4-20-2011

Measure Theory, Probability, and Martingales

Xin Ma
Trinity University, xma@trinity.edu

MEASURE THEORY, PROBABILITY, AND MARTINGALES

XIN MA

A DEPARTMENT HONORS THESIS SUBMITTED TO THE
DEPARTMENT OF MATHEMATICS AT TRINITY UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR GRADUATION WITH
DEPARTMENTAL HONORS

DATE ___APRIL 20, 2011_________________

______________________________ ______________________________

THESIS ADVISOR DEPARTMENT CHAIR

____

ASSOCIATE VICE PRESIDENT FOR ACADEMIC AFFAIRS,
CURRICULUM AND STUDENT ISSUES

Measure Theory, Probability, and Martingales

Xin Ma

April 20, 2011

Abstract

This paper serves as a concise and self-contained reference to measure-theoretical probability. We study the theory of expected values as integrals with respect to probability measures on abstract spaces and the theory of conditional expectations as Radon-Nikodym derivatives. Finally, the concept of martingale and its basic properties are introduced.

Acknowledgements

I would like to express my gratitude to the Mathematics department at Trinity University for all the training it offered me during the past four years, especially Dr. Peter Olofsson, for various crucial courses relating to this thesis, and Dr. Brian Miceli, for all the kind advice, without all of which this thesis would not be possible.

Contents

Introduction

I decided to study measure-theoretical probability so that I can gain a deeper understanding of probability and stochastic processes beyond the introductory level. In particular, I studied the definition and some basic properties of martingale, which requires the understanding of expectations as integrals with respect to probability measures and the conditional expectations as Radon-Nikodym derivatives.

The main book I used along my studies is *A Probability Path* by Sidney I. Resnick. My study followed the sequence of the chapters in this book and stopped after the chapter of martingales. I also used the books *Probability and Measure* by Patrick Billingsley and *Probability and Random Processes* by Geoffery R. Grimmett and David R. Stirzaker as references. Most of the results I studied come from *A Probability Path* since it contains a comprehensive list of definitions, theorems, propositions, and their proofs. *Probability and Random Processes* takes a more intuitive approach and helped me understand the application of martingale in branching processes. *Probability and Measure* studies the expectation in a more general context not limited to probability spaces and I relied on it most in my study of the expectations.

My study starts with the set theory and probability spaces and it moves into the definition of random variables as maps. Then it deals with properties of random variables such as independence and expectation and it finally concludes with the theory of martingales.

Chapter 1

Sets and Events

1.1 Basic Set Theory

We first need to introduce the basic notations necessary throughout the study. The notations used for sets are listed below:

Ω: An abstract set representing the sample space of some experiment. The points of Ω correspond to the outcomes of an experiment.

$\mathcal{P}(\Omega)$: The power set of Ω, that is, the set of all subsets of Ω.

Subsets A, B, ... of Ω which will usually be written with Roman letter at the beginning of the alphabet. Most subsets will be thought of as events.

Collections of subsets $\mathcal{A}, \mathcal{B}$, ... which are usually denoted by calligraphic letters.

An individual element of Ω: $\omega \in \Omega$.

The set operations we need to know for our study are the following:

1. Complementation: The complement of a subset $A \in \Omega$ is

$$A^c := \{\omega : \omega \notin A\}.$$

2. Intersection over arbitrary index sets: Suppose T is some index set and for each $t \in T$ we are given $A_t \subset \Omega$. We define

$$\bigcap_{t \subset T} A_t := \{\omega : \omega \in A_t, \forall t \in T\}.$$

3. Union over arbitrary index sets: As above, let T be an index set and suppose $A_t \subset \Omega$. Define the union as

$$\bigcup_{t \in T} A_t := \{\omega : \omega \in A_t, \text{ for some } t \in T\}.$$

4. Set difference: Given two sets A, B, the part that is in A but not in B is

$$A \backslash B := AB^c.$$

5. Symmetric difference: If A,B are two subsets, the set of the points that are in one but not in both is called the symmetric difference

$$A \triangle B = (A \backslash B) \bigcup (B \backslash A).$$

1.2 Indicator Functions

If $A \subset \Omega$, we define the *indicator function* of A as

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \in A^c. \end{cases}$$

We will later see that taking the expectation of an indicator function is theoretically equivalent to computing the probability of an event.

From the definition of an indicator function, we get

$$I_A \leq I_B \iff A \subset B,$$

and

$$I_{A^c} = 1 - I_A.$$

If f and g are two functions with domain Ω and range $\mathbb{R}$, we have

$$f \leq g \iff f(\omega) \leq g(\omega) \text{ for all } \omega \in \Omega$$

and

$$f = g \text{ if } f \leq f \text{ and } g \leq f.$$

1.3 Limits of Sets

To study the convergence concepts for random variables, we need to manipulate sequences of events, which requires the definitions of limits of sets. Let $A_n \subset \Omega$. We define

$$\inf_{k\geq n} A_k := \bigcap_{k=n}^{\infty} A_k, \quad \sup_{k\geq n} A_k := \bigcup_{k=n}^{\infty} A_k$$
$$\liminf_{n\to\infty} A_n = \lim_{n\to\infty}\left(\inf_{k\geq n} A_k\right) = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k,$$
$$\limsup_{n\to\infty} A_n = \lim_{n\to\infty}\left(\sup_{k\geq n} A_k\right) = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k.$$

Let A_n be a sequence of subsets of Ω, the sample space of events. An alternative interpretation of lim sup is

$$\limsup_{n\to\infty} A_n = \{\omega : \sum_{n=1}^{\infty} I_{A_n}(\omega) = \infty\} = \{\omega : \omega \in A_{n_k}, k = 1, 2, ...\}$$

for some subsequence n_k depending on ω. Consequently,

$$\limsup_{n\to\infty} A_n = [A_n\, i.o.].$$

where *i.o.* means "infinitely often".

For lim inf, we have

$$\begin{aligned}\liminf_{n\to\infty} A_n &= \{\omega : \omega \in A_n \text{ for all n except a finite number}\} \\ &= \{\omega : \sum_n I_{A_n^c}(\omega) < \infty\} \\ &= \{\omega : \omega \in A_n, \forall n \geq n_0(\omega)\}.\end{aligned}$$

The relationship between lim sup and lim inf is

$$\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$$

since $\{\omega : \omega \in A_n, \forall n \geq n_0(\omega)\} \subset \{\omega : \omega \in A_n \text{ infinitely often }\}$.

Another connection between lim sup and lim inf is via de Morgan's laws:

$$(\liminf_{n\to\infty} A_n)^c = \limsup_{n\to\infty} A_n^c,$$

which is obtained by applying de Morgan's laws to the definitions of lim sup and lim inf.

1.4 Monotone Sequences

A sequence of sets $\{A_n\}$ is monotone non-decreasing if $A_1 \subset A_2 \subset A_3 \subset ....$ The sequence $\{A_n\}$ is monotone non-increasing if $A_1 \supset A_2 \supset A_3 \supset ....$ We use the notation $A_n \uparrow$ for non-decreasing sets and $A_n \downarrow$ for non-increasing sets.

Recall that we wish to find the limits of sequences of sets. For a monotone sequence, the limit always exits. The limits of monotone sequences are found as follows.

Proposition 1.4.1. *Suppose $\{A_n\}$ is a monotone sequence of subsets.*

1. *If $A_n \uparrow$, then $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.*

2. If $A_n \downarrow$, then $\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Recall that for any sequence $\{B_n\}$, we have

$$\inf_{k\geq n} B_k \uparrow, \text{ and } \sup_{k\geq n} B_k \downarrow.$$

It follows that

$$\liminf_{n\to\infty} B_n = \lim_{n\to\infty}\left(\inf_{k\geq n} B_n\right) = \bigcup_{n=1}^{\infty} \inf_{k\geq n} B_k, \text{ and}$$
$$\limsup_{n\to\infty} B_n = \lim_{n\to\infty}\left(\sup_{k\geq n} B_n\right) = \bigcap_{n=1}^{\infty} \sup_{k\geq n} B_k.$$

1.5 Set Operations and Closure

In this section we consider some set operations and the notion of a class of sets to be closed under certain set operations. Suppose $\mathcal{C} \subset \mathcal{P}(\Omega)$ is a collection of subsets of Ω.

