# Notes in Stat 235
# Mathematics in Statistics
# from Prof. Dorado's Lecture

Joseph S. Tabadero, Jr.

# Contents

# Notations

| | |
|---|---|
| $\subset$ | subset |
| $\forall$ | for all, for every |
| $\exists$ | there exists, there exists at least one, for some |
| $\ni$ | such that |
| $\cup$ | union |
| $\cap$ | intersection |
| $\setminus$ | set difference operator |
| $\Lambda$ | Q.E.D, end of proof |
| $\mathcal{F}$ | A $\sigma$-algebra, unless stated otherwise |

CHAPTER 1

# Lecture 1: Introduction

## 1. Review on sets

1.0.1. *Set operations.*

**Union:** $A \cup B = \{x \mid x \in A \,\text{or}\, x \in B\}$
**Intersection:** $A \cap B = \{x \mid x \in A \,\text{and}\, x \in B\}$
**Difference:** $A \setminus B = \{x \mid x \in A \,\text{and}\, x \notin B\}$
**Subset:** $A \subset B$ if $x \in A$ implies $x \in B$

REMARKS 1.1. (1) *The following are some uses of the finite operations on sets in probability theory.*

(a) $P(|X| \leq a) = P(\{X \leq a\} \cap \{X \geq -a\})$
(b) $P(|X| > a) = P(\{X > a\} \cup \{X < -a\})$
(c) $P(\min(X_1, X_2, \ldots, X_n) > a) = P(\bigcap_{i=1}^n \{X_i > a\})$
(d) $P(\max(X_1, X_2, \ldots, X_n) > a) = P(\bigcup_{i=1}^n \{X_i > a\})$

(2) *However, there are examples of results that cannot be expressed as a finite operations on sets. For example, suppose* $S_n = X_1 + X_2 + \cdots + X_n$ *of random variables* $X_i$, $i = 1, 2, \ldots, n$. *Then* $P(\lim_{n\to\infty} \frac{S_n}{n} = \mu) = 1$. *This is known as the Strong Law of Large Numbers (SLLN). SLLN cannot be expressed as a finite operations on sets. De Morgan's Identities and the Distributive Laws on countably infinite sets enables us to express SLLN as a countable number of operations on sets.*

1.0.2. *De Morgan's Identities.*

$$\left[\bigcup_{i=1}^{\infty} A_i\right] = \bigcap_{i=1}^{\infty} A_i^c \tag{1}$$

and

$$\left[\bigcap_{i=1}^{\infty} A_i\right] = \bigcup_{i=1}^{\infty} A_i^c \tag{2}$$

1.0.3. *Distributive Law.*

$$A \cap \left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} (A \cap B_i) \tag{3}$$

and

$$A \cup \left(\bigcap_{i=1}^{\infty} B_i\right) = \bigcap_{i=1}^{\infty} (A \cup B_i) \tag{4}$$

## 2. Algebra

DEFINITION 1.1 (ALGEBRA). *Let* $\Omega$ *be a nonempty set. A class* $\mathcal{F}$ *of subsets of* $\Omega$ *is called an algebra if*

*i.* $\Omega \in \mathcal{F}$,
*ii.* $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$,
*iii.* $A, B \in \mathcal{F}$ *implies* $A \cup B \in \mathcal{F}$.

DEFINITION 1.2 ($\sigma$-ALGEBRA). *A class* $\mathcal{F}$ *of subsets of* $\Omega$ *is called a* $\sigma$*-algebra if* $\mathcal{F}$ *is an algebra and* $A_1, A_2 \ldots \in \mathcal{F}$ *implies* $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$. *The pair* $(\Omega, \mathcal{F})$ *is called a measurable space. The elements of a* $\sigma$*-algebra* $\mathcal{F}$ *are called measurable sets.*

REMARKS 1.2. (1) *Algebras are closed under any finite set of operations. E.g.* $A \setminus B = A \cap B^c \in \mathcal{F}$.

(2) *$\sigma$-algebras are collection of subsets that are closed under countably infinite operations.*
(3) *The smallest $\sigma$-algebra is the trivial $\sigma$-algebra, $\mathcal{F} = \{\Omega, \varnothing\}$. It is constructed as follows: Such smallest $\sigma$-algebra must contain $\Omega$ and its complement $\Omega^c = \varnothing$. So the smallest such $\sigma$-algebra must contain at least these two elements. It is easy to see that these two elements are enough to form a $\sigma$-algebra.*

PROOF. (*i*) $\Omega \in \mathcal{F}$ by definition.
(*ii*) Let $E \in \mathcal{F}$. Then

$$E^c = \begin{cases} \varnothing & \text{if} \quad E = \Omega \\ \Omega & \text{if} \quad E = \varnothing \end{cases}$$

In each of these cases, $E^c \in \mathcal{F}$.
(*iii*) Suppose $E_1, E_2, \ldots \in \mathcal{F}$, then

$$\bigcup_{i=1}^{\infty} E_i = \begin{cases} \varnothing & \text{if} \quad E_i = \varnothing \; \forall i \\ \Omega & \text{if} \quad E_i = \Omega \text{ for at least one } i \end{cases}$$

This completes the proof showing that the trivial $\sigma$-algebra is the smallest $\sigma$-algebra. □

(4) *The largest $\sigma$-algebra on $\Omega$ is the power set $\mathcal{P}(\Omega)$, the collection of all subsets of $\Omega$.*

EXAMPLE 1.1. *Let $A \subset \Omega$. Then $\mathcal{F} = \{\varnothing, A, A^c, \Omega\}$ is a $\sigma$-algebra.*

PROOF. i. $\Omega \in \mathcal{F}$ by definition.
ii. If $E \in \mathcal{F}$, then

$$E^c = \begin{cases} \Omega & \text{if} \quad E = \varnothing \\ A^c & \text{if} \quad E = A \\ A & \text{if} \quad E = A^c \\ \varnothing & \text{if} \quad E = \Omega \end{cases}$$

Whatever the case, $E^c \in \mathcal{F}$.
iii. Suppose a sequence of sets $E_1, E_2, E_3, \ldots \in \mathcal{F}$. Then,

$$\bigcup_{i=1}^{\infty} = \begin{cases} \varnothing & \text{if} \quad E_i \in \mathcal{F} \quad \forall i \\ \Omega & \text{if} \quad \exists i \ni E_i = \Omega \\ A & \text{if} \quad E_i \neq \Omega \, \forall i, E_i \neq A^c \, \forall i, \exists i \ni E_i = A \\ A^c & \text{if} \quad E_i \neq \Omega \, \forall i, E_i \neq A \, \forall i, \exists i \ni E_i = A^c \\ \Omega & \text{if} \quad \exists i \ni E_i = A, \exists j \ni E_j = A^c, E_i \neq \Omega \, \forall i \end{cases}$$

□

EXAMPLE 1.2. *Let $\mathcal{F}$ consists of the finite and cofinite ($A$ being cofinite if $A^c$ is finite) subsets of a set $\Omega$. Then $\mathcal{F}$ is an algebra. If $\Omega$ is finite then $\mathcal{F}$ is also a $\sigma$-algebra. If $\Omega$ is infinite, however, then $\mathcal{F}$ is not a $\sigma$-algebra.*

PROOF. i. $\Omega$ is cofinite since $\Omega^c = \varnothing$ which is finite. Thus, $\Omega \in \mathcal{F}$.
ii. Let $A \in \mathcal{F}$. Suppose $A$ is finite. Then $A^c$ is cofinite since $(A^c)^c = A$ and hence $A^c \in \mathcal{F}$. Now if $A$ is cofinite. Then $A^c$ is finite and hence, in $\mathcal{F}$.
iii. Let $A, B \in \mathcal{F}$. Then $A^c \in \mathcal{F}$ and $B^c \in \mathcal{F}$ from part (ii) above. By De Morgan's Law, we can replace the condition $A \cup B \in \mathcal{F}$ with $A \cap B \in F$. To see this, note that since $A^c \in \mathcal{F}$ and $B^c \in \mathcal{F}$, we may show that $\mathcal{F}$ is a $\sigma$-algebra if together with parts i and ii above, we have $A^c \cup B^c \in \mathcal{F}$ if and only if $(A^c \cup B^c)^c = (A^c)^c \cap (B^c)^c = A \cap B \in \mathcal{F}$. Now, if one of $A$ or $B$ is finite, say without loss of generality that $A$ is finite, then $A \cap B \subset A$ is finite. If both $A$ and $B$ are cofinite, then $A^c$ and $B^c$ are both finite, and so $A^c \cup B^c$ is also finite since the countable union of finite sets is finite. Therefore, $(A^c \cup B^c)^c = A \cap B$ is cofinite. In both cases, $A \cap B \in \mathcal{F}$.

Therefore, $\mathcal{F}$ is an algebra.

If $\Omega$ is finite, and suppose $A_1, A_2, A_3 \cdots \in \mathcal{F}$. Note that $A_i \in \mathcal{F}$ for every $i$ if and only if $A_i^c \in \mathcal{F}$ for every $i$. So that $\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$ if and only if $\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathcal{F} = \left(\bigcap_{i=1}^{\infty} A_i\right) \in \mathcal{F}$. If at least one of the $A_i$'s are finite, say $A_k$, then $\left(\bigcap_{i=1}^{\infty} A_i\right) \subset A_k$ and hence $\left(\bigcap_{i=1}^{\infty} A_i\right)$ is finite. Therefore, $\left(\bigcap_{i=1}^{\infty} A_i\right) \in \mathcal{F}$. If each $A_i$ is cofinite, then each $A_i^c$ is finite. Therefore $\bigcup_{i=1}^{\infty} A_i^c$ is finite since the union of a countable collection of finite sets is finite. Therefore, $(\bigcup_{i=1}^{\infty} A_i^c)^c = (\bigcap_{i=1}^{\infty} A_i)$ is cofinite and hence in $\mathcal{F}$. Therefore, $\mathcal{F}$ is a $\sigma$-algebra.

Suppose $\Omega$ is infinite. Without loss of generality, we can say that $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \ldots\}$. Let $A_i = \{\omega_{2i}\}$ for each $i$, then $A_i$ is finite for each $i$ and hence $A_i \in \mathcal{F}$. But $\bigcup A_i = \{\omega_2, \omega_4, \omega_6, \ldots\}$ which is infinite, and $(\bigcup A_i)^c = \{\omega_1, \omega_3, \omega_5, \ldots\}$ which is also infinite. Therefore, $\bigcup A_i = \{\omega_2, \omega_4, \omega_6, \ldots\}$ is neither finite nor cofinite and hence is not in $\mathcal{F}$. Therefore, if $\Omega$ is infinite, $\mathcal{F}$ is not a $\sigma$-algebra.

□

Note that instead of determining whether or not $A \cap B \in \mathcal{F}$ in part ii of the proof that $\mathcal{F}$ is an algebra, we can see if $A \cup B \in \mathcal{F}$. If $A$ and $B$ rae finite, then $A \cup B$ is finite. If $A^c$ or $B^c$ is finite, then $A^c \cap B^c = (A \cup B)^c$ is finite, and hence, $A \cup B$ is cofinite and thus in $\mathcal{F}$. In both cases, $A \cup B \in \mathcal{F}$.

EXAMPLE 1.3. *Let $\mathcal{F}$ consists of the countable and co-countable ($A$ being co-countable if $A^c$ is countable) subsets of a set $\Omega$. Then $\mathcal{F}$ is a $\sigma$-algebra.*

PROOF. (*i*) $\Omega \in \mathcal{F}$ because $\Omega^c = \varnothing$ is countable so $\Omega$ is co-countable.

(*ii*) Let $A \in \mathcal{F}$. If $A$ is countable, then $A^c$ is co-countable and hence $A^c \in \mathcal{F}$. If $A$ is co-countable, then $A^c$ is countable and hence in $\mathcal{F}$. In both cases, $A^c \in \mathcal{F}$.

(*iii*) Let $A_1, A_2, A_3, \ldots \in \mathcal{F}$. Suppose $A_i^c$ is countable for at least one $i$, say $k$, such that $A_k^c$ is countable. Then $\bigcap_{i=1}^{\infty} A_i^c \subset A_k^c$. Therefore, $\bigcap_{i=1}^{\infty} A_i^c$ is countable. Hence, $(\bigcap_{i=1}^{\infty} A_i^c)^c = \bigcup_{i=1}^{\infty} A_i$ is co-countable. If $A_i$ is countable for each $i$ then $\bigcup_{i=1}^{\infty} A_i$ is countable since the countable union of countable sets is countable. In either case, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. Therefore, $\mathcal{F}$ is a $\sigma$-algebra.

□

In part (iii), we may also argue (Prof. Dorado did) as follows. Suppose $A_i$ is countable for some $i$, say for $i = k$. Then $(\bigcup_{i=1}^{\infty} A_i^c)^c = \bigcap_{i=1}^{\infty} A_i \subset A_k$ is countable. This implies that $\bigcup_{i=1}^{\infty} A_i^c$ is co-countable. Suppose now that $A_i$ is co-countable for every $i$. Then $A_i^c$ is countable for every $i$. So $\bigcup_{i=1}^{\infty} A_i^c$ is countable and so $(\bigcup_{i=1}^{\infty} A_i^c)^c = \bigcap_{i=1}^{\infty}$ is co-countable. In both cases, $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

DEFINITION 1.3 ($\sigma$-ALGEBRA GENERATED BY $\mathcal{C}$). *Let $\mathcal{C}$ be a class of subsets of $\Omega$. Then $\sigma(\mathcal{C})$, the $\sigma$-algebra generated by $\mathcal{C}$, is the smallest $\sigma$-algebra $\mathcal{F}$ on $\Omega$ such that $\mathcal{C} \subset \mathcal{F}$.*

THEOREM 1.1. *The $\sigma$-algebra generated by $\mathcal{C}$ is the intersection of all $\sigma$-algebras on $\Omega$ containing $\mathcal{C}$.*

PROOF. Define $\bigcap \mathcal{F} = \bigcap \{\mathcal{F} \mid \mathcal{F}$ is a $\sigma$-algebra of subsets of $\Omega$ and $\mathcal{C} \subset \mathcal{F}\}$. We need to show that this is a $\sigma$-algebra and that it is the smallest $\sigma$-algebra on $\Omega$ containing $\mathcal{C}$.

