

Probability: Theory and Examples

Rick Durrett

Edition 4.1, April 21, 2013

Typos corrected, three new sections in Chapter 8.

Copyright 2013, All rights reserved.

4th edition published by Cambridge University Press in 2010

Probability: Theory and Examples

Rick Durrett

Edition 4.1, April 21, 2013

Typos corrected, three new sections in Chapter 8.

Copyright 2013, All rights reserved.

4th edition published by Cambridge University Press in 2010

Contents

1

Measure Theory

1

1.1

Probability Spaces

1

1.2

Distributions

8

1.3

Random Variables

12

1.4

Integration

.

15

1.5

Properties of the Integral

.....

21

1.6

Expected Value

24

1.6.1

Inequalities

24

1.6.2

Integration to the Limit

25

1.6.3

Computing Expected Values

.....

27

1.7

Product Measures, Fubini's Theorem

31

2

Laws of Large Numbers

37

2.1

Independence	37
2.1.1	
Sufficient Conditions for Independence	38
2.1.2	
Independence, Distribution, and Expectation	41
2.1.3	
Sums of Independent Random Variables	42
2.1.4	
Constructing Independent Random Variables	45
2.2	
Weak Laws of Large Numbers	47
2.2.1	
L2 Weak Laws	47
2.2.2	
Triangular Arrays	
.....	
	50
2.2.3	

Truncation	52
2.3	
Borel-Cantelli Lemmas	56
2.4	
Strong Law of Large Numbers	63
2.5	
Convergence of Random Series*	68
2.5.1	
Rates of Convergence	
.....	
71	
2.5.2	
Infinite Mean	73
2.6	
Large Deviations*	
.....	
75	
3	
Central Limit Theorems	
81	

3.1	
The De Moivre-Laplace Theorem	
81	
3.2	
Weak Convergence	
83	
3.2.1	
Examples	
83	
3.2.2	
Theory	
.	
86	
3.3	
Characteristic Functions	
91	
3.3.1	
Definition, Inversion Formula	
91	
3.3.2	
Weak Convergence	
97	
3.3.3	
Moments and Derivatives	
98	

3.3.4

Polya's Criterion*

..... 101

iii

iv

CONTENTS

3.3.5

The Moment Problem*

..... 103

3.4

Central Limit Theorems 106

3.4.1

i.i.d. Sequences 106

3.4.2

Triangular Arrays

..... 110

3.4.3

Prime Divisors (Erdős-Kac)* 114

3.4.4

Rates of Convergence (Berry-Esseen)* 118

3.5

Local Limit Theorems*

..... 121

3.6

Poisson Convergence 126

3.6.1	
The Basic Limit Theorem	126
3.6.2	
Two Examples with Dependence	130
3.6.3	
Poisson Processes	132
3.7	
Stable Laws*	135
3.8	
Infinitely Divisible Distributions*	144
3.9	
Limit Theorems in \mathbb{R}^d	147
4	
Random Walks	
153	
4.1	
Stopping Times	153
4.2	
Recurrence	
.	162
4.3	
Visits to 0, Arcsine Laws*	172
4.4	
Renewal Theory*	177
5	

Martingales	
189	
5.1	
Conditional Expectation	189
5.1.1	
Examples	191
5.1.2	
Properties	193
5.1.3	
Regular Conditional Probabilities*	197
5.2	
Martingales, Almost Sure Convergence	198
5.3	
Examples	204
5.3.1	
Bounded Increments	204
5.3.2	
Polya's Urn Scheme	205
5.3.3	
Radon-Nikodym Derivatives	206
5.3.4	
Branching Processes	209
5.4	
Doob's Inequality, Convergence in L_p	212
5.4.1	

Square Integrable Martingales*	216
5.5	
Uniform Integrability, Convergence in L^1	220
5.6	
Backwards Martingales	225
5.7	
Optional Stopping Theorems	229
6	
Markov Chains	
233	
6.1	
Definitions	233
6.2	
Examples	236
6.3	
Extensions of the Markov Property	240
6.4	
Recurrence and Transience	
.	245
6.5	
Stationary Measures	252
6.6	
Asymptotic Behavior	261
6.7	
Periodicity, Tail σ -field*	266