Some typical set operations include arbitrary union, countable union, finite union, arbitrary intersection, countable intersection, finite intersection, complementation, and monotone limits. The definition of closure is as follows.

Definition 1.5.1 (Closure). Let $\mathcal{C}$ be a collection of subsets of Ω. $\mathcal{C}$ is *closed* under one of the set operations listed above if the set obtained by performing the set operation on sets in $\mathcal{C}$ yields a set in $\mathcal{C}$.

Example Suppose $\Omega = \mathbb{R}$, and

$$\mathcal{C} = \text{finite intervals} = \{(a, b], -\infty < a \leq b < \infty\}.$$

$\mathcal{C}$ is not closed under finite unions since $(1, 2] \cup (3, 4]$ is not a finite interval. $\mathcal{C}$ is closed under finite intersections since $(a, b] \cup (c, d] = (a \vee c, d \wedge b]$.

Example Suppose $\Omega = \mathbb{R}$ and $\mathcal{C}$ consists of the open subsets of $\mathbb{R}$. Then $\mathcal{C}$ is not closed under complementation since the complement of an open set if not open.

An event is a subset of the sample space. Generally, we cannot assign probabilities to all subsets of the sample space Ω. We call the class of subsets of Ω to which we know how to assign probabilities the event space. By manipulating subsets in the event space (events) by set operations, we can get the probabilities of more complex events. To do this, we first need to make sure that the event space is closed under the set operations, in other words, that manipulating events under the set operation does not carry events outside the event space. This is why we need the idea of closure.

1.6 Fields and σ -fields

Definition 1.6.1. A *field* is a non-empty class of subsets of Ω closed under finite union, finite intersection and complements. A synonym for field is algebra.

A minimal set of postulates for $\mathcal{A}$ to be a field is

- $\Omega \in \mathcal{A}$.
- $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$.
- $A, B \in \mathcal{A}$ implies $A \cup B \in \mathcal{A}$.

Definition 1.6.2. A $\sigma-$*field* $\mathcal{B}$ is a non-empty class of subsets of Ω closed under countable union, countable intersection and complements. A synonym for $\sigma-$field is $\sigma-$algebra.

A minimal set of postulates for $\mathcal{B}$ to be a $\sigma-$field is

- $\Omega \in \mathcal{B}$.
- $B \in \mathcal{B}$ implies $B^c \in \mathcal{B}$.
- $B_i \in \mathcal{B}, i \geq 1$ implies $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$.

Example The countable/co-countable σ-field. Let $\Omega = \mathbb{R}$, and

$$\mathcal{B} = \{A \subset \mathbb{R} : A \text{ is countable}\} \cup \{A \subset \mathbb{R} : A^c \text{ is countable}\},$$

so $\mathcal{B}$ consists of the subsets of $\mathbb{R}$ that are either countable or have countable complements. By checking the three postulates, we see that $\mathcal{B}$ is a σ-field.

Example Let $\mathcal{F}_1, \mathcal{F}_2, ...$ be a class of subsets of Ω such that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq ...$ and let $\mathcal{F} = \bigcup_{n=1}^{\infty} \mathcal{F}_n$. We can show that if the $\mathcal{F}_n$ are fields, then $\mathcal{F}$ is also a field, by checking the three postulates of a field. However, if the $\mathcal{F}_n$ are σ-fields, $\mathcal{F}$ is not necessarily a σ-field, since a countable union of elements in $\mathcal{F}$ is not necessarily an element in $\mathcal{F}$.

From the above example, we see that a countable union of fields is a field, but a countable union of σ-fields is not necessarily a σ-field.

By checking the three postulates, it is not hard to show that the intersection of σ-fields is a σ-field, which we state as a corollary below.

Corollary 1.6.1. *The intersection of σ-fields is a σ-field.*

1.7 The σ-field Generated by a Given Class $\mathcal{C}$

We call a collection of subsets of Ω a class, denoted by $\mathcal{C}$.

Definition 1.7.1. Let $\mathcal{C}$ be a collection of subsets of Ω. The σ-field generated by $\mathcal{C}$, denoted $\sigma(\mathcal{C})$, is a σ-field satisfying

- $\sigma(\mathcal{C}) \supset \mathcal{C}$
- If $\mathcal{B}'$ is some other σ-field containing $\mathcal{C}$, then $\mathcal{B}' \supset \sigma(\mathcal{C})$

The σ-field generated by a certain class $\mathcal{C}$ is also called a minimal σ-field over $\mathcal{C}$.

Proposition 1.7.1. *Given a class $\mathcal{C}$ of subsets of Ω, there is a unique minimal σ-field containing $\mathcal{C}$.*

The proof for the above proposition is abstract and non-constructive. The idea is to take the intersection of all σ-fields that contain $\mathcal{C}$ and claim that the intersection σ-field is the minimal σ-field.

In probability, $\mathcal{C}$ will be a class of events to which we know how to assign probabilities and manipulations of events in $\mathcal{C}$ will not carry events out of $\sigma(\mathcal{C})$. By measure theory, we will late see that we can assign probabilities to events in $\sigma(\mathcal{C})$, given that we know how to assign probabilities to events in $\mathcal{C}$.

1.8 Borel Sets on the Real Line

Suppose $\Omega = \mathbb{R}$ and let

$$\mathcal{C} = \{(a, b], -\infty \leq a \leq b < \infty\}.$$

Define

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{C})$$

and call $\mathcal{B}(\mathbb{R})$ the Borel subsets of $\mathbb{R}$. Basically, Borel sets on the real line are σ-fields generated by half-open intervals. It can be proven that

$$
\begin{aligned}
\mathcal{B}(\mathbb{R}) &= \sigma((a,b), -\infty \le a \le b \le \infty) \\
&= \sigma([a,b), -\infty < a \le b \le \infty) \\
&= \sigma([a,b], -\infty < a \le b < \infty) \\
&= \sigma((-\infty, x], x \in \mathbb{R}) \\
&= \sigma(\text{open subsets of } \mathbb{R})
\end{aligned}
$$

Thus Borel sets are σ-fields of any open interval of the real line.

Chapter 2

Probability Spaces

2.1 Basic Definitions and Properties

Now that we have the idea of σ-field, a collection of subsets in Ω with certain closure properties, let us look at some collections of subsets in Ω with different properties. A *structure* is a class of subsets in Ω that satisfies certain closure properties. The two main structures we will study in this chapter are the π-system and the λ-system.

We start with the definition of a π-system.

Definition 2.1.1. (π-system) $\mathcal{P}$ is a π-system if it is closed under finite intersections: $A, B \in \mathcal{P}$ implies $A \cap B \in \mathcal{P}$.

The definition of a λ-system is as follows.

Definition 2.1.2. (λ-system) $\mathcal{L}$ is a λ-system if

- $\Omega \in \mathcal{L}$

- $A \in \mathcal{L} \Rightarrow A^c \in \mathcal{L}$

- $n \neq m, A_n \bigcap A_m = \emptyset, A_n \in \mathcal{L} \Rightarrow \bigcup_n A_n \in \mathcal{L}$

We see that a λ-system and a σ-field share the same postulates except the last one. A λ-system requires closure under countable disjoint unions while a σ-field requires closure under arbitrary countable unions, which is stricter. Thus a σ-field is always a λ-system, but not vice versa.

