SOUTH AFRICA ACTUARIAL JOURNAL
SAAJ **10** (2010) 159–62

BOOK REVIEW

Hidden Markov Models for Time Series: An Introduction using R
by Walter Zucchini and Iain L. MacDonald. Chapman & Hall (CRC Press), 2009

Hidden Markov models (HMMs) are statistical models in which the distribution that generates the observation depends on the state of an underlying but unobserved Markov process. They provide a general approach to modelling apparently non-stationary time series, where the non-stationarity originates from an underlying Markov process influencing the distribution of the observations.

An example used extensively in the book is the annual number of major earthquakes (magnitude 7 or greater) in the world over the period 1900 to 2006. An HMM analysis suggests that the number occurring in any year has a Poisson distribution, the average number varying according to a Markov process with three (or possibly four) states, presumably corresponding to different states of the Earth's crust. A corresponding example from the actuarial field could be the number of major accident claims received by an insurance company in a year. Here the different states of the Markov process (if indeed an HMM is appropriate) could correspond to different states of the world security. Many other examples are provided in the text.

This book is in a sense a follow-up of the authors' previous book (Macdonald & Zucchini, 1997), although its scope is much broader and it is not focused principally on discrete-valued series. Models for a wide range of types of time-series data are covered, including continuous, circular and multivariate, in addition to discrete and categorical data. In addition, the authors have made the models more accessible by illustrating how the open-source computing environment R can be used to carry out the necessary computations, such as for parameter estimation and inference, model selection, model checking, decoding and forecasting, and by providing R code for many of them. While it is claimed that only a modest level of probability and statistics is required of the reader, someone without at least three years of undergraduate statistics and a reasonable familiarity with applied matrix algebra could find some of the sections heavy going. On the other hand, since most of the proofs have been left as exercises at the end of the chapters, the reader wishing to pick up the essence and practical application of HMMs need not get bogged down in the theoretical aspects.

Chapter 1 introduces the reader to the two building blocks of HMMs, being independent (finite) mixture models, which would be appropriate if one were to ignore the time-series nature of the data, and (discrete-time) Markov chains, which describe the process by which a time series changes from one state to the next. Chapter 2 covers the basic theory of HMMs, including the general form of the marginal and joint distributions of the data and hence for the likelihood function of a whole data series generated by an m-state HMM. In one of the few results that are proved in the main part of the text, a very elegant general expression for the likelihood (in matrix form) is derived, which naturally leads to a convenient computational algorithm for it. An additional advantage is that the likelihood easily accommodates situations where data are missing at random or observations are interval-censored.

Despite the convenient algorithm for computing the likelihood, parameter estimation by maximising it is far from simple. Two alternative approaches are described, the first being by direct maximisation and the second via the expectation-maximisation (EM) algorithm. Because of the relatively modest number of operations involved in evaluating the likelihood, direct numerical maximisation with respect to the parameters is quite feasible, even when the series is long. The main problems are numerical underflow (since the likelihood is made up of a product of matrices, the usual solution of taking logs and converting the product to a sum is not possible), constraints on the parameters (in particular the elements of the Markov transition matrix) and the possible occurrence of multiple local maxima in the likelihood. Chapter 3 discusses several solutions to these problems, as well as the computation of standard errors and confidence intervals, via the Hessian matrix and using the asymptotic normality of the maximum-likelihood estimators. However, the latter can be unreliable for moderate-length series, or if some of the parameters are at or near the boundary of their parameter space, and as an alternative the parametric bootstrap is described for achieving more reliable confidence intervals.

Use of the EM algorithm for fitting HMMs is described in Chapter 4. Broadly speaking, EM is an iterative algorithm for maximum-likelihood estimation when some of the data are missing: in the E-step the missing data are estimated conditionally—the estimates being based on the observed data and the current parameter estimates—and in the M-step the completed log-likelihood, with the missing data replaced by their estimates, is maximised, giving updated estimates for the parameters. These two steps are repeated many times until some convergence criterion is satisfied. In the case of an HMM, the unknown states of the Markov process at each point in the series are considered as missing values that are estimated in the E-step, and then the likelihood with these 'known' states is maximised in the M-step. This turns out to be a relatively simple process, depending largely on how difficult it is to maximise the log-likelihood of the state-dependent distributions of the observations. Generally, use of the EM algorithm would be preferred to numerical maximisation, but as discussed in this chapter, this is not always the case for HMMs, so R code for implementing both approaches is given in an appendix.

