

Contents

Preface	<i>page</i> xiii
Notation	xv
Part I Bandits, Probability and Concentration	1
1 Introduction	3
1.1 The Language of Bandits	4
1.2 Applications	7
1.3 Notes	10
1.4 Bibliographic Remarks	10
2 Foundations of Probability (👉)	12
2.1 Probability Spaces and Random Elements	12
2.2 σ -Algebras and Knowledge	19
2.3 Conditional Probabilities	20
2.4 Independence	21
2.5 Integration and Expectation	23
2.6 Conditional Expectation	25
2.7 Notes	29
2.8 Bibliographic Remarks	32
2.9 Exercises	33
3 Stochastic Processes and Markov Chains (👉)	36
3.1 Stochastic Processes	37
3.2 Markov Chains	37
3.3 Martingales and Stopping Times	39
3.4 Notes	41
3.5 Bibliographic Remarks	42
3.6 Exercises	43
4 Stochastic Bandits	45
4.1 Core Assumptions	45
4.2 The Learning Objective	45
4.3 Knowledge and Environment Classes	46

	4.4 The Regret	48
	4.5 Decomposing the Regret	50
	4.6 The Canonical Bandit Model (↗)	51
	4.7 The Canonical Bandit Model for Uncountable Action Sets (↗)	53
	4.8 Notes	54
	4.9 Bibliographical Remarks	55
	4.10 Exercises	56
5	Concentration of Measure	60
	5.1 Tail Probabilities	60
	5.2 The Inequalities of Markov and Chebyshev	61
	5.3 The Cramér-Chernoff Method and Subgaussian Random Variables	62
	5.4 Notes	64
	5.5 Bibliographical Remarks	66
	5.6 Exercises	66
Part II	Stochastic Bandits with Finitely Many Arms	73
6	The Explore-Then-Commit Algorithm	75
	6.1 Algorithm and Regret Analysis	75
	6.2 Notes	78
	6.3 Bibliographical Remarks	78
	6.4 Exercises	79
7	The Upper Confidence Bound Algorithm	84
	7.1 The Optimism Principle	84
	7.2 Notes	91
	7.3 Bibliographical Remarks	92
	7.4 Exercises	92
8	The Upper Confidence Bound Algorithm: Asymptotic Optimality	97
	8.1 Asymptotically Optimal UCB	97
	8.2 Notes	100
	8.3 Bibliographic Remarks	100
	8.4 Exercises	101
9	The Upper Confidence Bound Algorithm: Minimax Optimality (↗)	103
	9.1 The MOSS Algorithm	103
	9.2 Two Problems	106
	9.3 Notes	108
	9.4 Bibliographic Remarks	108
	9.5 Exercises	109

10	The Upper Confidence Bound Algorithm: Bernoulli Noise (👁)	112
	10.1 Concentration for Sums of Bernoulli Random Variables	112
	10.2 The KL-UCB Algorithm	115
	10.3 Notes	118
	10.4 Bibliographic Remarks	119
	10.5 Exercises	119
 Part III Adversarial Bandits with Finitely Many Arms		123
11	The Exp3 Algorithm	127
	11.1 Adversarial Bandit Environments	127
	11.2 Importance-Weighted Estimators	129
	11.3 The Exp3 Algorithm	131
	11.4 Regret Analysis	131
	11.5 Notes	135
	11.6 Bibliographic Remarks	137
	11.7 Exercises	138
12	The Exp3-IX Algorithm	142
	12.1 The Exp3-IX Algorithm	142
	12.2 Regret Analysis	144
	12.3 Notes	148
	12.4 Bibliographic Remarks	149
	12.5 Exercises	149
 Part IV Lower Bounds for Bandits with Finitely Many Arms		153
13	Lower Bounds: Basic Ideas	155
	13.1 Main Ideas Underlying Minimax Lower Bounds	155
	13.2 Notes	158
	13.3 Bibliographic Remarks	158
	13.4 Exercises	159
14	Foundations of Information Theory (👁)	160
	14.1 Entropy and Optimal Coding	160
	14.2 Relative Entropy	162
	14.3 Notes	165
	14.4 Bibliographic Remarks	167
	14.5 Exercises	167
15	Minimax Lower Bounds	170
	15.1 Relative Entropy Between Bandits	170
	15.2 Minimax Lower Bounds	171

