KAI LAI CHUNG
JOHN B. WALSH

Volume 249

Grundlehren
der mathematischen
Wissenschaften

A Series of
Comprehensive Studies
in Mathematics

MARKOV PROCESSES, BROWNIAN MOTION, AND TIME SYMMETRY

SECOND EDITION

Springer

Kai Lai Chung and John B. Walsh

Markov Processes, Brownian Motion, and Time Symmetry

Second Edition

Springer

Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Kai Lai Chung
Department of Mathematics
Stanford University
Stanford, CA 94305
USA

John B. Walsh
Mathematics Department
University of British Columbia
Vancouver, BC V6T 1Z2
Canada

AMS Subject Classifications (2000): 60Jxx, 60J65

Library of Congress Cataloging in Publication Data
Chung, Kai Lai
Markov processes, Brownian motion, and time symmetry / Kai Lai Chung, John B. Walsh.
p. cm.
Includes bibliographical references and index.
ISBN 0-387-22026-7
1. Markov processes. 2. Brownian motion processes. I. Walsh, John B. II. Title.
QA274.7.C52 2004
519.2′33- -dc22 2004048547

ISBN-10: 0-387-22026-7 Printed on acid-free paper.
ISBN-13: 978-0387-22026-0

Printed in the United States of America. (TS/SB)

9 8 7 6 5 4 3 2 1 SPIN 10950685

Springer is a part of Springer Science+Business Media
springeronline.com

Grundlehren der mathematischen Wissenschaften 249

A Series of Comprehensive Studies in Mathematics

Preface to the New Edition

This book consists of two parts, to be called Part I and Part II. Part I, Chapters 1 through 5, is essentially a new edition of Kai Lai Chung's Lectures from Markov Processes to Brownian Motion (1982). He has corrected a number of misprints in the original edition, and has inserted a few references and remarks, of which he says, "The latter must be regarded as randomly selected since twenty-some years is a long time to retrace steps ..." This part introduces strong Markov processes and their potential theory. In particular, it studies Brownian motion, and shows how it generates classical potential theory.

Part II, Chapters 6 through 15, began life as a set of notes for a series of lectures on time reversal and duality given at the University of Paris. I originally planned to add the essential parts of these notes to this edition to show how the reversal of time—the retracing of steps—explained so much about Markov processes and their potential theory. But like many others, I learned that the inessential parts of a cherished manuscript form at most a fuzzy empty set, while the essential parts include everything that should have been in the original, even if it wasn't. In short, this, like Topsy, just grow'd.

Indeed, reversal and duality are best understood in light of Ray processes and the Ray–Knight compactification. But it is fitting that a study of symmetry be symmetrical itself, so once I had included the Ray compactification, I had to include its mirror image, the Martin boundary. This was followed by a host of examples, remarks and theorems to show how these new ideas influence the theory and practice developed in the first part. The result was the present Part II.

In a sense, Part II deals with the same subjects as Part I, but more narrowly: using Part I for a general understanding, we are free to focus on the effects of time reversal, duality, and time-symmetry on potential theory. Certain theorems in Part I are re-proved in Part II under slightly weaker hypotheses. This is not because I want to generalize the theorems, but because I want to show them in a different light: the proofs in Part II are quite different from those of Part I.

The class of Markov processes in Part I is slightly less general than it first appears—it does not include all Markov chains, for example, nor is it closed under time-reversal. Thus, after setting the stage with preliminary sections on general theory and Markov chains, I introduce Ray processes, which are in a sense the Platonic ideals of Markov processes: any nice Markov process with stationary transition probabilities has its ideal Ray process, and the two are connected by a procedure called Ray–Knight compactification. Ray processes introduce branching points. These are customarily ignored in the Markov process theory, but they arise naturally when one considers boundary conditions. Although they may initially seem to complicate the theory, they clarify some important things, such as the classification of stopping times and path discontinuities, and they help to explain some otherwise-mysterious behavior. I consider them an essential part of the subject.

