University of Campinas
Insitute of Mathematics, Statistics, and Scientific Computing
Department of Applied Mathematics

MATHEUS FERES TURCHETI

# Measure theoretic probability theory

Campinas
2023/06/28

MATHEUS FERES TURCHETI

# Measure theoretic probability theory

Monograph presented to the Institute of Mathematics, Statistics and Scientific Computing as part of the requirements for obtaining credits in the Supervised Project course, under the guidance of Prof. Christian Horacio Oliveira.

## Abstract

The main goal of the present work is to explore the underlying concepts of probability theory. This exploration is done via the study of measure theory, which allows the formalization of said concepts. Before introducing measure theory, we can already define basic ideas for finite probability spaces. To apply analogous concepts to general probability spaces — the ones that can actually be found in the real world — it is presented an introduction to measure theory with the sole purpose of achieving such generalization.

# Contents

# 1 Introduction

Probability is usually introduced in high school as a way of calculating chances of certain things happening among the many possibilities for a random event, such as rolling a die or flipping a coin. The exploration of more complex "random events" — such as the height of a man drawn at random from a group of people, or the size of a population throughout some time-frame, or the price of a certain asset in the stock market — make it obvious that probability, even as a metric of *chance*, is not quite as simple as counting how many outcomes are favorable to some guess.

The present work has the notable historical advantage of being done after all the theory needed is already developed, so the order and complexity by which the concepts appear in the text are far from a representation of how the field developed. Still, one should imagine that "high school probability" still has to be somehow valid under the generalization needed for the more complex systems (as the ones mentioned above, which are all of great interest for statisticians, physicists, economists, etc.).

For this reason, in the first section we define quite convenient concepts, such as $\sigma$-algebras, probability spaces and expectation. Some of these concepts can seem quite foreign and without reason at first, but they serve as a bridge between the tools used both in the simple flipping-coin-like systems and the complex quantum-mechanics-like systems.

Most of the definitions, lemmas, and theorems from this work are taken from [3] with a bit more mathematical care, when possible. In the second section, it is also used excerpts from [1] with changes made only to fit previous notation used. These excerpts are needed since this section represents an introduction to measure theory, so the concepts and results presented can be used outside the probability context. The second part of this section connects the dots: the probability space is viewed as a measure space, so the definitions and results from measure theory can extend the concepts of the first section to more general scenarios.

The big difference between the said simple and complex systems are the sample spaces from which the events are drawn. Without measure theory, one can only formally represent countable (not necessarily finite) sample spaces. Measure theory allows for general, in particular continuous, sample spaces to be treated analogously to the countable ones.

# 2 Recap of Probability Theory

The goal of this section is to simply define the concepts that will be most useful in the future of this work. All the definitions and results shown can be found in [2, 3], the presentation below tries basically to condense them.

## 2.1 Definitions

**Definition 2.1** ($\sigma$-algebra)**.** For a nonempty set $\Omega$, a $\sigma$-algebra is a collection $\mathcal{F}$ of subsets from $\Omega$ such that

1. $\varnothing \in \mathcal{F}$
2. If $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (closed under complementation)
3. If $A_1, A_2, A_3, \dots \in \mathcal{F} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ (closed under countable union)

**Lemma 2.1.** $\sigma$-algebras are also closed under countable intersection.

*Proof.* Let $A_1, A_2, A_3, \dots \in \mathcal{F}$ for a $\sigma$-algebra $\mathcal{F}$. Note that, for each $k \in \mathbb{N}$,

$$
\begin{aligned}
A_k \in \mathcal{F} &\Rightarrow A_k^c \in \mathcal{F} \text{ (closed under complementation)} \\
&\Rightarrow \bigcup_{k=1}^{\infty} A_k^c \in \mathcal{F} \text{ (closed under countable union)} \\
&\Rightarrow \left( \bigcap_{k=1}^{\infty} A_k \right)^c \in \mathcal{F} \text{ (DeMorgan's laws [2])} \\
&\Rightarrow \bigcap_{k=1}^{\infty} A_k \in \mathcal{F} \text{ (closed under complementation)}
\end{aligned}
$$

Therefore $\sigma$-algebras are closed under countable intersection. □

**Definition 2.2** (Filtration)**.** A finite sequence $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ of $\sigma$-algebras of a nonempty set $\Omega$ is called a filtration if $\mathcal{F}_i \subset \mathcal{F}_k$ for each $i \leq k$.

