Markov Chains and Stochastic Stability

S.P. Meyn and R.L. Tweedie

(Springer-Verlag, 1993; this version compiled September, 2005.)

Preface

Books are individual and idiosyncratic. In trying to understand what makes a good book, there is a limited amount that one can learn from other books; but at least one can read their prefaces, in hope of help.

Our own research shows that authors use prefaces for many different reasons.

Prefaces can be explanations of the role and the contents of the book, as in Chung [49] or Revuz [223] or Nummelin [202]; this can be combined with what is almost an apology for bothering the reader, as in Billingsley [25] or Çinlar [40]; prefaces can describe the mathematics, as in Orey [208], or the importance of the applications, as in Tong [267] or Asmussen [10], or the way in which the book works as a text, as in Brockwell and Davis [32] or Revuz [223]; they can be the only available outlet for thanking those who made the task of writing possible, as in almost all of the above (although we particularly like the familial gratitude of Resnick [222] and the dedication of Simmons [240]); they can combine all these roles, and many more.

This preface is no different. Let us begin with those we hope will use the book.

Who wants this stuff anyway?

This book is about Markov chains on general state spaces: sequences Φ_n evolving randomly in time which remember their past trajectory only through its most recent value. We develop their theoretical structure and we describe their application.

The theory of general state space chains has matured over the past twenty years in ways which make it very much more accessible, very much more complete, and (we at least think) rather beautiful to learn and use. We have tried to convey all of this, and to convey it at a level that is no more difficult than the corresponding countable space theory.

The easiest reader for us to envisage is the long-suffering graduate student, who is expected, in many disciplines, to take a course on countable space Markov chains.

Such a graduate student should be able to read almost all of the general space theory in this book without any mathematical background deeper than that needed for studying chains on countable spaces, provided only that the fear of seeing an integral rather than a summation sign can be overcome. Very little measure theory or analysis is required: virtually no more in most places than must be used to define transition probabilities. The remarkable Nummelin-Athreya-Ney regeneration technique, together with coupling methods, allows simple renewal approaches to almost all of the hard results.

Courses on countable space Markov chains abound, not only in statistics and mathematics departments, but in engineering schools, operations research groups and even business schools. This book can serve as the text in most of these environments for a one-semester course on more general space applied Markov chain theory, provided that some of the deeper limit results are omitted and (in the interests of a fourteen week semester) the class is directed only to a subset of the examples, concentrating as best suits their discipline on time series analysis, control and systems models or operations research models.

The prerequisite texts for such a course are certainly at no deeper level than Chung [50], Breiman [31], or Billingsley [25] for measure theory and stochastic processes, and Simmons [240] or Rudin [233] for topology and analysis.

Be warned: we have not provided numerous illustrative unworked examples for the student to cut teeth on. But we have developed a rather large number of thoroughly worked examples, ensuring applications are well understood; and the literature is littered with variations for teaching purposes, many of which we reference explicitly.

This regular interplay between theory and detailed consideration of application to specific models is one thread that guides the development of this book, as it guides the rapidly growing usage of Markov models on general spaces by many practitioners.

The second group of readers we envisage consists of exactly those practitioners, in several disparate areas, for all of whom we have tried to provide a set of research and development tools: for engineers in control theory, through a discussion of linear and non-linear state space systems; for statisticians and probabilists in the related areas of time series analysis; for researchers in systems analysis, through networking models for which these techniques are becoming increasingly fruitful; and for applied probabilists, interested in queueing and storage models and related analyses.

We have tried from the beginning to convey the applied value of the theory rather than let it develop in a vauum. The practitioner will find detailed examples of transition probabilities for real models. These models are classified systematically into the various structural classes as we define them. The impact of the theory on the models is developed in detail, not just to give examples of that theory but because the models themselves are important and there are relatively few places outside the research journals where their analysis is collected.

Of course, there is only so much that a general theory of Markov chains can provide to all of these areas. The contribution is in general qualitative, not quantitative. And in our experience, the critical qualitative aspects are those of stability of the models. Classification of a model as stable in some sense is the first fundamental operation underlying other, more model-specific, analyses. It is, we think, astonishing how powerful and accurate such a classification can become when using only the apparently blunt instruments of a general Markovian theory: we hope the strength of the results described here is equally visible to the reader as to the authors, for this is why we have chosen stability analysis as the cord binding together the theory and the applications of Markov chains.

