mathcal{F}$ be the σ-field of all subsets of Ω.
Measure: Let $p(\omega)$ be a nonnegative function on Ω such that $\sum_{\omega \in \Omega} p(\omega) = 1$. Then, define $P(A) = \sum_{\omega \in A} p(\omega)$.

We let $A = \bigcup_{i=1}^{\infty} A_i$, where the A_i are disjoint and $\omega_{i1}, \omega_{i2}, \ldots$ are the points in A_i
Then, by the definition of a non-negative series,

$p(A) = \sum_{ij} p(\omega_{ij}) = \sum_i \sum_j p(\omega_{ij}) = \sum_i P(A_i)$, thus P is countably additive.

This $(\Omega, \mathcal{F}, P)$ is the familiar **Discrete Probability Space.**

Suppose that P is a probability measure on a field $\mathcal{F}$, and that $A, B \in \mathcal{F}$ and $A \subset B$. Since $P(A) + P(B - A) = P(B)$

$$P(A) \leq P(B) \qquad \text{if } A \subset B \qquad \textbf{Monotonicity} \tag{2.2}$$

Furthermore, for any $A, B \in \mathcal{F}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.3}$$

Which is just the two variable case of the general **Inclusion-Exclusion Formula**:

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots \\ \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n) \tag{2.4}$$

If $B_1 = A_1$ and $B_k = A_k \cap A_1^c \cap \cdots \cap A_{k-1}^c$, then the B_k are disjoint and $\bigcup_{k=1}^n A_k = \bigcup_{k=1}^n B_k$, so that $P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(B_k)$. Since $P(B_k) \leq P(A_k)$ by monotonicity:

$$P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k). \qquad \textbf{Finite Subadditivity} \tag{2.5}$$

Theorem 2.1: Let P be a probability measure on a field $\mathcal{F}$.
i. Continuity from below: If A_n and A lie in $\mathcal{F}$ and $A_n \uparrow A$, then $P(A_n) \uparrow P(A)$.
ii. Continuity from above: If A_n and A lie in $\mathcal{F}$ and $A_n \downarrow A$, then $P(A_n) \downarrow P(A)$.
iii. Countable subadditivity: If $A_1, A_2, \ldots$ and $\bigcup_{k=1}^\infty A_k$ lie in $\mathcal{F}$ (the A_k need not be disjoint), then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k). \tag{2.6}$$

Proof:

i: Let $B_1 = A_1$ and $B_k = A_k - Ak - 1$. Then, B_k are disjoint, $A = \bigcup_{k=1}^\infty B_k$, and $A_n = \bigcup_{k=1}^\infty B_k$, thus by countable and finite additivity, $P(A) = \sum_{k=1}^\infty P(B_k) = \lim_n \sum_{k=1}^n P(B_k) = \lim_n P(A_n)$
ii. $A_n \downarrow A$ implies $A_n^c \uparrow A^c$, meaning that $1 - P(A_n) \uparrow 1 - P(A)$.
iii. Using the definition finite subadditivity we can then apply the result from **i.** give the desired result. **QED**

Lebesgue Measure on the Unit Interval

The current issue at hand is trying to prove that P is countably additive. For the sake of notational simplicity we will redefine P to λ.

We begin by defining $\mathfrak{B}_0$ as a field of finite disjoint unions of intervals in $(0, 1]$. From here we will define $\mathcal{J}$ to be the class of subintervals $(a, b]$ of $(0, 1]$; then $\lambda(I) = |I| = b - a$ is ordinary length, and $\varnothing$ as a zero length element in $\mathcal{J}$. If

$A = \cup_{i=1}^n I_i$, where I_i are disjoint $\mathcal{J}$-sets. We can redefine how we assign probability to A as:

$$P(A) = \sum_{i=1}^n |I_i| = \sum_{i=1}^n (b_i - a_i) \implies \lambda(A) = \sum_{i=1}^n \lambda(I_i) = \sum_{i=1}^n |I_i| \tag{2.7}$$

Applying Theorem 1.3.iii to this new probability measure yields:

$$\sum_{i=1}^n |I_i| = \sum_{i=1}^n \sum_{j=1}^m |I_i \cap J_j| = \sum_{j=1}^m |J_j| \tag{2.8}$$

Thus, (2.7) defines a set function λ on $\mathcal{B}_I$ that we call the **Lebesgue Measure**.

Theorem 2.2: The Lebesgue measure λ is a countably additive probability measure on the field $\mathfrak{B}_0$.

Proof:

Suppose that $A = \bigcup_{k=1}^\infty A_k$, where A and the A_k are $\mathfrak{B}_0$-sets and the A_k are disjoint. Then $A = \bigcup_{i=1}^n I_i$ and $A_k = \bigcup_{j=1}^{m_k} J_{kj}$ are disjoint unions of $\mathcal{J}$-sets. Applying Theorem 1.3.iii to the definition of the Lebesgue Measure yields:

$$\begin{aligned} \lambda(A) &= \sum_{i=1}^n |I_i| = \sum_{i=1}^n \sum_{k=1}^\infty \sum_{j=1}^{m_k} |I_i \cap J_{kj}| \\ &= \sum_{k=1}^\infty \sum_{j=1}^{m_k} |J_{kj}| = \sum_{k=1}^\infty \lambda(A_k) \end{aligned} \tag{2.9}$$

QED

1.3 - Extension & Uniqueness

This entire section is entirely a pursuit of proving the following theorem:

Theorem 3.1: A probability measure on a field has a unique extension to the generated σ-field

In other words, let P be a probability measure on a field $\mathbb{F}_0$ of subsets of Ω, and let $\mathcal{F} = \sigma(\mathcal{F}_0)$. Then there exists a probability measure Q on $\mathcal{F}$ such that $Q(A) = P(A)$ for $A \in \mathcal{F}_0$. Further, if Q' is another probability measure on $\mathcal{F}$ such that $Q'(A) = P(A)$ for $A \in \mathcal{F}_0$, then $Q'(A) = Q(A)$ for $A \in \mathcal{F}$

Extension - Carathéodory's Method

Let P be a probability measure on a field $\mathcal{F}_0$. We want to extend P to a more general class. For each subset A of Ω, we define its **Outer Measure** as the following:

$$P^\star(A) = \textbf{inf} \sum_n P(A_n) \qquad \textbf{Outer Measure} \tag{3.1}$$

Where the infimum extends over all finite and infinite sequences $A_1, A_2, ...$ of $\mathcal{F}_0$-sets satisfying $A \subset \bigcup_n A_n$. When considered together, the A_n form a good covering of A if they do not overlap very much, nor do they extend very far beyond A. When this is this case $\sum_n P(A_n)$ is a good outer approximation of the measure A. As a result of this, (3.1) acts as our initial attempt to assign a measure to A.
Conversely, the **Inner Measure** approximates A from the inside by subtracting the compliment of A from one:

$$P_\star(A) = 1 - P^\star(A^c) \qquad \textbf{Inner Measure} \tag{3.2}$$

Using these two pre-measures, we want to assign measure to those A that satisfy the following condition:

$$P^\star(A \cap E) + P^\star(A^c \cap E) = P^\star(E) \quad \text{for every set } E \tag{3.3}$$

A set A is called $\boldsymbol{P^\star}$**-Measurable** if (3.3) holds for all E. Let $\mathcal{M}$ be the class of $P^\star$-measurable sets. Our task then becomes to show that:

$\mathcal{M}$ contains $\sigma(\mathcal{F}_0)$ and

The restriction of $P^\star$ to $\sigma(\mathcal{F}_0)$ is the required extension of P

We accomplish this by employing four key properties of the set-function $P^\star$:

i. $P^\star(\emptyset) = 0$

ii. $P^\star$ is non-negative: $P^\star \geq 0$ for every $A \subset \Omega$
iii. $P^\star$ is monotone: $A \subset B$ implies $P^\star(A) \leq P^\star(B)$
iv. $P^\star$ is countably subadditive: $P^\star(\bigcup_n A_n) \leq \sum_n P^\star(A_n)$
The first three of these properties are true by definition.
Proof of iv.