**Definition 2.3** (Probability measure)**.** Let $\mathcal{F}$ be a $\sigma$-algebra of a nonempty set $\Omega$. A probability measure is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ such that

1. $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\varnothing) = 0$

2. If $A_1, A_2, \dots \in \mathcal{F}$ is a countable sequence of **disjoint** sets, then

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

**Definition 2.4** (Sample Space)**.** A sample space $\Omega$ is a collection of possible results $\omega$ for a certain *random experiment*.

**Definition 2.5** (Event)**.** Any subset $E$ of $\Omega$ is called an event.

**Definition 2.6** (Random variable)**.** Let $\Omega$ be a sample space. A (real) random variable is a function $X : \Omega \to \mathbb{R}$.

**Notation 2.1.** $\mathbb{P}(X = x) \triangleq \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\})$

**Definition 2.7** (Probability Space)**.** A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega$ is a sample space, $\mathcal{F}$ is a $\sigma$-algebra of $\Omega$ (sometimes called event space), and $\mathbb{P}$ is an associated probability measure.

**Definition 2.8** (Preimage)**.** Let $f : A \to B$ be a generic function. The preimage of $f$

- over $Y \subset B$ is the **set** $f^{-1}(Y) = \{x \in A : f(x) \in Y\}$
- for $y \in B$ is the **set** $f^{-1}(y) = \{x \in A : f(x) = y\}$

**Observation 2.1.** From notation (2.1), $\mathbb{P}(X = x) \equiv \mathbb{P}(X^{-1}(x))$

**Definition 2.9.** A $\sigma$-algebra generated by random variable is the set of all preimages of the random variable. For a random variable $X : \Omega \to \mathbb{R}$, we write

$$\sigma(X) = \{X^{-1}(Y) : Y \subset \mathbb{R}\} = \{\{\omega \in \Omega : X(\omega) \in Y\} : Y \subset \mathbb{R}\}$$

**Notation 2.2.** $\{X \in A\} \triangleq \{\omega \in \Omega : X(\omega) \in A\} \Rightarrow \sigma(X) = \{\{X \in Y\} : Y \subset \mathbb{R}\}$

**Definition 2.10.** Let $X$ be a random variable in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathcal{G} \subset \mathcal{F}$ be a sub-$\sigma$-algebra of $\mathcal{F}$. We say that $\sigma(X)$ is $\mathcal{G}$-measurable if every set in $\sigma(X)$ is also in $\mathcal{G}$.

**Definition 2.11** (Induced Measure)**.** For a random variable $X$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and for any given $A \subset \mathbb{R}$, we define the induced measure of $X$ on $A$ to be $\mathcal{L}_X(A) \triangleq \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\} \equiv \mathbb{P}(X \in A)$.

**Definition 2.12** (Cumulative Distribution Function)**.** For a random variable $X$, in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we associate a function $F_X : \mathbb{R} \to [0, 1]$, called cumulative distribution function (CDF), such that $F_X(x) \triangleq \mathbb{P}(X \leq x) \equiv \mathcal{L}_X(\infty, x]$.

## 2.2 Countable Probability Spaces

**Definition 2.13** (Expected Value)**.** The expected value of a random variable $X$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a **countable** $\Omega$, is defined as

$$\mathbb{E}(X) \triangleq \sum_{\omega \in \Omega} X(\omega)\mathbb{P}\{\omega\}$$

It's interesting to note that the idea behind the "expected value" (or *expectation*) is exactly that: if you were to guess the most likely result from you random variable, that would be it. Of course, any half-decent statistician would raise an eyebrow to this affirmation, since we know nothing of the *distribution* (we'll get back to this concept in time) of values of our random variable, so this interpretation could be massively flawed — in the sense that such guess could be a bad choice if the outcome of the guess mattered. For this reason, the expectation of a random variable is to be understood as a metric of its average value for all possible outcomes — in statistics terms, it is called the *mean*.

