

Contents

1	Statistics	13
1.1	Introduction	13
1.2	Probabilities	13
1.2.1	Discrete Variables	13
1.2.2	Continuous Variables	14
1.2.3	The Gaussian Density	15
1.2.4	Probability relations	16
1.3	Expectation and Moments	16
1.4	Maximum Likelihood Estimation	17
1.5	Correlation and Regression	18
1.5.1	Correlation	18
1.5.2	Linear regression	18
1.6	Bias and Variance	20
1.7	Minimum variance estimation	21
1.8	Statistical Inference	22
1.8.1	Means	23
1.8.2	Regression	24
1.8.3	Correlation	25
1.9	Discussion	26
2	Linear Algebra	27

2.1	Introduction	27
2.2	Transposes and Inner Products	27
	2.2.1 Properties of matrix multiplication	28
2.3	Types of matrices	29
	2.3.1 Covariance matrices	29
	2.3.2 Diagonal matrices	29
	2.3.3 The correlation matrix	30
	2.3.4 The identity matrix	30
2.4	The Matrix Inverse	31
2.5	Orthogonality	33
	2.5.1 Angles between vectors	33
	2.5.2 Projections	34
	2.5.3 Orthogonal Matrices	35
	2.5.4 Orthonormal Matrices	36
2.6	Subspaces	37
2.7	Determinants	37
2.8	Eigenanalysis	38
2.9	Gram-Schmidt	39
	2.9.1 Diagonalization	40
	2.9.2 Spectral Theorem	40
2.10	Complex Matrices	41
2.11	Quadratic Forms	41
	2.11.1 Ellipses	42
3	Multivariate Statistics	43
	3.1 Introduction	43
	3.2 Multivariate Linear Regression	43

3.2.1	Estimating the weights	44
3.2.2	Understanding the solution	44
3.2.3	Feature selection	45
3.2.4	Example	47
3.2.5	Partial Correlation	47
3.3	Principal Component Analysis	49
3.3.1	The Multivariate Gaussian Density	49
3.3.2	Dimensionality Reduction	50
3.3.3	Singular Value Decomposition	51
4	Information Theory	53
4.1	Introduction	53
4.2	Measures of Information	53
4.3	Entropy	54
4.4	Joint Entropy	55
4.5	Relative Entropy	56
4.6	Mutual Information	56
4.7	Minimum Description Length	57
5	Fourier methods	59
5.1	Introduction	59
5.2	Sinewaves and Samples	59
5.3	Sinusoidal models	60
5.3.1	Fitting the model	61
5.3.2	But sinewaves are orthogonal	61
5.4	Fourier Series	64
5.4.1	Example	65
5.5	Fourier Transforms	65

5.5.1	Discrete Fourier Transform	65
5.5.2	The Fourier Matrix	68
5.6	Time-Frequency relations	69
5.6.1	Power Spectral Density	70
5.6.2	Filtering	70
5.6.3	Autocovariance and Power Spectral Density	71
5.7	Spectral Estimation	72
5.7.1	The Periodogram	72
5.7.2	Autocovariance methods	72
5.7.3	Aliasing	73
5.7.4	Filtering	73
6	Stochastic Processes	77
6.1	Introduction	77
6.2	Autocorrelation	77
6.3	Autoregressive models	79
6.3.1	Random walks	80
6.3.2	Relation to autocorrelation	80
6.3.3	Relation to partial autocorrelation	82
6.3.4	Model order selection	83
6.3.5	Example: Sleep EEG	84
6.3.6	Discussion	84
6.4	Moving Average Models	85
6.5	Spectral Estimation	85
7	Multiple Time Series	87
7.1	Introduction	87
7.2	Cross-correlation	87

