Applications of Mathematics

1. Fleming/Rishel, Deterministic and Stochastic Optimal Control (1975)
7. Vorob'ev, Game Theory: Lectures for Economists and Systems Scientists (1977)
11. Hida, Brownian Motion (1980)
27. Winkler, Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: An Introduction to Mathematical Aspects (1994)
33. Embrechts/Kluppelberg/Mikosch, Modelling Extremal Events (1997)
34. Duflo, Random Iterative Models (1997)

(continued after index)
Søren Asmussen

Applied Probability and Queues

Second Edition

With 46 Illustrations

Springer
Preface

This book treats the mathematics of queueing theory and some related areas, as well as the basic mathematical tools for the study of such models. It thus aims to serve as an introduction to queueing theory, to provide a thorough treatment of tools such as Markov processes, renewal theory, random walks, Lévy processes, matrix–analytic methods and change of measure, and to treat in some detail basic structures such as the $GI/G/1$ and $GI/G/s$ queues, Markov–modulated models, queueing networks, and models within the areas of storage, inventory and insurance risk. Within this framework the choice of topics is, however, rather traditional. The aim has been to present what I consider the basic knowledge in the area, not to advocate special directions in which the area is at present developing.

The first edition was published in 1987. This second edition incorporates about 100 extra pages containing an extended treatment of queueing networks and matrix–analytic methods as well as a number of additional topics, in particular Poisson’s equation, the fundamental matrix, insensitivity, rare events and extreme values for regenerative processes, Palm theory, rate conservation, Lévy processes, reflection, Skorokhod problems, Loynes’s lemma, Siegmund duality, light traffic, heavy tails, the Ross conjecture and ordering, and finite buffer problems.

Also, the references, typically given in the Notes following the separate sections, have been thoroughly updated. It should be noted, however, that these Notes are mainly intended as a first guidance for further reading, not as a bibliography or history of the subject. When a textbook or a survey paper dealing with a topic is available, this is the preferred reference rather than the original papers. Thus, details of priority are treated rather sporad-
ically. The principle has been to cite only the most important milestones and classical texts, but otherwise to make the references as up-to-date as possible. Thus, compared to the first edition, many older references have been removed.

The reader should be familiar with probability theory at the level of Breiman (1968), Chung (1974), Durrett (1991) or Shiryaev (1996). Most readers are likely to know large parts of Chapters I–II, which therefore may serve mainly as a refresher or reference part. However, one should note that I.5–8 has much material not usually included in introductory texts. How to read the rest of the book is a question of particular interests. The reader oriented towards queueing theory may want to concentrate first on Chapters III–IV and next on X–XII after having skimmed Chapters V, VI and VIII for needed background; the reader with more general interests will find Chapters V–IX and XIII more relevant.

The writing of both the first and the second editions of this book has been an immense pleasure to me. This is due not least to the interest shared by friends, colleagues and students. Their impact cannot be overestimated, and the list of people who in some way have influenced the book would be huge. Let me just mention and thank a few who have contributed with detailed comments on the second edition: Niels Hansen, Masakiyo Miyazawa, Mats Pihlsgård, Tomasz Rolski, Volker Schmidt, Karl Sigman and Anders Tolver Jensen. Most figures were done by Jane Bjørn Vedel (supported by MaPhySto, Aarhus) and my mother, Hanna Asmussen, typed much of the material that is close to the first edition.

Finally, I gratefully acknowledge the permission of World Scientific Publishing Co., Singapore, to incorporate some parts (XI.2 and XIII.3) which are close to the exposition in Asmussen (2000).

Søren Asmussen
Aarhus
February 2003
## Contents

Preface \hspace{1cm} v

Notation and Conventions \hspace{1cm} xi

**Part A: Simple Markovian Models** \hspace{1cm} 1

I  Markov Chains 3
1 Preliminaries 3
2 Aspects of Renewal Theory in Discrete Time 7
3 Stationarity 11
4 Limit Theory 16
5 Harmonic Functions, Martingales and Test Functions 20
6 Nonnegative Matrices 25
7 The Fundamental Matrix, Poisson’s Equation and the CLT 29
8 Foundations of the General Theory of Markov Processes 32

II  Markov Jump Processes 39
1 Basic Structure 39
2 The Minimal Construction 41
3 The Intensity Matrix 44
4 Stationarity and Limit Results 50
5 Time Reversibility 56
### Contents