> For a given ε, choose $\mathcal{F}_0$-sets B_{nk} such that $A_n \subseteq \bigcup_k B_{nk}$ and $\sum_k P(B_{nk}) < P^\star(A_n) + \varepsilon/2^n$, which is possible by the definition (3.1). Now $\bigcup_n A_n \subseteq \bigcup_{n,k} B_{nk}$, so that $P^\star(\bigcup_n A_n) < \sum_{n,k} P(B_{nk}) < \sum_n (P^\star(A_n) + \varepsilon/2^n)$, and **iv** follows. **QED**.

A lies in the class $\mathcal{M}$ of $P^\star$-measurable sets if it splits each E in 2^ω in such a way that $P^\star$ adds for the pieces; mathematically that is:

$$P^\star(A \cap E) + P^\star(A^c \cap E) \leq P^\star(E) \tag{3.4}$$

↳ i.e., A splits E in a way that preserves the additivity of $P^\star$

Lemma 1: The class $\mathcal{M}$ is a field.
Proof:

> It is clear that $\Omega \in \mathcal{F}'$ and that $\mathcal{F}'$ is closed under complementation. Suppose that $A, B \in \mathcal{F}'$ and $E \subseteq \Omega$. Then
>
> $$\begin{aligned} P^\star(E) &= P^\star(B \cap E) + P^\star(B^c \cap E) \\ &= P^\star(A \cap B \cap E) + P^\star(A^c \cap B \cap E) \\ &\quad + P^\star(A \cap B^c \cap E) + P^\star(A^c \cap B^c \cap E) \\ &\geq P^\star(A \cap B \cap E) \quad \text{(by subadditivity)} \\ &\quad + P^\star\big((A^c \cap B \cap E) \cup (A \cap B^c \cap E) \cup (A^c \cap B^c \cap E)\big) \\ &= P^\star((A \cap B) \cap E) + P^\star((A \cap B)^c \cap E), \end{aligned} \tag{3.5}$$
>
> Hence $A \cap B \in \mathcal{F}'$, and $\mathcal{F}'$ is a field. **QED**

Lemma 2: If $A_1, A_2, ...$ is a finite sequence or infinite sequence of $\mathcal{M}$-sets, then for each:$E \subset \Omega$

$$P^\star\left(E \cap \left(\bigcup_k A_k\right)\right) = \sum_k P^\star(E \cap A_k) \tag{3.6}$$

Proof:

Consider first the case of finitely many A_k, say n of them. For $n = 1$, there is nothing to prove. In the case $n = 2$, if $A_1 \cup A_2 = \Omega$, then (3.6) is just (3.4) with A_1 (or A_2) in the role of A. If $A_1 \cup A_2 \subsetneq \Omega$, split $E \cap (A_1 \cup A_2)$ by A_1 and A_2 (or by A_2 and A_1) and use (3.4) and disjointness.

Assume (3.6) holds for the case of $n-1$ sets. By the case $n = 2$, together with the induction hypothesis,

$$P^\star\left(E \cap \bigcup_{k=1}^{n} A_k\right) = P^\star\left(E \cap \bigcup_{k=1}^{n-1} A_k\right) + P^\star(E \cap A_n) = \sum_{k=1}^{n} P^\star(E \cap A_k).$$

Thus, (3.6) holds in the finite case.

For the infinite case, use monotonicity:

$$P^\star\left(E \cap \bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} P^\star\left(E \cap \bigcup_{k=1}^{n} A_k\right) = \lim_{n\to\infty} \sum_{k=1}^{n} P^\star(E \cap A_k).$$

Hence, the left side of (3.6) is greater than or equal to the right. The reverse inequality follows by countable subadditivity. **QED**

Lemma 3: The class $\mathcal{M}$ is a σ-field, and $P^\star$ restricted to $\mathcal{M}$ is countably additive.

Proof:

Suppose that $A_1, A_2, \ldots$ are disjoint sets with union A. Since $F_n = \bigcup_{k=1}^{n} A_k$ lies in the field $\mathcal{F}'$, we have

$$P^\star(E) = P^\star(E \cap F_n) + P^\star(E \cap F_n^c).$$

To the first term on the right, apply (3.6), and to the second term apply monotonicity (since $F_n \subseteq A^c$):

$$P^\star(E) \geq \sum_{k=1}^{n} P^\star(E \cap A_k) + P^\star(E \cap A^c).$$

Let $n \to \infty$ and use (3.6) again:

$$P^\star(E) \geq \sum_{k=1}^{\infty} P^\star(E \cap A_k) + P^\star(E \cap A^c) = P^\star(E \cap A) + P^\star(E \cap A^c).$$

Hence A satisfies (3.5), and so lies in $\mathcal{F}'$, which is therefore closed under the formation of countable disjoint unions.

From the fact that $\mathcal{F}'$ is a field closed under the formation of countable disjoint unions, it follows that $\mathcal{F}'$ is a σ-field. (For sets $B_k \in \mathcal{F}'$, let $A_1 = B_1$, and for $k \geq 2$, let

$$A_k = B_k \cap B_1^c \cap \cdots \cap B_{k-1}^c;$$

then the A_k are disjoint $\mathcal{F}'$-sets and

$$\bigcup_k B_k = \bigcup_k A_k \in \mathcal{F}'.$$

The countable additivity of $P^\star$ on $\mathcal{F}'$ then follows from (3.6): take $E = \Omega$.

QED

Lemma 4: If $P^\star$ is defined by $P^\star(A) = \mathbf{inf} \sum_n P(A_n)$, then $\mathcal{F}_0 \in \mathcal{M}$

Proof:

Suppose that $A \in \mathcal{G}$. Given E and $\varepsilon > 0$, choose $\mathcal{F}_0$-sets A_n such that

$$E \subseteq \bigcup_n A_n \quad \text{and} \quad \sum_n P(A_n) < P^\star(E) + \varepsilon.$$

Define

$$B_n = A_n \cap A \quad \text{and} \quad C_n = A_n \cap A^c.$$

Since $\mathcal{F}_0$ is a field, both B_n and C_n lie in $\mathcal{F}_0$. Also,

$$E \cap A \subseteq \bigcup_n B_n, \quad E \cap A^c \subseteq \bigcup_n C_n.$$

By the definition of $P^\star$ and the finite additivity of P, we get:

$$P^\star(E \cap A) + P^\star(E \cap A^c) \leq \sum_n P(B_n) + \sum_n P(C_n) = \sum_n P(A_n) < P^\star(E) + \varepsilon.$$

Hence $A \in \mathcal{G} \subseteq \mathcal{F}'$. **QED**

Lemma 5: If $P^\star$ is defined by $P^\star(A) = \mathbf{inf} \sum_n P(A_n)$, then:

$$P^\star(A) = P(A) \qquad \text{for } A \in \mathcal{F}_0 \tag{3.7}$$

Proof:

$P^\star(A) \leq P(A)$ for $A \in \mathcal{F}_0$.

If $A \subseteq \bigcup_n A_n$, where A and each A_n are in $\mathcal{F}_0$, then by the countable subadditivity and monotonicity of P on $\mathcal{F}_0$, we have:

$$P(A) \leq \sum_n P(A \cap A_n) \leq \sum_n P(A_n).$$

Hence, taking the infimum over all such covers $\{A_n\} \subseteq \mathcal{F}_0$, we conclude:

$$P^\star(A) \leq P(A).$$

QED

⇒*Recall:*

Theorem 3.1: A probability measure on a field has a ~~unique~~* extension to the generated σ-field.

* We cannot yet prove the uniqueness; the proof below only shows that the extension from the probability measure to the generated σ-field exists.

Proof of Theorem 3.1:

Suppose that $P^\star$ is the outer measure of a (countably additive) probability measure P on the field $\mathcal{F}_0$. Let $\mathcal{F} = \sigma(\mathcal{F}_0)$.