Definition 2.1.3. A class $\mathcal{S}$ of subsets of Ω is a *semialgebra* if the following postulates hold:

- $\emptyset, \Omega \in \mathcal{S}$.
- $\mathcal{S}$ is a π-system; that is, it is closed under finite intersections.
- If $A \in \mathcal{S}$, then there exist some finite n and disjoint sets $C_1, ..., C_n$, with each $C_i \in \mathcal{S}$ such that $A^c = \sum_{i=1}^{n} C_i$.

Example Let $\Omega = \mathbb{R}$, and suppose $\mathcal{S}$ consists of intervals including $\emptyset$, the empty set:

$$\mathcal{S} = \{(a, b] : -\infty \leq a \leq b \leq \infty\}.$$

If $I_1, I_2 \in \mathcal{S}$, then $I_1 \bigcap I_2$ is an interval and in $\mathcal{S}$ and if $I \in \mathcal{S}$, then I^c is a union of disjoint intervals in $\mathcal{S}$.

2.2 Dynkin's Theorem

Now that we have the definitions of λ and π systems, we are ready to state Dynkin's theorem.

Theorem 2.2.1. *(a) If $\mathcal{P}$ is a π-system and $\mathcal{L}$ is a λ-system such that $\mathcal{P} \subset \mathcal{L}$, then $\sigma(\mathcal{P}) \subset \mathcal{L}$.*

(b) If $\mathcal{P}$ is a π-system

$$\sigma(\mathcal{P}) = \mathcal{L}(\mathcal{P})$$

that is, the minimal σ-field over $\mathcal{P}$ equals the minimal λ-system over $\mathcal{P}$.

Part (b) can be implied from part (a). Suppose (a) is true, recall that $\mathcal{P} \subset \mathcal{L}(\mathcal{P})$, by (a) we have $\sigma(\mathcal{P}) \subset \mathcal{L}(\mathcal{P})$. Remember that a σ-field is always a λ-system. So a σ-field over $\mathcal{P}$ must include the λ-system over the same $\mathcal{P}$, or, $\sigma(\mathcal{P}) \supset \mathcal{L}(\mathcal{P})$. Therefore (b) follows from (a).

2.3 Probability Spaces

Now let us get back to our σ-fields and define the probability space.

Definition 2.3.1. A *probability space* is a triple $(\Omega, \mathcal{B}, P)$ where

- Ω is the sample space corresponding to outcomes of some experiment.
- $\mathcal{B}$ is the σ-field of subsets of Ω. These subsets are called events.
- P is a probability measure; that is, P is a function with domain $\mathcal{B}$ and range [0, 1] such that

  (i) $P(A) \geq 0$ for all $A \in \mathcal{B}$

  (ii) P is σ-additive: If $\{A_n, n \geq 1\}$ are events in $\mathcal{B}$ that are disjoint, then

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

  (iii) $P(\Omega) = 1$.

Let us look at some consequences of the definition of a probability measure.

1. $P(\Omega) = 1 = P(A \cup A^c) = P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A)$.

2. $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$.

3.

$$
\begin{aligned}
P(A \cup B) &= P(A \cap B^c \cup B \cap A^c \cup A \cap B) \\
&= P(A \cap B^c + P(B \cap A^c + P(A \cap B))) \\
&= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \\
&= P(A) + P(B) - P(A \cap B).
\end{aligned}
$$

4. The inclusion-exclusion formula also follows from the definition of a probability space: If $A_1, ..., A_n$ are events, then

$$
\begin{aligned}
P(\bigcup_{j=1}^{n} A_j) &= \sum_{j=1}^{n} P(A_j) - \sum_{1 \leq i < j \leq n} P(A_i \cap Aj) \\
&\quad \sum_{1 \leq i < j < k \leq n} P(A_i \cap Aj \cap A_k) - ... \\
&\quad (-1)^{n+1} P(A_1 \cap ... \cap A_n).
\end{aligned}
$$

5. Since $P(B) = P(A) + P(B \backslash A) \geq P(A)$, $A \subset B \Rightarrow P(A) \leq P(B)$.

6. Since $\bigcup_{n=1}^{\infty} A_n = A_1 + A_1^c A_2 + A_3 A_1^c A_2^c + ...$,

$$
\begin{aligned}
P(\bigcup_{n=1}^{\infty} A_n) &= P(A_1 + A_1^c A_2 + A_3 A_1^c A_2^c + ...) \\
&= P(A_1) + P(A_1^c A_2) + P(A_3 A_1^c A_2^c) + ...) \\
&\leq P(A_1) + P(A_2) + P(A_3) + ... \text{ by (5).}
\end{aligned}
$$

7. The measure P is continuous for monotone sequences in the sense that

(i) If $A_n \uparrow A$, where $A_n \in \mathcal{B}$, then $P(A_n) \uparrow P(A)$.

(ii)If $A_n \downarrow A$, where $A_n \in \mathcal{B}$, then $P(A_n) \downarrow P(A)$.

8. Fatou's Lemma:

$$\begin{aligned}
P(\liminf_{n\to\infty} A_n) &\leq \liminf_{n\to\infty} P(A_n) \\
&\leq \limsup_{n\to\infty} P(A_n) \\
&\leq P(\limsup_{n\to\infty} A_n).
\end{aligned}$$

9. A stronger continuity result follows from Fatou's Lemma:

$$\text{If } A_n \to A, \text{ then } P(A_n) \to P(A).$$

2.4 Uniqueness of Probability Measures

In this section we will see that a probability measure is uniquely determined by its cumulative distribution function. We first start with the following proposition.

Proposition 2.4.1. *Let P_1, P_2 be two probability measures on $(\Omega, \mathcal{B})$. The class*

$$\mathcal{L} := \{A \in \mathcal{B} : P_1(A) = P_2(A)\}$$

is a λ-system.

Proof. Since $P_1(\Omega) = P_2(\Omega) = 1$, $\Omega \in \mathcal{L}$.

Let $A \in \mathcal{L}$, then $P_1(A) = P_2(A)$, then $P_1(A^c) = 1 - P_1(A) = 1 - P_2(A) = P_2(A^c)$. Thus $A \in \mathcal{L} \Rightarrow A^c \in \mathcal{L}$.

Finally, let $\{A_j\}$ be a mutually disjoint sequence of events in $\mathcal{L}$, then $P_1(A_j) = P_2(A_j)$ for all j. Hence,

$$P_1(\bigcup_j A_j) = \sum_j P_1(A_j) = \sum_j P_2(A_j) = P_2(\bigcup_j A_j)$$

so that

$$\bigcup_j A_j \in \mathcal{L}.$$

□

Corollary 2.4.2. *If P_1, P_2 are two probability measures on $(\Omega, \mathcal{B})$ and if $\mathcal{P}$ is a π-system such that*

$$\forall A \in \mathcal{P} : P_1(A) = P_2(A),$$

then

$$\forall B \in \sigma(\mathcal{P}) : P_1(B) = P_2(B).$$

Proof. Recall that $\mathcal{L} := \{A \in \mathcal{B} : P_1(A) = P_2(A)\}$ is a λ-system and note that $\mathcal{L} \supset \mathcal{P}$ clearly. By Dynkin's theorem, we have $\mathcal{L} \supset \sigma(\mathcal{P})$. Thus $\forall B \in \sigma(\mathcal{P})$: $P_1(B) = P_2(B)$. □

Corollary 2.4.3. *Let $\Omega = \mathbb{R}$. Let P_1, P_2 be two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that their cumulative distribution functions are equal:*

$$\forall x \in \mathbb{R} : F_1(x) = P_1((-\infty, x]) = F_2(x) = P_2((-\infty, x]).$$

Then

$$P_1 \equiv P_2$$

on $\mathcal{B}(\mathbb{R})$.