#### III Queueing Theory at the Markovian Level

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Generalities</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>General Birth–Death Processes</td>
<td>71</td>
</tr>
<tr>
<td>3</td>
<td>Birth–Death Processes as Queueing Models</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>The Phase Method</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>Renewal Theory for Phase–Type Distributions</td>
<td>88</td>
</tr>
<tr>
<td>6</td>
<td>Lindley Processes</td>
<td>92</td>
</tr>
<tr>
<td>7</td>
<td>A First Look at Reflected Lévy Processes</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>Time–Dependent Properties of $M/M/1$</td>
<td>98</td>
</tr>
<tr>
<td>9</td>
<td>Waiting Times and Queue Disciplines in $M/M/1$</td>
<td>108</td>
</tr>
</tbody>
</table>

#### IV Queueing Networks and Insensitivity

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Poisson Departure Processes and Series of Queues</td>
<td>114</td>
</tr>
<tr>
<td>2</td>
<td>Jackson Networks</td>
<td>117</td>
</tr>
<tr>
<td>3</td>
<td>Insensitivity in Erlang’s Loss System</td>
<td>123</td>
</tr>
<tr>
<td>4</td>
<td>Quasi–Reversibility and Single–Node Symmetric Queues</td>
<td>125</td>
</tr>
<tr>
<td>5</td>
<td>Quasi–Reversibility in Networks</td>
<td>131</td>
</tr>
<tr>
<td>6</td>
<td>The Arrival Theorem</td>
<td>133</td>
</tr>
</tbody>
</table>

#### Part B: Some General Tools and Methods

#### V Renewal Theory

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Renewal Processes</td>
<td>138</td>
</tr>
<tr>
<td>2</td>
<td>Renewal Equations and the Renewal Measure</td>
<td>143</td>
</tr>
<tr>
<td>3</td>
<td>Stationary Renewal Processes</td>
<td>150</td>
</tr>
<tr>
<td>4</td>
<td>The Renewal Theorem in Its Equivalent Versions</td>
<td>153</td>
</tr>
<tr>
<td>5</td>
<td>Proof of the Renewal Theorem</td>
<td>158</td>
</tr>
<tr>
<td>6</td>
<td>Second–Moment Results</td>
<td>159</td>
</tr>
<tr>
<td>7</td>
<td>Excessive and Defective Renewal Equations</td>
<td>162</td>
</tr>
</tbody>
</table>

#### VI Regenerative Processes

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Basic Limit Theory</td>
<td>168</td>
</tr>
<tr>
<td>2</td>
<td>First Examples and Applications</td>
<td>172</td>
</tr>
<tr>
<td>3</td>
<td>Time–Average Properties</td>
<td>177</td>
</tr>
<tr>
<td>4</td>
<td>Rare Events and Extreme Values</td>
<td>179</td>
</tr>
</tbody>
</table>

#### VII Further Topics in Renewal Theory and Regenerative Processes

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spread–Out Distributions</td>
<td>186</td>
</tr>
<tr>
<td>2</td>
<td>The Coupling Method</td>
<td>189</td>
</tr>
<tr>
<td>3</td>
<td>Markov Processes: Regeneration and Harris Recurrence</td>
<td>198</td>
</tr>
<tr>
<td>4</td>
<td>Markov Renewal Theory</td>
<td>206</td>
</tr>
<tr>
<td>5</td>
<td>Semi–Regenerative Processes</td>
<td>211</td>
</tr>
<tr>
<td>6</td>
<td>Palm Theory, Rate Conservation and PASTA</td>
<td>213</td>
</tr>
</tbody>
</table>
VIII Random Walks 220
1 Basic Definitions ........................................ 220
2 Ladder Processes and Classification .................... 223
3 Wiener–Hopf Factorization .............................. 227
4 The Spitzer–Baxter Identities ............................ 229
5 Explicit Examples. $M/G/1, GI/M/1, GI/PH/1$ .... 233

IX Lévy Processes, Reflection and Duality 244
1 Lévy Processes .............................................. 244
2 Reflection and Loynes’s Lemma ........................... 250
3 Martingales and Transforms for Reflected Lévy Processes ........................................ 255
4 A More General Duality .................................... 260

