

Contents

Dedication	v
Acknowledgments	vi
Glossary	vii
Notation	viii
1 Introduction	1
2 Probabilistic Inference in Signal Processing	6
2.1 Introduction	6
2.2 The likelihood function	7
2.2.1 Maximum likelihood	8
2.3 Bayesian data analysis	9
2.4 Prior probabilities	10
2.4.1 Flat priors	10
2.4.2 Smoothness priors	11
2.4.3 Convenience priors	12
2.5 The removal of nuisance parameters	12
2.6 Model selection using Bayesian evidence	13
2.6.1 Ockham's razor	14
2.7 The general linear model	15
2.8 Interpretations of the general linear model	17

2.8.1	Features	17
2.8.2	Orthogonalization	17
2.9	Example of marginalization	18
2.9.1	Results	19
2.10	Example of model selection	20
2.10.1	Closed form expression for evidence	21
2.10.2	Determining the order of a polynomial	22
2.10.3	Determining the order of an AR process	22
2.11	Concluding remarks	24
3	Numerical Bayesian Inference	26
3.1	The normal approximation	27
3.1.1	Effect of number of data on the likelihood function	27
3.1.2	Taylor approximation	28
3.1.3	Reparameterization	29
3.1.4	Jacobian of transformation	31
3.1.5	Normal approximation to evidence	31
3.1.6	Normal approximation to the marginal density	32
3.1.7	The delta method	33
3.2	Optimization	34
3.2.1	Local algorithms	35
3.2.2	Global algorithms	38
3.2.3	Concluding remarks	41
3.3	Integration	42
3.4	Numerical quadrature	43
3.4.1	Multiple integrals	44
3.5	Asymptotic approximations	46
3.5.1	The saddlepoint approximation and Edgeworth series	47
3.5.2	The Laplace approximation	47
3.5.3	Moments and expectations	48
3.5.4	Marginalization	49
3.6	The Monte Carlo method	51
3.7	The generation of random variates	54
3.7.1	Uniform variates	54
3.7.2	Non-uniform variates	55
3.7.3	Transformation of variables	55
3.7.4	The rejection method	56
3.7.5	Other methods	56
3.8	Evidence using importance sampling	57
3.8.1	Choice of sampling density	57
3.8.2	Orthogonalization using noise colouring	60
3.9	Marginal densities	61
3.9.1	Histograms	61
3.9.2	Jointly distributed variates	62
3.9.3	The dummy variable method	62

3.9.4	Marginalization using jointly distributed variates . . .	64
3.10	Opportunities for variance reduction	64
3.10.1	Quasi-random sequences	65
3.10.2	Antithetic variates	65
3.10.3	Control variates	66
3.10.4	Stratified sampling	66
3.11	Summary	67
4	Markov Chain Monte Carlo Methods	69
4.1	Introduction	69
4.2	Background on Markov chains	70
4.3	The canonical distribution	73
4.3.1	Energy, temperature and probability	73
4.3.2	Random walks	74
4.3.3	Free energy and model selection	75
4.4	The Gibbs sampler	75
4.4.1	Description	76
4.4.2	Discussion	77
4.4.3	Convergence	77
4.5	The Metropolis-Hastings algorithm	78
4.5.1	The general algorithm	78
4.5.2	Convergence	80
4.5.3	Choosing the proposal density	80
4.5.4	Relationship between Gibbs and Metropolis	81
4.6	Dynamical sampling methods	81
4.6.1	Derivation	82
4.6.2	Hamiltonian dynamics	83
4.6.3	Stochastic transitions	84
4.6.4	Simulating the dynamics	85
4.6.5	Hybrid Monte Carlo	86
4.6.6	Convergence to canonical distribution	87
4.7	Implementation of simulated annealing	87
4.7.1	Annealing schedules	87
4.7.2	Annealing with Markov chains	89
4.8	Other issues	91
4.8.1	Assessing convergence of Markov chains	92
4.8.2	Determining the variance of estimates	92
4.9	Free energy estimation	93
4.9.1	Thermodynamic integration	93
4.9.2	Other methods	95
4.10	Summary	95
5	Retrospective Changepoint Detection	96
5.1	Introduction	96
5.2	The simple Bayesian step detector	97

