

MIT Press • Open Encyclopedia of Cognitive Science

Hidden Markov Models

Ingmar Visser¹

¹University of Amsterdam

MIT Press

Published on: Nov 20, 2025

DOI: <https://doi.org/10.21428/e2759450.528dddc3>

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](#).

Hidden Markov models are a class of statistical model used to characterize time series and longitudinal data. When applied to longitudinal data, the model is also known as the latent Markov model. Hidden Markov models are very versatile models and have a wide range of applications. The model is appropriate when the underlying process is assumed to be discrete, and measurements are taken at discrete (equally spaced) points in time. In cognitive science, the model is primarily applied to data on learning and cognitive development as well as to psychophysiological data.

History

Baum and Petrie laid the statistical foundations of hidden Markov models (HMMs) in the 1960s. In their article, [Baum and Petrie \(1966\)](#) derived the main equations necessary to efficiently compute the likelihood for these models (informally, the probability of the data under the model as a function of the parameters) as well as outlining methods for parameter estimation. Note that, at the time, the models were not called HMMs but referred to as *probabilistic functions of Markov chains*. The groundwork for latent Markov models was done by Wiggins around the same time, presenting models for the analysis of longitudinal panel data (see [Wiggins, 1973](#)).

Early applications of time series analysis with HMMs were developed in engineering and economics. A key paper in the engineering literature is [Rabiner \(1989\)](#). In this work, Rabiner applied the HMM to the problem of speech recognition; a continuous stream of speech data needs to be parsed into discrete segments—words. In economic time series, HMMs have been (and continue to be) applied to characterize and predict different phases of the economy. For example, so-called “bear” and “bull” markets (with average stock prices going up or down) on the stock exchange are modeled by [Kim \(1994\)](#). In the context of economics, these models are often referred to as regime-switching models.

The latent Markov model (LMM) for longitudinal data was popularized in the social sciences by Langeheine and Van de Pol in the 1980s and 1990s (e.g., [Langeheine & Van de Pol, 1990](#)). The aim was to characterize groups of people and societal change. Applications in learning and developmental psychology are first presented in [Visser et al. \(2002\)](#) and [Schmittmann et al. \(2006\)](#). In these and later applications, the discrete states of the HMM or LMM are seen as cognitive states or strategies that people use to answer items or solve problems in (experimental) psychology tasks.

Core concepts

Two main assumptions characterize HMMs. First, the underlying process of interest is discrete, meaning that there are only a finite number of discrete states that the process can be in. Second, the evolution of the process (or development) is Markovian, meaning that the current state only depends on the previous state—and not on earlier states. In cognitive science, the primary use of HMMs and LMMs is in the areas of learning and cognitive development and in the area of psychophysiological data. HMMs capture

sudden changes in a system and are therefore very suitable for analyzing *phase transitions* in complex systems [see [Complex Dynamical Systems](#)].

