Discrete-time Stochastic Processes Module 1

MSc Financial Engineering

Table of Contents

1. Brief

This document contains the core content for Module 1 of Discrete-time Stochastic Processes, entitled Probability Theory. It consists of six sets of notes, four lecture transcripts, and a collaborative review task.

2. Course Context

Discrete-time Stochastic Processes is the third course presented in the WorldQuant University (WQU) Master of Science in Financial Engineering (MScFE) program. The course introduces derivative pricing in discrete time. It begins with measure-theoretic probability and stochastic processes, with an emphasis on discrete-time martingales. These ideas are then applied to the pricing of derivatives in discrete time, followed by an introduction to interest rate and credit risk modeling.

2.1 Course-level Learning Outcomes

After completing the Discrete-time Stochastic Processes course, you will be able to:

1. Understand the language of measure-theoretic probability.
2. Understand stochastic processes and their applications.
3. Understand the theory of discrete-time martingales.
4. Define trading strategies in discrete time.
5. Create replicating portfolios in discrete time.
6. Model stock price movements on a binomial tree.
7. Price and hedge European derivatives in discrete time.
8. Price and hedge exotic European derivatives in discrete time.
9. Price and hedge American derivatives on a binomial tree.
10. Construct a simple interest rate model on a tree.
11. Price interest rate derivatives on a tree.

2.2 Module Breakdown

The Discrete-time Stochastic Processes course consists of the following one-week modules:

1. Probability Theory
2. Stochastic Processes
3. Discrete Martingales
4. Trading in Discrete Time
5. The Binomial Model
6. American Derivatives
7. An Introduction to Interest Rate Models

3. Module 1:

Probability Theory

Module 1 introduces the measure-theoretic foundations of probability. It begins with an introduction to sigma-algebras and measures, then proceeds to introduce random variables and measurable functions, before concluding with Lebesgue integration and the Radon-Nikodym theorem.

3.1 Module-level Learning Outcomes

After completing this module, you will be able to:

1. Understand the language of measure-theoretic probability.
2. Define random variables, measurable functions and the Lebesgue integral.
3. Calculate integrals with respect to certain measures.
4. Apply the Radon-Nikodym theorem in changing measures.

3.2 Transcripts and Notes

3.2.1 Transcript: Sigma-Algebras and Measures

Hi, in this video, I'll be introducing the notion of **probability theory**, which is going to be one of the most important concepts in the course.

The study of probability theory begins with what we call a **random experiment**. A random experiment is just an experiment whose outcome cannot be predetermined. An example of such an experiment is the tossing of three coins. When tossing three coins, the outcome of this random experiment cannot be predetermined – when we toss the coins, we can never know what the outcome will be.

Each random experiment is associated with what we call the "sample space", which is just a set that represents all the possible outcomes of the random experiment. We denote it by Ω and we call it the set of all possible outcomes of the random experiment.

In our example of tossing three coins, the sample space can be defined as a set that looks like this: $\{\text{HHH},\text{HHT}, \dots ,\text{TTT}\}$ – it contains (HHH), which is the outcome when all three tosses give a head; (HHT) where the first two tosses give a head and the last one gives a tail; and so on, up until we get to (TTT). In total, we will have eight outcomes.

We also define the notion of an **event** in elementary probability as a subset of Ω. For instance, we can look at the event A which is the event that the outcome contains two heads. That event corresponds to the subset of Ω containing $\{\text{HHT}, \text{HTH}, \text{THH}\}$. Unfortunately, though, for the experiments that we are going to consider, not all subsets of Ω correspond to events. What we will need to do, then, is restrict the collection of events to a sub-collection, F which is a subset of the powerset of Ω.

This sub-collection will contain all the events that we are interested in considering and it should satisfy the following conditions:

> $\mathcal{A}1$ Firstly, it should contain the empty set, meaning that the empty set is always an event no matter what experiment you're looking at.
>
> $\mathcal{A}2$ The second condition is that if a set A is an event – a in other words if A belongs to $\mathcal{F}$, then its complement is also an event – i.e. it also belongs to F.
>
> $\mathcal{A}3$ The third condition is that if a set A and another set B are both events in $\mathcal{F}$ then their union is also an event in $\mathcal{F}$.
>
> $\sigma\mathcal{A}4$ The last condition, which we call $\sigma\mathcal{A}4$, is that if you have a countable collection of events – so if $\mathcal{A}1$, $\mathcal{A}2$, $\mathcal{A}3$, and so on, are all events in $\mathcal{F}$ – then their union is also an event in $\mathcal{F}$.

Any collection of sets that satisfies these four conditions is called a sigma-algebra (σ-algebra). So, the collection of events should form a σ-algebra. If a sub-collection satisfies only $\mathcal{A}1$, $\mathcal{A}2$, and $\mathcal{A}3$, we call it an algebra. Therefore, a σ-algebra is simply an algebra with the additional condition that if you take countably many events in $\mathcal{F}$, their union also belongs to $\mathcal{F}$.

We are going to call them the pair $(\Omega, \mathcal{F})$ where Ω is any set and $\mathcal{F}$ is a σ-algebra of subsets of Ω. We will call this a **measurable space**.

Now let $(\Omega, \mathcal{F})$ be a measurable space, which means that Ω is a set and $\mathcal{F}$ is a σ-algebra. We define $\mathbb{P}$ to be a probability measure if $\mathbb{P}$ is a function from Ω to $[0, 1]$. What that means is that it assigns to each event A a number between 0 and 1 and satisfies the following conditions:

> $\mathbb{P}1$ The first condition is that $\mathbb{P}$ of the empty set must be 0. Thus, the probability of the empty set is 0.
>
> $\mathbb{P}2$ The second condition is that $\mathbb{P}$ of any event A is always greater than or equal to 0.
>
> $\mathbb{P}3$ The third condition is that if you have a countable collection of pairwise disjoint events – so if $\mathcal{A}1$, $\mathcal{A}2$, and so on are pairwise disjoint events – then $\mathbb{P}$ of their union is the sum of $\mathbb{P}$ of each one of them, from n is 1 to infinity. This is an important condition that we call σ-**additivity**.
>
> $\mathbb{P}4$ The last condition that must be satisfied by probability is that $\mathbb{P}(\Omega) = 1$. This simply means that it assigns probability 1 to the sample space.
>
> Another way to visualize this is on a number line between 0 and 1. Probability in each event A assigns a real number between 0 and 1 in such a way that that assignment satisfies conditions $\mathbb{P}1$, $\mathbb{P}2$, $\mathbb{P}3$, and $\mathbb{P}4$.
>
> Returning to $\mathbb{P}4$ and what it means, the sample space Ω is assigned probability 1, which we usually translate as saying that it is a certain event.

Now, there are other set functions that satisfy conditions $\mathbb{P}1$, $\mathbb{P}2$, and $\mathbb{P}3$. Some examples of them include **length**: if you think about the lengths of subsets of the real line, the length of the empty set is 0, the length of any subset of the real line should be non-negative, and the length of disjoint subsets of the real line should be the sum of the length. Of course, the length function does not satisfy condition $\mathbb{P}4$ because the length of the real line itself, which is the sample space in this case, is infinite.

Another example is **area**, which also satisfies $\mathbb{P}1$, $\mathbb{P}2$, and $\mathbb{P}3$, with $\mathbb{P}$ representing the corresponding measure and volume.

What we are going to do in this course is study what is called a measure μ, which is just a set function that satisfies $\mathbb{P}1$, $\mathbb{P}2$, and $\mathbb{P}3$, but not necessarily $\mathbb{P}4$. What we will see is that a probability measure will simply be a special case of a measure whereby $\mu(\Omega) = 1$. That field is called Measure Theory.

If you want to go through this at your own pace, please refer to the notes.

3.2.2 Notes: Sigma-Algebras and Measures

The study of probability theory begins with what is called a **random experiment**: an experiment whose outcome cannot be pre-determined with certainty. An example of such a random experiment is the tossing of three fair coins, with the (uncertain) outcome being the face shown by each of them.

We associate with every random experiment a set Ω of all possible outcomes of the random experiment. An outcome ω is simply an element of Ω.

We define an event to be a subset of Ω and we say that an event A occurs if the outcome ω belongs to A, i.e. $\omega \in A$.

In our example of tossing three coins, the sample space can be chosen to be:

$$\Omega = \{\mathrm{HHH}, HHT, \dots, \mathrm{TTT}\} = \{\omega = \omega_1\omega_2\omega_3 : \omega_i = \mathrm{H} \text{ or } \mathrm{T} \text{ for } i = 1,2,3\},$$

where H means "heads" and T means "tails". For instance, the event A that exactly two heads come up corresponds to the set

$$\mathrm{A} = \{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}\}.$$

For complex models, it turns out that we have to restrict the class of events to a sub-collection $\mathcal{F}$ of subsets of Ω that satisfies the following properties:

$\mathcal{A}1.$ $\emptyset \in \mathcal{F}$;

$\mathcal{A}2.$ For every $A \subseteq \Omega$, $A \in \mathcal{F} \implies A^c \in \mathcal{F}$;

$\mathcal{A}3.$ For every $A, B \subseteq \Omega, A, B \in \mathcal{F} \Longrightarrow A \cup B \in \mathcal{F}$;

$\sigma\mathcal{A}4.$ If $\{A_n : n = 1,2,3, \dots\}$ is a collection of subsets of Ω with $A_n \in F$ for each $n \geq 1$, then

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

A collection of subsets of Ω that satisfies $\mathcal{A}1$, $\mathcal{A}2$ and $\mathcal{A}3$ is called an *algebra* in Ω, and an algebra that also satisfies $\sigma\mathcal{A}4$ is called a σ-*algebra*[1]. Thus, the collection of admissible events $\mathcal{F}$ should be a σ-*algebra.*

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

Once we have the collection of events $\mathcal{F}$, we want to assign probabilities to each one of them. We call a set function $\mathbb{P}: \mathcal{F} \to \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ a probability measure if $\mathbb{P}$ satisfies the following:

$\mathbb{P}1$. $\mathbb{P}(\emptyset) = 0$;

$\mathbb{P}2$. $\mathbb{P}(A) \geq 0$ for every $A \in \mathcal{F}$;

$\mathbb{P}3$. If $\{A_n : n = 1,2,3, \dots\}$ is a countable collection of pairwise disjoint events, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)\,;$$

$\mathbb{P}4$. $\mathbb{P}(\Omega) = 1$.