**Observation 2.2.** If $\Omega$ has cardinality $|\Omega| = n$ (abusing notation here so that either $n \in \mathbb{N}$ or $n = \infty$), then we can write $\{x_1, \ldots, x_n\} \triangleq \{X(\omega) : \omega \in \Omega\}$, therefore

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i)$$

Clearly, one can write the set of values for the random variable in terms of $x_i$ because such set is countable, that is, it is either finite (exactly $n$ elements) or has the cardinality of the natural numbers $\mathbb{N}$, thus it's always possible to find a bijection such that the equivalence remains. Since the numbers $x_i$ are fixed, the case where $|\Omega| = \infty$ implies that $\mathbb{E}(X)$ exists if the series converges, which depends on how the probability measure $\mathbb{P}$ is defined for the random variable.

**Definition 2.14** (Variance)**.** The variance of a random variable $X$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is defined as

$$\mathrm{Var}(X) \triangleq \mathbb{E}[(X - \mathbb{E}(X))^2]$$

**Lemma 2.2.** $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

*Proof.* The proof follows directly from the definition

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}(X))^2] \\
&= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + [\mathbb{E}(X)]^2) \\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + [\mathbb{E}(X)]^2 \\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + [\mathbb{E}(X)]^2 \\
&= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2
\end{aligned}$$

□

# 3 Introduction to Measure Theory

In this section the mathematical rigor is increased. For this reason, most of the definitions, results, and proofs extracted from [3] are rewritten with more mathematical care. Also, the more abstract results are not contained in the previous reference, but can be found in [1].

## 3.1 Lebesgue Integral

**Definition 3.1.** The *Borel $\sigma$-algebra* $\mathcal{B}(\mathbb{R})$ is the **smallest** $\sigma$-algebra containing all open intervals in $\mathbb{R}$. The sets in $\mathcal{B}(\mathbb{R})$ are called Borel.

**Lemma 3.1.** Every closed interval is Borel.

*Proof.* Since $\mathcal{B}(\mathbb{R})$ is a $\sigma$-algebra, then one can note that every countable union of open intervals is Borel, for instance, $(a, b) \cup (c, d)$ with $b < c$ is Borel. On another hand, the complement of a Borel set must also be a Borel set, so

$$[(a, b) \cup (c, d)]^c = (-\infty, a] \cup [b, c] \cup [d, \infty)$$

is Borel. Moreover, $[b, c]$ is Borel. Since $a, b, c, d \in \mathbb{R}$ were arbitrary, any closed interval (or half closed, for that matter) is Borel. □

**Lemma 3.2.** Every set with countably many real numbers is Borel.

*Proof.* Note that, $\forall a \in \mathbb{R}$, the limit

$$\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n}\right)$$

holds because of the Archimedean property of the real numbers, that is, $\forall \varepsilon > 0, \exists n \in \mathbb{N} : \varepsilon < 1/n$, which in turn only is true given $\mathbb{N}$ is not bounded above [4]. So, since $\sigma$-algebras are closed under countable intersections, then $\{a\}$ is Borel.
Now, since $\sigma$-algebras are also closed under countable union, we have that $\bigcup_{k \in M} \{a_k\}$ is Borel, considering $M$ countable and each $a_k \in \mathbb{R}$. □

**Notation 3.1.** Let's denote the extended real line as $\overline{\mathbb{R}} \triangleq \mathbb{R} \cup \{\infty\}$.

**Definition 3.2** (Measure)**.** Let $\mathcal{B}(\mathbb{R})$ be the $\sigma$-algebra of Borel subsets of $\mathbb{R}$. A measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a function $\mu : \mathcal{B}(\mathbb{R}) \to \overline{\mathbb{R}}$ such that

