PSYCHOMETRIKA—VOL. 89, NO. 2, 741–743
JUNE 2024
https://doi.org/10.1007/s11336-024-09958-5

BOOK REVIEW

VISSER, I., & SPEEKENBRINK, M. (2022). *Mixture and Hidden Markov Models with R. Springer, Cham, CH*.

The book by Visser and Speekenbrink (2022) provides an integrated framework for finite mixture (FM) and hidden Markov (HM) models. FM models (e.g., Frühwirth-Schnatter et al., 2019; McLachlan et al., 2019) have been proposed since many years and nowadays have become very popular due to the rise of many classification and clustering challenges as a consequence of the increasing availability of large and complex data, especially in behavioral science; see, among others, Everitt & Hand (1981). These models are typically specified by formulating the conditional distribution of the response variables given mixture components having certain probabilities, also known as weights, so that individuals assigned to the same component share a common distribution of these variables. The Gaussian FM model (see McLachlan and Peel 2000, for a thorough illustration) is the most common model-based clustering method for analyzing continuous responses that also allows for classification, which is "*…the process of assigning group membership labels to unlabeled observations. In this context, a group may be a class or a cluster…*" (McNicholas, 2016).

HM models are dependent mixture models proposed for analyzing time-series data; see, among others, MacDonald and Zucchini (1997), and the review provided in Ephraim and Merhav (2002). Then, they have been extensively developed and used for the analysis of longitudinal data starting from the work of Wiggins (1973); see, among others, Bartolucci et al. (2013) for detailed illustrations and extensions of these models and Visser and Speekenbrink (2014) for a related discussion. HM models assume that the response variables depend on a latent process that follows a Markov chain, typically of first order and homogeneous over time. Over the last two decades, these models have had a profound impact on applied research for analyzing multivariate longitudinal data in particular. Their popularity is due to mathematical tractability that allows them to handle the temporal structure of many observed phenomena. In this framework, it is also possible to account for missing data, which frequently arise under different structures. HM models provide a practical model-based dynamic clustering method (Bouveyron et al., 2019) for identifying latent group structures, also represented in terms of trajectories.

The book of Visser and Speekenbrink (2022) is a compendium of these models with examples based on simulated and real data arising from social sciences research fields, especially psychology. The authors are expert researchers who have contributed to the theoretical and computational development of the models at issue for many years. The book links theory and applications through the open-source statistical environment R (R Core Team, 2024).

1. Book Contents

Each chapter of the book provides a general introduction to specific models and methods, along with applications based on basic R functions and certain R packages. The output from the used software is discussed in some detail. A GitHub repository at https://depmix.github.io/hmmr/ conveniently collects all R codes and data proposed in the book. These are also available as an R package called `hmmr`. The authors make extensive use of the functions of the package `depmixS4` (Visser & Speekenbrink, 2010) developed and currently maintained by them.

Chapter 1 introduces basic R commands and illustrates the data used in the book. This chapter also discusses the main research questions related to proposed datasets that can be answered through these models.

Chapter 2 covers FMs, and latent class models, along with computational methods for estimating them and, specifically, maximum likelihood estimation using the expectation–maximization algorithm. This chapter also covers parameter identification and model selection (i.e., choosing the number of mixture components or latent classes).

Chapter 3 provides an overview through empirical applications of the univariate and multivariate FMs and latent class models. Bootstrap methods are introduced to compute standard errors.

Chapter 4 introduces HM models, along with maximum likelihood methods for estimating them. This chapter also covers how the models can accommodate different missing data mechanisms. Details on numerical optimization are illustrated with the package `depmixS4`.

Chapter 5 focuses on univariate HM models with empirical applications related to psychology, which have been investigated in previous authors' articles.

Chapter 6 illustrates applications of the HM models to multivariate time-series based on the R packages developed by the authors.

Chapter 7 concludes the book with some advanced topics in HM models, such as higher-order serial dependencies and Bayesian estimation. It illustrates these topics with reproducible R code examples. The problem of obtaining standard errors for assessing the uncertainty of the estimated parameters is treated in particular detail.

2. Intended Audience

Basic elements of statistical inference and computational procedures are a preliminary requirement, and therefore, the book is ideal for instructors, graduate and postgraduate students, and applied researchers. Throughout the book, the authors illustrate the implementation of the methods with several data examples, codes, visualization of data, and results using an expository style. The book is intended to guide the reader step by step on how to apply basic versions of FM and HM models and keep the text generally accessible to a broad audience, mainly from sociology, psychology, and economics, who have not used these models before and would apply them.

