

Preface

This text has evolved over some 20 years, starting as lecture notes for two first-year graduate subjects at M.I.T., namely, *Discrete Stochastic Processes* (6.262) and *Random Processes, Detection, and Estimation* (6.432). The two sets of notes are closely related and have been integrated into one text. Instructors and students can pick and choose the topics that meet their needs, and suggestions for doing this follow this preface.

These subjects originally had an application emphasis, the first on queueing and congestion in data networks and the second on modulation and detection of signals in the presence of noise. As the notes have evolved, it has become increasingly clear that the mathematical development (with minor enhancements) is applicable to a much broader set of applications in engineering, operations research, physics, biology, economics, finance, statistics, etc.

The field of stochastic processes is essentially a branch of probability theory, treating probabilistic models that evolve in time. It is best viewed as a branch of mathematics, starting with the axioms of probability and containing a rich and fascinating set of results following from those axioms. Although the results are applicable to many areas, they are best understood initially in terms of their mathematical structure and interrelationships.

Applying axiomatic probability results to a real-world area requires creating a probability model for the given area. Mathematically precise results can then be derived within the model and translated back to the real world. If the model fits the area sufficiently well, real problems can be solved by analysis within the model. However, since models are almost always simplified approximations of reality, precise results within the model become approximations in the real world.

Choosing an appropriate probability model is an essential part of this process. Sometimes an application area will have customary choices of models, or at least structured ways of selecting them. For example, there is a well developed taxonomy of queueing models. A sound knowledge of the application area, combined with a sound knowledge of the behavior of these queueing models, often lets one choose a suitable model for a given issue within the application area. In other cases, one can start with a particularly simple model and use the behavior of that model to gain insight about the application, and use this to iteratively guide the selection of more general models.

An important aspect of choosing a probability model for a real-world area is that a prospective choice depends heavily on prior understanding, at both an intuitive and mathematical level, of results from the range of mathematical models that might be involved. This partly explains the title of the text — Theory for applications. The aim is to guide the reader in both the mathematical and intuitive understanding necessary in developing and using stochastic process models in studying application areas.

Application-oriented students often ask why it is important to understand axioms, theorems, and proofs in mathematical models when the precise results in the model become approximations in the real-world system being modeled. One answer is that a deeper understanding of the mathematics leads to the required intuition for understanding the differences between model and reality. Another answer is that theorems are transferable between applications, and thus enable insights from one application area to be transferred to another.

Given the need for precision in the theory, however, why is an axiomatic approach needed? Engineering and science students learn to use calculus, linear algebra and undergraduate probability effectively without axioms or rigor. Why doesn't this work for more advanced probability and stochastic processes?

Probability theory has more than its share of apparent paradoxes, and these show up in very elementary arguments. Undergraduates are content with this, since they can postpone these questions to later study. For the more complex issues in graduate work, however, reasoning without a foundation becomes increasingly frustrating, and the axioms provide the foundation needed for sound reasoning without paradoxes.

I have tried to avoid the concise and formal proofs of pure mathematics, and instead use explanations that are longer but more intuitive while still being precise. This is partly to help students with limited exposure to pure math, and partly because intuition is vital when going back and forth between a mathematical model and a real-world problem. In doing research, we grope toward results, and successful groping requires both a strong intuition and precise reasoning.

The text neither uses nor develops measure theory. Measure theory is undoubtedly important in understanding probability at a deep level, but most of the topics useful in many applications can be understood without measure theory. I believe that the level of precision here provides a good background for a later study of measure theory.

The text does require some background in probability at an undergraduate level. Chapter 1 presents this background material as review, but it is too concentrated and deep for most students without prior background. Some exposure to linear algebra and analysis (especially concrete topics like vectors, matrices, and limits) is helpful, but the text develops the necessary results. The most important prerequisite is the mathematical maturity and patience to couple precise reasoning with intuition.

The organization of the text, after the review in Chapter 1 is as follows: Chapters 2, 3, and 4 treat three of the simplest and most important classes of stochastic processes, first Poisson processes, next Gaussian processes, and finally finite-state Markov chains. These are beautiful processes where almost everything is known, and they contribute insights, examples, and initial approaches for almost all other processes. Chapter 5 then treats renewal processes, which generalize Poisson processes and provide the foundation for the rest of the text.

