

fact, I am somewhat hopeful that this is the first part of a two part series.

TALEA L. MAYO
University of Central Florida

Hidden Markov Processes: Theory and Applications to Biology. By M. Vidyasagar. Princeton University Press, Princeton, NJ, 2014. \$59.50. xiv+287 pp., hardcover. ISBN 978-0-691-13315-7.

Vidyasagar's book provides a rigorous introduction to selected topics in the mathematical theory of hidden Markov processes (HMPs) and a sampling of applications of hidden Markov models (HMMs) in biology. The book's specific focus is on two main areas: realization theory, which deals with necessary and sufficient conditions for constructing a stochastic process as an HMP, and large deviation theory for Markov processes, which characterizes the (exponential) rate at which empirical estimates converge to their true values. Additional advanced topics in HMMs covered in the book are ergodicity and alternative representations for HMPs. The coverage of applications in biology is fairly sparse and not necessarily cohesive or representative. The book will appeal primarily to readers interested in the specific advanced concepts in the theory of HMPs that are the focus of the book, whereas readers interested in a broader background in HMMs or particularly focused on biological applications will find the book less useful.

The book is organized as nine chapters grouped into three parts. Part 1, comprising Chapters 1 through 3, provides background in probability and random variables, information theory, and nonnegative matrices. Part 2, comprising Chapters 4 through 7, covers Markov processes, introductory large deviation theory, basic properties of HMPs, and the complete realization problem for HMPs. Part 3, comprising Chapters 8 and 9, covers biological applications. The material in Chapter 1 ("Introduction to Probability and Random Variables") and Chapter 2 ("Introduction to Information Theory") is fairly standard and should be familiar to

most readers who have the mathematical background required for the book. Inclusion of this material is beneficial, nonetheless, because it makes the book self-contained and provides required background in a notational convention that is coherent with the rest of the book. Chapter 3 covers the theory of nonnegative matrices, anticipating the close connection between nonnegative matrices and the transition probability matrices for finite state Markov chains. The material in Part 2 is the core material of the book. Chapter 4 develops the theory of Markov processes, making extensive use of the infrastructure of nonnegative matrices developed in Chapter 3. The treatment of ergodicity is one of the advanced concepts included within Chapter 4 that is often missing from basic treatments of Markov processes. Chapter 5 is an introduction to large deviations theory; the key results in this chapter are the rate functions for iid and Markov processes, which derived using the method of types. Chapter 6 covers the basic properties of HMPs. Three alternative representations for defining HMMs are introduced and shown to be mathematically equivalent, and the fundamental problems of inference and parameter estimation for HMMs are discussed. Specifically, given an observed HMP sample, the Viterbi algorithm for obtaining the maximum a posteriori probability (MAP) estimate of the sequence of state transitions is derived and an algorithm is described for (reestimating) the HMM parameters based on the Viterbi estimate. Chapter 8 characterizes the necessary and sufficient conditions for a given stationary stochastic process with a finite sample space of outcomes to be an HMP. The necessary condition is shown to correspond to a finite rank requirement on an infinite Hankel style matrix formed from the statistics of the process. This condition is, however, shown to not be sufficient and a meaningful sufficient condition is obtained only under significant additional assumptions on the process. Chapter 8 initiates a discussion of applications in computational biology. After a very brief introduction to some of the relevant aspects of molecular biology, the basic sequence alignment problem is for-

mulated and a dynamic programming approach is presented for solving the problem. An overview is provided of the GLIMMER and GENSCAN methods, which make use of interpolated Markov models and HMMs, respectively, for gene finding. Finally, the use of profile HMMs for protein classification is briefly reviewed. Chapter 9 covers the theory behind the BLAST algorithm. In particular, the chapter focuses on how the theory of large deviations for i.i.d. processes is utilized in the BLAST methodology to ascertain the significance of high scoring sequence segments found by the algorithm.

HMPs and HMMs have been extensively studied for a long time and there are numerous other publications on the topic of HMMs. It is valuable to compare and contrast this book against those I am most familiar with. The book is almost entirely complementary to the landmark tutorial article by Rabiner [5]. Whereas Rabiner's article focuses entirely on the pragmatic aspects of inference and parameter estimation using HMMs and provides excellent guidance for practical implementation, the present book focuses entirely on the theory without delving into any of the practical aspects. For example, Rabiner's article has an extensive discussion of scaling, which is required in order to address dynamic range limitations of floating point representations in implementations of the forward-backward and Baum-Welch algorithms, whereas the current book does not even outline these algorithms adequately. For bioinformaticians, the standard reference on the topic of HMMs is the book by Durbin et al [3], which is also quite distinct from the current book. Coverage of biological applications is much more cohesive and representative in the Durbin book, whereas the sampling in the present book is much more limited and somewhat haphazard. Specifically, in a strange omission, the basic two-sequence alignment HMM (see [3, Chap. 4]) is not even developed and only the heuristically formulated Needleman Wunsch [4] and Smith-Waterman [7] dynamic programming algorithms are presented for sequence alignment in Chapter 8. The present book does have the advantage over both the Rabiner tutorial and the

Durbin book in its coverage of advanced topics in the theory of HMMs. The material on these advanced topics, which is included in Part 2, is largely reproduced from two of the author's published papers [8, 9].

