Hidden Markov Models for Time Series

An Introduction Using R

MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

J. Fan, V. Isham, N. Keiding, T. Louis, R. L. Smith, and H. Tong

1 Stochastic Population Models in Ecology and Epidemiology *M.S. Barlett* (1960)
2 Queues *D.R. Cox and W.L. Smith* (1961)
3 Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)
4 The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)
5 Population Genetics *W.J. Ewens* (1969)
6 Probability, Statistics and Time *M.S. Barlett* (1975)
7 Statistical Inference *S.D. Silvey* (1975)
8 The Analysis of Contingency Tables *B.S. Everitt* (1977)
9 Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)
10 Stochastic Abundance Models *S. Engen* (1978)
11 Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)
12 Point Processes *D.R. Cox and V. Isham* (1980)
13 Identification of Outliers *D.M. Hawkins* (1980)
14 Optimal Design *S.D. Silvey* (1980)
15 Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)
16 Classification *A.D. Gordon* (1981)
17 Distribution-Free Statistical Methods, 2nd edition *J.S. Maritz* (1995)
18 Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)
19 Applications of Queueing Theory, 2nd edition *G.F. Newell* (1982)
20 Risk Theory, 3rd edition *R.E. Beard, T. Pentikäinen and E. Pesonen* (1984)
21 Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)
22 An Introduction to Latent Variable Models *B.S. Everitt* (1984)
23 Bandit Problems *D.A. Berry and B. Fristedt* (1985)
24 Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)
25 The Statistical Analysis of Composition Data *J. Aitchison* (1986)
26 Density Estimation for Statistics and Data Analysis *B.W. Silverman* (1986)
27 Regression Analysis with Applications *G.B. Wetherill* (1986)
28 Sequential Methods in Statistics, 3rd edition
G.B. Wetherill and K.D. Glazebrook (1986)
29 Tensor Methods in Statistics *P. McCullagh* (1987)
30 Transformation and Weighting in Regression
R.J. Carroll and D. Ruppert (1988)
31 Asymptotic Techniques for Use in Statistics
O.E. Bandorff-Nielsen and D.R. Cox (1989)
32 Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)
33 Analysis of Infectious Disease Data *N.G. Becker* (1989)
34 Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)
35 Empirical Bayes Methods, 2nd edition *J.S. Maritz and T. Lwin* (1989)
36 Symmetric Multivariate and Related Distributions
K.T. Fang, S. Kotz and K.W. Ng (1990)
37 Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)
38 Cyclic and Computer Generated Designs, 2nd edition
J.A. John and E.R. Williams (1995)
39 Analog Estimation Methods in Econometrics *C.F. Manski* (1988)
40 Subset Selection in Regression *A.J. Miller* (1990)
41 Analysis of Repeated Measures *M.J. Crowder and D.J. Hand* (1990)
42 Statistical Reasoning with Imprecise Probabilities *P. Walley* (1991)
43 Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)

