

Contents

1	Notation and conventions	6
7	1.1 Some Useful Mathematical Facts	7
8	1.2 Acknowledgements	8
9	2 First Tools for Looking at Data	9
9	2.1 Datasets	9
11	2.2 What's Happening? - Plotting Data	11
12	2.2.1 Bar Charts	12
12	2.2.2 Histograms	12
13	2.2.3 How to Make Histograms	13
15	2.2.4 Conditional Histograms	15
15	2.3 Summarizing 1D Data	15
16	2.3.1 The Mean	16
17	2.3.2 Standard Deviation and Variance	17
20	2.3.3 Variance	20
21	2.3.4 The Median	21
22	2.3.5 Interquartile Range	22
24	2.3.6 Using Summaries Sensibly	24
25	2.4 Plots and Summaries	25
25	2.4.1 Some Properties of Histograms	25
27	2.4.2 Standard Coordinates and Normal Data	27
30	2.4.3 Boxplots	30
32	2.5 Whose is bigger? Investigating Australian Pizzas	32
36	2.6 What You Must Remember	36
38	3 Intermezzo - Programming Tools	38
43	4 Looking at Relationships	43
43	4.1 Plotting 2D Data	43
43	4.1.1 Categorical Data, Counts, and Charts	43
46	4.1.2 Series	46
48	4.1.3 Scatter Plots for Spatial Data	48
49	4.1.4 Exposing Relationships with Scatter Plots	49
53	4.2 Correlation	53
55	4.2.1 The Correlation Coefficient	55
60	4.2.2 Using Correlation to Predict	60
64	4.2.3 Confusion caused by correlation	64
65	4.3 Sterile Males in Wild Horse Herds	65
68	4.4 What You Must Remember	68
72	5 Basic ideas in probability	72
72	5.1 Experiments, Outcomes, Events, and Probability	72
74	5.1.1 The Probability of an Outcome	74

**

**

**

76	5.1.2	Events
78	5.1.3	The Probability of Events
81	5.1.4	Computing Probabilities by Counting Outcomes
84	5.1.5	Computing Probabilities by Reasoning about Sets
87	5.1.6	Independence
91	5.1.7	Permutations and Combinations
95	5.2	Conditional Probability
96	5.2.1	Evaluating Conditional Probabilities
102	5.2.2	The Prosecutors Fallacy
103	5.2.3	Independence and Conditional Probability
105	5.3	Example: The Monty Hall Problem
107	5.4	What you should remember
111	6	Random Variables and Expectations
111	6.1	Random Variables
113	6.1.1	Joint and Conditional Probability for Random Variables
116	6.1.2	Just a Little Conditional Probability
120	6.2	Expectations and Expected Values
120	6.2.1	Expected Values of Discrete Random Variables
121	6.2.2	Expected Values of Continuous Random Variables
122	6.2.3	Mean, Variance and Covariance
126	6.2.4	Expectations and Statistics
127	6.2.5	Indicator Functions
128	6.2.6	Two Inequalities
130	6.2.7	IID Samples and the Weak Law of Large Numbers
132	6.3	Using Expectations
133	6.3.1	Should you accept a bet?
135	6.3.2	Odds, Expectations and Bookmarking — a Cultural Diversion
136	6.3.3	Ending a Game Early
137	6.3.4	Making a Decision with Decision Trees and Expectations
139	6.3.5	Utility
141	6.4	What you should remember
145	7	Useful Probability Distributions
145	7.1	Discrete Distributions
145	7.1.1	The Discrete Uniform Distribution
145	7.1.2	The Geometric Distribution
146	7.1.3	Bernoulli Random Variables
147	7.1.4	The Binomial Probability Distribution
149	7.1.5	Multinomial probabilities
150	7.1.6	The Poisson Distribution
152	7.2	Continuous Distributions
152	7.2.1	The Continuous Uniform Distribution
152	7.2.2	The Beta Distribution
153	7.2.3	The Gamma Distribution
154	7.2.4	The Exponential Distribution
155	7.3	The Normal Distribution