We have adopted two novel approaches in writing this book. The reader will find key theorems announced at the beginning of all but the discursive chapters; if these are understood then the more detailed theory in the body of the chapter will be better motivated, and applications made more straightforward. And at the end of the book we have constructed, at the risk of repetition, "mud maps" showing the crucial equivalences between forms of stability, and giving a glossary of the models we evaluate. We trust both of these innovations will help to make the material accessible to the full range of readers we have considered.

What's it all about?

We deal here with Markov chains. Despite the initial attempts by Doob and Chung [68, 49] to reserve this term for systems evolving on countable spaces with both discrete and continuous time parameters, usage seems to have decreed (see for example Revuz [223]) that Markov chains move in discrete time, on whatever space they wish; and such are the systems we describe here.

Typically, our systems evolve on quite general spaces. Many models of practical systems are like this; or at least, they evolve on $\mathbb{R}^k$ or some subset thereof, and thus are not amenable to countable space analysis, such as is found in Chung [49], or Çinlar [40], and which is all that is found in most of the many other texts on the theory and application of Markov chains.

We undertook this project for two main reasons. Firstly, we felt there was a lack of accessible descriptions of such systems with any strong applied flavor; and secondly, in our view the theory is now at a point where it can be used properly in its own right, rather than practitioners needing to adopt countable space approximations, either because they found the general space theory to be inadequate or the mathematical requirements on them to be excessive.

The theoretical side of the book has some famous progenitors. The foundations of a theory of general state space Markov chains are described in the remarkable book of Doob [68], and although the theory is much more refined now, this is still the best source of much basic material; the next generation of results is elegantly developed in the little treatise of Orey [208]; the most current treatments are contained in the densely packed goldmine of material of Nummelin [202], to whom we owe much, and in the deep but rather different and perhaps more mathematical treatise by Revuz [223], which goes in directions different from those we pursue.

None of these treatments pretend to have particularly strong leanings towards applications. To be sure, some recent books, such as that on applied probability models by Asmussen [10] or that on non-linear systems by Tong [267], come at the problem from the other end. They provide quite substantial discussions of those specific aspects of general Markov chain theory they require, but purely as tools for the applications they have to hand.

Our aim has been to merge these approaches, and to do so in a way which will be accessible to theoreticians and to practitioners both.

So what else is new?

In the preface to the second edition [49] of his classic treatise on countable space Markov chains, Chung, writing in 1966, asserted that the general space context still had had "little impact" on the the study of countable space chains, and that this "state of mutual detachment" should not be suffered to continue. Admittedly, he was writing of continuous time processes, but the remark is equally apt for discrete time models of the period. We hope that it will be apparent in this book that the general space theory has not only caught up with its countable counterpart in the areas we describe, but has indeed added considerably to the ways in which the simpler systems are approached.

There are several themes in this book which instance both the maturity and the novelty of the general space model, and which we feel deserve mention, even in the restricted level of technicality available in a preface. These are, specifically,

(i) the use of the *splitting technique*, which provides an approach to general state space chains through regeneration methods;

(ii) the use of "Foster-Lyapunov" *drift criteria*, both in improving the theory and in enabling the classification of individual chains;

(iii) the delineation of appropriate *continuity conditions* to link the general theory with the properties of chains on, in particular, Euclidean space; and

(iv) the development of *control model* approaches, enabling analysis of models from their deterministic counterparts.

These are not distinct themes: they interweave to a surprising extent in the mathematics and its implementation.

The key factor is undoubtedly the existence and consequences of the Nummelin splitting technique of Chapter 5, whereby it is shown that if a chain $\{\Phi_n\}$ on a quite general space satisfies the simple "φ-irreducibility" condition (which requires that for some measure φ, there is at least positive probability from *any* initial point x that one of the Φ_n lies in any set of positive φ-measure; see Chapter 4), then one can induce an artificial "regeneration time" in the chain, allowing all of the mechanisms of discrete time renewal theory to be brought to bear.

Part I is largely devoted to developing this theme and related concepts, and their practical implementation.

The splitting method enables essentially all of the results known for countable space to be replicated for general spaces. Although that by itself is a major achievement, it also has the side benefit that it forces concentration on the aspects of the theory that depend, not on a countable space which gives regeneration at every step, but on a single regeneration point. Part II develops the use of the splitting method, amongst other approaches, in providing a full analogue of the positive recurrence/null recurrence/transience trichotomy central in the exposition of countable space chains, together with consequences of this trichotomy.

In developing such structures, the theory of general space chains has merely caught up with its denumerable progenitor. Somewhat surprisingly, in considering asymptotic results for positive recurrent chains, as we do in Part III, the concentration on a single regenerative state leads to stronger ergodic theorems (in terms of total variation convergence), better rates of convergence results, and a more uniform set of equivalent conditions for the strong stability regime known as positive recurrence than is typically realised for countable space chains.