44 Inspection Errors for Attributes in Quality Control
N.L. Johnson, S. Kotz and X. Wu (1991)

45 The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)

46 The Analysis of Quantal Response Data *B.J.T. Morgan* (1992)

47 Longitudinal Data with Serial Correlation—A State-Space Approach
R.H. Jones (1993)

48 Differential Geometry and Statistics *M.K. Murray and J.W. Rice* (1993)

49 Markov Models and Optimization *M.H.A. Davis* (1993)

50 Networks and Chaos—Statistical and Probabilistic Aspects
O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (1993)

51 Number-Theoretic Methods in Statistics *K.-T. Fang and Y. Wang* (1994)

52 Inference and Asymptotics *O.E. Barndorff-Nielsen and D.R. Cox* (1994)

53 Practical Risk Theory for Actuaries
C.D. Daykin, T. Pentikäinen and M. Pesonen (1994)

54 Biplots *J.C. Gower and D.J. Hand* (1996)

55 Predictive Inference—An Introduction *S. Geisser* (1993)

56 Model-Free Curve Estimation *M.E. Tarter and M.D. Lock* (1993)

57 An Introduction to the Bootstrap *B. Efron and R.J. Tibshirani* (1993)

58 Nonparametric Regression and Generalized Linear Models
P.J. Green and B.W. Silverman (1994)

59 Multidimensional Scaling *T.F. Cox and M.A.A. Cox* (1994)

60 Kernel Smoothing *M.P. Wand and M.C. Jones* (1995)

61 Statistics for Long Memory Processes *J. Beran* (1995)

62 Nonlinear Models for Repeated Measurement Data
M. Davidian and D.M. Giltinan (1995)

63 Measurement Error in Nonlinear Models
R.J. Carroll, D. Rupert and L.A. Stefanski (1995)

64 Analyzing and Modeling Rank Data *J.J. Marden* (1995)

65 Time Series Models—In Econometrics, Finance and Other Fields
D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (1996)

66 Local Polynomial Modeling and its Applications *J. Fan and I. Gijbels* (1996)

67 Multivariate Dependencies—Models, Analysis and Interpretation
D.R. Cox and N. Wermuth (1996)

68 Statistical Inference—Based on the Likelihood *A. Azzalini* (1996)

69 Bayes and Empirical Bayes Methods for Data Analysis
B.P. Carlin and T.A Louis (1996)

70 Hidden Markov and Other Models for Discrete-Valued Time Series
I.L. MacDonald and W. Zucchini (1997)

71 Statistical Evidence—A Likelihood Paradigm *R. Royall* (1997)

72 Analysis of Incomplete Multivariate Data *J.L. Schafer* (1997)

73 Multivariate Models and Dependence Concepts *H. Joe* (1997)

74 Theory of Sample Surveys *M.E. Thompson* (1997)

75 Retrial Queues *G. Falin and J.G.C. Templeton* (1997)

76 Theory of Dispersion Models *B. Jørgensen* (1997)

77 Mixed Poisson Processes *J. Grandell* (1997)

78 Variance Components Estimation—Mixed Models, Methodologies and Applications *P.S.R.S. Rao* (1997)

79 Bayesian Methods for Finite Population Sampling
G. Meeden and M. Ghosh (1997)

80 Stochastic Geometry—Likelihood and computation
O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (1998)

81 Computer-Assisted Analysis of Mixtures and Applications—
Meta-analysis, Disease Mapping and Others *D. Böhning* (1999)