Part C: Special Models and Methods 265

X Steady-State Properties of $GI/G/1$ 266
1 Notation. The Actual Waiting Time ..................... 266
2 The Moments of the Waiting Time ....................... 269
3 The Workload .............................................. 272
4 Queue Length Processes .................................. 276
5 $M/G/1$ and $GI/M/1$ ..................................... 279
6 Continuity of the Waiting Time ........................... 284
7 Heavy Traffic Limit Theorems ............................ 286
8 Light Traffic .............................................. 290
9 Heavy–Tailed Asymptotics ................................ 295

XI Markov Additive Models 302
1 Some Basic Examples ..................................... 302
2 Markov Additive Processes .............................. 309
3 The Matrix Paradigms $GI/M/1$ and $M/G/1$ ....... 316
4 Solution Methods .............................. 328
5 The Ross Conjecture and Other Ordering Results ... 336

XII Many–Server Queues 340
1 Comparisons with $GI/G/1$ .............................. 340
2 Regeneration and Existence of Limits ................... 344
3 The $GI/M/s$ Queue .................................... 348

XIII Exponential Change of Measure 352
1 Exponential Families .................................... 352
2 Large Deviations, Saddlepoints and the Relaxation Time ........................................ 355
3 Change of Measure: General Theory ................... 358
4 First Applications .................................... 362
<table>
<thead>
<tr>
<th>Sections</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cramér–Lundberg Theory</td>
<td>365</td>
</tr>
<tr>
<td>Siegmund’s Corrected Heavy Traffic Approximations</td>
<td>369</td>
</tr>
<tr>
<td>Rare Events Simulation</td>
<td>373</td>
</tr>
<tr>
<td>Markov Additive Processes</td>
<td>376</td>
</tr>
<tr>
<td>X  Dams, Inventories and Insurance Risk</td>
<td>380</td>
</tr>
<tr>
<td>1 Compound Poisson Dams with General Release Rule</td>
<td>380</td>
</tr>
<tr>
<td>2 Some Examples</td>
<td>387</td>
</tr>
<tr>
<td>3 Finite Buffer Capacity Models</td>
<td>389</td>
</tr>
<tr>
<td>4 Some Simple Inventory Models</td>
<td>396</td>
</tr>
<tr>
<td>5 Dual Insurance Risk Models</td>
<td>399</td>
</tr>
<tr>
<td>6 The Time to Ruin</td>
<td>401</td>
</tr>
<tr>
<td>Appendix</td>
<td>407</td>
</tr>
<tr>
<td>A1 Polish Spaces and Weak Convergence</td>
<td>407</td>
</tr>
<tr>
<td>A2 Right–Continuity and the Space $D$</td>
<td>408</td>
</tr>
<tr>
<td>A3 Point Processes</td>
<td>410</td>
</tr>
<tr>
<td>A4 Stochastical Ordering</td>
<td>411</td>
</tr>
<tr>
<td>A5 Heavy Tails</td>
<td>412</td>
</tr>
<tr>
<td>A6 Geometric Trials</td>
<td>412</td>
</tr>
<tr>
<td>A7 Semigroups of Positive Numbers</td>
<td>413</td>
</tr>
<tr>
<td>A8 Total Variation Convergence</td>
<td>413</td>
</tr>
<tr>
<td>A9 Transforms</td>
<td>414</td>
</tr>
<tr>
<td>A10 Stopping Times and Wald’s Identity</td>
<td>414</td>
</tr>
<tr>
<td>A11 Discrete Skeletons</td>
<td>415</td>
</tr>
<tr>
<td>Bibliography</td>
<td>416</td>
</tr>
<tr>
<td>Index</td>
<td>431</td>
</tr>
</tbody>
</table>
The basic principle for references within the book is to specify the chapter number only when it is not the current one. Thus, say, Proposition 1.3, formula (2.7) or Section 5 of Chapter IV are referred to as IV.1.3, IV.(2.7) and IV.5, respectively, in all chapters other than IV where we write Proposition 1.3, (2.7) and Section 5.

Symbols such as say \( A, \eta \), etc. do not of course have the same meaning throughout the book and may be used interchangeably for real numbers, measures and so on. For queueing processes, some effort has been made to make the notation (introduced in III.1) reasonably consistent throughout the book. One inconvenience is that the associated random walk becomes \( S_n = X_0 + \cdots + X_{n-1} \) and not \( X_1 + \cdots + X_n \) as in Chapter VIII. Of course, similar (hopefully minor) incidents occur at a number of other places.