The outcomes of this splitting technique approach are possibly best exemplified in the case of so-called "geometrically ergodic" chains.

Let τ_C be the hitting time on any set C: that is, the first time that the chain Φ_n returns to C; and let $P^n(x, A) = \mathsf{P}(\Phi_n \in A \mid \Phi_0 = x)$ denote the probability that the chain is in a set A at time n given it starts at time zero in state x, or the "n-step transition probabilities", of the chain. One of the goals of Part II and Part III is to link conditions under which the chain returns quickly to "small" sets C (such as finite or compact sets) , measured in terms of moments of τ_C, with conditions under which the probabilities $P^n(x, A)$ converge to limiting distributions.

Here is a taste of what can be achieved. We will eventually show, in Chapter 15, the following elegant result:

The following conditions are all equivalent for a φ-irreducible "aperiodic" (see Chapter 5) chain:

(A) *For some one "small" set C, the return time distributions have geometric tails; that is, for some $r > 1$*

$$\sup_{x \in C} \mathsf{E}_x[r^{\tau_C}] < \infty;$$

(B) *For some one "small" set C, the transition probabilities converge geometrically quickly; that is, for some $M < \infty, P^\infty(C) > 0$ and $\rho_C < 1$*

$$\sup_{x \in C} |P^n(x, C) - P^\infty(C)| \leq M \rho_C^n;$$

(C) *For some one "small" set C, there is "geometric drift" towards C; that is, for some function $V \geq 1$ and some $\beta > 0$*

$$\int P(x, dy) V(y) \leq (1 - \beta) V(x) + \mathbb{1}_C(x).$$

Each of these implies that there is a limiting probability measure π, a constant $R < \infty$ and some uniform rate $\rho < 1$ such that

$$\sup_{|f| \leq V} \ \left| \int P^n(x, dy) f(y) - \int \pi(dy) f(y) \right| \leq R V(x) \rho^n$$

where the function V is as in (C).

This set of equivalences also displays a second theme of this book: not only do we stress the relatively well-known equivalence of hitting time properties and limiting results, as between (A) and (B), but we also develop the equivalence of these with the one-step "Foster-Lyapunov" drift conditions as in (C), which we systematically derive for various types of stability.

As well as their mathematical elegance, these results have great pragmatic value. The condition (C) can be checked directly from P for specific models, giving a powerful applied tool to be used in classifying specific models. Although such drift conditions have been exploited in many continuous space applications areas for over a decade, much of the formulation in this book is new.

The "small" sets in these equivalences are vague: this is of course only the preface! It would be nice if they were compact sets, for example; and the continuity conditions we develop, starting in Chapter 6, ensure this, and much beside.

There is a further mathematical unity, and novelty, to much of our presentation, especially in the application of results to linear and non-linear systems on $\mathbb{R}^k$. We formulate many of our concepts first for deterministic analogues of the stochastic systems, and we show how the insight from such deterministic modeling flows into appropriate criteria for stochastic modeling. These ideas are taken from control theory, and forms of control of the deterministic system and stability of its stochastic generalization run in tandem. The duality between the deterministic and stochastic conditions is indeed almost exact, provided one is dealing with φ-irreducible Markov models; and the continuity conditions above interact with these ideas in ensuring that the "stochasticization" of the deterministic models gives such φ-irreducible chains.

Breiman [31] notes that he once wrote a preface so long that he never finished his book. It is tempting to keep on, and rewrite here all the high points of the book.

We will resist such temptation. For other highlights we refer the reader instead to the introductions to each chapter: in them we have displayed the main results in the chapter, to whet the appetite and to guide the different classes of user. Do not be fooled: there are many other results besides the highlights inside. We hope you will find them as elegant and as useful as we do.

Who do we owe?

Like most authors we owe our debts, professional and personal. A preface is a good place to acknowledge them.

The alphabetically and chronologically younger author began studying Markov chains at McGill University in Montréal. John Taylor introduced him to the beauty of probability. The excellent teaching of Michael Kaplan provided a first contact with Markov chains and a unique perspective on the structure of stochastic models.

He is especially happy to have the chance to thank Peter Caines for planting him in one of the most fantastic cities in North America, and for the friendship and academic environment that he subsequently provided.

In applying these results, very considerable input and insight has been provided by Lei Guo of Academia Sinica in Beijing and Doug Down of the University of Illinois. Some of the material on control theory and on queues in particular owes much to their collaboration in the original derivations.

