Chapman & Hall/CRC Interdisciplinary Statistics Series

Analysis of Capture-Recapture Data

 $1 - \chi_{c,j} = (1 - S_{c,j})\lambda_{c,j} + S_{c,j}\{1 - (1 - p_{c,j+1})\chi_{c,j+1}\}$

Rachel S. McCrea and Byron J. T. Morgan



Analysis of Capture-Recapture Data

CHAPMAN & HALL/CRC Interdisciplinary Statistics Series

Series editors: N. Keiding, B.J.T. Morgan, C.K. Wikle, P. van der Heijden

Published titles

DATA

AGE-PERIOD-COHORT ANALYSIS: NEW MODELS, METHODS, AND EMPIRICAL APPLICATIONS Y. Yang and K. C. Land

ANALYSIS OF CAPTURE-RECAPTURE R.S. McCrea and B.J.T. Morgan

AN INVARIANT APPROACH TO STATISTICAL ANALYSIS OF SHAPES

ASTROSTATISTICS

BAYESIAN ANALYSIS FOR POPULATION ECOLOGY

BAYESIAN DISEASE MAPPING: HIERARCHICAL MODELING IN SPATIAL EPIDEMIOLOGY, SECOND EDITION

BIOEQUIVALENCE AND STATISTICS IN CLINICAL PHARMACOLOGY

CLINICAL TRIALS IN ONCOLOGY, THIRD EDITION

CLUSTER RANDOMISED TRIALS

CORRESPONDENCE ANALYSIS IN PRACTICE, SECOND EDITION

DESIGN AND ANALYSIS OF QUALITY OF LIFE STUDIES IN CLINICAL TRIALS, SECOND EDITION

DYNAMICAL SEARCH

FLEXIBLE IMPUTATION OF MISSING DATA

GENERALIZED LATENT VARIABLE MODELING: MULTILEVEL, LONGITUDINAL, AND STRUCTURAL EQUATION MODELS S. Lele and J. Richtsmeier

G. Babu and E. Feigelson

R. King, B. J.T. Morgan, O. Gimenez, and S. P. Brooks

A. B. Lawson

S. Patterson and B. Jones

S. Green, J. Benedetti, A. Smith, and J. Crowley

R.J. Hayes and L.H. Moulton

M. Greenacre

D.L. Fairclough

L. Pronzato, H. Wynn, and A. Zhigljavsky

S. van Buuren

A. Skrondal and S. Rabe-Hesketh

GRAPHICAL ANALYSIS OF MULTI-RESPONSE DATA	K. Basford and J. Tukey
MARKOV CHAIN MONTE CARLO IN PRACTICE INTRODUCTION TO COMPUTATIONAL BIOLOGY: MAPS, SEQUENCES, AND GENOMES	W. Gilks, S. Richardson, and D. Spiegelhalter M. Waterman
MEASUREMENT ERROR AND MISCLASSIFICATION IN STATISTICS AND EPIDEMIOLOGY: IMPACTS AND BAYESIAN ADJUSTMENTS	P. Gustafson
MEASUREMENT ERROR: MODELS, METHODS, AND APPLICATIONS	J. P. Buonaccorsi
META-ANALYSIS OF BINARY DATA USING PROFILE LIKELIHOOD	D. Böhning, R. Kuhnert, and S. Rattanasiri
STATISTICAL ANALYSIS OF GENE EXPRESSION MICROARRAY DATA	T. Speed
STATISTICAL AND COMPUTATIONAL PHARMACOGENOMICS	R.Wu and M. Lin
STATISTICS IN MUSICOLOGY	J. Beran
STATISTICS OF MEDICAL IMAGING	T. Lei
STATISTICAL CONCEPTS AND APPLICATIONS IN CLINICAL MEDICINE	J. Aitchison, J.W. Kay, and I.J. Lauder
STATISTICAL AND PROBABILISTIC METHODS IN ACTUARIAL SCIENCE	P.J. Boland
STATISTICAL DETECTION AND SURVEILLANCE OF GEOGRAPHIC CLUSTERS	P. Rogerson and I. Yamada
STATISTICS FOR ENVIRONMENTAL BIOLOGY AND TOXICOLOGY	A. Bailer and W. Piegorsch
STATISTICS FOR FISSION TRACK ANALYSIS	R.F. Galbraith
VISUALIZING DATA PATTERNS WITH MICROMAPS	D.B. Carr and L.W. Pickle

This page intentionally left blank

Chapman & Hall/CRC Interdisciplinary Statistics Series

Analysis of Capture-Recapture Data

Rachel S. McCrea and Byron J. T. Morgan

National Centre for Statistical Ecology School of Mathematics, Statistics and Actuarial Science University of Kent Canterbury, UK



CRC Press is an imprint of the Taylor & Francis Group, an **informa** business A CHAPMAN & HALL BOOK MATLAB^{*} is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB^{*} software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB^{*} software.