Proof. Let

$$\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}.$$

Then $\mathcal{P}$ is a π-system since

$$(-\infty, x] \cap (-\infty, y] = (-\infty, x \bigwedge y] \in \mathcal{P}.$$

Remember that $\sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$ as we have seen in section Borel Sets. For all $x \in \mathcal{P}$, we have $P_1(x) = P_2(x)$. Then by the previous corollary, we have $P_1 = P_2, \forall x \in \sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$. □

Thus, if a probability measure extends from $\mathcal{P}$, a π-system, to $\sigma(\mathcal{P})$, then it is unique. In fact, we can say that a probability measure extends uniquely from $\mathcal{S}$, a semialgebra, to $\sigma(\mathcal{S})$, and we state this result as the following theorem.

Theorem 2.4.4. *(Combo Extension Theorem) Suppose $\mathcal{S}$ is a semialgebra of subsets of Ω and that P is a σ-additive set function mapping $\mathcal{S}$ into [0,1] such that $P(\Omega) = 1$. There is a unique probability measure on $\sigma(\mathcal{S})$ that extends P.*

2.5 Measure Constructions

2.5.1 Lebesgue Measure on (0, 1]

Suppose

$$\Omega = (0, 1],$$

$$\mathcal{B} = \mathcal{B}((0, 1]),$$

$$\mathcal{S} = \{(a, b] : 0 \leq a \leq b \leq 1\}.$$

Define on $\mathcal{S}$ the function $\lambda : \mathcal{S} \mapsto [0, 1]$ by

$$\lambda(\emptyset) = 0, \text{ and } \lambda(a, b] = b - a.$$

It is easy to show that λ is finite additive, but it is also σ-additive in fact. By the Combo Extension Theorem, we conclude that there is a unique probability measure on $\sigma(\mathcal{S}) = \mathcal{B}((0, 1]) = \mathcal{B}$ that extends P.

2.5.2 Construction of Probability Measure on $\mathbb{R}$ with Given $F(x)$

Now that we have Lebesgue measure constructed, let us consider the construction of probability measure on the real line given a certain distribution function $F(x) = P_F((-\infty, x])$. Let us start with the definition of the left continuous inverse of F as

$$F^{\leftarrow}(y) = \inf\{s : F(s) \geq y\}, 0 < y \leq 1\}.$$

If we define $A(y) := \{s : F(s) \geq y\}$, we can show some properties of $A(y)$. Few important ones are as follows.

$F(F^{\leftarrow}(y)) \geq y$, and $F^{\leftarrow}(y) > t \iff y > F(t)$, and equivalently $F^{\leftarrow}(y) \leq t \iff y \leq F(t)$. Now let us define for $A \in \mathbb{R}$,

$$\xi_F(A) = \{x \in (0, 1] : F^{\leftarrow}(x) \in A\}.$$

Now we can define P_F as

$$P_F(A) = \lambda(\xi_F(A)),$$

where λ is Legesgue measure on (0,1]. To check that P_F is a probability distribution,

$$\begin{aligned} P_F(-\infty, x] &= \lambda(\xi_F(-\infty, x]) = \lambda\{y \in (0, 1] : F^{\leftarrow}(y) \leq x\} \\ &= \lambda\{y \in (0, 1] : y \leq F(x)\} \\ &= \lambda((0, F(x)]) \\ &= F(x). \end{aligned}$$

Chapter 3

Random Variables and Measurable Maps

We start this chapter with the definition of a random variable. A random variable is a real valued function with domain Ω which has an extra property called measurability.

The sample space Ω might potentially carry a lot of information because it contains all of the outcomes of an experiment. Sometimes we only want to focus on some aspects of the outcomes. A random variable helps us summarizing the information contained in Ω.

Example Imagine a sequence of coin flips. $\Omega = \{(\omega_1, ..., \omega_n) : \omega_i = 0 \text{ or } 1, i = 1, ..., n\}$, where 0 means a head and 1 means a tail. The total number of tails appeared during the experiment is

$$X((\omega_1, ..., \omega_n)) = \omega_1 + ... + \omega_n,$$

a function with domain Ω.

3.1 Inverse Maps

Suppose Ω and Ω' are two sets. (Frequently $\Omega' = \mathbb{R}$). Suppose

$$X : \Omega \mapsto \Omega',$$

meaning X is a function with domain Ω and range Ω'. Then X determines a function

$$X^{-1} : \mathcal{P}(\Omega') \mapsto \mathcal{P}(\Omega)$$

defined by

$$X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\}$$

for $A' \subset \Omega'$. We will see that X^{-1} preserves complementation, union and intersection. For $A' \subset \Omega', A'_t \subset \Omega'$, where T is an arbitrary index set, we have:

- $X^{-1}(\emptyset) = \emptyset, X^{-1}(\Omega') = \Omega$.
- $X^{-1}(A'^c) = (X^{-1}(A'))^c$.
- $X^{-1}(\bigcup_{t \in T} A'_t) = \bigcup_{t \in T} X^{-1}(A'_t)$, and, $X^{-1}(\bigcap_{t \in T} A'_t) = \bigcap_{t \in T} X^{-1}(A'_t)$.

If $\mathcal{C}' \subset \mathcal{P}(\Omega')$ is a class of subsets of Ω', we define

$$X^{-1}(\mathcal{C}') := \{X^{-1}(C') : C' \in \mathcal{C}'\}.$$

We are now ready to state and prove the following proposition.

Proposition 3.1.1. *If $\mathcal{B}'$ is a σ-field of subsets of Ω', then $X^{-1}(\mathcal{B}')$ is a σ-field of subsets of Ω.*

Proof. We will verify the three postulates of a σ-field for $X^{-1}(\mathcal{B}')$.

(i) Since $\Omega' \in \mathcal{B}'$, we have

$$X^{-1}(\Omega') = \Omega \in X^{-1}(\mathcal{B}').$$

(ii) If $A' \in \mathcal{B}'$, then $(A')^c \in \mathcal{B}'$, and so if $X^{-1}(A') \in X^{-1}(\mathcal{B}')$, we have

$$X^{-1}((A')^c) = (X^{-1}(A'))^c \in X^{-1}(\mathcal{B}')$$

by the fact that X^{-1} preserves complementation.

(iii) If $X^{-1}(B'_n) \in X^{-1}(\mathcal{B}')$, then

$$\bigcup_n X^{-1}(B'_n) = X^{-1}(\bigcup B'_n) \in X^{-1}(\mathcal{B}')$$

by the fact that X^{-1} preserves union.

□

In fact, a stronger result follows.

Proposition 3.1.2. *If $\mathcal{C}'$ is a class of subsets of Ω' then*

$$X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}')).$$

3.2 Measurable Maps

We call $(\Omega, \mathcal{B})$, which is a pair of set and the σ-field of subsets in it, a *measurable space* in the sense that it is ready to have a measure assigned to it. If we have two measurable spaces $(\Omega, \mathcal{B})$ and $(\Omega', \mathcal{B}')$, then a map

$$X : \Omega \to \Omega'$$

is called measurable if

$$X^{-1}(\mathcal{B}') \subset \mathcal{B}.$$

We call X a *random element* of Ω' and denote it as

$$X : (\Omega, \mathcal{B}) \mapsto (\Omega', \mathcal{B}').$$

As a special case, when $(\Omega', \mathcal{B}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we call X a *random variable*.

3.3 Induced Probability Measures

Let $(\Omega, \mathcal{B}, P)$ be a probability space and suppose

$$X : (\Omega, \mathcal{B}) \mapsto (\Omega', \mathcal{B}')$$

is measurable. For $A' \in \Omega'$, we define

$$[X \in A'] := X^{-1}(A') = \{\omega : X(\omega) \in A'\}, \text{ and}$$
$$P \circ X^{-1}(A') = P(X^{-1}(A')).$$

Then $P \circ X^{-1}$ is a probability on $(\Omega', \mathcal{B}')$ called the *induced probability.* Let us now verify that $P \circ X^{-1}$ is a probability measure on $\mathcal{B}'$.