Let us prove first that $\bigcap \mathcal{F}$ is indeed a $\sigma$-algebra.

(*i*) $\Omega \in \mathcal{F}$ for any $\sigma$-agebra and hence, $\Omega \in \bigcap \mathcal{F}$.

(*ii*) Suppose $A \in \mathcal{F}$ with $\mathcal{F}$ a $\sigma$-algebra and $\mathcal{C} \subset \mathcal{F}$. Then $A^c \in \mathcal{F}$. Since $\mathcal{F}$ is arbitrary, then $A^c \in \bigcap \mathcal{F}$.

(*iii*) Now let $A_1, A_2, A_3, \ldots \in \bigcap \mathcal{F}$. Then for any $\mathcal{F}$ in the intersection, $A_1, A_2, A_3, \ldots \in \mathcal{F}$. Since $\mathcal{F}$ is a $\sigma$-algebra then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ and therefore, $\bigcup_{i=1}^{\infty} A_i \in \bigcap \mathcal{F}$. Therefore, $\bigcup \mathcal{F}$.

We now show that $\bigcap \mathcal{F}$ is the smallest such $\sigma$-algebra and hence $\sigma(\mathcal{C}) = \bigcap \mathcal{F}$.

Clearly, $\mathcal{C} \subset \bigcap \mathcal{F}$ by definition. Let $\mathcal{G}$ be a $\sigma$-algebra with $\mathcal{C} \subset \mathcal{G}$. Therefore $\mathcal{G}$ is one of the $\sigma$-algebras in the intersection and thus, since $\mathcal{F}$ is arbitrary, $\bigcap \mathcal{F} \subset \mathcal{G}$. Therefore, $\sigma(\mathcal{C}) = \bigcap \mathcal{F}$. □

EXAMPLE 1.4. *The Borel $\sigma$-algebra $\mathcal{B}$ on $\mathbb{R}$ is defined as the smallest $\sigma$-algebra containing the open sets in $\mathbb{R}$. An open set in $\mathbb{R}$ is either empty or a disjoint union of a countable collection of open intervals. This $\sigma$-algebra is generated by the collection $\pi(\mathbb{R})$ defined by*

$$\pi(\mathbb{R}) = \{(-\infty, x] \mid x \in \mathbb{R}\}. \tag{5}$$

*The elements of $\mathcal{B}$ are called Borel sets.*

REMARKS 1.3. 1. *Borel $\sigma$-algebra is the most important $\sigma$-algebra for mathematics in statistics.*

2. *To get an idea about the first note, we recall the standard normal density*

$$\begin{aligned} F(a) &= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{x^2/2}\,dx \\ &= P(X \le a) \\ &= P(X \in (-\infty, a]) \end{aligned}$$

Second meeting: February 17, 2016.

DEFINITION 1.4 ($\pi$-SYSTEM, $d$-SYSTEM). *Let $\Omega$ be a nonempty set. A collection $\mathcal{J}$ of subsets of $\Omega$ is called a $\pi$-system if whenever $A, B \in \mathcal{J}$ we have $A \cap B \in \mathcal{J}$. A collection $\mathcal{D}$ of subsets of $\Omega$ is called a $d$-system if*

*i.* $\Omega \in \mathcal{D}$,
*ii.* $A, B \in \mathcal{D}$ *and* $A \subset B$ *implies* $B \setminus A \in \mathcal{D}$,
*iii.* $A_1, A_2, \ldots \in \mathcal{D}$ *and* $A_1 \subset A_2 \subset \cdots$ *implies* $\bigcup_{i=i}^{\infty} A_i \in \mathcal{D}$.

REMARKS 1.4. $\pi(\mathbb{R})$ *of Example* (1.4) *is termed as such because it is a $\pi$-system.*

PROOF. It is easy to show that $\pi(\mathbb{R})$ is a $\pi$-system. We need to show that for any two sets $A, B \in \pi(\mathbb{R})$, we must have $A \cap B \in \pi(\mathbb{R})$.

Now, let $A, B \in \pi(\mathbb{R})$, then for some $x, y \in \mathbb{R}$ we have $A = (-\infty, x]$ and $B = (-\infty, y]$. Without loss of generality, assume that $x < y$. Then $A = (-\infty, x] \subset (-\infty, y] = B$ and so $A \cap B = (-\infty, x] \cap (-\infty, y] = (-\infty, x] \in \pi(\mathbb{R})$. Therefore, $\pi(\mathbb{R})$ is a $\pi$-system. ☐

$\sigma$-algebra is still a difficult concept and we wanted to work with some much easier concepts in order to deal more easily with some proofs. This is the motivation for working with with $d$-systems and $\pi$-systems because it is easier to check whether a collection of sets is a $d$-system or a $\pi$-system than it is to check directly if it is a $\sigma$-algebra. Such is the spirit of Theorem (1.2).

THEOREM 1.2. *A collection $\mathcal{F}$ of subsets of $\Omega$ is a $\sigma$-algebra if and only if $\mathcal{F}$ is both a $\pi$-system and a $d$-system.*

PROOF. Let us suppose first that $\mathcal{F}$ is a $\sigma$-algebra. We will show that it is both a $\pi$-system and a $d$-system. We have shown this a lot of times but we will show the reasoning again.

Suppose $A, B \in \mathcal{F}$ and that $\mathcal{F}$ is a $\sigma$-algebra. Then since $\mathcal{F}$ is a $\sigma$-algebra, we also have $A^c, B^c \in \mathcal{F}$, and, thus, $A^c \cup B^c \in \mathcal{F}$. This means that, by De Morgan's Law, $A \cap B = (A^c \cup B^c)^c \in \mathcal{F}$. Therefore, $\mathcal{F}$ is a $\pi$-system.

Condition (i) is the same for both $d$-system and $\sigma$-algebra. If $A, B \in \mathcal{F}$, then, again, using the same reasoning as above, we know that $B \cap A^c \in \mathcal{F}$. Therefore, $B \setminus A = B \cap A^c \in \mathcal{F}$ and so condition (ii) of the $d$-system is satisfied. Finally, condition (iii) of the $\sigma$-algebra is a more general statement than that of condition (iii) of the $d$-system. Therefore, condition (iii) of the $\sigma$-algebra satisfies condition (iii) of the $d$-system. Therefore, $\mathcal{F}$ is also a $d$-system.

Suppose now that $\mathcal{F}$ is both a $\pi$ and a $d$-system.

(*i*) $\Omega \in \mathcal{F}$ by definition of a $d$-system.
(*ii*) Now suppose $A \in \mathcal{F}$. Then $A^c = \Omega \setminus A$. But since $\Omega \in \mathcal{F}$ and $A \subset \Omega$, by property (ii) of a $d$-system, then $A^c \in \mathcal{F}$.
(*iii*) Now let $A_1, A_2, A_3, \ldots$ be in $\mathcal{F}$. Note that we don't know if $\{A_i\}$ is increasing. So we create a sequence of sets that are increasing as follows. Define

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_1 \cup A_2 \\ B_3 &= A_1 \cup A_2 \cup A_3 \\ &\vdots \\ B_n &= A_1 \cup A - 2 \cup A_3 \cup \cdots \cup A_n \\ &\vdots \end{aligned}$$

Then $B_1 \subset B_2 \subset B_3 \subset \cdots$. Now, for any $n \geq 1$, $B_n = A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i = \left(\bigcap_{i=1}^{n} A_i^c\right)^c = \Omega \setminus \left(\bigcap_{i=1}^{n} A_i^c\right)$. Now, $\mathcal{F}$ is a $\pi$-system, so $\bigcap_{i=1}^{n} A_i^c \in \mathcal{F}$. Also, $\bigcap_{i=1}^{n} A_i^c \subset \Omega$, therefore, by property (ii) of a $d$-system, $\Omega \setminus (\bigcap_{i=1}^n A_i^c) \in \mathcal{F}$. Now, $\{B_i\}$ is an increasing sequence of sets that are in $\mathcal{F}$. Therefore, $\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty B_i \in \mathcal{F}$ by property (iii) of a $d$-system. Therefore, $\mathcal{F}$ is a $\sigma$-algebra.

□

THEOREM 1.3. *If $\mathcal{J}$ is a $\pi$-system and $\mathcal{D}$ is a $d$-system with $\mathcal{J} \subset \mathcal{D}$, then $\sigma(\mathcal{J}) \subset \mathcal{D}$.*

REMARKS 1.5. *$\sigma(\mathcal{J}) \subset \mathcal{D}$ means that $\mathcal{D}$ contains the smallest $\sigma$-algebra contained in $\mathcal{J}$.*

PROOF. To show $\sigma(\mathcal{J}) \subset \mathcal{D}$, we need only show that $\mathcal{D}$ is a $\pi$-system. Why? If $\mathcal{D}$ is a $\pi$-system, then, since it is also a $d$-system, then it is a $\sigma$-algebra. $\sigma(\mathcal{J})$ and $\mathcal{D}$ both being $\sigma$-algebras and $\sigma(\mathcal{J})$ being the smallest $\sigma$-algebra containing $\mathcal{J}$ means $\sigma(\mathcal{J}) \subset \mathcal{D}$.

Now without loss of generality, we can also let $\mathcal{D}$ be the smallest such $d$-system that contains $\mathcal{J}$. If it holds that $\sigma(\mathcal{J}) \subset \mathcal{D}$ for this $\mathcal{D}$ then it will hold for other bigger $\mathcal{D}$'s.

Consider $\mathcal{F} = \{A \in \mathcal{D} \mid A \cap B \in \mathcal{D} \text{ for all } B \in \mathcal{J}\}$. We need to show that $\mathcal{J} \subset \mathcal{F}$ and $\mathcal{F}$ is a $d$-system.

Note than since $\mathcal{J}$ is a $\pi$-system, for $A \in \mathcal{J}$ then $A \cap B \in \mathcal{J} \subset \mathcal{D}$ for all $B \in \mathcal{J}$. Hence, $A \in \mathcal{F}$ and $\mathcal{J} \subset \mathcal{F}$.

We now proceed to show that $\mathcal{F}$ is a $d$-system.

(*i*) Now $\Omega \cap B = B \in \mathcal{J}$ for all $B \in \mathcal{J}$ and hence $\Omega \in \mathcal{F}$.

(*ii*) Moreover, if $A, C \in \mathcal{F}$ and $A \subset C$ then for any $B \in \mathcal{J}$, $(C \setminus A) \cap B = (C \cap B) \setminus (A \cap B)$. Now $C \cap B \in \mathcal{D}$ and $A \cap B\mathcal{D}$ by definition of $\mathcal{F}$. So $(C \setminus A) \cap B \in \mathcal{D}$ since $\mathcal{D}$ is a $d$-system. Thus $C \setminus A$ satisfies the condition of $\mathcal{F}$ so $C \setminus A \in \mathcal{F}$.

(*iii*) $A_1, A_2, \ldots \in \mathcal{F}$ and $A_1 \subset A_2 \subset \cdots$. Then for any $B \in \mathcal{J}$, $(\bigcup_{i=1}^\infty A_i) \cap B = \bigcup_{i=1}^\infty (A_i \cap B)$. Now $A_i \cap B \in \mathcal{D}$ for all $i$ since $A_i \in \mathcal{F}$ and $(A_1 \cap B) \subset (A_2 \cap B) \subset (A_3 \cap B) \subset \cdots$. This means that $(\bigcup_{i=1}^\infty A_i) \cap B = \bigcup_{i=1}^\infty (A_i \cap B) \in \mathcal{D}$, since $\mathcal{D}$ is a $d$-system. Therefore, $(\bigcup_{i=1}^\infty A_i) \in \mathcal{F}$. Therefore, $\mathcal{F}$ is a $d$-system.

Now since $\mathcal{D}$ is the smallest $d$-system containing $\mathcal{J}$, and $\mathcal{J} \subset \mathcal{D}$ it follows from the condition we set on $\mathcal{D}$ that $\mathcal{F} = \mathcal{D}$. Therefore, if $A \in \mathcal{D}$ and $B \in \mathcal{J}$, then $A \cap B \in \mathcal{D}$.

Now define $\mathcal{G} = \{B \in \mathcal{D} \mid A \cap B \in \mathcal{D} \text{ for all } A \in \mathcal{D}\}$. Note that $\mathcal{J} \subset \mathcal{G}$. Also $\mathcal{G} \subset \mathcal{D}$ by definition of $\mathcal{G}$. We can show in a similar fashion that $\mathcal{G}$ is a $d$-system (see Exercise 1). Hence, $\mathcal{G} = \mathcal{D}$. Therefore, if $A \in \mathcal{F} = \mathcal{D}$ and $B \in \mathcal{G} = \mathcal{D}$ then $A \cap B \in \mathcal{D}$ and hence $\mathcal{D}$ is a $\pi$-system. Now $\mathcal{D}$ is then both a $\pi$-system and a $d$-system, and, whence, a $\sigma$-algebra. Therefore, $\sigma(\mathcal{J}) \subset \mathcal{D}$. □

## Exercises

(1) Show that $\mathcal{G} = \{B \in \mathcal{D} \mid A \cap B \in \mathcal{D} \text{ for all } A \in \mathcal{D}\}$ in the proof of Theorem (5.3) is a $d$-system.

CHAPTER 2

# Measures

DEFINITION 2.1 (MEASURE). *A set function $\mu$ on an algebra $\mathcal{F}$ in $\Omega$ is a measure if*

*i.* $\mu(A) \in [0, \infty]$ *for* $A \in \mathcal{F}$,

*ii.* $\mu(\varnothing) = 0$,

*iii. if* $\{A_n\}$ *is a disjoint collection of sets in* $\mathcal{F}$ *with* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ *then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \tag{6}$$

*If $\mu$ is a measure on a $\sigma$-algebra $\mathcal{F}$ then the triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space.*

DEFINITION 2.2. *Let $\mu$ be a measure on $\mathcal{F}$. Then $\mu$ is called*

*i. finite if* $\mu(\Omega) < \infty$.

