


dsmmR: Estimation and Simulation of Drifting Semi-Markov Models


Vlad Stefan Barbu ¹, Ioannis Mavrogiannis ², and Nicolas Vergne¹

¹ Laboratory of Mathematics Raphaël Salem (LMRS), UMR CNRS 6085, University of Rouen Normandy, France ² Laboratoire de Mécanique, Modélisation & Procédés Propres (M2P2), UMR 7340, Aix Marseille Université, CNRS, Centrale Méditerranée, France  Corresponding author

DOI: [10.21105/joss.05572](https://doi.org/10.21105/joss.05572)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: Hugo Gruson 

Reviewers:

- [@Athene-ai](#)
- [@RomainAzais](#)
- [@strevezas](#)

Submitted: 22 January 2023

Published: 28 July 2024

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

Markov models are a common approach in the researcher's toolbox for aiding with the modeling of many real-life phenomena, often presented as a sequence of states in discrete time, known as a Markov chain. However, this assumes that the sequence under inspection is time-homogeneous and that the sojourn times always follow a geometric distribution. Notice that this is not always true in practice, for example when modeling DNA sequences.

Drifting semi-Markov models (DSMM) aim to be a discrete-time pairing of semi-Markov models with drifting Markov models. Semi-Markov models allow an arbitrary choice for the distribution of the sojourn times. The drifting Markov models describe the non-homogeneity of a sequence through a smooth, known shape that is gradually evolving, expressed through a polynomial function. As a result, DSMM are best suited to capture non-homogeneities which occur from the intrinsic evolution of the system or from the interactions between the system and the environment. For a detailed introduction to semi-Markov models see Barbu & Limnios (2009). Drifting Markov models were first introduced in Vergne (2008).

dsmmR is an R package which allows the user to perform parametric and non-parametric estimation and simulation of drifting semi-Markov processes. The user can also define their own parametric and non-parametric DSMM specifications, allowing for a necessary degree of freedom when dealing with a research question. Furthermore, three different types of DSMM are considered. These three models differ in the way they characterize the drifting semi-Markov kernel. Specifically, the first model allows both the transition matrix and the sojourn time distribution to vary (i.e. to “drift”), the second model allows only the transition matrix to drift, while the third model allows only the sojourn time distribution to drift.

The main functions of dsmmR are the following:

- `fit_dsmm()` : estimates a DSMM given a sequence.
- `parametric_dsmm()` : defines a parametric DSMM.
- `nonparametric_dsmm()` : defines a non-parametric DSMM.
- `simulate()` : simulates a sequence given a dsmm object.
- `get_kernel()` : computes the drifting semi-Markov kernel given a dsmm object.

The estimation of the DSMM is parametric or non-parametric and can be defined through the function `fit_dsmm()`, which returns an object of the S3 class (`fit_dsmm_parametric`, `dsmm`) or (`fit_dsmm_nonparametric`, `dsmm`). In the parametric estimation case, several sojourn time distributions are considered for the sojourn times: Uniform, Geometric, Poisson, Discrete Weibull (of type 1) and Negative Binomial. The parametric DSMM specification can be defined through the function `parametric_dsmm()`, which returns an object of the S3 class (`dsmm_parametric`, `dsmm`). The non-parametric DSMM specification can be defined through the function `nonparametric_dsmm()`, which returns an object of the S3 class

(`dsmm_nonparametric`, `dsmm`). It allows for the sojourn time distributions to be of an arbitrary shape.

The `dsmm` class acts like a wrapper class, enabling the handling of all above objects, which encompass the three models and parametric or non-parametric cases for any degree, for:

- Simulating a sequence of states under a drifting semi-Markov kernel through the S3 method `simulate.dsmm()`.
- Calculating the drifting semi-Markov kernel through the generic function `get_kernel()`.