CRC Press Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works Version Date: 20140623

International Standard Book Number-13: 978-1-4398-3660-6 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright. com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at http://www.taylorandfrancis.com

and the CRC Press Web site at http://www.crcpress.com

Contents

Α	ckno	wledge	ements	xv
P	refac	e		xvii
N	otati	on		xix
1 Introduction			ion	1
	1.1	Histor	ry and motivation	1
	1.2	Marki	ing	2
	1.3	Introd	luction to the Cormorant data set	3
		1.3.1	Capture-recapture-recovery data	4
		1.3.2	Nest counts and productivity	6
	1.4	Mode	lling population dynamics	6
	1.5	Summ	hary	7
	1.6	Furth	er reading	8
		1.6.1	History and aspects of bird ringing	8
		1.6.2	Reviews, books and websites	8
		1.6.3	Of mice and men	9
		1.6.4	Design	9
2	Mo	del fit	ting, averaging and comparison	11
	2.1	Introd	luction	11
		2.1.1	Forming likelihoods	11
		2.1.2	Bayes theorem and methods	12
		2.1.3	Iterative methods and parameterisation	12
	2.2	Classi	ical inference	13
		2.2.1	Point and interval estimation using maximum likelihood	13
		2.2.2	Flat likelihoods	15
		2.2.3	The EM algorithm	15
		2.2.4	Model comparison	15
		2.2.5	Model averaging	18
		2.2.6	Goodness-of-fit	19
	2.3	Bayes	ian inference	19
		2.3.1	Introduction	19
		2.3.2	Metropolis Hastings	20
		2.3.3	Gibbs sampling	21

		2.3.4	Using MCMC simulations	21
		2.3.5	Model probabilities and model averaging	22
		2.3.6	Reversible jump Markov chain Monte Carlo	22
		2.3.7	Hierarchical models	22
		2.3.8	Goodness-of-fit: Bayesian p-values and calibrated	
			simulation	23
	2.4	Compu	uting	23
	2.5	Summ	ary	24
	2.6	Furthe	r reading	24
3	\mathbf{Esti}	mating	g the size of closed populations	27
	3.1	Introd	uction	27
		3.1.1	Background	27
		3.1.2	Assumptions	27
	3.2	The So	chnabel census	28
		3.2.1	General notation	28
	3.3	Analys	sis of Schnabel census data	30
		3.3.1	Likelihoods based on the multinomial distribution	30
		3.3.2	Modelling based on the Poisson distribution, and	
			relationship to the multinomial	30
		3.3.3	Chao's lower-bound estimator	31
		3.3.4	Conditional analysis	33
		3.3.5	Relationship between conditional and unconditional	
			Poisson and multinomial analyses	34
	3.4	Model	classes	35
		3.4.1	Model M_0	35
		3.4.2	Time-dependent capture probability: model class, M_t	36
		3.4.3	Behavioural capture probability: model class, M_b	37
	3.5	Accourt	nting for unobserved heterogeneity	38
		3.5.1	Mixture models	38
		3.5.2	Horvitz-Thompson estimator	40
		3.5.3	Horvitz-Thompson-like estimator	42
		3.5.4	Other approaches to modelling heterogeneity	43
	3.6	Logisti	ic-linear models	43
	3.7	Spurio	usly large estimates	45
	3.8	Bayesi	an modelling	45
	3.9	Medica	al and social applications	46
		3.9.1	Log-linear models	47
	3.10	Testing	g for closure	47
	3.11	N-mix	ture estimators	47
	3.12	Spatia	l capture-recapture models	48
		3.12.1	Likelihood formation for a spatial capture-recapture	
			model	48
		3.12.2	Modelling animal locations	49
		3.12.3	Modelling the capture probability	50

CO	ONTE	INTS	ix
	3.13 3.14 3.15 3.16	 3.12.4 Alternative estimation methods 3.12.5 Applications of spatial capture-recapture models Computing Summary Further reading Exercises 	$50 \\ 51 \\ 51 \\ 52 \\ 52 \\ 52 \\ 54$
4	Sur	vival modelling: single-site models	57
	4.1	Introduction	57
	4.2	Mark-recovery models	59
		4.2.1 Time-dependence	60
		4.2.2 Incorporating age	61
		4.2.3 Full age-dependence in survival: the Cormack-Seber	69
		4.2.4 Model selection for mark recovery models	63
		4.2.4 Model selection for mark-recovery models	68
	13	Mark-recenture models	60
	4.0	4.3.1 Cormack-Jolly-Seber model	70
		4.3.2 Explicit maximum-likelihood estimates for the CJS	10
		model	70
		4.3.3 Bayesian estimation for CJS model	72
		4.3.4 Incorporating heterogeneity in the CJS model	73
		4.3.5 Age-dependent models in capture recapture	73
	4.4	Combining separate mark-recapture and recovery data sets	74
	4.5	Joint recapture-recovery models	75
		4.5.1 Extensions of joint recapture-recovery models	80
	4.6	Computing	81
	4.7	Summary	82
	4.8	Further reading	82
	4.9	Exercises	83
5	Surv	vival modelling: multisite models	87
	5.1	Introduction	87
	5.2	Matrix representation	88
	5.3	Multisite joint recapture-recovery models	89
	5.4	Multistate models as a unified framework	92
	5.5	Extensions to multistate models	94
		5.5.1 Models with unobservable states	94
		5.5.2 Memory models	96
	5.6	Model selection for multisite models	97
	5.7	Multievent models	99
		5.7.1 A unified framework	100
	F ^	5.7.2 Applications of multievent models	102
	5.8	Computing	102
	5.9	Summary	103