1. $\forall A \in \mathcal{B}(\mathbb{R}),\ \mu(A) \geq 0$
2. $\mu(\varnothing) = 0$
3. If $A_1, A_2, \dots \in \mathcal{B}(\mathbb{R})$ is a countable sequence of **disjoint** sets, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

Note that, by definition (2.3), a probability measure only adds the requirement of an upper bound of 1; thus it is possible to define $\mathbb{P}(A) = \frac{\mu(A)}{\mu(\Omega)}$, $\forall A \in \mathcal{F}$ $\sigma$-algebra of $\Omega$, for a certain measure $\mu$. This covers definition (2.3) and resembles the frequentist approach to probability (but let's not get down that statistics rabbit hole).

**Definition 3.3.** The *length* of an interval $[a, b] \subset \mathbb{R}$ (not necessarily closed), will be given by the well known metric

$$\begin{aligned} d : \overline{\mathbb{R}} \times \overline{\mathbb{R}} &\to [0, \infty] \\ (a, b) &\mapsto |b - a| \end{aligned}$$

**Definition 3.4.** The *Lebesgue Measure* $\mu_0$ of an interval is the length of the interval.

**Lemma 3.3.** A set with countably many points has Lebesgue measure equal to zero.

*Proof.* Let $A = \{a_1, a_2, a_3, \dots\} \subset \mathbb{R}$. Note that $A = \bigcup_{n=1}^{\infty} \{a_n\}$, so $\mu_0(A) = \sum_{n=1}^{\infty} \mu_0\{a_k\}$. Now it suffices to show that, $\forall k$, $\mu_0\{a_k\} = 0$ since $\mu_0(B) \geq 0$ for any set $B$. And that follows from

$$\begin{aligned} \{a_k\} &= \bigcap_{n=1}^{\infty} \left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right) \\ \Rightarrow \{a_k\} &\subset \left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right) \text{ for any } n \in \mathbb{N} \\ \Rightarrow \mu_0\{a_k\} &\leq \mu_0\left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right) \\ \Rightarrow 0 \leq \mu_0\{a_k\} &\leq \frac{2}{n} \\ n \to \infty \Rightarrow \mu_0\{a_k\} &= 0 \end{aligned}$$

Therefore any set with countably many points has Lebesgue measure of zero. □

**Definition 3.5.** We say that a function $f : \mathbb{R} \to \mathbb{R}$ is *Borel-measurable* if

$$\forall A \in \mathcal{B}(\mathbb{R}),\ f^{-1}(A) \in \mathcal{B}(\mathbb{R})$$

In other words, $f$ is Borel-measurable if $\sigma(f) \subset \mathcal{B}(\mathbb{R})$.

**Definition 3.6.** We say that a property is valid **almost everywhere** if the set on which such property fails to hold has measure zero.

**Observation 3.1.** From this point on, **unless explicit stated**, all real functions $f$ will be Borel-measurable, and all *subsets* of $\mathbb{R}$ that will be considered will be Borel.

Now we begin defining the Lebesgue integral by somewhat of an standard procedure [1, 3], which is considering basic functions and assigning interpretations to the integral, in order to then use such functions to build the Lebesgue integral of a general function.

**Definition 3.7.** Let $\mathbb{I}_A : \mathbb{R} \to \mathbb{R}$ such that $\mathbb{I}_A = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$. This is called an **Indicator function**, and its Lebesgue integral is defined as

$$\int_{\mathbb{R}} \mathbb{I}_A \, d\mu_0 \triangleq \mu_0(A)$$

**Definition 3.8.** Let $h(x) = \sum_{k=1}^{n} c_k \mathbb{I}_{A_k}$ for $c_k \in \mathbb{R}$ and $A_k \subset \mathbb{R}$. This is called a **simple function**, and it is such that its Lebesgue integral is defined as

$$\int_{\mathbb{R}} h \, d\mu_0 \triangleq \sum_{k=1}^{n} c_k \mu_0(A_k)$$

**Definition 3.9.** Let $f : \mathbb{R} \to \overline{\mathbb{R}}_+$ be a nonnegative real function, its Lebesgue integral is

$$\int_{\mathbb{R}} f \, d\mu_0 \triangleq \sup \left\{ \int_{\mathbb{R}} h \, d\mu_0 : h \text{ simple and } h(x) \leq f(x), \, \forall x \in \mathbb{R} \right\}$$

