Texts in Statistical Science

Nonlinear Time Series

Theory, Methods, and Applications with R Examples

Randal Douc
Telecom SudParis
Evry, France

Eric Moulines
Telecom ParisTech
Paris, France

David S. Stoffer
University of Pittsburgh
Pennsylvania, USA

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

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

Printed on acid-free paper
Version Date: 20131111

International Standard Book Number-13: 978-1-4665-0225-3 (Hardback)

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

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

Contents

Preface **xiii**

Frequently Used Notation **xvii**

I Foundations 1

1 Linear Models **3**

1.1 Stochastic processes 3
1.2 The covariance world 5
1.2.1 Second-order stationary processes 5
1.2.2 Spectral representation 9
1.2.3 Wold decomposition 13
1.3 Linear processes 15
1.3.1 What are linear Gaussian processes? 15
1.3.2 ARMA models 16
1.3.3 Prediction 19
1.3.4 Estimation 21
1.4 The multivariate cases 25
1.4.1 Time domain 25
1.4.2 Frequency domain 27
1.5 Numerical examples 28
Exercises 30

2 Linear Gaussian State Space Models **33**

2.1 Model basics 33
2.2 Filtering, smoothing, and forecasting 36
2.3 Maximum likelihood estimation 42
2.3.1 Newton–Raphson 42
2.3.2 EM algorithm 43
2.4 Smoothing splines and the Kalman smoother 45
2.5 Asymptotic distribution of the MLE 47
2.6 Missing data modifications 49
2.7 Structural component models 50
2.8 State-space models with correlated errors 53
2.8.1 ARMAX models 54

2.8.2 Regression with autocorrelated errors 56
Exercises 56

3 Beyond Linear Models 61
3.1 Nonlinear non-Gaussian data 62
3.2 Volterra series expansion 68
3.3 Cumulants and higher-order spectra 69
3.4 Bilinear models 72
3.5 Conditionally heteroscedastic models 73
3.6 Threshold ARMA models 77
3.7 Functional autoregressive models 78
3.8 Linear processes with infinite variance 79
3.9 Models for counts 81
3.9.1 Integer valued models 81
3.9.2 Generalized linear models 83
3.10 Numerical examples 84
Exercises 89

4 Stochastic Recurrence Equations 91
4.1 The Scalar Case 93
4.1.1 Strict stationarity 93
4.1.2 Weak stationarity 98
4.1.3 GARCH(1, 1) 102
4.2 The Vector Case 107
4.2.1 Strict stationarity 109
4.2.2 Weak stationarity 111
4.2.3 GARCH(p, q) 114
4.3 Iterated random function 118
4.3.1 Strict stationarity 118
4.3.2 Weak stationarity 121
Exercises 123

II Markovian Models 131

5 Markov Models: Construction and Definitions 133
5.1 Markov chains: Past, future, and forgetfulness 133
5.2 Kernels 134
5.3 Homogeneous Markov chain 136
5.4 Canonical representation 138
5.5 Invariant measures 139
5.6 Observation-driven models 142
5.7 Iterated random functions 143
5.8 MCMC methods 152
5.8.1 Metropolis-Hastings algorithm 152
5.8.2 Gibbs sampling 155

Exercises 157

6 Stability and Convergence 165
6.1 Uniform ergodicity 166
6.1.1 Total variation distance 166
6.1.2 Dobrushin coefficient 167
6.1.3 The Doeblin condition 169
6.1.4 Examples 169
6.2 V-geometric ergodicity 173
6.2.1 V-total variation distance 173
6.2.2 V-Dobrushin coefficient 174
6.2.3 Drift and minorization conditions 175
6.2.4 Examples 180
6.3 Some proofs 186
6.4 Endnotes 188
Exercises 189

