

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Practical Statistics for Astronomers, Second Edition

Astronomy needs statistical methods to interpret data, but statistics is a many-faceted subject that is difficult for non-specialists to access. This handbook helps astronomers analyze the complex data and models of modern astronomy.

This Second Edition has been revised to feature many more examples using Monte Carlo simulations, and now also includes Bayesian inference, Bayes factors and Markov chain Monte Carlo integration. Chapters cover basic probability, correlation analysis, hypothesis testing, Bayesian modelling, time series analysis, luminosity functions and clustering. Exercises at the end of each chapter guide readers through the techniques and tests necessary for most observational investigations. The data tables, solutions to problems, and other resources are available online at www.cambridge.org/9780521732499.

Bringing together the most relevant statistical and probabilistic techniques for use in observational astronomy, this handbook is a practical manual for advanced undergraduate and graduate students and professional astronomers.

JASPER V. WALL is Adjunct Professor in the Department of Physics and Astronomy, University of British Columbia, and Visiting Professor at the University of Oxford, UK.

CHARLES R. JENKINS is a Research Scientist in Earth Sciences and Resource Engineering at the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia.

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Cover insets

Front cover:

A 500-ksec Chandra X-ray image of the Bullet cluster, tracing the bulk of the cluster baryons in the form of the hot intergalactic plasma (Clowe *et al.*, 2006). This gas is centralized and shocked in the collision between the two subclusters, while the dissipationless stars and dark matter (contours) retain the initial shape of their gravitational potentials. Reproduced by kind permission of the authors and the AAS.

Back cover:

(1) An all-sky image from the Planck mission, part of the early release data set. This image of the microwave sky has been synthesized from 12 months of data spanning the full frequency range of Planck, 30–857 GHz. Reproduced with the kind permission of ESA and the Planck Mission Team.

(2) Constraints on the mass densities of dark energy (Ω_Λ) and matter (Ω_m) provided by the supernovae type Ia Hubble diagram, measurements of baryon acoustic oscillations, and the power spectrum of the cosmic microwave background (CMB) (Suzuki *et al.*, 2011, see Chapter 11). The contours show the 68 per cent, 95 per cent and 99.7 per cent confidence regions. Reproduced with the kind permission of the authors, the Supernova Cosmology Project Team and the AAS.

(3) The power spectrum of the fluctuations in the CMB, as determined from the Wilkinson Microwave Anisotropy Probe (WMAP) 7-year data, together with data from other experiments (Komatsu *et al.*, 2010, see Chapter 11). Image reproduced courtesy of the authors, the WMAP team, the NASA GSFC public archive and the AAS.

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Cambridge Observing Handbooks for Research Astronomers

Today's professional astronomers must be able to adapt to use telescopes and interpret data at all wavelengths. This series is designed to provide them with a collection of concise, self-contained handbooks that covers the basic principles peculiar to observing in a particular spectra region, or to using a special technique or type of instrument. The book can be used as an introduction to the subject and as a handy reference for use at the telescope, or in the office.

Series editors

Professor Richard Ellis, Department of Astronomy, *California Institute of Technology*

Professor Steve Kahn, Department of Physics, *Stanford University*

Professor George Rieke, Steward Observatory, *University of Arizona*, Tucson

Dr Peter B. Stetson, Herzberg Institute of Astrophysics, *Dominion Astrophysical Observatory*, Victoria, British Columbia

Books currently available in this series:

1. *Handbook of Infrared Astronomy*

I. S. Glass

4. *Handbook of Pulsar Astronomy*

D. R. Lorimer, M. Kramer

5. *Handbook of CCD Astronomy*, Second Edition

Steve B. Howell

6. *Introduction to Astronomical Photometry*, Second Edition

Edwin Budding, Osman Demircan

7. *Handbook of X-ray Astronomy*

Edited by Keith Arnaud, Randall Smith and Aneta Siemiginowska

8. *Practical Statistics for Astronomers*, Second Edition

J. V. Wall, C. R. Jenkins

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Practical Statistics for Astronomers

Second Edition

J. V. WALL

University of British Columbia, Vancouver, Canada

C. R. JENKINS

*Commonwealth Scientific and Industrial Research Organization
(CSIRO), Australia*