By Lemmas 3 and 4, $\mathcal{F}_0 \subseteq \mathcal{F}' \subseteq 2^\Omega$.

By Lemma 5, $P^\star(\Omega) = P(\Omega) = 1$. By Lemma 3, $P^\star$ (which is defined on all of 2^Ω) restricted to $\mathcal{F}'$ is therefore a probability measure on $\mathcal{F}'$. Then, $P^\star$ further restricted to $\mathcal{F}$ is clearly a probability measure on that class as well.

This measure on $\mathcal{F}$ is the required extension, because by Lemma 5 it agrees with P on $\mathcal{F}_0$.

QED

Uniqueness and Dynkin's π-λ Theorem

To prove the extension in Theorem 3.1 is unique requires some auxiliary concepts: Let $\mathcal{P}$ and $\mathcal{L}$ be classes of subsets of Ω. We call $\mathcal{P}$ a **π-System** if it is closed under the formation of finite intersections, and we call $\mathcal{L}$ a **λ-Systems** if it contains Ω and is closed under the formation of both complements and countable disjoint unions.

π-Systems ($\mathcal{P}$)	**λ-Systems ($\mathcal{L}$)**
A class $\mathcal{P} \subseteq \mathcal{P}(\Omega)$ is a π-system if: **π.i** $A, B \in \mathcal{P} \Rightarrow A \cap B \in \mathcal{P}$ Note: All fields are π-systems, but not all π-systems are fields.	A class $\mathcal{L} \subseteq \mathcal{P}(\Omega)$ is a λ-system if: **λ.i** $\Omega \in \mathcal{L}$ **λ.ii** $A \in \mathcal{L} \Rightarrow A^c \in \mathcal{L}$ ↳ **λ.ii′** $A, B \in \mathcal{L}$ and $A \subset B \Rightarrow B \setminus A \in \mathcal{L}$ **λ.iii** If $A_1, A_2, \ldots \in \mathcal{L}$ are disjoint, then $\bigcup_n A_n \in \mathcal{L}$

Since the definition of a λ-system is just a weaker version of the definition of a σ-field, all σ-fields are also λ-systems but not all λ-systems are also σ-fields.

Lemma 6: A class that is both a π-system and a λ-system is a σ-field.
Proof:

> The class contains Ω by **λ.i** and is closed under the formation of complements and finite intersection by **λ.ii** and **π.i**, meaning that it is a field. Furthermore, it is a σ-field because if it contains sets A_n then it also contains the disjoint sets $B_n = A_n \cap A_1^c \cap \cdots \cap A_{n-1}^c$ and by **λ.iii** contains $\bigcup_n A_n = \bigcup_n B_n$. **QED**

Dynkin's π-λ Theorem
Theorem 3.2: If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subset \mathcal{L}$, then it follows that $\sigma(\mathcal{P}) \subset \mathcal{L}$.

Proof:

> We start by considering the smallest λ-system containing $\mathcal{P}$. Let $\mathcal{L}_0$ be the λ-system generated by $\mathcal{P}$(i.e. let $\mathcal{L}_0$ be the intersection of all λ-systems containing $\mathcal{P}$). Thus, every λ-system that contains $\mathcal{P}$ also contains $\mathcal{L}_0$ meaning that $\mathcal{P} \subset \mathcal{L}_0 \subset \mathcal{L}$. If it can be show that $\mathcal{L}_0$ is also a π-system, then it follows from lemma 6 that it is a σ-field. From the minimality of $\sigma(\mathcal{P})$ it then follows that $\sigma(\mathcal{P}) \subset \mathcal{L}$. Thus , it suffices to show that $\mathcal{L}_0$ is a π-system.

$\mathcal{L}_A$ is a λ-system:

λ.i: For each A, let $\mathcal{L}_A$ be the class of sets B such that $A \cap B \in \mathcal{L}_0$. If A is assumed to lie in $\mathcal{P}$, or even if A is merely assumed to lie in $\mathcal{L}_0$, then $\mathcal{L}_A$ is a λ-system; $A \cap \Omega = A \in \mathcal{L}_0$ by the assumption. $\checkmark$

λ.ii′: If $B_1, B_2 \in \mathcal{L}_A$ and $B_1 \subset B_2$ then the λ-system $\mathcal{L}_0$ contains $A \cap B_1$ and $A \cap B_2$ and hence it contains the proper difference $(A \cap B_2) - (A \cap B_1) = A \cap (B_2 - B_1)$, so that $\mathcal{L}_A$ contains $B_2 - B_1$. $\checkmark$

λ.iii: If B_n are disjoint $\mathcal{L}_A$-sets, then $\mathcal{L}_0$ contains the disjoint sets $A \cap B_n$ and hence $\mathcal{L}_0$ contains their union $A \cap (\bigcup_n B_n)$. $\checkmark$

If $A \in \mathcal{P}$, then since $B \in \mathcal{P}$, and $\mathcal{P}$ is a π-system, we have $A \cap B \in \mathcal{P} \subset \mathcal{L}_0$. Therefore, since $\mathcal{L}_A$ is defined as the class of sets B such that $A \cap B \in \mathcal{L}_0$, it follows that $B \in \mathcal{L}_A$. Thus, $A \in \mathcal{P}$ implies $\mathcal{P} \subset \mathcal{L}_A$, and since $\mathcal{L}_A$ is a λ-system, and we defined $\mathcal{L}_0$ to be the smallest λ-system containing $\mathcal{P}$, it follows that $\mathcal{L}_0 \subset \mathcal{L}_A$.

Hence, for all $A \in \mathcal{P}$ and $B \in \mathcal{L}_0$, we have $B \in \mathcal{L}_A$, meaning that the intersection $A \cap B$ lies in $\mathcal{L}_0$. Since $\mathcal{P} \subset \mathcal{L}_0$, $A \in \mathcal{P}$ implies that $A \in \mathcal{L}_0$, meaning that for all $A, B \in \mathcal{L}_0$, we have $A \cap B \in \mathcal{L}_0$. Therefore, $\mathcal{L}_0$ is closed under finite intersections, i.e., it is a π-system. **QED**

Since all fields are π-systems, we have one final theorem to prove that guarantees the uniqueness of the extension asserted in theorem 3.1.

Theorem 3.3: Suppose that P_1 and P_2 are probability measures on $\sigma(\mathcal{P})$, where $\mathcal{P}$ is a π-system. If P_1 and P_2 agree on $\mathcal{P}$, then they agree on $\sigma(\mathcal{P})$.

Proof:

Let $\mathcal{L}$ be the class of sets A in $\sigma(\mathcal{P})$ such that $P_1(A) = P_2(A)$. Clearly $\Omega \in \mathcal{L}$. If $A \in \mathcal{L}$, then $P_1(A^c) = 1 - P_1(A) = 1 - P_2(A) = P_2(A^c)$, and hence $A^c \in \mathcal{L}$. If A_n are disjoint sets in $\mathcal{L}$, then $P_1(\bigcup_n A_n) = \sum_n P_1(A_n) = \sum_n P_2(A_n) = P_2(\bigcup_n A_n)$, and hence $\bigcup_n A_n \in \mathcal{L}$. Therefore $\mathcal{L}$ is a λ-system. Since by hypothesis $\mathcal{P} \subset \mathcal{L}$ and $\mathcal{P}$ is a π-system, the π-λ theorem gives $\sigma(\mathcal{P}) \subset \mathcal{L}$, as required. **QED**

Finally, with seperate proofs of both existence and uniqueness established we can finally combine them into a complete proof of the existence and the uniqueness of the extension from a probability measure to the σ-field that it generates.

⇒*Recall:*

Theorem 3.1: A probability measure on a field has a *unique* extension to the generated σ-field.

Proof of Theorem 3.1:

Let $\mathcal{F}_0$ be a field on Ω, and suppose that P is a probability measure on $\mathcal{F}_0$. Define $\mathcal{F} = \sigma(\mathcal{F}_0)$, the smallest σ-field containing $\mathcal{F}_0$.