Note that the above definition for a nonnegative function matches the *idea* of a lower sum of the *Riemann integral.*

**Definition 3.10** (Lebesgue Integral of any function)**.** Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be a general function. We define its *positive* and *negative parts*, respectively, as

$$f^+(x) \triangleq \max\{f(x), 0\} \text{ and } f^-(x) \triangleq \max\{-f(x), 0\}$$

With this we define the Lebesgue integral of $f$ to be

$$\int_{\mathbb{R}} f\, d\mu_0 \triangleq \int_{\mathbb{R}} f^+\, d\mu_0 - \int_{\mathbb{R}} f^-\, d\mu_0$$

**Definition 3.11.** A function $f$ is said **integrable** if $\int_{\mathbb{R}} |f|\, d\mu_0 < \infty$. That is, neither terms in the right-hand side of definition (3.10) are $\infty$.

**Definition 3.12** (Lebesgue integral on a set)**.** For $f : X \subset \mathbb{R} \to \mathbb{R}$ with $A \subset X$, we define

$$\int_A f\, d\mu_0 \triangleq \int_{\mathbb{R}} \mathbb{I}_A f\, d\mu_0$$

A famous example of non-integrable functions when studying the *Riemann integral* is the function $\mathbb{I}_{\mathbb{Q}}$, also known as the Dirichlet function. Let's consider both the Lebesgue and the Riemann integrals on the interval $[0, 1]$.
Starting with Lebesgue, since $\mathbb{Q}$ is countable, $\mu_0(\mathbb{Q}) = 0$, so

$$\int_{[0,1]} \mathbb{I}_{\mathbb{Q}}\, d\mu_0 = 0$$

As for the Riemann, we need to define an *upper* and a *lower* sum, and then check if their limits coincide. Note that the upper and lower sums are, respectively,

$$U = \sum_{k=1}^{n} 1 \cdot (x_k - x_{k-1}) \text{ and } L = \sum_{k=1}^{n} 0 \cdot (x_k - x_{k-1})$$

Since $\lim_{n\to\infty} U = 1 \neq 0 = \lim_{n\to\infty} L$, we can affirm that $\mathbb{I}_{\mathbb{Q}}$ is *not Riemann integrable*, but it is Lebesgue integrable.

**Definition 3.13.** For a sequence $(a_n), n \in \mathbb{N}$, we say that it has **monotone** convergence to $a$ if either

1. $(a_n) \uparrow a \Leftrightarrow \lim_{n\to\infty} a_n = a$ **and** $\forall n, a_n \leq a_{n+1}$

2. $(a_n) \downarrow a \Leftrightarrow \lim_{n\to\infty} a_n = a$ **and** $\forall n, a_n \geq a_{n+1}$

The proofs for the following theorems that are not shown here can be found on [1].

**Theorem 3.1.** For $f, g$ functions in a measure space $(\Omega, \mathcal{F}, \mu)$

1. If $\alpha, \beta \in \mathbb{R}$ are scalars, then

$$\int_{\Omega} (\alpha f + \beta g)\, d\mu_0 = \alpha \int_{\Omega} f\, d\mu_0 + \beta \int_{\Omega} g\, d\mu_0$$