*ii. $\sigma$-finite if there is a sequence $\{A_n\}$ of elements of $\mathcal{F}$ such that $\mu(A_n) < \infty$ for all $n$ and $\bigcup A_n = \Omega$.*

*iii. a probability measure if* $\mu(\Omega) = 1$.

EXAMPLE 2.1. *A measure $\mu$ on $(\Omega, \mathcal{F})$ is discrete if there are countably many points $\omega_i$ in $\Omega$ and numbers $m_i$ in $[0, \infty]$ such that $\mu(A) = \sum_{\omega_i \in A} m_i$ for $A$ in $\mathcal{F}$. If $\mathcal{F}$ contains each singleton $\{\omega_i\}$, then $\mu$ is $\sigma$-finite if and only if $m_i < \infty$ for all $i$.*

PROOF. We first prove the first statement.

(*i*) The first condition is satisfied since if $m_i \in [0, \infty]$ then $\mu(A) = \sum_{\omega_i \in A} m_i \in [0, \infty]$ for every $A$ in $\mathcal{F}$ by closure property of addition over the real numbers.

(*ii*) Suppose we have a sequence of disjoint sets $A_1, A_2, A_3 \ldots$. Then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{\omega_k \in \bigcup_{i=1}^{\infty}} A_i \tag{7}$$

$$= \sum_{i=1}^{\infty} \left( \sum_{\omega_k \in A_i} m_k \right) \tag{8}$$

$$= \sum_{i=1}^{\infty} \mu(A_i) \tag{9}$$

This is what is happening in Eq. (8). The disjoint sets are the following:

$$\begin{aligned} A_1 &= \{\omega_{11}, \omega_{12}, \omega_{13}, \ldots\} \\ A_2 &= \{\omega_{21}, \omega_{22}, \omega_{23}, \ldots\} \\ &\vdots \end{aligned}$$

Now,

$$\sum_{\omega_k \in \bigcup_{i=1}^{\infty} A_i} m_k = (m_{11} + m_{12} + \cdots) + (m_{21} + m_{22} + m_{23} + \cdots) + \cdots$$

Note that the $m_{ij}'s$ are not necessarily different.

Note also that this part is what is used in elementary probability (discrete).

We now prove the second statement.

Suppose $\mu$ is $\sigma$-finite. Then there exists $A_1, A_2, A_3, \ldots$ such that $\Omega = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$ for all $i$. Fix $k$ and consider the singleton $\omega_k$. Then for some $i$, $\omega_k \in A_i$ and thus $m_k \leq \sum_{\omega_j \in A_j} m_j = \mu(A_i) < \infty$ since $\mu$ being $\sigma$-finite means that $\mu(A_i)$ is finite for all $i$. And so $m_k$ is finite. But $k$ is arbitrary so $m_i$ is finite for every $i$.

Conversely, let $m_i < \infty$ for all $i$. Now,

$$\Omega = \bigcup_{i=1}^{\infty} \{\omega_i\} \cup \left\{ \Omega \setminus \bigcup_{i=1}^{\infty} \{\omega_i\}. \right\}$$

Note that

$$\mu(\{\omega_i\}) = \sum_{\omega_k \in \{\omega_i\}} m_k = m_i < \infty.$$

Now

$$\mu \left\{ \Omega \setminus \bigcup_{i=1}^{\infty} \{\omega_i\} \right\} = \sum_{\omega_k \in \Omega \setminus \bigcup_{i=1}^{\infty} \{\omega_i\}} m_k = 0.$$

Therefore, $\mu$ is $\sigma$-finite. □

EXAMPLE 2.2. *Let $\mathcal{F}$ be the $\sigma$-algebra of all subsets of $\Omega$, and let $\mu(A)$ be the number of elements in $A$, where $\mu(A) = \infty$ if $A$ is not finite. This $\mu$ is counting measure; it is finite if and only if $\Omega$ is finite and is $\sigma$-finite if and only if $\Omega$ is countable.*

PROOF. We first prove that $\mu$ is $\sigma$-finite.

(*i*) $\mu(\varnothing) = 0$ insce $\varnothing$ is finite and it has no elements.
(*ii*) Let $A_1, A_2, \ldots$ be disjoint. Then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \begin{cases} \sum_{i=1}^{\infty} \mu(A_i) & \text{if } \bigcup_{i=1}^{\infty} A_i \text{ is finite.} \\ \infty \; (= \sum_{i=1}^{\infty} \mu(A_i)) & \text{if } \bigcup_{i=1}^{\infty} \text{ is infinite and alll } A_i\text{'s are finite(? Medyo duda ako rito. Hehe)} \\ \infty \; (= \sum_{i=1}^{\infty} \mu(A_i)) & \text{if } \bigcup_{i=1}^{\infty} \text{ is infinite and at least one } A_i \text{ is infinite.} \end{cases}$$

We shall now prove that $\Omega$ is $\sigma$-finite if and only if $\Omega$ is countable.

Note that if $\Omega$ is countable then it is $\sigma$-finite by definition of a $\sigma$-finite.

Suppose that $\Omega$ is $\sigma$-finite. Then there exists a sequence of disjoint sets $A_1, A_2, \ldots$ such that $\bigcup A_i = \Omega$ and $\mu(A_i) < \infty$. Now, this means that $\bigcup A_i = \Omega$ is a countable union of finite sets. Therefore, $\Omega$ is finite and hence countable.

□

EXAMPLE 2.3. *Let $\Omega = (0,1]$. For $A \subset \Omega$, say that $A \in \mathcal{F}$ if it may be written as a finite union*

$$A = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$$

*where $0 \leq a_1 \leq b_1 \leq \cdots \leq a_n \leq b_n \leq 1$. Then $\mathcal{F}$ is an algebra on $(0,1]$. For $A \subset \mathcal{F}$ define*

$$\mu(A) = \sum_{k=1}^{n} (b_k - a_k).$$

*Then $\mu$ is a measure on $\mathcal{F}$.*

Example 3 is much difficult to prove. We shall postpone its proof until we have tackled Caratheodory theory. The application of this example is the uniform distribution and is simply the application of the Lebesque measure. In other words, Lebesque measure is simply the uniform distribution[1].

THEOREM 2.1. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then*

*i.* $\mu(A) \leq \mu(B)$ *if $A \subset B$*
*ii.* $\mu(\bigcup_{i \leq n} A_i) \leq \sum_{i \leq n} \mu(A_i)$ *($A_1, A_2, \ldots, A_n \in \mathcal{F}$). Furthermore, if $\mu(\Omega) < \infty$, then*

[1] I am just making this up at this point. It is not so clear if this is what Prof. Dorado really said. Hehe.

*iii. for* $A_1, A_2, \ldots, A_n \in F$,

$$
\begin{aligned}
\mu\left(\bigcup_{i\le n} A_i\right) &= \sum_{i\le n}\mu(A_i) - \sum\sum_{i<j\le n}\mu(A_i\cap A_j) \\
&+ \sum\sum\sum_{i<j<k\le n}\mu(A_i\cap A_j\cap A_k) - \\
&\cdots + (-1)^{n-1}\mu(A_1\cap A_2\cap\cdots\cap A_n)
\end{aligned} \tag{10}
$$

REMARKS 2.1. *(1) Part (ii) is called countable subadditivity.*

*(2) Part (iii) is called the inclusion-exclusion principle.*

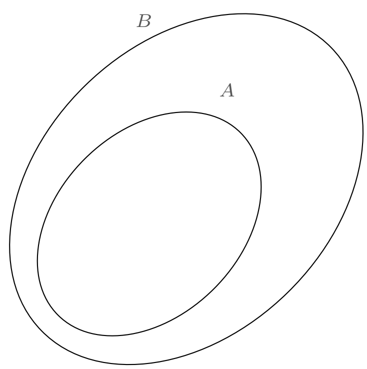
*(3) If* $A_1, A_2, \ldots, A_n$ *are disjoint then* $\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$.

*(4) In the definition of the measure, take* $B_i = A_i$, $i = 1, 2, \ldots, n$ *and* $B_i = \varnothing$ *for* $i > n$. *Then*

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

*since* $\mu(B_i) = 0$ *for* $i > n$.

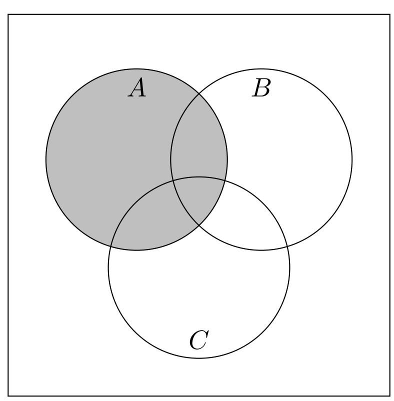
To get an intuition about the proof, consider the figure below.



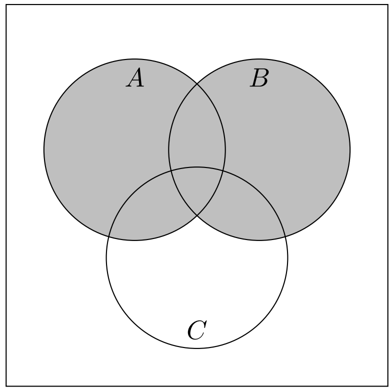

We now prove Theorem (2.1). Now the motivation for building the sequence of sets $B_n$'s is that we are building disjoint sets from the given sets, which might not be disjoint.

Note that $\mu(A \cup B) \le \mu(A) + \mu(B)$ should make sense because when we are adding $\mu(A)$ and $\mu(B)$, we are actually adding the intersection $\mu(A \cap B)$ twice in our count.

We shall develop some more intuition about the procedure of building the disjoint sets $B_n$'s from the $A_n$'s. Consider the Venn Diagram below. We can shade the whole of set $A$ first and call this set $B_1$.




Part of $B$ and $C$ are taken and we don't want to add those parts again in the count later on. So in considering set $B$, we take only its part which does not intersect set $A$. And that is $B \setminus A$.




Finally, we take that part of $C$ that does not intersect $A$ and $B$ and that is $C \setminus (A \cup B)$.

PROOF. (i) $B = A \cup B \setminus A$ and $A$ and $B \setminus A$ are disjoint. Hence

$$\begin{aligned}\mu(B) &= \mu(A \cup (B \setminus A)) \\ &= \mu(A) + \mu(B \setminus A) \\ &\geq \mu(A)\end{aligned}$$

We shall now build a sequence of sets that are decreasing and disjoint.
Let

$$\begin{aligned}B_1 &= A_1 \\ B_2 &= A_2 \setminus A_1 \\ B_3 &= A_3 \setminus (A_1 \cup A_2) \\ &\vdots \\ B_k &= A_k \setminus (A_1 \cup A_2 \cup \cdots \cup A_{k-1})\end{aligned}$$

The $B_n$'s are disjoint and $B_k \subset A_k$. This implies that from part (i), $\mu(B_n) \leq \mu(A_n)$. Also by definition of the $B_n$'s, we have $\bigcup_{k=1}^n B_k = \bigcup_{k=1}^n A_n$. Hence,

$$\mu\left(\bigcup_{k=1}^n A_k\right) = \mu\left(\bigcup_{k=1}^n B_k\right) = \sum_{k=1}^n \mu(B_i) \leq \sum_{k=1}^n \mu(A_i).$$

The proof is by mathematical induction.

REMARKS 2.2. (1) *We have shown earlier that* $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2)$.

(2) *Note also that* $\mu(A_1 \cup A_2 \cup A_3) = \mu(A_1) + \mu(A_2) + \mu(A_3) - \mu(A_1 \cap A_2) - \mu(A_2 \cap A_3) - \mu(A_1 \cap A_3) + \mu(A_1 \cap A_2 \cap A_3)$. *This should be intuitively explained.* $A_1 \cap A_2$ *is both in* $A_1$ *and* $A_2$ *so it should be deducted since it was counted twice in* $\mu(A_1) + \mu(A_2) + \mu(A_3)$. *The same is true for* $A_2 \cap A_3$ *and* $A_1 \cap A_3$. *However,* $A_1 \cap A_2 \cap A_3$ *have been added deducted thrice from* $\mu(A_1) + \mu(A_2) + \mu(A_3)$, *and deducted thrice from* $-\mu(A_1 \cap A_2) - \mu(A_2 \cap A_3) - \mu(A_1 \cap A_3)$ *so it should be added once to balance the equation.*

Since we have shown our inductive step (we showed the statement is true for $n = 2$), we only need to show that the statement holds for $n = k + 1$ if it holds for $n = k$.

