

Springer Series in Statistics

Series editors

Peter Bickel, Berkeley, CA, USA

Peter Diggle, Lancaster, UK

Stephen E. Fienberg, Pittsburgh, PA, USA

Ursula Gather, Dortmund, Germany

Ingram Olkin, Stanford, CA, USA

Scott Zeger, Baltimore, MD, USA

For further volumes:

<http://www.springer.com/series/692>

Robert E. Kass · Uri T. Eden
Emery N. Brown

Analysis of Neural Data

Robert E. Kass
Carnegie Mellon University
Pittsburgh, PA
USA

Emery N. Brown
Massachusetts Institute of Technology
Cambridge, MA
USA

Uri T. Eden
Boston University
Boston, MA
USA

ISSN 0172-7397
ISBN 978-1-4614-9601-4
DOI 10.1007/978-1-4614-9602-1
Springer New York Heidelberg Dordrecht London

ISSN 2197-568X (electronic)
ISBN 978-1-4614-9602-1 (eBook)

Library of Congress Control Number: 2013955054

© Springer Science+Business Media New York 2014, Corrected at 2nd printing 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To our families

Preface

This book serves as a guide and reference for anyone who wishes to understand analysis of neural data generated from studies that range from molecules, to circuits, to systems, to behavior.

Its origins may be traced to the decision by two of us (E.N.B. and R.E.K.), in 1998, to write a review article on statistical analysis of spike train data. Shortly after commencing we realized that some of the methods we thought we ought to be reviewing had, in fact, not yet been developed. After we and others rectified this situation, we published a pair of reviews (Brown et al. 2004; Kass et al. 2005). During this time we also broadened our interests to other experimental modalities, such as neuroimaging, and we began teaching workshops and semester-long courses on statistical methods for neuroscience. In addition, we met the third author of this book (U.E.), who came to share our interests in research and pedagogy (and who pursued his Ph.D. thesis under the guidance of E.N.B.).

It became clear that a book on this subject was desperately needed, and we agreed to write one. While this turned into a longer project than we anticipated, numerous research collaborations, conversations with colleagues at meetings, and extensive comments from students gave us many insights into the content and presentation of the principles and techniques that evolved to form this volume. We feel we are much wiser than when we started, and we hope we have succeeded in imparting a good deal of what we have learned in the process.

Some readers may expect a book organized by type of neural data. We decided, instead, to organize by analysis, with each chapter devoted to broadly categorized statistical concepts described succinctly in section headings that are available in the extended version of the table of contents. Each chapter, however, also contains multiple examples of the way these analytical ideas have been used in the brain sciences: there are more than 100 such examples throughout the book, and they are indexed. A reader wishing to see how we have discussed fMRI data, for instance, should start with the example index. More specific organizational guidelines are given in [Chapter 1](#).

The book is intended as either a reference, or a text. R.E.K. has used preliminary versions of the manuscript in classes populated by graduate students of varying backgrounds, ranging from biologists with minimal mathematical knowledge, who were looking for conceptual understanding, to engineers, who needed to see derivations. We opted to try to satisfy both kinds of audiences.

An appendix is provided as a reminder of key mathematical ideas, and derivations are often marked as optional by indenting them. To those who wish to use the book as a text, R.E.K. would suggest the following ordering of topics:

Part I (Elementary Statistics): Chapters 1–7, 10, 12.1–12.4, 13.1.

Part II (Basic Statistical Theory): Chapters 8, 9, 11, 12.5, 13.2–13.4.

Part III (Advanced Topics): Selections from Chapters 14–19.

In his experience, Parts I and II take approximately 12 and 7 classes, respectively.

Many readers will want to see computer code for the methods we have described. We ourselves used both Matlab and R to produce figures. Although we decided not to inject Matlab or R code into the body of the book, we have put code up on our the website <http://www.stat.cmu.edu/~kass/KEB>.

In addition to the many colleagues and students who made suggestions along the way, including those who are acknowledged within the text, we are indebted to Spencer Koerner, who helped clean up and create much code and many figures, Patrick Foley, who created the website, Heidi Sestrich, who fixed numerous defects in our LATEX, and Matthew Marler, who read the whole manuscript carefully and provided extremely helpful comments. We are also grateful to Elan Cohen and Ryan Sieberg, who each created several figures.