The expression \( \mathbb{E}[X; A] \) means \( \mathbb{E}X I(A) \), where \( I(A) \) is the indicator of \( A \) (if say \( A = \{X > 0\} \), we write \( \mathbb{E}[X; X > 0] \)). By \( X \overset{D}{=} Y \) we mean equality in distribution and by \( X_n \overset{d}{\to} X \) convergence in distribution (weak convergence). The relation \( a_n \sim b_n \) means that \( a_n/b_n \to 1 \) as \( n \to \infty \) (other limits may also occur), whereas \( a_n \approx b_n \) indicates various different types of asymptotics, often just at the heuristical level. We use occasionally \( \lim \) instead of \( \lim \sup \), and similarly for \( \lim \), \( \lim \inf \). Ends of proofs, examples or remarks are marked by the symbol \( \Box \).

The typeface \( \mathbb{P}, \mathbb{E} \) is used for probability and expectation; \( \mathbb{P}_e, \mathbb{E}_e \) have a special meaning by referring to stationarity (equilibrium or steady state, cf. III.1). Matrices and vectors are in boldface \( A, \mathbf{t}, \pi, \) etc.; usually, matrices have uppercase Roman letters (occasionally Greek), column vectors lower-
case Roman letters and row vectors lowercase Greek letters. The column vector with all entries equal to 1 is denoted \( \mathbf{1} \), the \( i \)th unit vector \( \mathbf{1}_i \). The transpose of \( \mathbf{A} \) is written \( \mathbf{A}^T \).

The standard sets are denoted as follows:

- \( \mathbb{N} = \{0, 1, 2, \ldots\} \) the natural numbers
- \( \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\} \) the integers
- \( \mathbb{Q} = \{p/q : p \in \mathbb{Z}, q \in \mathbb{N} \setminus \{0\}\} \) the rationals
- \( \mathbb{R} = (-\infty, \infty) \) the real numbers
- \( \mathbb{C} = \{x + iy : x, y \in \mathbb{R}\} \) the complex numbers

(no special notation like \( \mathbb{R}_+ \) is used for \((0, \infty)\) or \([0, \infty)\)). The index set for the time parameter of a stochastic process, usually \( \mathbb{N}, \mathbb{Z}, [0, \infty) \) or \((-\infty, \infty)\), is denoted by \( \mathbb{T} \) if more than one possibility may occur.

The set \( D \) of functions \( \{x_t\} \) which are right–continuous \((x_s \to x_t, s \downarrow t)\) and have left–hand limits \( x_{t-} = \lim_{s \uparrow t} x_s \) is frequently encountered. If, say, \( t \) varies in \([0, 1]\) and \( x_t \) is \( E \)-valued, we may specify this by writing \( D([0, 1], E) \). Most often \( D \) stands for \( D[0, \infty) = D([0, \infty), \mathbb{R}) \). \( D_0 \) is the set of \( D \)-functions with finite lifelength; see A2.

Some main abbreviations are given in the following list (others occur locally):

- LLN law of large numbers
- CLT central limit theorem
- LIL law of the iterated logarithm
- l.h.s. left–hand side
- r.h.s. right–hand side
- a.s. almost surely
- i.i.d. independent identically distributed
- i.o. infinitely often
- r.v. random variable
- t.v. total variation
- w.l.o.g. without loss of generality
- w.p. with probability
- w.r.t. with respect to
- d.R.i. directly Riemann integrable
- ch.f. characteristic function
- m.g.f. moment generating function
- c.g.f. cumulant generating function
- g.c.d. greatest common divisor
- supp support
- spr spectral radius

The notation \( \hat{F} \) for the transform of a probability distribution may denote either of the probability generating function, the m.g.f. or the ch.f.; see A9.

The delta function is \( \delta_{ij} = I(i = j) \), whereas \( \delta_x \) often denotes the measure degenerate at \( x \).
Part A:
Simple Markovian Models
I
Markov Chains

1 Preliminaries

We consider a Markov chain $X_0, X_1, \ldots$ with discrete (i.e. finite or countable) state space $E = \{i, j, k, \ldots\}$ and specified by the transition matrix $P = (p_{ij})_{i,j \in E}$. By this we mean that $P$ is a given $E \times E$ matrix such that $p_i \cdot = (p_{ij})_{j \in E}$ is a probability (vector) for each $i$, and that we study $\{X_n\}$ subject to exactly those governing probability laws $\mathbb{P} = \mathbb{P}_{\mu}$ (Markov probabilities) for which

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \mu_i p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} \quad (1.1)$$

where $\mu_i = \mathbb{P}(X_0 = i)$. The particular value of the initial distribution $\mu$ is unimportant in most cases and is therefore suppressed in the notation. An important exception is the case where $X_0$ is degenerate, say at $i$, and we write then $\mathbb{P}_i$ so that $\mathbb{P}_i(X_0 = i) = 1$.