He is now especially fortunate to work in close proximity to P.R. Kumar, who has been a consistent inspiration, particularly through his work on queueing networks and adaptive control. Others who have helped him, by corresponding on current research, by sharing enlightenment about a new application, or by developing new theoretical ideas, include Venkat Anantharam, A. Ganesh, Peter Glynn, Wolfgang Kliemann, Laurent Praly, John Sadowsky, Karl Sigman, and Victor Solo.

The alphabetically later and older author has a correspondingly longer list of influences who have led to his abiding interest in this subject. Five stand out: Chip Heathcote and Eugene Seneta at the Australian National University, who first taught the enjoyment of Markov chains; David Kendall at Cambridge, whose own fundamental work exemplifies the power, the beauty and the need to seek the underlying simplicity of such processes; Joe Gani, whose unflagging enthusiasm and support for the interaction of real theory and real problems has been an example for many years; and probably most significantly for the developments in this book, David Vere-Jones, who has shown an uncanny knack for asking exactly the right questions at times when just enough was known to be able to develop answers to them.

It was also a pleasure and a piece of good fortune for him to work with the Finnish school of Esa Nummelin, Pekka Tuominen and Elja Arjas just as the splitting technique was uncovered, and a large amount of the material in this book can actually be traced to the month surrounding the First Tuusula Summer School in 1976. Applying the methods over the years with David Pollard, Paul Feigin, Sid Resnick and Peter Brockwell has also been both illuminating and enjoyable; whilst the ongoing stimulation and encouragement to look at new areas given by Wojtek Szpankowski, Floske Spieksma, Chris Adam and Kerrie Mengersen has been invaluable in maintaining enthusiasm and energy in finishing this book.

By sheer coincidence both of us have held Postdoctoral Fellowships at the Australian National University, albeit at somewhat different times. Both of us started much of our own work in this field under that system, and we gratefully acknowledge those most useful positions, even now that they are long past.

More recently, the support of our institutions has been invaluable. Bond University facilitated our embryonic work together, whilst the Coordinated Sciences Laboratory of the University of Illinois and the Department of Statistics at Colorado State University have been enjoyable environments in which to do the actual writing.

Support from the National Science Foundation is gratefully acknowledged: grants ECS 8910088 and DMS 9205687 enabled us to meet regularly, helped to fund our students in related research, and partially supported the completion of the book.

Writing a book from multiple locations involves multiple meetings at every available opportunity. We appreciated the support of Peter Caines in Montréal, Bozenna and Tyrone Duncan at the University of Kansas, Will Gersch in Hawaii, Götz Kersting and Heinrich Hering in Germany, for assisting in our meeting regularly and helping with far-flung facilities.

Peter Brockwell, Kung-Sik Chan, Richard Davis, Doug Down, Kerrie Mengersen, Rayadurgam Ravikanth, and Pekka Tuominen, and most significantly Vladimir Kalashnikov and Floske Spieksma, read fragments or reams of manuscript as we produced them, and we gratefully acknowledge their advice, comments, corrections and encouragement. It is traditional, and in this case as accurate as usual, to say that any remaining infelicities are there despite their best efforts.

Rayadurgam Ravikanth produced the sample path graphs for us; Bob MacFarlane drew the remaining illustrations; and Francie Bridges produced much of the bibliography and some of the text. The vast bulk of the material we have done ourselves: our debt to Donald Knuth and the developers of LaTeX is clear and immense, as is our debt to Deepa Ramaswamy, Molly Shor, Rich Sutton and all those others who have kept software, email and remote telematic facilities running smoothly.

Lastly, we are grateful to Brad Dickinson and Eduardo Sontag, and to Zvi Ruder and Nicholas Pinfield and the Engineering and Control Series staff at Springer, for their patience, encouragement and help.

And finally ...

And finally, like all authors whether they say so in the preface or not, we have received support beyond the call of duty from our families. Writing a book of this magnitude has taken much time that should have been spent with them, and they have been unfailingly supportive of the enterprise, and remarkably patient and tolerant in the face of our quite unreasonable exclusion of other interests.

They have lived with family holidays where we scribbled proto-books in restaurants and tripped over deer whilst discussing Doeblin decompositions; they have endured sundry absences and visitations, with no idea of which was worse; they have seen come and go a series of deadlines with all of the structure of a renewal process.

They are delighted that we are finished, although we feel they have not yet adjusted to the fact that a similar development of the continuous time theory clearly needs to be written next.

So to Belinda, Sydney and Sophie; to Catherine and Marianne: with thanks for the patience, support and understanding, this book is dedicated to you.

Table of Contents

List of Figures