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Upon careful observation, we see that there are several other set functions that behave like a probability. Here are some examples of set functions that satisfy conditions $\mathbb{P}1 - \mathbb{P}3$:

- Length of subsets of the real line. This is the function λ_1 that assigns to certain subsets of $\mathbb{R}$ the "length" of that subset. If $A = (a, b)$ is a bounded interval, then $\lambda_1(A) = b - a$. However, λ_1 is defined on a larger collection than intervals (the

[1] Note that for a σ-algebra, condition $\mathcal{A}3$ can be removed since it is implied by condition $\sigma\mathcal{A}4$.

intervals by themselves do not form a σ-algebra). This *measure*, as we will refer to it, satisfies conditions $\mathbb{P}1 - \mathbb{P}3$, but not $\mathbb{P}4$. Also, $\lambda_1(\Omega) = \lambda_1(\mathbb{R}) = \infty$, hence λ_1 takes values on the *extended real numbers* $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. This measure is called the *Lebesgue measure*, after the French mathematician Henri Lebesgue, who was instrumental in laying the foundations of the field of *Measure Theory*.

- The above example can be extended to arbitrary dimensions to construct the N-*dimensional Lebesgue measure* λ_N on subsets of $\mathbb{R}^N$, where N is a positive integer. For $N = 2$ this corresponds to area in the plane, while λ_3 measures volume in $\mathbb{R}^3$.

- Mass for quantities of matter.

- Cardinality of sets also satisfies the same properties.

Given an arbitrary set Ω and a σ-*algebra* $\mathcal{F}$ of subsets of Ω, we say that a set function $\mu: \mathcal{F} \to \overline{\mathbb{R}}$ is a *positive measure* (or *measure*) on $(\Omega, \mathcal{F})$ if it satisfies $\mathbb{P}1$, $\mathbb{P}2$, and $\mathbb{P}3$ (with μ replacing $\mathbb{P}$, of course). If, in addition, the measure μ also satisfies $P4$, then μ is a *probability measure*. This means that the study of measure theory contains probability theory as a special case! The pair $(\Omega, \mathcal{F})$ is called a *measurable space* and the triple $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*. If μ is a probability measure, then $(\Omega, \mathcal{F}, \mu)$ is called a *probability space*. Thus, on our road to understanding probability theory we are going to take a detour through abstract measure theory. We will prove many important results that hold in general measure spaces and later apply them to probability spaces.

Here are some examples of algebras and σ-algebras. Check that these are indeed algebras or σ-algebras.

1 The "smallest" σ-algebra on any set Ω is $\mathcal{A} := \{\emptyset, \Omega\}$. Any other algebra (or σ-algebra) must necessarily contain it.

2 On the other hand, the "largest" σ-algebra on a set Ω is the power set of Ω, 2^{Ω}. (A power set is a set of all subsets of a set.) Every algebra (or σ-algebra) is contained in 2^{Ω}.

3 Let $A \subseteq \Omega$, then $\mathcal{F}_A = \{A, A^c, \emptyset, \Omega\}$ is a σ-algebra (and, therefore, an algebra) on Ω. We call this the σ-algebra generated by $\{A\}$.

4 For a concrete example, let $\Omega = \{1,2,3,4\}$ and

$$\mathcal{F} = \{\{1\}, \{2,3\}, \{4\}, \{2,3,4\}, \{1,2,3\}, \{1,4\}, \emptyset, \Omega\}.$$

One can easily check that $\mathcal{F}$ is a σ-algebra on Ω.

Notice that in all the examples above the number of elements in each algebra is a power of 2. In general, we say that a σ-algebra $\mathcal{F}$ is generated by blocks if there exists a countable partition $\{\mathcal{B}_n : n = 0,1,2, \dots\}$ of Ω such that

$$\mathcal{F} = \left\{ \bigcup_{n \in J} \mathcal{B}_n : J \subseteq N \right\}.$$

Now we consider more complicated examples.

1 Let $\Omega = \mathbb{R}$ and choose $\mathcal{F} = \{A \subseteq \mathbb{R}: \text{ for every } x \in \mathbb{R},\ x \in A \Leftrightarrow -x \in A\}$ be the collection of all symmetric subsets of $\mathbb{R}$. This is a σ-algebra on $\mathbb{R}$.

2 Let $\Omega = \mathbb{R}$ again and define

$$\mathcal{A} = \{A \subseteq \mathbb{R}: A = \cup_{k=1}^{n} I_k \text{ for some intervals } I_k \subseteq \mathbb{R}, k = 1,2, \dots, n, n \in \mathbb{N}^+\}$$

to be subsets of $\mathbb{R}$ that are finite unions of (open, closed, bounded, empty, etc.) intervals in $\mathbb{R}$. Then $\mathcal{A}$ is an algebra but not a σ-algebra, since (for instance) $A_n := \{n\} = [n, n] \in A$ for each $n \in \mathbb{N}$, yet $\cup_{n \in \mathbb{N}} A_n = N \notin \mathcal{A}$.

3 Let Ω be an infinite set. Define:

$$\mathcal{A} \coloneqq \{A \subseteq \Omega : A \text{ is finite or } A^c \text{ is finite}\}.$$

Then $\mathcal{A}$ is an algebra but not a σ-algebra.

4 Let Ω be an infinite set. Define:

$$\mathcal{F} \coloneqq \{A \subseteq \Omega : A \text{ is countable or } A^c \text{ is countable}\}.$$

Then $\mathcal{F}$ is a σ-algebra.

Here are some examples of measures.

1 **The Dirac Measure**. Let Ω be any non-empty set and pick $a \notin \Omega$. If we let $\mathcal{F} = 2^{\Omega}$ and define $\delta_a : \mathcal{F} \to \mathbb{R}$ by

$$\delta_a(A) \coloneqq \begin{cases} 1 \text{ if } a \in A \\ 0 \text{ otherwise,} \end{cases}$$

then δ_a is a probability measure on $(\Omega, \mathcal{F})$, called the *Dirac measure*.

2 **The Counting Measure**. Let Ω be a set and define $\#: 2^{\Omega} \to \overline{\mathbb{R}}$ by

$$\#(A) \coloneqq \begin{cases} |A| \text{ if } A \text{ is finite} \\ \infty \text{ otherwise ,} \end{cases}$$

where $|A|$ represents the cardinality (number of elements) of A. Then this is a measure on $(\Omega, 2^{\Omega})$, called the *counting measure*.

3 **The Lebesgue Measure** on R. This measure λ_1 is defined on a σ-algebra that includes all the intervals and it extends the notion of length of an interval to such sets. If $I = (a, b)$ is an interval, then

$$\lambda_1(I) = b - a.$$

The N-dimensional analog of Lebesgue measure is a measure λ_N on some subsets of $\mathbb{R}^N$ that extends area for $N = 2$ and volume for $N = 3$.

A property P is said to be true almost everywhere with respect to a measure μ if the set of points where P is false is contained in a set of measure zero. We write $P - \mu$ a.e., and we say $\mu -$ a.s. if μ is a probability measure.

Let $\mathcal{A} \subseteq 2^{\Omega}$ be a collection of subsets of Ω. In general, $\mathcal{A}$ will not be a σ-algebra. What we want is to find the smallest σ-algebra containing $\mathcal{A}$ and we denote it by $\sigma(\mathcal{A})$.

A special σ-algebra is the *Borel σ-algebra* on $\mathbb{R}$ denoted by $B(\mathbb{R})$. It is the σ-algebra generated by intervals in $\mathbb{R}$, i.e.,

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\{I : I \subseteq \mathbb{R} \text{ is an interval}\}\right).$$

The elements of $\mathcal{B}(\mathbb{R})$ are called *Borel* sets. We also define the Borel σ-algebra of a subset $A \subseteq \mathbb{R}$ as the restriction of $\mathcal{B}(\mathbb{R})$ to A; i.e.,

$$\mathcal{B}(A) \coloneqq \{B \cap A : B \in \mathcal{B}(\mathbb{R})\}.$$

3.2.3 Notes: Random Variables

Intuitively, a random variable is a numerical variable whose value is determined by the outcome of a random experiment. This description suggests that we define a random variable X as a function $X: \Omega \to \mathbb{R}$. Unfortunately, defining a random variable this way is not sufficient in calculating probabilities about X. Indeed, to calculate $\mathbb{P}(\{X \leq 2\})$ for instance, we need to make sure that the set $X^{-1}\big((-\infty, 2]\big)$ is in the σ-algebra $\mathcal{F}$.

We define a *random variable* as a function $X: \Omega \to \mathbb{R}$ such that:

$$X^{-1}(B) \in \mathcal{F} \quad \text{for every } B \in \mathcal{F}.$$

This definition of a random variable does not require a measure to be defined on $(\Omega, \mathcal{F})$; all we need is Ω and the σ-algebra $\mathcal{F}$. We now extend this definition to general measurable spaces.

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces and $f: \Omega \to E$ be a function. (We will write $f: (\Omega, \mathcal{F}) \to (E, \mathcal{E})$.) We say that f is $\mathcal{F}/\mathcal{E}$-measurable if $f^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{E}$.

1. If $(E, \mathcal{E}) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$, we say that f is *Borel measurable*, or simply, *measurable*.
2. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(E, \mathcal{E}) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$, then f is called a *random variable* and we usually use the letter X instead of f.
3. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(E, \mathcal{E}) = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$, then f is called a *random vector* and we also use the letter X instead of f.