2. If $f \leq g$ everywhere, then

$$\int_{\Omega} f\, d\mu_0 \leq \int_{\Omega} g\, d\mu_0$$

3. If $A, B$ are disjoint, then

$$\int_{A \cup B} f\, d\mu_0 = \int_A f\, d\mu_0 + \int_B f\, d\mu_0$$

**Theorem 3.2** (Monotone Convergence Theorem)**.** Let $(f_n)$ be a sequence of functions $f_n : D \subset \mathbb{R} \to \mathbb{R}$ such that $0 \leq f_n \uparrow f$ almost everywhere. Then $\int_D f_n d\mu_0 \uparrow \int_D f d\mu_0$, that is

$$\int_D f\, d\mu_0 = \lim_{n \to \infty} \int_D f_n\, d\mu_0$$

*Proof.* Since $(f_n)$ converges monotonically to $f$ for all $x \in A \subset D$ with $\mu_0(A^c) = 0$, so $f_n \mathbb{I}_A \uparrow f \mathbb{I}_A$ everywhere, then

$$\begin{aligned}
\int_D f_n\, d\mu_0 &= \int_A f_n\, d\mu_0 + \int_{A^c} f_n\, d\mu_0 \\
&= \int_D f_n \mathbb{I}_A\, d\mu_0 \ \because\ \mu_0(A^c) = 0 \\
&\uparrow \int_D f \mathbb{I}_A\, d\mu_0 \\
&= \int_A f\, d\mu_0 + \int_{A^c} f\, d\mu_0 \ \because\ \mu_0(A^c) = 0 \\
\therefore \int_D f_n\, d\mu_0 &\uparrow \int_D f\, d\mu_0
\end{aligned}$$

Which is effectively equivalent to swapping the integral with the limit. □

**Theorem 3.3** (Fatou's Lemma)**.** Let $(f_n)$ be a sequence of *non-negative* functions, then

$$\int_{\mathbb{R}} \liminf_{n\to\infty} f_n \, d\mu_0 \leq \liminf_{n\to\infty} \int_{\mathbb{R}} f_n \, d\mu_0$$

**Theorem 3.4** (Dominated Convergence Theorem)**.** Let $(f_n)$ be a sequence of functions $f_n : D \subset \mathbb{R} \to \mathbb{R}$ such that $|f_n| \leq g$ almost everywhere, with $g$ integrable. If $f_n \to f$ almost everywhere, then $(f_n)$ and $f$ are integrable with

$$\int_D f_n \, d\mu_0 \to \int_D f \, d\mu_0$$

**Notation 3.2.** From this point on, unless explicit stated, all integrals will be Lebesgue, so we can use the same notation of the Riemann integral without confusion

$$\int_{[a,b]} f \, d\mu_0 \triangleq \int_{[a,b]} f(x) d\mu_0(x) \equiv \int_a^b f(x) dx$$

Specially considering that the Lebesgue integral is defined for more functions and, if the Riemann integral exists, it will yield the same result.

## 3.2 General Probability Spaces

Equipped with the Lebesgue integral, we can define more general probability spaces by allowing the sample space $\Omega$ to be uncountable.

**Definition 3.14** (Integral of a Random Variable)**.** Let $X$ be a random variable in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

1. if $X(\omega) = \mathbb{I}_A(\omega)$, for $A \in \mathcal{F}$, then $X$ is an indicator and

$$\int_\Omega X \, d\mathbb{P} = \mathbb{P}(A)$$

2. if $X(\omega) = \sum_{k=1}^{n} c_k \mathbb{I}_{A_k}(\omega)$, then $X$ is simple and

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

3. if $X : \Omega \to \mathbb{R}_+$, then

$$\int_\Omega X \, d\mathbb{P} \triangleq \sup \left\{ \int_\Omega Y \, d\mathbb{P} : Y \text{ simple and } Y(\omega) \leq X(\omega), \, \forall \omega \in \Omega \right\}$$

4. if $X$ is a general random variable, we define its integral the same way we define a real function, provided that the right-hand side does not have any $\infty$:

$$\int_\Omega X \, d\mathbb{P} = \int_\Omega X^+ \, d\mathbb{P} - \int_\Omega X^- \, d\mathbb{P}$$

5. for $A \in \mathcal{F}$, we have

$$\int_A X \, d\mathbb{P} = \int_\Omega \mathbb{I}_A X \, d\mathbb{P}$$

**Observation 3.2.** All the theorems for a general *measure* space can be translated to the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, including the convergence ones — the only change in language needed is that *almost everywhere* becomes *almost surely*.