The key tools in this study are the fundamental theorem of time-reversal in §10 and Doob's theory of h-transforms in §11. These figure systematically in the proofs.

We study probabilistic potential theory under duality hypotheses. In one sense, duality and time-reversal are two aspects of the same phenomenon: the reversed process is an h-transform of the dual. In another, deeper sense, discussed in §15, duality simply implies the existence of a cofine topology. In any case, duality is a natural setting for a study of time-symmetry.

One of the themes of Part II is that the strength and elegance of the potential theory increase as the process and its dual become more alike, i.e. as the process becomes more symmetric. However, some say that near-symmetry is beautiful but perfect symmetry is bland, and I feel that the most interesting theory arises when the process and its dual are nearly alike, but when there is still some tension between symmetry and assymmetry.

Another theme is the importance of left limits, particularly at the lifetime. (The whole structure of the Martin boundary can be viewed as an attempt to understand this quantity in depth.) This leads to an interplay between the strong Markov property, the moderate Markov property, and time-reversal. Watch for it.

I have included several chapters dubbed "Fireside chats" on subjects which are relevant and illuminating, but not strictly necessary for the rest of the material. They contain careful statements of results, but no rigorous proofs. The treatment is informal and is more concerned with why the results should be true rather than why they are true. The reader can treat these as extended remarks, or simply ignore them entirely.

Finally, Part II uses the same notation and terminology as Part I, with a few notable exceptions: for instance, "optional times" become "stopping times" and "superaveraging functions" become "supermedian functions." There is no deep reason for this, other than a misjudgement of the difficulty of changing these in the final computer file by a simple search-and-replace. It should not cause the reader any problems.

John B. Walsh
Vancouver, B.C., Canada
July 4, 2004

Preface to the First Edition

This book evolved from several stacks of lecture notes written over a decade and given in classes at slightly varying levels. In transforming the overlapping material into a book, I aimed at presenting some of the best features of the subject with a minimum of prerequisites and technicalities. (Needless to say, one man's technicality is another's professionalism.) But a text frozen in print does not allow for the latitude of the classroom; and the tendency to expand becomes harder to curb without the constraints of time and audience. The result is that this volume contains more topics and details than I had intended, but I hope the forest is still visible with the trees.

The book begins at the beginning with the Markov property, followed quickly by the introduction of optional times and martingales. These three topics in the discrete parameter setting are fully discussed in my book *A Course In Probability Theory* (second edition, Academic Press, 1974).* The latter will be referred to throughout this book as the *Course*, and may be considered as a general background; its specific use is limited to the material on discrete parameter martingale theory cited in §1.4. Apart from this and some dispensable references to Markov chains as examples, the book is self-contained. However, there are a very few results which are explained and used, but not proved here, the first instance being the theorem on projection in §1.6. The fundamental regularity properties of a Markov process having a Feller transition semigroup are established in Chapter 2, together with certain measurability questions which must be faced. Chapter 3 contains the basic theory as formulated by Hunt, including some special topics in the last three sections. Elements of a potential theory accompany the development, but a proper treatment would require the setting up of dual structures. Instead, the relevant circle of ideas is given a new departure in Chapter 5. Chapter 4 grew out of a short compendium as a particularly telling example, and Chapter 5 is a splinter from unincorporated sections of Chapter 4. The venerable theory of Brownian motion is so well embellished and ramified that once begun it is hard to know where to stop. In the end I have let my own propensity and capability make the choice. Thus the last three sections of the book treat several recent developments which have engaged me lately. They are included here with the hope of inducing further work in such fascinating old-and-new themes as equilibrium, energy, and reversibility.

*Third edition, 2001.

I used both the Notes and Exercises as proper non-trivial extensions of the text. In the Notes a number of regretably omitted topics are mentioned, and related to the text as a sort of guide to supplementary reading. In the Exercises there are many alternative proofs, important corollaries and examples that the reader will do well not to overlook.