HMMs and LMMs in learning and development

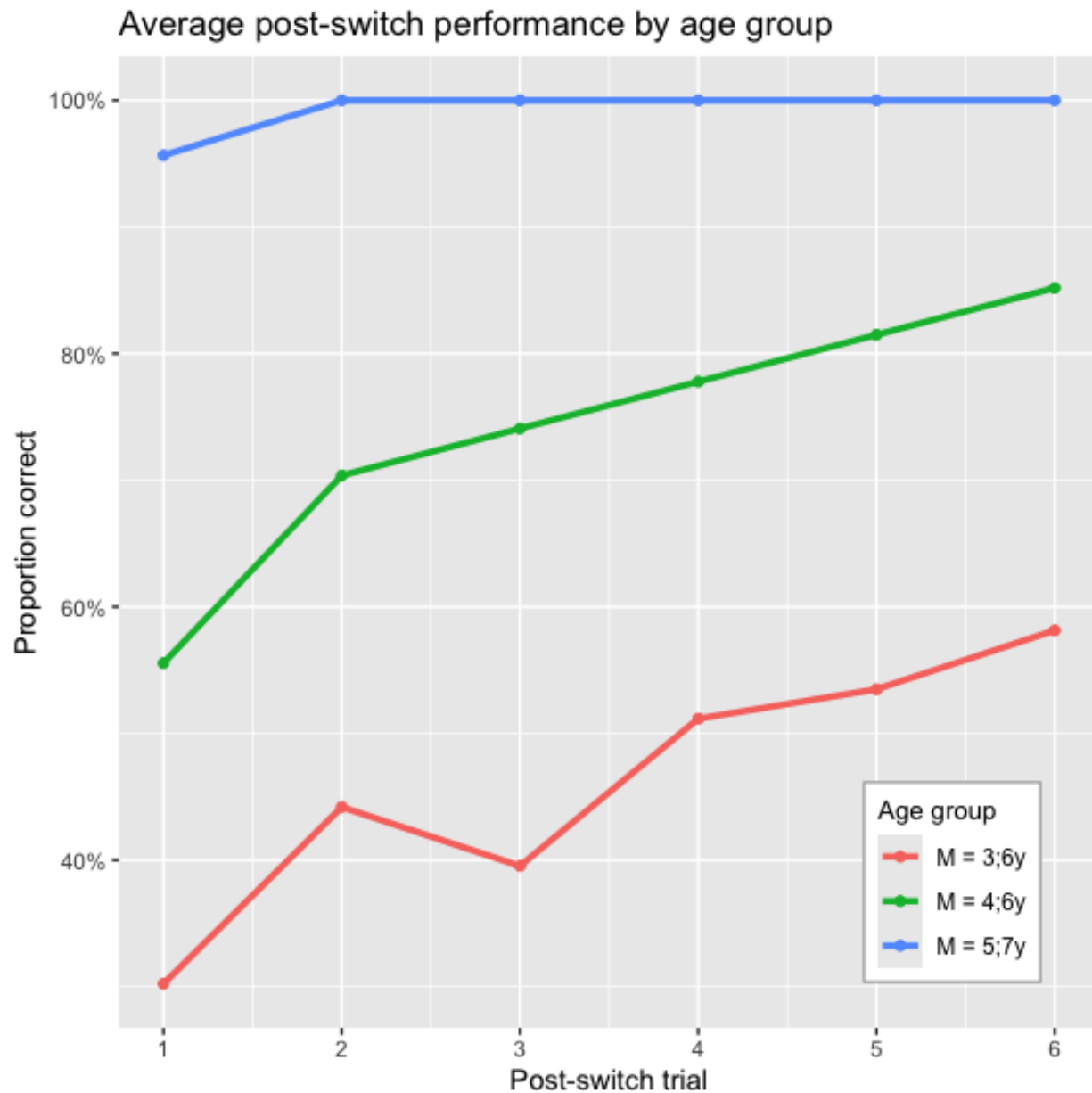
The theoretical perspective of stages in developmental psychology has been very influential ([Inhelder & Piaget, 1958](#)). In this framework, cognitive development proceeds through several discrete stages, which cannot be directly observed [see [Cognitive Development](#)]. Such stages are visible in typical Piagetian tasks such as the conservation of liquid task and the balance scale task. When children engage in such tasks repeatedly, a developmental sequence may be observed in which they switch from one strategy to the next. In the balance scale task, a typical switch concerns including the distance dimension instead of only considering the number of weights. In such situations, the LMM is an appropriate model for the data and can help detect the switches the children go through ([Jansen et al., 2007](#)).

Similarly, participants in experimental paradigms may switch between different strategies. A typical case in category learning is the switch between rule-based and exemplar-based categorization ([Johansen & Palmeri, 2002](#)), which was analyzed using HMMs ([Raijmakers et al., 2014](#)). Another application is the analysis of *regime switches* in response times in experimental tasks ([Van der Maas et al., 2008](#)), in which a discrete switch occurs between different settings of the speed–accuracy trade-off.

In all these examples, the construct of interest is assumed to be a discrete state, in particular, a cognitive state or a developmental stage, which is observed through a set of measurements that have a probabilistic or noisy relationship with the underlying state. In this context, the probabilistic relationship is also referred to as the *measurement model*, as it measures the underlying states.