**Definition 3.15.** An event $E \in \mathcal{F}$ happens **almost surely** if $\mathbb{P}(E) = 1$.

**Definition 3.16** (Expected Value)**.** The expectation of a random variable $X$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is defined as

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

Note that the **variance** definition (2.14) still works for a *general* random variable. Also, item 1 from definition (3.14) gives away an often useful identity of an indicator random variable, which is that its expected value is the same as the probability of the event it indicates.

A very useful way of characterizing random variables is via their "distribution". Let's look at an example to build an intuition about an important theorem from measure theory.

**Example 3.1** (Standard normal distribution)**.** Consider the function $\varphi : \mathbb{R} \to \mathbb{R}$ such that

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

If we consider a random variable $X$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\Omega = \mathbb{R} \Leftrightarrow \mathcal{F} = \mathcal{B}(\mathbb{R})$, and $\mathbb{P}$ such that

$$\mathbb{P}(X \in A) \equiv \mathcal{L}_X(A) = \int_A \varphi \, d\mu_0$$

then we can say that $X$ has a standard normal distribution.
We don't know what a "standard normal distribution" means, but it's certainly related to something we imposed on the probability space defined. Since there's nothing special with considering the real line as the sample space $\Omega$ (and thus having the Borel $\sigma$-algebra as $\mathcal{F}$), meaning we could imagine many different random variables with that constraint, we have to assume that such concept of distribution is strictly related to the way we defined the probability measure.
Note that, if $A = (-\infty, x]$ for some $x \in \mathbb{R}$, then

$$\begin{aligned}
\int_A \varphi \, d\mu_0 &= \mathbb{P}(X \in A) \\
\int_{-\infty}^{x} \varphi \, d\mu_0 &= \mathbb{P}(X \leq x) \equiv \mathcal{L}_X(-\infty, x] \\
\int_{-\infty}^{x} \varphi(t) dt &= F_X(x) \\
\Rightarrow \varphi(x) &= \frac{dF_X}{dx} \equiv \frac{d\mathcal{L}_X}{d\mu_0}\bigg|_A = \frac{d\mathbb{P}}{d\mu_0}\bigg|_{\{X \leq x\}}
\end{aligned}$$

The last equation definitely has some abuse of notation, given we didn't define the derivative with respect to a *measure*, which is called the **Radon-Nikodym derivative** [1]. Properly defining and exploring the Radon-Nikodym derivative is outside the scope of this work, however we can interpret it as the way the measure $\mathbb{P}$ is weighted along the real line according to the density $\varphi$ [3].

**Definition 3.17.** If $\varphi : \mathbb{R} \to \mathbb{R}_+$ is a function such that $\int_{\mathbb{R}} \varphi \, d\mu_0 = 1$ then $\varphi$ is called a **density**.

Note that, in the example above, it was sufficient to use the fundamental theorem of calculus [4] to obtain a relationship between the cumulative *distribution* function (definition 2.12) and the function of interest $\varphi$. So indeed, the distribution of a random variable $X$ is determined by how the probability measure is defined for it, and that can be given by the function $F_X$ or by its derivative [2].

**Definition 3.18.** For a random variable $X$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with cumulative distribution function $F_X$, we define the **probability density function** (PDF) of $X$ as

$$f_X(x) = \frac{dF_X}{dx}$$

**Definition 3.19.** Given a random variable $X$ with a PDF $f_X$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the probability of $X$ falling into $A \in \mathcal{F}$ is given by

$$\mathbb{P}(X \in A) \triangleq \int_A f_X \, d\mu_0$$

Obviously, we used this definition in the example (3.1), but the more useful information here is that this is general, meaning we can use this definition to calculate the probabilities for any random variable $X$ for which we know its PDF or CDF.