The manuscript was prepared over a span of time apparently too long for me to maintain a uniform style and consistent notation. For instance, who knows whether "semipolar" should be spelled with or without a hyphen? And if both $|x|$ and $\|x\|$ are used to denote the same thing, does it really matter? Certain casual remarks and repetitions are also left in place, as they are permissible, indeed desirable, in lectures. Despite considerable pains on the part of several readers, it is perhaps too much to hope that no blunders remain undetected, especially among the exercises. I have often made a point, when assigning homework problems in class, to say that the correction of any inaccurate statement should be regarded as part of the exercise. This is of course not a defense for mistakes but merely offered as prior consolation.

Many people helped me with the task. To begin with, my first formal set of notes, contained in five folio-size, lined, students' copybooks, was prepared for a semester course given at the Eidgenössiche Technische Hochschule in the spring of 1970. My family has kept fond memories of a pleasant sojourn in a Swiss house in the great city of Zürich, and I should like to take this belated occasion to thank our hospitable hosts. Another set of notes (including the lectures given by Doob mentioned in §4.5) was taken during 1971–2 by Harry Guess, who was kind enough to send me a copy. Wu Rong, a visiting scholar from China, read the draft and the galley proofs, and checked out many exercises. The comments by R. Getoor, N. Falkner, and Liao Ming led to some final alterations. Most of the manuscript was typed by Mrs. Gail Stein, who also typed some of my other books. Mrs. Charlotte Crabtree, Mrs. Priscilla Feigen, and my daughter Marilda did some of the revisions. I am grateful to the National Science Foundation for its support of my research, some of which went into this book.

August 1981 Kai Lai Chung

Contents

Chapter 1

Markov Process

1.1. Markov Property

We begin by describing a general Markov process running on continuous time and living in a topological space. The time parameter is the set of positive numbers, considered at first as just a linearly ordered set of indices. In the discrete case this is the set of positive integers and the corresponding discussion is given in Chapter 9 of the *Course*. Thus some of the proofs below are the same as for the discrete case. Only later when properties of sample functions are introduced will the continuity of time play an essential role. As for the living space we deal with a general one because topological properties of sets such as "open" and "compact" will be much used while specific Euclidean notions such as "interval" and "sphere" do not come into question until much later.

We must introduce some new terminology and notation, but we will do this gradually as the need arises. Mathematical terms which have been defined in the *Course* will be taken for granted, together with the usual symbols to denote them. The reader can locate these through the Index of the *Course*. But we will repeat certain basic definitions with perhaps slight modifications.

Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let

$$\mathbf{T} = [0, \infty).$$

Let $\mathbf{E}$ be a locally compact separable metric space; and let $\mathscr{E}$ be the minimal Borel field in $\mathbf{E}$ containing all the open sets. The reader is referred to any standard text on real analysis for simple topological notions. Since the Euclidean space $\mathbf{R}^d$ of any dimension d is a well known particular case of an $\mathbf{E}$, the reader may content himself with thinking of $\mathbf{R}^d$ while reading about $\mathbf{E}$, which is not a bad practice in the learning process.

For each $t \in \mathbf{T}$, let

$$X_t(\omega) = X(t, \omega)$$

be a function from Ω to $\mathbf{E}$ such that

$$X_t^{-1}(\mathscr{E}) \subset \mathscr{F}.$$

This will be written as

$$X_t \in \mathscr{F}/\mathscr{E};$$

and we say that X_t is a random variable taking values in $(\mathbf{E}, \mathscr{E})$. For $\mathbf{E} = \mathbf{R}^1$, $\mathscr{E} = \mathscr{B}^1$, this reduces to the familiar notion of a real random variable. Now any family $\{X_t, t \in \mathbf{T}\}$ is called a *stochastic process*. In this generality the notion is of course not very interesting. Special classes of stochastic processes are defined by imposing certain conditions on the random variables X_t, through their joint or conditional distributions. Such conditions have been formulated by pure and applied mathematicians on a variety of grounds. By far the most important and developed is the class of Markov processes that we are going to study.