Assume the result is true for $n = k$, that is, it is true that

$$\mu\left(\bigcup_{i \leq k} A_i\right) = \sum_{i \leq n} \mu(A_i) - \sum \sum_{i < j \leq k} \mu(A_i \cap A_j) \cdots + (-1)^{k-1} \mu(A_1 \cap A_2 \cap \cdots \cap A_k) \tag{11}$$

Now,

$$
\begin{aligned}
\mu\left(\bigcup_{i=1}^{k+1} A_i\right) =& \mu\left(\bigcup_{i=1}^{k} A_i \cup A_{k+1}\right) \\
=& \mu\left(\bigcup_{i=1}^{k} A_i\right) + \mu(A_{k+1}) - \mu\left(\left(\bigcup_{i=1}^{k} A_i\right) \cap A_{k+1}\right) \\
=& \mu\left(\bigcup_{i=1}^{k} A_i\right) + \mu(A_{k+1}) - \mu\left(\bigcup_{i=1}^{k} (A_i \cap A_{k+1})\right) \\
=& \sum_{i\leq k} \mu(A_i) - \sum_{i\leq j\leq k} \mu(A_i \cap A_j) + \cdots + (-1)^{k-1}\mu(A_1 \cap A_2 \cap \cdots \cap A_k) + \mu(A_{k+1}) \\
& - \sum_{i=1k}^{max} \mu(A_i \cap A_{k+1}) - \sum \sum_{i<j<k} \mu(A_i \cap A_j \cap A_{k+1}) - \cdots \\
& - (-1)^{k-1}\mu(A_1 \cap A_2 \cap \cdots \cap A_{k+1}) \\
& = \sum_{i=1}^{k+1} \mu(A_i) - \sum \sum_{i<j\leq k+1} \mu(A_i \cap A_j) + \cdots + (-1)^{k}\mu(A_1 \cap A_2 \cap \cdots \cap A_k \cap A_{k+1})
\end{aligned}
$$

which is the form of Eq. (10) when $n = k + 1$. Therefore, the result holds by mathematical induction. $\square$

CHAPTER 3

# Continuity and uniqueness

THEOREM 3.1. *Let $\mu$ be a measure on a $\sigma$-algebra $\mathcal{F}$.*

(1) *If $\{A_n\}$ is an increasing sequence in $\mathcal{F}$ with $A = \bigcup A_n$ then $\lim\limits_{n\to\infty} \mu(A_n) = \mu(A)$.*
(2) *If $\{A_n\}$ is a decreasing sequence in $\mathcal{F}$ with $A = \bigcap A_n$ and $\mu(A_k) < \infty$ for some $k$ then $\lim\limits_{n\to\infty} \mu(A_n) = \mu(A)$.*
(3) *If $\{A_n\}$ is a sequence in $\mathcal{F}$ then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

REMARKS 3.1. *Recall from function theory that $f(x)$ is continuous at $x_0$ if $\lim\limits_{x\to x_0} f(x) = f(x_0) = f(\lim\limits_{x\to x_0} x)$. Note that the* $\lim$ *notation go in and out of the function notation. Think of $\bigcup$ as a function so that*

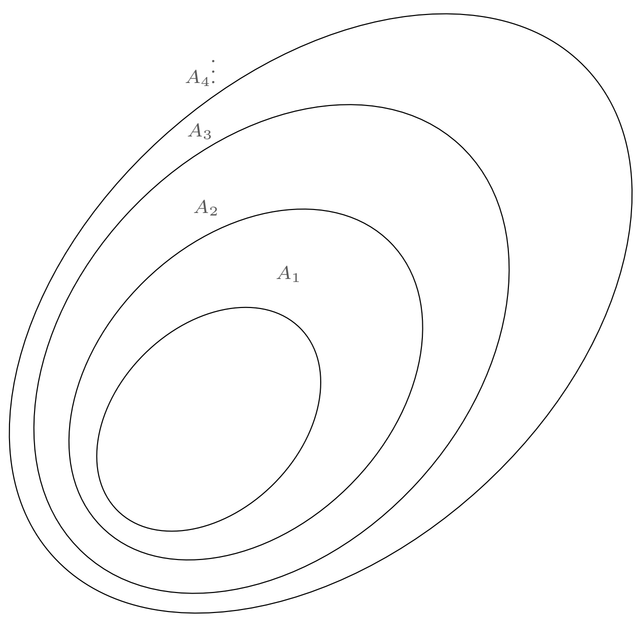
$$\bigcup_{n=1}^{\infty} A_n = \lim_n A_n$$

*and*

$$Pr(\lim_n A_n) = \lim_n Pr(A_n).$$

We now prove Theorem 3.1.

PROOF. (*i*) Use the figure below to visualize the construction of sets.



Let $B_1 = A_1$, $B_n = A_n \setminus A_{n-1}$, $n > 1$. Then the $B_n$'s are disjoint with $\bigcup B_n = \bigcup A_n$. Henc

$$\mu(A) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A_i \setminus A_{i-1}) \tag{12}$$

$$= \lim_{n\to\infty} \sum_{i=1}^{\infty} [\mu(A_i) - \mu(A_{i-1})] \quad \text{since } A_i \subset A_i \tag{13}$$

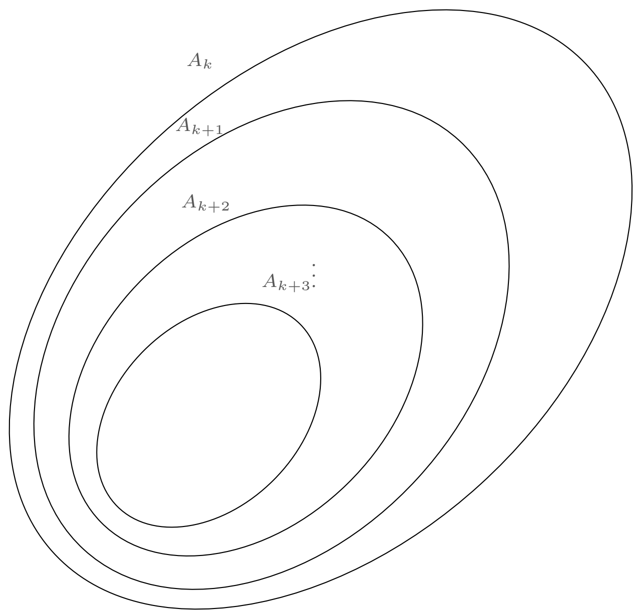
$$= \lim_{n\to\infty} \mu(A_n) \tag{14}$$

To see how (14) was arrived at from (13), the summation in (13) expands to

$$\mu(A_1) + [\mu(A_2) - \mu(A_1)] + [\mu(A_3) - \mu(A_2)] + \cdots + [\mu(A_{n-1}) - \mu(A_{n-2})] + [\mu(A_n) - \mu(A_{n-1})] = \mu(A_n)$$

This technique is called *telescoping* and is used a lot in probability theory.

(*ii*) Consider the figure below.




Let $B_n = A_k \setminus A_n$, $n > k$. Then $\{B_n\}_{n>k}$ is an increasing sequence and

$$\mu(A_k) - \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} [\mu(A_k) - \mu(A_n)] = \lim_{n\to\infty} \mu(A_k \setminus A_n) = \lim_{n\to\infty} \mu(B_n) \tag{15}$$

$$= \mu\left(\bigcup_{n=k+1}^{\infty} B_n\right) = \mu\left(\bigcup_{n=k+1}^{\infty} (A_k \setminus A_n)\right) = \mu\left(\bigcup_{i=k+1}^{\infty} (A_k \cup A_n^c)\right) \tag{16}$$

$$= \mu\left(A_k \cap \bigcup_{n=k+1}^{\infty} A_n^c\right) = \mu\left(A_k \cap \left(\bigcap_{n=k+1}^{\infty} A_n\right)^c\right) \tag{17}$$

$$= \mu\left(A_k \setminus \left(\bigcap_{n=k+1}^{\infty} A_n\right)\right) = \mu(A_k) - \mu\left(\bigcup_{n=k+1}^{\infty} A_n\right) \tag{18}$$

$$= \mu(A_k) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right) \quad \text{since } \bigcap_{n=1}^{\infty} A_n \subset \bigcap_{n=k+1}^{\infty} A_n \tag{19}$$

(*iii*) Let $B_n = \bigcup_{i=1}^n A_i$ then by continuity

$$\mu(B_n) = \mu\left(\bigcup_{n=1}^\infty A_n\right) = \lim_{n\to\infty} \mu(B_n) \leq \lim_{n\to\infty} \sum_{i=1}^n \mu(A_i) = \sum_{n=1}^\infty \mu(A_n)$$

□

THEOREM 3.2. *Let $\Omega$ be a set. Let $\mathcal{J}$ be a $\pi$-system on $\Omega$. Let $\mathcal{F} = \sigma(\mathcal{J})$. Suppose that $\mu_1$ and $\mu_2$ are measures on $(\Omega, \mathcal{F})$ such that $\mu_1(\Omega) = \mu_2(\Omega) < \infty$ and $\mu_1 = \mu_2$ on $\mathcal{J}$. Then $\mu_1 = \mu_2$ on $\mathcal{F}$.*

REMARKS 3.2. *This guarantees as far as Borel sets are concerned that probability distributions are unique*[1].

Here is the proof of Theorem 3.2.

PROOF. Let $\mathcal{G} = \{A \in \mathcal{F} \mid \mu_1(A) = \mu_2(A)\}$.

(*i*) Let $A, B \in \mathcal{G}$ with $A \subset B$ then

$$\begin{aligned} m_1(B \setminus A) &= \mu_1(B) - \mu_1(A) && (20)\\ &= \mu_2(B) - \mu_2(A) && (21)\\ &= \mu(B \setminus A) && (22) \end{aligned}$$

Hence, $B \setminus A \in \mathcal{G}$.

(*ii*) Let $\{A_n\}$ be an increasing sequence in $\mathcal{G}$. Then

$$\begin{aligned} \mu\left(\bigcup_{n=1}^\infty A_n\right) &= \lim_{n\to\infty} (A_n)\\ &= \lim_{n\to\infty} \mu_2(A_n)\\ &= \mu_2\left(\bigcup_{n=1}^\infty A_n\right) \end{aligned}$$

Hence $\bigcup_{n=1}^\infty A_n \in \mathcal{G}$. Then $\mathcal{G}$ is a $d$-system containing the $\pi$-system containing $\mathcal{J}$. Hence, $\mathcal{F} = \sigma(\mathcal{J})$.

□

What follows is an immediate result.

COROLLARY 3.1. *If two probability measures agree on a $\pi$-system, then they agree on the $\sigma$-algebra by that $\pi$-system.*

[1] Again, I am partly making this up. I don't know if this is what is exactly what Prof. Dorado said. Hehe.

CHAPTER 4

# Caratheodory's Extension Theorem

Throughout our lessons, one of our aims is to make the concept of the $\sigma$-algebra concept easier, smaller.

THEOREM 4.1 (CARATHEODORY'S EXTENSION THEOREM). *Let $\Omega$ be a set, let $\mathcal{A}$ be an algebra on $\Omega$, and let $\mathcal{F} = \sigma(\mathcal{A})$. If $\mu_0$ is a countably additive map $\mu_0 : \mathcal{A} \to [0, \infty]$, then there exists a measure $\mu$ on $(\Omega, \mathcal{F})$ such that $\mu = \mu_0$ on $\mathcal{A}$. If $\mu_0(S) < \infty$ then this extension is unique.*

DEFINITION 4.1 (OUTER MEASURE). *For any set $A \subseteq \Omega$ define the outer measure $\mu^*(A)$ of $A$ by*

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu_0(A_n) : A_n \in \mathcal{A}, A \subseteq \bigcup_{n=1}^{\infty} A_n \right\}. \tag{23}$$

LEMMA 4.1.1. *For any sets $A$ and $A_n$ in $\Omega$, if $A \subseteq \bigcup_{n=1}^{\infty} A_n$ then*

$$\mu^* \leq \sum_{n=1}^{\infty} \mu^*(A_n). \tag{24}$$

PROOF. If the latter sum is $+\infty$, there is no problem. Otherwise, given $\varepsilon > 0$, let $A_n \bigcup_m A_{nm}$ and $\sum_m \mu_0(A_{nm}) < \mu^*(A_n) + \varepsilon/2^n$ with $A_{nm} \in \mathcal{A}$ for all $m, n$. Then $A \subseteq \bigcup_{m,n} A_{nm}$ and

$$\mu^*(A) \leq \sum_n \sum_m \mu_0(A_{nm}) \leq \sum_n \mu^*(A_n) + \varepsilon. \tag{25}$$

Letting $\varepsilon \downarrow 0$ proves the result. □

LEMMA 4.1.2. *For any $A \in \mathcal{A}$, $\mu^*(A) = \mu_0(A)$.*

PROOF. Clearly, $\mu^*(A) \leq \mu_0(A)$. Let $A \subseteq \bigcup_n A_n$ and $B_n = A_n \setminus \bigcup_{j<n} A_j$ with $A_n \in \mathcal{A}$.

$$\mu_0(A) \leq \mu_0(\cup B_n) = \sum \mu(B_n) \leq \sum \mu(A_n). \tag{26}$$

Since the $A_n$'s are arbitrary then $\mu_0(A) \leq \mu^*(A)$. □

DEFINITION 4.2 ($\mu^*$-MEASURABLE). *A set $E \subset \Omega$ is called $\mu^*$-measurable if for every set $A \subset \Omega$,*

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c) \tag{27}$$

*We will denote the collection of $\mu^*$-measurable sets by $\mathcal{M}(\mu^*)$.*

LEMMA 4.1.3. *All sets in $\mathcal{A}$ are $\mu^*$-measurable, i.e., $\mathcal{A} \subset \mathcal{M}(\mu^*)$.*

PROOF. Let $A \in \mathcal{A}$ and $B \subset \Omega$. Given $\varepsilon > 0$, take $B \subset \cup B_n$ with $B_n \in \mathcal{A}$ and $\sum \mu_0(B_n) \leq \mu^*(B) + \varepsilon$. Then $B \cap A \subset \bigcup(B_n \cap A)$, $B \cap A^c \subset \bigcup(B_n \cap A^c)$ with $B_n \cap A \in \mathcal{A}$ and $B_n \cap A^c \in \mathcal{A}$. So that

$$\begin{aligned} mu^*(B \cap A) + \mu^*(B \cap A^c) &\leq \sum [\mu^*(B_n \cap A) + \mu^*(B_n \cap A^c)] \\ &\leq \sum \mu_0(B_n) \leq \mu^*(B) + \varepsilon. \end{aligned}$$

Letting $\varepsilon \downarrow 0$ gives the result. □

LEMMA 4.1.4. *$\mathcal{M}(\mu^*)$ is a $\sigma$-algebra and $\mu^*$ is a measure on it.*

PROOF. Clearly $A \in \mathcal{M}(\mu^*)$ then $A^c \in \mathcal{M}(\mu^*)$. If $A, B \in \mathcal{M}(\mu^*)$ then for $C \subset \Omega$,

$$\begin{aligned}\mu^*(C) &= \mu^*(C \cap A) + \mu^*(C \cap A^c)\\ &= \mu^*(C \cap A \cap B) + \mu^*(C \cap A \cap B^c) + \mu^*(C \cap A^c)\\ &= \mu^*(C \cap (A \cap B)) + \mu^*(C \cap (A \cap B)^c).\end{aligned}$$

Thus $\mathcal{M}(\mu^*)$ is an algebra. Let $A_n \in \mathcal{M}(\mu^*)$, $A = \bigcup_n A_n$, $B_n = \bigcup_{j<n} A_j$, and assume, without loss of generality, that the $A_n$'s are disjoint.