6.8	
General State Space*	270
6.8.1	
Recurrence and Transience	
.....	273
6.8.2	
Stationary Measures	274
6.8.3	
Convergence Theorem	275
6.8.4	
GI/G/1 queue	
.....	276

CONTENTS

v

7

Ergodic Theorems

279

7.1

Definitions and Examples 279

7.2

Birkhoff's Ergodic Theorem 283

7.3

Recurrence

..... 287

7.4

A Subadditive Ergodic Theorem* 290

7.5

Applications* 294

8

Brownian Motion

301

8.1

Definition and Construction 301

8.2

Markov Property, Blumenthal's 0-1 Law 307

8.3

Stopping Times, Strong Markov Property 312

8.4

Path Properties 315

8.4.1

Zeros of Brownian Motion 316

8.4.2

Hitting times 316

8.4.3

Lévy's Modulus of Continuity 319

8.5

Martingales 320

8.5.1

Multidimensional Brownian Motion

..... 324

8.6	
Itô's formula*	327
8.7	
Donsker's Theorem	333
8.8	
CLT's for Martingales*	340
8.9	
Empirical Distributions, Brownian Bridge	346
8.10 Weak convergence*	351
8.10.1 The space C	
.....	351
8.10.2 The Space D	353
8.11 Laws of the Iterated Logarithm*	355
A Measure Theory Details	
359	
A.1 Carathe	
'eodory's Extension Theorem	359
A.2 Which Sets Are Measurable?	
.....	364
A.3 Kolmogorov's Extension Theorem	366
A.4 Radon-Nikodym Theorem	368
A.5 Differentiating under the Integral	371

vi

CONTENTS

Chapter 1

Measure Theory

In this chapter, we will recall some definitions and results from measure theory. Our purpose here is to provide an introduction for readers who have not seen these concepts

before and to review that material for those who have. Harder proofs, especially those that do not contribute much to one's intuition, are hidden away in the appendix.

Readers with a solid background in measure theory can skip Sections 1.4, 1.5, and 1.7, which were previously part of the appendix.

1.1

Probability Spaces

Here and throughout the book, terms being defined are set in boldface. We begin with the most basic quantity. A probability space is a triple (Ω, \mathcal{F}, P) where Ω is a set of "outcomes," \mathcal{F} is a set of "events," and $P : \mathcal{F} \rightarrow [0, 1]$ is a function that assigns probabilities to events. We assume that \mathcal{F} is a σ -field (or σ -algebra), i.e., a (nonempty) collection of subsets of Ω that satisfy

(i) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, and

(ii) if $A_i \in \mathcal{F}$ is a countable sequence of sets then $\bigcup_i A_i \in \mathcal{F}$.

Here and in what follows, countable means finite or countably infinite. Since $\bigcap_i A_i = (\bigcup_i A_i^c)^c$, it follows that a σ -field is closed under countable intersections. We omit the

i

last property from the definition to make it easier to check.

Without P , (Ω, \mathcal{F}) is called a measurable space, i.e., it is a space on which we

can put a measure. A measure is a nonnegative countably additive set function; that

is, a function $\mu : F \rightarrow \mathbb{R}$ with

(i) $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in F$, and

(ii) if $A_i \in F$ is a countable sequence of disjoint sets, then

$$\mu(\cup_i A_i) =$$

$$\mu(A_i)$$

i

If $\mu(\Omega) = 1$, we call μ a probability measure. In this book, probability measures are usually denoted by P .

The next result gives some consequences of the definition of a measure that we will need later. In all cases, we assume that the sets we mention are in F .

1

2

CHAPTER 1. MEASURE THEORY

Theorem 1.1.1. Let μ be a measure on (Ω, F)

(i) monotonicity. If $A \subset B$ then $\mu(A) \leq \mu(B)$.

