# Contents

*Preface* *page* 1

**1 Probability basics** 4
1.1 The three doors problem 4
1.1.1 Earthquake variation 5
1.2 Probability 6
1.2.1 Fundamental properties 6
1.2.2 Some problems 7
1.2.3 Frequency 8
1.2.4 Bayes' theorem and the principle of inference 8
1.2.5 Discrete and continuous probability distributions 9
1.2.6 Normalization 10
1.3 Expectation, variance, and moments 10
1.4 Univariate probability distributions 13
1.4.1 Binomial 13
1.4.2 Poisson 16
1.4.3 Beta 19
1.4.4 Uniform 20
1.4.5 Gaussian 20
1.4.6 Gamma 21
1.4.7 Cauchy 22
1.5 Cumulative distribution function and quantile function 23
1.6 Multiple variables 24
1.6.1 Covariance 24
1.6.2 Multivariate probability distributions 25
1.7 Combinations and permutations 28
1.7.1 Using R 30
1.8 Random number generation 31
1.9 Change of variables 33
1.9.1 One-dimensional 33
1.9.2 Multi-dimensional 33
1.10 The three doors problem revisited 35

**2 Estimation and uncertainty** 36
2.1 Estimators 36
2.2 Noise, measurement errors, and measurement models 37



Downloaded from https://www.cambridge.org/core. Universitaet Heidelberg, on 06 Aug 2018 at 07:59:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108123891

2.3 The central limit theorem and $\sqrt{N}$ reduction from repeated measurements 39
2.3.1 Demonstration of the central limit theorem 40
2.3.2 A counter-example: estimating the mean of the Cauchy distribution 42
2.4 Population properties vs sample estimates 44
2.5 The mean is not necessarily the most efficient estimator 45
2.6 Outliers and robust statistics 47
2.7 Errors, accuracy, and precision 49
2.8 Propagation and combination of uncertainties 50
2.8.1 Linear function of one variable 50
2.8.2 Arbitrary function of one variable 50
2.8.3 Linear function of two variables 51
2.8.4 Arbitrary function of two variables 52
2.8.5 Arbitrary function of many variables 52
2.8.6 Fractional errors 53
2.8.7 Weighted mean 53

**3 Statistical models and inference** 55
3.1 Introduction to data modelling 55
3.2 Bayesian model comparison 56
3.2.1 Theory 56
3.2.2 Does a positive test result mean you have a disease? 58
3.2.3 Thinking in terms of frequencies may help 61
3.3 Data modelling with parametric models 61
3.3.1 The likelihood 62
3.3.2 The prior 63
3.3.3 The posterior 63
3.3.4 The evidence 65
3.4 Making decisions 66
3.5 Estimating one parameter from one data point 67
3.5.1 Misleading intuition 69
3.5.2 The inference problem 70
3.5.3 An improper uniform prior 71
3.5.4 A proper uniform prior 72
3.5.5 What is a good prior? 73
3.6 An inference story 74

**4 Linear models, least squares, and maximum likelihood** 76
4.1 One-dimensional linear regression 76
4.1.1 Uncertainty estimates on the model parameters 77
4.1.2 Uncertainties on predictions 79
4.1.3 Measures for goodness of fit 79
4.1.4 Linear model fitting with measurement errors 80
4.1.5 Using R: the `lm` function 81
4.2 Linear models 83

Downloaded from https://www.cambridge.org/core. Universitaet Heidelberg, on 06 Aug 2018 at 07:59:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108123891

4.3 Degrees of freedom 84
4.4 Maximum likelihood and maximum posterior probability 84
4.4.1 Maximum likelihood estimate of the weighted mean of Gaussian variables 86
4.4.2 Least squares and $\chi^2$ as a case of maximum likelihood 86
4.4.3 Minimizing $L^1$ and $L^2$ losses 87
4.4.4 Maximum posterior 87
4.4.5 Maximum posterior estimate for Poisson data 88
4.5 Multi-dimensional linear regression 89
4.5.1 Generalized least squares 90
4.6 One-dimensional regression with nonlinear functions 91
4.7 Least squares with errors on both axes 94
4.8 Bias-variance decomposition 95

**5 Parameter estimation: single parameter** 98
5.1 Bayesian analysis of coin tossing 98
5.1.1 Uniform prior 99
5.1.2 Beta prior 101
5.2 Likelihoods can be arbitrarily small and their absolute values are irrelevant 109
5.3 Assigning priors 112
5.3.1 Location and scale parameters 115
5.3.2 Jeffreys prior 116
5.3.3 Are priors subjective? 119
5.4 Some other conjugate priors 120
5.5 Summarizing distributions 121