Robert E. Kass
Uri T. Eden
Emery N. Brown

Short Table of Contents

Preface	vii
1 Introduction	1
2 Exploring Data	23
3 Probability and Random Variables	37
4 Random Vectors	71
5 Important Probability Distributions	105
6 Sequences of Random Variables	137
7 Estimation and Uncertainty	149
8 Estimation in Theory and Practice	179
9 Propagation of Uncertainty and the Bootstrap	221
10 Models, Hypotheses, and Statistical Significance	247
11 General Methods for Testing Hypotheses	287
12 Linear Regression	309
13 Analysis of Variance	361
14 Generalized Linear and Nonlinear Regression	391

15	Nonparametric Regression	413
16	Bayesian Methods.	439
17	Multivariate Analysis	491
18	Time Series	513
19	Point Processes.	563
	Appendix: Mathematical Background	605
	References	623
	Example Index	635
	Index	639

Contents

1	Introduction	1
1.1	Data Analysis in the Brain Sciences	1
1.1.1	Appropriate analytical strategies depend crucially on the purpose of the study and the way the data are collected.	3
1.1.2	Many investigations involve a response to a stimulus or behavior.	6
1.2	The Contribution of Statistics.	8
1.2.1	Statistical models describe regularity and variability of data in terms of probability distributions.	9
1.2.2	Statistical models are used to express knowledge and uncertainty about a signal in the presence of noise, via inductive reasoning.	13
1.2.3	Statistical models may be either parametric or nonparametric.	14
1.2.4	Statistical model building is an iterative process that incorporates assessment of fit and is preceded by exploratory analysis.	17
1.2.5	All models are wrong, but some are useful.	17
1.2.6	Statistical theory is used to understand the behavior of statistical procedures under various probabilistic assumptions.	19
1.2.7	Important data analytic ideas are sometimes implemented in many different ways.	20
1.2.8	Measuring devices often pre-process the data.	20
1.2.9	Data analytic techniques are rarely able to compensate for deficiencies in data collection.	21
1.2.10	Simple methods are essential.	21
1.2.11	It is convenient to classify data into several broad types.	21

2	Exploring Data	23
2.1	Describing Central Tendency and Variation	23
2.1.1	Alternative displays and summaries provide different views of the data	23
2.1.2	Exploratory methods can be sophisticated	26
2.2	Data Transformations	28
2.2.1	Positive values are often transformed by logarithms	28
2.2.2	Non-logarithmic transformations are sometimes applied	33
3	Probability and Random Variables	37
3.1	The Calculus of Probability	38
3.1.1	Probabilities are defined on sets of uncertain events	38
3.1.2	The conditional probability $P(A B)$ is the probability that A occurs given that B occurs	40
3.1.3	Probabilities multiply when the associated events are independent	41
3.1.4	Bayes' theorem for events gives the conditional probability $P(A B)$ in terms of the conditional probability $P(B A)$	42
3.2	Random Variables	46
3.2.1	Random variables take on values determined by events	47
3.2.2	Distributions of random variables are defined using cumulative distribution functions and probability density functions, from which theoretical means and variances may be computed	48
3.2.3	Continuous random variables are similar to discrete random variables	52
3.2.4	The hazard function provides the conditional probability of an event, given that it has not yet occurred	61
3.2.5	The distribution of a function of a random variable is found by the change of variables formula	62
3.3	The Empirical Cumulative Distribution Function	64
3.3.1	P-P and Q-Q plots provide graphical checks for gross departures from a distributional form	65
3.3.2	Q-Q and P-P plots may be used to judge the effectiveness of transformations	69