	15.3 Notes	173
	15.4 Bibliographic Remarks	174
	15.5 Exercises	174
16	Instance-Dependent Lower Bounds	177
	16.1 Asymptotic Bounds	177
	16.2 Finite-Time Bounds	180
	16.3 Notes	181
	16.4 Bibliographic Remarks	181
	16.5 Exercises	181
17	High-Probability Lower Bounds	185
	17.1 Stochastic Bandits	186
	17.2 Adversarial Bandits	188
	17.3 Notes	190
	17.4 Bibliographic Remarks	190
	17.5 Exercises	190
Part V	Contextual and Linear Bandits	191
18	Contextual Bandits	193
	18.1 Contextual Bandits: One Bandit per Context	193
	18.2 Bandits with Expert Advice	195
	18.3 Exp4	197
	18.4 Regret Analysis	198
	18.5 Notes	200
	18.6 Bibliographic Remarks	201
	18.7 Exercises	202
19	Stochastic Linear Bandits	205
	19.1 Stochastic Contextual Bandits	205
	19.2 Stochastic Linear Bandits	207
	19.3 Regret Analysis	209
	19.4 Notes	211
	19.5 Bibliographic Remarks	213
	19.6 Exercises	214
20	Confidence Bounds for Least Squares Estimators	219
	20.1 Martingales and the Method of Mixtures	221
	20.2 Notes	225
	20.3 Bibliographic Remarks	226
	20.4 Exercises	227

21	Optimal Design for Least Squares Estimators	231
	21.1 The Kiefer–Wolfowitz Theorem	231
	21.2 Notes	233
	21.3 Bibliographic Remarks	235
	21.4 Exercises	235
22	Stochastic Linear Bandits with Finitely Many Arms	236
	22.1 Notes	237
	22.2 Bibliographic Remarks	238
	22.3 Exercises	238
23	Stochastic Linear Bandits with Sparsity	240
	23.1 Sparse Linear Stochastic Bandits	240
	23.2 Elimination on the Hypercube	241
	23.3 Online to Confidence Set Conversion	244
	23.4 Sparse Online Linear Prediction	248
	23.5 Notes	248
	23.6 Bibliographical Remarks	249
	23.7 Exercises	249
24	Minimax Lower Bounds for Stochastic Linear Bandits	250
	24.1 Hypercube	250
	24.2 Unit Ball	252
	24.3 Sparse Parameter Vectors	253
	24.4 Misspecified Models	255
	24.5 Notes	256
	24.6 Bibliographic Remarks	257
	24.7 Exercises	257
25	Asymptotic Lower Bounds for Stochastic Linear Bandits	258
	25.1 An Asymptotic Lower Bound for Fixed Action Sets	258
	25.2 Clouds Looming for Optimism	262
	25.3 Notes	263
	25.4 Bibliographic Remarks	264
	25.5 Exercises	264
Part VI Adversarial Linear Bandits		265
26	Foundations of Convex Analysis (↻)	267
	26.1 Convex Sets and Functions	267
	26.2 Jensen’s Inequality	269
	26.3 Bregman Divergence	269
	26.4 Legendre Functions	271
	26.5 Optimisation	273