Whenever we say f is measurable without mentioning $(E, \mathcal{E})$, we will assume that we mean Borel measurable – i.e. $(E, \mathcal{E}) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$.

Here are some examples:

1 Let $(\Omega, \mathcal{F})$ be any measurable space and $A \subseteq \Omega$. We define the *indicator function* $I_A: \Omega \to \mathbb{R}$ of A as

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$$

Then I_A is measurable (Borel measurable) if and only if $A \in \mathcal{F}$. To see this, note that for any $B \in \mathcal{B}(\mathbb{R})$,

$$I_A^{-1}B = \begin{cases} \emptyset & 0,1 \notin B \\ A & 1 \in B, 0 \notin B \\ A^c & 0 \in B, 1 \notin B \\ \Omega & 0,1 \in B, \end{cases}$$

which is always in $\mathcal{F}$ if and only if $A \in \mathcal{F}$.

2 On $(\Omega, 2^\Omega)$, every function $f: \Omega \to \mathbb{R}$ is measurable, since $f^{-1}(B) \subseteq \Omega$ for each $B \in \mathcal{B}(\mathbb{R})$.

3 On $(\Omega, \{\emptyset, \Omega\})$, the only (Borel) measurable functions $f: \Omega \to \mathbb{R}$ are constants.

4 On $(\Omega, \{\emptyset, \Omega, A, A^c\}) A \subseteq \Omega$, the only measurable functions $f: \Omega \to \mathbb{R}$ are constant on A and constant on A^c. That is, $f: \Omega \to \mathbb{R}$ is measurable if and only if $f = \alpha I_A + \beta I_{A^c}$ for some constants $\alpha, \beta \in \mathbb{R}$.

Let us now look at a concrete example. Consider the example where two fair coins are tossed and only the outcome of the first coin is recorded. Let A be the event that the outcome of the first coin is a head and B be the event that the outcome of the first coin is a tail. Then $\Omega = \{HH, HT, TH, TT\}$ and

$$\mathcal{F} := \sigma(\{A, B\}) = \sigma\big(\{\{HH, HT\}, \{TH, TT\}\}\big) = \big\{\{HH, HT\}, \{TH, TT\}, \emptyset, \Omega\big\}.$$

Notice that since $A^c = B$, then $\mathcal{B} = \{A, B\}$ are the blocks that generate $\mathcal{F}$. Both I_A and I_B are random variables. In fact, the only random variables are linear combinations of the two. In other words, X is a random variable if it is constant on both A and B. This simply means that X is only dependent on the outcome of the first toss.

The preceding example suggests that if $\mathcal{F}$ is generated by (countably many) blocks, then the only $\mathcal{F}$-measurable functions are the ones that are constant on the blocks. The following result makes this more precise.

Theorem 1

Let $(\Omega, \mathcal{F})$ be a measurable space, where $\mathcal{F}$ is generated by countably many blocks $\mathcal{B} = \{B_1, B_2, \ldots\}$. Then $f: \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable if and only if f is constant on each block $B_n (n \geq 1)$. In this case, f can be written as

$$f(\omega) = \sum_{n=1}^{\infty} a_n I_{B_n}(\omega), \quad \omega \in \Omega,$$

where a_n is the value of f on B_n – i.e. $B_n \subseteq f^{-1}(\{a_n\})$.

Let $\Omega = \mathbb{R}$ and choose $\mathcal{F} = \sigma(\{\{\omega\}: \omega \in \mathbb{Z}\})$. Note that $\mathcal{F}$ is generated by the following (countably many) blocks:

$$\mathcal{B} = \{\{\omega\}: \omega \in \mathbb{Z}\} \cup \{\mathbb{R}\backslash\mathbb{Z}\} = \{\ldots\{-2\}, \{-1\}, \{0\}, \{1\}, \{2\}, \ldots, \mathbb{R}\backslash\mathbb{Z}\}.$$

Therefore, a function $f: \Omega \to \mathbb{R}$ is measurable if it is constant on the non-integers $\mathbb{R}\backslash\mathbb{Z}$. That is, $f: \Omega \to \mathbb{R}$ is measurable if and only if there exist a function $g: \mathbb{Z} \to \mathbb{R}$ and a constant $c \in \mathbb{R}$ such that

$$f(\omega) = \begin{cases} g(\omega) & \omega \in \mathbb{Z} \\ c & \omega \in \mathbb{R}\backslash\mathbb{Z}. \end{cases}$$

If the σ-algebra $\mathcal{F}$ is not generated by countably many blocks, it is generally difficult to prove that a function f is measurable. The difficulty is caused by the fact that we don't have a general form of an arbitrary Borel set B. (B could be an interval, a countably infinite set, or even the irrational numbers.) Thus, checking the pre-image condition for arbitrary Borel sets might be challenging.

The next theorem tells us that we do not need to check this condition for every Borel set, but it is enough to check the condition for a sub-collection that generates $\mathcal{B}(\mathbb{R})$.

Theorem 2

Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces and $\mathcal{A} \subseteq \mathcal{E}$ be a collection of subsets of E such that $\sigma(\mathcal{A}) = \mathcal{E}$. A function $f: \Omega, \mathcal{F} \to (E, \mathcal{E})$ is $\mathcal{F}/\mathcal{E}$-measurable if and only if for each $B \in \mathcal{A},\ f^{-1}(B) \in \mathcal{F}$.

As a corollary, the following are equivalent:

1. f is Borel measurable
2. $\{f < c\} := f^{-1}\big((-\infty, c)\big) \in \mathcal{F}$ for every $c \in \mathbb{R}$
3. $\{f \leq c\} := f^{-1}\big((-\infty, c]\big) \in \mathcal{F}$ for every $c \in \mathbb{R}$
4. $\{f > c\} := f^{-1}\big((c, \infty)\big) \in \mathcal{F}$ for every $c \in \mathbb{R}$
5. $\{f \geq c\} := f^{-1}([c, \infty)) \in \mathcal{F}$ for every $c \in \mathbb{R}$
6. $\{a < f < b\} := f^{-1}\big((a, b)\big) \in \mathcal{F}$ for every $a, b \in \mathbb{R}$
7. $\{a \leq f < b\} := f^{-1}([a, b)) \in \mathcal{F}$ for every $a, b \in \mathbb{R}$
8. $\{a < f \leq b\} := f^{-1}\big((a, b]\big) \in \mathcal{F}$ for every $a, b \in \mathbb{R}$
9. $\{a \leq f \leq b\} := f^{-1}([a, b]) \in \mathcal{F}$ for every $a, b \in \mathbb{R}$

Let $\Omega = [0,1], \mathcal{F} = \mathcal{B}([0,1]) := \sigma(\{(a, b): 0 \leq a < b \leq 1\}) =$ Borel subsets of $[0,1]$ and $\mathbb{P}(A) = \lambda_1(A)$ for each $A \in \mathcal{F}$. Define $X: \Omega \to \mathbb{R}$ as

$$X(\omega) := \begin{matrix} \frac{1}{4} & 0 \leq \omega < \frac{1}{2} \\ \omega & \frac{1}{2} \leq \omega \leq 1. \end{matrix}$$

Then X is a random variable. For each $c \in \mathbb{R}$, we have

$$\{X \leq c\} = \begin{cases} \Omega & c \geq 1 \\ [0, c] & \frac{1}{2} \leq c < 1 \\ \left[0, \frac{1}{2}\right] & 0 \leq c < \frac{1}{2} \\ \emptyset & c < 0 \end{cases}$$

which is always a Borel subset of $[0,1]$.

We denote by $m\mathcal{F}$ the set of all measurable functions $f: (\Omega, \mathcal{F}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$.

Let (Ω, F) be a measurable space and $f, g, f_n \in m\mathcal{F}$ be real-valued Borel measurable functions. Then:

1. $\alpha f, f + g, f^2, fg, f/g \in m\mathcal{F}$ where $\alpha \in \mathbb{R}$
2. $sup_n f_n, inf_n f_n \in m\mathcal{F}$ (See footnote.)[2]
3. $\lim_{n\to\infty} sup f_n, \lim_{n\to\infty} \inf f_n \in m\mathcal{F}$
4. If $\lim_{n\to\infty} f_n$ exists, then $\lim_{n\to\infty} f_n \in m\mathcal{F}$.

Let $(\Omega, \mathcal{F})$, $(E, \mathcal{E})$ and $(G, \mathcal{G})$ be measurable spaces and $f: (\Omega, \mathcal{F}) \to (E, \mathcal{E})$, $g: (E, \mathcal{E}) \to (G, \mathcal{G})$ be measurable functions. Then $h \coloneqq g \circ f: (\Omega, \mathcal{F}) \to (G, \mathcal{G})$ is $\mathcal{F}/\mathcal{G}$-measurable.

Let Ω be a set and $X: \Omega \to \mathbb{R}$ be a function. We define the σ-algebra generated by $X, \sigma(X)$, to be the smallest σ-algebra with respect to which X is a random variable.

Consider the random experiment of tossing two fair coins.

The sample space is $\Omega = \{HH, HT, TH, TT\}$.
Define $X: \Omega \to \mathbb{R}$ by $X(\omega) \coloneqq$ number of heads in ω. We have $X(HH) = 2, X(HT) = X(TH) = 1$ and $X(TT) = 0$. If $B \in \mathcal{B}(\mathbb{R})$, then $X^{-1}(B)$ will be one of $\{HH\}, \{HT, TH\}, \{TT\}$ or a union of these sets, depending on which of the numbers 0, 1, and 2 the set B contains. Thus:

[2] Strictly speaking, these are extended random variables. See the next set of notes.

$$\sigma(X) = \sigma(\{\{HH\},\{HT,TH\},\{TT\}\}).$$

That is, $\sigma(X)$ is generated by the blocks $\mathcal{B} = \{X^{-1}(\{2\}), X^{-1}(\{1\}), X^{-1}(\{0\})\}$.

