

Probability tools, tricks, and miracles

©David Pollard

1	What and why	1
1.1	How I chose material for this book	1
1.2	Naming and notation	2
1.3	Notes	3
2	It's just calculus and convexity	6
2.1	Introduction	6
2.2	Polynomial approximations	7
2.3	Convexity of the log MGF	9
2.4	Conjugate convex functions	11
2.5	Approximation of factorials	13
2.6	Distances between probability measures	17
2.7	Bounding sums by integrals	19
2.8	Problems	19
2.9	Notes	22
3	The moment generating function method	26
3.1	Tail bounds from the moment generating function	27
3.2	Behavior of the tail bound near the origin	28
3.3	Normal	30
3.4	Global behavior of the normal hazard rate	35
3.5	Poisson	41
3.6	Gamma and chi-squared	43
3.7	Binomial	46
3.8	Sampling and the hypergeometric	50
3.9	Problems	52
3.10	Notes	57
4	Poisson-Binomial	63
4.1	The extreme Binomial	63
4.2	The PGF of the Poisson-Binomial	64
4.3	Polynomials with real roots	66
4.4	Sampling and the Poisson-Binomial	69
4.5	Ratios and mode(s)	73
4.6	Mode(s) of the Poisson-Binomial	74
4.6.1	Zeros and ones	77
4.7	Binomial tails versus Poisson-Binomial tails	79
4.8	The median of the Poisson-Binomial	84
4.9	Problems	85
4.10	Notes	87

5	Orlicz spaces	90
5.1	What is an orlicz space?	90
5.2	Facts about orlicz functions	92
5.3	Orlicz functions that grow exponentially fast	95
5.4	Norms of maxima of finitely many variables	98
5.5	Conjugacy and Young's inequality	100
5.6	Comparison of $\mathcal{L}^p$ and $\mathcal{L}^\Psi$	104
5.7	Problems	106
5.8	Notes	110
6	Multivariate normal	113
6.1	Introduction	113
6.1.1	Subgaussian concentration	114
6.1.2	Comparison inequalities	116
6.2	The path method	117
6.3	Concentration of LIPSCHITZ functionals	119
6.3.1	The PISIER-MAUREY method	120
6.3.2	The smart path method	121
6.3.3	The stochastic calculus method	123
6.4	The gaussian isoperimetric inequality	127
6.5	Tail bound comparisons	128
6.6	An application of the GORDON inequality	132
6.7	A generalized FERNIQUE inequality	135
6.8	Problems	137
6.9	Notes	140
7	Subgaussian distributions	146
7.1	Definition of subgaussian	146
7.2	Characterizations of subgaussianity	148
7.3	The Hoeffding inequality for independent summands	150
7.4	From independence to martingales	152
7.5	A cautionary example	157
7.6	Problems	159
7.7	Notes	160
8	Between subgaussian and subexponential	162
8.1	Introduction	162
8.2	A BENNETT inequality for independent summands	163
8.3	A BERNSTEIN inequality	166
8.4	Quadratics in independent subgaussians	170
8.5	A BENNETT inequality for martingales	175
8.6	Concentration of random polynomials	177
8.7	Problems	185
8.8	Notes	189

9	Measurability woes	192
9.1	The difficulty	192
9.2	Essential supremum and infimum	193
9.3	Outer integrals and measurable cover functions	194
9.4	Separable versions	195
9.5	Problems	198
9.6	Notes	198
10	Chaining	201
10.1	Maximal inequalities	201
10.2	A reminder about ORLICZ norms	203
10.3	What is chaining?	205
10.4	Covering and packing numbers	210
10.5	Chaining with packing numbers	213
10.5.1	Expected values	214
10.5.2	Orlicz norms	214
10.5.3	Conditional expected values	215
10.5.4	Tail probabilities	215
10.6	Oscillation and continuity of sample paths	217
10.7	An example of a classical chaining argument	219
10.8	Problems	220
10.9	Notes	223
11	Chaining via partitions	227
11.1	Overview	227
11.2	Partitions, nested and not nested	230
11.2.1	Case: Nested partitions	230
11.2.2	Case: Non-nested partitions	231
11.2.3	Chaining bounds	231
11.3	Fernique and majorizing measures	232
11.4	Majorizing measures as smoothers	234
11.5	Partitions with weights	236
11.6	From majorizing measures to weighted frameworks	239
11.7	Talagrand and admissible partitions	242
11.8	From majorizing measures to admissible partitions	245
11.9	An illuminating example	248
11.9.1	Paths bounded but not continuous	249
11.9.2	Failure of the packing argument	250
11.9.3	Majorizing measure	250
11.9.4	Weighted partitions	250
11.9.5	Admissible partitions	251
11.10	A comparison of chaining methods	252
11.11	Problems	253
11.12	Notes	256

12 Majorizing measures for gaussian processes	259
12.1 Introduction	259
12.2 A gaussian growth property	260
12.3 From functional to nested partitions with weights	262
12.4 Uniformly continuous gaussian sample paths	265
12.5 From functional to admissible partitions	267
12.6 Beyond gaussian	272
12.7 Problems	272
12.8 Notes	273
 13 Symmetrization	 275
13.1 Symmetry?	275
13.2 Stochastic processes as random objects	283
13.3 A generic tail bound by symmetrization	286
13.4 Empirical processes	286
13.5 Expected supremum of an empirical process	288
13.6 Symmetrization of independent processes	292
13.7 Maximal tail bounds for an empirical process	293
13.8 Oscillation control for an empirical process	297
13.9 U-processes	300
13.10 Problems	303
13.11 Notes	306
 16 Convergence in distribution	 310
16.1 Difficulties with the uniform metric	310
16.2 Convergence in distribution, new style	314
16.3 Approximation versus limit theorem	318
16.4 Convergence of the uniform empirical process	320
16.5 Problems	323
16.6 Notes	324
 14 VC Sets	 326
14.1 An introductory example	326
14.2 VC-dimension, shatter-dimension	330
14.3 Proof of the shatter theorem	335
14.4 Another way to think about binary matrices	337
14.5 Haussler's improvement of the packing bound	340
14.6 Problems	343
14.7 Notes	346
 15 Fat shattering	 349
15.1 VC-subgraphs	349
15.2 Fat shattering	353
15.3 Packing for $[0, 1]$ -valued functions	357
15.4 Packing for subsets of an $\mathcal{L}^q$ ball	359
15.5 Chaining bounds	362
15.6 Lattice cubes surrounded by well separated vectors	364

15.7	Problems	367
15.8	Notes	371
17	Concentration: Talagrand's inductive method	373
17.1	Introduction	373
17.2	Examples	375
17.3	Convexity and convex hulls	378
17.4	Control by a single point	380
17.5	Proof of the convex hull theorem	383
17.6	Problems	388
17.7	Notes	391
18	Concentration: the entropy method	393
18.1	Introduction	394
18.2	Tensorization	395
18.3	Marginal calculations	397
18.4	Massart et al	402
18.5	Concentration for empirical processes	406
18.6	Problems	406
18.7	Notes	407
19	Bracketing methods	408
19.1	What is bracketing?	409
19.2	Bounded independent summands	413
19.3	Independent summands with truncation	417
19.4	A generic bracketing bound	422
19.5	Phi mixing	427
19.6	Absolute regularity	430
19.7	Strong mixing	432
19.8	Tail probability bounds	433
19.9	Problems	433
19.10	Notes	434