Proof. 1. $P \circ X^{-1}(\Omega') = P(\Omega) = 1$

2. $P \circ X^{-1}(A') \geq 0$, for all $A' \in \mathcal{B}'$.

3. If $\{A'_n, n \geq 1\}$ are disjoint,

$$\begin{aligned} P \circ X^{-1}(\bigcup_n A'_n) &= P(\bigcup_n X^{-1}(A'_n)) \\ &= \sum_n P(X^{-1}(A'_n)) \\ &= \sum_n P \circ X^{-1}(A'_n) \end{aligned}$$

since $\{X^{-1}(A'_n)\}_{n \geq 1}$ are disjoint in $\mathcal{B}$. □

Thus we showed that $P \circ X^{-1}$ is a probability measure on $(\Omega', \mathcal{B}')$. As a special case, if X is a random variable, $P \circ X^{-1}$ is the probability measure induced on $\mathbb{R}$ by

$$P \circ X^{-1}(-\infty, x] = P[X \leq x] = P(\omega : X(\omega) \leq x).$$

Since P knows how to assign probabilities to elements in $\mathcal{B}$, by the concept if measurability, we know how to assign probabilities to $\mathcal{B}'$.

Example Let us consider an example of tossing two independent dice. Let

$$\Omega = \{(i,j) : 1 \leq i, j \leq 6\}.$$

Define

$$X : \Omega \mapsto \{2, 3, 4, ..., 12\} =: \Omega'$$

by

$$X((i,j)) = i + j.$$

Then

$$X^{-1}(\{4\}) = \{X = 4\} = \{(1,3), (3,1), (2,2)\} \subset \Omega.$$

We can now assign probabilities to elements in Ω'. For example,

$$\begin{aligned} P(X = 4) &= P(\omega : X(\omega) = 4) \\ &= P(\omega \in \{(1,3), (3,1), (2,2)\}) \\ &= \tfrac{1}{6} \times \tfrac{1}{6} \times 3 \\ &= \tfrac{3}{36}. \end{aligned}$$

Recall that the definition of measurability requires $X^{-1}(\mathcal{B}) \subset \mathcal{B}'$. In fact, we can show that it suffices to check that X^{-1} is well behaved on a smaller class than $\mathcal{B}'$.

Proposition 3.3.1. *(Test for measurability) Suppose*

$$X : \Omega \mapsto \Omega'$$

where $(\Omega, \mathcal{B})$ *and* $(\Omega', \mathcal{B}')$ *are two measurable spaces. Suppose* $\mathcal{C}'$ *generates* $\mathcal{B}'$*, that is,*

$$\mathcal{B}' = \sigma(\mathcal{C}').$$

Then X *is measurable if and only if*

$$X^{-1}(\mathcal{C}') \subset \mathcal{B}.$$

Proof. Suppose

$$X^{-1}(\mathcal{C}') \subset \mathcal{B}.$$

Notice that $\mathcal{B}$ is a σ-field, and it contains $X^{-1}(\mathcal{C}')$. Recall that $\sigma(X^{-1}(\mathcal{C}'))$ is the smallest σ-field that contains $X^{-1}(\mathcal{C}')$. By minimality, we must have $\sigma(X^{-1}(\mathcal{C}')) \subset \mathcal{B}$.

Recall a previous result:

$$X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}')).$$

Thus,

$$\sigma(X^{-1}(\mathcal{C}')) = X^{-1}(\sigma(\mathcal{C}')) = X^{-1}(\mathcal{B}') \subset \mathcal{B},$$

which is the definition of measurability. □

Corollary 3.3.2. *(Special case of random variable) The real valued function*

$$X : \Omega \mapsto \mathbb{R}$$

is a random variable if and only if

$$X^{-1}((-\infty, \lambda]) = [X \leq \lambda] \in \mathcal{B}, \forall \lambda \in \mathbb{R}.$$

Proof. Since $\sigma((-\infty, \lambda], \lambda \in \mathbb{R}) = \mathcal{B}(\mathbb{R})$, the corollary follows directly from the proposition for the general case. □

3.4 σ-Fields Generated by Maps

Let $X : (\Omega, \mathcal{B}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. The σ-field generated by X, denoted as $\sigma(X)$, is defined as

$$\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R})).$$

Another equivalent definition of $\sigma(X)$ is

$$\sigma(X) = \{[X \in A], A \in \mathcal{B}(\mathbb{R})\}.$$

Generally, if X is map, namely,

$$X : (\Omega, \mathcal{B}) \mapsto (\Omega', \mathcal{B}'),$$

we define

$$\sigma(X) = X^{-1}(\mathcal{B}').$$

Remember the definition of measurability, if $\sigma(X) = X^{-1}(\mathcal{B}') \subset \mathcal{F}$, where $\mathcal{F}$ is a sub-σ-field of $\mathcal{B}$, we say X is measurable with respect to $\mathcal{F}$.

Let us first look at an extreme example of the concept of σ-field induced from a random variable.

Example Let $X(\omega) \equiv 17$ for all ω. Then

$$\sigma(X) = \{[X \in B], B \in \mathcal{B}(\mathbb{R})\} = \{\emptyset, \Omega\}.$$

Since $X(\omega) \equiv 17$, there are only two cases of B. Either B contains 17 or it does not.

Let us consider a more general example now.

Example Suppose $X = I_A$ for some $A \in \mathcal{B}$. Note X has range {0,1}.

$$X^{-1}(\{0\}) = A^c, \text{ and } X^{-1}(\{1\}) = A.$$

$$\sigma(X) = \{[X \in B], B \in \mathcal{B}(\mathbb{R})\} = \{\emptyset, \Omega, A, A^c\}.$$

There are four cases of B to consider:

- $1 \in B, 0 \notin B$,
- $1 \in B, 0 \in B$,
- $1 \notin B, 0 \notin B$,
- and $1 \notin B, 0 \notin B$.

Let us look at a more complex and useful example.

Example Suppose a random variable X has range $\{a_1, ..., a_k\}$, where the a's are distinct. Define

$$A_i := X^{-1}(\{a_i\}) = [X = a_i].$$

Then $\{A_i\}$ partitions Ω. We can then represent X as

$$X = \sum_{i=1}^{k} a_i I_{A_i}, \text{ and}$$

$$\sigma(X) = \sigma(A_1, ..., A_k).$$

Chapter 4

Independence

In probability, independence is an important property that says the occurrence or non-occurrence of an event has no effect on the occurrence or non-occurrence of an independent event. This intuition works well in most cases but sometimes it fails to agree with technical definitions of independence in some examples.

In this chapter we look at a series of definitions of independence with increasing sophistication.

4.1 Two Events, Finitely Many Events, Classes, and σ-Fields

For two events only, the definition of independence is stated as follows.

Definition 4.1.1. (Independence for two events) Suppose $(\Omega, \mathcal{B}, P)$ is a fixed probability space. Events $A, B \in \mathcal{B}$ are independent if

$$P(A \cap B) = P(A)P(B).$$

For finitely many events, we have the following result.

Definition 4.1.2. (Independence for finitely many events) The events $A_1, ..., A_n (n \geq 2)$ are independent if

$$P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i), \text{ for all finite } I \subset \{1, ..., n\}.$$

We then define the meaning of independent classes.

Definition 4.1.3. (Independent classes) Let $\mathcal{C}_i \subset \mathcal{B}, i = 1, ..., n$. The classes $\mathcal{C}_i$ are independent, if for any choice $A_1, ..., A_n$, with $A_i \in \mathcal{C}_i, i = 1, ..., n$, we have the events $A_1, ..., A_n$ independent events.

A σ field is a structure that contains a certain class of subsets. The independence of σ-fields is defined below.