Borel field is also called σ-field or σ-algebra. As a general notation, for any family of random variables $\{Z_\alpha, \alpha \in A\}$, we will denote the σ-field generated by it by $\sigma(Z_\alpha, \alpha \in A)$. Now we put specifically

$$\mathscr{F}_t^0 = \sigma(X_s, s \in [0, t]); \qquad \mathscr{F}_t' = \sigma(X_s, s \in [t, \infty)).$$

Intuitively, an event in $\mathscr{F}_t^0$ is determined by the behavior of the process $\{X_s\}$ up to the time t; an event in $\mathscr{F}_t'$ by its behavior after t. Thus they represent respectively the "past" and "future" relative to the "present" instant t. For technical reasons, it is convenient to enlarge the past, as follows.

Let $\{\mathscr{F}_t, t \in \mathbf{T}\}$ be a family of σ-fields of sets in $\mathscr{F}$, such that

(a) if $s < t$, then $\mathscr{F}_s \subset \mathscr{F}_t$;
(b) for each t, $X_t \in \mathscr{F}_t$.

Property (a) is expressed by saying that "$\{\mathscr{F}_t\}$ is increasing"; property (b) by saying that "$\{X_t\}$ is adapted to $\{\mathscr{F}_t\}$". Clearly the family $\{\mathscr{F}_t^0\}$ satisfies both conditions and is the minimal such family in the obvious sense. Other instances of $\{\mathscr{F}_t\}$ will appear soon. The general definition of a Markov process involves $\{\mathscr{F}_t\}$ as well as $\{X_t\}$.

Definition. $\{X_t, \mathscr{F}_t, t \in \mathbf{T}\}$ is a Markov process iff one of the following equivalent conditions is satisfied:

(i) $\forall t \in \mathbf{T}, A \in \mathscr{F}_t, B \in \mathscr{F}_t'$:

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

(ii) $\forall t \in \mathbf{T}, B \in \mathscr{F}_t'$:

$$P(B \mid \mathscr{F}_t) = P(B \mid X_t).$$

(iii) $\forall t \in \mathbf{T}, A \in \mathscr{F}_t$:

$$P(A \mid \mathscr{F}_t') = P(A \mid X_t).$$

The reader is reminded that a conditional probability or expectation is an equivalence class of random variables with respect to the measure P. The equations above are all to be taken in this sense.

We shall use the two basic properties of conditional expectations, for arbitrary σ-fields $\mathscr{G}$, $\mathscr{G}_1$, $\mathscr{G}_2$ and integrable random variables Y and Z:

(a) If $Y \in \mathscr{G}$, then $E\{YZ|\mathscr{G}\} = YE\{Z|\mathscr{G}\}$;

(b) If $\mathscr{G}_1 \subset \mathscr{G}_2$, then

$$E\{E(Y|\mathscr{G}_1)|\mathscr{G}_2\} = E\{E(Y|\mathscr{G}_2)|\mathscr{G}_1\} = E\{Y|\mathscr{G}_1\}.$$

See Chapter 9 of *Course*.

Let us prove the equivalence of (i), (ii) and (iii). Assume that (i) holds, we will deduce (ii) in the following form. For each $A \in \mathscr{F}_t$ and $B \in \mathscr{F}'_t$ we have

$$E\{1_A P(B|X_t)\} = P(A \cap B). \tag{1}$$

Now the left member of (1) is equal to

$$\begin{aligned} E\{E[1_A P(B|X_t)]|X_t\} &= E\{P(A|X_t)P(B|X_t)\} \\ &= E\{P(A \cap B|X_t)\} = P(A \cap B). \end{aligned}$$

Symmetrically, we have

$$E\{1_B P(A|X_t)\} = P(A \cap B),$$

which implies (iii).