Let $\Omega = [0,1]$ and define $X:\Omega \to \mathbb{R}$ by

$$X(\omega) := \begin{cases} 2 & 0 \leq \omega < \frac{1}{2} \\ \omega & \frac{1}{2} \leq \omega \leq 1. \end{cases}$$

For each $c \in \mathbb{R}$ we have

$$\{X \leq c\} = \begin{cases} \Omega & c \geq 2 \\ \left[\frac{1}{2}, 1\right] & 1 \leq c < 2 \\ \left[\frac{1}{2}, \sqrt{c}\right] & \frac{1}{4} \leq c \leq 1 \\ \emptyset & c < \frac{1}{4}. \end{cases}$$

Thus

$$\sigma(\mathrm{X}) = \sigma(\{\{\mathrm{X} \leq \mathrm{c}\} : \mathrm{c} \in \mathbb{R}\}) = \sigma\left(\left\{\Omega, \left[\frac{1}{2}, 1\right], \emptyset\right\} \cup \left\{\left[\frac{1}{2}, b\right] : \frac{1}{2} \leq b \leq 1\right\}\right)$$
$$= \sigma\left\{\left\{\left[\frac{1}{2}, \mathrm{b}\right] : \frac{1}{2} \leq b \leq 1\right\}\right\}$$
$$= \left\{\mathrm{B} \subseteq [0,1] : B \in \mathcal{B}\left(\left[\tfrac{1}{2}, 1\right]\right) \text{ or } \mathrm{B} = \left[0, \tfrac{1}{2}\right) \cup \mathrm{B}' \text{ where } \mathrm{B}' \in \mathcal{B}\left(\left[\tfrac{1}{2}, 1\right]\right)\right\}.$$

The set $\left[0, \frac{1}{2}\right)$ is a block of $\sigma(X)$ since X does not distinguish between elements of this set, implying that if a function $Y:\Omega \to \mathbb{R}$ is $\sigma(X)$-measurable, then Y must also be constant on $\left[0, \frac{1}{2}\right)$.

We end with a powerful result:

Theorem 3 [Doob-Dynkin Lemma]

Let Ω be a set and $X, Y: \Omega \to \mathbb{R}$ be functions. Then Y is $\sigma(X)$-measurable if and only if there exists a Borel measurable function $g: (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $Y = g(X)$.

3.2.4 Transcript: A Worked Example of a Random Variable

Now that we have introduced the basics of probability theory, we will begin the study of random variables.

So, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space – recall that this means that Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure.

A function, X, from Ω to $\mathbb{R}$ is called a random variable if it satisfies an additional condition which we call a **measurability condition**. This condition states that the pre-image of every Borel set is measurable:

$$X^{-1}(B) \in \mathcal{F} \text{ for every } B \in \mathcal{F}.$$

So, if you take any Borel subset of $\mathbb{R}$ and interval – for instance, one could be the natural numbers, or it can be any other Borel set – and you take the set of all outcomes for which X maps into that Borel set, that collection must be in $\mathcal{F}$. You will see why we need this – so that we can calculate the probability that X belongs to the Borel set.

Now, we can generalize this to any measurable function. So, in general, if we have a measurable space – for example, $(\Omega, \mathcal{F})$ – we define $\mathcal{F}$ as a function from $\Omega \to \mathbb{R}$ to be measurable if it satisfies a similiar condition. So, in other words, if the pre-image of every Borel set belongs to $\mathcal{F}$. So, essentially, a random variable is just a measurable function whereby the domain space has a probability measure defined on it, which makes it a probability space.

Let's look at an example.

Consider $(\Omega, \mathcal{F})$ to be any measurable space – remember that Ω is a set and $\mathcal{F}$ is a σ-algebra – and take A to be a subset of Ω. We define the indicator function of the set A as a function I_A from $\Omega \to \mathbb{R}$ that satisfies the following condition:

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$$

To visualize this, you have Ω and you have A as a subset of Ω. The indicator function takes the value 1 within the set A and 0 outside the set A.

The question is: is this a random variable or in general a measurable function? Let's check that.

If we take B to be any Borel subset of $\mathbb{R}$, then the pre-image of B will look like this:

$$I_A^{-1}(B) = \begin{cases} \Omega & 0, 1 \in \mathrm{B} \\ A & 1 \in B, 0 \notin B \\ A^c & 0 \in B, 1 \notin B \\ \emptyset & 0, 1 \notin B. \end{cases}$$

- In the first case, if the set B contains both 0 and 1, then the pre-image will be everything in Ω because it is the set of everything that is mapped to B, and, in this case, B contains both 0 and 1. Again, to visualize this, you will have 0 and 1 on the y-axis of a number line and Ω on the x-axis (even though it need not be the real line). I_A maps elements of Ω to either 0 or 1. Now, if the Borel set contains both 0 and 1, then the pre-image will be all of Ω.
- Next, if the Borel set contains 1 but not 0, the pre-image will be A.
- Similarly, it will be A^c if the Borel set contains 0 but does not contain 1.
- Finally, it will be the empty set if the Borel set contains neither 0 nor 1.

So, those are the pre-images.

Now, for I^A to be measurable with respect to the σ-algebra, we need the pre-image of any Borel set to be in $\mathcal{F}$ for each of these four cases. As you can see, this will happen if and only if A itself belongs to $\mathcal{F}$, because if A belongs to $\mathcal{F}$ then – Ω is always in $\mathcal{F}$ because it's Ω and the empty set is always in $\mathcal{F}$ because it's the empty set – A is in $\mathcal{F}$ and A^c is also in $\mathcal{F}$. So, I_A is measurable if and only if A belongs to $\mathcal{F}$.

We now introduce the law of a random variable, which is also called the **probability distribution** of X. So, let $X: \Omega \to \mathbb{R}$ be a random variable on the probability space (Ω, $\mathcal{F}$, $\mathbb{P}$). We define the law of X as a function $\mathbb{P}_X$ from the Borel subsets to $[0,1]$. This means that $\mathbb{P}_X$ is actually a probability measure and it is defined as $\mathbb{P}_X(B) := \mathbb{P}\big(X^{-1}(B)\big)$. Written in full:

$$\mathbb{P}_X: \mathcal{B}(\mathbb{R}) \to [0,1], \qquad \mathbb{P}_X(B) := \mathbb{P}\big(X^{-1}(B)\big).$$

If we look at this on a number line, we can take the Borel subset of $\mathbb{R}$, take its pre-image, or inverse image, via X, and then you measure the probability of the resulting set $\mathbb{P}\big(X^{-1}(B)\big)$ in Ω. That is how the law of X is defined.

Let's consider the same example where A is a measurable subset and X is the indicator of A. Recall that we found the pre-image of any Borel set, B, to be the following:

$$X^{-1}(B) = \begin{cases} \Omega & 0,1 \in B \\ A & 1 \in B, 0 \notin B \\ A^c & 1 \notin B, 0 \in B \\ \emptyset & 0,1 \notin B. \end{cases}$$

So, this is the pre-image. Therefore, we can calculate the law of X, as the above equation implies that the law of X will be as follows:

$$\mathbb{P}(B) = \begin{cases} 1 & 0,1 \in B \\ \mathbb{P}(A) & 1 \in B, 0 \notin B \\ \mathbb{P}(A^c) & 1 \notin B, 0 \in B \\ \mathbb{P}(\emptyset) & 0,1 \notin B. \end{cases}$$

Another way of writing this is simply as follows:

$$\mathbb{P}_X = \mathbb{P}(A)\delta_1 + \mathbb{P}(A^c)\delta_0.$$

3.2.5 Notes: The Law of a Random Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $X: (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable (i.e. $X \in m\mathcal{F}$) and $B \in \mathcal{B}(\mathbb{R})$. To calculate the probability that $X \in B$, we first find $X^{-1}(B)$ and then evaluate $\mathbb{P}\left(X^{-1}(B)\right)$. This means that we can define a set function $\mathbb{P}_X: \mathcal{B}(\mathbb{R}) \to [0,1]$ by:

$$\mathbb{P}_X(B) = P\big(X^{-1}(B)\big), \quad B \in \mathcal{B}(\mathbb{R}).$$

The function $\mathbb{P}_X$ is a probability measure on $\mathcal{B}(\mathbb{R})$.Indeed,

1. $\mathbb{P}_X(\emptyset) = \mathbb{P}\big(X^{-1}(\emptyset)\big) = 0.$
2. $\mathbb{P}_X(B) = \mathbb{P}\big(X^{-1}(B)\big) \geq 0, \quad B \in \mathcal{B}(\mathbb{R}).$
3. If $B_1, B_2, \ldots \in \mathcal{B}(\mathbb{R})$ are pairwise disjoint elements of $\mathcal{B}(\mathbb{R})$, then $X^{-1}(B_1), X^{-1}(B_2), \ldots$ are pairwise disjoint elements of $\mathcal{F}$ and

$$\mathbb{P}_x\left(\bigcup_{n=1}^{\infty} B_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\right) = \sum_{n=1}^{\infty} \mathbb{P}\big(X^{-1}(B_n)\big) = \sum_{n=1}^{\infty} \mathbb{P}_X(B_n).$$

4. $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}\big(X^{-1}(\mathbb{R})\big) = \mathbb{P}(\Omega) = 1.$

Thus, each random variable induces a measure on $\mathcal{B}(\mathbb{R})$. We will call $\mathbb{P}_X$ the *probability distribution* or *Law of* X.

Consider the random experiment of tossing 2 fair coins and let X count the number of heads.