I encountered several errors and typos in the book. The most significant of the observed errors were in Chapter 6 dealing with the basic properties of HMPs. In section 6.2.3 the description of the Baum-Welch algorithm is not correct. The Baum-Welch algorithm works in conjunction with the forward-backward (BCJR) algorithm [1] and not with the Viterbi algorithm (for the correct version, readers should see either [5, section III.C] or [3, pp. 63–65]). The algorithm, as described in section 6.2.3, is sometimes referred to as Viterbi training [3, pp. 64–65] and is a crude approximation of the Baum-Welch algorithm, which is an instance of expectation maximization [2]. Also, in Chapter 6, equation (6.15) that lists the so-called “backward term” in the forward-backward (BCJR) algorithm is incorrect. The correct equation should be

$$\beta_i(\mathbf{u}, k) = \Pr \{ (Y_{k+1}, Y_{k+2}, \dots, Y_l) \mid X_k = i \}.$$

Another “less serious but irksome all the same” error is in Example 3.1 in Chapter 3 on nonnegative matrices, where the reachability matrix M in the example does not appear to be correct for the nonnegative (connectivity) matrix A . Furthermore, the text description also appears to be inconsistent with the version of the reachability matrix as provided in the example. Because a number of errors and typos were found upon detailed examination of only selected portions of the book, it is likely that additional errors also exist, several of which may not be easy for readers to immediately identify. Therefore, a useful service to readers would be an errata provided by the author or publisher that is updated periodically.

The coverage of material in the book is quite uneven in multiple aspects and does not appear to clearly align with a specific focus or objective. For instance, it is unclear why the uniqueness of the entropy function is belabored in Chapter 2. This material appears to be drawn directly from Shannon's epic paper [6] and bears no connection at all with the topics covered in

the rest of the book. Some of the more elementary chapters and sections feature exercises and abundant examples, whereas there are no exercises and only sparse examples for the advanced topics that are the particular focus of the book, where readers would benefit most from meaningful exercises and examples. As already indicated, the coverage of applications in biology is haphazard and not particularly representative. While the significance and impact of the BLAST algorithm in computational biology cannot be overemphasized, the material in Chapter 9 appears to be an oddball inclusion in the book as it does not relate to HMMs per se—the results used come entirely from the theory of large deviations for i.i.d. sequences. Also, key linkages between topics, and of the material to the central theme of the book, are missing in several places. For instance, the Viterbi estimation and forward-backward recursions for HMMs are also instances of the dynamic programming that was referred to in connection with the heuristic Needleman Wunsch [4] and Smith–Waterman [7] algorithms. Furthermore, by considering scores defined as log-odds ratios of the transition and emission probabilities for HMMs, the pair alignment HMM (which is strangely not included in the book) reduces to an additive score maximization alignment algorithm analogous to the heuristic schemes considered.

The writing style in the book is verbose and desultory. Tangential observations, historical background, and personal opinions are inserted at various random locations within the main thread of presentation instead of being more appropriately organized as notes at the end of each chapter. Currently, many of these observations are also anticipatory and do not necessarily make sense unless the reader has prior familiarity with the topic or until the following material has been read. The writing is not characterized by the frugality of words and precision that is the hallmark of classic mathematical texts. For instance, statements regarding the properties of convex and concave functions, in Chapter 2, are repetitive because of the inclusion of essentially the same results, separately, for both convex and concave functions. Greater economy of space

and clarity would be accomplished by simply noting that concave functions are those whose negatives yield convex functions, and then stating results only for convex functions. Another useful feature absent from the book is a table summarizing the notational conventions and acronyms adopted. Such a table would help the reader to keep track of the many different concepts, terms, notation, and acronyms that are used throughout the book, allowing them to better follow results and proofs that are built up sequentially over multiple chapters.