7 Sample Paths and Limit Theorems 195
7.1 Law of large numbers 196
7.1.1 Dynamical system and ergodicity 196
7.1.2 Markov chain ergodicity 203
7.2 Central limit theorem 211
7.3 Deviation inequalities for additive functionals 218
7.3.1 Rosenthal type inequality 218
7.3.2 Concentration inequality 221
7.4 Some proofs 225
Exercises 231

8 Inference for Markovian Models 239
8.1 Likelihood inference 239
8.2 Consistency and asymptotic normality of the MLE 245
8.2.1 Consistency 245
8.2.2 Asymptotic normality 247
8.3 Observation-driven models 254
8.4 Bayesian inference 263
8.5 Some proofs 271
8.6 Endnotes 274
Exercises 275

III State Space and Hidden Markov Models 285

9 Non-Gaussian and Nonlinear State Space Models 287
9.1 Definitions and basic properties 287
9.1.1 Discrete-valued state space HMM 287
9.1.2 Continuous-valued state-space models 295

9.1.3 Conditionally Gaussian linear state-space models 297
9.1.4 Switching processes with Markov regimes 300
9.2 Filtering and smoothing 302
9.2.1 Discrete-valued state-space HMM 303
9.2.2 Continuous-valued state-space HMM 310
9.3 Endnotes 314
Exercises 315

10 Particle Filtering 321
10.1 Importance sampling 321
10.2 Sequential importance sampling 329
10.3 Sampling importance resampling 334
10.3.1 Algorithm description 335
10.3.2 Resampling techniques 336
10.4 Particle filter 337
10.4.1 Sequential importance sampling 337
10.4.2 Auxiliary sampling 339
10.5 Convergence of the particle filter 341
10.5.1 Exponential deviation inequalities 341
10.5.2 Time-uniform bounds 343
10.6 Endnotes 349
Exercises 350

11 Particle Smoothing 361
11.1 Poor man's smoother algorithm 362
11.2 FFBSm algorithm 365
11.3 FFBSi algorithm 367
11.4 Smoothing functionals 369
11.5 Particle independent Metropolis-Hastings 370
11.6 Particle Gibbs 376
11.7 Convergence of the FFBSm and FFBSi algorithms 381
11.7.1 Exponential deviation inequality 384
11.7.2 Asymptotic normality 386
11.7.3 Time uniform bounds 390
11.8 Endnotes 394
Exercises 397

12 Inference for Nonlinear State Space Models 405
12.1 Monte Carlo maximum likelihood estimation 407
12.1.1 Particle approximation of the likelihood function 407
12.1.2 Particle stochastic gradient 410
12.1.3 Particle Monte Carlo EM algorithms 412
12.1.4 Particle stochastic approximation EM 415
12.2 Bayesian analysis 418
12.2.1 Gaussian linear state space models 419

12.2.2 Gibbs sampling for NLSS model 423
12.2.3 Particle marginal Markov chain Monte Carlo 428
12.2.4 Particle Gibbs algorithm 431
12.3 Endnotes 433
Exercises 435

13 Asymptotics of the MLE for NLSS 441
13.1 Strong consistency of the MLE 442
13.1.1 Forgetting the initial distribution 442
13.1.2 Approximation by conditional likelihood 444
13.1.3 Strong consistency 445
13.1.4 Identifiability of mixture densities 452
13.2 Asymptotic normality 453
13.2.1 Convergence of the observed information 458
13.2.2 Limit distribution of the MLE 460
13.3 Endnotes 461
Exercises 462

IV Appendices 467

Appendix A Some Mathematical Background 469
A.1 Some measure theory 469
A.2 Some probability theory 471

Appendix B Martingales 475
B.1 Definitions and elementary properties 475
B.2 Limits theorems 477

Appendix C Stochastic Approximation 483
C.1 Robbins–Monro algorithm: Elementary results 484
C.2 Stochastic gradient 487
C.3 Stepsize selection and averaging 488
C.4 The Kiefer–Wolfowitz procedure 488