Existence: Define the outer measure $P^\star$ via Carathéodory's construction from P using (3.1). By Lemmas 3 and 4, the class

$$\mathcal{F}' = \{A \subseteq \Omega : P^\star \text{ is additive on subsets of } A\}$$

is a σ-field that contains $\mathcal{F}_0$, and $P^\star$ restricted to $\mathcal{F}'$ is a probability measure that extends P. In particular, the restriction of $P^\star$ to $\mathcal{F} \subseteq \mathcal{F}'$ defines a probability measure on $\mathcal{F}$ that agrees with P on $\mathcal{F}_0$. This proves existence.

Uniqueness: Suppose P_1 and P_2 are two extensions of P to probability measures on $\mathcal{F} = \sigma(\mathcal{F}_0)$. Since $\mathcal{F}_0$ is a field, it is in particular a π-system. Since P_1 and P_2 agree on $\mathcal{F}_0$, Theorem 3.3 implies that P_1 and P_2 agree on $\sigma(\mathcal{F}_0) = \mathcal{F}$. Therefore, the extension is unique. **QED**

Monotone Classes

A class $\mathcal{M}$ of subsets of Ω is **Monotone** if it is closed under the formation of monotone unions and intersections:

i. $A_1, A_2, ... \in \mathcal{M}$ and $A_n \uparrow A$ imply $A \in \mathcal{M}$

ii. $A_1, A_2, ... \in \mathcal{M}$ and $A_n \downarrow A$ imply $A \in \mathcal{M}$

Halmo's Monotone Class Theorem is similar to the π-λ theorem, however Billingsley says that he uses it with less frequency. However, I'll note it down just for the sake of having comprehensive notes.

Halmo's Monotone Class Theorem:

Theorem 3.4: If $\mathcal{F}_0$ is a field and $\mathcal{M}$ is a monotone class, then $\mathcal{F}_0 \in \mathcal{M}$ implies that $\sigma(\mathcal{F}_0) \subset \mathcal{M}$

Proof:

Let $m(\mathcal{F}_0)$ be the minimal monotone class over $\mathcal{F}_0$—the intersection of all monotone classes containing $\mathcal{F}_0$. It is enough to prove $\sigma(\mathcal{F}_0) \subseteq m(\mathcal{F}_0)$; this will follow if $m(\mathcal{F}_0)$ is shown to be a field, because a monotone field is a σ-field.

Consider the class $\mathcal{G} = \{A : A^c \in m(\mathcal{F}_0)\}$. Since $m(\mathcal{F}_0)$ is monotone, so is $\mathcal{G}$.

Since $\mathcal{F}_0$ is a field, $\mathcal{F}_0 \subseteq \mathcal{G}$, and so $m(\mathcal{F}_0) \subseteq \mathcal{G}$. Hence $m(\mathcal{F}_0)$ is closed under complementation.

Define $\mathcal{G}_1$ as the class of A such that $A \cup B \in m(\mathcal{F}_0)$ for all $B \in \mathcal{F}_0$. Then $\mathcal{G}_1$ is a monotone class and $\mathcal{F}_0 \subseteq \mathcal{G}_1$; from the minimality of $m(\mathcal{F}_0)$, it follows $m(\mathcal{F}_0) \subseteq \mathcal{G}_1$.

Define $\mathcal{G}_2$ as the class of B such that $A \cup B \in m(\mathcal{F}_0)$ for all $A \in m(\mathcal{F}_0)$. Then $\mathcal{G}_2$ is a monotone class. Now from $m(\mathcal{F}_0) \subseteq \mathcal{G}_1$, it follows that $A \in m(\mathcal{F}_0)$ and $B \in \mathcal{F}_0$ together imply that $A \cup B \in m(\mathcal{F}_0)$; in other words, $B \in \mathcal{F}_0$ implies that $B \in \mathcal{G}_2$. Thus $\mathcal{F}_0 \subseteq \mathcal{G}_2$; by minimality, $m(\mathcal{F}_0) \subseteq \mathcal{G}_2$, and hence $A, B \in m(\mathcal{F}_0)$ implies that $A \cup B \in m(\mathcal{F}_0)$. **QED**

Lebesgue Measure & Borel Sets

Consider the unit interval (0,1] together with the field $\mathfrak{B}_0$ of finite disjoint unions of sub intervals, and the σ-field $\mathfrak{B} = \sigma(\mathfrak{B}_0)$ which is comprised of Borel sets in (0,1]. Theorem 2.2 defines a probability measure λ on $\mathfrak{B}_0$. Theorem 3.1 states that λ can be extended to $\mathfrak{B}$, with the extended λ being the Lebesgue Measure. The following probability space will be our foundation for what follows:

$$\begin{aligned} &\textbf{Space: } \Omega = (0,1], \quad \boldsymbol{\sigma}\textbf{-field: } \mathfrak{B} = \sigma(\mathfrak{B}_0), \ \& \\ &\textbf{Measure: } \lambda(A) = \sum_{i=1}^{n} \lambda(I_i) = \sum_{i=1}^{n} (b_i - a_i) \end{aligned} \tag{3.8}$$

Since the intervals in $(0,1]$ form a π-system generating $\mathfrak{B}$, λ is the only probability measure on $\mathfrak{B}$ that assigns to each interval its length as its measure.
The outer measure corresponding to λ on $\mathfrak{B}_0$ is the infimum of the sums $\sum_n \lambda(A_n)$ for which $A_n \in \mathfrak{B}_0$ and $A \subset \bigcup_n A_n$. Since each A_n is a finite disjoint union of intervals, this outer measure is:

$$\lambda^\star(A) = \mathbf{inf} \sum_n |I_n| \tag{3.9}$$

where the infimum extends over covers of A by intervals I_n. The notion of negligibility can therefore be reformulated:

> A is **negligible** if and only if $\lambda^\star(A) = 0$. For A in $\mathfrak{B}$, this is the same thing as $\lambda(A) = 0$. This covers the set N of normal numbers: since the complement N^c is negligible and lies in $\mathfrak{B}$, $\lambda(N^c) = 0$. Therefore, the Borel set N itself has probability 1, that is $\lambda(N) = 1$

Below are nested collections of subsets of $\mathbb{R}$: intervals, Borel sets $\mathcal{B}$, Lebesgue measurable sets $\mathcal{L}$, and the power set $2^{\mathbb{R}}$, with annotations indicating whether each collection is a σ-algebra and whether the measure μ is countably additive.

Intervals
Not a σ-algebra ✗
$I(I) = \mu^\star(I) = \mu(I)$

$\mathcal{B}$: **Borel Sets**
σ-algebra ✓
$\mu(E) = \mu^\star(E)$
μ countably additive ✓

$\mathcal{L}$: **Lebesgue Measurable Sets**
σ-algebra ✓
$\mu(E) = \mu^\star(E)$
μ countably additive ✓

$2^{\mathbb{R}}$: **all subsets of** $\mathbb{R}$
σ-algebra ✓
$\mu^\star(E)$ **Not** countably additive. ✗

⇒ **Intervals:** $\mu^\star(I) = \mu(I)$ for any interval I, since intervals are Lebesgue measurable and the outer measure agrees with the measure.