82 Classification, 2nd edition *A.D. Gordon* (1999)

83 Semimartingales and their Statistical Inference *B.L.S. Prakasa Rao* (1999)

84 Statistical Aspects of BSE and vCJD—Models for Epidemics
C.A. Donnelly and N.M. Ferguson (1999)

85 Set-Indexed Martingales *G. Ivanoff and E. Merzbach* (2000)

86 The Theory of the Design of Experiments *D.R. Cox and N. Reid* (2000)

87 Complex Stochastic Systems
O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg (2001)

88 Multidimensional Scaling, 2nd edition *T.F. Cox and M.A.A. Cox* (2001)

89 Algebraic Statistics—Computational Commutative Algebra in Statistics
G. Pistone, E. Riccomagno and H.P. Wynn (2001)

90 Analysis of Time Series Structure—SSA and Related Techniques
N. Golyandina, V. Nekrutkin and A.A. Zhigljavsky (2001)

91 Subjective Probability Models for Lifetimes
Fabio Spizzichino (2001)

92 Empirical Likelihood *Art B. Owen* (2001)

93 Statistics in the 21st Century
Adrian E. Raftery, Martin A. Tanner, and Martin T. Wells (2001)

94 Accelerated Life Models: Modeling and Statistical Analysis
Vilijandas Bagdonavicius and Mikhail Nikulin (2001)

95 Subset Selection in Regression, Second Edition *Alan Miller* (2002)

96 Topics in Modelling of Clustered Data
Marc Aerts, Helena Geys, Geert Molenberghs, and Louise M. Ryan (2002)

97 Components of Variance *D.R. Cox and P.J. Solomon* (2002)

98 Design and Analysis of Cross-Over Trials, 2nd Edition
Byron Jones and Michael G. Kenward (2003)

99 Extreme Values in Finance, Telecommunications, and the Environment
Bärbel Finkenstädt and Holger Rootzén (2003)

100 Statistical Inference and Simulation for Spatial Point Processes
Jesper Møller and Rasmus Plenge Waagepetersen (2004)

101 Hierarchical Modeling and Analysis for Spatial Data
Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand (2004)

102 Diagnostic Checks in Time Series *Wai Keung Li* (2004)

103 Stereology for Statisticians *Adrian Baddeley and Eva B. Vedel Jensen* (2004)

104 Gaussian Markov Random Fields: Theory and Applications
Håvard Rue and Leonhard Held (2005)

105 Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition
Raymond J. Carroll, David Ruppert, Leonard A. Stefanski,
and Ciprian M. Crainiceanu (2006)

106 Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood
Youngjo Lee, John A. Nelder, and Yudi Pawitan (2006)

107 Statistical Methods for Spatio-Temporal Systems
Bärbel Finkenstädt, Leonhard Held, and Valerie Isham (2007)

108 Nonlinear Time Series: Semiparametric and Nonparametric Methods
Jiti Gao (2007)

109 Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis
Michael J. Daniels and Joseph W. Hogan (2008)

110 Hidden Markov Models for Time Series: An Introduction Using R
Walter Zucchini and Iain L. MacDonald (2009)

Monographs on Statistics and Applied Probability 110

Hidden Markov Models for Time Series

An Introduction Using R

Walter Zucchini
Georg–August–Universität
Göttingen, Germany

Iain L. MacDonald
University of Cape Town
South Africa

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

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

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-58488-573-3 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Zucchini, W.
Hidden Markov models for time series : an introduction using R / Walter Zucchini, Iain L. MacDonald.
p. cm. -- (Monographs on statistics and applied probability ; 110)
Includes bibliographical references and index.
ISBN 978-1-58488-573-3 (hardcover : alk. paper)
1. Time-series analysis. 2. Markov processes. 3. R (Computer program language) I. MacDonald, Iain L. II. Title. III. Series.

QA280.Z83 2009
519.5'5--dc22 2009007294

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

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

Für Hanne und Werner,
mit herzlichem Dank für Eure Unterstützung
bei der Suche nach den versteckten Ketten.

Contents

Preface

In the eleven years since the publication of our book *Hidden Markov and Other Models for Discrete-valued Time Series* it has become apparent that most of the 'other models', though undoubtedly of theoretical interest, have led to few published applications. This is in marked contrast to hidden Markov models, which are of course applicable to more than just *discrete-valued* time series. These observations have led us to write a book with different objectives.

Firstly, our emphasis is no longer principally on discrete-valued series. We have therefore removed Part One of the original text, which covered the 'other models' for such series. Our focus here is exclusively on hidden Markov models, but applied to a wide range of types of time series: continuous-valued, circular, multivariate, for instance, in addition to the types of data we previously considered, namely binary data, bounded and unbounded counts and categorical observations.

Secondly, we have attempted to make the models more accessible by illustrating how the computing environment **R** can be used to carry out the computations, e.g., for parameter estimation, model selection, model checking, decoding and forecasting. In our previous book we used proprietary software to perform numerical optimization, subject to linear constraints on the variables, for parameter estimation. We now show how one can use standard **R** functions instead. The **R** code that we used to carry out the computations for some of the applications is given, and can be applied directly in similar applications. We do not, however, supply a ready-to-use package; packages that cover 'standard' cases already exist. Rather, it is our intention to show the reader how to go about constructing and fitting application-specific variations of the standard models, variations that may not be covered in the currently available software. The programming exercises are intended to encourage readers to develop expertise in this respect.

