
Preface

Hidden Markov models—most often abbreviated to the acronym “HMMs”—are one of the most successful statistical modelling ideas that have come up in the last forty years: the use of hidden (or unobservable) states makes the model generic enough to handle a variety of complex real-world time series, while the relatively simple prior dependence structure (the “Markov” bit) still allows for the use of efficient computational procedures. Our goal with this book is to present a reasonably complete picture of statistical inference for HMMs, from the simplest finite-valued models, which were already studied in the 1960’s, to recent topics like computational aspects of models with continuous state space, asymptotics of maximum likelihood, Bayesian computation and model selection, and all this illustrated with relevant running examples. We want to stress at this point that by using the term *hidden Markov model* we do not limit ourselves to models with finite state space (for the hidden Markov chain), but also include models with continuous state space; such models are often referred to as *state-space models* in the literature.

We build on the considerable developments that have taken place during the past ten years, both at the foundational level (asymptotics of maximum likelihood estimates, order estimation, etc.) and at the computational level (variable dimension simulation, simulation-based optimization, etc.), to present an up-to-date picture of the field that is self-contained from a theoretical point of view and self-sufficient from a methodological point of view. We therefore expect that the book will appeal to academic researchers in the field of HMMs, in particular PhD students working on related topics, by summing up the results obtained so far and presenting some new ideas. We hope that it will similarly interest practitioners and researchers from other fields by leading them through the computational steps required for making inference in HMMs and/or providing them with the relevant underlying statistical theory.

The book starts with an introductory chapter which explains, in simple terms, what an HMM is, and it contains many examples of the use of HMMs in fields ranging from biology to telecommunications and finance. This chapter also describes various extensions of HMMs, like models with autoregression

or hierarchical HMMs. Chapter 2 defines some basic concepts like transition kernels and Markov chains. The remainder of the book is divided into three parts: *State Inference*, *Parameter Inference* and *Background and Complements*; there are also three appendices.

Part I of the book covers inference for the unobserved state process. We start in Chapter 3 by defining smoothing, filtering and predictive distributions and describe the forward-backward decomposition and the corresponding recursions. We do this in a general framework with no assumption on finiteness of the hidden state space. The special cases of HMMs with finite state space and Gaussian linear state-space models are detailed in Chapter 5. Chapter 3 also introduces the idea that the conditional distribution of the hidden Markov chain, given the observations, is Markov too, although non-homogeneous, for both ordinary and time-reversed index orderings. As a result, two alternative algorithms for smoothing are obtained. A major theme of Part I is simulation-based methods for state inference; Chapter 6 is a brief introduction to Monte Carlo simulation, and to Markov chain Monte Carlo and its applications to HMMs in particular, while Chapters 7 and 8 describe, starting from scratch, so-called sequential Monte Carlo (SMC) methods for approximating filtering and smoothing distributions in HMMs with continuous state space. Chapter 9 is devoted to asymptotic analysis of SMC algorithms. More specialized topics of Part I include recursive computation of expectations of functions with respect to smoothed distributions of the hidden chain (Section 4.1), SMC approximations of such expectations (Section 8.3) and mixing properties of the conditional distribution of the hidden chain (Section 4.3). Variants of the basic HMM structure like models with autoregression and hierarchical HMMs are considered in Sections 4.2, 6.3.2 and 8.2.

Part II of the book deals with inference for model parameters, mostly from the maximum likelihood and Bayesian points of views. Chapter 10 describes the expectation-maximization (EM) algorithm in detail, as well as its implementation for HMMs with finite state space and Gaussian linear state-space models. This chapter also discusses likelihood maximization using gradient-based optimization routines. HMMs with continuous state space do not generally admit exact implementation of EM, but require simulation-based methods. Chapter 11 covers various Monte Carlo algorithms like Monte Carlo EM, stochastic gradient algorithms and stochastic approximation EM. In addition to providing the algorithms and illustrative examples, it also contains an in-depth analysis of their convergence properties. Chapter 12 gives an overview of the framework for asymptotic analysis of the maximum likelihood estimator, with some applications like asymptotics of likelihood-based tests. Chapter 13 is about Bayesian inference for HMMs, with the focus being on models with finite state space. It covers so-called reversible jump MCMC algorithms for choosing between models of different dimensionality, and contains detailed examples illustrating these as well as simpler algorithms. It also contains a section on multiple imputation algorithms for global maximization of the posterior density.

