

Received: 25 June, 2024

Accepted: 18 July, 2024

Published: 19 July, 2024

***Corresponding author:** Orchidea Maria Lecian
Sapienza, University of Rome, Rome, Italy,
E-mail: orchideamaria.lecian@uniroma1.it

Keywords: Chains; Markov chains; Enveloping algebras; Genomic events; Allele-specific copy-number abnormalities

Copyright License: © 2024 Sapienza OML. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://www.clinsurgroup.us>



Short Communication

Markov Models of Genomic Events

Orchidea Maria Lecian Sapienza*

University of Rome, Rome, Italy

Abstract

The Markov Models of genomic elements are newly considered. The representation of the fundamental matrix of the Markov model is newly theorised. The order of magnitude of the initial conditions for the elements of the transition probabilities is newly hypothesised.

The model is compared with a sub-Hidden Markov Model of genomic events. The chosen representation of the states is newly proven to consist of an enveloping algebra. The new condition is posed on the Markovian feature of the originating chain from the study of the elements of the loci of the state space; in this case, the choice of the representation of the probability matrix is analytically spelled out, and Monte Carlo methods are not necessitated.

Introduction

The present report is aimed at further improving the mathematical definitions in the Markov models of genomic elements, such as that recently presented in [1].

The present paper is aimed at improving from [1] the long-standing interrogations raised in [2-5] about the analytical modellings of algorithms of oncogenesis.

Eq. (1) form [1] is here imposed a new hypothesis, for which the comparison holds also with the (alternative) numerical (Monte Carlo) methods developed in [6] and more recently improved in [7]. More in detail, the new analysis is pointed out, which ensures the new choice of the representation of the probability matrix, for which the confrontation with the numerical methods (if/where necessitated) is compliant. The comparison with numerical methods can be of interest i.e. in the case envisaged in [8] for the numerical test of inference parameters.

Furthermore, the method is compared with the analysis of

the sub-Hidden Markov Model (subHMM), which is used in [9] to understand the study of the copy number abnormalities in the allele-specific analyses; in this case, the states of the Markov models are newly proven to consist of an enveloping algebra. Furthermore, the relevance of the hypothesis of a constant number of Markov states in the definition of the fundamental matrix of the originating chain is newly demonstrated to define the Markovian feature. Accordingly, the enveloping algebra defines the committors, which characterise the Markov State Model, from which the subHMMs can be issued. After these proofs, it is possible to analytically calculate the Mean-First Passage Times, the time evolutions of the eigenvalues, and those of the modellisation errors.

Low-rank-tensor methods

The evolution of cancer phenomena can be modelled as continuous-time Markov chains.

Transition rates are hypothesised as separable functions in [1], i.e. such that convergent 'iteration methods' can be made use of, for which the notion of distribution is retrieved.



Non-stationarity is due to the fact that the age of the tumors might be unknown, for which the marginalisation of the time variable is needed.

The necessity of the low-rank tensor methods is justified from the evidence that given d the number of 'genomic events', there are $n = 2^d$ number of Markov states of the tumor; as from the recent understandings, there are $d \sim 299$ known genes which determine the evolution of the tumors [10]. The functional dependence of the state space on 2^d is named 'state space explosion' after [11]; it is tamed after the introduction of the 'marginal distributions', by which operators that act on the low-rank tensors are defined.

The 'Hierarchical Tucker format is adopted.

Let \hat{Q} be the fundamental matrix of the chosen Markov chain on a discrete state space S with initial distributions assumed as defined.

It is here newly requested that for Eq. (1) from [1] to hold, the hypothesis that the entries of \hat{Q} be infinitesimal must newly be requested.

Let P be the probability matrix associated with the fundamental matrix \hat{Q} and after the new hypothesis; the distributions from \hat{P} are defined from the initial value p , where the latter is written as

$$p = \int_0^{\infty} e^{\tau[\hat{Q}-\hat{I}]} p(0) d\tau \quad (1)$$

The new hypothesis $p(0) \sim o(0)$ is here therefore newly requested for the proper definition. In Eq. (1), τ is a time variable, and $[\hat{Q}-\hat{I}]$ is a regular operator. The spectrum $\sigma([\hat{Q}-\hat{I}])$ of the operator $[\hat{Q}-\hat{I}]$ is written as from the states $x \in S$ from the definition

$$\sigma([\hat{Q}-\hat{I}]) \subseteq \bigcup_{x \in S} \{z \in \mathbb{C}: |z - Q_{xx}| \leq |Q_{xx}|\} \subseteq \{z \in \mathbb{C}: \operatorname{Re} z \leq 0\} \quad (2)$$

It is important to remark that the marginalisation procedures originating from Eq. (2), from which the Markov models descend, therefore differ from the 'dominant-eigenvalue' technique with

$$\sigma([\hat{Q}-\hat{I}]) \subseteq \{z \in \mathbb{C}: \operatorname{Re}(z) \leq -1\}. \quad (3)$$

The method of the 'stochastic automata networks' is further discussed in [12].

Allele-specific copy number methods

Allele-specific copy-number methods allow one to study copy-number abnormalities, as from [9].

For this sake, a sub-Hidden Markov Model (subHMM) is implemented: it allows one to consider both the 'subclone region' and the 'region-specific genotype'. The hidden-state variable W_k of the state k represents the 'conglomeration of the subclone genotype' and the 'clonal proportion'.