The book is intended to illustrate the wonderful plasticity of hidden Markov models as general-purpose models for time series. We hope that readers will find it easy to devise for themselves 'customized' models that will be useful in summarizing and interpreting their data. To this end we offer a range of applications and types of data — Part Two is entirely devoted to applications. Some of the applications appeared in the original text, but these have been extended or refined.

Our intended readership is applied statisticians, students of statistics, and researchers in fields in which time series arise that are not amenable to analysis by the standard time series models such as Gaussian ARMA models. Such fields include animal behaviour, epidemiology, finance, hydrology and sociology. We have tried to write for readers who wish to acquire a general understanding of the models and their uses, and who wish to apply them. Researchers primarily interested in developing the theory of hidden Markov models are likely to be disappointed by the lack of generality of our treatment, and by the dearth of material on specific issues such as identifiability, hypothesis testing, properties of estimators and reversible jump Markov chain Monte Carlo methods. Such readers would find it more profitable to refer to alternative sources, such as Cappé, Moulines and Rydén (2005) or Ephraim and Merhav (2002). Our strategy has been to present most of the ideas by using a single running example and a simple model, the Poisson–hidden Markov model. In Chapter 8, and in Part Two of the book, we illustrate how this basic model can be progressively and variously extended and generalized.

We assume only a modest level of knowledge of probability and statistics: the reader is assumed to be familiar with the basic probability distributions such as the Poisson, normal and binomial, and with the concepts of dependence, correlation and likelihood. While we would not go as far as Lindsey (2004, p. ix) and state that 'Familiarity with classical introductory statistics courses based on point estimation, hypothesis testing, confidence intervals [. . .] will be a definite handicap', we hope that extensive knowledge of such matters will not prove necessary. No prior knowledge of Markov chains is assumed, although our coverage is brief enough that readers may wish to supplement our treatment by reading the relevant parts of a book such as Grimmett and Stirzaker (2001). We have also included exercises of a theoretical nature in many of the chapters, both to fill in the details and to illustrate some of the concepts introduced in the text. All the datasets analysed in this book can be accessed at the following address: http://134.76.173.220/hmm-with-r/data.

This book contains some material which has not previously been published, either by ourselves or (to the best of our knowledge) by others. If we have anywhere failed to make appropriate acknowledgement of the work of others, or misquoted their work in any way, we would be grateful if the reader would draw it to our attention. The applications described in Chapters 14, 15 and 16 contain material which first appeared in (respectively) the *South African Statistical Journal*, the *International Journal of Epidemiology* and *Biometrics*. We are grateful to the editors of these journals for allowing us to reuse such material.

We wish to thank the following researchers for giving us access to their data, and in some cases spending much time discussing it with us: David Bowie, Graham Fick, Linda Haines, Len Lerer, Frikkie Potgieter, David Raubenheimer and Max Suster.

We are especially indebted to Andreas Schlegel and Jan Bulla for their important inputs, particularly in the early stages of the project; to Christian Gläser, Oleg Nenadić and Daniel Adler, for contributing their computing expertise; and to Antony Unwin and Ellis Pender for their constructive comments on and criticisms of different aspects of our work. The second author wishes to thank the Institute for Statistics and Econometrics of Georg–August–Universität, Göttingen, for welcoming him on many visits and placing facilities at his disposal. Finally, we are most grateful to our colleague and friend of many years, Linda Haines, whose criticism has been invaluable in improving this book.

Göttingen
November 2008

Notation and abbreviations

> Since the underlying mathematical ideas are the important quantities, no notation should be adhered to slavishly. It is all a question of who is master.
>
> Bellman (1960, p. 82)

> [...] many writers have acted as though they believe that the success of the Box–Jenkins models is largely due to the use of the acronyms.
>
> Granger (1982)

Notation

Although notation is defined as it is introduced, it may also be helpful to list here the most common meanings of symbols, and the pages on which they are introduced. Matrices and vectors are denoted by bold type. Transposition of matrices and vectors is indicated by the prime symbol: $'$. All vectors are row vectors unless indicated otherwise.