4	Random Vectors.	71
4.1	Two or More Random Variables	71
4.1.1	The variation of several random variables is described by their joint distribution.	73
4.1.2	Random variables are independent when their joint pdf is the product of their marginal pdfs.	75
4.2	Bivariate Dependence	76
4.2.1	The linear dependence of two random variables may be quantified by their correlation.	77
4.2.2	A bivariate normal distribution is determined by a pair of means, a pair of standard deviations, and a correlation coefficient.	82
4.2.3	Conditional probabilities involving random variables are obtained from conditional densities.	84
4.2.4	The conditional expectation $E(Y X = x)$ is called the regression of Y on X .	85
4.3	Multivariate Dependence.	90
4.3.1	The mean of a random vector is a vector and its variance is a matrix.	90
4.3.2	The dependence of two random vectors may be quantified by mutual information.	92
4.3.3	Bayes' theorem for random vectors is analogous to Bayes' theorem for events.	98
4.3.4	Bayes classifiers are optimal.	99
5	Important Probability Distributions	105
5.1	Bernoulli Random Variables and the Binomial Distribution.	105
5.1.1	Bernoulli random variables take values 0 or 1.	105
5.1.2	The binomial distribution results from a sum of independent and homogeneous Bernoulli random variables.	106
5.2	The Poisson Distribution	110
5.2.1	The Poisson distribution is often used to describe counts of binary events.	110
5.2.2	For large n and small p the binomial distribution is approximately the same as Poisson.	113
5.2.3	The Poisson distribution results when the binary events are independent.	115
5.3	The Normal Distribution	116
5.3.1	Normal random variables are within 1 standard deviation of their mean with probability $2/3$; they are within 2 standard deviations of their mean with probability .95.	116

5.3.2	Binomial and Poisson distributions are approximately normal, for large n or large λ	118
5.4	Some Other Common Distributions	119
5.4.1	The multinomial distribution extends the binomial to multiple categories.	119
5.4.2	The exponential distribution is used to describe waiting times without memory.	120
5.4.3	Gamma distributions are sums of exponentials.	123
5.4.4	Chi-squared distributions are special cases of gamma distributions.	124
5.4.5	The beta distribution may be used to describe variation on a finite interval.	124
5.4.6	The inverse Gaussian distribution describes the waiting time for a threshold crossing by Brownian motion.	125
5.4.7	The t and F distributions are defined from normal and chi-squared distributions.	128
5.5	Multivariate Normal Distributions	129
5.5.1	A random vector is multivariate normal if linear combinations of its components are univariate normal.	129
5.5.2	The multivariate normal pdf has elliptical contours, with probability density declining according to a χ^2 pdf.	130
5.5.3	If X and Y are jointly multivariate normal then the conditional distribution of Y given X is multivariate normal.	132
6	Sequences of Random Variables	137
6.1	Random Sequences and the Sample Mean	137
6.1.1	The standard deviation of the sample mean decreases as $1/\sqrt{n}$	139
6.1.2	Random sequences may converge according to several distinct criteria.	142
6.2	The Law of Large Numbers.	143
6.2.1	As the sample size n increases, the sample mean converges to the theoretical mean.	143
6.2.2	The empirical cdf converges to the theoretical cdf.	144
6.3	The Central Limit Theorem	145
6.3.1	For large n , the sample mean is approximately normally distributed.	145
6.3.2	For large n , the multivariate sample mean is approximately multivariate normal.	147

7	Estimation and Uncertainty	149
7.1	Fitting Statistical Models	149
7.2	The Problem of Estimation	151
7.2.1	The method of moments uses the sample mean and variance to estimate the theoretical mean and variance.	153
7.2.2	The method of maximum likelihood maximizes the likelihood function, which is defined up to a multiplicative constant.	154
7.3	Confidence Intervals	158
7.3.1	For scientific inference, estimates are useless without some notion of precision.	158
7.3.2	Estimation of a normal mean is a paradigm case.	160
7.3.3	For non-normal observations the central limit theorem may be invoked.	162
7.3.4	A large-sample confidence interval for μ is obtained using the standard error s/\sqrt{n} .	162
7.3.5	Standard errors lead immediately to confidence intervals.	164
7.3.6	Estimates and standard errors should be reported to two digits in the standard error.	169
7.3.7	Appropriate sample sizes may be determined from desired size of standard error.	169
7.3.8	Confidence assigns probability indirectly, making its interpretation subtle.	170
7.3.9	Bayes' theorem may be used to assess uncertainty.	173
7.3.10	For small samples it is customary to use the t distribution instead of the normal.	176
8	Estimation in Theory and Practice	179
8.1	Mean Squared Error	181
8.1.1	Mean squared error is bias squared plus variance.	181
8.1.2	Mean squared error may be evaluated by computer simulation of pseudo-data.	186
8.1.3	In estimating a theoretical mean from observations having differing variances a weighted mean should be used, with weights inversely proportional to the variances.	190
8.1.4	Decision theory often uses mean squared error to represent risk.	195