More in detail, the state $W_k[z_k, U_k, T_k]$ is defined as giving rise to time-dependent transition probabilities which can be represented as 'multinomial distribution'.

The states W_k are specified after Z_k the 'mainclone genotype' of the locus k , U_k the 'indicator' about whether there is a subclone in k , and T_k the 'subclone genotype' (i.e. if the considered subclone exists).

The transition of the states W_k is considered in [9] only for consecutive 'loci'.

A maximum number of copies is assumed.

Therefore, the elements of the subHMM are here newly proven to compose an enveloping algebra.

Under the hypothesis of the 'constant clonal proportion', the transition probabilities $P_i(z)$ from Eq. (2) in [9] determine that the hidden states are not observed, and 'allele-specific' elements are considered.

Conclusion

The Markov model of genomic events is newly further analysed.

More in detail, the choice of the representation of the transition probabilities is reconducted to be well-posed only under the new hypothesis that the entries of the fundamental matrix be infinitesimal.

The new hypothesis on the initial conditions of the transition elements is requested for the time-marginalisation technique to be consistent. The difference with the 'dominant-eigenvalue approach' is stressed. The case of the sub-Hidden Markov Model in the study of allele-specific copy number analysis is newly approached.

The elements of the Markov models are therefore here newly proven to consist of an enveloping algebra.

Furthermore, it aims to focus on the hypothesis of a constant number of 'constant clonal proportions: in this case, the Markovian feature of the originating chain is newly proven after the study of the entries of the fundamental matrix.

It has to be stressed that the proof of the Markovian property of the originating chain is fundamental in the definition of the Markov State Model(s) from which the subHMM is taken. In the case of the Markovian feature, the possibility to define the committor is necessitated for the study of the Mean-First Passage times and that of the time evolution of the eigenvalues, as from [13] and [14], respectively.

References

- Georg P, Grasedyck L, Klever M, Schill R, Spang R, Wettig T. Low-rank tensor methods for Markov chains with applications to tumor progression models. *J Math Biol.* 2022;86(1):7. Available from: <https://doi.org/10.1007/s00285-022-01846-9>.
- Hjelm M, Hoeglund M, Lagergren J. New probabilistic network models and algorithms for oncogenesis. *J Comput Biol.* 2006:853-865.



Available from: <https://doi.org/10.1089/cmb.2006.13.853>.

3. Beerenwinkel N, Sullivan S. Markov models for accumulating mutations. *Biometrika*. 2009;96:645-661. Available from: <https://doi.org/10.1093/biomet/asp023>.
4. Schill R, Solbrig S, Wettig T, Spang R. Modelling cancer progression using Mutual Hazard Networks. *Bioinformatics*. 2019;36:241-249. Available from: <https://doi.org/10.1093/bioinformatics/btz513>.
5. Gotovos A, Burkholz R, Quackenbush J, Jegelka S. Scaling up continuous-time Markov chains helps resolve underspecification. *arXiv*. 2021. arXiv:2107.02911. Available from: <https://doi.org/10.48550/arXiv.2107.02911>.
6. Ji H, Mascagni M, Li Y. Convergence analysis of Markov chain Monte Carlo linear solvers using Ulam-von Neumann Algorithm. *SIAM J Numer Anal*. 2013;51(4):2107-2122. Available from: <https://doi.org/10.1137/130904867>.
7. Fathi-Vajargah B, Hassanzadeh Z. Improvements on the hybrid Monte Carlo algorithms for matrix computations. *Sa'dhana*. 2019;44(1):1. Available from: <https://doi.org/10.1007/s12046-018-0983-y>.
8. Beerenwinkel N, Schwarz RF, Gerstung M, Markowitz F. Cancer Evolution: Mathematical Models and Computational Inference. *Syst Biol*. 2015;64. Available from: <https://doi.org/10.1093/sysbio/syu081>.
9. Choo-Wosoba H, Albert PS, Zhu B. A hidden Markov modeling approach for identifying tumor subclones in next-generation sequencing studies. *Biostatistics*. 2022;23:69-82. Available from: <https://doi.org/10.1093/biostatistics/kxaa013>.
10. Bailey MH, Tokheim C, Porta-Pardo EL, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;173(2):371-385.e18. Available from: <https://doi.org/10.1016/j.cell.2018.02.060>.
11. Buchholz P, Dayar T. On the convergence of a class of multilevel methods for large sparse Markov chains. *SIAM J Matrix Anal Appl*. 2007;29(3):1025-1049. Available from: <https://doi.org/10.1137/060651161>.
12. Plateau B, Stewart WJ. Stochastic automata networks. In: *International series in operations research and management science*. New York: Springer; 2000. p. 113-151. Available from: https://link.springer.com/chapter/10.1007/978-1-4757-4828-4_5
13. Lecian OM. Analytical results from the two-states Markov-states model and applications to validation of molecular dynamics. *Int J Math Comput Res*. 2023;11(9):3746-3754. Available from: <https://doi.org/10.47191/ijmcr/v11i9.08>
14. Lecian OM. Laplace Kernels with Radon measures in Galerkin Markov-State Models: new theorems about analytical expressions of time evolution of eigenvalues and about errors. e-print. 2024. Available from: <http://dx.doi.org/10.13140/RG.2.2.24311.39841/1>