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press
978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition
J. V. Wall and C. R. Jenkins
Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Mexico City
Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521732499

© J. V. Wall and C. R. Jenkins 2012

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2012

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-73249-9 Paperback

Additional resources for this publication at www.cambridge.org/9780521732499

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

Contents

<i>Foreword to first edition</i>	<i>page</i> xiii
<i>Foreword to second edition</i>	xv
<i>Note on notation</i>	xvii
1 Decision	1
1.1 How is science done?	4
1.2 Probability; probability distributions	6
1.3 Bolt-on statistics?	7
1.4 Probability and statistics in inference: an overview of this book	8
1.5 How to use this book	17
2 Probability	20
2.1 What is probability?	21
2.2 Conditionality and independence	24
2.3 ... and Bayes' theorem	26
2.4 Probability distributions	33
2.5 Bayesian inferences with probability	46
2.6 Monte Carlo generators	48
3 Statistics and expectations	55
3.1 Statistics	55
3.2 What should we expect of our statistics?	59
3.3 Simple error analysis	61
3.4 Some useful statistics, and their distributions	66
3.5 Uses of statistics	68
4 Correlation and association	71
4.1 The fishing trip	71

viii	<i>Contents</i>	
	4.2 Testing for correlation	74
	4.3 Partial correlation	83
	4.4 But what next?	84
	4.5 Principal component analysis	86
5	Hypothesis testing	92
	5.1 Methodology of classical hypothesis testing	94
	5.2 Parametric tests: means and variances, t and F tests	96
	5.3 Non-parametric tests: single samples	106
	5.4 Non-parametric tests: two independent samples	113
	5.5 Summary, one- and two-sample non-parametric tests	119
	5.6 Statistical ritual	121
6	Data modelling and parameter estimation: basics	126
	6.1 The maximum-likelihood method	128
	6.2 The method of least squares: regression analysis	134
	6.3 The minimum chi-square method	138
	6.4 Weighting combinations of data	141
	6.5 Bayesian likelihood analysis	142
	6.6 Bootstrap and jackknife	147
7	Data modelling and parameter estimation: advanced topics	151
	7.1 Model choice and Bayesian evidence	151
	7.2 Model simplicity and the Ockham factor	153
	7.3 The integration problem	154
	7.4 Pitfalls in model choice	157
	7.5 The Akaike and Bayesian information criteria	159
	7.6 Monte Carlo integration: doing the Bayesian integrals	160
	7.7 The Metropolis–Hastings algorithm	161
	7.8 Computation of the evidence by MCMC	166
	7.9 Models of models, and the combination of data sets	170
	7.10 Broadening the range of models, and weights	173
	7.11 Press and Kochanek’s method	176
	7.12 Median statistics	180
8	Detection and surveys	182
	8.1 Detection	183
	8.2 Catalogues and selection effects	189
	8.3 Luminosity functions	196
	8.4 Tests on luminosity functions	208
	8.5 Survival analysis	211
	8.6 The confusion limit	224

Contents ix

9	Sequential data – 1D statistics	230
9.1	Data transformations, the Karhunen–Loeve transform, and others	231
9.2	Fourier analysis	234
9.3	Filtering	241
9.4	Correlating	247
9.5	Unevenly sampled data	252
9.6	Wavelets	254
9.7	Detection difficulties: $1/f$ noise	257
10	Statistics of large-scale structure	262
10.1	Statistics on a spherical surface	262
10.2	Sky representation: projection and contouring	265
10.3	The sky distribution	268
10.4	Two-point angular correlation function	269
10.5	Counts in cells	275
10.6	The angular power spectrum	281
10.7	Galaxy distribution statistics: interpretation	287
11	Epilogue: statistics and our Universe	291
11.1	The galaxy universe	291
11.2	The weak lensing universe	301
11.3	The cosmic microwave background universe	307
	<i>Appendix A The literature</i>	316
	<i>Appendix B Statistical tables</i>	321
	<i>References</i>	335
	<i>Index</i>	347

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Cambridge University Press
978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition
J. V. Wall and C. R. Jenkins
Frontmatter
[More information](#)