Conversely, to show for instance that (ii) implies (i), we have

$$\begin{aligned} P(A \cap B|X_t) &= E\{E(1_A \cdot 1_B|\mathscr{F}_t)|X_t\} \\ &= E\{1_A P(B|\mathscr{F}_t)|X_t\} = E\{1_A P(B|X_t)|X_t\} \\ &= P(B|X_t)E\{1_A|X_t\} = P(B|X_t)P(A|X_t). \end{aligned}$$

From here on we shall often omit such qualifying phrases as "$\forall t \in \mathbf{T}$". As a general notation, we denote by $b\mathscr{G}$ the class of bounded real-valued $\mathscr{G}$-measurable functions; by C_c the class of continuous functions on $\mathbf{E}$ with compact supports.

Form (ii) of the Markov property is the most useful one and it is equivalent to any of the following:

(iia) $\forall Y \in b\mathscr{F}'_t$:

$$E\{Y|\mathscr{F}_t\} = E\{Y|X_t\}.$$

(iib) $\forall u \geq t, f \in b\mathscr{E}$:

$$E\{f(X_u)|\mathscr{F}_t\} = E\{f(X_u)|X_t\}.$$

(iic) $\forall u \geq t, f \in C_c(E)$:

$$E\{f(X_u)|\mathscr{F}_t\} = E\{f(X_u)|X_t\}.$$

It is obvious that each of these conditions is weaker than the preceding one. To show the reverse implications we state two lemmas. As a rule, a self-evident qualifier "nonempty" for a set will be omitted as in the following proposition.

Lemma 1. *For each open set G, there exists a sequence of functions $\{f_n\}$ in C_c such that*

$$\lim_n \uparrow f_n = 1_G.$$

This is an easy consequence of our topological assumption on **E**, and gives the reader a good opportunity to review his knowledge of such things.

Lemma 2. *Let S be an arbitrary space and $\mathbb{D}$ a class of subsets of S. $\mathbb{D}$ is closed under finite intersections. Let $\mathbb{C}$ be a class of subsets of S such that $S \in \mathbb{C}$ and $\mathbb{D} \subset \mathbb{C}$. Furthermore suppose that $\mathbb{C}$ has the following closure properties*:

(a) *if $A_n \in \mathbb{C}$ and $A_n \subset A_{n+1}$ for $n \geq 1$, then $\bigcup_{n=1}^{\infty} A_n \in \mathbb{C}$;*
(b) *if $A \subset B$ and $A \in \mathbb{C}$, $B \in \mathbb{C}$, then $B - A \in \mathbb{C}$.*

Then $\mathbb{C} \supset \sigma(\mathbb{D})$.

Here as a general notation $\sigma(\mathbb{D})$ is the σ-field generated by the class of sets $\mathbb{D}$. Lemma 2 is Dynkin's form of the "monotone class theorem"; the proof is similar to Theorem 2.1.2 of *Course*, and is given as an exercise there. The reader should also figure out why the cited theorem cannot be applied directly in what follows.

Let us now prove that (iic) implies (iib). Using the notation of Lemma 1, we have by (iic)

$$E\{f_n(X_u) \mid \mathscr{F}_t\} = E\{f_n(X_u) \mid X_t\}.$$

Letting $n \to \infty$ we obtain by monotone convergence

$$P\{X_u \in G \mid \mathscr{F}_t\} = P\{X_u \in G \mid X_t\}. \tag{2}$$

Now apply Lemma 2 to the space **E**. Let $\mathbb{D}$ be the class of open sets, $\mathbb{C}$ the class of sets A satisfying

$$P\{X_u \in A \mid \mathscr{F}_t\} = P\{X_u \in A \mid X_t\}. \tag{3}$$

Of course $\mathbb{D}$ is closed under finite intersections, and $\mathbb{D} \subset \mathbb{C}$ by (2). The other properties required of $\mathbb{C}$ are simple consequences of the fact that each member of (3), as function of A, acts like a probability measure; see p. 301 of *Course* for a discussion. Hence we have $\mathbb{C} \supset \mathscr{E}$ by Lemma 2, which means that (3) is true for each A in $\mathscr{E}$, or again that (iib) is true for $f = 1_A$, $A \in \mathscr{E}$. The class of f for which (iib) is true is closed under addition, multiplication by a constant, and monotone convergence. Hence it includes $b\mathscr{E}$ by a standard approximation. [We may invoke here Problem 11 in §2.1 of *Course.*]