Let $\mathcal{F} = 2^{\Omega}$ and define the probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ by:

$$\mathbb{P} = \sum_{\omega \in \Omega} \frac{1}{4} \delta_{\omega}\,.$$

Then X is automatically a random variable. It can be shown that:

$$\mathbb{P}_X = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2 \,.$$

We will later see that if X is a "discrete" random variable, then $\mathbb{P}_X$ is a linear combination of Dirac measures.

Let $\Omega = (0,1]$, $\mathcal{F} = \mathcal{B}((0,1])$ and $\mathbb{P} = \lambda_1$.Then $X(\omega) = 4\omega\ \forall\omega \in \Omega$ is a random variable.

We now calculate $\mathbb{P}_X$.

First, if $B = (-\infty, x]$, then:

$$\mathbb{P}_X\big((-\infty, x]\big) = \mathbb{P}(\{X \leq x\}) = \mathbb{P}(\{\omega \in (0,1]: 4\omega \leq x\}) = \mathbb{P}\big((0,1] \cap (-\infty, x/4]\big)$$
$$= \begin{cases} 0 & x \leq 0 \\ \dfrac{x}{4} & 0 < x < 4 \\ 1 & x \geq 4\,. \end{cases}$$

Now, since $\mathrm{Range}(X) = (0,4]$, $\mathbb{P}_X$ is concentrated on (0,4] and

$$\mathbb{P}_X(B) = \frac{\lambda_1\big((0,4] \cap B\big)}{4}, \quad B \in \mathcal{B}(\mathbb{R}).$$

Let $X: (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be a random variable. The *Cumulative Distribution Function (CDF)* of X is the function $F_X: \mathbb{R} \to \mathbb{R}$ defined by:

$$F_X(x) \coloneqq \mathbb{P}_X\big((-\infty, x]\big) = \mathbb{P}(\{X \leq x\}), \quad x \in \mathbb{R}.$$

The CDF characterizes the distribution of a random variable. That is, if $X, Y: (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are random variables, then:

$$\mathbb{P}_X = \mathbb{P}_Y \Leftrightarrow F_X = F_Y.$$

The CDF of any random variable X satisfies the following properties:

1. F_X is increasing; i.e. if $x \leq$ then $F_X(x) \leq F_X(y)$.
2. F_X is right continuous; i.e. $\mathrm{F_X}(x^+) \coloneqq \lim_{y \to x^+} F_X(y) = F(x)$.
3. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.

The converse of the above is also true. That is, if you start with an arbitrary function that satisfies the above properties, then that function is a CDF of some random variable.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X \in m\mathcal{F}$ be a random variable.

1. We say that X is *discrete* iff $\mathbb{P}_X$ is concentrated on a countable set $\{x_n : n = 1,2,3, \dots\}$.
2. We say that X is *continuous* iff F_X is a continuous function.
3. We say that X is *absolutely continuous* iff F_X is an absolutely continuous function, in the sense that for every $\epsilon > 0$ there exists $\delta > 0$ such that if $(a_1, b_1), \dots, (a_n, b_n)$ is a finite collection of disjoint intervals in $\mathbb{R}$ with

$$\sum_{i=1}^{n} |b_i - a_i| < \delta,$$

then

$$\sum_{i=1}^{n} |F(b_i) - F(a_i)| < \epsilon.$$

If X is a discrete random variable with $\mathbb{P}_X$ concentrated on $\{x_n : n = 1,2,3\}$, then

$$\mathbb{P}_X = \sum_{n=1}^{\infty} p_X(x_n) \delta_{x_n},$$

where the function $p_X: \mathbb{R} \to \mathbb{R}$, called the *probability mass function (PMF)* of X, is defined by

$$p_X(x) = \mathbb{P}_X(\{x\}), \quad x \in \mathbb{R}.$$

The coin toss random variable counting the number of heads is obviously discrete. In fact, every random variable with a countable range is discrete.

3.2.6 Notes: Expectation

In this section, we develop the Lebesgue integral of a measurable function f with respect to a measure μ. In the special case when $\mu = \mathbb{P}$ is a probability measure and $f = X$ is a random variable, this integral is the expected value of X.

Our intuition about integration is that it represents “aggregation” or “summing up”. We will see that by an appropriate choice of μ and Ω, the integral will cover the many forms of aggregation we are used to in mathematics, including summation, expectation, and calculating area/volume using the Riemann integral. The advantage now is that we can look at all these forms of aggregation as essentially doing the same thing: integrating with respect to a measure. This point of view will help us prove one theorem that applies to all these special cases. In applications to probability theory, it will enable us to treat discrete, continuous, and other random variables the same way. There will be no need, for instance, to prove a theorem for discrete random variables using summation and then prove the same theorem again for absolutely continuous random variables using integrals.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f\colon \Omega \to \mathbb{R}$ be a measurable function. In this section, we want to define the integral of f with respect to the measure μ, denoted by

$$\int_\Omega f \, d\mu.$$

When $\mu = \mathbb{P}$ is a probability measure and $f = X$ is a random variable, we will define the expectation of X as

$$\mathbb{E}(X) \coloneqq \int_\Omega X \, d\mathbb{P}.$$

For reasons we will see later, it will be useful to define the integral for *extended measurable function*. First, on the extended real numbers $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty, -\infty\}$, we define the Borel σ-algebra $\mathcal{B}(\overline{\mathbb{R}})$ to be

$$\mathcal{B}(\overline{\mathbb{R}}) := \{B \subseteq \overline{\mathbb{R}}: B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\}$$
$$= \{B \subseteq \overline{\mathbb{R}}: B \in \mathcal{B}(\mathbb{R}) \text{ or } B = B' \cup \{\infty\} \text{ or } B = B' \cup \{-\infty\} \text{ or } B = B' \cup \{\infty, -\infty\}, B' \in \mathcal{B}(\mathbb{R})\}.$$

A function $f: (\Omega, \mathcal{F}) \to \left(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\right)$ that is $\mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$-measurable will be called an *extended measurable function*. If $f = X$ is a random variable, we will call X an *extended random variable*. Throughout this chapter we will assume that all measurable functions are extended measurable functions.

We will carry out the definition in four steps:

1. We first define the integral for the case when $f = I_A$ is an indicator function of a measurable set $A \in \mathcal{F}$.
2. We will the extend the definition to when f is positive[3] and is a (finite) linear combination of indicators. Such functions are called (positive) *simple* functions.
3. We then extend the definition to all positive measurable functions and prove useful results that aid with the calculation of the integral in many concrete examples.
4. Finally, we extend the definition to general measurable (sign-changing) functions f and define the notion of integrability. This will be covered in the next section.

We begin with indicators. Throughout the rest of the chapter, $(\Omega, \mathcal{F}, \mu)$ is a measure space and $f: (\Omega, \mathcal{F}) \to \left(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})\right)$ is $\mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$-measurable.

If $f = I_A$ for some $A \in \mathcal{F}$, then we define the *integral* of f with respect to the measure μ to be

[3] We should really be saying "non-negative".

$$\int_{\Omega} f \, d\mu = \int_{\Omega} I_A \, d\mu \coloneqq \mu(A).$$

In the case of a probability space, we have

$$\mathbb{E}(I_A) = \int_{\Omega} I_A \, d\mathbb{P} \coloneqq \mathbb{P}(A).$$

There is nothing surprising about this definition; it is consistent with our intuition of aggregation. We move on to simple functions.

A measurable function $s: \Omega \to \overline{\mathbb{R}}$ is called *simple* if it has a finite range. We will denote by $\mathcal{S}$ the set of all (measurable) simple functions. We will also write $\mathcal{S}^+$ for the class of positive simple functions and $m\mathcal{F}^+$ for the class of positive measurable functions. Note that if $s \in \mathcal{S}$ with $\text{Range}(s) = \{s_1, \dots, s_n\}$, then s can be written as

$$s = \sum_{k=1}^{n} s_k \, I_{A_k},$$

where $A_k = s^{-1}\{(s_k)\} \in \mathcal{F}$ for $k = 1,2, \dots, n$ are pairwise disjoint. We will call this the canonical representation of a simple function and assume this representation when we talk about a simple function.

Let $f = s$, where $s \in \mathcal{S}^+$ is a positive simple function with canonical representation

$$s = \sum_{k=1}^{n} s_k \, I_{A_k}.$$

We define the integral of f with respect to the measure μ as

$$\int_{\Omega} f \, d\mu = \int_{\Omega} s \, d\mu \coloneqq \sum_{k=1}^{n} s_k \, \mu(A_k).$$

Note that the value of this integral could be ∞, even if f is finite. The probability space equivalent of this definition is that if X is a simple random variable with

$$X = \sum_{k=1}^{n} s_k \, I_{A_k},$$

then

$$\mathbb{E}(X) = \int_{\Omega} X \, d\mathbb{P} \coloneqq \sum_{k=1}^{n} s_k \, \mathbb{P}(A_k).$$

Note that if $s \in \mathcal{S}^+$ is a positive simple function with canonical representation

$$s = \sum_{k=1}^{n} s_k \, I_{A_k},$$

then

$$\int_{\Omega} s \, d\mu = \sum_{k=1}^{n} s_k \int_{\Omega} I_{A_k} \, d\mu.$$

We extend this definition to all positive measurable functions. Let $f \in m\mathcal{F}^+$ be a positive measurable function. We define the integral of f with respect to μ as

$$\int_{\Omega} f \, d\mu \coloneqq sup\left\{\int_{\Omega} s \, d\mu : s \in \mathcal{S}^+,\ 0 \le s \le f\right\}.$$

When $\mu = \mathbb{P}$ is a probability measure and $f = X$ is a random variable, then this definition means

$$\mathbb{E}(X) = \int_{\Omega} X\, d\mathbb{P} := sup\{\mathbb{E}(s): s \in \mathcal{S}^+,\ 0 \le s \le X\}.$$

If $f \in \mathcal{S}^+$, then this definition gives the same result as the previous definition of the integral for positive simple functions. That is

$$\int_{\Omega} f\, d\mu := sup\left\{\int_{\Omega} s\, d\mu: s \in \mathcal{S}^+,\ 0 \le s \le f\right\}$$

even when $f \in \mathcal{S}^+$.