**6 Parameter estimation: multiple parameters** 125
6.1 Conditional and marginal distributions 125
6.2 Inferring the parameters of a Gaussian 126
6.2.1 Standard deviation known 127
6.2.2 Mean known 127
6.2.3 Neither standard deviation nor mean known 128
6.3 A two-parameter problem: estimating amplitude and background 133
6.3.1 R code for fitting the amplitude and background 137
6.3.2 Suggested experiments 139
6.3.3 A note on computation with finite precision 140

**7 Approximating distributions** 142
7.1 The quadratic approximation 142
7.1.1 One dimension 142
7.1.2 Two dimensions 144
7.1.3 Higher dimensions 144
7.2 Density estimation 145
7.2.1 Histograms 145

Downloaded from https://www.cambridge.org/core. Universitaet Heidelberg, on 06 Aug 2018 at 07:59:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108123891

7.2.2 Kernel density estimation 147

**8 Monte Carlo methods for inference** 154
8.1 Why we need efficient sampling 154
8.2 Uses of integration in Bayesian inference 156
8.2.1 Marginal parameter distributions 156
8.2.2 Expectation values (parameter estimation) 156
8.2.3 Model comparison (marginal likelihood) 156
8.2.4 Prediction 156
8.3 Monte Carlo integration 157
8.4 Monte Carlo sampling 159
8.4.1 Rejection sampling 160
8.4.2 Importance sampling 161
8.5 Markov Chain Monte Carlo 161
8.5.1 Metropolis–Hastings algorithm 162
8.5.2 Analysing the chains 163
8.5.3 Summarizing the distribution from the sample 165
8.5.4 Parameter transformations 167
8.6 R code 167
8.6.1 Demonstration of the Metropolis algorithm in one dimension 167
8.6.2 Implementation of the Metropolis algorithm 170

**9 Parameter estimation: Markov Chain Monte Carlo** 173
9.1 Fitting a straight line with unknown noise 173
9.1.1 Priors 175
9.1.2 Sampling the posterior 175
9.1.3 Posterior predictive distribution 178
9.1.4 R code for fitting a straight line 181
9.1.5 Discussion 186
9.1.6 Centring the data 188
9.2 Fitting a quadratic curve with unknown noise 190
9.3 A mixture model: fitting a straight line with an outlier model 193
9.3.1 Application and results 195
9.3.2 Discussion 197
9.4 Fitting curves with arbitrary error bars on both axes 199

**10 Frequentist hypothesis testing** 204
10.1 The principles of frequentist statistics and hypothesis testing 204
10.2 One-sample hypothesis tests 206
10.2.1 Gaussian z test 206
10.2.2 Student's t test 210
10.2.3 General t testing in R and frequentist confidence intervals 212
10.2.4 Summary: z test vs t test 215
10.3 Two-sample hypothesis tests 215

Downloaded from https://www.cambridge.org/core. Universitaet Heidelberg, on 06 Aug 2018 at 07:59:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108123891

10.3.1 Gaussian z test 215
10.3.2 Student's t test 216
10.4 Hypothesis testing in linear modelling 217
10.5 Goodness of fit and the $\chi^2$ distribution 218
10.5.1 The $\chi^2$ distribution 219
10.5.2 Using $\chi^2$ as a measure of goodness of fit 220
10.5.3 Derivation of the $\chi^2$ distribution 221
10.6 Issues with frequentist hypothesis testing 222

**11 Model comparison** 225
11.1 Bayesian model comparison 225
11.1.1 The evidence as a marginal likelihood 227
11.2 Example of an analytic evidence calculation: is a coin fair? 227
11.3 Example of a numerical evidence calculation: is there evidence for a non-zero gradient? 231
11.3.1 R code 235
11.3.2 A frequentist hypothesis testing approach 237
11.4 Comparing Gaussians (or other distributions) 238
11.5 How the evidence accounts for model complexity 239
11.6 Other ways of computing the evidence 242
11.6.1 The cross-validation likelihood 243
11.7 Other measures for model comparison: AIC and BIC 245
11.8 The stopping problem 248
11.8.1 The frequentist approach(es) 249
11.8.2 The Bayesian resolution 250
11.9 Issues with Bayesian model comparison 252

**12 Dealing with more complicated problems** 255
12.1 Cross-validation 255
12.2 Regularization in regression 260
12.2.1 Regularization 260
12.2.2 Ridge regression 261
12.2.3 Probabilistic interpretation of regularization 265
12.3 Regression with basis functions 266
12.3.1 Splines 266
12.3.2 Smoothing splines 269
12.3.3 R code 270
12.4 Regression kernels 273
12.5 A non-parametric smoothing problem 277
12.6 Numerical optimization (mode finding) 280
12.7 Bootstrap resampling 282
*References* 289
*Index* 291

Downloaded from https://www.cambridge.org/core. Universitaet Heidelberg, on 06 Aug 2018 at 07:59:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108123891

Downloaded from https://www.cambridge.org/core. Universitaet Heidelberg, on 06 Aug 2018 at 07:59:04, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108123891