Symbol	Meaning	Page
$\mathbf{A}(,i)$	i th column of any matrix $\mathbf{A}$	86
$A_n(\kappa)$	$I_n(\kappa)/I_0(\kappa)$	160
$\mathbf{B}_t$	$\mathbf{\Gamma P}(x_t)$	37
C_t	state occupied by Markov chain at time t	16
$\mathbf{C}^{(t)}$	$(C_1, C_2, \dots, C_t)$	16
$\{g_t\}$	parameter process of a stochastic volatility model	190
I_n	modified Bessel function of the first kind of order n	156
l	log-likelihood	21
L or L_T	likelihood	21, 35
log	logarithm to the base e	
m	number of states in a Markov chain,	17
	or number of components in a mixture	7
$\mathbb{N}$	the set of all positive integers	
N_t	nutrient level	220
$N(\bullet;\mu,\sigma^2)$	distribution function of general normal distribution	191
$n(\bullet;\mu,\sigma^2)$	density of general normal distribution	191
p_i	probability mass or density function in state i	31
$\mathbf{P}(x)$	diagonal matrix with i th diagonal element $p_i(x)$	32
$\mathbb{R}$	the set of all real numbers	

T	length of a time series	35
$\mathbf{U}$	square matrix with all elements equal to 1	19
$\mathbf{u}(t)$	vector $(\Pr(C_t = 1), \ldots, \Pr(C_t = m))$	17
$u_i(t)$	$\Pr(C_t = i)$, i.e. i th element of $\mathbf{u}(t)$	32
w_t	$\boldsymbol{\alpha}_t \mathbf{1}' = \sum_i \alpha_t(i)$	46
X_t	observation at time t, or just t th observation	30
$\mathbf{X}^{(t)}$	$(X_1, X_2, \ldots, X_t)$	30
$\mathbf{X}^{(-t)}$	$(X_1, \ldots, X_{t-1}, X_{t+1}, \ldots X_T)$	76
$\mathbf{X}_a^b$	$(X_a, X_{a+1}, \ldots, X_b)$	61
$\boldsymbol{\alpha}_t$	(row) vector of forward probabilities	38
$\alpha_t(i)$	forward probability, i.e. $\Pr(\mathbf{X}^{(t)} = \mathbf{x}^{(t)}, C_t = i)$	59
$\boldsymbol{\beta}_t$	(row) vector of backward probabilities	60
$\beta_t(i)$	backward probability, i.e. $\Pr(\mathbf{X}_{t+1}^T = \mathbf{x}_{t+1}^T \mid C_t = i)$	60
$\boldsymbol{\Gamma}$	transition probability matrix of Markov chain	17
γ_{ij}	(i, j) element of $\boldsymbol{\Gamma}$; probability of transition from state i to state j in a Markov chain	17
$\boldsymbol{\delta}$	stationary or initial distribution of Markov chain, or vector of mixing probabilities	18, 7
$\boldsymbol{\phi}_t$	vector of forward probabilities, normalized to have sum equal to 1, i.e. $\boldsymbol{\alpha}_t / w_t$	46
Φ	distribution function of standard normal distribution	
$\mathbf{1}$	(row) vector of ones	19

Abbreviations

ACF	autocorrelation function
AIC	Akaike's information criterion
BIC	Bayesian information criterion
CDLL	complete-data log-likelihood
c.o.d.	change of direction
c.v.	coefficient of variation
HM	hidden Markov
HMM	hidden Markov model
MC	Markov chain
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
MLE	maximum likelihood estimator or estimate
PACF	partial autocorrelation function
qq-plot	quantile-quantile plot
SV	stochastic volatility
t.p.m.	transition probability matrix