⇒ **Borel Sets $\mathcal{B}$:** Sets generated from open intervals and are measurable.

$$\mu(E) = \mu^\star(E) \text{ for all } E \in \mathcal{B} \quad \Rightarrow \quad \text{Countable Additivity Holds.}$$

⇒ **Lebesgue Measurable Sets $\mathcal{L}$:** Sets satisfying Carathéodory's criterion:

$$\mu^\star(A) = \mu^\star(A \cap E) + \mu^\star(A \setminus E), \quad \forall A \subseteq \mathbb{R}. \text{ Then } \mu^\star(E) = \mu(E).$$

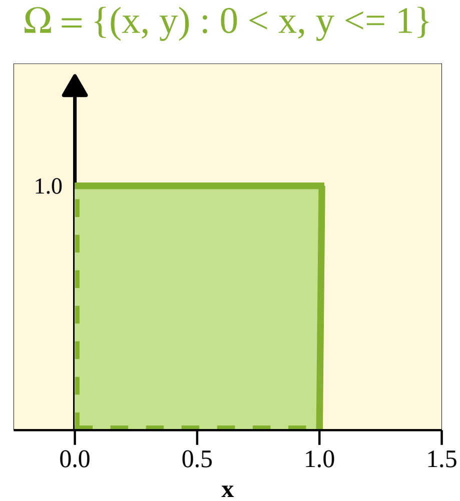
For such sets E, the outer measure $\mu^\star(E)$ coincides with the Lebesgue measure $\mu(E)$.

⇒ $2^{\mathbb{R}}$ **(all subsets of $\mathbb{R}$):** $\mu^\star(E)$ is defined for every E, but many are not measurable. Countable additivity fails for non-measurable sets.

Lebesgue Measure & Borel Sets Example

Exercise 3.5

Let Ω be the unit square $\{(x, y) : 0 < x, y \leq 1\}$.
Let $\mathcal{F}$ be the class of sets of the form $\{(x, y) : x \in A,\ 0 < y \leq 1\}$, where $A \in \mathfrak{B}$.
↳ ($\mathfrak{B}$ is the Borel σ-algebra on the interval $(0, 1]$.)
Let P have value $\lambda(A)$ at this set. ($\lambda(A)$ is the Lebesgue Measure.)



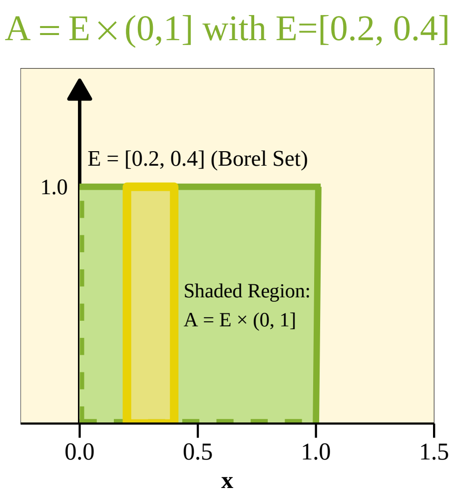
$$\Omega = \{(x, y) : 0 < x, y \leq 1\}$$
$$\mathcal{F} = \{(x, y) : x \in A,\ 0 < y \leq 1 \,|\, A \in \mathfrak{B}\}$$
$$P(A) = \lambda(A) \text{ for } A \in \mathcal{F}$$

a. Show that $(\Omega, \mathcal{F}, P)$ is a probability measure space.

In order to show that $(\Omega, \mathcal{F}, P)$ is a probability measure space, we need to show that $\mathcal{F}$ is a σ-field on Ω and that P is a probability measure on $\mathcal{F}$.

Each set $A \in \mathcal{F}$ is of the form:

$$A = \{(x, y) \in \Omega : x \in E\} = E \times (0, 1] \quad \text{for some Borel set } E \subset (0, 1].$$



The shaded region to the left represents such a set in $\mathcal{F}$, where $E = [0.2, 0.4] \in \mathfrak{B}$. This vertical strip includes every point $(x, y) \in \Omega$ such that $x \in E$ and $0 < y \leq 1$.

Thus, we can generalize this and say that every set in $\mathcal{F}$ is of the form:

$$A = E \times (0, 1] \quad \text{for some } E \in \mathfrak{B}.$$

These are the sets that our probability measure is defined on.

Our probability measure P is defined on sets $A = E \times (0, 1] \in \mathcal{F}$, where $E \in \mathfrak{B}$ is a Borel subset of $(0, 1]$. We define P in terms of the Lebesgue measure λ on $(0, 1]$ as:

$$P(A) = \lambda(E).$$

That is, the measure P assigns to each vertical strip A the length of its base set E. Intuitively, since the strip A spans the entire height of the unit square from $y = 0$ to $y = 1$, its measure corresponds to the one-dimensional Lebesgue measure of the set of x-values it covers.

Formally, this corresponds to defining a product measure:

$$\lambda_1(E \times (0,1]) = \lambda(E) \cdot \lambda((0,1]) = \lambda(E) \cdot 1 = \lambda(E),$$

Thus, λ_1 denotes the product measure formed from the Lebesgue measure on $(0,1]$ applied to the product space $(0,1] \times (0,1]$.

In other words, the measure P on $\mathcal{F}$ is consistent with the one-dimensional Lebesgue measure of the projection of A onto the x-axis.

$\mathcal{F}$ is a σ-Field on Ω:

$\Omega \in \mathcal{F}$: $\Omega = (0,1] \times (0,1] = E \times (0,1]$ where $E = (0,1] \in \mathfrak{B}$, so $\Omega \in \mathcal{F}$.

$\mathcal{F}$ is closed under complements: Let $A = E \times (0,1] \in \mathcal{F}$ for some $E \in \mathfrak{B}$. Then

$$\Omega \setminus A = E^c \times (0,1], \text{ where } E^c \in \mathfrak{B} \Rightarrow A^c \in \mathcal{F}.$$

$\mathcal{F}$ is closed under countable unions: Let $A_n = E_n \times (0,1] \in \mathcal{F}$ with $E_n \in \mathfrak{B}$. Then

$$\bigcup_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} E_n\right) \times (0,1], \text{ which is in } \mathcal{F}.$$

P is a Probability Measure on $\mathcal{F}$:

$P(\Omega) = 1$: $P(\Omega) = P((0,1] \times (0,1]) = \lambda((0,1]) = 1.$

P is countably additive: Let $A_n = E_n \times (0,1]$ with $E_n \in \mathfrak{B}$. Then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\left(\bigcup_{n=1}^{\infty} E_n\right) \times (0,1]\right) = \lambda\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

QED

b. Show for $A = \{(x,y) : 0 < x \leq 1,\ y = \frac{1}{2}\}$ that $P_\star(A) = 0$ and $P^\star(A) = 1$.
$\Rightarrow$*Recall:*

For each subset A of Ω, we define its **Outer Measure** as the following:

$$P^\star(A) = \mathbf{inf} \sum_n P(A_n) \qquad \textbf{Outer Measure}$$

where the infimum is taken over all countable collections of sets $A_n \in \mathcal{F}$ such that $A \subseteq \bigcup_{n=1}^{\infty} A_n$.
Conversely, we define the **Inner Measure** as:

$$P_\star(A) = 1 - P^\star(A^c) \qquad \textbf{Inner Measure}$$

Let $A = \{(x,y) : 0 < x \leq 1,\ y = \frac{1}{2}\}$.
Step 1: Show that $P^\star(A) = 1$

We can cover A with a countable collection of sets $A_n = (0, \frac{1}{n}] \times \left\{\frac{1}{2}\right\}$ for $n = 1, 2, \ldots$. Then,

$$P(A_n) = \lambda\left(\left(0, \frac{1}{n}\right]\right) = \frac{1}{n}.$$

Thus,

$$P^\star(A) = \inf \sum_{n=1}^{\infty} P(A_n) = \inf \sum_{n=1}^{\infty} \frac{1}{n} = 1.$$

Step 2: Show that $P_\star(A) = 0$

We can cover A with a countable collection of sets $A_n = (0, \frac{1}{n}] \times \left\{\frac{1}{2}\right\}$ for $n = 1, 2, \ldots$. Then,

$$P(A_n) = \lambda\left(\left(0, \frac{1}{n}\right]\right) = \frac{1}{n}.$$

Thus,

$$P_\star(A) = 1 - P^\star(A^c) = 1 - \inf \sum_{n=1}^{\infty} P(A_n) = 0.$$

QED

Non-Measurable Sets

There exists sets in $(0, 1]$ that lie outside $\mathfrak{B}$.