Chapters 6 and 7 use renewal theory to generalize Markov chains to countable state spaces and continuous time. Chapters 8 and 10 then study decision making and estimation, which in a sense gets us out of the world of theory and back to using the theory. Finally Chapter 9 treats random walks, large deviations, and martingales and illustrates many of their applications.

Most results here are quite old and well established, so I have not made any effort to attribute results to investigators. My treatment of the material is indebted to the texts by Bertsekas and Tsitsiklis [2], Sheldon Ross [22] and William Feller [8] and [9].

Contents

1	INTRODUCTION AND REVIEW OF PROBABILITY	1
1.1	Probability models	1
1.1.1	The sample space of a probability model	3
1.1.2	Assigning probabilities for finite sample spaces	4
1.2	The axioms of probability theory	5
1.2.1	Axioms for events	7
1.2.2	Axioms of probability	8
1.3	Probability review	9
1.3.1	Conditional probabilities and statistical independence	9
1.3.2	Repeated idealized experiments	11
1.3.3	Random variables	12
1.3.4	Multiple random variables and conditional probabilities	15
1.4	Stochastic processes	16
1.4.1	The Bernoulli process	17
1.5	Expectations and more probability review	20
1.5.1	Random variables as functions of other random variables	23
1.5.2	Conditional expectations	25
1.5.3	Typical values of rv's; mean and median	28
1.5.4	Indicator random variables	29
1.5.5	Moment generating functions and other transforms	30
1.6	Basic inequalities	32
1.6.1	The Markov inequality	32
1.6.2	The Chebyshev inequality	33

1.6.3	Chernoff bounds	33
1.7	The laws of large numbers	37
1.7.1	Weak law of large numbers with a finite variance	37
1.7.2	Relative frequency	40
1.7.3	The central limit theorem	40
1.7.4	Weak law with an infinite variance	45
1.7.5	Convergence of random variables	46
1.7.6	Convergence with probability 1	49
1.8	Relation of probability models to the real world	51
1.8.1	Relative frequencies in a probability model	52
1.8.2	Relative frequencies in the real world	53
1.8.3	Statistical independence of real-world experiments	55
1.8.4	Limitations of relative frequencies	56
1.8.5	Subjective probability	57
1.9	Summary	57
1.10	Exercises	59
2	POISSON PROCESSES	74
2.1	Introduction	74
2.1.1	Arrival processes	74
2.2	Definition and properties of a Poisson process	76
2.2.1	Memoryless property	77
2.2.2	Probability density of S_n and joint density of S_1, \dots, S_n	80
2.2.3	The PMF for $N(t)$	81
2.2.4	Alternate definitions of Poisson processes	82
2.2.5	The Poisson process as a limit of shrinking Bernoulli processes	84
2.3	Combining and splitting Poisson processes	86
2.3.1	Subdividing a Poisson process	88
2.3.2	Examples using independent Poisson processes	89
2.4	Non-homogeneous Poisson processes	91
2.5	Conditional arrival densities and order statistics	94
2.6	Summary	98
2.7	Exercises	100

3	GAUSSIAN RANDOM VECTORS AND PROCESSES	109
3.1	Introduction	109
3.2	Gaussian random variables	109
3.3	Gaussian random vectors	111
3.3.1	Generating functions of Gaussian random vectors	112
3.3.2	IID normalized Gaussian random vectors	112
3.3.3	Jointly-Gaussian random vectors	113
3.3.4	Joint probability density for Gaussian n -rv's (special case)	116
3.4	Properties of covariance matrices	118
3.4.1	Symmetric matrices	118
3.4.2	Positive definite matrices and covariance matrices	119
3.4.3	Joint probability density for Gaussian n -rv's (general case)	121
3.4.4	Geometry and principal axes for Gaussian densities	123
3.5	Conditional PDF's for Gaussian random vectors	125
3.6	Gaussian processes	129
3.6.1	Stationarity and related concepts:	131
3.6.2	Orthonormal expansions	132
3.6.3	Continuous-time Gaussian processes	135
3.6.4	Gaussian sinc processes	136
3.6.5	Filtered Gaussian sinc processes	139
3.6.6	Filtered continuous-time stochastic processes	141
3.6.7	Interpretation of spectral density and covariance	142
3.6.8	White Gaussian noise	144
3.6.9	The Wiener process / Brownian motion	146
3.7	Circularly-symmetric complex random vectors	149
3.7.1	Circular symmetry and complex Gaussian rv's	149
3.7.2	Covariance and pseudo-covariance of complex n -rv's	150
3.7.3	Covariance matrices of complex n -rv	152
3.7.4	Linear transformations of $\mathbf{W} \sim \mathcal{CN}(0, [I_\ell])$	153
3.7.5	Linear transformations of $\mathbf{Z} \sim \mathcal{CN}(0, [K])$	154
3.7.6	The PDF of circularly-symmetric Gaussian n -rv's	155

