

Chain and Multi-Spatial Markov Chains of Time Dependence of Allele Manifestation in Cancer: Metric, Algebras and Expectations of the Reward

Orchidea Maria Lecian^{1*}

¹Sapienza University of Rome, Rome, Italy.

Received date: 05 August 2024; Accepted date: 21 August 2024; Published date: 31 August 2024

Corresponding author: Orchidea Maria Lecian, Sapienza University of Rome, Rome, Italy.

Citation: Orchidea Maria Lecian. Chain and Multi-Spatial Markov Chains of the Time Dependence of Allele Manifestation in Cancer: Metric Algebras, and Expectations of the Reward. Journal of Medical and Clinical Case Reports 1(8). <https://doi.org/10.61615/JMCCR/2024/AUG027140831>

Copyright: © 2024 Orchidea Maria Lecian. 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.

Abstract

Time-continuous Chain and Multi-spatial Markov chains of time dependence of allele manifestation in cancer are newly written. The allele manifestation is associated with 'non-cancer states' and 'cancer states' which constitute nodes (and pathways). The allele manifestation is described as directed graphs (on the opportune manifold), which are modeled according to Markov Decision Processes. The measure from the algebra on the manifold of the directed graphs is newly proven to be matched with that of the Kantorovich metric on the Borel (sub-)sets of the Markov states space. The expectation of the reward is calculated in the case of 'non-cancer states' and in that of 'cancer states'; the scheme is ready for a path-integral approach.

Keywords: Cancer; chains; first-order Markov processes; allele manifestation; allele loci; metric distances; Markov models; Kantorovich measure; directed graphs; groupoids; free semi-groupoids.

Introduction

In [1], the transition probabilities of objects (i.e. ones from biological samples i.e. tissues, cells, sequences) from the 'non-cancer state' to the 'cancer states' are described for neighboring objects as far as both distance and state of the neighbors are concerned.

The manifestation of the alleles is described in [2].

In [1], the presence of uncertainties is mentioned, from which the standard deviations can be newly calculated; nevertheless, no numerical values are reported, i.e. such that it is not possible, within the presented data, to assess more specific Markov models.

The uncertainties on the identification of objects depending on the loci of the alleles within the Markov models are further discussed in [2].

The existence of experimental uncertainties due to the error propagation is recalled to be consistent with the Kantorovich measure from the Kantorovich metric of Markov Decision Processes.

The possibility to define metrics and algebras is studied in [3]. The calculation of the reward in [4] is here newly analytically proposed for 'non-cancer states' and 'cancer states'.

The standard deviation of the experimental apparati and techniques are recapitulated in [5] and in [6]; as a result, the 'similarity' of states obtained after the use of the Kantorovich metric is here understood to be consistent for a broader class of phenomena; such phenomena were previously sampled as Monte Carlo [5]. Moreover, the distances for the first-order Markov process extracted from the stayer-mover scheme are framed within the analysis of [5]. Starting from the description in [17], the allele manifestation is in the present work newly written as a landscape of oriented graphs with measure theory from a σ -algebra. More in detail, the directed graphs [7] are recalled to be

endowed with a measure from a σ -algebra (within the framework of semigroupoids) [8]; as a new consequence, the Kantorovich measure is newly demonstrated to apply to the directed graphs which admit measure theory from a σ -algebra. The role of the propagated experimental errors is therefore newly outlined.

Furthermore, from [3], the QoS (Quality of Service) metrics are proposed in order for the metrics to consist of the sum of probabilities. This method might apply to the comments in [5]. Within the present analysis, unaccordingly, the chosen measure is the Kantorovich measure. The expectation value of the reward will be expressed also from the definition of the first hitting time.

The directed graphs allow one to use the path-integral approach in Markov models. The paper is organized as follows.