Note that for $E \subset \Omega$

$$\begin{aligned}\mu^* &= \mu^*(E \cap B_n^c) + \mu^*(E \cap B_n)\\ &= \mu^*(E \cap B_n^c) + \mu^*(E \cap A_n) + \mu^*\left(\bigcup_{j<n}(E \cap A_i)\right).\end{aligned}$$

From this it follows by induction on $n$ that

$$\mu^*(E) \geq \mu^*(E \cap A^c) + \sum_{j=1}^{n} \mu^*(E \cap A_j).$$

Letting $n \to \infty$

$$\mu^*(E) \geq \mu^*(E \cap A^c) + \mu^*(E \cap A).$$

Thus $A \in \mathcal{M}(\mu^*)$ and

$$\mu^*(E) = \mu^*(E \cap A^c) + \sum_{j=1}^{\infty} \mu^*(E \cap A_j) \tag{28}$$

Letting $E = A$ shows that $\mu^*$ is countably additive on $\mathcal{M}(\mu^*)$. □

Here is the proof of the Caratheodory's Theorem

PROOF. Note that by Lemmas 3 and 4, $\mathcal{M}(\mu^*)$ is a $\sigma$-algebra containing $\mathcal{A}$ and, hence, $\mathcal{F} \subset \mathcal{M}(\mu^*)$. Take $\mu = \mu^*$ on $\mathcal{F}$. Then $\mu$ is a measure on $\mathcal{F}$ and for $A \in \mathcal{A}$,

$$\mu(A) = \mu^*(A) = \mu_0 * (A).$$

Uniqueness follows from Theorem 2 of Lecture 3. □

CHAPTER 5

# Measurable Functions and Mappings

DEFINITION 5.1 ($\mathcal{F}/F'$-MEASURABLE). *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces. For a mapping $T : \Omega \to \Omega'$ and $A' \subset \Omega'$, define*

$$T^{-1}(A') = \{\omega \in \Omega : T(\omega) \in A'\}$$

THEOREM 5.1. *The map $T^{-1}$ preserves set operations:*

(1) $T^{-1}(\bigcup_\alpha A_\alpha) = \bigcup_\alpha T^{-1}(A_\alpha)$
(2) $T^{-1}(\bigcap_\alpha A_\alpha) = \bigcap_\alpha T^{-1}(A_\alpha)$
(3) $T^{-1}(A^c) = (T^{-1}(A))^c$

PROOF. (1) $\omega \in T^{-1}\left(\bigcup_\alpha A_\alpha\right)$ if and only if $T(\omega) \in A_\alpha$ for some $\alpha$. This is equivalent to $\omega \in T^{-1}(A_\alpha)$ for some $\alpha$, i.e., $\omega \in \bigcup_\alpha T^{-1}(A_\alpha)$.
(2) Similar to (a) with union replaced by intersection.
(3) $\omega \in T^{-1}(A^c)$ if and only if $T(\omega) \notin A$. This is equivalent to $\omega \notin T^{--1}(A)$, i.e., $\omega \in (T^{-1}(A))^c$. □

THEOREM 5.2. *Let $T$ be a mapping from the measurable space $(\Omega, \mathcal{F})$ into the measurable space $(\Omega', \mathcal{F}')$. Then $T^{-1}(\mathcal{F}') = \{A' \subseteq \Omega' : T^{-1}(A') \in \mathcal{F}\}$ is a $\sigma$-algebra on $\Omega'$.*

PROOF. $\Omega = T^{-1}(\Omega')$ so that $\Omega \in T^{-1}(\mathcal{F}')$. Since $(T^{-1}(A'))^c = T^{-1}(A'^c)$ then $T^{-1}(\mathcal{F}')$ is closed under complementation. Lastly, $\bigcup_{n=1}^\infty T^{-1}(A'_n) = T^{-1}\left(\bigcup_{n=1}^\infty A'_n\right)$. Hence, $T^{-1}(\mathcal{F}')$ is a $\sigma$-algebra. Now, let $\mathcal{G} = \{A' \subseteq \Omega' : T^{-1}(A') \in \mathcal{F}\}$. $T^{-1}(\Omega') = \Omega \in \mathcal{F}$ so that $\Omega' \in \mathcal{G}$. $T^{-1}(A'^c) = (T^{-1}(A'))^c \in \mathcal{F}$ whenever $A' \in \mathcal{G}$. Hence, $\mathcal{G}$ is closed under complementation. Finally, $T^{-1}\left(\bigcup A'_n\right) = \bigcup T^{-1}(A'_n) \in \mathcal{F}$ whenever $A'_n \in \mathcal{G}$. Therefore, $\mathcal{G}$ is a $\sigma$-algebra. □

THEOREM 5.3. *If $\mathcal{C} \subseteq \mathcal{F}'$, $\sigma(\mathcal{C}) = \mathcal{F}'$ and $T^{-1}(A') \in \mathcal{F}$ for every $A' \in \mathcal{C}$, then $T$ is $\mathcal{F}/\mathcal{F}'$-measurable.*

PROOF. Let $\mathcal{G} = \{A' \in \mathcal{F}' \mid T^{-1}(A') \in \mathcal{F}\}$. Then $\mathcal{G}$ is a $\sigma$-algebra containing $\mathcal{C}$ and, hence, $\mathcal{G} = \mathcal{F}'$. □

THEOREM 5.4. *If $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$ and $(\Omega'', \mathcal{F}'')$ are measurable spaces, and if $T$ is $\mathcal{F}/F'$-measurable and $T'$ is $\mathcal{F}'/\mathcal{F}''$-measurable, then $T' \circ T$ is $\mathcal{F}/\mathcal{F}''$-measurable.*

PROOF. Let $A'' \in \mathcal{F}''$. It follows that $(T' \circ T)^{-1}(A'') = T^{-1}(T'^{-1}(A'')) \in \mathcal{F}$ since $T'^{-1}(A'') \in \mathcal{F}'$. □

DEFINITION 5.2 (MEASURABLE FUNCTION). *Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R}$ is called measurable if $f^{-1}(B) \in \mathcal{F}$ for every Borel set $B$.*

THEOREM 5.5. *The function $f : \Omega \mapsto \mathbb{R}$ is measurable if and only if $\{f \le c\} := \{\omega \in \Omega \mid f(\omega) \le c\} \in \mathcal{F}$ for all $c \in \mathbb{R}$.*

PROOF. The class $\{(-\infty, c] : c \in \mathbb{R}\}$ generates the Borel $\sigma$-algebra. Also, note that $\{f \le c\} = f^{-1}((-\infty, c])$. The result then follows from Theorem 5.3. □

THEOREM 5.6. *Let $\lambda \in \mathbb{R}$ and $f$, $f_1$, $f_2$ be measurable. Then $f_1 + f_2$, $f_1 f_2$ and $\lambda f$ are measurable.*

PROOF. Note that $\{f_1 + f_> c\} = \bigcup_{r \in \mathbb{Q}}(\{f_1 > r\}) \cap \{f_2 > c - r\}$. Hence, $f_1 + f_2$ is measurable. If $\lambda = 0$, then $\{\lambda f \le c\}$ is neither $\varnothing$ (if $c < 0$) or $\Omega$ (if $c \ge 0$). Hence $\lambda f$ is measurable. If $\lambda \ne 0$, note that

$$\{\lambda f > c\} = \{f > c/\lambda\} \quad \text{if} \quad \lambda > 0$$

and

$$\{\lambda f > c\} = \{f < c/\lambda\} \quad \text{if} \quad \lambda < 0$$

This shows that $\lambda f$ is measurable.

To show that $f_1 f_2$ is measurable, first note that

$$\{f^2 > c\} = \{f > \sqrt{c}\} \cup \{f < -\sqrt{c}\}.$$

This shows that $f^2$ is measurable whenever $f$ is measurable. Observe that

$$f_1 f_2 = \frac{1}{2}[(f_1 + f_2)^2 - f_1^2 - f_2^2]$$

and, hence, $f_1 f_2$ is measurable. □

THEOREM 5.7. *Let* $\{f_n : n \in \mathbb{N}\}$ *be a sequence of measurable functions. Then* $\inf f_n$, $\liminf f_n$ *and* $\limsup f_n$ *are measurable into* $([-\infty, +\infty], \mathcal{B}[-\infty, +\infty])$. *Further* $\{\omega \in \Omega : \lim f_n(\omega) \text{ exists}\} \in \mathcal{F}$.

PROOF. Note that $\{\inf f_n > c\} = \bigcap_n \{f_n > c\} \in \mathcal{F}$. Hence $\inf f_n$ is measurable. Let $g_n = \inf_{k \geq n} f_k$. Then

$$\{\liminf f_n \leq c\} = \{\sup g_n \leq c\} = \bigcap \{g_n \leq c\} \in \mathcal{F}.$$

Hence, $\liminf f_n$ is measurable. Now, $\limsup f_n = -\liminf -f_n$ and so $\limsup f_n$ is measurable. Finally,

$$\{\lim f_n \text{ exists}\} = \{\limsup f_n < \infty\} \cap \{\liminf f_n > -\infty\} \cap g^{-1}(\{0\}),$$

where $g = \limsup f_n - \liminf f_n$. This is in $\mathcal{F}$. □

CHAPTER 6

# Probability Spaces, Random Variables and Distribution Fuunctions

DEFINITION 6.1 (RANDOM VARIABLE). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A measurable function $X : \Omega \mapsto \Re$ is called a random variable. The law of the random variable is the probability measure $\mu$ on $(\Re, \mathcal{B})$ defined by*

$$\mu(A) = \mathbb{P}(X \in A), A \in \mathcal{B}. \tag{29}$$

*The distribution function of $X$ is defined by*

$$F(x) = \mu((-\infty, x]) = \mathbb{P}(X \leq x) \tag{30}$$

*for real $x$.*

REMARKS 6.1. *(1) $\mathbb{P}(\Omega) = 1$ and $X^{-1}(B) \in \mathcal{F}$ for any $B \in \mathcal{B}$ implies that $\mathbb{P}(X^{-1}(B))$ makes sense. This is the law of the random variable $X$*
*(2) $B = (-\infty, x]$ is the Borel set. $F(x)$ is a probability distribution*

THEOREM 6.1. *Let $F$ be a distribution function of some random variable $X$. Then*
*(1) $F(x) \leq F(y)$ whenever $x \leq y$,*
*(2) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$,*
*(3) $F$ is right-continuous*

Recall the standard normal distribution

$$\phi(x) = \frac{1}{2\pi} e^{-x^2/2}, \qquad -\infty < x < \infty$$

The distribution function of the standard normal distribution is right continuous. Another example is the Poisson distribution. Refer to Figure 1.

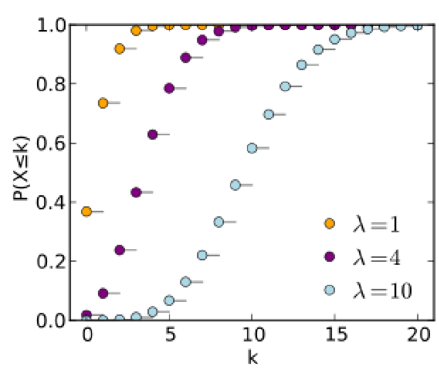



FIGURE 1. Graph of the Poisson cdf with $\lambda = 1$, $\lambda = 2$, and $\lambda = 4$. Source: `https://en.wikipedia.org/wiki/Poisson_distribution`

We now prove Theorem 6.1.

PROOF. (*i*) If $x \leq y$ then $(-\infty, x] \subseteq (-\infty, y]$ and it follows that $F(x) = \mu((-\infty, x]) \leq \mu((-\infty, y]) = F(y)$.

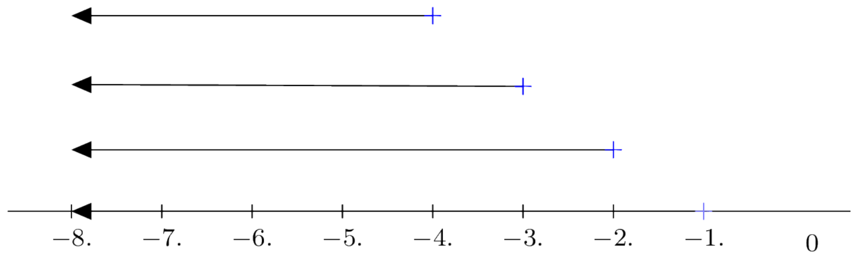



FIGURE 2. Intervals $(-\infty, -n]$

(*ii*) Get the intersections of the intervals $r \in \bigcup_{n=1}^{\infty} (-\infty, -n]$. (See Fig. 2) That is, $r \le -n$ for all $n$.