3.7.7	Conditional PDF's for circularly symmetric Gaussian rv's	157
3.7.8	Circularly-symmetric Gaussian processes	158
3.8	Summary	160
3.9	Exercises	161
4	FINITE-STATE MARKOV CHAINS	167
4.1	Introduction	167
4.2	Classification of states	169
4.3	The matrix representation	173
4.3.1	Steady state and $[P^n]$ for large n	174
4.3.2	Steady state assuming $[P] > 0$	177
4.3.3	Ergodic Markov chains	178
4.3.4	Ergodic unichains	180
4.3.5	Arbitrary finite-state Markov chains	182
4.4	The eigenvalues and eigenvectors of stochastic matrices	182
4.4.1	Eigenvalues and eigenvectors for $M = 2$ states	183
4.4.2	Eigenvalues and eigenvectors for $M > 2$ states	184
4.5	Markov chains with rewards	187
4.5.1	Expected first-passage times	187
4.5.2	The expected aggregate reward over multiple transitions	191
4.5.3	The expected aggregate reward with an additional final reward	194
4.6	Markov decision theory and dynamic programming	196
4.6.1	Dynamic programming algorithm	197
4.6.2	Optimal stationary policies	201
4.6.3	Policy improvement and the search for optimal stationary policies	204
4.7	Summary	208
4.8	Exercises	210
5	RENEWAL PROCESSES	223
5.1	Introduction	223
5.2	The strong law of large numbers and convergence WP1	226
5.2.1	Convergence with probability 1 (WP1)	226

5.2.2	Strong law of large numbers (SLLN)	228
5.3	Strong law for renewal processes	230
5.4	Renewal-reward processes; time averages	235
5.4.1	General renewal-reward processes	238
5.5	Stopping times for repeated experiments	242
5.5.1	Wald's equality	244
5.5.2	Applying Wald's equality to $E[N(t)]$	247
5.5.3	Generalized stopping trials, embedded renewals, and G/G/1 queues	247
5.5.4	Little's theorem	251
5.5.5	M/G/1 queues	254
5.6	Expected number of renewals; ensemble averages	257
5.6.1	Laplace transform approach	259
5.6.2	The elementary renewal theorem	260
5.7	Renewal-reward processes; ensemble averages	262
5.7.1	Age and duration for arithmetic processes	264
5.7.2	Joint age and duration: non-arithmetic case	267
5.7.3	Age $Z(t)$ for finite t : non-arithmetic case	269
5.7.4	Age $Z(t)$ as $t \rightarrow \infty$; non-arithmetic case	271
5.7.5	Arbitrary renewal-reward functions: non-arithmetic case	273
5.8	Delayed renewal processes	276
5.8.1	Delayed renewal-reward processes	278
5.8.2	Transient behavior of delayed renewal processes	278
5.8.3	The equilibrium process	279
5.9	Summary	280
5.10	Exercises	282
6	COUNTABLE-STATE MARKOV CHAINS	300
6.1	Introductory examples	300
6.2	First passage times and recurrent states	303
6.3	Renewal theory applied to Markov chains	307
6.3.1	Renewal theory and positive recurrence	308

6.3.2	Steady state	310
6.3.3	Blackwell's theorem applied to Markov chains	313
6.3.4	Age of a renewal process	314
6.4	Birth-death Markov chains	315
6.5	Reversible Markov chains	317
6.6	The M/M/1 sample-time Markov chain	320
6.7	Branching processes	323
6.8	Round-robin and processor sharing	326
6.9	Summary	331
6.10	Exercises	333
7	MARKOV PROCESSES WITH COUNTABLE STATE SPACES	338
7.1	Introduction	338
7.1.1	The sampled-time approximation to a Markov process	342
7.2	Steady-state behavior of irreducible Markov processes	343
7.2.1	Renewals on successive entries to a given state	345
7.2.2	The limiting fraction of time in each state	345
7.2.3	Finding $\{p_j(i); j \geq 0\}$ in terms of $\{\pi_j; j \geq 0\}$	347
7.2.4	Solving for the steady-state process probabilities directly	349
7.2.5	The sampled-time approximation again	350
7.2.6	Pathological cases	351
7.3	The Kolmogorov differential equations	351
7.4	Uniformization	355
7.5	Birth-death processes	356
7.5.1	The M/M/1 queue again	357
7.5.2	Other birth/death systems	358
7.6	Reversibility for Markov processes	358
7.7	Jackson networks	364
7.7.1	Closed Jackson networks	370
7.8	Semi-Markov processes	371
7.8.1	Example — the M/G/1 queue	375
7.9	Summary	376
7.10	Exercises	378

