

Notes: Practical Statistics for Physics & Astronomy

R.B. Metcalf

Alma Mater Studiorum - Università di Bologna

February 15, 2018

Contents

1	What is Probability?	5
1.1	Frequentist	5
1.2	Bayesian	5
1.3	Quantum mechanical probability	5
1.4	the rules of probability	6
2	Some warm up problems	7
2.1	Rolling Dice	8
2.2	Birthday Paradox	9
2.3	Poker	10
2.4	The Monty Hall Problem	12
3	Probability distributions	13
3.1	properties of a probability distribution function (PDF)	13
3.2	mean, median, mode	13
3.3	moment generating function	15
3.3.1	changing of variables	15
3.4	Binomial and Bernoulli	15
3.4.1	drawing without replacement, the hypergeometric distribution	17
3.5	Poisson distribution	17
3.5.1	as a limit of the binomial distribution	18
3.6	Gaussian and normal	19
3.7	central limit theorem	20
3.7.1	The distribution of the sum of independent random variables	21
3.8	lognormal	23
3.9	Power law distribution	23
3.10	multivariate distributions	24
3.11	multinomial distributions	25
3.12	multivariate gaussian	25
3.12.1	conditional Gaussian distribution	27
3.12.2	marginalized Gaussian distribution	27
3.12.3	combining two multivariate Gaussians	28

3.13	χ^2 distribution	28
3.14	student's t-distribution	30
4	Sampling	31
4.1	estimating the mean	31
4.2	estimating the variance	34
4.3	estimating the mean when the variance is known	35
4.4	median	36
4.5	extreme values	37
4.6	quintile estimation	37
5	The Bayesian method	38
5.1	Posterior, likelihood, prior and evidence	38
5.2	Parameter estimation	39
5.2.1	example:	39
5.2.2	example:	40
5.2.3	example:	42
5.3	Marginalization	45
5.3.1	example:	45
5.4	Choice of prior	46
5.4.1	example:	47
5.4.2	example:	48
5.5	Model selection	48
5.5.1	Occam's factor	49
5.5.2	example:	51
5.6	Calculating the evidence	51
5.7	Example: A spectral line	51
5.8	Example: The luminosity function	51
6	Bayesian inference with Gaussian distributions & Gaussian approximations	52
7	Hypothesis testing & frequentist parameter fitting	53
7.0.1	mean of two populations are the same	53
7.0.2	the variance of two populations are the same	53
7.1	linear fitting, regression	53
7.2	p-values, χ^2 , degrees of freedom, etc.	54
7.3	principle components	54
7.4	nonlinear fitting	54
7.5	hypothesis testing	54
7.5.1	example: test isotropy of sources	54
7.6	The early stop problem & shortcoming of frequentist hypothesis testing	54
8	Frequentist non-parametric tests	55
8.1	Kolmogorov-Smirnov test	55
8.2	rank statistics	55
8.2.1	Wilcoxon test	55
8.2.2	Spearman's correlation statistic	55
8.3	bootstrap & jackknife sampling methods	55

9	Frequentist vs Bayesian	55
9.1	frequentist & Bayesian confidence levels	55
9.1.1	Example: the highest redshift QSO	55
9.2	numerical methods, Monte Carlo	55
9.2.1	Creating a random number in a computer	55
10	Estimators	56
10.1	bias	56
10.1.1	Example: Eddington-Malmquist bias	56
10.2	maximum likelihood estimator	56
10.3	least squares estimator	56
10.3.1	weighting data	56
10.4	minimum variance estimator	56
11	The Fisher matrix and information	57
11.1	the Fisher matrix	57
11.1.1	the Gaussian case	57
11.1.2	marginalized error estimates	57
11.2	Rao-Cramer inequality	57
11.3	information content of data	57
11.3.1	different definitions of information	57
11.3.2	Kullback–Leibler divergence (information gain)	57
11.3.3	agreement/disagreement of data sets	57
11.4	error forecasting	57
12	Random Fields	58
12.1	correlation function & power spectrum	58
12.2	going between discrete and continuous variables	58
12.3	Gaussian random fields	58
12.4	Poisson noise	58
13	Estimating correlation functions, power spectra & time-delays from irregularly sampled data	59
13.1	power spectrum estimation	59
13.2	data compression, Karhunen-Loeve modes	59
14	Image reconstruction and map making	60
14.1	Wiener filtering	60
14.2	maximum entropy	60
14.3	other filters & estimators	60
15	Numerical methods for the Bayesian Inference Problem	61
15.1	finding maximum likelihood	61
15.2	Markov Chain Monte Carlo (MCMC)	61
15.3	nested sampling & other strategies	61
15.4	meaning & calculation of evidence	61
15.5	displaying results graphically	61