To prove that (iib) implies (iia), we consider first

$$Y = f_1(X_{u_1}) \cdots f_n(X_{u_n})$$

where $t \leq u_1 < \cdots < u_n$, and $f_j \in b\mathscr{E}$ for $1 \leq j \leq n$. For such a Y with $n = 1$, (iia) is just (iib). To make induction from $n-1$ to n, we write

$$E\left\{\prod_{j=1}^{n} f_j(X_{u_j}) \,\middle|\, \mathscr{F}_t\right\} = E\left\{E\left[\prod_{j=1}^{n} f_j(X_{u_j}) \,\middle|\, \mathscr{F}_{u_{n-1}}\right] \middle|\, \mathscr{F}_t\right\}$$
$$= E\left\{\prod_{j=1}^{n-1} f_j(X_{u_j}) E[f_n(X_{u_n}) \,|\, \mathscr{F}_{u_{n-1}}] \,\middle|\, \mathscr{F}_t\right\}. \tag{4}$$

Now we have by (iib)

$$E[f_n(X_{u_n}) \,|\, \mathscr{F}_{u_{n-1}}] = E[f_n(X_{u_n}) \,|\, X_{u_{n-1}}] = g(X_{u_{n-1}})$$

for some $g \in b\mathscr{E}$. Substituting this into the above and using the induction hypothesis with $f_{n-1} \cdot g$ taking the place of f_{n-1}, we see that the last term in (4) is equal to

$$E\left\{\prod_{j=1}^{n-1} f_j(X_{u_j}) E[f_n(X_{u_n}) \,|\, \mathscr{F}_{u_{n-1}}] \,\middle|\, X_t\right\} = E\left\{E\left[\prod_{j=1}^{n} f_j(X_{u_j}) \,\middle|\, \mathscr{F}_{u_{n-1}}\right] \middle|\, X_t\right\}$$
$$= E\left\{\prod_{j=1}^{n} f_j(X_{u_j}) \,\middle|\, X_t\right\}$$

since $X_t \in \mathscr{F}_{u_{n-1}}$. This completes the induction.

Now let $\mathbb{D}$ be the class of subsets of Ω of the form $\bigcap_{j=1}^{n} \{X_{u_j} \in B_j\}$ with the u_j's as before and $B_j \in \mathscr{E}$. Then $\mathbb{D}$ is closed under finite intersections. Let $\mathbb{C}$ be the class of subsets Λ of Ω such that

$$P\{\Lambda \,|\, \mathscr{F}_t\} = P\{\Lambda \,|\, X_t\}.$$

Then $\Omega \in \mathbb{C}$, $\mathbb{D} \subset \mathbb{C}$ and $\mathbb{C}$ has the properties (a) and (b) in Lemma 2. Hence by Lemma 2, $\mathbb{C} \supset \sigma(\mathbb{D})$ which is just $\mathscr{F}'_t$. Thus (iia) is true for any indicator $Y \in \mathscr{F}'_t$ [that is (ii)], and so also for any $Y \in b\mathscr{F}'_t$ by approximations. The equivalence of (iia), (iib), (iic), and (ii), is completely proved.

Finally, (iic) is equivalent to the following: for arbitrary integers $n \geq 1$ and $0 \leq t_1 < \cdots t_n < t < u$, and $f \in C_c(\mathbf{E})$ we have

$$E\{f(X_u) \,|\, X_t, X_{t_n}, \ldots, X_{t_1}\} = E\{f(X_u) \,|\, X_t\}. \tag{5}$$