**

**

* *

155	7.3.1	The Standard Normal Distribution	
156	7.3.2	The Normal Distribution	
157	7.3.3	Properties of the Normal Distribution	**
158	7.4	Approximating Binomials with Large N	**
159	7.4.1	Large N	
161	7.4.2	Getting Normal	
163	7.4.3	So What?	
164	7.5	What you should remember	
169	8	Markov Chains and Simulation	
169	8.1	Markov Chains	
169	8.1.1	Motivating Example: Multiple Coin Flips	
171	8.1.2	Motivating Example: The Gambler's Ruin	
173	8.1.3	Motivating Example: A Virus	
173	8.1.4	Markov Chains	
176	8.1.5	Example: Particle Motion as a Markov Chain	
178	8.2	Simulation	**
178	8.2.1	Obtaining Uniform Random Numbers	
178	8.2.2	Computing Expectations with Simulations	
178	8.2.3	Computing Probabilities with Simulations	
180	8.2.4	Simulation Results as Random Variables	
182	8.2.5	Obtaining Random Samples	
184	8.3	Simulation Examples	
185	8.3.1	Simulating Experiments	
187	8.3.2	Simulating Markov Chains	
188	8.3.3	Example: Ranking the Web by Simulating a Markov Chain	
191	8.3.4	Example: Simulating a Complicated Game	
201	9	Inference: Making Point Estimates	
202	9.1	Estimating Model Parameters with Maximum Likelihood	
203	9.1.1	The Maximum Likelihood Principle	***
211	9.1.2	Cautions about Maximum Likelihood	
211	9.2	Incorporating Priors with Bayesian Inference	
212	9.2.1	Constructing the Posterior	
215	9.2.2	The Posterior for Normal Data	*
218	9.2.3	MAP Inference	
220	9.2.4	Cautions about Bayesian Inference	
220	9.3	Samples, Urns and Populations	
221	9.3.1	Estimating the Population Mean from a Sample	
222	9.3.2	The Variance of the Sample Mean	***
226	9.3.3	The Probability Distribution of the Sample Mean	
226	9.3.4	When The Urn Model Works	
230	10	Intervals and Testing	
230	10.1	Straightforward Constructions of Confidence Intervals	
231	10.1.1	Confidence Intervals for Population Means	**
234	10.1.2	Bayesian Confidence Intervals	**

236	10.2 Using Simulation to Construct Intervals	*
236	10.2.1 Constructing Confidence Intervals for Parametric Models	
239	10.2.2 Estimating Standard Error	
242	10.3 Using the Standard Error to Evaluate Hypotheses	***
245	10.3.1 Does this Population have this Mean?	
249	10.3.2 Do Two Populations have the same Mean?	
254	10.3.3 Variants on the Basic Test	
255	10.4 χ^2 Tests: Is Data Consistent with a Model?	*
260	11 Extracting Important Relationships in High Dimensions	
260	11.1 Summaries and Simple Plots	
260	11.1.1 The Mean	****
261	11.1.2 Parallel Plots	
266	11.2 Blob Analysis of High-Dimensional Data	
266	11.2.1 Understanding Blobs with Scatterplot Matrices	
267	11.2.2 Transforming High Dimensional Data	**
269	11.2.3 Transforming Blobs	
272	11.2.4 Whitening Data	
275	11.3 Principal Components Analysis	
275	11.3.1 The Blob Coordinate System and Smoothing	
277	11.3.2 The Low-Dimensional Representation of a Blob	****
278	11.3.3 Smoothing Data with a Low-Dimensional Representation	
282	11.3.4 The Error of the Low-Dimensional Representation	
283	11.3.5 Example: Representing Spectral Reflectances	
285	11.3.6 Example: Representing Faces with Principal Components	
287	11.4 Multi-Dimensional Scaling	
287	11.4.1 Principal Coordinate Analysis	*
289	11.4.2 Example: Mapping with Multidimensional Scaling	
291	11.5 Example: Understanding Height and Weight	
296	12 Models of High Dimensional Data	
296	12.1 The Curse of Dimension	
298	12.2 The Multivariate Normal Distribution	
299	12.2.1 Affine Transformations and Gaussians	****
299	12.3 Agglomerative and Divisive Clustering	
301	12.3.1 Clustering and Distance	
301	12.3.2 Example: Agglomerative Clustering	
302	12.4 The K-Means Algorithm	
307	13 Regression	
307	13.1 Least Squares and Linear Models	***
310	13.1.1 Interpreting a Regression	
310	13.1.2 How Good is a Fit?	
313	13.1.3 Regularizing Linear Regressions	

318	14.1	Classification, Error, and Loss	
318	14.1.1	Using Loss to Determine Decisions	
320	14.1.2	Example: Building a Classifier out of Histograms	*
322	14.1.3	Example: Classifying Using Nearest Neighbors	*
323	14.1.4	Training Error, Test Error, and Overfitting	
324	14.1.5	Error Rate and Cross-Validation	
325	14.2	Basic Ideas for Numerical Minimization	
326	14.2.1	Overview	
327	14.2.2	Gradient Descent	*
328	14.2.3	Stochastic Gradient Descent	
329	14.2.4	Newton's Method	
330	14.3	Logistic Regression, Minimizing Loss, and Regularization	
330	14.3.1	Logistic Regression	
334	14.3.2	Regularization	
336	14.4	The Linear Support Vector Machine	****
336	14.4.1	Classifying with a Hyperplane	
338	14.4.2	The Hinge Loss	
340	14.4.3	Example: Training a Support Vector Machine with Stochastic Gradient Descent	
342	14.5	Practical Methods for Building Classifiers	
343	14.5.1	Manipulating Training Data to Improve Performance	
343	14.5.2	Building Multi-Class Classifiers Out of Binary Classifiers	*
344	14.5.3	Finding the Nearest Neighbors and Hashing	
347	14.5.4	Software for SVMs	
348	14.6	Notes	
351	15	Math Resources	
351	15.1	Useful Material about Matrices	
352	15.1.1	Approximating A Symmetric Matrix	