If $f, g \in m\mathcal{F}^+$ are such that $f(\omega) \le g(\omega)$ for every $\omega \in \Omega$, then

$$\int_{\Omega} f\, d\mu \le \int_{\Omega} g\, d\mu.$$

This definition is not user-friendly. We would like to avoid having to use it to calculate integrals in concrete cases. The following two results will help us do just that. First, we need some definitions. Let (f_n) be a sequence of measurable functions. We say that (f_n) is *increasing* if for each $\omega \in \Omega$ and $n \ge 1$, $f_n(\omega) \le f_{n+1}(\omega)$. If $f \in m\mathcal{F}$, we say that (f_n) *converges point-wise* to f if for each $\omega \in \Omega$, $\lim_{n\to\infty} f_n(\omega) = f(\omega)$. We will simply write $f_n \to f$ to denote point-wise convergence of (f_n) to f.

Here's the big theorem:

Theorem 4: [Monotone Convergence Theorem (MON), Beppo-Levi]
Let $f \in m\mathcal{F}^+$ and (f_n) be an increasing sequence of positive measurable functions that converges to f point-wise. Then

$$\int_{\Omega} f\, d\mu = \lim_{n\to\infty} \int_{\Omega} f_n\, d\mu.$$

That is,

$$\lim_{n\to\infty} \int_\Omega f_n \, d\mu = \int_\Omega \lim_{n\to\infty} f_n \, d\mu.$$

The probability space equivalent statement is that if (X_n) is an increasing sequence of non-negative random variables that converges point-wise to a non-negative random variable X, then

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) = \mathbb{E}(X).$$

This result says that we can interchange between taking limits and evaluating integrals in the case when the sequence of functions is increasing and positive. We will see other theorems of this nature in the next section. The result will be helpful in calculating the integral of $f \in m\mathcal{F}^+$ if we already know how to calculate the integrals of the f_n's. In other words, we need to be able to find a sequence (f_n) such that:

1. (f_n) is increasing,
2. $f_n \in m\mathcal{F}^+$ for each $n \geq 1$,
3. (f_n) converges point-wise to f, and
4. $\int_\Omega f_n \, d\mu$ is easy to calculate.

Here is one case where all these requirements are met.

Theorem 5: [Simple Function Approximation]
If $f \in m\mathcal{F}^+$, then there exists an increasing sequence of positive simple functions (s_n) such that (s_n) converges point-wise to f.

It then follows that

$$\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega s_n \, d\mu,$$

by (MON).

Let us see this theorem in action. Consider the measure space $([0,1], \mathcal{B}([0,1]), \lambda_1)$ and $f: [0,1] \to \overline{\mathbb{R}}$ defined by $f(x) = x$. We have

$$s_n = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} I_{\left\{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\right\}} + nI_{\{f \geq n\}},$$

where

$$\left\{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\right\} = \left\{x \in [0,1]: \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\right\} = [0,1] \cap \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right),$$

which is a null set when $k \geq 2^n$. Hence,

$$s_n = \sum_{k=0}^{2^n - 1} \frac{k}{2^n} I_{\left\{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\right\}} = \sum_{k=0}^{2^n - 1} \frac{k}{2^n} I_{\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)}$$

$$\int_{[0,1]} s_n \, d\lambda_1 = \sum_{k=0}^{2^n - 1} \frac{k}{2^n} \lambda_1\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right) = \sum_{k=0}^{2^n - 1} \frac{k}{2^{2n}} = \frac{1}{2}\left(1 - \frac{1}{2^n}\right).$$

Hence, (MON) gives

$$\int_{[0,1]} f d\mu = \lim_{n \to \infty} \int_{[0,1]} s_n \, d\mu = \lim_{n \to \infty} \frac{1}{2}\left(1 - \frac{1}{2^n}\right) = \frac{1}{2}.$$

We will see later that in general, if $f: [a, b] \to \mathbb{R}$ is Riemann-integrable, then

$$\int_{[a,b]} f \, d\lambda_1 = \int_a^b f(x) dx.$$

That is, integrating with respect to the Lebesgue measure is the same as ordinary Riemann integration when f is Riemann integrable on a bounded interval.

Also note that $([0,1], \mathcal{B}([0,1]), \lambda_1)$ is actually a probability space, therefore f is random variable. It is easy to show that $f \sim U(0,1)$, which is consistent with the fact the integral represents the expected value of f.

We now extend the definition of the integral to functions in $m\mathcal{F}$ that may change sign. If $f \in m\mathcal{F}$, we define the positive part of f as the function $f^+ := max\{f, 0\}$, and the negative part of f as $f^- := max\{-f, 0\}$. It is easy to show that both $f^+, f^- \in m\mathcal{F}^+$, $f = f^+ - f^-$, and $|f| = f^+ + f^-$.

Let $f \in m\mathcal{F}$. We define the *integral* of f with respect to μ as

$$\int_\Omega f \, d\mu := \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu,$$

provided at least one of the integrals on the right is finite. In that case, we say that the integral exists, otherwise, we say that the integral does not exist. We will say that f is *integrable* if both integrals on the right are finite, i.e. if

$$\int_\Omega |f| \, d\mu < \infty.$$

We will denote by $\mathcal{L}^1 := \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ the set of all integrable functions. In general, if $p \in [1, \infty)$, we define

$$\mathcal{L}^p := \mathcal{L}^p(\Omega, \mathcal{F}, \mu) := \left\{ f \in m\mathcal{F} \colon \int_\Omega |f|^p \, d\mu < \infty \right\}.$$

We will also be interested in integrating functions over subsets of Ω. If $A \in \mathcal{F}$, we define the integral of f over A as

$$\int_A f \, d\mu := \int_\Omega f I_A \, d\mu,$$

provided the integral exists. If $\mu = \mathbb{P}$ is a probability space and $f = X$ is a random variable, we will write this as

$$\mathbb{E}(X; A) := \mathbb{E}(XI_A).$$

Here is a powerful theorem that allows us to interchange limits and integrals.

Theorem 6: [Lebesgue Dominated Convergence Theorem (DCT)]
Let (f_n) be a sequence of measurable functions such that there exists $g \in \mathcal{L}^1 \cap m\mathcal{F}^+$ such that $|f_n(\omega)| \leq g(\omega)$ for every $\omega \in \Omega$ and $n \geq 1$. If $f \in m\mathcal{F}$ and $\lim_{n\to\infty} f_n(\omega) = f(\omega)$ for every $\omega \in \Omega$, then

1. $f \in \mathcal{L}^1$,
2. $\lim_{n\to\infty} \int_\Omega |f_n - f| \, d\mu = 0$, *and*
3. $\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega f_n$.

In the language of probability, the theorem says that if (X_n) is a sequence of random variables such that there exists a random variable $Y \in \mathcal{L}^1 \cap m\mathcal{F}^+$ such that $|X_n(\omega)| \leq Y(\omega)$ for every $\omega \in \Omega, n \geq 1$, and $X \in m\mathcal{F}$ is such that $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for every $\omega \in \Omega$, then

1. $X \in \mathcal{L}^1$,
2. $\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|\right) = 0$, and
3. $\lim_{n\to\infty} \mathbb{E}\left(X_n\right) = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) = \mathbb{E}(X)$.

Here is a very important special case of how to evaluate Lebesgue integrals:

Theorem 7
Let $f : [a, b] \to \mathbb{R}$ be bounded.

1. *f is Riemann integrable if and only if f is continuous λ_1-almost everywhere.*
2. *If f is Riemann integrable, then it is Lebesgue integrable and*

$$\int_a^b f(x)\ dx = \int_{[a,b]} f\, d\lambda_1.$$

On $(\mathbb{N}, 2^{\mathbb{N}}, \#)$, integration reduces to summation:

$$\int_{\mathbb{N}} f d\# = \sum_{n\in\mathbb{N}} f(n),$$

provided the series converges.

For any set Ω and $a \in \Omega$, we have

$$\int_{\Omega} f\, d\delta_a = f(a).$$

Finally, we mention that if $\mu = \sum_k a_k\, \mu_k$, where $a_k \geq 0$, we have

$$\int_{\Omega} f\, d\mu = \sum_k a_k \int_{\Omega} f\, d\mu_k,$$

provided f is μ-integrable.

3.2.7 Transcript: A Worked Example of an Expectation

Hi, in this video we introduce the expectation of a random variable X and the integral of a function f with respect to a measure μ.

So, let $(\Omega, \mathcal{F}, \mu)$ be a measure space and f, where $f : \Omega \to \mathbb{R}$, be a measurable function.

What we want to do is define what we mean by the integral of f with respect to the measure μ. Written in full:

$$\int_{\Omega} f \, d\mu.$$

We are going to do this in four steps:

1. $f - I_A, \quad A \in \mathcal{F}$
   Firstly, we will assume that f is the indicator of a set A, where A is measurable. We will define what integration is in that special case.

2. Secondly, we will assume that f is what we call a **simple function**. A simple function is simply a function that has a finite combination of indicators.

3. $f \geq 0$
   Next, we are going to assume that f is positive and measurable. We will use a definition from simple functions to define integration in that special case.

4. $f \in M\mathcal{F}$
   Finally, we are going to assume that f is a general measurable function and define the notion of integrability.

We will begin with step one of indicators. So, let $f = I_A$. We define the integral of f with respect to the measure μ as simply the measure of A. Written in full:

$$\int_{\Omega} f \, d\mu \coloneqq \mu(A).$$

So, in that special case, the definition is simple.