Note that $\bigcap_{n=1}^{\infty} (-\infty, -n] = \varnothing$ so that

$$\begin{aligned} \lim_{x\to\infty} F(x) &= \lim_{n\to\infty} F(-n) = \lim_{n\to\infty} \mu((-\infty, -n]) \\ &= \mu\left(\bigcap_{n=1}^{\infty} (-\infty, -n]\right) \qquad (decreasing) \\ &= \mu(\varnothing) = 0 \end{aligned}$$

Similarly,

$$\bigcup_{n=1}^{\infty} (-\infty, n] = \Re$$

and

$$\lim_{x\to+\infty} F(x) = \mu\left(\bigcup_{n=1}^{\infty} (-\infty, n]\right) = 1$$

(*iii*) For right continuity[1], note that

$$\lim_{n\to\infty} F\left(x + \frac{1}{n}\right) = \mu\left(\bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right]\right) = F(x)$$

To see this, not that

$$\begin{aligned} \lim_{n\to\infty} F\left(x + \frac{1}{n}\right) &= \lim_{n\to\infty} \mu\left(\left(-\infty, x + \frac{1}{n}\right]\right) \\ &= \mu\left(\bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right]\right) \end{aligned}$$

If $y \in \left(-\infty, x + \frac{1}{n}\right]$ for all $n$, then $y \le x$. To see this, if $y > x$ then if $x + 1 > y > x$ then there is an $m$ such tht $y > x + \frac{1}{m} > x$ contradicting the original assumption. So if $y \le x + \frac{1}{n}$ for all $n$, then $y \le x$. Therefore,

$$\begin{aligned} \lim_{n\to\infty} F\left(x + \frac{1}{n}\right) &= \mu\left(\bigcap_{n=1}^{\infty} \left(-\infty, x + \frac{1}{n}\right]\right) \\ &= \mu((-\infty, x]) \\ &= F(x) \end{aligned}$$

[1]Remember from calculus that right continuity means $\lim_{y\to x^+} F(y) = F(x)$

□

Is the converse of Theorem 6.1 true? To prove this, we need to find $x$ and show that $F$ is really its distribution function. This is the content of the next theorem.

THEOREM 6.2. *If $F$ is a nondecreasing, right-continuous function with $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$, then there exists on some probability space a random variable $X$ for which $F(x) = \mathbb{P}(X \le x)$.*

PROOF. Define the random variable $X$ on $([0,1], \mathcal{B}[0,1], \lambda)$ by $X(\omega) = \inf\{z : F(z) \ge \omega\}$. Now, if $z > X(\omega)$ then $F(z) \ge \omega$, so by right-continuity, $F(X(\omega)) \ge \omega$. If, in addition $X(\omega) \le c$ then $\omega \le F(X(\omega)) \le F(c)$. Thus $\omega \le F(c)$ if and only if $X(\omega) \le c$, so that $\mathbb{P}(X \le c) = \lambda([0, F(c)]) = F(c)$. □

PROOF. Define the random variable $X$ on $([0,1], B[0,1], \lambda)$, (where $\lambda$ is Lebesque measure) by $X(\omega) = \inf\{z : F(z) \ge \omega\}$. Now, if $z > X(\omega)$ then $F(z) > \omega$ so by right-continuity, $F(X(\omega)) \ge \omega$. If, in addition, $X(\omega) \le c$, then $F(X(\omega)) \le F(c)$. Thus $\omega \le F(c)$ if and only if $X(\omega) \le c$ so that $\mathbb{P}(X \le c) = \lambda([0, F(c)]) = F(c)$.

Now, whenever $z > X(\omega)$, $F(z) \ge \omega$. Thus $\lim_{z \to X(\omega)^+} F(z) \ge \omega$. But since $F(z)$ is right-continuous, we have $\lim_{z \to X(\omega^+)} F(z) \ge \omega$

So $\mathbb{P}(X(\omega) \le c) = \mathbb{P}(X \le c)$. Note that $\mathbb{P}(X \le c) = \lambda(\{\omega : \omega \le F(c)\}) = \lambda([0, F(c)]) = F(c)$. □

The foregoing theory shows that we don't need to know about the probability space as long as we know the distribution. that is why we an talk about a normal distribution without knowing the space of concern.

CHAPTER 7

# Construction of the Integral

Definition 7.1 (Simple function). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A nonnegative measurable function $f$ is called simple if $f$ may be written as a finite sum $f = \sum_{k=1}^{m} a_k I_{A_k}$ where $a_k \in [0, \infty]$ and $A_k \in \mathcal{F}$. We then define the integral of $f$ with respect to $\mu$ by $\int f d\mu = \sum_{k=1}^{m} a_k \mu(A_k)$.*

Note that

$$I_{A_k} = \begin{cases} 1 & \text{if} \quad \omega \in A_k \\ 0 & \text{if} \quad \omega \notin A_k \end{cases}$$

This means that in $\sum_{k=1}^{m} a_k I_{A_k}$, you are adding a number $a_k$ assigned to $A_k$ as long as $\omega \in A_k$.

Example 7.1. *Let $\Omega = [0, 5]$ and*

$$\begin{aligned} f(\omega) &= \begin{cases} 2 & \textit{if} \quad 0 \le \omega < 3 \\ 5 & \textit{if} \quad 5 \le \omega < 4 \\ 3 & \textit{if} \quad 3 \le \omega \le 5 \end{cases} \\ &= 2I_{[0,3)} + 5I_{[3,4)} + 3I_{[4,5]} \end{aligned}$$

*Now*

$$f = \sum_{k=1}^{m} a_k \mu(A_k) \implies \int f d\mu = \sum_{k=1}^{m} a_k \mu(A_k)$$

*For this example, this means that*

$$\int f d\lambda = 2\lambda([0,3)) + 5\lambda([3,4)) + 3\lambda([4,5]) = 2(3) + 5(1) + 3(1) = 14$$

The question is, is $f$ well defined? No. For instance,

$$f = 2I_{[0,4)}(\omega) + 3I_{[3,5]}(\omega)$$

however,

$$\int f d\lambda = 2(4) + 3(2) = 14$$

gives the same result. That is, $f$ can be written in other ways but $\int f d\mu$ will always give the same result.
The following theorem states that $\int f d\mu$ is indeed well-defined.

Theorem 7.1. *(1) The integral $\int f d\mu$ os well-defined.*
*(2) If $f$, $g$ are simple functions and $c \ge 0$ then $f + g$ and $cf$ are simple functions and $\int (f + g) d\mu = \int f d\mu + \int g d\mu$, $\int cf d\mu = c \int f d\mu$.*
*(3) If $f$ and $g$ are simple functions with $f \le g$ then $\int f d\mu \le \int g d\mu$.*
*(4) If $f$ and $g$ are simple functions then $f \vee g$ and $f \wedge g$ are simple functions.*

Proof. (*i*) consider the simple function $f = \sum_{i=1}^{m} a_i I_{A_i}$. Note that there are no restricitons for the $A_i$. In fact, they might not be disjoint. Now define

$$B_j = \left\{ \omega \in \bigcup_{i=1}^{m} A_i : f(\omega) = b_j \right\}$$

The $B_j$'s are disjoint and the $b_j$'s are distinct values corresponding to the $a_j$'s.

Then

$$
\begin{aligned}
\int f d\mu &= \sum_{i=1}^{m} a_i \mu(A_i) \\
&= \sum_{i=1}^{m} a_i \mu\left(A_i \cap \bigcup_{j=1}^{n} B_j\right) \\
&= \sum_{i=1}^{m} a_i \mu\left(\bigcup_{j=1}^{n} (A_i \cap B_j)\right) && (A_i \cap B_j)\text{'s are disjoint} \\
&= \sum_{i=1}^{m} a_i \left(\sum_{j=1}^{n} \mu(A_i \cap B_j)\right) \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_i \mu(A_i \cap B_j)\right)
\end{aligned}
$$

Now,

$$
\begin{aligned}
\sum_j b_j \mu(B_j) &= \sum_j \mu(B_j) \sum_{A_k \supseteq B_j} a_k && (\text{Add all } a_k \text{ where } A_k \supseteq B_j) \tag{31} \\
&= \sum_k a_k \sum_{B_j \subseteq A_k} \mu(B_j) && \text{Note: the } B_i\text{'s are disjoint} \tag{32} \\
&= \sum a_k \mu(A_k) \tag{33}
\end{aligned}
$$

This proves that $f$ is well-defined.

To illustrate what we mean by the $b_j$'s and $B_j$'s, we give an example.

EXAMPLE 7.2.

$$
\begin{aligned}
f &= \overset{a_1}{2} I_{[0,3)} + \overset{a_2}{2} I_{[2,4]} + \overset{a_3}{1} I_{[3,5]} \\
&= \begin{cases} \overset{b_1}{2} & \omega \in [0,2) \\ \overset{b_2}{5} & \omega \in [2,3) \\ \overset{b_3}{4} & \omega \in [3,4] \\ \overset{b_4}{1} & \omega \in (4,5] \end{cases}
\end{aligned}
$$

*Insert figure here*

To illustrate Eqs. 31 to 33, we have the following example.

EXAMPLE 7.3.

$$
\begin{aligned}
&\mu(B_1)[a_1 + a_2] + \mu(B_2)[a_3] + \mu(B_3)[a_3 + a_4 + a_5] + \mu(B_4)[a_4 + a_5] \\
&\quad = a_1\mu(B_1) + a_2\mu(B_1) + a_3[\mu(B_2) + \mu(B_3)] + a_4[\mu(B_3) + \mu(B_4)] + a_5[\mu(B_3) + \mu(B_4)]
\end{aligned}
$$

(*ii*) Note that $f + g = \sum (a_i + b_j)_{I_{A_i \cap B_j}}$ and

□

PROOF. (1) For distinct nonnegative values $b_j$ of $f$ define $B_j = \{\omega \in \Omega : f(\omega) = b_j\}$. Then

$$
\begin{aligned}
\sum b_j \mu(B_j) &= \sum \mu(B_j) \sum \{a_k : A_k \supseteq B_j\} \\
&= \sum a_k \sum \{\mu(B_j) : B_j \subseteq A_k\} \\
&= \sum a_k \mu(A_k).
\end{aligned}
$$

This proves that $f$ is well-defined.

(2) Note that $f+g=\sum(a_i+b_j)I_{A_i\cap B_j}$ and $cf=\sum ca_iI_{A_i}$. So that $f+g$ and $cf$ are both simple. Now

$$\begin{aligned}\int(f+g)d\mu &= \sum_i a_i \sum \mu(A_i\cap B_j)+\sum_j b_j\sum_i \mu(A_i\cap B_j)\\ &= \sum_i a_i\mu(A_i)+\sum_j b_j\mu(B_j)\\ &= \int fd\mu+\int gd\mu\end{aligned}$$

and

$$\int cfd\mu = c\sum a_iI_{A_i}=c\int fd\mu.$$

(3) Note that for fixed $i,j$ we have $a_i\mu(A_i\cap B_j)\leq b_j\mu(A_i\cap B_j)$. It follows from this that

$$\begin{aligned}\int fd\mu=\sum a_i\mu(A_i) &= \sum_i\sum_j a_i\mu(A_i\cap B_j)\\ &= \sum_i\sum_j b_j\mu(A_i\cap B_j)\leq\sum_j b_j\mu(B_j)=\int gd\mu.\end{aligned}$$

□

DEFINITION 7.2. *Let $(\Omega,\mathcal{F},\mu)$ be a measure space and $f$ be a nonnegative measurable function, we define $\int fd\mu=\sum\{\int gd\mu : 0\leq g\leq f, g$ simple$\}$.*

THEOREM 7.2. *For any measurable $f\geq 0$, there exists an increasing sequence of simple functions $f_n$ with $f=\lim\limits_{nto\infty} f_n$. Moreover for any such sequence $\{f_n\}$, $\int fd\mu=\lim\limits_{n\to\infty}\int f_nd\mu$.*

PROOF. Let $A_{nj}=f^{-1}$ □

CHAPTER 8

# Limit Theorems on Integrals

THEOREM 8.1 (MONOTONE CONVERGENCE THEOREM). *If $\{f_n\}$ is an increasing sequence of nonnegative measurable functions such that $\lim_{n\to\infty} f_n = f$, then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.*

PROOF. For each $n$, take increasing simple functions $\{f_{mn}\}$ with $\lim_{m\to\infty} f_{mn} = f_n$. Define $g_n = \max(f_{1n}, f_{2n}, \ldots, f_{nn})$. Then $\{g_n\}$ is an increasing sequence of simple functions with $\lim_{n\to\infty} g_n = f$. It follows that $\lim_{n\infty} \inf g_n d\mu = \inf f d\mu$. Since $g_n \leq f_n \leq f$, we get $\int f d\mu = \lim_{n\to\infty} \int g_n \leq \lim_{n\to\infty} \int f_n d\mu \leq \int f d\mu$. This implies that $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$. $\square$

LEMMA 8.1.1. *Let $f$ be a nonnegative measurable function and $N \in \mathcal{F}$ with $\mu(N) = 0$. Then $\int f I_N d\mu = 0$.*

PROOF. Let $\{f_n\}$ be a sequence of simple functions with $\lim_{n\to\infty} f_n = f$. Then $\{f_n I_N\}$ is an increasing sequence of simple functions with $\lim_{n\to\infty} f_N I_N = f I_N$. If $f_n = \sum a_{ni} I_{A_{ni}}$ then $f_n IN = \sum a_{ni} I_{A_{ni}\cap N}$. This implies that $\int f_n IN d\mu = \sum a_{ni}\mu(A_{ni} \cap N) = 0$ since $\mu(A_{ni} \cap N) \leq \mu(N) = 0$. Therefore $\int f I_N d\mu = \lim_{n\to\infty} f_n IN d\mu = 0$. $\square$

THEOREM 8.2. (1) *If $f$ and $g$ are nonnegative measurable functions and $f = g$ a.e., then $\int f d\mu = \int g d\mu$.*
(2) *If $\{f_n\}$ and $f$ are nonnegative measurable functions with $\{f_n\}$ increasing and $\lim_{n\to\infty} f_n = f$ a.e.. Then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.*

PROOF. (1) $f_n = \sum_{j=0}^{n2^n-1} j2^{-n} I_{f^{-1}((j2^{-n},(j+12^{-n})])} + nI_{f^{-1}((n,\infty])}$ and $g_n = \sum_{j=0}^{n2^n-1} j2^{-n} I_{g^{-1}((j2^{-n},(j+12^{-n})])} + nI_{g^{-1}((n,\infty])}$. Then $f_n = g_n$ a.e. and so $\int f_n d\mu = \int g_n d\mu$. The result now follows by monotone convergence.
(2) Let $N = \{\omega \in \Omega : \lim_{n\to\infty} f_n(\omega) \neq f(\omega)\}$. Then $\lim_{n\to\infty} f_n I_{N^c} = f I_{N^c}$. From this it follows that

$$\lim_{n\to\infty} \int f_n d\mu = \lim_{n\to\infty} \int f_n I_{N^c} d\mu = \int f I_{N^c} d\mu = \int f d\mu$$