In affectionate memory of Peter Scheuer (1930–2001)
mentor and friend
' $2 + 2 \simeq 5$ '

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Foreword to first edition

Peter Scheuer started this. In 1977 he walked into JVW's office in the Cavendish Lab and quietly asked for advice on what further material should be taught to the new intake of Radio Astronomy graduate students (that year including the hapless CRJ). JVW, wrestling with simple Chi-square testing at the time, blurted out '*They know nothing about practical statistics . . .*'. Peter left thoughtfully. A day later he returned. 'Good news! *The Management Board has decided that the students are going to have a course on practical statistics.*' Can I sit in, JVW asked innocently. 'Better news! *The Management Board has decided that you're going to teach it . . .*'.

So, for us, began the notion of practical statistics. A subject that began with gambling is not an arcane academic pursuit, but it is certainly subtle as well. It is fitting that Peter Scheuer was involved at the beginning of this (lengthy) project; his style of science exemplified both subtlety and pragmatism. We hope that we can convey something of both. If an echo of Peter's booming laugh is sometimes heard in these pages, it is because we both learned from him that a useful answer is often much easier – and certainly much more entertaining – than you at first think.

After the initial course, the material for this book grew out of various further courses, journal articles and the abundant personal experience that results from understanding just a little of any field of knowledge that counts Gauss and Laplace amongst its originators. More recently, the invigorating polemics of Jeffreys and Jaynes have been a great stimulus; although we have tried in this book not to engage too much with 'old, unhappy, far-off things / and battles long ago'.

Amongst today's practitioners of practical statistics, we have had valued discussions with Mark Birkinshaw, Phil Charles, Eric Feigelson, Pedro Ferreira, Paul Francis, Steve Gull, Dave Jauncey, Ofer Lahav, Robert Laing, Tony Lynas-Gray, Donald Lynden-Bell, Louis Lyons, Andrew Murray, John Peacock, Chris

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Pritchett, Prasenjit Saha and Adrian Webster. We are very grateful to Chris Blake, whose excellent D.Phil. thesis laid out clearly the interrelation of 2D descriptive statistics; and who has allowed us to borrow extensively from this opus. CRJ particularly acknowledges the Bayesian convictions of the Real Time Decisions group at Schlumberger; Dave Hargreaves, Iain Tuddenham and Tim Jervis. Try betting lives on your interpretation of the Kolomogorov axioms.

JVW is indebted to the Astrophysics Department of the University of Oxford for the enjoyable environment in which much of this was pulled together. The hospitality of the Department Heads – Phil Charles and then Joe Silk – is greatly appreciated; the stimulation, kindness, technical support and advice of colleagues there has been invaluable. Jenny Wall gave total support and encouragement throughout; the writing benefited greatly from the warmth and happiness of her companionship.

CRJ wishes to acknowledge the support of Schlumberger Cambridge Research for the writing of this book, as part of its ‘Personal Research Time’ initiative. The encouragement of the lab’s director, Mike Sheppard, catalysed its completion. Programme manager Ashley Johnson created the necessary space in a busy research group. Fiona Hall listened, helped with laughter through the long period of gestation and took time out from many pressing matters to support that final burst of writing.

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Foreword to second edition

Teaching is highly educational for teachers. Teaching from the first edition revealed to us how much students enjoyed Monte Carlo methods, and the ability with such methods to test and to check every derivation, test, procedure or result in the book. Thus, a change in the second edition is to introduce Monte Carlo as early as possible (Chapter 2). Teaching also revealed to us areas in which we assumed too much (and too little). We have therefore aimed for some smoothing of learning gradients where slope changes have appeared to be too sudden. Chapters 6 and 7 substantially amplify our previous treatments of Bayesian hypothesis testing/modelling, and include much more on model choice and Markov chain Monte Carlo (MCMC) analysis. Our previous chapter on 2D (sky distribution) analysis has been significantly revised. We have added a final chapter sketching the application of statistics to some current areas of astrophysics and cosmology, including galaxy formation and large-scale structure, weak gravitational lensing, and the cosmological microwave background (CMB) radiation.