We move onto the next step where f is a function that we call a simple function. This can be represented as a finite linear combination of indicators as follows:

$$f = \sum_{k=1}^{n} \alpha_k I_{A_k}.$$

We are going to define the integral of f as the sum α_k times μ of A_k as follows:

$$\int_{\Omega} f \, d\mu \coloneqq \sum_{k=1}^{n} \alpha_k \mu(A_k).$$

Again, this is a linear combination of the measures of the sets A_k.

Now, one might ask why, if we represent this function in a different way, will the integral necessarily be the same. We can show that it is indeed the same and so, this definition is independent of the representation of f.

We move on to the third step where the function f is now positive. So, f is positive and measurable at the same time, which we can write as:

$$f \geq 0, \quad f \in M\mathcal{F}^+.$$

We are going to define the integral of f in that special case as the supremum of the integrals of simple functions that are less than f. Written in full:

$$\int_{\Omega} f\, d\mu := sup\left\{\int_{n} s\, d\mu : 0 \leq s \leq f\right\}.$$

So, we take the integral of every simple function that is less than f, greater than or equal to 0, and we take the supremum of that set. Note that the first integral could be infinite.

In the next step, we take f to be a measurable function that may be sign-changing and, if that is the case, we define f^+ to be the positive part of f, so, the maximum of f and 0. To visualize this, we can imagine that we have a function that looks like this. f^+ will take simply take f when f is positive, and 0 when f is negative. So, it will look like this – so that's what f^+ is. Similarly, we define f^-, which is called the negative part of f, and that will be the maximum of negative f and 0. What happens is, when f is negative, it reflects it, so, it takes the negative of that, and, when f is positive, it remains at 0, and so on. So, f^- will look like this. Note that f can be written as a difference between f^+ and f^-. The absolute value of f is actually a sum of f^+ plus f^-.

We are therefore going to define the integral of f in this case as the integral of f^+ minus the integral of f^-, provided at least one of them is finite. Otherwise, the integral is undefined. Written in full:

$$\int_{\Omega} f\, d\mu := \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu.$$

We are allowed to make this definition because we know the integral of f^+ since f^+ is a positive measurable function. Similarly, we can also define the integral of f^- since f^- is a positive and measurable function.

We say that f is integrable if both of these integrals are finite and that is equivalent to the integral of the absolute value of f, which is less than infinity:

$$\int_{\Omega} |f| d\mu < \infty.$$

We will now look at some examples of integration.

1 The first one, which is probably the most important for us, is the special case where μ is actually a probability measure. We are going to denote it by $\mathbb{P}$.

In that case, $(\Omega, \mathcal{F}, \mathbb{P})$is a probability space and X is a random variable, where $X: \Omega \to \mathbb{R}$. In that case, the integral of X with respect to $\mathbb{P}$ is defined as the expected value of the random variable X. Written in full:

$$\int_{\Omega} X \, d\mathbb{P} =: E(X).$$

This is a very important special case for this course.

2 In the second example, we will take Ω to be the real line $\mathbb{R}$, $\mathcal{F}$ to be the σ-algebra of Borel subsets of $\mathbb{R}$, and μ to be the Lebesgue measure, λ_1.

If we consider an interval, $[a, b]$, as a subset of $\mathbb{R}$, then the integral over $[a, b]$ of a function f, with respect to the Lebesgue measure, turns out to be very simple in many special cases. It is equal to the Riemann integral of $f(x)$, provided that f is Riemann integrable on the interval $[a, b]$. Written in full:

$$\int_{[a,b]} f d\lambda_1 = \int_a^b f(x) dx.$$

3 In another example, we take Ω to be the natural numbers, and $\mathcal{F}$ to be the powerset of the natural numbers, and consider μ to be equal to the counting measure that counts the number of elements in each set. Then, the integral of a function, f, with respect to the counting measure, is simply equal to a sum of $f(n)$. So, summation is a special case of integration. Written in full:

$$\int_{\Omega} f d\,\# = \sum_{n \in \mathbb{N}} f(n).$$

4 As a last example, we consider any measure space and equip it with the Dirac measure – for example, $(\Omega, \mathcal{F}, \delta_a)$ – so that a belongs to $\mathcal{F}$ and the Dirac measure, as you will recall, is a measure that assigns 1 if a belongs to A and 0 otherwise. Then the integral of a function with respect to the Dirac measure is simply equal to a function evaluated at that point. Written in full:

$$\int_{\Omega} f d\delta_a = f(a).$$

Now that we have understood what expectation is, we can move on to the Radon-Nikodym theorem.

3.2.8 Notes: The Radon-Nikodym Theorem

Let $f \in m\mathcal{F}^+$ and define $\nu_f^\mu : \mathcal{F} \to \overline{\mathbb{R}}$ by

$$\nu_f^\mu(A) = \int_A f \, d\mu, \quad A \in \mathcal{F}.$$

Then ν_f^μ is a positive measure (exercise). Also, if $\mu(A) = 0$ then $\nu_f^\mu(A) = 0$. Likewise, if $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $X \in m\mathcal{F}^+$ is a non-negative random variable with $\mathbb{E}(X) = 1$, then the function $Q : \mathcal{F} \to [0,1]$ defined by

$$Q(A) = \mathbb{E}(X; A), \quad A \in \mathcal{F}$$

is a probability measure such that for every $A \in \mathcal{F}, \mathbb{P}(A) = 0 \Longrightarrow Q(A) = 0$.

Let μ, ν be positive measures on $(\Omega, \mathcal{F})$.

1 We say that ν is *absolutely continuous* with respect to μ (written $\nu \ll \mu$) if for each $A \in \mathcal{F}, \mu(A) = 0 \Longrightarrow \nu(A) = 0$.
2 We say that μ and ν are *equivalent* (written $\mu \equiv \nu$) if $\nu \ll \mu$ and $\mu \ll \nu$.
3 We say that ν and μ are *mutually singular* (written $\nu \perp \mu$) if there exists $A \in \mathcal{F}$ such that $\nu(A) = \mu(A^c) = 0$.

Here are some examples:

1 On $(\mathbb{R}, L(\mathbb{R})), \delta_0 \perp \lambda_1$ since $\lambda_1(\{0\}) = \delta_0(\mathbb{R} \backslash \{0\}) = 0$.
2 If $(\Omega, \mathcal{F}, \mu)$ is a measurable space, then $\mu \ll \#$, since $\#(A) = 0 \Longrightarrow A = \emptyset \Longrightarrow \mu(A) = 0$.
3 If $(\Omega, \mathcal{F}, \mu)$ is a measure space, then for any constants $a, b > 0, a\mu \equiv b\mu$.
4 Note that in general, if $\mu = \sum_{k=1}^{\infty} a_k \mu_k$ where $a_k > 0$ for each $k \geq 1$, then $\mu_k \ll \mu$ for every $k \geq 1$.

We have seen above that $\nu_f^\mu \ll \mu$. The Radon-Nikodym Theorem tells us that all measures that are absolutely continuous with respect to μ arise in this way (at least in the σ −finite case).

Theorem 8 [Radon-Nikodym]

Let $(\Omega, \mathcal{F})$ *be a measurable space, and* μ, ν *be two* σ *−finite measures on* $(\Omega, \mathcal{F})$ *such that* $\nu \ll \mu$. *Then there exists* $f \in m\mathcal{F}^+$ *such that* $\nu = \nu_f^\mu$, *in the sense that*

$$\nu(A) = \int_A f \, d\mu, \quad A \in \mathcal{F}.$$

Furthermore, if $g \in m\mathcal{F}^+$is such that $\nu = \nu_g^\mu$, then $f = g \; \mu - a.e.$. Thus, f is unique up to a null set.

The function f such that $\nu = \nu_f^\mu$ is called the *Radon-Nikodym derivative* or *density* of ν with respect to μ and is denoted by

$$f =: \frac{d\nu}{d\mu} \; .$$

Here are some examples:

1 If $(\Omega, \mathcal{F}, \mu)$ is a measure space, then define (for any constant $a > 0$), $\nu := a\mu$. We have that $\nu \ll \mu$ and

$$\frac{d\nu}{d\mu} = a, \; \mu - a.e.$$

To see this, note that for each $A \in \mathcal{F}$,

$$\nu(A) = a\mu(A) = a \int_A 1 d\mu = \int_A a \, d\mu.$$

2 If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $F \in \mathcal{F}$, define $\nu(A) := \mu(A \cap F)$ for every $A \in \mathcal{F}$. Then ν is also a measure on $(\Omega, \mathcal{F})$ and $\nu \ll \mu$ with

$$\frac{d\nu}{d\mu} = I_F, \qquad \mu - a.e..$$

Theorem 9

Let $(\Omega, \mathcal{F})$ be a measurable space and ν, μ be measures with $\nu \ll \mu$. If $g \in \mathcal{L}^1(\Omega, \mathcal{F}, \nu)$, then $g d\nu/d\mu \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ and

$$\int_A g\, d\nu = \int_A g \frac{d\nu}{d\mu}\, d\mu, \text{for every } A \in \mathcal{F}.$$

Let $(\Omega, \mathcal{F})$ be a measurable space and μ, ν, and λ be measures.

1 If $\nu \ll \mu$ and $\lambda \ll \mu$, then $a\nu + b\lambda \ll \mu$ for every $a, b \geq 0$. Furthermore,

$$\frac{d(a\nu + b\lambda)}{d\mu} = a\frac{d\nu}{d\mu} + b\frac{d\lambda}{d\mu}, \quad \mu - a.e..$$

2 If $\lambda \ll \nu$ and $\nu \ll \mu$, then $\lambda \ll \mu$ and

$$\frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu}\frac{d\nu}{d\mu}.$$

3 If $\nu \equiv \mu$, then $d\nu/d\mu > 0 \quad \mu - a.e.,\ d\mu/d\nu > 0 \quad \nu - a.e.$, and

$$\frac{d\mu}{d\nu} = \frac{1}{\frac{d\nu}{d\mu}} \quad \nu - a.e.$$

A random variable X is absolutely continuous if and only if $\mathbb{P}_X \ll \lambda_1$.