In **Section 3**, the allele manifestation is defined as nodes within 'non-cancer states' and 'cancer states'. The transition from the different states is recalled to define pathways. The pathways are framed within the Markov Decision Process. The pathways are identified with directed graphs. The Kantorovich measure is reminded. The directed graphs are reviewed on the opportune manifold endowed with a metric from a σ -algebra from the groupoid technique. In **Section 4**, a new choice of the probability matrix of the Markov chain of cancer progression is newly taken and explained.

In **Section 5**, a new time-continuous chain inspired by the mover-stayer model is analytically written for the allele manifestation methods.

In **Section 6**, the allele manifestation phenomena are proposed as Markov Decision

Processes. The allele manifestation is identified with paths that correspond to directed graphs; the Kantorovich metric is newly proven to the manifold of

the directed graphs as, in the latter case, a measure issued from a σ -algebra on Borel (sub-)set holds.

In **Section 7**, a new mathematical application for the nodes as loci of the allele

manifestation and measure theory on directed graphs is prepared. The expectation of the rewards is calculated in the 'non-cancer Markov states' and in the 'cancer Markov states'. The results are discussed in **Section 8**.

Introductory Material

The Markov Models

In [1], the definition of 'node points' which have to be referred to as 'non-cancer states' and 'cancer states' is proposed.

The existence of uncertain sources is pointed out.

N is called the nodal points, with $N = 1, N$

This way, the set N consists of $N_i = \{j \in N \mid j \text{ is the nearest neighbor of the node } i\}$. Each one of the nodes can be either in the 'non-cancer state' or in the 'cancer state'; the items of information are encoded within the 'switch' parameters S_i , i.e. such that $S_i = 0$ corresponds to the description according to which the node i is in the non-cancer state, and $S_i = 1$ corresponds to the fact that the node i is in the cancer state.

There exists a 'memory-less' exponential probability distribution with probability rates denominated Λ_i , which evolve in the time variable t as $\Lambda_i(t)$. The probability distribution $f(s, \Lambda_i)$ depends on the time variables t and s such as

$$f(s, \Lambda_i) = \Lambda_i e^{-\Lambda_i(s)}(s-t), \quad s \geq t. \quad (1)$$

The probability p to pass from the mode $\{S_i = 0\}$ to the mode $\{S_i = 1\}$ is therefore obtained as

$$p(S_i(t + \tau) = 1 \mid S_i(t) = 0) \simeq 1 - e^{-\Lambda_i(t)\tau} \simeq \Lambda_i\tau + O(\Lambda_i\tau)^2 \quad (2)$$

The probability matrix \hat{P}_τ over the time interval τ is

$$\hat{P}_\tau = \begin{bmatrix} 1 - \Lambda(t)\tau & 0 \\ \Lambda t\tau & 1 \end{bmatrix} \quad (3)$$

The probabilities of finding cancer states close to non-cancer states can be calculated by composition after **Eq. (3)**.

From [9], the irreversible cancer process is studied from the 'mover-stayer model' described in [10].

The time-dependent probability matrix $\hat{P}(T)$ is decomposed as

$$\hat{P}(T) \equiv \hat{s} + (\hat{I} - \hat{s})M(T) \quad (4)$$

In this study, a probability matrix is issued from the complete 'mover-stayer model'; indeed, the complete 'mover-stayer' model does not correspond to a Markov process: nevertheless, a first-order Markov process is obtained in [9]. For these purposes, the Sojourn Time is defined as the lapse of time during which the tumor is staying in the 'pre-clinical' state, but the phase is 'screen detectable'.

Accordingly, the Mean Sojourn Time is the time during which there is the possibility of diagnosis.

From [9] page. 267 ibidem, the 'malignancy grades' are defined and here reported as

1. No detectable disease.
2. Pre-clinical screen-detectable grade 1-2.
3. Pre-clinical screen-detectable grade 3.
4. Clinical grade 1-2.
5. Clinical grade 3.

A Markov chain $X(t)$ is described as a continuous-time T Markov chain, whose states are denominated S_{ij} .