8.2	Estimation in Large Samples	196
8.2.1	In large samples, an estimator should be very likely to be close to its estimand.	196
8.2.2	In large samples, the precision with which a parameter may be estimated is bounded by Fisher information.	196
8.2.3	Estimators that minimize large-sample variance are called efficient.	200
8.3	Properties of ML Estimators	202
8.3.1	In large samples, ML estimation is optimal.	202
8.3.2	The standard error of the MLE is obtained from the second derivative of the loglikelihood function.	203
8.3.3	In large samples, ML estimation is approximately Bayesian.	207
8.3.4	MLEs transform along with parameters.	208
8.3.5	Under normality, ML produces the weighted mean.	209
8.4	Multiparameter Maximum Likelihood.	209
8.4.1	The MLE solves a set of partial differential equations.	210
8.4.2	Least squares may be viewed as a special case of ML estimation.	212
8.4.3	The observed information is the negative of the matrix of second partial derivatives of the loglikelihood function, evaluated at $\hat{\theta}$	213
8.4.4	When using numerical methods to implement ML estimation, some care is needed.	214
8.4.5	MLEs are sometimes obtained with the EM algorithm.	215
8.4.6	Maximum likelihood may produce bad estimates.	219
9	Propagation of Uncertainty and the Bootstrap.	221
9.1	Propagation of Uncertainty	223
9.1.1	Simulated observations from the distribution of the random variable X produce simulated observations from the distribution of the random variable $Y = f(X)$	223
9.1.2	In large samples, transformations of consistent and asymptotically normal random variables become approximately linear.	229
9.2	The Bootstrap	237
9.2.1	The bootstrap is a general method of assessing uncertainty.	237

9.2.2	The parametric bootstrap draws pseudo-data from an estimated parametric distribution.	239
9.2.3	The nonparametric bootstrap draws pseudo-data from the empirical cdf.	241
9.3	Discussion of Alternative Methods	245
10	Models, Hypotheses, and Statistical Significance	247
10.1	Chi-Squared Statistics	248
10.1.1	The chi-squared statistic compares model-fitted values to observed values.	249
10.1.2	For multinomial data, the chi-squared statistic follows, approximately, a χ^2 distribution.	250
10.1.3	The rarity of a large chi-squared is judged by its p -value.	253
10.1.4	Chi-squared may be used to test independence of two traits.	254
10.2	Null Hypotheses	256
10.2.1	Statistical models are often considered null hypotheses.	256
10.2.2	Null hypotheses sometimes specify a particular value of a parameter within a statistical model.	257
10.2.3	Null hypotheses may also specify a constraint on two or more parameters.	257
10.3	Testing Null Hypotheses	258
10.3.1	The hypothesis $H_0: \mu = \mu_0$ for a normal random variable is a paradigm case.	258
10.3.2	For large samples the hypothesis $H_0: \theta = \theta_0$ may be tested using the ratio $(\hat{\theta} - \theta_0)/SE(\hat{\theta})$	260
10.3.3	For small samples it is customary to test $H_0: \mu = \mu_0$ using a t statistic.	262
10.3.4	For two independent samples, the hypothesis $H_0: \mu_1 = \mu_2$ may be tested using the t -ratio.	264
10.3.5	Computer simulation may be used to find p -values.	266
10.3.6	The Rayleigh test can provide evidence against a uniform distribution of angles.	268
10.3.7	The fit of a continuous distribution may be assessed with the Kolmogorov-Smirnov test.	270
10.4	Interpretation and Properties of Tests	271
10.4.1	Statistical tests should have the correct probability of falsely rejecting H_0 , at least approximately.	271
10.4.2	A confidence interval for θ may be used to test $H_0: \theta = \theta_0$	274