Given $\mu$, it is readily checked that (1.1) uniquely determines a probability distribution on $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. Appealing to basic facts from the foundational theory of Markov processes (to be discussed in Section 8), this set of probabilities can be uniquely extended to a probability law $\mathbb{P}_{\mu}$ governing the whole chain. Thus, since the transition matrix $P$ is fixed here and in the following, the Markov probabilities are in one–to–one correspondence with the set of initial distributions.

If $\mathbb{P}$ is a Markov probability, then (with the usual a.s. interpretation of conditional probabilities and expectations)

$$p_{ij} = \mathbb{P}_i(X_1 = j) = \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad (1.2)$$
\begin{align}
P(X_{n+1} = j \mid \mathcal{F}_n) &= p_{X_nj} = \mathbb{P}_{X_n}(X_1 = j), \quad (1.3) \\
\mathbb{E}[h(X_n, X_{n+1}, \ldots) \mid \mathcal{F}_n] &= \mathbb{E}_{X_n} h(X_0, X_1, \ldots). \quad (1.4)
\end{align}

Conversely\(^1\), either (1.3) or (1.4) is sufficient for \(\mathbb{P}\) to be a Markov probability. The formal proof of these facts is an easy (though in part lengthy) exercise in conditioning arguments and will not be given here. However, equations (1.2), (1.3), (1.4) have important intuitive contents. Thus (1.4) means that at time \(n\), the chain is restarted with the new initial value \(X_n\). Equivalently, the post-\(n\)-chain \(X_n, X_{n+1}, \ldots\) evolves as the Markov chain itself, started at \(X_n\) but otherwise independent of the past. Similarly, in simulation terminology (1.3) means that the chain can be stepwise constructed by at step \(n\) drawing \(X_{n+1}\) according to \(p_{X_n}\). (to get started, draw \(X_0\) according to \(\mu\)).

Recall from A10 (the Appendix) that a \textit{stopping time} \(\sigma\) is a r.v. with values in \(N \cup \{\infty\}\) and satisfying \(\{\sigma = n\} \in \mathcal{F}_n\) for all \(n\), that \(\mathcal{F}_\sigma\) denotes the \(\sigma\)-algebra which consists of all disjoint unions of the form \(\bigcup_0^\infty A_n\) with \(A_n \in \mathcal{F}_n\), \(A_n \subseteq \{\sigma = n\}\) (here \(n = \infty\) is included with the convention \(\mathcal{F}_\infty = \sigma(X_0, X_1, \ldots)\)), and that \(\sigma\) and \(X_\sigma\) are measurable w.r.t. to \(\mathcal{F}_\sigma\). The important \textit{strong Markov property} states that for the sake of predicting the future development of the chain a stopping time may be treated as a fixed deterministic point of time. For example, we have the following extension of (1.4):

\textbf{Theorem 1.1 \textit{(Strong Markov Property)}} \ Let \(\sigma\) be a stopping time. Then a.s. on \(\{\sigma < \infty\}\) it holds that

\[ \mathbb{E}[h(X_\sigma, X_{\sigma+1}, \ldots) \mid \mathcal{F}_\sigma] = \mathbb{E}_{X_\sigma} h(X_0, X_1, \ldots). \quad (1.5) \]

\textbf{Proof.} We must show that for \(A \in \mathcal{F}_\sigma\), \(A \subseteq \{\sigma < \infty\}\) we have

\[ \mathbb{E}[h(X_\sigma, X_{\sigma+1}, \ldots) \mid A] = \mathbb{E}[\mathbb{E}_{X_\sigma} h(X_0, X_1, \ldots) \mid A]. \]

However, if \(A \in \mathcal{F}_n\) and \(\sigma = n\) on \(A\), this is immediate from (1.4). Replace \(A\) by \(A \cap \{\sigma = n\}\) and sum over \(n\). \(\Box\)

The \(m\)th power (iterate) of the transition matrix is denoted by \(P^m = (p_{ij}^m)\). An easy calculation (e.g. let \(n = nm\) in (1.1) and sum over the \(i_k\) with \(k \not\in \{0, m, \ldots, nm\}\)) shows that \(X_0, X_m, X_{2m}, \ldots\) is a Markov chain and that its transition matrix is simply \(P^m\).