CO	NT	ΈN	TS
$\sim \circ$	T I T		- N

	5.10	Furth	er reading	103
	5.11	Exerc	ises	104
6 Occupancy modelling		y modelling	107	
	6.1	Introd	luction	107
	6.2	The t	wo-parameter occupancy model	109
		6.2.1	Survey design	110
	6.3	Exten	sions	111
		6.3.1	Multiple seasons	111
		6.3.2	Multiple species	111
		6.3.3	Citizen science and presence-only data	111
	6.4	Movin	g from species to individual: abundance-induced hetero-	
		geneit	y	112
	6.5	Accou	nting for spatial information	112
		6.5.1	Poisson process model	113
		6.5.2	Markov modulated Poisson process	114
	6.6	Accou	nting for spatial information	115
		6.6.1	Abundance-induced heterogeneity along a transect	116
		6.6.2	Incorporating clustering	116
	6.7	Comp	uting	117
	6.8	Summ	arv	117
	6.9	Furth	er reading	118
	6.10	Exerc	ises	118
7	Cov	ariate	s and random effects	121
	7.1	Introd	luction	121
	7.2	Extern	nal covariates	122
		7.2.1	Regressions	122
	7.3	Thres	hold models	124
	7.4	Indivi	dual covariates	126
		7.4.1	Estimating N: observed heterogeneity	127
		7.4.2	Dealing with missing information: a classical approach	
			to dealing with time-varying individual covariates in	
			open populations	128
		7.4.3	Dealing with missing information: a Bayesian approach	
			and comparison	132
	7.5	Rando	om effects	132
	7.6	Measu	irement error	136
	7.7	Use of	f P-splines	136
	7.8	Senes	cence	138
	79	Variał	ole selection	138
	1.0	7.91	Lasso	139
		792	The Bayesian approach to model selection	140
	710	Snatia	l covariates	141
	7 11	Comp	uting	1/12
	1.11	Comp	uung	140

х

CONTENTS			
	7.12	Summary	144
	7.13	Further reading	144
	7.14	Exercises	145
8	\mathbf{Sim}	ultaneous estimation of survival and abundance	149
	8.1	Introduction	149
	8.2	Estimating abundance in open populations	149
		8.2.1 The original Jolly-Seber model	150
		8.2.2 The Schwarz and Arnason formulation	151
		8.2.3 Alternative Jolly-Seber formulations	152
	8.3	Batch marking	153
	8.4	Robust design	154
		8.4.1 Closed robust design model	154
		8.4.2 Open robust design model	157
	8.5	Stopover models	158
	8.6	Computing	161
	8.7	Summary	161
	8.8	Further reading	161
	8.9	Exercises	162
9	Goo	odness-of-fit assessment	165
	9.1	Introduction	165
	9.2	Diagnostic goodness-of-fit tests	165
		9.2.1 Contingency table tests of homogeneity	165
		9.2.2 Single-site goodness-of-fit tests	166
		9.2.3 Equivalence of score tests	174
		9.2.4 Multisite goodness-of-fit tests	175
		9.2.5 Joint recapture and recovery goodness-of-fit tests	179
	9.3	Absolute goodness-of-fit tests	180
		9.3.1 Sufficient statistic goodness-of-fit tests	180
		9.3.2 Bayesian p-values and calibrated simulation	183
	9.4	Computing	183
	9.5	Summary	183
	9.6	Further reading	184
	9.7	Exercises	184
10	Par	ameter redundancy	187
	10.1	Introduction	187
	10.2	Using symbolic computation	190
		10.2.1 Determining the deficiency of a model	190
		10.2.2 Determining estimable parameters for parameter-	100
		redundant models	191
	10.3	Parameter redundancy and identifiability	192
	10.4	Decomposing the derivative matrix of full rank models	192
	10.5	Extension	193