$\square$

LEMMA 8.2.1 (FATOU LEMMA). *For a sequence $\{f_n\}$ nonnegative measurable functions*

$$\int \liminf f_n d\mu \leq \liminf \int f_n d\mu$$

PROOF. Let $g_n = \inf_{k\geq n} f_k$. Then $\{g_n\}$ is an increasing sequence with $\lim_{n\to\infty} g_n = \liminf f_n$. For $k \geq n$, we have $f_k \geq g_n$, so that $\int f_k d\mu \geq \int g_n d\mu$. This implies that $\inf_{k\geq n} \int f_k d\mu \geq \int g_n d\mu$. Thus $\int \liminf f_n \leq \lim_n \inf_{k\geq n} \int f_k d\mu = \liminf \int f_n d\mu$. $\square$

LEMMA 8.2.2 (REVERSE FATOU). *If $\{f_n\}$ is a sequence of nonnegative measurable functions such that for some nonnegative measurable function $g$, we have $f_n < g$ for all $n$ and $\int g d\mu$, then*

$$\limsup f_n d\mu \geq \limsup \int f_n d\mu.$$

PROOF. For each $n$, let $h_n = g - f_n$. Then $\{h_n\}$ is a sequence of nonnegative measurable functions and by Fatou lemma we have

$$\begin{aligned}\int g d\mu - \lim \sum \int f_n d\mu &= \liminf \int h_n d\mu \\ &\geq \int \liminf h_n d\mu \\ &= \int g d\mu - \int \limsup f_n d\mu.\end{aligned}$$

Subtracting $\int g d\mu$ on both sides of the inequality gives the result. $\square$

THEOREM 8.3 (DOMINATED CONVERGENCE THEOREM). *Suppose that $\{f_n\}$, $f$ are measurable functions, that $\lim_{n\to\infty} f_n = f$ on $\Omega$ and that $|f_n| \leq g$ for some integrable function $g$. Then*

$$\lim_{n\to\infty} \inf |f_n - f| d\mu = 0$$

*and, hence, $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.*

PROOF. Note that $|f_n - f| \leq 2g$ and so by reverse Fatou lemma

$$\limsup \inf |f_n - f|\, d\mu \leq \inf \limsup |f_n - f| = 0.$$

Consequently,

$$\lim_{n\to\infty} \left| \int f_n d\mu - \int f d\mu \right| = \lim_{n\to\infty} \left| \int (f_n - f) d\mu \right| \leq \lim_{n\to\infty} \int |f_n - f|\, d\mu = 0.$$

$\square$

COROLLARY 8.1 (BOUNDED CONVERGENCE THEOREM). *Suppose that $\{f_n\}$, $f$ are measurable functions, that $\lim_{n\to\infty} f_n = f$ on $\Omega$ and that $|f_n| \leq M$ for some real number $M < \infty$. Then $\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$.*

THEOREM 8.4. *(1) If $f_n \geq 0$ for all $n$, then $\int \sum_n f_n d\mu = \sum_n \int f_n d\mu$.*
*(2) If $\sum_n f_n$ converges a.e. and $|\sum_{k=1}^n f_k| \leq g$ a.e., where $g$ is integrable, then $\sum_n f_n$ and the $f_n$ are integrable and $\int \sum_n f_n d\mu = \sum_n \int f_n d\mu$.*

PROOF. (1) Let $g_k = \sum_{n=1}^k f_n$. Then $\{g_k\}$ is an increasing sequence of nonnegative functions with $\lim_{k\to\infty} g_k = \sum f_n$. It follows from the Monotone Convergence theorem that

$$\int \sum f_n d\mu = \lim \int g_k d\mu = \lim_{k\to\infty} \sum_{n=1}^k \int f_n d\mu = \sum \int f_n d\mu.$$

(2) Same proof as above ecept that it follows from DCT instead of MCT.

$\square$

CHAPTER 9

# Integrals over Subsets

LEMMA 9.0.1. *Let $T$ be a measurable map from $(\Omega_1, \mathcal{F}_1, \mu)$ to $(\Omega, \mathcal{F}_2)$. For $A \in \mathcal{F}_2$ define*

$$(\mu \circ T^{-1})(A) = \mu(T^{-1}(A)).$$

*Then $\mu \circ T^{-1}$ is a measure on $(\Omega_2, \mathcal{F}_2)$.*

PROOF. For disjoint elements $A_1, A_2, \ldots$ of $\mathcal{F}_2$ we have

$$\begin{aligned}(\mu \circ T^{-1})\left(\bigcup_n A_n\right) &= \mu\left(T^{-1}\left(\bigcup_n A_n\right)\right) \\ &= \mu\left(\bigcup_n T^{-1}(A_n)\right) \\ &= \sum_n \mu(T^{-1}(A_n)) \\ &= \sum_n (\mu \circ T^{-1})(A_n).\end{aligned}$$

□

THEOREM 9.1 (CHANGE OF VARIABLE). *Let $T$ be a measurable map from $(\Omega_1, \mathcal{F}_1, \mu)$ to $(\Omega_2, \mathcal{F}_2)$. Let $f$ be a measurable function on $\Omega_2$. Then*

$$\int f d(\mu \circ T^{-1}) = \int f \circ T d\mu$$

*if either integral is defined (possibly infinite).*

PROOF. For $f = I_A$, note that $f \circ T = I_{T^{-1}(A)}$ and, hence,

$$\int f d(\mu \circ T^{-1}) = (\mu \circ T^{-1})(A) = \mu(T^{-1}(A)) = \int f \circ T d\mu.$$

By linearity of the integral, the result is then true for simple function $f$. Let $f$ be a nonnegative measurable function and $g_1, g_2, \ldots$ be an increasing sequence of simple functions with limit $f$. Then by MCT

$$\begin{aligned}\int f d(\mu \circ T^{-1}) &= \int f^+ d(\mu \circ T^{-1}) - \int f^- d(\mu \circ T^{-1}) \\ &= \int f^+ \circ T d\mu - \int f^- \circ T d\mu \\ &= f \circ T d\mu.\end{aligned}$$

□

DEFINITION 9.1 (LEBESQUE INTEGRAL). *The integral of $f$ over a set $A$ in $\mathcal{F}$ is defined by*

$$\int_A f d\mu = \int f I_A d\mu$$

*If $\lambda$ is Lebesque measure then the integral $\int_{[a,b]} f d\lambda$ is called the Lebesque integral of $f$ and is usually denoted by $\int_a^b f(x)dx$.*

THEOREM 9.2. *Let $f$ be a bounded function defined on the closed, bounded interval $[a, b]$. If $f$ is Riemann integrable over $[a, b]$, then it is Lebesque integrable over $[a, b]$ and the two integrals are equal.*

DEFINITION 9.2 (LEBESQUE-STIELTJES INTEGRAL). *Let $F$ be a nondecreasing and right-continuous function on $\Re$. Define $\mu_F((x,y]) = F(y) - F(x)$ for $x \leq y$. Then the integral $\int_{[a,b]} f d\mu_F$ is called the Lebesque-Stieltjes integral of $f$ with respect to $F$ and is usually denoted by $\int_a^b f(x) dF(x)$.*

LEMMA 9.2.1. *If $A_1, A_2, \ldots$ are disjoint, and if $f$ is either nonnegative or integrable, then*

$$\int_{\bigcup_n A_n} f d\mu = \sum_n \int_{A_n} f d\mu.$$

PROOF.

$$\begin{aligned}
\int_{\bigcup_n A_n} f d\mu &= \int f I_{\bigcup_n A_n} d\mu \\
&= \lim_{N \to \infty} \inf f I_{\bigcup_{n=1}^N A_n} d\mu \\
&= \lim_{N \to \infty} \sum_{n=1}^N \int f I_{A_n} d\mu \\
&= \sum_n \int_{A_n} f d\mu.
\end{aligned}$$

□

THEOREM 9.3. *Let $f$ be a nonnegative measurable function. Define the set function $\nu$ on $\mathcal{F}$ by*

$$\nu(A) = \int_A f d\mu.$$

*Then $\nu$ is a measure.*

PROOF. From the lemma

$$\nu\left(\bigcup A_n\right) = \sum \int_{A_n} f d\mu = \sum \nu(A_n)$$

for disjoint $A_1, A_2, \ldots$ in $\mathcal{F}$. □

REMARK 9.3.1. *We say that the measure $\nu$ has density $f$ relative to $\mu$, and write $d\nu/d\mu = f$. Moreover, note that for $E$ in $\mathcal{F}$, $\mu(E) = 0$ implies that $\nu(E) = 0$.*

THEOREM 9.4 (CHAIN RULE). *Let $f$ be a nonnegative measurable function and suppose that $f = d\nu/d\mu$. If $g$ is a measurable function. Then*

$$\int g d\nu = \int g f d\mu.$$

*whenever either side exists.*

PROOF. For $g = I_A$, we have

$$\int g d\nu = \nu(A) = \int_A f d\mu = \int g f d\mu.$$

By linearity of integrals, it follows that the result is true for simple functions. That result follows for nonnegative unctions now follows from MCT the fact that any nonnegative function is an increasing limit of simple functions. finally, the result follows for integrable $f$ by applying the equation to $f^+$ and $f^-$, separately. □

CHAPTER 10

# Product Measure

DEFINITION 10.1 (PRODUCT $\sigma$-ALGEBRA). *Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be any two measure spaces. In $\Omega_1 \times \Omega_2$ let $\Re$ be the collection of all rectangles $B \times C$ with $B \in \mathcal{F}_1$ and $C \in \mathcal{F}_2$. The product $\sigma$-algebra $\mathcal{F}_1 \times F_2$ is the smallest $\sigma$-algebra on $\Omega_1 \times \Omega_2$ containing $\Re$.*

DEFINITION 10.2 (SOME NOTATIONS). *Throughout, we will assume that $\mu_1$ and $\mu_2$ are finite measures. Also, we let*

$$\begin{aligned}
\Omega &= \Omega_1 \times \Omega_2 \\
\mathcal{F} &= \mathcal{F}_1 \times \mathcal{F}_2 \\
I_1^f(\omega_1) &= \int f(\omega_1, \cdot) d\mu_2 \\
I_2^f(\omega_2) &= \int f(\cdot, \omega_2) d\mu_1.
\end{aligned}$$

LEMMA 10.0.1. *Let $f$ be a bounded $\mathcal{F}$-measurable function on $\Omega$ then*

(1) *for each $\omega_1 \in \Omega_1$, the map $f(\omega_1, \cdot) : \omega_2 \mapsto f(\omega_1, \omega_2)$ is $\mathcal{F}_2$-measurable on $\Omega_2$, and*
(2) *for each $\omega_2 \in \Omega_2$, the map $f(\cdot, \omega_2) : \omega_1 \mapsto f(\omega_1, \omega_2)$ is $\mathcal{F}_1$-measurable on $\Omega_1$.*

PROOF. Let $g : \omega_2 \mapsto (\omega_1, \omega_2)$ and note that $f(\omega_1, \cdot) = f \circ g$. Now $f$ is $\mathcal{F}$-measurable and $g$ is $\mathcal{F}_2/F$-measurable. Hence $f(\omega_1, \cdot)$ is $\mathcal{F}_2$-measurable. Similarly, we can show that $f(\cdot, \omega_2)$ is $\mathcal{F}_1$-measurable. □

LEMMA 10.0.2. *Let $f$ be a bounded $\mathcal{F}$-measurable function on $\Omega$ then*

(1) *the function $I_1^f$ is $\mathcal{F}_1$-measurable,*
(2) *the function $I_2^f$ is $\mathcal{F}_2$-measurable, and*
(3) $\int I_1^f d\mu_1 = \int I_2^f d\mu_2$.

PROOF. Let $f = I_{A \times B}$. Then $I_1^f(\omega - 1) = I_A(\omega_1)\mu_2(B)$ which is clearly $\mathcal{F}_1$-measurale. Define

$$\mathcal{G} = \{F \in \mathcal{F} : I_1^f \text{ is } F_1\text{-measurable}, f = I_F\}.$$

Then $\mathcal{G}$ is a $\sigma$-algebra containing $\Re$. It follows that $I_1^f$ is $\mathcal{F}_1$-measurable for $f = I_F, F \in \mathcal{F}$. That $I_1^f$ is $\mathcal{F}_1$-measurable for simple function $f$ follows from linearity of the integral and that it is $\mathcal{F}_1$-measurable for nonnegative $f$ follows from MCT. Finally that $I_1^f$ is $\mathcal{F}_1$-measurable for bounded measurable $f$ follows from measurability of $I_1^{f^+}$ and $I_1^{f^-}$. The measurability of $I_2^f$ is proved similarly.