Associated with each state is the hitting time

\[ \tau(i) = \inf \{n \geq 1 : X_n = i\} \]

(with the usual convention \(\tau(i) = \infty\) if no such \(n\) exists) and the number of visits \(N_i = \sum_1^\infty I(X_n = i)\) to \(i\). Clearly, \(\{\tau(i) < \infty\} = \{N_i > 0\}\) and we

\(^1\)The meaning of (1.4) is that this should hold for any \(h : E \times E \times \cdots \to \mathbb{R}\) for which (1.4) makes sense, say \(h\) is bounded or nonnegative; similarly, (1.5) should hold for all \(n\) and \(j\). In (1.3), \(\mathbb{P}_{X_n}(X_1 = j)\) means \(g(x) = \mathbb{P}_x(X_1 = j)\) evaluated at \(x = X_n\).
call $i$ recurrent if the recurrence time distribution $\mathbb{P}_i(\tau(i) = k)$ is proper, i.e. if $\mathbb{P}_i(\tau(i) < \infty) = 1$, and transient otherwise. The chain itself is recurrent (transient) if all states are so.

**Proposition 1.2** Let $i$ be some fixed state. Then:

(i) The following assertions (a), (b), (c) are equivalent: (a) $i$ is recurrent;  
(b) $N_i = \infty \mathbb{P}_i$-a.s.; (c) $\mathbb{E}_i N_i = \sum_1^\infty p_{ii}^m = \infty$;

(ii) the following assertions (a$'$), (b$'$), (c$'$) are equivalent as well: (a$'$) $i$ is transient; (b$'$) $N_i < \infty \mathbb{P}_i$-a.s.; (c$'$) $\mathbb{E}_i N_i = \sum_1^\infty p_{ii}^m < \infty$.

**Proof.** Define $\tau(i; 1) = \tau(i)$,

$$
\tau(i; k+1) = \inf\{n > \tau(i; k) \colon X_n = i\}, \quad \theta = \mathbb{P}_i(\tau(i; 1) < \infty).$

Then $N_i$ is simply the number of $k$ with $\tau(i; k) < \infty$, and by the strong Markov property and $X_{\tau(i; k)} = i$,

$$
\mathbb{P}_i(\tau(i; k+1) < \infty) = \mathbb{E}_i \mathbb{P}(\tau(i; k+1) < \infty, \tau(i; k) < \infty \mid \mathcal{F}_{\tau(i; k)})
= \mathbb{E}_i[\mathbb{P}(\tau(i; k+1) < \infty \mid \mathcal{F}_{\tau(i; k)}); \tau(i; k) < \infty]
= \mathbb{E}_i[\mathbb{P}_{X_{\tau(i; k)}}(\tau(i; 1) < \infty); \tau(i; k) < \infty]
= \theta \mathbb{P}_i(\tau(i, k) < \infty) = \cdots = \theta^{k+1}. \tag{1.6}
$$

If (a) holds, then $\theta = 1$ so that it follows that all $\tau(i; k) < \infty \mathbb{P}_i$-a.s., and (b) also holds. Clearly, (b$'$)$\Rightarrow$(c$'$) so that for part (i) it remains to prove (c$'$)$\Rightarrow$(a$'$) or equivalently (a$'$)$\Rightarrow$(c$'$). But if $\theta < 1$, then

$$
\mathbb{E}_i N_i = \sum_{k=0}^\infty \mathbb{P}_i(N_i > k) = \sum_{k=1}^\infty \mathbb{P}_i(\tau(i; k) < \infty) = \sum_{k=1}^\infty \theta^k < \infty.
$$

For part (ii), it follows by negation that (a$'$)$\iff$(c$'$)$\iff$(b$''$)$ \mathbb{P}_i(N_i < \infty) > 0$. However, clearly (b$'$)$\Rightarrow$(b$''$) and from (1.6) it is seen that if (b$''$) holds, then $\theta < 1$. Thus (b$''$) $\Rightarrow$ (a$'$). \hfill \square

It should be noted that though Proposition 1.2 gives necessary and sufficient conditions for recurrence/transience, the criteria are almost always difficult to check: even for extremely simple transition matrices $P$, it is usually impossible to find closed expressions for the $p_{ii}^m$. Some alternative general approaches are discussed in Section 5, but in many cases the recurrence/transience classification leads into arguments particular for the specific model.