An objective of the book, as stated in the preface, is to provide a treatment of HMPs that includes advanced notions such as ergodicity, representations, analysis of large deviations, and realization of HMPs, while “keeping technicalities to an absolute minimum.” In fact, the author states in the preface, “By restricting attention to Markov processes with finite state spaces, I try to capture most of the interesting phenomena such as ergodicity and large deviation theory, while giving elementary proofs that are accessible to anyone who knows undergraduate-level linear algebra.” Conceptually, by restricting attention to the setting of a finite state space, the presentation is more broadly accessible, for example, to readers without a background in measure theory. In practice, however, readers of the book will quickly realize that a high level of mathematical maturity is required to follow the material. Readers with only an undergraduate exposure to linear algebra will invariably find it extremely challenging to understand the material. In fact, parts of this review also need to rely on technical terminology that assumes that the reader has prior familiarity with the concepts being discussed.

Overall, the book would have benefited from more careful editing to ensure a clear focus and a terser and more uniform treatment of the material throughout. In its present form, the book appears to be the result of a not-particularly-concerted effort to build a scaffolding for the advanced concepts in the theory of HMPs explored in two of the author’s papers [8, 9]. Readers wishing to understand the results of these two papers, who also want the necessary background material to be introduced in con-

text, would benefit most from the current book, whereas those looking for a broader exposition with greater context for theory, practice, and applications in biology will not be particularly well-served.

REFERENCES

- [1] L. R. BAHL, J. COCKE, F. JELINEK, AND J. RAVIV, *Optimal decoding of linear codes for minimizing symbol error rate*, IEEE Trans. Inform. Theory, IT-20 (1974), pp. 284–287.
- [2] A. P. DEMPSTER, N. M. LAIRD, AND D. B. RUBIN, *Maximum likelihood from incomplete data via the EM algorithm*, J. Roy. Statist. Soc., 39 (1977), pp. 1–38.
- [3] R. DURBIN, S. R. EDDY, A. KROGH, AND G. MITCHISON, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1999.
- [4] S. B. NEEDLEMAN AND C. D. WUNSCH, *A general method applicable to the search for similarities in the amino acid sequence of two proteins*, J. Mol. Biol., 48 (1970), pp. 443–453.
- [5] L. RABINER, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proc. IEEE, 77 (1989), pp. 257–286.
- [6] C. E. SHANNON, *A mathematical theory of communication*, Bell Syst. Tech. J., 27 (1948), pp. 379–423, 623–656.
- [7] T. F. SMITH AND M. S. WATERMAN, *Identification of common molecular sub-sequences*, J. Mol. Biol., 147 (1981), pp. 195–197.
- [8] M. VIDYASAGAR, *The complete realization problem for hidden Markov models: A survey and some new results*, Math. Control Signals Systems, 23 (2011), pp. 1–65.
- [9] M. VIDYASAGAR, *An elementary derivation of the large deviation rate function for finite state Markov chains*, Asian J. Control, 16 (2014), pp. 1–19.

GAURAV SHARMA
University of Rochester

An Introduction to Mathematical Epidemiology. By Maia Martcheva. Springer, New York, 2015. \$79.99. xiv+453 pp., hardcover. ISBN 978-1-4899-7611-6.

In 2008, I published an introductory textbook explaining the basics of mathematical epidemiology for a nonexpert audience [1]. It covered simple epidemic models, the basic reproductive ratio (and, in particular, its failure, which was one of the first publications to do so), vector-borne diseases, fitting models to data, and discrete epidemic models. I did so with a lightness of touch with respect to the mathematical details. *An Introduction to Mathematical Epidemiology* covers simple epidemic models, the basic reproductive ratio (and, in particular, its failure), vector-borne diseases, fitting models to data, and discrete epidemic models. To be fair, it also covers global stability, multistrain disease dynamics, optimal control, age- and class-structure, and immuno-epidemic modeling.

This isn't necessarily a problem, as there's utility in presenting these topics with the mathematical details filled in, as in the first half of the book. The new material is also worthy of consideration. It's unacceptable that the source material isn't referenced at all, however. There are even statements like "some researchers believe that it should not be called a reproduction number" (p. 110) without any attributions.

In general, the book presents a variety of ordinary differential equation (ODE) models. These are fine, although it does become a bit samey after a while. However, the models chosen are somewhat simplistic. In particular, a variety of deadly diseases are modeled—e.g., malaria (p. 70) or TB (p. 165)—with the assumption that there is no death rate due to disease. This appears to be done so that the models are more tractable, but they ignore the biological reality of the situation. Indeed, the concept of a disease-specific death rate is not introduced in any of the models until over halfway through. This is highly unrealistic.

The definition of chaos is missing topological transitivity (p. 81), which is entirely misleading: wholly unstable systems are also aperiodic and have sensitive dependence on initial conditions, but they aren't chaotic. There are also numerous issues with the proofreading, such as comma splices (e.g., pp. 116, 224, 389, 422) and glaring typos like "is always grater than" (p. 153). Sloppiness like this and the lack of