We received very helpful comments from anonymous referees whom CUP consulted about our proposals for the second edition. These reviewers requested that we keep the book (a) practical and (b) concise and – *small*, or ‘back-packable’, as one of them put it. We have additional colleagues to thank either for further discussions, finding errata or because we just plain missed them from our first edition list: Matthew Colless, Jim Condon, Mike Disney, Alan Heavens, Martin Hendry, Jim Moran, Douglas Scott, Robert Smith and Malte Tewes. Jonathan Benjamin, Chris Blake, Adam Moss, John Peacock and Sanaz Vafaei provided valuable input on additional material in this second edition.

During preparation of the second edition, JVW is grateful to the National Science and Engineering Council of Canada for support, and to the Physics

and Astronomy Department of the University of British Columbia for its hospitality. Matt Wall provided JVW with welcome and expert computer-system assistance.

As before, our greatest debt of gratitude is to our partners Jenny Wall and Fiona Hall for their unflagging understanding and support.

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)

Note on notation

Here are some of the symbols used in the mathematical parts of this book. The list is not complete, but does include notation of more than localized interest. Some symbols are used with different meanings in different parts of the book, but in context there should be no possibility of confusion.

a_{lm} : coefficients of a spherical harmonic expansion

\mathcal{B} : Bayes factor

C : usually the covariance (or error) matrix, characterizing a multivariate Gaussian

c_l : coefficients of the angular power spectrum

$\text{cov}[x, y]$: covariance of two random variables x and y .

D : Kolmogorov–Smirnov test statistic

$E[X]$: expectation or ensemble average. Also denoted $\langle X \rangle$

f, F : probability density distributions and cumulative probability density distributions, respectively; in Chapter 9, Fourier pairs

\mathcal{F} : variable distributed according to the F distribution

\mathcal{H} : Hessian matrix

H_0, H_1 : null hypothesis and alternative hypothesis

\mathcal{K} : Kaplan–Meier estimator

L : intrinsic luminosity

\mathcal{L} : likelihood

ML : maximum likelihood, maximum likelihood method

MLE : maximum likelihood estimator

$N(S)$: flux density distribution, or source count

- \mathcal{P} : posterior odds
- $P(N)$: counts-in-cells probability of finding N objects in a cell
- P_l : Legendre polynomials
- $\text{prob}(\dots)$: probability of the indicated event. In the case of a continuous variable, the probability density
- $\text{prob}(A \mid B)$: probability of A , given B
- R : distance
- r : product-moment coefficient
- \mathcal{R} : Rayleigh test statistic
- S : mean square deviation of a set of data; in Chapter 8, flux density
- S : test statistic for a particular orientation of the principal axis of the orientation matrix
- S_e : sample cumulative distribution, as used in the Kolmogorov–Smirnov test
- t : variable distributed according to the t distribution
- U : Wilcoxon–Mann–Whitney test statistic
- V, V_{\max} : volume contained within R ; the maximum volume, corresponding to the greatest distance consistent with an object still appearing in a catalogue
- $\text{var}[x]$: variance of a random variable x
- $w(\theta)$: two-point angular correlation function
- \bar{X} : sample average of a set of data
- X_1, X_2, \dots : usually a specific set of data; instances of possible data, denoted x . We try to keep to this distinction by using upper case for particular values and lower case for algebraic variables (although not with Greek letters, or statistics like t where lower case is standard)
- y, z : excess variance and skewness of clustered counts-in-cells
- Y_{lm} : spherical harmonics
- $\vec{\alpha}$: vector, usually a vector of parameters
- Γ : Gehan test statistic
- η : luminosity distribution
- κ : Kendall test statistic
- μ, σ : usually the mean and standard deviation of a Gaussian distribution; μ may also be the parameter of a Poisson distribution
- μ_n : n th central moment of a distribution

Note on notation

xix

ρ : covariance coefficient of a bivariate Gaussian; in Chapter 8, the luminosity function

$\zeta(\theta, \phi)$: surface density of objects on the sky

σ_s : sample standard deviation

ϕ : space distribution

χ^2 : variable distributed according to the χ^2 distribution

Cambridge University Press

978-0-521-73249-9 - Practical Statistics for Astronomers: Second Edition

J. V. Wall and C. R. Jenkins

Frontmatter

[More information](#)