Proof (by Vitali Set):

Begin by defining a custom operator that uses addition modulo 1 in (0,1].

$$x \oplus y = \begin{cases} x + y & \text{if } x + y \leq 1 \\ x + y - 1 & \text{if } x + y > 1 \end{cases}$$

Let $\mathcal{L}$ be the class of Borel sets $A \subset (0, 1]$ such that for all $x \in (0, 1]$, the translated set $A \oplus x$ is Borel and

$$\lambda(A \oplus x) = \lambda(A).$$

Then $\mathcal{L}$ is a λ-system containing the intervals. Since intervals form a π-system generating the Borel σ-field $\mathcal{B}$, the π-λ theorem implies $\mathcal{B} \subseteq \mathcal{L}$. Thus, Lebesgue measure λ is translation-invariant.

Define $x \sim y$ if $x = y \oplus r$ for some rational $r \in (0, 1]$. This partitions $(0, 1]$ into disjoint equivalence classes.

Use the axiom of choice to pick a set $H \subset (0, 1]$ containing exactly one representative from each equivalence class under $\sim$.

For each rational $r \in (0, 1] \cap \mathbb{Q}$, define $H \oplus r = \{h \oplus r : h \in H\}$. Then

$$(0, 1] = \bigcup_{r \in \mathbb{Q} \cap (0,1]} H \oplus r,$$

is a countable disjoint union.

If H were measurable, then so would each $H \oplus r$ be, and by countable additivity,

$$\lambda((0, 1]) = \sum_{r \in \mathbb{Q} \cap (0,1]} \lambda(H \oplus r).$$

But all these sets are translates of H, so each term equals $\lambda(H)$. The sum is either:

$\Rightarrow$ 0, if $\lambda(H) = 0$, contradicting $\lambda((0, 1]) = 1$,
or
$\Rightarrow \infty$, if $\lambda(H) > 0$, also a contradiction.

Thus, no such $\lambda(H)$ can exist, and we conclude that $H \notin \mathcal{B}$. As a result, there exist sets $H \subset (0, 1]$ that are not Lebesgue measurable. **QED**

1.4 - Denumerable Probabilities

We will now further scrutinize the details concerning infinite sequences of events in a probability space. The following examples are given in the **Unit interval**, meaning that $(\Omega, \mathcal{F}, P)$ is the triple where

$$\Omega = (0,1]$$
$$\mathcal{F} = \mathfrak{B} = \sigma(\mathfrak{B}_0)$$
$$P(A) = \lambda(A) \text{ for } A \in \mathcal{F}$$

However, the definitions and theorems in the subsequent section apply to *all* probability spaces.

General Formulas

If $P(A) > 0$, the **Conditional Probability** of B given A is defined familiarly:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{4.1}$$

The following define the **Chain Rule Formulas**:

$$\begin{gathered} P(A \cap B) = P(A)P(B|A), \\ P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap P) \\ \vdots \end{gathered} \tag{4.2}$$

If $A_1, A_2, \ldots$ partition Ω, then $$P(B) = \sum_n P(A_n)P(B|A_n) \tag{4.3}$$

For fixed A the function $P(B|A)$ defines a probability measure as B varies over $\mathcal{F}$.

Limit Sets

For a sequence $A_1, A_2, \ldots$ of sets, define the sets

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad \textbf{(Limit Superior of } \boldsymbol{A_n}\textbf{)} \tag{4.4}$$

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k. \quad \textbf{(Limit Inferior of } \boldsymbol{A_n}\textbf{)} \tag{4.5}$$

Example: Consider the functions $d_n(\omega)$ defined on the unit interval by the dyadic expansion (1.1), and let $\boldsymbol{\ell}(\omega)$ be the length of the run of 0's starting at $d_n(\omega)$:

$$\boldsymbol{\ell}_n(\omega) = k \text{ if both } \begin{array}{l} d_n(\omega) = \cdots = d_{n+k-1}(\omega) = 0 \\ \text{and} \quad d_{n+k}(\omega) = 1 \end{array} \quad \Rightarrow \ \boldsymbol{\ell}_n(\omega) = 0 \text{ if } d_n(\omega) = 1$$

Probabilities can be computed using the following formula:

$$P\{\omega : d_i(\omega) = u_i,\ i = 1, \ldots, n\} = \frac{1}{2^n} \tag{4.6}$$

Since $\{\omega : \boldsymbol{\ell}_n(\omega) = k\}$ is a union of 2^{n-1} disjoint intervals of length 2^{-n-k}, it lies in $\mathcal{F}$ and has probability 2^{-k-1}. Therefore, $\{\omega : \boldsymbol{\ell}_n(\omega) \geq r\} = \{\omega : d_i(\omega) = 0,\ n \leq i < n + r\}$ lies also in $\mathcal{F}$ and has probability $\sum_{k \geq r} 2^{-k-1}$.

$$P(\omega : \boldsymbol{\ell}_n(\omega) \geq r) = 2^{-r}. \tag{4.7}$$

If A_n is the event in (4.7), then (4.4) is the set of ω such that $l_n(\omega) \geq r$ for infinitely many n, or, n being regarded as a time index, such that $\boldsymbol{\ell}_n(\omega) \geq r$ *infinitely often.*

Independent Events

Events A and B are **Independent** if $P(A \cap B) = P(A)P(B)$. For events with positive probability, this is the same thing as requiring $P(B|A) = P(B)$ or $P(A|B) = P(A)$. More generally, a finite collection $A_1, ..., A_n$ of events is independent if:

$$P(A_{k_1} \cap \cdots \cap A_{k_j}) = P(A_{k_j}) \cdots P(A_{k_j}) \tag{4.7}$$

for $2 \leq j \leq n$ and $1 \leq k_1 < \cdots < k_j \leq n$. Reordering the sets clearly has no effect on the condition for independence, and a subcollection of independent events is also independent. An infinite (perhaps uncountable) collection of events is defined to be independent in each of its finite subcollections is.

$$\text{If } n = 3 \begin{cases} \text{for } j = 2 \text{ the three constraints are:} \\ \qquad P(A_1 \cap A_2) = P(A_1)P(A_2) \\ \qquad P(A_1 \cap A_3) = P(A_1)P(A_3) \\ \qquad P(A_2 \cap A_3) = P(A_2)P(A_3) \\ \text{and if } j = 3 \text{ the single constraint is:} \\ \qquad P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \end{cases} \tag{4.8}$$

Theorem 4.2: If $\mathcal{A}_1, ..., \mathcal{A}_n$ are independent and each $\mathcal{A}_i$ is a π-system, them, $\sigma(\mathcal{A}_1), ..., \sigma\mathcal{A}_n$ are independent.

Proof:

Let $\mathcal{B}_i$ be the class $\mathcal{A}_i$ augmented by Ω. Then for each $\mathfrak{B}_i$ is a π-system, then the independence hypothesis holds if $B_i \in \mathfrak{B}_i,\ i = 1, ..., n$. For fixed sets $B_2, ..., B_n$ lying respectively in $\mathfrak{B}_2, ..., \mathfrak{B}_n$, let $\mathcal{L}$ be the class of $\mathcal{F}$-sets B_1 for which the hypothesis of independence holds. Then $\mathcal{L}$ is a λ-system containing the π-system $\mathfrak{B}_1$ and hence containing $\sigma(\mathfrak{B}_1) = \sigma(\mathcal{A}_1$. Therefore, they are independent if $B_1, B_2, ..., B_n$ lie respectively in $\sigma(\mathcal{A}_1), \mathfrak{B}_2, ..., \mathfrak{B}_n$, which means that $\sigma(\mathcal{A}_1), \mathcal{A}_2, ..., \mathcal{A}_n$ are independent. Clearly, the argument holds for any other choice of index that is less than or equal to n. **QED**

- **Corollary 1:** If $\mathcal{A}_\theta$, $\theta \in \Theta$ are independent and each $\mathcal{A}_\theta$ is a π-system, then $\sigma(\mathcal{A}_\theta)$, $\theta \in \Theta$, are independent.
- **Corollary 2:** Suppose that the array $\Rightarrow$ of events is independent; here each row is a finite or infinite sequence, and there are finitely or infinitely many rows.

$$\begin{matrix} A_{11} & A_{12} & \cdots \\ A_{21} & A_{22} & \cdots \\ \vdots & \vdots & \ddots \end{matrix} \tag{4.9}$$

If $\mathcal{F}_i$ is the σ-field generated by the ith row, then $\mathcal{F}_1, \mathcal{F}_2, \ldots$ are independent.