8	DETECTION, DECISIONS, AND HYPOTHESIS TESTING	391
8.1	Decision criteria and the MAP criterion	392
8.2	Binary MAP detection	395
8.2.1	Sufficient statistics I	397
8.2.2	Binary detection with a one-dimensional observation	398
8.2.3	Binary MAP detection with vector observations	402
8.2.4	Sufficient statistics II	407
8.3	Binary detection with a minimum-cost criterion	412
8.4	The error curve and the Neyman-Pearson rule	413
8.4.1	The Neyman-Pearson detection rule	419
8.4.2	The min-max detection rule	420
8.5	Finitely many hypotheses	420
8.5.1	Sufficient statistics with $M \geq 2$ hypotheses	423
8.5.2	More general min-cost tests	425
8.6	Summary	426
8.7	Exercises	428
9	RANDOM WALKS, LARGE DEVIATIONS, AND MARTINGALES	435
9.1	Introduction	435
9.1.1	Simple random walks	436
9.1.2	Integer-valued random walks	437
9.1.3	Renewal processes as special cases of random walks	437
9.2	The queuing delay in a G/G/1 queue:	438
9.3	Threshold crossing probabilities in random walks	441
9.3.1	The Chernoff bound	441
9.3.2	Tilted probabilities	443
9.3.3	Large deviations and compositions	446
9.3.4	Back to threshold crossings	449
9.4	Thresholds, stopping rules, and Wald's identity	451
9.4.1	Wald's identity for two thresholds	452
9.4.2	The relationship of Wald's identity to Wald's equality	453

9.4.3	Zero-mean random walks	454
9.4.4	Exponential bounds on the probability of threshold crossing	455
9.5	Binary hypotheses with IID observations	456
9.5.1	Binary hypotheses with a fixed number of observations	457
9.5.2	Sequential decisions for binary hypotheses	460
9.6	Martingales	462
9.6.1	Simple examples of martingales	463
9.6.2	Scaled branching processes	465
9.6.3	Partial isolation of past and future in martingales	465
9.7	Submartingales and supermartingales	466
9.8	Stopped processes and stopping trials	468
9.8.1	The Wald identity	471
9.9	The Kolmogorov inequalities	472
9.9.1	The strong law of large numbers (SLLN)	474
9.9.2	The martingale convergence theorem	475
9.10	A simple model for investments	477
9.10.1	Portfolios with constant fractional allocations	480
9.10.2	Portfolios with time-varying allocations	485
9.11	Markov modulated random walks	487
9.11.1	Generating functions for Markov random walks	489
9.11.2	Stopping trials for martingales relative to a process	490
9.11.3	Markov modulated random walks with thresholds	490
9.12	Summary	492
9.13	Exercises	494
10	ESTIMATION	510
10.1	Introduction	510
10.1.1	The squared-cost function	511
10.1.2	Other cost functions	512
10.2	MMSE estimation for Gaussian random vectors	514
10.2.1	Scalar iterative estimation	517

10.2.2	Scalar Kalman filter	518
10.3	Linear least-squares error estimation	520
10.4	Filtered vector signal plus noise	522
10.4.1	Estimate of a single rv in IID vector noise	524
10.4.2	Estimate of a single rv in arbitrary vector noise	524
10.4.3	Vector iterative estimation	525
10.4.4	Vector Kalman filter	526
10.5	Estimation for circularly-symmetric Gaussian rv's	528
10.6	The vector space of rv's; orthogonality	529
10.7	MAP estimation and sufficient statistics	535
10.8	Parameter estimation	536
10.8.1	Fisher information and the Cramer-Rao bound	539
10.8.2	Vector observations	541
10.8.3	Information	542
10.9	Summary	544
10.10	Exercises	547
A	Solutions to selected exercises	554