

This book provides an introduction to heavy-traffic stochastic-process limits for queues, with special emphasis on nonstandard scaling and nonstandard limit processes. In addition to limit processes related to Brownian motion, the book discusses limit processes related to stable Lévy motion and fractional Brownian motion, which appear with heavy-tailed probability distributions and strong dependence.

To set the stage, the first four chapters present an informal introduction to stochastic-process limits. These chapters convey the unifying power of stochastic-process limits without sinking into the mathematical details. Simulation is used to help build intuition. Throughout, the book emphasizes the applied significance as well as the underlying mathematics. Hence, the book should appeal to a wide audience, including researchers and graduate students in operations research, industrial engineering, electrical engineering, computer science, business, economics and statistics, as well as specialists in probability theory.

ISBN 978-0-38795-358-8



9 780387 953588

ISBN 0-387-95358-2

www.springer-ny.com

1	Experiencing Statistical Regularity	1
1.1	A Simple Game of Chance	1
1.1.1	Plotting Random Walks	2
1.1.2	When the Game is Fair	3
1.1.3	The Final Position	7
1.1.4	Making an Interesting Game	12
1.2	Stochastic-Process Limits	16
1.2.1	A Probability Model	16
1.2.2	Classical Probability Limits	20
1.2.3	Identifying the Limit Process	22
1.2.4	Limits for the Plots	25
1.3	Invariance Principles	27
1.3.1	The Range of Brownian Motion	28
1.3.2	Relaxing the IID Conditions	30
1.3.3	Different Step Distributions	31
1.4	The Exception Makes the Rule	34
1.4.1	Explaining the Irregularity	35
1.4.2	The Centered Random Walk with $p = 3/2$	37
1.4.3	Back to the Uncentered Random Walk with $p = 1/2$	43
1.5	Summary	45

2	Random Walks in Applications	49
2.1	Stock Prices	49
2.2	The Kolmogorov-Smirnov Statistic	51
2.3	A Queueing Model for a Buffer in a Switch	55
2.3.1	Deriving the Proper Scaling	56
2.3.2	Simulation Examples	59
2.4	Engineering Significance	63
2.4.1	Buffer Sizing	64
2.4.2	Scheduling Service for Multiple Sources	68
3	The Framework for Stochastic-Process Limits	75
3.1	Introduction	75
3.2	The Space $\mathcal{P}$	76
3.3	The Space D	78
3.4	The Continuous-Mapping Approach	84
3.5	Useful Functions	86
3.6	Organization of the Book	89
4	A Panorama of Stochastic-Process Limits	95
4.1	Introduction	95
4.2	Self-Similar Processes	96
4.2.1	General CLT's and FCLT's	96
4.2.2	Self-Similarity	97
4.2.3	The Noah and Joseph Effects	99
4.3	Donsker's Theorem	101
4.3.1	The Basic Theorems	101
4.3.2	Multidimensional Versions	104
4.4	Brownian Limits with Weak Dependence	106
4.5	The Noah Effect: Heavy Tails	109
4.5.1	Stable Laws	111
4.5.2	Convergence to Stable Laws	114
4.5.3	Convergence to Stable Lévy Motion	116
4.5.4	Extreme-Value Limits	118
4.6	The Joseph Effect: Strong Dependence	120
4.6.1	Strong Positive Dependence	121
4.6.2	Additional Structure	122
4.6.3	Convergence to Fractional Brownian Motion	124
4.7	Heavy Tails Plus Dependence	130
4.7.1	Additional Structure	130
4.7.2	Convergence to Stable Lévy Motion	131
4.7.3	Linear Fractional Stable Motion	132
4.8	Summary	136