Subfields

The definition of independence provided above involves several different σ-fields at once, which is not all that uncommon to see within probability theory. However, measure theory for its own sake typically involves a single all-encompassing σ-field $\mathcal{F}$. A subclass $\mathcal{A}$ of $\mathcal{F}$ corresponds to **partial information**.

Imagine that a point ω is drawn from Ω according to the probabilities given by P ("ω lies in A with probability $P(A)$") Imagine also that an observer who does not know which ω has been drawn but who does know for each $A \in \mathcal{A}$ whether $\omega \in A$ or $\omega \notin A$ (meaning that they do not know ω but they do know the values of $I_A\omega$ for each $A \in \mathcal{A}$). Identifying this partial information with the class $\mathcal{A}$? itself will illuminate the connection between various measure-theoretic concepts and the premathematical ideas lying behind them.

The set B is by definition independent of the class $\mathcal{A}$ if $P(A|B) = P(B)$ for all sets $A \in \mathcal{A}$ for which $P(A) > 0$. Thus if B is independent of $\mathcal{A}$, then the observer's probability for B is $P(B)$ even after he has received the information in $\mathcal{A}$; in this case $\mathcal{A}$ contains no information about B. This remains true even if the observer has all the information contained in $\sigma(\mathcal{A})$.

The notion of partial information can be looked at in terms of partitions. We say that points ω and ω' are **$\mathcal{A}$-equivalent**, if for every A in $\mathcal{A}$, ω and ω' lie either both in A or both in A^c, that is if:

$$I_A(\omega) = I_A(\omega'), \quad \forall A \in \mathcal{A} \tag{4.10}$$

This relation partitions Ω into sets of equivalent points; call this the **$\mathcal{A}$-partition**.

Example: If ω and ω' are $\sigma(\mathcal{A})$-equivalent, then certainly they are $\mathcal{A}$-equivalent. For fixed ω and ω', the class of A such that $I_A(\omega) = I_A(\omega')$ is a σ-field; if ω and ω' are $\mathcal{A}$-equivalent, then this σ-field contains $\mathcal{A}$ and hence $\sigma(\mathcal{A})$, so that ω and ω' are also $\sigma(\mathcal{A})$-equivalent. Thus $\mathcal{A}$-equivalence and $\sigma(\mathcal{A})$-equivalence are the same thing, and the $\mathcal{A}$-partition coincides with the $\sigma(\mathcal{A})$-partition. **QED**

An observer with the information in $\sigma(\mathcal{A})$ knows, not the point ω drawn, but only the equivalence class containing it. That is exactly the information they have. One who knows that ω lies in a set A has more information about ω the smaller A is. One who knows $I_A(\omega)$ for each A in class $\mathcal{A}$, however has more information about ω the larger $\mathcal{A}$ is. Furthermore, to have the information that $\mathcal{A}_1$ and the information in $\mathcal{A}_2$ is to have the information in $\mathcal{A}_1 \cup \mathcal{A}_2$, not that in $\mathcal{A}_1 \cap \mathcal{A}_2$.

Example: In the unit interval $(\Omega, \mathcal{F}, P)$ let $\mathcal{G}$ be the σ-field consisting of the countable and the co-countable sets. Since $P(G)$ is 0 or 1 for each G in $\mathcal{G}$, each set H in $\mathcal{F}$ is independent of $\mathcal{G}$. But in this case the $\mathcal{G}$-partition consists of the singletons, and so the information in $\mathcal{G}$ tells the observer exactly which ω in Ω has been drawn.

i. The σ-field $\mathcal{G}$ contains *no* information about H—in the sense that H and $\mathcal{G}$ are independent.

ii. The σ-field $\mathcal{G}$ contains *all* the information about H—in the sense that it tells the observer exactly which ω was drawn. **QED**

The two conclusions of the above example appear to stand in direct contradiction of one another. However, the first conclusion (**i**) is mathematically rigorous – H and $\mathcal{G}$ are independent – whereas the second conclusion (**ii**) only concerns a particular interpretation. This is a result of the unusual structure of the σ-field $\mathcal{G}$ rather than a shortcoming in the definition of independence. The heuristic of equating σ-fields and information is helpful, even if sometimes it breaks down. It is important to remember that the definition of independence takes precedence when the heuristic interpretation appears to contradict it.

The Borel-Cantelli Lemmas

The First Borel-Cantelli Lemma
Theorem 4.3: if $\sum_n P(A_n)$ converges, then $P(\mathbf{lim\,sup}_n\ A_n) = 0$
Proof:

$$\limsup_n A_n \subset \bigcup_{k=m}^{\infty} A_k \ = \text{implies} \Rightarrow \ P(\limsup_n A_n) \le P\left(\bigcup_{k=m}^{\infty} A_k\right) \le \sum_{k=m}^{\infty} P(A_k)$$

and this sum tends to 0 as $m \to \infty$ if $\sum_n P(A_n)$ converges. **QED**

The Second Borel-Cantelli Lemma
Theorem 4.4: if $\{A_n\}$ is an independent sequence of events and $\sum_n P(A_n)$ diverges, then $P(\mathbf{lim\,sup}_n\ A_n) = 1$
Proof:

It is enough to prove that $P\left(\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k^c\right) = 0$ and hence enough to prove that $P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = 0$ for all n. Since $1 - x \le e^{-x}$,

$$P\left(\bigcap_{k=n}^{n+j} A_k^c\right) = \prod_{k=n}^{n+j}(1 - P(A_k)) \le \exp\left[-\sum_{k=n}^{n+j} P(A_k)\right].$$

Since $\sum_k P(A_k)$ diverges, the last expression tends to 0 as $j \to \infty$, and hence $P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_j P\left(\bigcap_{k=n}^{n+j} A_k^c\right) = 0.$ **QED**

The Zero-One Law For a sequence $A_1, A_2, ...$ of events in a probability space $(\Omega, \mathcal{F}, P)$ consider the σ-fields $\sigma(A_n, A_{n+1}, ...)$ annd their intersection

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, ...) \tag{4.11}$$

This is called the **tail σ-field** associated with the sequence $\{A_n\}$, and its elements are called **tail events**. The idea is that a tail event is determined solely by the A_n for arbitrarily large n.

Theorem 4.5 (Kolmogorov's zero-one law): If $A_1, A_2, \dots$ is an independent sequence of events, then for each event A in the tail σ-field (4.10), $P(A)$ is either 0 or 1.