	10.6	The moderating effect of data	194
		10.6.1 Missing data	195
		10.6.2 Near singularity	195
	10.7	Covariates	195
	10.8	Exhaustive summaries and model taxonomies	197
	10.9	Bayesian methods	198
		10.9.1 A Bayesian fit for a parameter-redundant model	198
		10.9.2 Weakly identifiable parameters	200
	10.10) Computing	202
	10.11	l Summary	202
	10.12	2 Further reading	202
	10.13	3 Exercises	203
11	Stat	e-space models	207
	11.1	Introduction	207
	11.2	Definitions	207
		11.2.1 Constructing Leslie matrices: the order of biological	
		processes	211
	11.3	Fitting linear Gaussian models	212
		11.3.1 The stochastic Gompertz model	212
		11.3.2 Linear Gaussian models	213
		11.3.3 The Kalman filter	214
		11.3.4 Initialising the filter	215
	11.4	Models which are not linear Gaussian	215
		11.4.1 Conditionally Gaussian models	216
	11.5	Bayesian methods for state-space models	217
	11.6	Formulation of capture-reencounter models	218
		11.6.1 Capture-recapture data	219
		11.6.2 Ring-recovery data	220
		11.6.3 Multistate capture-recapture models	221
	11.7	Formulation of occupancy models	221
	11.8	Computing	222
	11.9	Summary	222
	11.1() Further reading	223
	11.11	L Exercises	224
12	Inte	grated population modelling	227
	12.1	Introduction	227
	10.0	12.1.1 Multiplying likelihoods	229
	12.2	Normal approximations of component likelihoods	229
		12.2.1 Multivariate normal approximation to capture reen-	0.00
	10.0	counter likelihoods	230
	12.3	Model selection	231
	12.4	Goodness-of-fit for integrated population modelling; calibrated	001
		SITUATION	2.5

CONTENTS	xiii
12.5 Previous applications	231
12.6 Hierarchical modelling	235
12.7 Computing	236
12.8 Summary	237
12.9 Further reading	237
12.10 Exercises	238
A Distributions Reference	241
References	243
Index	279

This page intentionally left blank

Acknowledgements

We would like to thank the many people who have helped us complete this book.

In particular, we are grateful to our colleagues Takis Besbeas, Dankmar Böhning, David Borchers, Steve Buckland, Ted Catchpole, Rémi Choquet, Diana Cole, Laura Cowen, Emily Dennis, Rachel Fewster, David Fletcher, Stephen Freeman, Olivier Gimenez, Gurutzeta Guillera-Arroita, Ben Hubbard, Anita Jeyam, Ruth King, Eleni Matechou, Chiara Mazzetta, José Lahoz-Monfort, Jean-Dominique Lebreton, Kelly Moyes, Beth Norris, Lauren Oliver, Shirley Pledger, Roger Pradel, Martin Ridout, Giacomo Tavecchia, David Thomson, Anne Viallefont, Hannah Worthington and the anonymous referees who commented on an early draft. We also thank James McCrea, for his tireless work on the bibliography and website and the ecologists, Thomas Bregnballe, Tim Coulson, Tim Clutton-Brock, Marco Festa-Bianchet, Richard Griffiths, Mike Harris, Warwick Nash, David Roberts, David Sewell, Sarah Wanless, Lydia Yeo and the Euring community who have produced the data, provided invaluable input and insights into the populations being studied and have even taken us on occasional fieldwork trips.

The material of this book has evolved through many workshops given at Cambridge, St Andrews, Rothamsted and elsewhere, as well as through an MSc in Statistics module at the University of Kent. We would like to thank the workshop participants and MSc students for helping the content develop.

The work of this book would not have been possible without the funding provided by EPSRC, NERC, the British Council and the Royal Society.

Finally we would like to thank Chapman and Hall, CRC Press for their patience and advice.

This page intentionally left blank

Preface

In this book we aim to cover the many modern developments in the area of capture-recapture and related models, and to set them in the historical context of relevant research over the past 100 years. The main area of application is ecology.

Much of the material of the book has been taught to students on the MSc course in Statistics at the University of Kent over several years. Parts of the book have also been covered in a number of workshops at both the University of Kent and St Andrews University, over many years, as well as in workshops for the *National Centre for Statistical Ecology*, at Cambridge, the British and Irish Region of the *International Biometric Society* at Rothamsted Experimental Station, the *International Biometric Conference* in Florence, Italy, at Point Reyes and Patuxent, USA, Radolfzel and Rostock, Germany and Dunedin, New Zealand.

The book may be used as an advanced undergraduate and higher-level textbook, for students with a knowledge of statistics and probability such as that typically provided in a second-year mathematics/statistics undergraduate course in the UK. Chapter 2 of the book provides a brief revision of the basic ideas of classical and Bayesian inference that are used and referenced in the later chapters. An appendix lists the distributions used in the book.