10.4.3	Statistical tests are evaluated in terms of their probability of correctly rejecting H_0	276
10.4.4	The performance of a statistical test may be displayed by the ROC curve.	278
10.4.5	The p -value is not the probability that H_0 is true.. . . .	281
10.4.6	Borderline p -values are especially worrisome with low power.	282
10.4.7	The p -value is conceptually distinct from type one error..	283
10.4.8	A non-significant test does not, by itself, indicate evidence in support of H_0	283
10.4.9	One-tailed tests are sometimes used..	285
11	General Methods for Testing Hypotheses	287
11.1	Likelihood Ratio Tests	288
11.1.1	The likelihood ratio may be used to test $H_0: \theta = \theta_0$	288
11.1.2	P -values for the likelihood ratio test of $H_0: \theta = \theta_0$ may be obtained from the χ^2 distribution or by simulation.	290
11.1.3	The likelihood ratio test of $H_0: (\omega, \theta) = (\omega, \theta_0)$ plugs in the MLE of ω , obtained under H_0	291
11.1.4	The likelihood ratio test reproduces, exactly or approximately, many commonly-used significance tests.	293
11.1.5	The likelihood ratio test is optimal for simple hypotheses.	293
11.1.6	To evaluate alternative non-nested models the likelihood ratio statistic may be adjusted for parameter dimensionality..	294
11.2	Permutation and Bootstrap Tests	297
11.2.1	Permutation tests consider all possible permutations of the data that would be consistent with the null hypothesis.	297
11.2.2	The bootstrap samples with replacement..	300
11.3	Multiple Tests	301
11.3.1	When multiple independent data sets are used to test the same hypothesis, the p -values are easily combined.	301
11.3.2	When multiple hypotheses are considered, statistical significance should be adjusted..	302

12 Linear Regression.	309
12.1 The Linear Regression Model	310
12.1.1 Linear regression assumes linearity of $f(x)$ and independence of the noise contributions at the various observed x values.	315
12.1.2 The relative contribution of the linear signal to the total response variation is summarized by R^2 .	316
12.1.3 Theory shows that if the model were correct then the least-squares estimate would be likely to be accurate for large samples.	318
12.2 Checking Assumptions	319
12.2.1 Residuals should represent unstructured noise.	319
12.2.2 Graphical examination of (x, y) data can yield crucial information.	320
12.2.3 Failure of independence among the errors can have substantial consequences.	321
12.3 Evidence of a Linear Trend	323
12.3.1 Confidence intervals for slopes are based on SE, according to the general formula.	323
12.3.2 Evidence in favor of a linear trend can be obtained from a t -test concerning the slope.	325
12.3.3 The fitted relationship may not be accurate outside the range of the observed data.	326
12.4 Correlation and Regression	327
12.4.1 The correlation coefficient is determined by the regression coefficient and the standard deviations of x and y .	327
12.4.2 Association is not causation.	328
12.4.3 Confidence intervals for ρ may be based on a transformation of r .	328
12.4.4 When noise is added to two variables, their correlation diminishes.	330
12.5 Multiple Linear Regression	332
12.5.1 Multiple regression estimates the linear relationship of the response with each explanatory variable, while adjusting for the other explanatory variables.	334
12.5.2 Response variation may be decomposed into signal and noise sums of squares.	335
12.5.3 Multiple regression may be formulated concisely using matrices.	339
12.5.4 The linear regression model applies to polynomial regression and cosine regression.	346

12.5.5	Effects of correlated explanatory variables cannot be interpreted separately..	350
12.5.6	In multiple linear regression interaction effects are often important.	352
12.5.7	Regression models with many explanatory variables often can be simplified.	353
12.5.8	Multiple regression can be treacherous.	358
13	Analysis of Variance.	361
13.1	One-Way and Two-Way ANOVA	361
13.1.1	ANOVA is based on a linear model.	363
13.1.2	One-way ANOVA decomposes total variability into average group variability and average individual variability, which would be roughly equal under the null hypothesis.	365
13.1.3	When there are only two groups, the ANOVA F -test reduces to a t -test.	368
13.1.4	Two-way ANOVA assesses the effects of one factor while adjusting for the other factor.	369
13.1.5	When the variances are inhomogeneous across conditions a likelihood ratio test may be used.	371
13.1.6	More complicated experimental designs may be accommodated by ANOVA.	371
13.1.7	Additional analyses, involving multiple comparisons, may require adjustments to p -values.	372
13.2	ANOVA as Regression	374
13.2.1	The general linear model includes both regression and ANOVA models.	374
13.2.2	In multi-way ANOVA, interactions are often of interest.	377
13.2.3	ANOVA comparisons may be adjusted using analysis of covariance.	380
13.3	Nonparametric Methods	381
13.3.1	Distribution-free nonparametric tests may be obtained by replacing data values with their ranks.	382
13.3.2	Permutation and bootstrap tests may be used to test ANOVA hypotheses.	385
13.4	Causation, Randomization, and Observational Studies.	385
13.4.1	Randomization eliminates effects of confounding factors.	385
13.4.2	Observational studies can produce substantial evidence.	387