Our emphasis in the following is on the recurrent case and we shall briefly discuss some aspects of the set–up. Two states $i, j$ are said to communicate, written $i \leftrightarrow j$, if $i$ can be reached from $j$ (i.e. $p_{ji}^m > 0$ for some $m$) and vice versa. Clearly, the relation is transitive and symmetric. Now suppose $i$ is recurrent and that $j$ can be reached from $i$. Then also $i$ can be reached from $j$. In fact even $\tau(i) < \infty \mathbb{P}_j$-a.s. since otherwise $\mathbb{P}_i(\tau(i) = \infty) > 0$.  

1. Preliminaries 5
Furthermore, \( j \) is recurrent since
\[
\sum_{m=1}^{\infty} p_{jj}^{m} \geq \sum_{m=1}^{\infty} p_{ji}^{m1} p_{ii}^{m2} p_{ij}^{m2} = \infty
\]
if \( m_1, m_2 \) are chosen with \( p_{ji}^{m1} > 0, p_{ij}^{m2} > 0 \). Obviously \( i \leftrightarrow i \) by recurrence, and it follows that \( \leftrightarrow \) is an equivalence relation on the recurrent states so that we may write
\[
E = T \cup R_1 \cup R_2 \cdots ,
\]
where \( R_1, R_2, \ldots \) are the equivalence classes (recurrent classes) and \( T \) the set of transient states. It is basic to note that the recurrent classes are closed (or absorbing), i.e.
\[
\mathbb{P}_i(X_n \in R_k \text{ for all } n) = 1 \quad \text{when } i \in R_k
\]
(this follows from the above characterization of \( R_k \) as the set of all states that can be reached from \( i \)). When started at \( i \in R_k \) the chain therefore evolves within \( R_k \) only, and the state space may be reduced to \( R_k \). If, on the other hand, \( X_0 = i \) is transient, two types of paths may occur: either \( X_n \in T \) for all \( n \) or at some stage the chain enters a recurrent class \( R_k \) and is absorbed, i.e. evolves from then on in \( R_k \).

Most often one can restrict attention to irreducible chains, defined by the requirement that all states in \( E \) communicate. Such a chain is either transient or \( E \) consists of exactly one recurrent class. In fact, if a recurrent state, say \( i \), exists at all, it follows from the above that any other state \( j \) is in the same recurrence class as \( i \).

A recurrent state is called positive recurrent if the mean recurrence time \( \mathbb{E}_i \tau(i) \) is finite. Otherwise \( i \) is null recurrent. The period \( d = d(i) \) is the period of the recurrence–time distribution, i.e. the greatest integer \( d \) such that \( \mathbb{P}_i(\tau(i) \in L_d) = 1 \) where \( L_d = \{d, 2d, 3d, \ldots \} \). If \( d = 1 \), \( i \) is aperiodic.

**Proposition 1.3** Let \( R \) be a recurrent class. Then the states in \( R \) (i) are either all positive recurrent or all null recurrent; (ii) have all the same period.

**Proof.** (i) is deferred to Section 3. Let \( i, j \in R \) and choose \( r, s \) with \( p_{ij}^r > 0, p_{ji}^s > 0 \). Then \( p_{ii}^{r+s} > 0 \), i.e. \( r+s \in L_{d(i)} \), and whenever \( p_{jj}^n > 0, p_{ii}^{r+s+n} > 0 \) also, i.e. \( r+s+n \in L_{d(i)} \) so that \( n \in L_{d(i)} \) also. It follows that \( \mathbb{P}_j(\tau(j) \in L_{d(i)}) = 1 \), i.e. \( d(j) \geq d(i) \). By symmetry, \( d(i) \geq d(j) \). \( \square \)

**Proposition 1.4** Let \( i \) be aperiodic and recurrent. Then: (a) there exists \( n_i \) such that \( p_{ii}^m > 0 \) for all \( m \geq n_i \); (b) if \( j \) can be reached from \( i \), then there exists \( n_j \) such that \( p_{ij}^m > 0 \) for all \( m \geq n_j \).

**Proof.** For (a), see A7.1(a). For (b), choose \( k_j \) with \( p_{ij}^{k_j} > 0 \) and let \( n_j = n_i + k_j \). \( \square \)