The book may also be used as a source of reference for quantitative ecologists as well as statisticians. Many of the methods described in the book have been devised in collaboration between statisticians and ecologists, notably through the series of successful *Euring* technical meetings over the past 20 years. Capture-recapture and related models are designed for particular types of ecological, social, epidemiological and medical data, and applications to real and simulated data sets provide fine examples of biometry and applied statistics at work.

One of the reasons that the capture-recapture area has seen a recent explosion of new applications and methods is the availability of free user-friendly computer software. Another reason is the appearance of important long-term data sets, as well as a wide range of new data resulting from technological advances in the marking, identification and tracking of wild animals. Several capture-recapture data sets now have classic status, and we include discussion and analyses of several of these. However, in order to provide continuity, where possible we illustrate methods on a single data set on Great cormorants, studied in Denmark. In order to prevent the book becoming very large we have had to be selective. Supplementary material is provided in Exercises and on the book web site, www.capturerecapture.co.uk, which also contains data, computer programs, pictures, and an indication of how the book has been used as the basis of a lecture course.

Rachel McCrea and Byron Morgan, National Centre for Statistical Ecology, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent. June 23, 2014

MATLAB is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc. 3 Apple Hill Drive Natick, MA 01760-2098 USA Tel: 508 647-7000 Fax: 508-647-7001 E-mail: info@mathworks.com Web: www.mathworks.com

xviii

Notation

Where possible within the book we adopt the following notation.

S: survival probability

 $\phi:$ apparent survival probability

 λ : recovery probability

p: capture probability

 $\psi :$ transition probability (in the context of multistate capture-recapture models, see Chapter 5)

 $\varphi :$ survival-transition probability

 $\psi :$ occupancy probability (in the context of occupancy models, see Chapter 6)

 $T{:}$ total number of capture occasions

N : population size

K: number of sites/states

 $G\!\!:\mathrm{number}$ of groups

 $\epsilon:$ emigration probability

 $\rho {:} \ {\rm productivity}$

 $\pmb{\theta} \text{:}$ model parameters

 \boldsymbol{x} : data

 $M(\boldsymbol{\theta})$: model

 $L(\pmb{\theta}; \pmb{x}):$ likelihood

 $\ell(\boldsymbol{\theta}; \boldsymbol{x}) = \log\{L(\boldsymbol{\theta}; \boldsymbol{x})\}$

 $\pi(\boldsymbol{\theta})$: prior distribution

 $\pi(\boldsymbol{\theta}|\boldsymbol{x})$: posterior distribution

 $\mathbf{U} = \partial \ell(\boldsymbol{\theta}; \boldsymbol{x}) / \partial \boldsymbol{\theta}$: scores vector

A: Hessian matrix

 $oldsymbol{J} = -\mathbb{E}[oldsymbol{A}(oldsymbol{ heta})]$: Fisher information matrix

 $-\boldsymbol{A}(\boldsymbol{\theta}):$ observed Fisher information matrix

 $N_d(\boldsymbol{\theta}, \boldsymbol{J}^{-1}(\boldsymbol{\theta}))$: multivariate normal distribution

 χ^2_{ν} : chi-square distribution

- $-2\log\lambda$: likelihood-ratio test statistic
- M: number of models in a model set
- q(a, b) proposal function in Metropolis-Hastings
- AIC: Akaike information criterion
- AIC_c : AIC adapted for small sample sizes
- $QAIC:\,AIC$ adapted for over dispersion
- BIC: Bayesian information criterion
- O_i : observed value
- E_i : expected value
- c^2 : over dispersion coefficient
- $D(\mathbf{x}, \boldsymbol{\theta})$: discrepancy statistic
- $\Lambda:$ population projection matrix
- $I(\cdot)$: indicator function

Generally, we will denote time-dependence of parameters by a subscript, state/site-dependence using a superscript and age-dependence will be denoted in parentheses. Note that when a parameter depends on a covariate, we will include this within the parentheses, and this will be made clear in the context of the example. Individual effects will also be denoted by a subscript and any further deviations from this notation will be described when introduced. "Right from the beginning, I always strived to capture everything I saw as completely as possible."

Norman Rockwell

This page intentionally left blank

Introduction

1.1 History and motivation

The Independent newspaper published in England on the 6^{th} of May, 2010, contained an article with the title: "End of the Alaotra grebe is further evidence of sixth great extinction: species are vanishing faster than at any point in the last 65 million years." It is generally supposed that this loss of species is due to anthropogenic change, and there is great current concern regarding the consequent loss of biodiversity, which motivates the work of this book; see Kolbert (2014).

The aims of human demography and of population ecology are similar, with key demographic parameters relating to birth, death and movement in each case. The Births and Deaths Registration Act of 1836 resulted in all registrations of birth, death and marriage being kept centrally in England and Wales from the 1^{st} of July 1837. As a result of such regulation, it is usually far easier to determine human mortality than that of wild animals, and the different data that arise for humans and animals require different methods of analysis.