14	Generalized Linear and Nonlinear Regression	391
14.1	Logistic Regression, Poisson Regression, and Generalized Linear Models	392
14.1.1	Logistic regression may be used to analyze binary responses..	392
14.1.2	In logistic regression, ML is used to estimate the regression coefficients and the likelihood ratio test is used to assess evidence of a logistic-linear trend with x .	395
14.1.3	The logit transformation is one among many that may be used for binomial responses, but it is the most commonly applied.	398
14.1.4	The usual Poisson regression model transforms the mean λ to $\log \lambda$.	400
14.1.5	In Poisson regression, ML is used to estimate coefficients and the likelihood ratio test is used to examine trends.	401
14.1.6	Generalized linear models extend regression methods to response distributions from exponential families.	402
14.2	Nonlinear Regression	405
14.2.1	Nonlinear regression models may be fitted by least squares.	405
14.2.2	Generalized nonlinear models may be fitted using maximum likelihood.	409
14.2.3	In solving nonlinear optimization problems, good starting values are important, and it can be helpful to reparameterize.	411
15	Nonparametric Regression	413
15.1	Smoothers	414
15.1.1	Linear smoothers are fast.	415
15.1.2	For linear smoothers, the fitted function values are obtained via a “hat matrix,” and it is easy to apply propagation of uncertainty.	415
15.2	Basis Functions	416
15.2.1	Splines may be used to represent complicated functions..	418
15.2.2	Splines may be fit to data using linear models.	418
15.2.3	Splines are also easy to use in generalized linear models.	421
15.2.4	With regression splines, the number and location of knots controls the smoothness of the fit.	422

15.2.5	Smoothing splines are splines with knots at each x_i , but with reduced coefficients obtained by penalized ML.	423
15.2.6	A method called BARS chooses knot sets automatically, according to a Bayesian criterion. . . .	424
15.2.7	Spline smoothing may be used with multiple explanatory variables.	425
15.2.8	Alternatives to splines are often used in nonparametric regression.	427
15.3	Local Fitting	429
15.3.1	Kernel regression estimates $f(x)$ with a weighted mean defined by a pdf.	430
15.3.2	Local polynomial regression solves a weighted least squares problem with weights defined by a kernel.	432
15.3.3	Theoretical considerations lead to bandwidth recommendations for linear smoothers. .	434
15.4	Density Estimation	435
15.4.1	Kernels may be used to estimate a pdf.	435
15.4.2	Other nonparametric regression methods may be used to estimate a pdf.	436
16	Bayesian Methods.	439
16.1	Posterior Distributions.	440
16.1.1	Bayesian inference equates descriptive and epistemic probability.	441
16.1.2	Conjugate priors are convenient.	442
16.1.3	For exponential families with conjugate priors the posterior mean is a weighted combination of the MLE and the prior mean.	443
16.1.4	There is no compelling choice of prior distribution. . .	447
16.1.5	For large samples, posteriors are approximately normal and centered at the MLE.	448
16.1.6	Powerful methods exist for computing posterior distributions.	450
16.2	Latent Variables.	457
16.2.1	Hierarchical models produce estimates of related quantities that are pulled toward each other.	459
16.2.2	For hierarchical models, posterior distributions are often computed by Gibbs sampling.	466
16.2.3	Penalized regression may be viewed as Bayesian estimation.	469