Chapter 5 applies the material of the previous chapters to the practical topics of forecasting, decoding and state prediction. Having fitted an HMM to a time series, the user can forecast the series (including forecast intervals), and 'decode' it. Decoding involves the allocation of the most likely Markov state underlying each of the points in the series. Convenient formulae exist for these purposes. Decoding comprises 'local decoding' and 'global decoding'. Local decoding allocates the states separately for each time point, whereas global decoding (achieved by dynamic programming via the Viterbi algorithm) finds the most likely sequence of states over all time points. State prediction is the prediction of the probabilities of the different states at any time in the future. Again, R code for implementing the required computations is provided in an appendix.

Chapter 6 covers the important topic of model selection, generally with respect to the number of states in the underlying Markov process, but possibly also with respect to assumptions regarding the structure of its transfer-probability matrix, or of the state-dependent distributions. This is most readily achieved through the Akaike information criterion or the Bayesian information criterion, and use of both is illustrated by applying them to the earthquake data. Model checking is achieved through the use of pseudo-residuals, which are used both to help decide whether a selected model is indeed adequate and to check for outliers. These are defined and, as is the practice in this book, are applied to data sets to illustrate their use and extensions.

Bayesian inference is covered in the next chapter, which focuses on Poisson HMMs. As with other aspects of HMMs, there are several obstacles to overcome in the Bayesian approach, some of which are illustrated in the choice of prior distributions of the parameters and use of the Gibbs sampler to compute the corresponding posterior distributions.

The last chapter in part 1 of the book considers extensions of the basic HMM, the first being the use of state-dependent distributions other than the Poisson, which is hardly an extension and follows naturally from the definition of HMMs. One problem that arises with continuous-valued time series is that the likelihood function may be unbounded, in particular with the normal distribution, but an easy solution via discretisation is available if that occurs. Other generalisations that are discussed, and which illustrate the general flexibility of HMMs, are:

— HMMs based on second-order Markov processes;
— HMMs for multivariate time series;
— HMMs that depend on covariates:
  – in the state-dependent distributions, or
  – in the transition probabilities of the underlying Markov process;

  or both;
— models with additional forms of dependence, viz.:
  – direct dependence between the observations of the time series;
  – dependence of the time series, not only on the current (hidden) state, but also on the previous state; or
  – dependence of the following state on the current observation.

The first eight chapters make up part 1 of the book; part 2, comprising another eight chapters, is entirely devoted to applications. In addition to illustrating the wide variety of

time series that can fruitfully be analysed as HMMs, a number of different types of data, models and extensions that are not covered in part 1, or are only briefly mentioned, are illustrated here. All these applications are based on real data series, so they are also worth reading for their own sakes. Chapter 13 describes models for financial series, which should be of direct interest to actuaries. The three series are:
— thinly traded shares on the JSE Securities Exchange (univariate and multivariate models);
— multivariate normal HMM for the returns on four shares; and
— discrete state-space stochastic volatility models.

There are two appendices, the first giving R code for most of the computations required in part 1, mostly, but not entirely, with respect to Poisson HMMs. Generally, the modifications required to adapt them to other distributions will be straightforward for someone versed in R. A second appendix presents proofs of some results used in part 1. There is an extensive list of references at the end.

This is an excellent book, which should be of great interest to the research-oriented actuary. Indeed, it could provide an entirely new and fruitful approach to analysing actuarial time series, providing insights that would not be available from other forms of analysis. For the student, or anyone wishing to study HMMs in depth, the exercises at the end of each chapter in part 1, and the first four chapters in part 2, provide very useful challenges and tests of understanding.

L. Paul Fatti

REFERENCE

MacDonald, IL and Zucchini, W (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman & Hall