Definition 4.1.4. (Independent σ-Fields)

If for each $i = 1, ..., n$, $\mathcal{C}_i$ is a non-empty class of events satisfying

- $\mathcal{C}_i$ is a π-system,
- $\mathcal{C}_i, i = 1, ..., n$ are independent,

then,

$$\sigma(\mathcal{C}_1), ..., \sigma(\mathcal{C}_n)$$

are independent.

4.2 Arbitrary Index Space

In this section, we introduce the concept of independence on an arbitrary index space.

Definition 4.2.1. (Arbitrary number of independent classes) Let T be an arbitrary index set. The classes $\mathcal{C}_t, t \in T$ are independent families if for each finite $I, I \subset T, \mathcal{C}_t, t \in I$ is independent.

Corollary 4.2.1. *If* $\{\mathcal{C}_t, t \in T\}$ *are non-empty* π*-systems that are independent, then* $\{\sigma(\mathcal{C}_t), t \in T\}$ *are independent.*

4.3 Random Variables

Now that we have the definition of independence of σ-fields, let us look at the concept for random variables. The independence of random variable can be defined in two ways. We look at both ways in this section.

4.3.1 Definition from Induced σ-Field

Definition 4.3.1. (Independent random variables) $\{X_t, t \in T\}$ is an independent family of random variables if $\{\sigma(X_t), t \in T\}$ are independent σ-fields.

In other words, the independence of random variables are determined by the independence of of their induced σ-fields. Let us look at an example of the indicator functions.

Example Suppose $A_1, ..., A_n$ are independent events. Let $I_{A_1}, ..., I_{A_n}$ be indicator functions on $A_1, ..., A_n$. For any A_i, where $i \in \{1, ..., n\}$, $\sigma(I_{A_i}) = \{\emptyset, \Omega, A_i, A_i^c\}$. If $A_1, ..., A_n$ are independent, then $\sigma(I_{A_i})$'s are independent. By the definition of independence of random variables, we have $I_{A_1}, ..., I_{A_n}$ independent.

4.3.2 Definition by Distribution Functions

An alternate way of defining independence of random variables is in terms of the distribution functions.

For a family of random variables $\{X_t, t \in T\}$, define

$$F_J(x_t, t \in J) = P(X_t \leq x_t, t \in J)$$

for all finite subsets $J \subset T$.

Theorem 4.3.1. *A family of random variables $\{X_t, t \in T\}$ indexed by a set T, is independent iff for all finite $J \subset T$*

$$F_J(x_t, t \in J) = \prod_{t \in J} P(X_t \leq x_t), \forall x_t \in \mathbb{R}.$$

Corollary 4.3.2. *The finite collection of random variables $X_1, ..., X_k$ is independent iff*

$$P(X_1 \leq x_1, ..., X_k \leq x_k) = \prod_{i=1}^{k} P(X_i \leq x_i)$$

for all $x_i \in \mathbb{R}, i = 1, ..., k$.

Corollary 4.3.3. *The discrete random variables $X_1, ..., X_k$ with countable range $\mathcal{R}$ are independent iff*

$$P(X_i = x_i, i = 1, ..., k) = \prod_{i=1}^{k} P(X_i = x_i),$$

for all $x_i \in \mathcal{R}, i = 1, ..., k$.

4.4 Borel-Cantelli Lemma

In this section we study the Borel-Cantelli Lemma, which is useful for proving convergence.

4.4.1 First Borel-Cantelli Lemma

Proposition 4.4.1. *Let A_n be any events. If*

$$\sum_n P(A_n) < \infty,$$

then

$$P([A_n i.o.]) = P(\limsup_{n\to\infty} A_n) = 0.$$

Proof. Since

$$\begin{aligned}\sum_n P(A_n) &= P(A_1) + P(A_2) + ...\\ &= \sum_{j=1}^{n-1} P(A_j) + \sum_{j=n}^{\infty} P(A_j)\\ &< \infty,\end{aligned}$$

we have

$$\begin{aligned}\sum_{j=n}^{\infty} P(A_j) &= \sum_n P(A_n) - \sum_{j=1}^{n-1} P(A_j)\\ &\to 0\end{aligned}$$

as $n \to \infty$. Now, by the definition of lim sup,

$$\begin{aligned}P([A_n i.o.]) &= P(\lim_{n\to\infty} \bigcup_{j\geq n} A_j)\\ &= \lim_{n\to\infty} P(\bigcup_{j\geq n} A_j) \text{ by the continuity of P}\\ &\leq \limsup_{n\to\infty} \sum_{j=n}^{\infty} P(A_j) \text{ by subadditivity of } A_j\\ &= 0,\end{aligned}$$

since $\sum_{j=n}^{\infty} P(A_j) \to 0$ as $n \to \infty$.

□

4.4.2 Second Borel-Cantelli Lemma

The previous result does not require independence among events, but the Second Borel-Cantelli Lemma does require independence.

Proposition 4.4.2. *If A_n is a sequence of independent events, then*

$$P([A_n i.o.]) = \begin{cases} 0 & \text{iff } \sum_n P(A_n) < \infty, \\ 1 & \text{iff } \sum_n P(A_n) = \infty. \end{cases}$$

Proof of the above lemma can be found on page 103 in *A Probability Path* by Sidney I. Resnick.

Chapter 5

Integration and Expectation

In this chapter we study the expectation of a random variable and its relation to the Riemann integral.

5.1 Simple Functions

To study the expectation, we first need the definition of a simple function.

Definition 5.1.1. Give a probability space $(\Omega, \mathcal{B}, P)$,

$$X : \Omega \mapsto \mathbb{R}$$

is simple if it has a finite range.

If a simple function is $\mathcal{B}/\mathcal{B}(\mathbb{R})$ measurable, the it can be written in the form

$$X(\omega) = \sum_{i=1}^{k} a_i I_{A_i}(\omega),$$

where a_i's are the possible values of X and $A_i := X^{-1}(\{a_i\})$ partitions Ω.

5.2 Measurability

Let $\mathcal{E}$ be the set of all simple functions on Ω.

We then have a result that shows that any nonnegative measurable function can be approximated by simple functions.

Theorem 5.2.1. *(Measurability Theorem) Suppose $X(\omega) \geq 0$, for all ω. Then $X \in \mathcal{B}/\mathcal{B}(\mathbb{R})$ (X is measurable) iff there exist simple functions $X_n \in \varepsilon$ and*

$$0 \leq X_n \uparrow X.$$

5.3 Expectation of Simple Functions

Definition 5.3.1. Suppose X is a simple random variable of the form

$$X = \sum_{i=1}^{n} a_i I_{A_i}$$

where $|a_i| < \infty$, and $\sum_{i=1}^{k} A_i = \Omega$. Then for $X \in \mathcal{E}$, we have

$$E[X] \equiv \int X dP =: \sum_{i=1}^{k} a_i P(A_i).$$

Notice that the definition agrees with our intuition that, for discrete random variables, the expectation is the weighted average of all its possible values, according to the probabilities assigned to each value.

5.4 Properties

Some nice properties arise from the above definition.

1. $E[1] = 1$ and $E[I_A] = P(A)$, since $E[I_A] = 1 \times I_A + 0 \times I_{A^c}$.

2. $X \geq 0 \Rightarrow E[X] \geq 0$.

Recall the definition of $E[X] = \sum_{i=1}^{k} a_i P(A_i)$. If $a_i \geq 0$ for all i, then $E[X] \geq 0$.

3. The expectation is linear, or, $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$ for $\alpha, \beta \in \mathbb{R}$.

4. The expectation is monotone on ε in the sense that for $X, Y \in \mathcal{E}$, if $X \leq Y$, then $E[X] \leq E[Y]$.

To show this, note that $Y - X \geq 0$. By 2, we have $E[Y - X] \geq 0$. By 3, we have $E[Y - X] = E[Y] - E[X] \geq 0 \Rightarrow E[X] \leq E[Y]$.