Example: Learning in the Dimensional Change Card Sorting task

In [Figure 1](#), data is presented from the Dimensional Change Card Sorting task ([van Bers et al., 2011](#)). In this task, six trials are presented in which children have to make a choice where to sort a particular card using a different rule than they used previously, for example, shifting from sorting on color to sorting on shape. The results show the performance over the six trials averaged within three age groups (93 children ranging from 3 to 5 years old were tested). Overall, performance is higher as the children are older. Moreover, as [Figure 1](#) indicates, the probability correct is not constant across trials, certainly in the 3- and 4-year-old children.

**Figure 1**

Post-switch performance in the Dimensional Change Card Sorting task from **van Bers et al. (2011)**.

An HMM is applied to this data by assuming that participants can be in either of two cognitive states: the perseverator state (P) or the switching state (S). That is, some children switch to the new rule correctly, whereas others have trouble and persevere using the old rule, and there is some mixed behavior. Between trials, they can potentially switch between states—this is assumed to happen only occasionally, and the model is used to estimate these probabilities. In this HMM, the measurement model within the states is straightforward; it represents the probability of sorting an item correctly given the state. In the switching

state, this probability is expected to be high (close to 1.0); in the perseverator state, it is expected to be low (close to zero).

A depiction of the fitted model for these data is shown in [Figure 2](#). The model has two states that are labeled P and S for perseverator and switcher, respectively. Within the nodes, the probabilities correct conditional on those states are provided, and they are, as expected, close to zero and one. The arrows between the states indicate the probabilities of transitioning between the states (from trial to trial). The probability of transitioning from P to S is 0.15; this means that there is some probability for participants to switch from being in the P state to being in the S state. The model parameters can be used to assess which participants switch at which trial (through the posterior probabilities), revealing that, in this case, some participants indeed switch (and that this mostly happens after trial one or two). The developmental interpretation is that children experience a sudden insight after one or two trials and then perform correctly from then on. Reports from children and experimenters corroborate such insights. The probability of switching from the S state to the P state is tiny and negligible in practice; the developmental interpretation is that regression does not occur in this task.

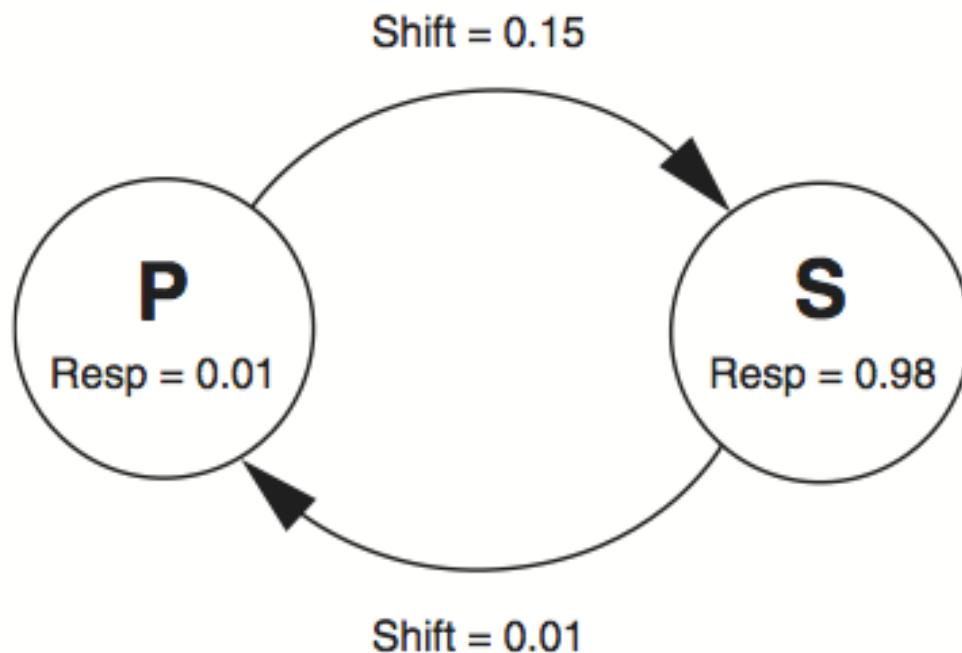


Figure 2
Two-state HMM for Dimensional Change Card Sorting task data.