5.2.1	Derivation of the step detector	97
5.2.2	Application of the step detector	100
5.3	The detection of changepoints using the general linear model	101
5.3.1	The general piecewise linear model	101
5.3.2	Simple step detector in generalized matrix form . . .	103
5.3.3	Changepoint detection in AR models	104
5.3.4	Application of AR changepoint detector	105
5.4	Recursive Bayesian estimation	106
5.4.1	Update of position	107
5.4.2	Update given more data	108
5.5	Detection of multiple changepoints	109
5.6	Implementation details	110
5.6.1	Sampling changepoint space	113
5.6.2	Sampling linear parameter space	114
5.6.3	Sampling noise parameter space	116
5.7	Multiple changepoint results	116
5.7.1	Synthetic step data	116
5.7.2	Well log data	117
5.8	Concluding Remarks	119
6	Restoration of Missing Samples in Digital Audio Signals	122
6.1	Introduction	122
6.2	Model formulation	123
6.2.1	The likelihood and the excitation energy	124
6.2.2	Maximum likelihood	125
6.3	The EM algorithm	126
6.3.1	Expectation	128
6.3.2	Maximization	129
6.4	Gibbs sampling	130
6.4.1	Description	130
6.4.2	Derivation of conditional densities	131
6.4.3	Conditional density for the missing data	132
6.4.4	Conditional density for the autoregressive parameters	132
6.4.5	Conditional density for the standard deviation . . .	133
6.5	Implementation issues	134
6.5.1	Estimating AR parameters	134
6.5.2	Implementing the ML algorithm	135
6.5.3	Implementing the EM algorithm	136
6.5.4	Implementation of Gibbs sampler	136
6.6	Relationship between the three restoration methods	138
6.6.1	ML vs Gibbs	139
6.6.2	Gibbs vs EM	139
6.6.3	EM vs ML	139
6.7	Simulations	140

6.7.1	Autoregressive model with poles near unit circle . . .	140
6.7.2	Autoregressive model with poles near origin	141
6.7.3	Sine wave	145
6.7.4	Evolution of sample interpolants	147
6.7.5	Hairy sine wave	149
6.7.6	Real data: Tuba	149
6.7.7	Real data: Sinéad O'Connor	152
6.8	Discussion	155
6.8.1	The temperature of an interpolant	155
6.8.2	Data augmentation	156
6.9	Concluding remarks	158
6.9.1	Typical interpolants	158
6.9.2	Computation	159
6.9.3	Modelling issues	160
7	Integration in Bayesian Data Analysis	161
7.1	Polynomial data	161
7.1.1	Polynomial data	163
7.1.2	Sampling the joint density	163
7.1.3	Approximate evidence	166
7.1.4	Approximate marginal densities	170
7.1.5	Conclusion	176
7.2	Decay problem	177
7.2.1	The Lanczos problem	178
7.2.2	Biomedical data	184
7.2.3	Concluding remarks	187
7.3	General model selection	189
7.3.1	Model selection in an impulsive noise environment . .	190
7.3.2	Model selection in a Gaussian noise environment . .	191
7.4	Summary	192
8	Conclusion	194
8.1	A review of the work	194
8.2	Further work	197
A	The General Linear Model	201
A.1	Integrating out model amplitudes	202
A.1.1	Least squares	202
A.1.2	Orthogonalization	203
A.2	Integrating out the standard deviation	204
A.3	Marginal density for a linear coefficient	204
A.4	Marginal density for standard deviation	205
A.5	Conditional density for a linear coefficient	205
A.6	Conditional density for standard deviation	206

B	Sampling from a Multivariate Gaussian Density	207
C	Hybrid Monte Carlo Derivations	209
	C.1 Full Gaussian likelihood	209
	C.2 Student-t distribution	210
	C.3 Remark	211
D	EM Algorithm Derivations	212
	D.1 Expectation	212
	D.2 Maximization	214
E	Issues in Sampling Based Approaches to Integration	217
	E.1 Marginalizing using the conditional density	217
	E.2 Approximating the conditional density	218
	E.3 Gibbs sampling from the joint density	220
	E.4 Reverse importance sampling	220
F	Detailed Balance	223
	F.1 Detailed balance in the Gibbs sampler	223
	F.2 Detailed balance in the Metropolis-Hastings algorithm . . .	224
	F.3 Detailed balance in the Hybrid Monte Carlo algorithm . . .	224
	F.4 Remarks	225
	References	227
	Index	237