(ii) subadditivity. If $A \subset \cup_{m=1}^{\infty} A_m$

A

$$\mu(A)$$

$$\mu(A_m)$$

$$\mu(A) \leq \sum_{m=1}^{\infty} \mu(A_m)$$

∞

$$\sum_{m=1}^{\infty} \mu(A_m)$$

m).

(iii) continuity from below. If $A_i \uparrow A$ (i.e., $A_1 \subset A_2 \subset \dots$ and $\cup_i A_i = A$) then

$$\mu(A_i) \uparrow \mu(A).$$

(iv) continuity from above. If $A_i \downarrow A$ (i.e., $A_1 \supset A_2 \supset \dots$ and $\bigcap A_i = A$), with $\mu(A_1) < \infty$ then $\mu(A_i) \downarrow \mu(A)$.

Proof. (i) Let $B - A = B \cap A^c$ be the difference of the two sets. Using $+$ to denote disjoint union, $B = A + (B - A)$ so

$$\mu(B) = \mu(A) + \mu(B - A) \geq \mu(A).$$

(ii) Let $A = \bigcup_{n=1}^{\infty} A_n$

and for $n > 1$, $B_n = A_n - \bigcup_{m=1}^{n-1} A_m$. Since the B_n

B_n

$B_n \cap A_m = \emptyset$, $B_1 = A_1$

$B_n = A_n - \bigcup_{m=1}^{n-1} A_m$

$B_m \cap A_n = \emptyset$

B_m

B_m are

disjoint and have union A we have using (ii) of the definition of measure, $B_m \subset A_m$,

and (i) of this theorem

$\mu(A) = \mu(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$

$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$

$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$

$\mu(B_m) \leq \mu(A_m)$

$\mu(A_m)$

$\mu(B_m) \leq \mu(A_m)$

$\mu(B_m) \leq \mu(A_m)$

(iii) Let $B_n = A_n - \bigcup_{m=1}^{n-1} A_m$. Then the B_n are disjoint and have $\bigcup_{n=1}^{\infty} B_n = A$

$\mu(A) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$

$$m = A,$$

$$\cup_n$$

$$B$$

$$m=1$$

$$m = A_n \text{ so}$$

$$\infty$$

$$n$$

$$\mu(A) =$$

$$\mu(B_m) = \lim$$

$$\mu(B_m) = \lim \mu(A_n)$$

$$n \rightarrow \infty$$

$$n \rightarrow \infty$$

$$m=1$$

$$m=1$$

(iv) $A_1 - A_n \uparrow A_1 - A$ so (iii) implies $\mu(A_1 - A_n) \uparrow \mu(A_1 - A)$. Since $A_1 \supset B$ we have

$\mu(A_1 - B) = \mu(A_1) - \mu(B)$ and it follows that $\mu(A_n) \downarrow \mu(A)$.

The simplest setting, which should be familiar from undergraduate probability, is:

Example 1.1.1. Discrete probability spaces. Let Ω = a countable set, i.e., finite

or countably infinite. Let F = the set of all subsets of Ω . Let

$$P(A) =$$

$p(\omega)$ where $p(\omega) \geq 0$ and

$$p(\omega) = 1$$

$$\omega \in A$$

$$\omega \in \Omega$$

A little thought reveals that this is the most general probability measure on this space.

In many cases when Ω is a finite set, we have $p(\omega) = 1/|\Omega|$ where $|\Omega|$ = the number of points in Ω .

For a simple concrete example that requires this level of generality consider the astragali, dice used in ancient Egypt made from the ankle bones of sheep. This die could come to rest on the top side of the bone for four points or on the bottom for three points. The side of the bone was slightly rounded. The die could come to rest on a flat and narrow piece for six points or somewhere on the rest of the side for one point. There is no reason to think that all four outcomes are equally likely so we need probabilities $p_1, p_3, p_4,$ and p_6 to describe P .