Appendix D Data Augmentation 491
D.1 The EM algorithm in the incomplete data model 492
D.2 The Fisher and Louis identities 494
D.3 Monte Carlo EM algorithm 495
D.3.1 Stochastic approximation EM 496
D.4 Convergence of the EM algorithm 498
D.5 Convergence of the MCEM algorithm 500
D.5.1 Convergence of perturbed dynamical systems 500
D.5.2 Convergence of the MCEM algorithm 502

References **505**

Index **527**

Preface

This book is designed for researchers and students who want to acquire advanced skills in nonlinear time series analysis and their applications. Before reading this text, we suggest a solid knowledge of linear Gaussian time series, for which there are many texts. At the advanced level, texts that cover both the time and frequency domains are Anderson (1994), Brockwell and Davis (1991), and Fuller (1996). At the intermediate level, we mention Hamilton (1994), Kitagawa (2010), and Shumway and Stoffer (2011), which cover both the time and frequency domains, and Box and Jenkins (1970), which covers primarily the time domain. Hannan and Deistler (2012) is an advanced text on the statistical theory of linear state space systems. There are a number of texts that cover time series at a more introductory level, but the material covered in this text requires at least an intermediate level of understanding of the time domain.

While it is not sensible to view statistics simply as a branch of mathematics, we believe that statistical modeling and inference need to be firmly grounded in theory. Although we avoid delving into sophisticated mathematical derivations, most of the statements of the book are rigorously established. The reader is therefore expected to have some background in measure theory (covering the construction of the measure and Lebesgue integrals), and in probability theory (including conditional expectations, the construction of discrete time stochastic processes and martingales). Examples of courses covering this material are Doob (1953), Billingsley (1995), Shiryaev (1996), and Durrett (2010), among many others. Although we constantly use measure-theoretic concepts and notations, nothing excessively deep is used. An introduction to discrete state space Markov chains is clearly a plus, but is not needed.

The book represents a biased selection of topics in nonlinear time series, reflecting our own inclinations toward state-space representations. Our focus on principles is intended to provide readers with a solid background to craft their own stochastic models, numerical methods, and software, and to be able to assess the advantages and disadvantages of different approaches. We do not believe in pulling mathematical formulas out of thin air, or establishing a catalog of models and methods. Of course, this attitude reflects our mathematical orientation and our willingness to postpone the statistical discussion to pay attention to rigorous theoretical foundations.

There are a number of texts that cover nonlinear and non-Gaussian models from a variety of points-of-view. Because financial series tend to be nonlinear, there are many texts that focus primarily on finance such as Chan (2002), Franses and Van Dijk (2000), and Tsay (2005). The text by Teräsvirta et al. (2011), while focusing primarily on finance, is a rather comprehensive and approachable text on the subject. Other texts that present general statistical approaches to nonlinear time series models are Fan and Yao (2003) and Gao (2007), which take a nonparametric or semiparametric smoothing approach, Priestley (1988), which focuses on nonlinear models and spectral analysis for nonstationary processes, and Tong (1983), which introduces threshold models. Tong (1990) and Kantz and Schreiber (2004) take a dynamical systems approach and present a wide array of nonlinear time series models. Two other texts that focus primarily on a state-space approach to nonlinear and non-Gaussian time series are Kitagawa and Gersch (1996) and the second part of Durbin and Koopman (2012). MacDonald and Zucchini (2009), Fraser (2008) and Kitagawa (2010) present a state-space approach to modeling linear and nonlinear time series at an introductory level. Kitagawa (2010) could serve as a supplement for readers seeking a more gentle initial approach to the subject.

We are agnostic about the nature of statistical inference. The reader must definitely look elsewhere for a philosophical discussion of the relative merits of frequentist versus Bayesian inference. Our belief is that nonlinear time series generally benefit from analysis using a variety of frequentist and Bayesian methods. These different perspectives strengthen the conclusions rather than contradict one another.