The index $i = 0, 1, 2$ is explicated as

$i = 0$ no detectable disease;

$i = 1$ pre-clinical screen-detectable disease;

$i = 2$ clinical disease.

The index j is specified as

$j = 1$ 1-2 grade;

$j = 2$ 3 grades.

The Markov chain $X(T)$ is defined on the state space $\Omega(T)$ with

$$\Omega(T) = \{S_{00}, S_{11}, S_{12}, S_{21}, S_{22}\}. \quad (5)$$

The matrices in **Eq. (4)** are here spelled out.

The matrix \hat{s} is a diagonal matrix, whose entries s_{ii} , $i = 1, 2, \dots, 5$ are proportional to the 'stayer' of each of the states as

$$\hat{S} = \begin{pmatrix} S_{11} & 0 & 0 & 0 & 0 \\ 0 & S_{22} & 0 & 0 & 0 \\ 0 & 0 & S_{33} & 0 & 0 \\ 0 & 0 & 0 & S_{44} & 0 \\ 0 & 0 & 0 & 0 & S_{55} \end{pmatrix} \quad (6)$$

The matrix \hat{M}_1 is defined as

$$\hat{M}_1 = \begin{bmatrix} -(l_1 + l_2) & l_1 & l_2 & 0 & 0 \\ 0 & -l_4 & 0 & l_4 & 0 \\ 0 & 0 & -l_5 & 0 & l_5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (7)$$

The matrix \hat{M}_2 is defined as

$$\hat{M}_2 = \begin{bmatrix} -(l_1 + l_2) & l_1 & l_2 & 0 & 0 \\ 0 & -(l_3 + l_4) & l_3 & l_4 & 0 \\ 0 & 0 & -l_5 & 0 & l_5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (8)$$

Each columns of the two matrices \hat{M}_1 and \hat{M}_2 represent the possibilities $\{S_{00}, S_{11}, S_{12}, S_{21}, S_{22}\}$, respectively.

From these objects, the first-order Markov process X^\wedge is issued.

3.2 About the Bounding of the Probabilistic Metrics

The determination of the convergence of a probability metric is studied in [11]. Let Ω be a measurable space, with a σ -algebra on the Borel subset B; furthermore, let Σ be the space of all the probability measures on (Ω, B) . Several notions of distances can be considered in Σ .

Metrics in probability spaces are schematized in **Table 1** ibidem.

In **Figure 1** ibidem, the relations between the metrics and the triples are studied accordingly; the diameter of the probability space $diam(\Omega)$ is taken into account.

As an example, the discrepancy metric can be considered. Given μ and ν two probability measures, the distance d_D is defined as

$$d_D(\mu, \nu) = \sup_{\text{all closed spherical neighborhoods}} |\mu(B) - \nu(B)| \quad (9)$$

The Diaconis measure [12] Eq. (9) is of use in the analysis of the experimental errors to be attributed to the committer found in [13] here to be applied within the framework of bisimulation metrics (which are illustrated in [4]).

3.3 Measures for Directed Graphs

In [8], measure theory for directed graphs is studied.

The study makes use of the method of semigroupoids. The analysis is shown to apply to σ -algebras.

The aim of [8] is the definition of a locally bounded positive measure for directed graphs; as an application, integration with this measure is performed.

Given a graph G, the free semi-groupoid is considered.

From Lemma 1.1 ibidem, the basic diagrams of finite paths on G are defined for the needed purpose. The cases of non-basic diagrams on G are reconducted to finite paths.