Part III of the book contains a chapter on discrete and general Markov chains, summarizing some of the most important concepts and results and applying them to HMMs. The other chapter of this part focuses on order estimation for HMMs with both finite state space and finite output alphabet; in particular it describes how concepts from information theory are useful for elaborating on this subject.

Various parts of the book require different amounts of, and also different kinds of, prior knowledge from the reader. Generally we assume familiarity with probability and statistical estimation at the levels of Feller (1971) and Bickel and Doksum (1977), respectively. Some prior knowledge of Markov chains (discrete and/or general) is very helpful, although Part III does contain a primer on the topic; this chapter should however be considered more a brush-up than a comprehensive treatise of the subject. A reader with that knowledge will be able to understand most parts of the book. Chapter 13 on Bayesian estimation features a brief introduction to the subject in general but, again, some previous experience with Bayesian statistics will undoubtedly be of great help. The more theoretical parts of the book (Section 4.3, Chapter 9, Sections 11.2–11.3, Chapter 12, Sections 14.2–14.3 and Chapter 15) require knowledge of probability theory at the measure-theoretic level for a full understanding, even though most of the results as such can be understood without it.

There is no need to read the book in linear order, from cover to cover. Indeed, this is probably the wrong way to read it! Rather we encourage the reader to first go through the more algorithmic parts of the book, to get an overall view of the subject, and then, if desired, later return to the theoretical parts for a fuller understanding. Readers with particular topics in mind may of course be even more selective. A reader interested in the EM algorithm, for instance, could start with Chapter 1, have a look at Chapter 2, and then proceed to Chapter 3 before reading about the EM algorithm in Chapter 10. Similarly a reader interested in simulation-based techniques could go to Chapter 6 directly, perhaps after reading some of the introductory parts, or even directly to Section 6.3 if he/she is already familiar with MCMC methods. Each of the two chapters entitled “Advanced Topics in...” (Chapters 4 and 8) is really composed of three disconnected complements to Chapters 3 and 7, respectively. As such, the sections that compose Chapters 4 and 8 may be read independently of one another. Most chapters end with a section entitled “Complements” whose reading is not required for understanding other parts of the book—most often, this section mostly contains bibliographical notes—although in some chapters (9 and 11 in particular) it also features elements needed to prove the results stated in the main text.

Even in a book of this size, it is impossible to include all aspects of hidden Markov models. We have focused on the use of HMMs to model long, potentially stationary, time series; we call such models *ergodic HMMs*. In other applications, for instance speech recognition or protein alignment, HMMs are used to represent short variable-length sequences; such models are often called

left-to-right HMMs and are hardly mentioned in this book. Having said that we stress that the computational tools for both classes of HMMs are virtually the same. There are also a number of generalizations of HMMs which we do not consider. In Markov random fields, as used in image processing applications, the Markov chain is replaced by a graph of dependency which may be represented as a two-dimensional regular lattice. The numerical techniques that can be used for inference in hidden Markov random fields are similar to some of the methods studied in this book but the statistical side is very different. Bayesian networks are even more general since the dependency structure is allowed to take any form represented by a (directed or undirected) graph. We do not consider Bayesian networks in their generality although some of the concepts developed in the Bayesian networks literature (the graph representation, the sum-product algorithm) are used. Continuous-time HMMs may also be seen as a further generalization of the models considered in this book. Some of these “continuous-time HMMs”, and in particular partially observed diffusion models used in mathematical finance, have recently received considerable attention. We however decided this topic to be outside the range of the book; furthermore, the stochastic calculus tools needed for studying these continuous-time models are not appropriate for our purpose.

We acknowledge the help of Stéphane Boucheron, Randal Douc, Gersende Fort, Elisabeth Gassiat, Christian P. Robert, and Philippe Soulier, who participated in the writing of the text and contributed the two chapters that compose Part III (see next page for details of the contributions). We are also indebted to them for suggesting various forms of improvement in the notations, layout, etc., as well as helping us track typos and errors. We thank François Le Gland and Catherine Matias for participating in the early stages of this book project. We are grateful to Christophe Andrieu, Søren Asmussen, Arnaud Doucet, Hans Künsch, Steve Levinson, Ya’acov Ritov and Mike Titterton, who provided various helpful inputs and comments. Finally, we thank John Kimmel of Springer for his support and enduring patience.

Paris, France
& Lund, Sweden
March 2005

Olivier Cappé
Eric Moulines
Tobias Rydén



<http://www.springer.com/978-0-387-40264-2>

Inference in Hidden Markov Models

Cappé, O.; Moulines, E.; Ryden, T.

2005, XVII, 653 p., Hardcover

ISBN: 978-0-387-40264-2