Notes: Practical Statistics for Physics & Astronomy

R.B. Metcalf

Alma Mater Studiorum - Universitá di Bologna

February 19, 2018

Contents

1 What is Probability? 5
- 1.1 Frequentist . . . 5
- 1.2 classical interpretation of probability . . . 5
- 1.3 subjective or Bayesian probability . . . 6
- 1.4 Quantum mechanical probability . . . 6
- 1.5 the rules of probability . . . 7

2 Some warm up problems 10
- 2.1 Rolling Dice . . . 10
- 2.2 Birthday Paradox . . . 11
- 2.3 Poker . . . 13
- 2.4 The Monty Hall Problem . . . 14

3 Probability distributions 15
- 3.1 properties of a probability distribution function (PDF) . . . 15
- 3.2 mean, median, mode 15
- 3.3 moment generating function . . . 17
  - 3.3.1 changing of variables . . . 17
- 3.4 Binomial and Bernoulli . . . 17
  - 3.4.1 drawing without replacement, the hypergeometric distribution . . . 19
- 3.5 Poisson distribution . . . 19
  - 3.5.1 as a limit of the binomial distribution . . . 20
- 3.6 Gaussian and normal . . . 21
- 3.7 central limit theorem . . . 22
  - 3.7.1 The distribution of the sum of independent random variables . . . 23
- 3.8 lognormal . . . 25
- 3.9 Power law distribution . . . 25
- 3.10 multivariant distributions . . . 26
- 3.11 multinomial distributions . . . 27
- 3.12 multivariant gaussian . . . 27
  - 3.12.1 conditional Gaussian distribution . . . 29
  - 3.12.2 marginalized Gaussian distribution . . . 29

3.12.3 combining two multivariant Gaussians . . . 30
3.13 χ^2 distribution . . . 30
3.14 student's t-distribution . . . 32

4 Sampling **33**
4.1 estimating the mean . . . 33
4.2 estimating the variance . . . 36
4.3 estimating the mean when the variance is known . . . 37
4.4 median . . . 38
4.5 extreme values . . . 39
4.6 quintile estimation . . . 39

5 The Bayesian method **40**
5.1 Posterior, likelihood, prior and evidence . . . 40
5.2 Parameter estimation . . . 41
5.2.1 example: Poisson radiation . . . 41
5.2.2 example: estimating mean . . . 42
5.2.3 example: estimating mean and variance . . . 44
5.3 Marginalization . . . 47
5.3.1 example: the mean without the variance . . . 47
5.4 Choice of prior . . . 48
5.4.1 example: Jeffreys prior . . . 49
5.4.2 example: radiation with Jeffreys prior . . . 50
5.5 Model selection . . . 50
5.5.1 Occam's factor . . . 51
5.5.2 example: detection with Poisson errors . . . 53
5.6 Calculating the evidence . . . 53
5.7 Example: A spectral line . . . 53
5.8 Example: The luminosity function . . . 53

6 Hypothesis testing & frequentist parameter fitting **54**
6.0.1 mean of two populations are the same . . . 54
6.0.2 the variance of two populations are the same . . . 54
6.1 linear fitting, regression . . . 54
6.2 p-values, χ^2, degrees of freedom, etc. . . . 55
6.3 principle components . . . 55
6.4 nonlinear fitting . . . 55
6.5 hypothesis testing . . . 55
6.5.1 example: test isotropy of sources . . . 55
6.6 The early stop problem & shortcoming of frequenist hypothesis testing . . . 55

7 Frequentist non-parametric tests **56**
7.1 Kolmogorov-Smirnov test . . . 56
7.2 rank statistics . . . 56
7.2.1 Wilcoxon test . . . 56
7.2.2 Spearman's correlation statistic . . . 56
7.3 bootstrap & jacknife sampling methods . . . 56

8 Frequentist vs Bayesian **56**
8.1 frequentist & Bayesian confidence levels . . . 56
8.1.1 Example: the highest redshift QSO . . . 56
8.2 numerical methods, Monte Carlo . . . 56
8.2.1 Creating a random number in a computer . . . 56
8.3 the stop problem . . . 56
8.4 When does a parameter count? . . . 56

9 Estimators **57**
9.1 bias . . . 57
9.1.1 Example: Eddington-Malmquist bias . . . 57
9.2 maximum likelihood estimator . . . 57
9.3 least squares estimator . . . 57
9.3.1 weighting data . . . 57
9.4 minimum variance estimator . . . 57

10 The Fisher matrix and information **58**
10.1 the Fisher matrix . . . 58
10.2 The Gaussian approximation . . . 58
10.2.1 marginalized error estimates . . . 58
10.3 Rao-Cramer inequality . . . 58
10.4 information content of data . . . 58
10.4.1 different definitions of information . . . 58
10.4.2 the maximum entropy principle for choosing a distribution . . . 58
10.4.3 Kullback–Leibler divergence (information gain) . . . 58
10.4.4 agreement/disagreement of data sets . . . 58
10.5 error forecasting . . . 58

11 Random Fields **59**
11.1 correlation function & power spectrum . . . 59
11.2 going between discrete and continuous variables . . . 59
11.3 Gaussian random fields . . . 59
11.4 Poisson noise . . . 59

12 Estimating correlation functions, power spectra & time-delays from irregularly sampled data **60**
12.1 power spectrum estimation . . . 60
12.2 data compression, Karhunen-Loeve modes . . . 60

13 Image reconstruction and map making **61**
13.1 Wiener filtering . . . 61
13.2 maximum entropy . . . 61
13.3 other filters & estimators . . . 61

14 Numerical methods for the Bayesian Inference Problem **62**
14.1 finding maximum likelihood . . . 62
14.2 Markov Chain Monte Carlo (MCMC) . . . 62
14.3 nested sampling & other strategies . . . 62
14.4 meaning & calculation of evidence . . . 62

14.5 displaying results graphically 62

15 Classification and regression from a machine learning prospective **63**
15.1 prediction vs inference 63
15.2 dividing objects into classes objectively: unsupervised learning 63
15.3 logistic regression and classification: supervised learning 63
15.4 artificial neural networks & convolutional neural networks in astronomy 63

A Matrix basics **64**

B notation **65**

C Some Useful Integrals and mathematical definitions **65**