	26.6 Projections	274
	26.7 Notes	274
	26.8 Bibliographic Remarks	275
	26.9 Exercises	275
27	Exp3 for Adversarial Linear Bandits	278
	27.1 Exponential Weights for Linear Bandits	278
	27.2 Regret Analysis	280
	27.3 Continuous Exponential Weights	281
	27.4 Notes	283
	27.5 Bibliographic Remarks	283
	27.6 Exercises	284
28	Follow-the-regularised-Leader and Mirror Descent	286
	28.1 Online Linear Optimisation	286
	28.2 Regret Analysis	290
	28.3 Application to Linear Bandits	294
	28.4 Linear Bandits on the Unit Ball	295
	28.5 Notes	298
	28.6 Bibliographic Remarks	301
	28.7 Exercises	301
29	The Relation between Adversarial and Stochastic Linear Bandits	306
	29.1 Unified View	306
	29.2 Reducing Stochastic Linear Bandits to Adversarial Linear Bandits	307
	29.3 Stochastic Linear Bandits with Parameter Noise	308
	29.4 Contextual Linear Bandits	309
	29.5 Notes	310
	29.6 Bibliographic Remarks	311
	29.7 Exercises	311
Part VII	Other Topics	313
30	Combinatorial Bandits	317
	30.1 Notation and Assumptions	317
	30.2 Applications	318
	30.3 Bandit Feedback	319
	30.4 Semi-bandit Feedback and Mirror Descent	320
	30.5 Follow-the-Perturbed-Leader	321
	30.6 Notes	326
	30.7 Bibliographic Remarks	327
	30.8 Exercises	328

31	Non-stationary Bandits	331
	31.1 Adversarial Bandits	331
	31.2 Stochastic Bandits	334
	31.3 Notes	336
	31.4 Bibliographic Remarks	337
	31.5 Exercises	338
32	Ranking	340
	32.1 Click Models	341
	32.2 Policy	343
	32.3 Regret Analysis	345
	32.4 Notes	349
	32.5 Bibliographic Remarks	351
	32.6 Exercises	351
33	Pure Exploration	353
	33.1 Simple Regret	353
	33.2 Best-Arm Identification with a Fixed Confidence	355
	33.3 Best-Arm Identification with a Budget	361
	33.4 Notes	362
	33.5 Bibliographical Remarks	364
	33.6 Exercises	365
34	Foundations of Bayesian Learning	369
	34.1 Statistical Decision Theory and Bayesian Learning	369
	34.2 Bayesian Learning and the Posterior Distribution	370
	34.3 Conjugate Pairs, Conjugate Priors and the Exponential Family	374
	34.4 The Bayesian Bandit Environment	377
	34.5 Posterior Distributions in Bandits	378
	34.6 Bayesian Regret	379
	34.7 Notes	380
	34.8 Bibliographic Remarks	382
	34.9 Exercises	382
35	Bayesian Bandits	386
	35.1 Bayesian Optimal Regret for k -Armed Stochastic Bandits	386
	35.2 Optimal Stopping (♣)	387
	35.3 One-armed bandits	389
	35.4 Gittins Index	393
	35.5 Computing the Gittins Index	399
	35.6 Notes	399
	35.7 Bibliographical Remarks	401
	35.8 Exercises	402

36	Thompson Sampling	404
	36.1 Finite-Armed Bandits	404
	36.2 Frequentist Analysis	406
	36.3 Linear Bandits	409
	36.4 Information Theoretic Analysis	411
	36.5 Notes	414
	36.6 Bibliographic Remarks	416
	36.7 Exercises	417
 Part VIII Beyond Bandits		 421
37	Partial Monitoring	423
	37.1 Finite Adversarial Partial Monitoring Problems	424
	37.2 The Structure of Partial Monitoring	426
	37.3 Classification of Finite Adversarial Partial Monitoring	430
	37.4 Lower Bounds	430
	37.5 Policy and Upper Bounds	435
	37.6 Proof of Theorem 37.16	439
	37.7 Proof of Theorem 37.17	440
	37.8 Proof of the Classification Theorem	444
	37.9 Notes	445
	37.10 Bibliographical Remarks	447
	37.11 Exercises	449
38	Markov Decision Processes	452
	38.1 Problem Set-Up	452
	38.2 Optimal Policies and the Bellman Optimality Equation	455
	38.3 Finding an Optimal Policy (♣)	458
	38.4 Learning in Markov Decision Processes	462
	38.5 Upper Confidence Bounds for Reinforcement Learning	463
	38.6 Proof of Upper Bound	465
	38.7 Proof of Lower Bound	468
	38.8 Notes	471
	38.9 Bibliographical Remarks	473
	38.10 Exercises	475
	 Bibliography	 484
	Index	513