5. If $X_n, X \in \mathcal{E}$ and either $X_n \uparrow X or X_n \downarrow X$, then

$$E[X_n] \uparrow E[X] \text{ or } E[X_n] \downarrow E[X].$$

5.5 Monotone Convergence Theorem

Theorem 5.5.1. *(Monotone Convergence Theorem) If*

$$0 \leq X_n \uparrow X,$$

then

$$E[X_n] \uparrow E[X],$$

or equivalently,

$$E[\lim_{n\to\infty} \uparrow X_n] = \lim_{n\to\infty} \uparrow E[X_n].$$

Corollary 5.5.2. *(Series Version of MCT)*

If $\xi_j \geq 0$ are non-negative random variables for $n \geq 1$, then

$$E[\sum_{j=1}^{\infty} \xi_j] = \sum_{j=1}^{\infty} E[\xi_j].$$

Proof. We can see that this corollary derives directly from MCT:

$$
\begin{aligned}
E[\sum_{j=1}^{\infty} \xi_j] &= E[\lim_{n\to\infty} \sum_{j=1}^{n} \xi_j] \\
&= \lim_{n\to\infty} \uparrow E[\sum_{j=1}^{n} \xi_j] \\
&= \lim_{n\to\infty} \uparrow \sum_{j=1}^{n} E[\xi_j] \\
&= \sum_{j=1}^{\infty} E[\xi_j].
\end{aligned}
$$

□

5.6 Fatou's Lemma

Theorem 5.6.1. *(Fatou's Lemma) If there exists $Z \in L_1$ and $X_n \geq Z$, then*

$$E[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} E[X_n].$$

Corollary 5.6.2. *(More Fatou) If there exists $Z \in L_1$ and $X_n \leq Z$, then*

$$E[\limsup_{n\to\infty} X_n] \geq \limsup_{n\to\infty} E[X_n].$$

5.7 Dominated Convergence Theorem

We now present a stronger convergence result that arises from Fatou's Lemma.

Theorem 5.7.1. *(Dominated Convergence Theorem) If*

$$X_n \to X,$$

and there exists a dominating random variable $Z \in L_1$ such that

$$|X_n| \leq Z,$$

then

$$E[X_n] \to E[X].$$

Proof. Since

$$-Z \leq X_n \leq Z,$$

both parts of Fatou Lemma apply. We get

$$\begin{aligned} E[X] &= E[\liminf_{n\to\infty} X_n] \\ &\leq \liminf_{n\to\infty} E[X_n] \\ &\leq \limsup_{n\to\infty} E[X_n] \\ &\leq E[\limsup_{n\to\infty} X_n] \\ &= E[X]. \end{aligned}$$

Thus,

$$\liminf_{n\to\infty} E[X_n] = \limsup_{n\to\infty} E[X_n] = \lim_{n\to\infty} E[X_n] = E[X].$$

□

5.8 The Riemann vs Lebesgue Integral

5.8.1 Definition of Lebesgue Integral

Consider

$$\textstyle\sum_i [\inf_{\omega \in A_i} f(\omega)]\lambda(A_i).$$

Notice that it undertakes a very similar form to its simple function counterpart.

The Lebesgue integral is defined as the supremum of the sums:

$$\int f d\lambda = \sup \sum_i [\inf_{\omega \in A_i} f(\omega)]\lambda(A_i).$$

Alternatively, it can also be define as the infimum of the sums:

$$\inf \sum_i [\sup_{\omega \in A_i} f(\omega)]\lambda(A_i).$$

For general f, consider its positive part

$$f^+(\omega) = \begin{cases} f(\omega) & \text{if } 0 \le f(\omega) \le \infty, \\ 0 & \text{if } -\infty \le f(\omega) \le 0. \end{cases}$$

and its negative part,

$$f^-(\omega) = \begin{cases} -f(\omega) & -\infty \le f(\omega) \le 0, \\ 0 & \text{if if } 0 \le f(\omega) \le \infty. \end{cases}$$

These functions are nonnegative measurable, and $f = f^+ - f^-$, $|f| = f^+ + f^-$. Thus the general integral is defined by

$$\int f d\lambda = \int f^+ d\lambda - \int f^- d\lambda.$$

Similarly, the general expectation for any random variable is defined as

$$E[X] = E[X^+ - X^-] = E[X^+] - E[X^-].$$

5.8.2 Comparison with Riemann Integral

We are familiar with computing expectations as Riemann integral from introductory probability course. Namely,

$$E[X] = \int x f(x) dx.$$

How does the Riemann integral compare with the Legesgue integral?

Theorem 5.8.1. *(Riemann and Lebesgue) Suppose* $f : (a, b] \mapsto \mathbb{R}$ *and*

1. f is $\mathcal{B}((a,b])/\mathcal{B}(\mathbb{R})$ measurable,

2. f is Riemann-integrable on (a, b].

Let λ be Lebesgue measure on (a, b]. Then

The Riemann integral of f equals the Lebesgue integral.

For Legesgue integral, the linearity, monotonicity, monotone convergence theorem, Fatou's lemma, and dominated convergence theorem still hold. Since the Riemann integral can only integrate functions that are bounded, and the Lebesgue integral does not have such restriction, the Lebesgue integral can be applied to more functions.

Chapter 6

Martingales

6.1 The Radon-Nikodym Theorem

Let $(\Omega, \mathcal{B})$ be a measurable space. Let μ and λ be positive bounded measures on $(\Omega, \mathcal{B})$. We say that λ is absolutely continuous (AC) with respect to μ, written $\lambda << \mu$, if $\mu(A) = 0$ implies $\lambda(A) = 0$.

In order to study the concept of conditional expectation, we first need the Randon-Nikodym Theorem.

Theorem 6.1.1. *(Radon-Nikodym Theorem)*

Let $(\Omega, \mathcal{B}, P)$ be the probability space. Suppose ν is a positive bounded measure and $\nu << P$. Then there exists an integrable random variable $X \in \mathcal{B}$, such that

$$\nu(E) = \int_E X dP, \text{ for all } E \in \mathcal{B}$$

X is unique and is written

$$X = \frac{d\nu}{dP}.$$

A corollary follows from the theorem.

Corollary 6.1.2. *If* μ, ν *are* σ*-finite measures* $(\Omega, \mathcal{B})$*, there exists a measurable* $X \in \mathcal{B}$ *such that*

$$\nu(A) = \int_A X d\mu, \forall A \in \mathcal{B}$$

iff

$$\nu << \mu.$$

The next corollary is important for the definition of conditional expectation.

Corollary 6.1.3. *Suppose* Q *and* P *are probability measures on* $(\Omega, \mathcal{B})$ *such that* $Q << P$*. Let* $\mathcal{G} \subset \mathcal{B}$ *be a sub-*σ*-field. Let* $Q|\mathcal{G}, P|\mathcal{G}$ *be the restrictions of* Q *and* P *to* $\mathcal{G}$*. Then in* $(\Omega, \mathcal{G})$

$$Q|\mathcal{G} << P|\mathcal{G}$$

and

$$\frac{dQ|\mathcal{G}}{dP|\mathcal{G}} \text{ is } \mathcal{G}\text{-measurable.}$$

6.2 Conditional Expectation

Suppose $X \in L_1(\Omega, \mathcal{B}, P)$ and let $\mathcal{G} \subset \mathcal{B}$ be a sub-σ-field. Then there exists a random variable $E[X|\mathcal{G}]$, called the conditional expectation of X with respect to $\mathcal{G}$, such that

(i)$E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable and integrable.