An initial step in studying wild animals is often to ensure that they can be identified uniquely. When such animals are reencountered then this gives rise to different forms of capture-recapture data. For instance, an issue of the journal *Ringing & Migration*, published in 2009, is devoted to celebrating 100 years of bird ringing in Britain and Ireland. The first paper in the issue, by Greenwood (2009), provides interesting background to the ringing scheme which is now administered by the British Trust for Ornithology (BTO). The bird rings have unique numbers, indicating when the rings were attached, and also contain an address. If a dead bird is found to have a ring, and the ring is returned to the address, then information is obtained on how long the bird lived; an illustration is on the book web site, www.capturerecapture.co.uk. This approach is generally regarded as having been initiated in Denmark in 1899, by Hans Chr. Mortensen, and is now used widely throughout the world.

Over the past 100 years a body of theory and methods has been developed for the analysis of capture-recapture data, and the area remains one of active research. Books on the subject include Williams et al. (2002), Amstrup et al. (2005), Royle and Dorazio (2008), Link and Barker (2009) and King et al. (2010). In addition there are specialised computer packages such as Program Mark (White and Burnham 1999) and E-SURGE (Choquet et al. 2009) for model-fitting by maximum likelihood, and Gimenez et al. (2008) and Kéry and Schaub (2012) describe how WinBUGS may be used for Bayesian inference. Capture-recapture data also arise in medical, epidemiological and social areas, and we shall describe some of these applications-see Chao et al. (2001). The paper by Borchers et al. (2014) provides a modelling framework which links capture-recapture and distance sampling surveys (Buckland et al. 2001).

1.2 Marking

Some animals can be identified by their physical characteristics, and this is true for example of the spot patterns of cheetahs, the belly patterns of great crested newts, (Figure 1.1(a)) and permanent marks produced by aggressive interactions between bottlenose dolphins. However, it is often the case that identifying marks have to be attached to wild animals. Marks may take the form of neck bands, leg rings, wing clips, ear tags, etc. Rings on the legs of cormorants are illustrated in Figure 1.1(c), and tags on the ears of ibex are shown in Figure 1.1(d). Butterflies may be marked using coloured marking pens, under standardised protocols. Also, more than one device may be used at the same time for any one animal, for instance colour rings for resighting live birds and metal rings with an address for the recovery of dead birds; this is the case for the cormorant study described in the next section. Double marking is sometimes also used if there is a danger of marks being lost or becoming illegible due to wear. It is usually assumed that the process of capturing animals and marking them does not affect their subsequent survival. Modern marking procedures also include radio-tagging such as the collar shown on a straw-coloured fruit bat in Figure 1.1(b), the use of passive integrated transponder (PIT) tags, like those used to identify animal pets, temperature depth recorder (TDR) tags, which can measure depth for marine animals, and miniature geolocators. In some cases genetic information can be gleaned from hair or fur deposited in the field on features such as barbed wire fences, and from faeces. In data analysis, account sometimes needs to be taken of the possibility of an animal losing a mark, of marks being damaged through wear and tear, and of misreporting of marks. It is often the case that marks from dead animals are reported soon after death, as otherwise the corpse may have decayed or been eaten. However, if marks are attached to shellfish, for example, then they can be discovered and reported a long time after death. We use the terms *reencounter* and *recapture* exchangeably, to cover recovery, recapture and resighting. There are often major differences in scale between the areas over which recoveries are reported and those for recaptures and resightings.



(a) Great crested newt (b) Straw-coloured fruit bat ©D.Hayman ©D.Sewell



(c) Cormorant ©T.Bregnballe

(d) Ibex ©A.Brambilla

Figure 1.1 Examples of natural and applied markings: 1.1(a) shows the belly pattern of a Great crested newt, Triturus cristatus; 1.1(b) shows a Straw-coloured fruit bat, Eidolon helvum with a radio-transmitter collar; 1.1(c) shows a Great cormorant, Phalacrocorax carbo sinensis, with one ring on each of its legs and 1.1(d) illustrates an Alpine ibex, Capra ibex, with an ear tag.