5	Heavy-Traffic Limits for Fluid Queues	137
5.1	Introduction	137
5.2	A General Fluid-Queue Model	139
5.2.1	Input and Available-Processing Processes	139
5.2.2	Infinite Capacity	140
5.2.3	Finite Capacity	143
5.3	Unstable Queues	145
5.3.1	Fluid Limits for Fluid Queues	146
5.3.2	Stochastic Refinements	149
5.4	Heavy-Traffic Limits for Stable Queues	153
5.5	Heavy-Traffic Scaling	157
5.5.1	The Impact of Scaling Upon Performance	158
5.5.2	Identifying Appropriate Scaling Functions	160
5.6	Limits as the System Size Increases	162
5.7	Brownian Approximations	165
5.7.1	The Brownian Limit	166
5.7.2	The Steady-State Distribution.	167
5.7.3	The Overflow Process	170
5.7.4	One-Sided Reflection	173
5.7.5	First-Passage Times	176
5.8	Planning Queueing Simulations	178
5.8.1	The Standard Statistical Procedure	180
5.8.2	Invoking the Brownian Approximation	181
5.9	Heavy-Traffic Limits for Other Processes	183
5.9.1	The Departure Process	183
5.9.2	The Processing Time	184
5.10	Priorities	187
5.10.1	A Heirarchical Approach	189
5.10.2	Processing Times	190
6	Unmatched Jumps in the Limit Process	193
6.1	Introduction	193
6.2	Linearly Interpolated Random Walks	195
6.2.1	Asymptotic Equivalence with M_1	195
6.2.2	Simulation Examples	196
6.3	Heavy-Tailed Renewal Processes	200
6.3.1	Inverse Processes	201
6.3.2	The Special Case with $m = 1$	202
6.4	A Queue with Heavy-Tailed Distributions	205
6.4.1	The Standard Single-Server Queue	206
6.4.2	Heavy-Traffic Limits	208
6.4.3	Simulation Examples	210
6.5	Rare Long Service Interruptions	216
6.6	Time-Dependent Arrival Rates	220

7	More Stochastic-Process Limits	225
7.1	Introduction	225
7.2	Central Limit Theorem for Processes	226
7.2.1	Hahn's Theorem	226
7.2.2	A Second Limit	230
7.3	Counting Processes	233
7.3.1	CLT Equivalence	234
7.3.2	FCLT Equivalence	235
7.4	Renewal-Reward Processes	238
8	Fluid Queues with On-Off Sources	243
8.1	Introduction	243
8.2	A Fluid Queue Fed by On-Off Sources	245
8.2.1	The On-Off Source Model	245
8.2.2	Simulation Examples	248
8.3	Heavy-Traffic Limits for the On-Off Sources	250
8.3.1	A Single Source	251
8.3.2	Multiple Sources	255
8.3.3	$M/G/\infty$ Sources	259
8.4	Brownian Approximations	260
8.4.1	The Brownian Limit	260
8.4.2	Model Simplification	263
8.5	Stable-Lévy Approximations	264
8.5.1	The RSLM Heavy-Traffic Limit	265
8.5.2	The Steady-State Distribution	268
8.5.3	Numerical Comparisons	270
8.6	Second Stochastic-Process Limits	272
8.6.1	$M/G/1/K$ Approximations	273
8.6.2	Limits for Limit Processes	277
8.7	Reflected Fractional Brownian Motion	279
8.7.1	An Increasing Number of Sources	279
8.7.2	Gaussian Input	280
8.8	Reflected Gaussian Processes	283
9	Single-Server Queues	287
9.1	Introduction	287
9.2	The Standard Single-Server Queue	288
9.3	Heavy-Traffic Limits	292
9.3.1	The Scaled Processes	292
9.3.2	Discrete-Time Processes	294
9.3.3	Continuous-Time Processes	297
9.4	Superposition Arrival Processes	301
9.5	Split Processes	305
9.6	Brownian Approximations	306
9.6.1	Variability Parameters	307