Proof:

> By Corollary 2 to Theorem 4.2, $\sigma(A_1), \dots, \sigma(A_{n-1}), \sigma(A_n, A_{n+1}, \dots)$ are independent. If $A \in \mathcal{T}$, then $A \in \sigma(A_n, A_{n+1}, \dots)$ and therefore $A_1, \dots, A_{n-1}, A$ are independent. Since independence of a collection of events is defined by independence of each finite subcollection, the sequence $A, A_1, A_2, \dots$ is independent. By a second application of Corollary 2 to Theorem 4.2, $\sigma(A)$ and $\sigma(A_1, A_2, \dots)$ are independent. But $A \in \mathcal{T} \subset \sigma(A_1, A_2, \dots)$; from $A \in \sigma(A)$ and $A \in \sigma(A_1, A_2, \dots)$ it follows that A is independent of itself: $P(A \cap A) = P(A)P(A)$. This is the same as $P(A) = (P(A))^2$ and can hold only if $P(A)$ is 0 or 1. **QED**

1.5 - Simple Random Variables

Let $(\Omega, \mathcal{F}, P)$ be an arbitrary probability measure space, and let X be a real-valued function on Ω. X is a **Simple Random Variable** if it has finitely many possible values, and if it satisfies:

$$[\omega : X(\omega) = x] \in \mathcal{F} \quad \text{for all } x \tag{5.1}$$

To actually decide if X satisfies this condition we really only need to look at $\mathcal{F}$ rather than P, but before we can do that we have to ensure that $P[\omega : X(\omega) = x]$ is defined. Our previously discussed $d_n(\omega)$, the function that characterizes the digits of a dyadic expansion, is an example of a simple random variable on the unit interval. For ease of notation it is very common in probability theory to drop the trailing (ω) in $X(\omega)$, as a result X is used to denote the general value of $[w : X(\omega) = x]$. A finite sum of the form:

$$X = \sum_i x_i I_{A_i} \tag{5.2}$$

is a random variable if the A_i form a finite partition of Ω into $\mathcal{F}$-sets. Furthermore, all simple random variables can be represented as a finite sum of the above form. Simply take the range of X and for each x_i let $A_i = [X = x_i]$. However it is often more natural to replace $x_i I_{A_i}$ with $\sum_j x_i I_{A_{ij}}$ when the A_{ij} form a finite decomposition of A_i into $\mathcal{F}$-sets.

The σ-field $\sigma(X)$ generated by X is the smallest σ-field with respect to which X is measurable; that is $\sigma(X)$ is the intersection of all σ-fields with respect to which X is measurable. For a sequence $X_1, X_2, ...$ of simple random variables , $\sigma(X_1, X_2, ...)$ is the smallest σ-field with respect to which each X_i is measurable.

Theorem 5.1: Let $X_1, ..., X_n$ be simple random variables.

i. The σ-field $\sigma(X_1, ..., X_n)$ consists of the sets:

$$[\ (X_1, ..., X_n) \in H\ ] = [\ \omega : (X_1(\omega), ..., X_n(\omega)) \in H\ ] \tag{5.3}$$

for $H \subset \mathbb{R}^n$; H may be taken to be finite in this representation.

ii. A simple random variable Y is measurable $\sigma(X_1, ..., X_n)$ if and only if:

$$Y = f(X_1, ..., X_n) \quad \text{for some } f : \mathbb{R}^n \to \mathbb{R}^1 \tag{5.4}$$

Proof:

Let $\mathcal{M}$ be the class of sets of the form (5.3):
$[\ (X_1, ..., X_n) \in H\ ] = [\ \omega : (X_1(\omega), ..., X_n(\omega)) \in H\ ]$. Sets of the form:
$[(X_1, \ldots, X_n) = (x_1, \ldots, x_n)] = \bigcap_{i=1}^{n} [X_i = x_i]$ must line in $\sigma(X_1, \ldots, X_n)$; each set (5.3) is a finite union of sets of this form because $(X_1, \ldots, X_n)$, as a mapping from Ω to $\mathbb{R}^n$, has finite range. Thus $\mathcal{M} \subset \sigma(X_1, \ldots, X_n)$.

On the other hand, $\mathcal{M}$ is a σ-field because:

i. $\Omega = [(X_1, \ldots, X_n) \in \mathbb{R}^n]$

ii. $[(X_1, \ldots, X_n) \in H]^c = [(X_1, \ldots, X_n) \in H^c]$,

iii. $\bigcup_j [(X_1, \ldots, X_n) \in H_j] = [(X_1, \ldots, X_n) \in \bigcup_j H_j]$.

But each X_i is measurable with respect to $\mathcal{M}$, because $[X_i = x]$ can be put in the form of (5.3) by taking H to consist of those $(x_1, ..., x_n)$ in $\mathbb{R}^n$ for which $x_i = x$. It follows $\sigma(X_1, ..., X_n)$ is contained in $\mathcal{M}$ and therefore equals $\mathcal{M}$. As intersecting H with the (finite) range of $(X_1, ..., X_n)$ in $\mathcal{R}^n$ does not affect (5.3), H may be taken to be finite. This proves **i.**

Assume that Y has the form (5.4)—that is, $Y(\omega) = f(X_1(\omega), \ldots, X_n(\omega))$, for every ω. Since $[Y = y]$ can be put in the form (5.3) by taking H to consist of those $x = (x_1, \ldots, x_n)$ for which $f(x) = y$, it follows that Y is measurable $\sigma(X_1, \ldots, X_n)$.

Now assume that Y is measurable $\sigma(X_1, \ldots, X_n)$. Let $y_1, \ldots, y_r$ be the distinct values Y assumes. By part (i), there exist sets $H_1, \ldots, H_r$ in $\mathbb{R}^n$ such that:
$[\omega : Y(\omega) = y_i] = [\omega : (X_1(\omega), \ldots, X_n(\omega)) \in H_i]$. Take $f = \sum_{i=1}^{r} y_i I_{H_i}$. Although the H_i need not be disjoint, if H_i and H_j share a point of the form

$(X_1(\omega), \ldots, X_n(\omega))$, then $Y(\omega) = y_i$ and $Y(\omega) = y_j$, which is impossible if $i \neq j$. Therefore each $(X_1(\omega), \ldots, X_n(\omega))$ lies in exactly one of the H_i, and it follows that $f(X_1(\omega), \ldots, X_n(\omega)) = Y(\omega)$. **QED**

This theorem tells us that functions of simple random variables are themselves simple random variables; that is a Y of the form (5.4) is measurable $\sigma(X_1, ..., X_n)$, it follows that X^2, e^{tX}, and so one are simple random variables.

Convergence of Random Variables

Theorem 5.2:

i. There is convergence $\lim_{n \to p} X_n = X$ with probability 1 if and only if:

$$P[\ |X_n - X| \geq \varepsilon \text{ i.o. }] = 0 \quad \text{(where 'i.o.' stands for } \textit{infinitely often} \text{)}. \tag{5.5}$$

ii. Convergence with probability 1 implies convergence in probability.

Proof:

Part i: We characterize the event that $X_n \not\to X$ almost surely. That occurs if and only if there exists some $\varepsilon > 0$ such that the deviation $|X_n - X| \geq \varepsilon$ occurs infinitely often. That is,

$$\left[\lim_n X_n = X\right]^c = \bigcup_{\varepsilon > 0} [|X_n - X| \geq \varepsilon \text{ i.o.}], \tag{5.6}$$

Since the right-hand side is a countable union (we can restrict ε to raionals), it is measurable and belongs to the basic σ-field $\mathcal{F}$. Taking complements, we obtain:

$$\left[\lim_n X_n = X\right] = \bigcap_{\varepsilon > 0} [|X_n - X| < \varepsilon \text{ eventually}], \tag{5.7}$$

which has probability 1 if and only if $P\,[|X_n - X| \geq \varepsilon \text{ i.o.}] = 0$.for all $\varepsilon > 0$

This completes the proof of part i.

Part ii: Suppose $X_n \to X$ with probability 1. Then for every $\varepsilon > 0$, we have

$$P\,[|X_n - X| \geq \varepsilon \text{ i.o.}] = 0.$$

The event on the left is the limit superior of the sequence of events $\{|X_n - X| \geq \varepsilon\}$. Therefore, by Theorem 4.1 (as referenced), we conclude:

$$\lim_{n \to \infty} P\,(|X_n - X| \geq \varepsilon) = 0.$$

This is precisely the definition of convergence in probability. Hence, convergence with probability 1 implies convergence in probability. **QED**

Example of Divergence

Take $X = 0$ and $X_n = I_{A_n}$. Then $X_{n \to p} X$ is equivalent to $P(A_n) \to 0$, and $[\ \lim_n X_n = X\ ]^c = [\ A_n \text{ i.o. }]$. Any sequence $\{A_n\}$ such that $P(A_n) \to 0$ but $P[\ A_n \text{ i.o. }] > 0$ therefore gives a counter example to the converse of theorem 5.2.ii.