1.3 Introduction to the Cormorant data set

Throughout the book we shall consider applications to a range of different animals, however for continuity we shall demonstrate several methods using data from a population of Great cormorants, Phalacrocorax carbo sinensis. We have kindly been given permission to use these data collected by the National Environmental Research Institute at Aarhus University, Denmark. The data are part of a much larger ringing programme started in Denmark in 1977, which continues to the present day. Analyses have been provided by, for example, Frederiksen et al. (2001). There are many colonies of Great cormorants around Denmark and the data that we use in this book correspond to a period of population expansion (1981-1993) in six colonies located 32-234km apart. The location of the colonies is shown in Figure 1.2. The oldest of the six colonies, Vorsö (VO), was established (est.) in 1944 and along with Ormö (OR, est. 1972) and Brændegård Sö (BR, est. 1973) comprised the only colonies present in Denmark at the start of the ringing study. Colonies Toft Sö (TO), Dyreford (DY) and Mågeöerne (MA) were established during the study in 1982, 1984 and 1985 respectively. Nest locations vary between colonies: the

nests at colony MA are only located on the ground, whilst nests at BR are located both on the ground and in trees. Nests in the remaining four colonies are only found in trees.



Figure 1.2 Location of cormorant colonies in Denmark. Figure provided by T. Bregnballe.

For illustration, throughout the book we shall extract components of this substantial data set for particular examples. The multisite/state nature of the recapture data, and the multiple types of data collected, including live recaptures, dead recoveries and population counts makes this an ideal data set to demonstrate several of the modelling approaches of the book. Additional features, such as known age at capture, allow us to explore a wide range of complex model structures.

1.3.1 Capture-recapture-recovery data

Chicks were ringed using a standard metal ring on one leg and a coloured plastic band with a unique combination of three alphanumeric characters on the other when they were between 25 and 38 days old. In this study recaptures are in fact resightings. Regular trips (often daily) are made to the colonies between February and October and attempts are made to resight and identify individuals; illustrative histories are shown in Table 1.1. We observe that within a single breeding season individuals can be resighted multiple times, and also that in some cases birds are not seen in a colony for several years after ringing. Generally, resighting is done from a distance; for example, at

Individual Code	Date of Ringing	Dates of Resighting
BL 089	31/05/1986	13/06/1994,
		19/06/1994,
		29/06/1994
GU 191	10/05/1980	10/03/1994,
		17/03/1994,
		22/03/1994,
		11/05/1994
HV 04L	25/05/1992	10/05/1994,
		12/05/1994,
		19/05/1994,
		20/05/1994,
		23/05/1994

Table 1.1: Extract of 1994 resignting data from colony VO.

the largest colony, VO, this is conducted from an observation tower within the breeding cormorant population. Resightings are recorded only for individuals identified as breeders, and birds were considered breeders at VO if they had been observed copulating, incubating chicks or rearing chicks. The timing of onset of breeding varied among the years but 71-97% of clutches that were laid at VO were initiated between 1 March and 19 April. This observation allowed criteria to be developed for identifying breeding cormorants at the other sites, where resighting intensity was not as high. Individuals at the remaining five colonies were judged to be likely breeders if they had been observed in the colony:

- 1. on a nest between 1 March and 19 April, or
- 2. between 1 March and 19 April and identified as a likely breeder in an earlier year, or
- 3. between 20 April and 30 June.

The birds were recovered dead within a large geographical area ranging from the United Kingdom to southern Algeria and from western Spain to eastern Romania, and separate data on recoveries of dead birds were provided by the Ringing Department at the Zoological Museum, Copenhagen. The approximate age of fledging is 30 days, and recoveries of birds found dead in their natal colony within this period were excluded; in addition, recovery information when only a ring was found within the natal colony less than 30 days after ringing was excluded.

Depending on the type of model we wish to fit, we may analyse information on several captures within a single year, or we may record data on an annual basis, and thereby develop an annual encounter history for an individual. We note that since individuals are ringed as non-breeders and not resignted until they start to breed, we may decide to model the data using multiple states, where the state denotes the breeding status of an individual. Such modelling can allow us to estimate biological processes of interest, such as recruitment probabilities which describe when animals become breeders.

1.3.2 Nest counts and productivity

Colony sizes were also estimated during each year of the study, by counting the number of nests in early May, which is the time of year when the nest numbers peaked; the data are given in Figure 1.3. The marked individuals make up a small proportion of the total population as the ringing of nestlings takes place in a small zone of each colony. Productivity is normally assessed through direct measurements, for instance through egg counting.



Figure 1.3 Cormorant colony size counts per colony, in terms of nests, between 1981 and 1992

1.4 Modelling population dynamics

Estimates of annual survival probability and of productivity are important components of population dynamics models, and as we shall see below, they contribute to entries in what are called Leslie matrices. For example, suppose that we have a bird species in which we can detect 5 age-classes, and, ignoring the numbers of individuals in their first year of life, that $N_t(1)$, $N_t(2)$, $N_t(3)$, and $N_t(a)$ denote, respectively, the numbers of females aged one year, two years, three years, and greater than three years at time t. We let $S_t(j)$ denote the annual survival probability at time t of a bird in age class j, where j takes