16.2.4	State-space models allow parameters to evolve dynamically.	470
16.2.5	The Kalman filter may be used to estimate state variables for linear Gaussian state-space models.	473
16.3	Bayes Factors.	475
16.3.1	Bayes factors can provide evidence in favor of hypotheses.	477
16.3.2	Bayes factors provide an interpretation of scientific progress.	480
16.3.3	Bayes factors can be difficult to use when there is little information about unknown parameters.	481
16.3.4	Bayes factors can be used to calibrate p-values.	482
16.4	Derivations of Results on Latent Variables	482
17	Multivariate Analysis	491
17.1	Introduction	491
17.2	Multivariate Analysis of Variance	492
17.2.1	MANOVA provides a multivariate extension of ANOVA.	492
17.2.2	When the variance matrices across conditions are unequal, the likelihood ratio test may be applied.	496
17.3	Dimensionality Reduction	498
17.3.1	A variance matrix may be decomposed into principal components.	498
17.3.2	Methods other than PCA may be used to reduce dimensionality.	503
17.4	Classification and Clustering	505
17.4.1	Bayes classifiers for multivariate normal distributions take a simple form.	505
17.4.2	Bayes classifiers are not always practical.	506
17.4.3	Multivariate observations may be clustered into groups.	510
18	Time Series	513
18.1	Introduction	513
18.2	Time Domain and Frequency Domain.	518
18.2.1	Fourier analysis is one of the great achievements of mathematical science.	522
18.2.2	The periodogram is both a scaled representation of contributions to R^2 from harmonic regression and a scaled power function associated with the discrete Fourier transform of a data set.	525
18.2.3	Autoregressive models may be fitted by lagged regression.	530

18.3	The Periodogram for Stationary Processes	535
18.3.1	The periodogram may be considered an estimate of the spectral density function.	535
18.3.2	For large samples, the periodogram ordinates computed from a stationary time series are approximately independent of one another and chi-squared distributed.	537
18.3.3	Consistent estimators of the spectral density function result from smoothing the periodogram. . . .	539
18.3.4	Linear filters can be fast and effective.	541
18.3.5	Frequency information is limited by the sampling rate.	544
18.3.6	Tapering reduces the leakage of power from non-Fourier to Fourier frequencies.	546
18.3.7	Time-frequency analysis describes the evolution of rhythms across time.	548
18.4	Propagation of Uncertainty for Functions of the Periodogram	550
18.4.1	Confidence intervals and significance tests may be carried out by propagating the uncertainty from the periodogram.	550
18.4.2	Uncertainty about functions of time series may be obtained from time series pseudo-data.	552
18.5	Bivariate Time Series	553
18.5.1	The coherence $\rho_{XY}(\omega)$ between two series X and Y may be considered the correlation of their ω -frequency components.	555
18.5.2	In examining cross-correlation or coherence of two time series it is advisable first to pre-whiten the series.	557
18.5.3	Granger causality measures the linear predictability of one time series by another.	559
19	Point Processes.	563
19.1	Point Process Representations	566
19.1.1	A point process may be specified in terms of event times, inter-event intervals, or event counts.	566
19.1.2	A point process may be considered, approximately, to be a binary time series.	567
19.1.3	Point processes can display a wide variety of history-dependent behaviors.	568

19.2	Poisson Processes	570
19.2.1	Poisson processes are point processes for which event probabilities do not depend on occurrence or timing of past events.	570
19.2.2	Inhomogeneous Poisson processes have time-varying intensities.	573
19.3	Non-Poisson Point Processes	578
19.3.1	Renewal processes have i.i.d. inter-event waiting times.	578
19.3.2	The conditional intensity function specifies the joint probability density of spike times for a general point process.	582
19.3.3	The marginal intensity is the expectation of the conditional intensity.	584
19.3.4	Conditional intensity functions may be fitted using Poisson regression.	586
19.3.5	Graphical checks for departures from a point process model may be obtained by time rescaling. . . .	594
19.3.6	There are efficient methods for generating point process pseudo-data.	597
19.3.7	Spectral analysis of point processes requires care. . . .	599
19.4	Additional Derivations	601
	Appendix: Mathematical Background	605
	References	623
	Example Index	635
	Index	639