3. Suggested Improvements for Future Editions

While we believe that the book by Visser and Speekenbrink (2022) is a valuable contribution to the development of knowledge and applications of FM, latent class, and HM models, there are some issues that we hope the authors could consider for future editions of the book. First, the initial introduction to the software could be avoided and the authors may refer to specific references. At the same time, the authors could clearly define the different data structures (i.e., cross-section, time-series, and longitudinal data), which can be analyzed by the models illustrated in the book, and they could also avoid to illustrate Gaussian FM models using time-series data characterized by serial dependence.

Second, while the book introduces FM and HM models, along with their interpretation, a similar introduction is absent for the latent class model, although this is one of the topics of the book, and this could be added. Certain concepts of Bayesian inference could be dealt with more detail, as well as how to include covariates in HM models, which could be part of one of the advanced sections.

Third, the authors could add some important references about popular models proposed in the literature, such as the multilevel latent class model (Vermunt 2003) and latent class item response theory models, which may be employed to perform hierarchical clustering (Bartolucci

et al., 2015). Certain R packages mentioned in the book, such as **poLCA** (Linzer & Lewis, 2011), should be properly cited, and the authors could also mention some other relevant packages for estimating FM models, such as **flexmix** (Grün & Leisch, 2008).

Finally, the first three chapters present an interesting section on further readings useful for scholars exploring more advanced versions of the proposed models. However, this section is not present in the last three chapters. Due to the gained popularity of the HM models, it could be really useful to provide more details for the readers in the later chapters as well. Moreover, the last chapter could be stronger if it will also refer to many of the other popular statistical software to estimate FM models, such as Latent GOLD (Vermunt and Magidson, 2021), Stata (GLAMM, (Rabe-Hesketh et al., 2004), and other R packages to estimate FM or HM models such as **Mclust** (Scrucca et al., 2023) and **LMest** (Bartolucci et al., 2017).

UNIVERSITY OF PERUGIA *Francesco Bartolucci*

UNIVERSITY OF MILANO-BICOCCA *Fulvia Pennoni*

References

Bartolucci, F., Bacci, S., & Gnaldi, M. (2015). *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata*. Chapman & Hall/CRC.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC Press.

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2017). LMest: An R package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software, 81*, 1–38.

Bouveyron, C., Celeux, G., Murphy, T. B., & Raftery, A. E. (2019). *Model-Based Clustering and Classification for Data Science: With Applications in R*. Cambridge University Press.

Ephraim, Y., & Merhav, N. (2002). Hidden Markov processes. *IEEE Transactions on Information Theory, 48*, 1518–1569.

Everitt, B. S., & Hand, D. J. (1981). *Finite Mixture Distributions*. Chapman and Hall/CRC Press.

Frühwirth-Schnatter, S., Celeux, G., & Robert, C. P. (2019). *Handbook of Mixture Analysis*. Chapman and Hall/CRC Press.

Grün, B., & Leisch, F. (2008). FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software, 28*, 1–35.

Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software, 42*, 1–29.

MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall/CRC Press.

McLachlan, G., & Peel, D. (2000). *Finite Mixture Models*. Wiley.

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application, 6*, 355–378.

McNicholas, P. D. (2016). *Mixture Model-Based Classification*. Chapman and Hall/CRC Press.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). GLLAMM manual. *Work. Pap., Div. Biostat., Univ. Calif., Berkeley*.

Scrucca, L., Fraley, C., Murphy, T. B., & Raftery, A. E. (2023). *Model-Based Clustering, Classification, and Density Estimation Using Mclust in R*. Chapman and Hall/CRC Press.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology, 33*, 213–239.

Vermunt, J. K., & Magidson, J. (2021). Upgrade manual for latent GOLD basic, advanced, syntax, and choice Version 6.0. *Statistical Innovations Inc.*.

Visser, I., & Speekenbrink, M. (2010). depmixS4: An R package for hidden Markov models. *Journal of Statistical Software, 36*, 1–21.

Visser, I., & Speekenbrink, M. (2014). The happy marrige between latent and hidden Markov models. Comments on: Latent Markov models: A review of a general framework for the analysis of longitudinal data with covariates. *Test, 23*, 478–483.

Visser, I., & Speekenbrink, M. (2022). *Mixture and hidden Markov models with R*. Springer.

Wiggins, L. (1973). *Panel analysis: Latent probability models for attitude and behaviour processes*. Elsevier.

Published Online Date: 3 APR 2024