To prepare for our next definition, we need

Exercise 1.1.1. (i) If $F_i, i \in I$ are σ -fields then $\cap_{i \in I} F_i$ is. Here $I = \emptyset$ is an arbitrary index set (i.e., possibly uncountable). (ii) Use the result in (i) to show if we are given a set Ω and a collection A of subsets of Ω , then there is a smallest σ -field containing A . We will call this the σ -field generated by A and denote it by $\sigma(A)$.

1.1. PROBABILITY SPACES

3

Let R^d be the set of vectors (x_1, \dots, x_d) of real numbers and R^d be the Borel sets, the smallest σ -field containing the open sets. When $d = 1$ we drop the superscript.

Example 1.1.2. Measures on the real line. Measures on (R, R) are defined by giving probability a Stieltjes measure function with the following properties:

- (i) F is nondecreasing.
- (ii) F is right continuous, i.e. $\lim_{y \downarrow x} F(y) = F(x)$.

Theorem 1.1.2. Associated with each Stieltjes measure function F there is a unique measure μ on (R, R) with $\mu((a, b]) = F(b) - F(a)$

$$\mu((a, b]) = F(b) - F(a)$$

(1.1.1)

When $F(x) = x$ the resulting measure is called Lebesgue measure.

The proof of Theorem 1.1.2 is a long and winding road, so we will content ourselves to describe the main ideas involved in this section and to hide the remaining details in the appendix in Section A.1. The choice of “closed on the right” in $(a, b]$ is dictated by the fact that if $b_n \downarrow b$ then we have

$$\bigcap_n (a, b_n] = (a, b]$$

The next definition will explain the choice of “open on the left.”

A collection S of sets is said to be a semialgebra if (i) it is closed under intersection, i.e., $S, T \in S$ implies $S \cap T \in S$, and (ii) if $S \in S$ then S^c is a finite disjoint union of sets in S . An important example of a semialgebra is

Example 1.1.3. $S_d =$ the empty set plus all sets of the form

$$(a_1, b_1] \times \cdots \times (a_d, b_d] \subset \mathbb{R}^d$$

where $-\infty \leq a_i < b_i \leq \infty$

The definition in (1.1.1) gives the values of μ on the semialgebra S_1 . To go from semialgebra to σ -algebra we use an intermediate step. A collection A of subsets of Ω is called an algebra (or field) if $A, B \in A$ implies A^c and $A \cup B$ are in A . Since $A \cap B = (A^c \cup B^c)^c$, it follows that $A \cap B \in A$. Obviously a σ -algebra is an algebra.

An example in which the converse is false is:

Example 1.1.4. Let $\Omega = \mathbb{Z}$ = the integers. $A =$ the collection of $A \subset \mathbb{Z}$ so that A or A^c is finite is an algebra.

Lemma 1.1.3. If S is a semialgebra then

$$S = \{\text{finite disjoint unions of sets in } S\}$$

is an algebra, called the algebra generated by S .

Proof. Suppose $A = \bigcup_i S_i$ and $B = \bigcup_j T_j$, where \bigcup denotes disjoint union and we assume the index sets are finite. Then $A \cap B = \bigcup_{i,j} S_i \cap T_j \in \mathcal{S}$

S. As for complements,

if $A = \bigcup_i S_i$ then $A^c = \bigcap_i S_i^c$. The definition of \mathcal{S} implies $S_i^c \in \mathcal{S}$

S. We have shown

\mathcal{S}

\mathcal{S}

that \mathcal{S}

\mathcal{S} is closed under intersection, so it follows by induction that $A^c \in \mathcal{S}$

S.

Example 1.1.5. Let $\Omega = \mathbb{R}$ and $\mathcal{S} = \mathcal{S}_1$ then \mathcal{S}

$\mathcal{S}_1 =$ the empty set plus all sets of the

form

$\bigcup_{i=1}^k (a_i,$

$b_i]$

where $-\infty \leq a_i < b_i \leq \infty$

Given a set function μ on \mathcal{S} we can extend it to \mathcal{S}

by

μ

μ

μ

μ

μ

μ

μ