7.2.1	Cross-correlation is asymmetric	88
7.2.2	Windowing	88
7.2.3	Time-Delay Estimation	90
7.3	Multivariate Autoregressive models	90
7.3.1	Model order selection	91
7.3.2	Example	91
7.4	Cross Spectral Density	92
7.4.1	More than two time series	94
7.4.2	Coherence and Phase	94
7.4.3	Welch's method for estimating coherence	95
7.4.4	MAR models	95
7.5	Example	96
7.6	Partial Coherence	97
8	Subspace Methods	99
8.1	Introduction	99
8.2	Singular Spectrum Analysis	99
8.2.1	Embedding	99
8.2.2	Noisy Time Series	100
8.2.3	Embedding Sinewaves	102
8.3	Spectral estimation	103
8.3.1	Model Order Selection	106
8.3.2	Comparison of methods	106
9	Nonlinear Methods	107
9.1	Introduction	107
9.2	Lyapunov Exponents	108
9.3	Measures of Information	109

9.3.1	Continuous variables	109
9.3.2	Measures of Information for Time Series	110
9.3.3	Marginal Mutual Information	111
9.3.4	Source Entropy	112
9.3.5	Correlation Sums	113
9.4	Nonlinear Prediction	114
9.4.1	Local methods	115
9.4.2	Global methods	116
9.5	Discusion	117
10	Bayesian Methods	119
10.1	Introduction	119
10.2	Bayes Rule	119
10.2.1	Example	120
10.3	Gaussian Variables	120
10.3.1	Combining Estimates	121
10.3.2	Sequential Estimation	121
10.4	Multiple Gaussian Variables	122
10.5	General Linear Models	123
10.5.1	The evidence framework	124
10.5.2	Example	125
11	Kalman Filters	127
11.1	Introduction	127
11.1.1	Sequential Estimation of Nonstationary Mean	127
11.1.2	A single state variable	128
11.1.3	Multiple state variables	130
11.1.4	Dynamic Linear Models	131

11.1.5	Recursive least squares	132
11.1.6	Estimation of noise parameters	133
11.1.7	Comparison with steepest descent	135
11.1.8	Other algorithms	136
11.1.9	An example	137
11.1.10	Discussion	138
12	EM algorithms	141
12.1	Gaussian Mixture Models	141
12.2	General Approach	142
12.3	Probabilistic Principal Component Analysis	143
12.4	Linear Dynamical Systems	144
A	Series and Complex Numbers	149
A.1	Power series	149
A.2	Complex numbers	150
A.3	Complex exponentials	151
A.4	DeMoivre's Theorem	152
A.5	Argand Diagrams	152
B	Linear Regression	153
B.1	Univariate Linear Regression	153
B.1.1	Variance of slope	154
B.2	Multivariate Linear Regression	155
B.2.1	Estimating the weight covariance matrix	155
B.3	Functions of random vectors	156
B.3.1	Estimating the weight covariance matrix	156
B.3.2	Equivalence of t-test and F-test for feature selection	157

C	Matrix Identities	159
C.1	Multiplication	159
C.2	Transposes	159
C.3	Inverses	159
C.4	Eigendecomposition	160
C.5	Determinants	160
C.6	Traces	160
C.7	Matrix Calculus	160
D	Probability Distributions	161
D.1	Transforming PDFs	161
	D.1.1 Mean and Variance	161
D.2	Uniform Distribution	163
D.3	Gaussian Distribution	163
	D.3.1 Entropy	163
	D.3.2 Relative Entropy	164
D.4	The Gamma distribution	164
	D.4.1 Entropy	165
	D.4.2 Relative Entropy	165
D.5	The χ^2 -distribution	165
D.6	The t-distribution	166
D.7	Generalised Exponential Densities	166
D.8	PDFs for Time Series	168
	D.8.1 Sampling	168
	D.8.2 Sine Wave	168
E	Multivariate Probability Distributions	171
E.1	Transforming PDFs	171

E.1.1	Mean and Covariance	171
E.2	The Multivariate Gaussian	172
E.2.1	Entropy	172
E.2.2	Relative Entropy	173
E.3	The Multinomial Distribution	173
E.4	The Dirichlet Distribution	173
E.4.1	Relative Entropy	173