SUMMARY

the values 1, 2, 3, 4 and a (denoting 4^{th} and older years of life) and denote productivity by the constant ρ . If we assume that the process of birth in transition (t - 1, t) relates to year t - 1 and operates before the process of survival, and that only birds aged 2 and older reproduce, then we can write down the following transition equation, based on a Leslie matrix, relating the numbers of birds at time t in each of four age classes to those at the previous time.

$$\begin{pmatrix} N_t(1)\\ N_t(2)\\ N_t(3)\\ N_t(a) \end{pmatrix} = \begin{pmatrix} 0 & \rho S_{t-1}(1) & \rho S_{t-1}(1) & \rho S_{t-1}(1)\\ S_{t-1}(2) & 0 & 0 & 0\\ 0 & S_{t-1}(3) & 0 & 0\\ 0 & 0 & S_{t-1}(4) & S_{t-1}(a) \end{pmatrix} \begin{pmatrix} N_{t-1}(1)\\ N_{t-1}(2)\\ N_{t-1}(3)\\ N_{t-1}(a) \end{pmatrix}.$$
(1.1)

Applying this equation several times allows predictions to be made about the future behaviour of the population.

The Perron-Frobenius theorem states that for appropriate constant transition matrices $\mathbf{\Lambda}$, if for some k, all elements of $\mathbf{\Lambda}^k$ are strictly positive, then there exists a real positive eigenvalue λ that is greater in absolute value (or in modulus, if some of the other eigenvalues are complex) than all other eigenvalues. The dominant eigenvalue λ represents the asymptotic growth rate of the population, and the normalised right eigenvector \mathbf{v} associated with λ represents the asymptotic proportion of every age class in the total population, called the stable age distribution; see Caswell (2001, p.83).

A simplifying feature of the model of Equation (1.1) is that it is deterministic, without any random components, and we shall consider stochastic transition equations, which allow for random variation, in Chapter 11.

1.5 Summary

There is much current concern at the loss of biodiversity. We often study the behaviour of wild animals after giving them marks which identify them uniquely. It is usually assumed that marking does not affect their survival, although there is some evidence to the contrary in certain special cases. Marked animals may be found dead, or seen again alive, providing information on demographic features such as mortality and movement. Estimates of mortality and of productivity are used in models of population dynamics, which can provide understanding of how populations change, and provide predictions of future behaviour. The analyses in the book will frequently be illustrated by means of data on a particular population of cormorants. In certain cases, similar models are used in application areas such as epidemiology.

1.6 Further reading

1.6.1 History and aspects of bird ringing

The paper by Pradel (2009) provides graphs showing the increase in the use of methods of capture-recapture over time. A poster giving a time line for the main developments in the area of capture-recapture is on www.capturerecapture.co.uk.

Greenwood (pers com) observes that British and Irish ornithologists were ringing birds from 1890, and that in fact marking birds dates from medieval times. A wide-ranging review is provided by Lebreton (2001). The book by Balmer et al. (2008) provides an excellent introduction to bird ringing, and concludes with an interesting list of record breakers (for example, the species with the highest recovery probability, the oldest ringed bird (over 50 years), etc.). It is observed that there is a decline in the reporting of rings from dead birds in recent time; see Balmer et al. (2008, p.24). Some rings now also contain a web address; see Balmer et al. (2008, p.57). The paper by Saraux et al. (2010) found that flipper bands attached to penguins affected both the survival and reproduction of the birds. Although this is generally regarded as an extreme example, it should always be a consideration that marked animals might survive and reproduce differently from unmarked ones. The paper by Aebischer (1983) provides a way to determine ring numbers after their superficial features have been eroded. The effect of misreading marks is the topic of Schwarz and Stobo (1999); such errors can be a problem when using genetic information, see for example Link et al. (2010). Genetic information can reveal information other than just survival; see Okland et al. (2010). Goudie and Goudie (2007) give the history of marking fish.

1.6.2 Reviews, books and websites

The book by Seber (1982) is a classical text in statistical ecology, and contains a wealth of applications and analyses of challenging and interesting data sets, as well as much historical material. There is for example discussion of ways of dealing with mark loss, on p.488. An early review of capture-recapture methods is provided by Pollock (1991). More recent work can be appreciated from Pollock (2000) and Schwarz and Seber (1999), Morgan and Thomson (2002), Senar et al. (2004) and Schaub and Kendall (2012). The volume by Williams et al. (2002) is encyclopaedic, and essential reading for statistical ecologists. The estimation of wild animal abundance is the subject of the book by Borchers et al. (2002). The book by Caswell (2001) is the classical reference for population dynamics modelling, and provides information regarding the formation of Leslie matrices. A modern approach to sensitivity analysis in population modelling is provided by Miller et al. (2011a). A wider ecological perspective is provided by references such as Morgan et al. (1997), Furness and Greenwood (1993), Clutton-Brock (1988), Clutton-Brock and Albon (1989) and Clutton-Brock et al. (1982).