This HMM representation of the data highlights that cognitive development can progress through sudden insight and as a result of a sudden increase in performance rather than a gradual increase resulting in slight improvements over time.

Other applications

More recently, HMMs have been widely applied to psychophysiological data. An early example is [Flexer et al. \(2002\)](#), who aimed to model sleep stages by analysis of electroencephalography data with HMMs. The underlying constructs of interest are the sleep stages, which cannot be directly observed. The electroencephalography signal, however, has typical patterns associated with the sleep stages that can be measured and segmented using HMMs.

Another area of application is that of eye-tracking data. Eye-movement data is typically characterized by eye-movement *events* such as fixations and saccades. To extract such events from raw eye-tracking data, HMMs can be appropriate tools by using as input the sample-to-sample velocity and acceleration as measurements. The states of the HMM correspond to the different events (see, e.g., [Houpt et al., 2018](#) and [Lüken et al., 2022](#)).

Questions, controversies, and new developments

Conceptually, the HMM and LMM are identical; both are characterized by discrete states and probabilistic measurement of those states. In one statistical aspect they are different: although the domain of generalization for the LMM is the population of the individuals (that was measured repeatedly), the domain of generalization in the HMM is time. That is, in a time series model, the aim is to predict future time points of the evolution or development of the system.

HMMs are versatile models as shown by their wide application. An important limitation is the Markov assumption itself, that is, the assumption that the current state of a process only depends on the previous state and not on multiple earlier states. Many extensions of HMMs are concerned with alleviating this, either by adding multiple layers of states such as in the hierarchical HMM ([Fine et al., 1998](#)) or by adding additional dependency such as in the hidden semi-Markov model ([Bulla & Bulla, 2006](#)).

In recent years, Bayesian estimation has become much more common and accessible also in the application of HMMs ([Cappe et al., 2005](#)). As a result, more extensive measurement models are possible than is the case with the classical estimation approaches (e.g., [Kucharsky et al., 2021](#)).

Broader connections

The concept of Markov models is ubiquitous in mathematics and plays a role in many areas of statistics. Closer to cognitive science and most relevant is the connection between Markov models and linguistic analysis.

[Chomsky \(1959\)](#) and [Miller and Chomsky \(1963\)](#) developed a framework for the characterization of formal languages. The simplest type of language (in the Chomsky hierarchy) is the *regular* language or *finite state* language. The connection with Markov models becomes clear in comparison with finite state languages; there is a finite number of discrete underlying states that is necessary to (completely) describe the (regular) language. When such a language is stochastic, the class of finite state languages corresponds

very closely to the class of HMMs such that each regular language is represented by an HMM. These models have also been implemented in connectionist models of language processing [see [Recurrent Neural Networks](#)].

Besides playing a role in theoretical discussions in the foundations of cognitive science, finite state languages have also been used in experimental psychology in artificial grammar learning experiments. Participants are exposed to sequences formed by a regular language and then tested on their discrimination and processing of sequences from that language ([Reber, 1967](#); [Cleeremans & McClelland, 1991](#)).

Further reading

- Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, 44(2), 457–467. <https://doi.org/10.1037/0012-1649.44.2.457>
- Visser, I., & Speekenbrink, M. (2022). *Mixture and hidden Markov models with R*. Springer.

References

- Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, 37(6), 1554–1540. <https://doi.org/10.1214/aoms/1177699147>
- Bulla, J., & Bulla, I. (2006). Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics & Data Analysis*, 51(4): 2192–2209. <https://doi.org/10.1016/j.csda.2006.07.021>
- Cappe, O., Moulines, E., & Ryden, T. (2005). *Inference in hidden Markov models*. Springer.
- Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, 2(2), 137–167. [https://doi.org/10.1016/S0019-9958\(59\)90362-6](https://doi.org/10.1016/S0019-9958(59)90362-6)
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120(3), 235–253. <https://doi.org/10.1037//0096-3445.120.3.235>
- Fine, S., Singer, Y., & Tishby, N. (1998). The hierarchical hidden Markov model: Analysis and applications. *Machine Learning*, 32, 41–62. <https://doi.org/10.1023/A:1007469218079>