Statement of need

The present R package consists of a novel approach to model the gradually evolving non-homogeneity of a discrete-time Markov chain through a polynomial function, while it simultaneously allows for the flexibility of choosing arbitrary sojourn time distributions. Drifting semi-Markov processes represent a versatile modeling tool, which can be readily applied in many fields of science like bio-informatics, engineering, finance and also potentially in reliability and survival analysis settings. In the literature, existing R packages deal with semi-Markov models, hidden semi-Markov models and drifting Markov models.

Such R packages specializing in semi-Markov models include the `SemiMarkov` (Król & Saint-Pierre, 2015) package, which performs maximum likelihood estimation for parametric continuous-time semi-Markov processes, while they allow a choice of distribution for the sojourn times between the Exponential, the Weibull or the exponentiated Weibull. Also associated hazard rates are computed, and covariates can also be taken into account through the Cox proportional hazard model. The R package `smmR` (Barbu, Lecocq, et al., 2021) deals with discrete-time multi-state semi-Markov models, performing parametric and non-parametric estimation and simulation, with k -th order Markov chains also considered. For the parametric estimation the following sojourn times were considered: Uniform, Geometric, Poisson, Discrete Weibull and Negative Binomial. One or more sample paths were considered, with or without censoring at the beginning or the end of the sample paths. There are also a few R packages also dedicated to hidden semi-Markov models, where methods for estimation and prediction are implemented, such as the `HMM` R package (Himmelmann, 2022), the `HiddenMarkov` R package (Harte, 2021) and the `mhsmm` R package (O'Connell & Højsgaard, 2011).

Furthermore, for drifting Markov models the R package `drimmR` (Barbu, Lothode, et al., 2021) has been developed, performing non-parametric estimation and simulation, also allowing for the exact computation of the associated reliability measures such as reliability, availability, maintainability and failure rates. Also several statistical frameworks were considered, allowing for estimation from multiple sequences, complete or incomplete, with the same or unequal model size. In this package, a polynomial function was also used to capture the non-homogeneity of the sequences considered.

Acknowledgements

We acknowledge the DATALAB Project (financed by the European Union with the European Regional Development fund (ERDF) and by the Normandy Region) and the HSMM-INCA Project (financed by the French Agence Nationale de la Recherche (ANR) under grant ANR-21-CE40-0005).

References

Barbu, V. S., Lecocq, F., Lothode, C., & Vergne, N. (2021). *smmR: Simulation, estimation and reliability of semi-Markov models*. <https://doi.org/10.32614/CRAN.package.smmR>

- Barbu, V. S., & Limnios, N. (2009). *Semi-Markov chains and hidden semi-Markov models toward applications: Their use in reliability and DNA analysis* (Vol. 191). Springer Science & Business Media. <https://doi.org/10.1007/978-0-387-73173-5>
- Barbu, V. S., Lothode, C., & Vergne, N. (2021). *drimmR: Estimation, simulation and reliability of drifting Markov models*. <https://doi.org/10.32614/CRAN.package.drimmR>
- Harte, D. (2021). *HiddenMarkov: Hidden markov models*. Statistics Research Associates. <https://doi.org/10.32614/CRAN.package.HiddenMarkov>
- Himmelman, L. (2022). *HMM: Hidden markov models*. <https://doi.org/10.32614/CRAN.package.HMM>
- Król, A., & Saint-Pierre, P. (2015). SemiMarkov: An R package for parametric estimation in multi-state semi-markov models. *Journal of Statistical Software*, 66(6), 1–16. <https://doi.org/10.18637/jss.v066.i06>
- O'Connell, J., & Højsgaard, S. (2011). Hidden semi markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software*, 39(4), 1–22. <https://doi.org/10.18637/jss.v039.i04>
- Vergne, N. (2008). Drifting Markov models with polynomial drift and applications to DNA sequences. *Statistical Applications in Genetics and Molecular Biology*, 7(1). <https://doi.org/10.2202/1544-6115.1326>