3.2.9 Transcript: The Radon-Nikodym Theorem

Hi, in this video we introduce the Radon-Nikodym theorem, which is a theorem about the relationship between measures on a measurable space.

Let (Ω, F) be a measurable space and take μ and λ (λ does not mean the Lebesgue measure in this case) to be measures on this measurable space: $[0, \infty]$. We say that λ is absolutely continuous with respect to μ and we write it in this form $\lambda \ll \mu$, which means if $\mu(A) = 0$ then $\lambda(A) = 0$. We say that λ and μ are equivalent if and only if $\lambda(A) = 0$ is equivalent to $\mu(A) = 0$. Another way of saying this is that λ is absolutely continuous with respect to μ, and μ is also absolutely continuous with respect to λ. Finally, we say that λ and μ are mutually singular, and we write it like this ($\lambda \perp \mu$) to mean that there exists a set A such that $\lambda(A) = 0$ is equal to $\mu(A^C)$. So, it sends both of them to 0.

As an example, let's take Ω to be $\mathbb{R}$ and F to be the Borel σ-algebra on $\mathbb{R}$. Consider the following measures: δ_0 and λ_1, where λ_1 is the Lebesgue measure and δ_0 is a Dirac measure concentrated at zero. These two measures are mutually singular since the set A can be picked to be the singleton 0. In that case, $\delta_0\,(\{0\}^C) = 0$ and that's the same as $\lambda_1(\{0\})$ because the Lebesgue measure sends singletons to zero.

In another example, if we take f to be a positive measurable function on any measurable space, we define $\lambda_1(A)$ to be the integral over A of f with respect to μ, where μ is a measure on the measurable space $\mathcal{F}$. This turns out to be a measure: one can show that it is the σ-additive and therefore satisfies all the axioms of a measure and it's absolutely continuous with respect to μ. If the measure of A is zero, then the integral over A will also be equal to zero, hence this will be absolutely continuous with respect to μ.

The Radon-Nikodym theorem says that the converse of the following statement is also true. In other words, if λ is absolutely continuous with respect to μ then, in fact, there exists a function that is positive and measurable such that $\lambda(A)$ is equal to the integral of f with respect to μ for every A in $\mathcal{F}$. So it is a converse of this result here that all measures λ that are absolutely continuous with respect to μ are in fact of this form, so they arise as an integral. This is true provided that the measures λ and μ are σ-finite measures. This function f that exists is called the Radon-Nikodym derivative and it's unique almost everywhere. f denoted by $\frac{d\lambda}{d\mu}$ is a function called a Radon-Nikodym derivative.

Let us look at how we can apply the Radon-Nikodym theorem. In this example, we will take (Ω, F) again to be any measurable space. Consider a measure μ and let $\lambda = A$ times $\mu(a\mu)$, where a is a positive constant ($a > 0$). The measure λ is just a scalar multiple of μ. Now, if $\mu\ (A) = 0$, then $\lambda(A)$, which is a times μ of A $(a\mu(A))$, will also be equal to 0. This implies that λ is absolutely continuous with respect to μ. And, assuming that μ is σ-finite, that would imply that λ is also σ -finite; hence a Radon-Nikodym derivative exists.

We can calculate it as follows: $\lambda(A)$ is equal to a times μ of A, which is a times the integral of 1 with respect to μ, which is equal to the integral over A of a times $d\mu$. Written in full:

$$\lambda(A) = a\mu(A) = a\int_A 1 d\mu$$

$$= \int_A a d\mu \Rightarrow \frac{d\lambda}{d\mu} = a.$$

This implies that the Radon-Nikodym derivative of λ with respect to μ is equal to a itself.

Next, we will look at stochastic processes.

3.2.10 Notes: Problem Set

Problem 1

Consider the function $f: \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = (x + 2)^3.$$

What is the set $f^{-1}\big((-1,1)\big)$?

Solution:

We just have to compute the inverse of the following function:

$$y = f(x) = (x + 2)^3.$$

First, we need to isolate x,

$$x = y^{1/3} - 2.$$

Now, we get the value for the extremes of the given set, $f^{-1}\big((-1,1)\big)$. For -1, we have $y^{1/3} - 2 = -1 - 2 = -3$ and for 1 we get $y^{1/3} - 2 = 1 - 2 = -1$. Thus, the solution is the interval $(-3, -1)$.

Problem 2

Consider the measure space $([0,1], \mathcal{B}([0,1]), \mu \coloneqq \lambda_1 + 2\delta_0)$ and the measurable function $f(\omega) = 2\omega$. Find

$$\int_{[0,1]} f \, d\mu.$$

Solution:

We will need to apply *Theorem 7:*

Let $f\colon [a,b] \to \mathbb{R}$ be bounded.

1. f is Riemann integrable if and only if f is continuous λ_1-almost everywhere.
2. If f is Riemann integrable, then it is Lebesgue integrable and

$$\int_a^b f(x)\ dx = \int_{[a,b]} f\, d\lambda_1.$$

We will need to apply this together with the following property: for any set Ω and $a \in \Omega$, we have

$$\int_{\Omega} f\ d\delta_a = f(a).$$

Thus,

$$\int_{[0,1]} f\, d\mu = \int_{[0,1]} 2x\,(d\lambda_1 + 2d\delta_0) = \int_{[0,1]} 2x\, d\lambda_1 + \int_{[0,1]} 4x\, d\delta_0.$$

We can solve the first integral applying Theorem 7 (see above):

$$\int_{[0,1]} 2x\, d\lambda_1 = \int_0^1 2x\, dx = 1.$$

On the other hand, the second integral will be solved as follows:

$$\int_{[0,1]} 4x\, d\delta_0 = f(0) = 4 * 0 = 0$$

Thus, the final solution is,

$$\int_{[0,1]} f\,d\mu = \int_{[0,1]} 2x\,(d\lambda_1 + 2d\delta_0) = \int_{[0,1]} 2x\,d\lambda_1 + \int_{[0,1]} 4x\,d\delta_0 = 1 + 0 = 1.$$

Problem 3

Consider the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and the random variable $X(\omega) = |\omega|$. Define the measures $\mathbb{P}$ and $\mathbb{P}^*$ as

$$\mathbb{P} \coloneqq 0.3\delta_0 + 0.4\delta_1 + 0.3\delta_2, \frac{d\mathbb{P}^*}{d\mathbb{P}} = X.$$

Then what is $\mathbb{P}^*([0,1])$?

Solution:

First of all, we should compute $\mathbb{P}^*$ as follows:

$$\mathbb{P}^* = \mathbb{P} * X = 0.3\delta_0 X(0) + 0.4\delta_1 X(1) + 0.3\delta_2 X(2).$$

Thus, as we are interested in the interval $\mathbb{P}^*([0,1])$, we need to find the probability of $\omega = 0$ and $\omega = 1$.

$$\mathbb{P}^*([0,1]) = 0.3\delta_0 X(0) + 0.4\delta_1 X(1) = 0.3 * 0 + 0.4 * |1| = 0.4.$$

The solution is 0.4*.*

Problem 4

Consider the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and the random variables X and Y by $X(\omega) = 2I_{(-\infty,0]}(\omega)$ and $Y(\omega) = \omega^2$. Define the measures $\mathbb{P}$ and $\mathbb{P}^*$ as

$$\mathbb{P} := 0.3\delta_0 + 0.5\delta_1 + 0.2\delta_{-1}, \frac{d\mathbb{P}^*}{d\mathbb{P}} := X.$$

Then, compute $\mathbb{E}^*(Y)$.

Solution:

First, we need to define the expected value of Y,

$$\mathbb{E}^*(Y) = \sum_{\omega} Y(\omega) * P^*(\omega).$$

Thus, as we already know $Y(\omega)$ (since it is given), we just need to compute P^* (see Problem 3) as follows:

$$\mathbb{P}^* = \mathbb{P} * X = 2 * 0.3\delta_0 + 2 * 0.2\delta_{-1} = 0.6\delta_0 + 0.4\delta_{-1}.$$

Finally, we compute $\mathbb{E}^*(Y)$,

$$\mathbb{E}^*(Y) = \sum_{\omega} Y(\omega) * P^*(\omega) = Y(0) * 0.6 + Y(-1) * 0.4 = 0.4$$

The solution is 0.4.

Problem 5

Consider the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \mu = \#)$ and the measurable function $f(\omega) = e^{-\omega}$. Find

$$\int_{\mathbb{N}} f \, d\mu.$$

Solution:

From the lecture notes we know that on $(\mathbb{N}, 2^{\mathbb{N}}, \#)$, integration reduces to summation:

$$\int_{\mathbb{N}} f \, d\# = \sum_{n \in \mathbb{N}} f(n),$$

provided the series converges. Thus, our problem could be re-written as follows,

$$\int_{\mathbb{N}} f d\# = \sum_{n \in \mathbb{N}} f(n) = \sum_{n \in \mathbb{N}} e^{-n} = \sum_{n \in \mathbb{N}} \left(\frac{1}{e}\right)^{n}.$$

The last expression above is just a geometric series with $r = 1/e$, thus the infinity sum is computed as:

$$\sum_{n \in \mathbb{N}} e^{-n} = \sum_{n \in \mathbb{N}} \left(\frac{1}{e}\right)^{n} = \frac{1}{1 - e^{-1}},$$

which is the solution of the proposed problem.

3.4 Collaborative Review Task

In this module, you are required to complete a collaborative review task, which is designed to test your ability to apply and analyze the knowledge you have learned during the week.

Question

A random variable X is said to have a standard normal distribution if X is absolutely continuous with density given by

$$\frac{d\mathbb{P}_X}{d\lambda_1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad x \in \mathbb{R}.$$

Construct (i.e. give an example of) a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X\colon \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that X has a standard normal distribution. In your example, be sure to verify that X does indeed have a standard normal distribution.