- Flexer, A., Sykacek, P., Rezek, I., & Dorffner, G. (2002). An automatic, continuous and probabilistic sleep stager based on a hidden Markov model. *Applied Artificial Intelligence*, 16(3), 199–207. <https://doi.org/10.1080/088395102753559271>

↩

- Houpt, J. W., Frame, M. E., & Blaha, L. M. (2018). Unsupervised parsing of gaze data with a beta-process vector auto-regressive hidden Markov model. *Behavior Research Methods*, 50(5), 2074–2096. <https://doi.org/10.3758/s13428-017-0974-7>

↩

- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. Basic Books.

↩

- Jansen, B. R. J., Raijmakers, M. E. J., & Visser, I. (2007). Rule transition on the balance scale task: A case study in belief change. *Synthese*, 155, 211–236. <https://doi.org/10.1007/s11229-006-9142-9>

↩

- Johansen, M. K., & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, 45(4), 482–553. [https://doi.org/10.1016/s0010-0285\(02\)00505-4](https://doi.org/10.1016/s0010-0285(02)00505-4)

↩

- Kim, C. J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1–2): 1–22. [https://doi.org/10.1016/0304-4076\(94\)90036-1](https://doi.org/10.1016/0304-4076(94)90036-1)

↩

- Kucharskỳ, Š., Tran, N. H., Veldkamp, H., Raijmakers, M. E. J., & Visser, I. (2021). Hidden Markov models of evidence accumulation in speeded decision tasks. *Computational Brain & Behavior*, 4, 416–441. <https://doi.org/10.1007/s42113-021-00115-0>

↩

- Langeheine, R., & Van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods and Research*, 18(4), 416–441. <https://doi.org/10.1177/0049124190018004002>

↩

- Lüken, M., Kucharskỳ, Š., & Visser, I. (2022). Characterising eye movement events with an unsupervised hidden Markov model. *Journal of Eye Movement Research*, 15(1), 1–29. <https://doi.org/10.16910/jemr.15.1.4>

↩

- Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 419–491). Wiley.

↩

- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286. <https://doi.org/10.1109/5.18626>

↩

- Raijmakers, M. E. J., Schmittmann, V. D., & Visser, I. (2014). Costs and benefits of automatization in category learning of ill-defined rules. *Cognitive Psychology*, 69, 1–24. <https://doi.org/10.1016/j.cogpsych.2013.12.002>

↩

- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855–863. [https://doi.org/10.1016/S0022-5371\(67\)80149-X](https://doi.org/10.1016/S0022-5371(67)80149-X)

↩

- Schmittmann, V. D., Visser, I., & Raijmakers, M. E. J. (2006). Multiple learning modes in the development of performance on a rule-based category-learning task. *Neuropsychologia*, 44(11), 2079–2091. <https://doi.org/10.1016/j.neuropsychologia.2005.12.011>

↩

- van Bers, B. M. C. W., Visser, I., van Schijndel, T. J. P., Mandell, D. J., & Raijmakers, M. E. J. (2011). The dynamics of development on the Dimensional Change Card Sorting task. *Developmental Science*, 14(5), 960–971. <https://doi.org/10.1111/j.1467-7687.2011.01045.x>

↩

- Van der Maas, H. L. J., Dutilh, G., Visser, I., Grasman, R. P. P. P., & Wagenmakers, E. J. (2008). Phase transitions in the trade-off between speed and accuracy in choice reaction time tasks. [Manuscript in preparation].

↩

- Visser, I., Raijmakers, M. E. J., & Molenaar, P. (2002). Fitting hidden Markov models to psychological data. *Scientific Programming*, 10(3), 185–199. <https://doi.org/10.1155/2002/874560>

↩

- Wiggins, L. M. (1973). *Panel analysis: Latent probability models for attitude and behavior processes*. Jossey-Bass.

↩