(ii) For all $G \in \mathcal{G}$ we have

$$\int_G X dP = \int_G E[X|\mathcal{G}] dP.$$

To show this, define

$$\nu(A) = \int_A X dP, A \in \mathcal{B}.$$

Then ν is finite and $\nu << P$. So

$$\nu|\mathcal{G} << P|\mathcal{G}.$$

By Radon-Nikodym theorem, there exists random variable $X|\mathcal{G}$ such that

$$E[X|\mathcal{G}] = \frac{d\nu|\mathcal{G}}{dP|\mathcal{G}}.$$

So for all $G \in \mathcal{G}$

$$\nu|\mathcal{G}(G) = v(G) = \int_G \frac{d\nu|\mathcal{G}}{dP|\mathcal{G}} dP|\mathcal{G} = \int_G \frac{d\nu|\mathcal{G}}{dP|\mathcal{G}} dP = \int_G E[X|\mathcal{G}]dP.$$

6.3 Martingales

Loosely speaking, a martingale is a stochastic process such that the conditional expected value of an observation at some time t, given all the observations up to some earlier time s, is equal to the observation at that earlier time s.

Next, we give the technical definition of a martingale.

Suppose we are given integrable random variables $\{X_n, n \geq 0\}$ and σ-fields $\{\mathcal{B}_n, n \geq 0\}$ which are sub σ-fields of $\mathcal{B}$. Then $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a martingale if

(i) Information accumulates as time progresses in the sense that

$$\mathcal{B}_0 \subset \mathcal{B}_1 \subset \mathcal{B}_2 \subset ... \subset \mathcal{B}.$$

(ii) X_n is adapted in the sense that for each n, $X_n \in \mathcal{B}_n$, or, X_n is $\mathcal{B}_n$-measurable.

(iii) For $0 \leq m < n$,

$$E[X_n|\mathcal{B}_m] = X_m.$$

6.3.1 Submartingale

If in (iii) equality is replaced by $\geq$, then $\{X_n\}$ is called a submartingale. In other words, things are "getting better" on average.

6.3.2 Supermartingale

Similarly, if the equality is replaced by $\leq$, then $\{X_n\}$ is called a supermartingale. In other words, things are "getting worse" on average.

6.3.3 Remarks

A few notes on the definition of martingale include:

1. $\{X_n\}$ is a martingale if it is both a sub and supermartingale.
2. $\{X_n\}$ is a supermartingale iff $\{-X_n\}$ is a submartingale.
3. Postulate (iii) in the definition could be replace by

$$E[X_{n+1}|\mathcal{B}_n] = X_n, \forall n \geq 0.$$

which states that the expectation of state X_{n+1}, given the values of all the past states, stays the same as the value of state X_n.

4. If $\{X_n\}$ is a martingale, then $E[X_n]$ is constant. In the case of a submartingale, the mean increases, and for supermartingale, it decreases. Let us consider a simple example of martingale.

Example Let $X_1, X_2, ...$ be independent variables with zero means. We claim that the sequence of partial sums $S_n = X_1 + ... + X_n$ is a martingale (with respect to $\{X_n\}$).

To see this, note that

$$
\begin{aligned}
E[S_{n+1}|X_1, ..., X_n] &= E[S_n + X_{n+1}|X_1, ..., X_n] \\
&= E[S_n|X_1, ..., X_n] + E[X_{n+1}|X_1, ..., X_n] \\
&= S_n, \text{ by independence.}
\end{aligned}
$$

6.4 Martingale Convergence

In this section we study the convergence property of a martingale. Two important types of convergence are almost sure convergence and mean square convergence. We define both of them as follows.

Definition 6.4.1. (Almost Sure Convergence) Suppose we are given a probability space $(\Omega, \mathcal{B}, P)$. We say that a statement about random elements holds almost surely, if there exists an event $N \in \mathcal{B}$ with $P(N) = 0$ such that the statement holds if $\omega \in N^c$.

Definition 6.4.2. (Mean Square Convergence) If we have a sequence of random variables $X_1, ...X_n$. We X_n converges in mean squares to X if $E[(X_n - X)^2]$ converges to 0, as $n \to \infty$.

We are now ready to state the martingale convergence theorem.

Theorem 6.4.1. *(Martingale Convergence Theorem)*

If $\{S_n\}$ is a martingale with $E[S_n^2] < M < \infty$ for some M and all n, then there exist a random variable S such that S_n converges to S almost surely and in mean square.

This theorem has a more general version that deals with submartingales.

Theorem 6.4.2. *(Submartingale Convergence Theorem)*

If $\{(X_n, \mathcal{B}_n), n \geq 0\}$ is a submartingale satisfying

$$\sup_{n\in\mathbb{N}} E[X_n^+] < \infty,$$

then there exists $X_\infty \in L_1$ *such that*

$$X_n \to X_\infty.$$

The submartingale convergence theorem has a lot of applications including branching processes.

6.5 Branching Processes

A branching process is a stochastic process that models a population in which each individual in generation n produces some random number of individuals in generation $n+1$.

Each individual reproduces according to an offspring distribution. Let Z_n be the number of individuals in generation n.Then $Z_n = \sum_{k=1}^{Z_{n-1}} X_k$ where X_k's are i.i.d. with the offspring distribution. Let μ to be the mean of the offspring distribution, then it can be shown that

$$E[Z_n] = \mu^n.$$

Let $W_n = \dfrac{Z_n}{E[Z_n]}$.

Since

$$E[Z_{n+1}|Z_1,...,Z_n] = Z_n\mu,$$

we have

$$\begin{aligned} E[W_{n+1}|Z_1,...,Z_n] &= E[\frac{Z_{n+1}}{E[Z_{n+1}]}|Z_1,...,Z_n] \\ &= \frac{Z_n}{\mu^n} \\ &= W_n. \end{aligned}$$

Thus $\{W_n\}$ is a martingale.

It can be shown that

$$E[W_n^2] = 1 + \frac{\sigma^2(1-\mu^{-n})}{\mu(\mu-1)} \text{ if } \mu \neq 1, \text{ where } \sigma^2 = Var[Z_1].$$

Thus, by martingale convergence theorem, we have

$$W_n = \frac{Z_n}{\mu^n} \to W a.s., \text{ where W is a finite random variable.}$$

6.6 Stopping Time

In both real life and the mathematical world, we sometimes come across events that only depend on the past and the present, not the future. For example, the stock price fluctuations do not rely on the prices in the future.

To state this property mathematically, let us consider a probability space $(\Omega, \mathcal{B}, P)$. Let $\mathcal{B} = \{\mathcal{B}_0, ..., \mathcal{B}_n\}$ be a filtration. We think of $\mathcal{B}_n$ as representing the information which is available at time n, or more precisely, the smallest σ-field with respect to which all observations up to and including time n are measurable.

The definition of stopping time is as follows.

Definition 6.6.1. A random variable T taking values in $\{0, 1, 2...\} \cup \{\infty\}$ is called a stopping time if $\{T = n\} \in \mathcal{B}_n$ for all $n \geq 0$.

6.6.1 Optional Stopping

Theorem 6.6.1. *(Optional Stopping Theorem) Let* $(Y, \mathcal{B})$ *be a martingale and let* T *be a stopping time. Then* $E[Y_T] = E[Y_0]$ *if:*

- $P(T < \infty) = 1$,

- $E|Y_t| < \infty,$ *and*

- $E[Y_n I_{\{T>n\}}] \to 0$ *as* $n \to \infty$.

The theorem states that, under the three above conditions, the expected value of a martingale at a stopping time is equal to its initial value.

Bibliography

[1] Patrick Billingsley, *Probability and Measure*, New York, John Wiley & Sons, Inc., 1995.

[2] Sidney I. Resnick, *A Probability Path*, Birkhauser Boston, c/o Springer Science+Business Media Inc., 2005.

[3] Geoffery R. Grimmett and David R. Stirzaker, *Probability and Random Processes*, Oxford, Oxford University Press, 2001.