Our hope is to acquaint readers with the main principles behind nonlinear time series models without overwhelming them with difficult mathematical developments. To keep the book length within acceptable limits, we have avoided the use of sophisticated probabilistic arguments, which underlie most of the recent developments of continuous state space Markov chains and sequential Monte Carlo methods. For the statistical part, we cover mostly the basics; other important concepts like the local asymptotic normality (e.g., Taniguchi and Kakizawa, 2000), empirical process techniques (e.g., Dehling et al., 2002), cointegration (e.g., Fuller, 1996), multivariate time series (e.g., Reinsel, 2003, Lütkepohl, 2005), model selection, semiparametric and nonparametric inference (e.g., Fan and Yao, 2003), and so on, may be found in other texts. We are, of course, responsible for any and all mistakes, misconceptions and omissions.

Although there is a logical progression through the chapters, the three parts can be studied independently. Some chapters within each part may also be read as independent surveys. Several chapters highlight recent developments such as explicit rate of convergence of Markov chains (we use the techniques outlined in Hairer and Mattingly, 2011 to discuss the geometric ergodicity of Markov chains), or sequential Monte Carlo techniques (covering for example the recently introduced particle Markov chain Monte Carlo methods found in Andrieu et al., 2010).

Any instructor contemplating a one-semester course based on this book will have to decide which chapters to cover and which to omit. The first part can be seen as a crash course on "classical" time series, with a special emphasis on linear state space models and a rather detailed coverage on random coefficient autoregressions, covering both ARCH and GARCH models. The second part is a self-contained introduction to Markov chain, discussing stability, the existence of a stationary distribution, ergodicity, limit theorems and statistical inference. Many examples are provided with the objective to develop empirical skills. We have already covered parts I and II in a fast-paced one semester advanced master level course. Part III is a self-contained account of nonlinear state space and sequential Monte Carlo methods. It is an elementary introduction to nonlinear state space modeling and sequential Monte Carlo, but it touches on many current topics in this field, from the theory of statistical inference to advanced computational methods. This has been used as a support to an advanced course on these methods, and can be used by readers who want to have an introduction to this field before studying more specialized texts such as Del Moral (2004) or Del Moral et al. (2010).

As with any textbook, the exercises are nearly as important as the main text. Statistics is not a spectator sport, so the book contains more than 200 exercises to challenge the readers. Most problems merely serve to strengthen intellectual muscles strained by the introduction of new theory; some problems extend the theory in significant ways.

We acknowledge the help of Julien Cornebise and Fredrik Lindsten who participated in the writing of the text and contributed to Chapter 11 and Chapter 12 in Part III. Julien also helped us considerably in the development of R code. We are also indebted to Pierre Priouret for suggesting various forms of improvement in the presentation, layout, and so on, as well as helping us track typos and errors. We are grateful to Christophe Andrieu, Pierre Del Moral, and Arnaud Doucet, who generously gave us some of their time to help to decipher many of the intricacies of the particle filters. This work would have not been possible without the continuous support of our colleagues and friends Olivier Cappé, Gersende Fort, Jimmy Olsson, François Roueff, and Philippe Soulier, who provided various helpful insights and comments. We also acknowledge Hedibert Freitas Lopes and Fredrik Lindsten for distributing code that became the basis of some of the R scripts used in Section 12.2. Finally, we thank John Kimmel for his support and enduring patience.

R. Douc thanks Telecom SudParis for a six month sabbatical in 2012 to work on the text. E. Moulines thanks Telecom ParisTech for giving him the opportunity to work on this project. D.S. Stoffer thanks the U.S. National Science Foundation and Telecom ParisTech for partial support during the preparation of this manuscript. In addition, most of the time spent on the text was during a visiting professorship at the Booth School of Business, University of Chicago, and their support is gratefully acknowledged.

The webpage for the text, www.stat.pitt.edu/stoffer/nltsa, contains the R scripts used in the examples and other useful information such as errata.

Paris	*Randal Douc*
Paris	*Eric Moulines*
Chicago	*David S. Stoffer*
September, 2013	