To prove (c), first assume that $f = I_{A \times B}$ with $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Then $I_1^f(\omega_1) = I_A(\omega_1)\mu_2(B)$ and $I_2^f(\omega_2) = \mu_1(A) I_B(\omega_2)$. Now,

$$\int I_1^f d\mu_1 = \int I_A d\mu_1 \cdot \mu_2(B) = \mu_1(A)\mu_2(B)$$

and

$$\int I_2^f d\mu_2 = \mu_1(A) \int I_B d\mu_2 = \mu_1(A)\mu_2(B).$$

Hence, the result is true for sets of the form $A \times B$ with $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Since these sets generate $\mathcal{F}$ then the result must also be true for $I_F$ with $F \in \mathcal{F}$. By the linearity of integrals, the rsult is also true for simple $f$. The result can then be extended to nonnegative $f$ by MCT. Finally, it can be extended to bounded functions by noting that $f = f^+ - f^-$. □

THEOREM 10.1 (FUBINI'S THEOREM). *For $A \in \mathcal{F}$ define*

$$\mu(A) = \int I_1^f \, d\mu_1 = \int I_2^f \, d\mu_2$$

*with $f = I_A$. Then $\mu$ is a measure on $\mathcal{F}$. Moreover, $\mu$ is the unique measure for which*

$$\mu(A \times B) = \mu_1(A)\mu_2(B)$$

*for $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. We write this as $\mu = \mu_1 \times \mu_2$. For any nonnegative $\mathcal{F}$-measurable function $f$, we have*

$$\int f d\mu = \int I_1^f \, d\mu_1 = \int I_2^f \, d\mu_2.$$

PROOF. The previous lemma guarantees that $\mu$ as given in the theorem is well-defined. Now that $\mu$ is a measure follows from the linearity of the integral and the MCT. That the measure $\mu$ is unique follows from the fact that the collection $\Re$ of rectangles is a $\pi$-system. Now let $f = I_{A\times B}$ with $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Then

$$\int f d\mu = \mu(A \times B) = \mu_1(A)\mu_2(B) = \int I_1^f \, d\mu_1 = \int I_2^f \, d\mu_2.$$

Since sets of the form $A \times B$ with $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ generate $\mathcal{F}$ then the result must also be true for $I_F$ with $F \in \mathcal{F}$. By the linearity of integrals, the result is also true for simple $f$ and this can then be extended to nonnegative $f$ by MCT. □

DEFINITION 10.3 (INDEPENDENCE). *Let $X$ and $Y$ be two random variables deined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $X$ and $Y$ are said to be independent if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

*for $A \in \sigma(X)$ and $B \in \sigma(Y)$.*

DEFINITION 10.4 (JOINT LAW AND JOINT DISTRIBUTION). *Let $X$ and $Y$ be two random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The joint law $\mu_{X,Y}$ of the pair $(X, Y)$ is defined by*

$$\mu_{X,Y}(A) = \mathbb{P}\{(X, Y) \in A\}$$

*for $A \in \mathcal{B} \times \mathcal{B}$. The joint distribution $F_{X,Y}$ of $X$ and $Y$ is defined by*

$$F_{X,Y}(x, y) = \mathbb{P}\{X \le x, Y \le y\}.$$

THEOREM 10.2. *Let $X$ and $Y$ be two random variables with laws $\mu_X$ and $\mu_Y$, respectively and distribution functions $F_X$ and $F_Y$, respectively. The following three statements are equivalent:*

1. *$X$ and $Y$ are independent.*
2. *$\mu_{X,Y} = \mu_X \times \mu_Y$*
3. *$F_{X,Y}(x, y) = F_X(x)F_Y(y)$.*

PROOF. Assume that $X$ and $Y$ are independent. Then

$$\mu_{X,Y}(A \times B) = \mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) = \mu_X(A)\mu_Y(B).$$

hence, (1) implies (2). Now suppose that (2) is true. Then

$$F_{X,Y}(x, y) = \mu_{X,Y}((-\infty, x]\times(-\infty, y]) = \mu_X((-\infty, x])\mu_Y((-\infty, y]) = F_X(x)F_Y(y). This shows that (2) implies (3).$$

Finally to show that (3) implies (1) note that

$$\mathbb{P}(X \le x, Y \le y) = F_{X,Y}(x, y) = F_X(x)F_Y(y) = \mathbb{P}(X \le x)\mathbb{P}(Y \le y).$$

Hence $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for $A$ of the form $\{X \le x\}$ for some $x \in \Re$ and $B$ of the form $\{Y \le y\}$ for some $y \in \Re$. But sets of the form $\{X \le x\}$ generate $\sigma(X)$ and sets of the form $\{Y \le y\}$ generate $\sigma(Y)$ and so the result extends to $A \in \sigma(X)$ and $B \in \sigma(Y)$. This shows that $X$ and $Y$ are independent. Therefore, all three statements are equivalent. □

CHAPTER 11

# $\mathcal{L}^p$ Spaces

DEFINITION 11.1 ($\mathcal{L}^p$ NORM OF $f$). *For any measurable space $(\Omega, \mathcal{F}, \mu)$ and $0 < p < \infty$, $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ denotes the set of all measurable functions $f$ on $\Omega$ such that $\int |f|^p \, d\mu < \infty$ and the values of $f$ are real numbers except possibly on a set of measure 0, where $f$ may be undefined or infinite. For $1 \le p < \infty$, let $\|f\|_p = \left(\int |f|^p \, d\mu\right)^{1/p}$, called the $\mathcal{L}^p$ norm of $f$.*

THEOREM 11.1 (HÖLDER INEQUALITY). *If $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^q$, then $fg \in \mathcal{L}^1$ and*

$$\left|\int fg d\mu\right| \le \int |fg| \, d\mu \le \|f\|_p \, \|g\|_q .$$

PROOF. If $\|f\|_p = 0$, then $f = 0$ a.e., so $fg = 0$ a.e., $\int |fg| \, d\mu = 0$, and the inequality holds, and likewise if $\|g\|_q = 0$. Assume that the norms are not 0. Now for any constant $c > 0$, $\|cf\|_p = c \|f\|_p$, dividing out y the norms, we can assume $\|f\|_p = \|g\|_q = 1$. We will use the fact that for any numbers $u, v$ and $0 < \alpha < 1$, $u^\alpha v^{1-\alpha} \le au + (1-\alpha)v$. This implies that with $\alpha = 1/p$, $u = |f|^p$ and $v = |g|^q$,

$$|fg| \le \alpha \, |f|^p + (1-\alpha) \, |g|^q .$$

Integrating gives

$$|fg| \, d\mu \le \alpha \int |f|^p \, d\mu + (1-\alpha) \int |g|^q \, d\mu = \alpha \, \|f\|_p^p + (1-\alpha) \, \|g\|_q^q = 1.$$

□

COROLLARY 11.1 (CAUCHY-SCHWARZ). *For any $f$ and $g$ in $\mathcal{L}^2$, we have $fg \in \mathcal{L}^1$, and*

$$\left|\int fg d\mu\right| \le \|f\|_2 \, \|g\|_2 .$$

THEOREM 11.2 (MINKOWSKI'S). *for $1 \le p < \infty$, if $f$ and $g$ are in $\mathcal{L}^p(S, \mathcal{M}, \mu)$, then $f + g \in \mathcal{L}^p(S, \mathcal{M}, \mu)$ and*

$$\|f+g\|_p \le \|f\|_p + \|g\|_p .$$

PROOF. Since $|f+g| \le |f| + |g|$, we can replace $f$ and $g$ by their absolute values and so assume $f \ge 0$, $g \ge 0$. If $f = 0$ a.e. or $g = 0$ a.e., the inequality is clear. If $p = 1$ or $\infty$ the inequality is straightforward. For $1 < p < \infty$ we have

$$(f+g)^p \le 2^p \max(f^p, g^p) \le 2^p(f^p + g^p).$$

Applying Hölder inequality gives

$$\begin{aligned} \|f+g\|_p^p &= \int (f+g)^p d\mu \\ &= \int f(f+g)^{p=1} d\mu + \int g(f+g)^{p-1} d\mu \\ &\le \|f\|_p \left\|(f+g)^{p-1}\right\|_q + \|g\|_p \left\|(f+g)^{p-1}\right\|_q . \end{aligned}$$

Now $(p-1)q = p$, so

$$\|f+g\|_p^p \le (\|f\|_p + \|g\|_p) \, \|f+g\|_p^{p/q}$$

which leads to the result. □

DEFINITION 11.2 (SEMINORM). *A seminorm on a real vector space $X$ is a function $\|\cdot\|$ from $X$ into $[0, \infty)$ such that*

(1) $\|cx\| = |c|\,\|x\|$ *for all* $c \in \Re$ *and* $x \in X$, *and*
(2) $\|x+y\| \leq \|x\| + \|y\|$ *for all* $x, y \in X$. *A seminorm* $\|\cdot\|$ *is called a norm if and only if* $\|x\| = 0$ *only for* $x = 0$. *A normed space* $(X, \|\cdot\|)$ *is complete if for any sequence* $\{x_n\}$ *in* $X$ *with* $\lim_{m,n\to\infty} \|x_m - x_n\| = 0$ *there exists an* $x \in X$ *with* $\lim_{n\to\infty} \|x_n - x\| = 0$.

REMARK 11.2.1. *The function* $\|\cdot\|_p$ *is a seminorm on* $\mathcal{L}^p$ *but not a norm. For each* $f$, *we define the equivalence class* $[f]$ *by* $[f] = \{g : f = g \text{ a.e}\}$. *For each* $\mathcal{L}^p$ *by* $\mathcal{L}^p = \{[f] : f \in \mathcal{L}^p\}$. *On* $\mathcal{L}^p$ *we define the real valued function* $\|\cdot\|_p$ *by* $\|[f]\|_p = \|f\|_p$. *Then* $(\mathcal{L}^p, \|\cdot\|_p)$ *is a complete normed linear space.*

DEFINITION 11.3 (SEMI-INNER PRODUCT). *A semi-inner product on a real vector space* $H$ *is a function* $(\cdot,\cdot)$ *from* $H \times H$ *into* $\Re$ *such that*
(1) $(cf + g, h) = c(f, h) + (g, h)$ *for all* $c \in \Re$ *and* $f, g \in H$.
(2) $(f, g) = (g, f)$ *for all* $f, g \in H$.
(3) $(f, f) \geq 0$ *for* $f \in H$.
*A semi-inner product* $(\cdot,\cdot)$ *is called an inner product if and only if* $(f, f) = 0$ *implies* $f = 0$.

DEFINITION 11.4 (HILBERT SPACE). *Let* $(H, (\cdot,\cdot))$ *be an inner product space. Then* $H$ *is called a Hilbert space if it is complete for the norm* $\|x\| = (x, x)^{1/2}$.

THEOREM 11.3. $\mathcal{L}^2(\Omega, \mathcal{F}, \mu)$ *is a Hilbert space with inner product*

$$([f], [g]) = (f, g) = \int fg\,d\mu.$$

THEOREM 11.4. *A function* $f$ *from a Hilbert space* $H$ *into* $\Re$ *is linear and continuous if and only if for some* $h \in H$, $f(x) = (x, h)$ *for all* $x \in H$. *If so, then* $h$ *is unique.*

THEOREM 11.5. *Let* $(\Omega, \mathcal{F}, \mu)$ *be any finite measure space. Then for* $1 \leq r < s < \infty$, $\mathcal{L}^S(\Omega, \mathcal{F}, \mu) \subseteq \mathcal{L}^r(\Omega, \mathcal{F}, \mu)$, *and the identity function from* $\mathcal{L}^S$ *into* $\mathcal{L}^r$ *is continuous.*

PROOF. By Hölder inequality with $p = s/r$,

$$\int |f|^r\,d\mu = \int |f|^r \cdot 1 d\mu \leq \left(\int |f|^{r(s/r)}\right)^{r/s} \mu(\Omega)^{1/q} < \infty.$$

Thus $\|f\|_r \leq \|f\|_S\,\mu(\Omega)^{\frac{1}{qr}}$ for all $f \in \mathcal{L}^S$. This implies continuity. □

DEFINITION 11.5 (ABSOLUTELY CONTINUOUS). *Let* $\nu$ *and* $\mu$ *be two measures on the same measurable space* $(\Omega, \mathcal{F})$. *We say that* $\nu$ *is absolutely continuous with respect to* $\mu$, *denoted* $\nu \prec \mu$, *if and only if* $\nu(A) = 0$ *whenever* $\mu(A) = 0$. *We say that* $\mu$ *and* $\nu$ *are singular, denoted* $\nu \perp \mu$, *if and only if there is an* $A \in \mathcal{F}$ *with* $\mu(A) = \nu(A^c) = 0$.

THEOREM 11.6 (LEBESQUE DECOMPOSITION). *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $\mu$ *and* $\nu$ *two* $\sigma$*-finite measures on it. Then there are unique measures* $\nu_{ac}$ *and* $\nu_S$ *such that* $\nu = \nu_{ac} + \nu_s$, $\nu_{ac} \prec \mu$ *and* $\nu_S \perp \mu$.

THEOREM 11.7 (RADON-NIKODYM). *On the measurable space* $(\Omega, \mathcal{F})$ *let* $\mu$ *be a* $\sigma$*-finite measure. Let* $\nu$ *be a finite measure, absolutely continuous with respect to* $\mu$. *Then there exists a nonnegative* $f$ *such that* $\nu(A) = \int_A f d\mu$ *for all* $A$ *in* $\mathcal{F}$. *Any two such* $f$ *are equal a.e.* $(\mu)$.

PROOF. Form the Hilbert space $H = \mathcal{L}^2(\Omega, \mathcal{F}, \mu + \nu)$ then $\mathcal{L}^2 \subseteq \mathcal{L}^1$ and the identity function from $\mathcal{L}^2$ into $\mathcal{L}^1$ is continuous. The linear function $h \mapsto \int h d\nu$ is continuous from $H$ to the set of real numbers and, hence, there is a $g \in H$, such that $\int h d\nu = \int hg d(\mu + \nu)$ for all $h \in H$. Note that the above can be written as

$$\int h(1-g)d(\mu+\nu) = \int h d\mu$$

for all $h \in H$.

Let $A = \{\omega : g(\omega) = 1\}$. For all $E \in \mathcal{F}$, let $\nu_S(E) = \nu(E \cap A)$ and $\nu_{ac}(E) = \nu(E \cap A^c)$. Then $\nu_S$ and $\nu_{ac}$ are measures with $\nu = \nu_S + \nu_{ac}$ and $\nu_S = 0$. Let $f = g/(1-g)$ on $A^c$ and $f = 0$ on $A$. Then $\int_E f d\mu = \int_{E \cap A^c} g d(\mu + \nu) = \nu(E \cap A^c) = \nu(E)$. □