Another interesting approach to densities and how they connect two different measures is considering a nonnegative $\mu_0$-measurable function $\delta : \mathbb{R} \to \mathbb{R}_+$ such that it defines another measure $\nu$ via

$$\nu(A) = \int_A \delta \, d\mu_0, \; A \in \mathcal{B}(\mathbb{R})$$

By this alternate definition, we say that $\nu$ has *density* $\delta$ with respect to $\mu_0$ [1]. We are interested in the cases where $\delta$ is also $\mu_0$-integrable with unitary integral (that is, for the whole real line).

**Theorem 3.5.** Consider a function $f$ in a measure space $(\Omega, \mathcal{F}, \nu)$, where $\nu$ has density $\delta$ with respect to $\mu_0$, with $\delta$ integrable with respect to $\mu_0$, then, for $A \in \mathcal{F}$

$$\int_A f\, d\nu = \int_A f\delta\, d\mu_0$$

*Proof.* For simplicity, let's consider $A = \Omega$ at first. If $f = \mathbb{I}_B$ then

$$\int_\Omega f\, d\nu = \nu(B) = \int_B \delta\, d\mu_0$$

Which holds by definition. If $f$ is a simple function

$$\begin{aligned}
\int_\Omega f\, d\nu &= \int_\Omega \sum_{k=1}^{n} c_k \mathbb{I}_k\, d\nu \\
&= \sum_{k=1}^{n} c_k \int_\Omega \mathbb{I}_k\, d\nu \\
&= \sum_{k=1}^{n} c_k \int_\Omega \mathbb{I}_k \delta\, d\mu_0 \\
&= \int_\Omega \sum_{k=1}^{n} c_k \mathbb{I}_k \delta\, d\mu_0 \\
&= \int_\Omega f\delta\, d\mu_0
\end{aligned}$$

If $f$ is non-negative, consider a sequence $(f_n)$ of simple functions such that $f_n \uparrow f$ almost everywhere.

$$\begin{aligned}
\lim_{n\to\infty} \int_\Omega f_n\, d\nu &= \lim_{n\to\infty} \int_\Omega f_n\delta\, d\mu_0 \ \because\ f_n \text{ is simple} \\
\int_\Omega f\, d\nu &= \int_\Omega f\delta\, d\mu_0 \ \because\ \text{monotone convergence theorem}
\end{aligned}$$

Finally, if $f$ is a general *integrable* function, then apply the logic above for the negative and positive parts.

Now, if we do $f \mapsto f\mathbb{I}_A$ then we can consider any set $A$ with the same arguments. □

Clearly, if $\nu = \mathbb{P}$ we are basically talking about a practical way of calculating expected

values and probabilities for a random variable $f$ for which we know the PDF $\delta$.
For a random variable $X$ which represents a simple numerical result from an experiment, that is, the probability space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$, and we can always re-write the random variable so that $X(\omega) = \omega$. Considering that $X$ has PDF $f_X$ we can write the expected value definition in the well known manner [2]

$$\mathbb{E}(X) \triangleq \int_{\mathbb{R}} X \, d\mathbb{P} = \int_{\mathbb{R}} X f_X \, d\mu_0 \equiv \int_{-\infty}^{\infty} x f_X(x) \, dx$$

# 4 Conclusions

The purpose of this work was to build the necessary basis of advanced probability theory necessary for the study of interesting applications, such as modelling stochastic processes in continuous time and space. In particular, the choice of [3] as bibliography gives away this intent, as the goal is to proceed studying stochastic calculus in the context of option pricing models, eventually arising at the infamous Black-Scholes-Merton model.
It's also a hope that this report could serve as a simple exposition of measure theoretic probability for applied mathematics students with starting knowledge in analysis and probability [2, 4].

# References


[1] Patrick Billingsley. *Probability and Measure*. Third Edition. Wiley, 1995.

[2] Sheldon Ross. *A First Course in Probability*. $8^{th}$*edition*. Prentice Hall, 2010.

[3] Steven E. Shreve. *Stochastic Calculus for Finance I: The Binomial Asset Pricing Model*. Springer Finance, 2004.

[4] Michael Spivak. *Calculus*. Fourth Edition. Publish or Perish, Inc., 2008.