9.6.2	Models with More Structure	310
9.7	Very Heavy Tails	313
9.7.1	Heavy-Traffic Limits	314
9.7.2	First Passage to High Levels	316
9.8	An Increasing Number of Arrival Processes	318
9.8.1	Iterated and Double Limits	318
9.8.2	Separation of Time Scales	322
9.9	Approximations for Queueing Networks	326
9.9.1	Parametric-Decomposition Approximations	326
9.9.2	Approximately Characterizing Arrival Processes	330
9.9.3	A Network Calculus	331
9.9.4	Exogenous Arrival Processes	337
9.9.5	Concluding Remarks	338
10	Multiserver Queues	341
10.1	Introduction	341
10.2	Queues with Multiple Servers	342
10.2.1	A Queue with Autonomous Service	342
10.2.2	The Standard m -Server Model	345
10.3	Infinitely Many Servers	348
10.3.1	Heavy-Traffic Limits	349
10.3.2	Gaussian Approximations	352
10.4	An Increasing Number of Servers	355
10.4.1	Infinite-Server Approximations	356
10.4.2	Heavy-Traffic Limits for Delay Models	357
10.4.3	Heavy-Traffic Limits for Loss Models	360
10.4.4	Planning Simulations of Loss Models	361
11	More on the Mathematical Framework	367
11.1	Introduction	367
11.2	Topologies	368
11.2.1	Definitions	368
11.2.2	Separability and Completeness	371
11.3	The Space $\mathcal{P}$	372
11.3.1	Probability Spaces	372
11.3.2	Characterizing Weak Convergence	373
11.3.3	Random Elements	375
11.4	Product Spaces	377
11.5	The Space D	380
11.5.1	J_2 and M_2 Metrics	381
11.5.2	The Four Skorohod Topologies	382
11.5.3	Measurability Issues	385
11.6	The Compactness Approach	386

12 The Space D	391
12.1 Introduction	391
12.2 Regularity Properties of D	392
12.3 Strong and Weak M_1 Topologies	394
12.3.1 Definitions	394
12.3.2 Metric Properties	396
12.3.3 Properties of Parametric Representations	398
12.4 Local Uniform Convergence at Continuity Points	401
12.5 Alternative Characterizations of M_1 Convergence	403
12.5.1 SM_1 Convergence	403
12.5.2 WM_1 Convergence	408
12.6 Strengthening the Mode of Convergence	409
12.7 Characterizing Convergence with Mappings	410
12.8 Topological Completeness	413
12.9 Noncompact Domains	414
12.10 Strong and Weak M_2 Topologies	416
12.11 Alternative Characterizations of M_2 Convergence	418
12.11.1 M_2 Parametric Representations	418
12.11.2 SM_2 Convergence	419
12.11.3 WM_2 Convergence	421
12.11.4 Additional Properties of M_2 Convergence	422
12.12 Compactness	424
13 Useful Functions	427
13.1 Introduction	427
13.2 Composition	428
13.3 Composition with Centering	431
13.4 Supremum	435
13.5 One-Dimensional Reflection	439
13.6 Inverse	441
13.6.1 The Standard Topologies	442
13.6.2 The M'_1 Topology	444
13.6.3 First Passage Times	446
13.7 Inverse with Centering	447
13.8 Counting Functions	453
14 Queueing Networks	457
14.1 Introduction	457
14.2 The Multidimensional Reflection Map	460
14.2.1 A Special Case	460
14.2.2 Definition and Characterization	462
14.2.3 Continuity and Lipschitz Properties	465
14.3 The Instantaneous Reflection Map	473
14.3.1 Definition and Characterization	474
14.3.2 Implications for the Reflection Map	480

14.4	Reflections of Parametric Representations	482
14.5	M_1 Continuity Results and Counterexamples	485
14.5.1	M_1 Continuity Results	485
14.5.2	Counterexamples	487
14.6	Limits for Stochastic Fluid Networks	490
14.6.1	Model Continuity	492
14.6.2	Heavy-Traffic Limits	493
14.7	Queueing Networks with Service Interruptions	495
14.7.1	Model Definition	495
14.7.2	Heavy-Traffic Limits	499
14.8	The Two-Sided Regulator	505
14.8.1	Definition and Basic Properties	505
14.8.2	With the M_1 Topologies	509
14.9	Related Literature	511
15	The Spaces E and F	515
15.1	Introduction	515
15.2	Three Time Scales	516
15.3	More Complicated Oscillations	519
15.4	The Space E	523
15.5	Characterizations of M_2 Convergence in E	527
15.6	Convergence to Extremal Processes	530
15.7	The Space F	533
15.8	Queueing Applications	535
	References	541
	Appendix A Regular Variation	569
	Appendix B Contents of the Internet Supplement	573
	Notation Index	577
	Author Index	579
	Subject Index	585