The use of the free-semi-groupoids allows one to consider the vertices on G. It is our care to anticipate that the vertices on G will be identified with the nodes from [1]

4. New Developments of the Spatial Markov Chain Model of Cancer Progression

From [1], the probability matrix of the Markov chain $\hat{P}_\tau(t)$ is here newly rewritten as a function of the time t with τ as a parameter from the new position

$$\hat{P}_\tau(t) = \begin{bmatrix} e^{-\mu t} & 0 \\ R(t) & 1 \end{bmatrix} \quad (10)$$

It is newly found that there corresponds to a Markov chain of fundamental matrix $\hat{Q}_\tau(t)$ which is here newly written as

$$\hat{Q}_\tau(t) = \begin{bmatrix} \Lambda(t) & o(0) \\ o(R(t)) & 0 \end{bmatrix} \quad (11)$$

Therefore, it is newly found that

$$\Lambda(t) \simeq o(0), \quad (12a)$$

$$o(R(t)) \simeq 0 \quad (12b)$$

5. Chain of the Mover-Stayer Model

The complete mover-stayer model is not a Markov process: the corresponding chain is not assured to be a Markovian one. From the items of information used in [9], the fundamental matrix of the chain $\hat{Q}_{m-s}(T)$ is newly spelled as

$$\frac{d\hat{Q}_{m-s}(T)}{dT} T + \hat{Q}_{m-s} \simeq \frac{d(\hat{P}(T) - \hat{I})}{dT} \equiv \frac{d\hat{P}(T)}{dT} \quad (13)$$

The solution of **Eq. (13)** depends on the matrix $\hat{M}(T)$.

The one possible choice of the representation crucially depends on the numerical values of the entries of the matrices in **Eq. (4)**.

The Markov chain of the first-order Markov process can therefore be extracted accordingly.

6.0 Allele Manifestation as a Finite Markov Decision Process

The allele manifestation(s) of cancer states in objects is here newly described within the framework of a Markov Decision Process (MDP).

From [4], the similarity of the states involved in a Markov Decision Process (MDP) allows one to introduce (experimental) uncertainties within the framework of multistate decision-making' in the probability spaces. Therefore, the manifestation of the allele(s) and the choice of the loci are here newly schematized as a MDP.

In [4], distance functions (and metrics) of the states of the MDP (the allele/loci Markov model) are defined.

Accordingly, the metrics can be made use of in order to 'aggregate' some (chosen) sets of states (alleles) which are here newly proposed to be used in several Markov models, i.e. such as (but not necessarily only) Hidden Markov State Models and subHidden Markov State Models.

The time dependence of the allele manifestation of different loci in cancer can therefore be studied within the framework of the definition of the 'similarity of states' in a finite MDP; the 'similarity' is understood within the numerical values of the standard deviations which characterize the choices of the representation of the matrices in **Eq. (13)**; the experimental uncertainties are discussed throughout the present paper.

6.1 Allele loci: States of the MDPs and Related Quantities

An MDP is a finite set of states W and a finite set of actions A on the states., such that, for every pair of states w and w' and the corresponding action an after which one can define

- The probability matrix \hat{P}_{ww^a} ; and
- A numerical reward r_w^a

The details are to be reviewed from [14].

The calculation of the rewards allows one from [4] to find families of

The Expectation Values of the Reward

The 'values of w from [4] are therefore the expectation of the reward. The algebra here newly defined allows one to specify the one expectation value of the reward for the time spent in each chosen subset (in this case, the chosen subset in the state w).

equations that provide one with the definition of the 'values of w .

7.0 Novel Mathematical Description of the Nodes as Loci of the Allele Manifestation and Measure Theory on Directed Graphs

The analysis of the pathways given in [17] according to which it is possible to trace the cancer states and the non-cancer states uniquely allows one to interpret such pathways in the state landscape as directed graphs. A comprehensive study of directed graphs is presented in [7]. It is here newly pointed out, from [8], that the measure of directed graphs is from the σ -algebra on the subset. It is therefore here newly pointed out that the Kantorovich measure applies to directed graphs as follows. From **Eq. (9)**, the probabilities of finding cancer states close to non-cancer states are here newly analyzed as forming directed graphs (on the opportune manifold). The findings of [6] can therefore be applied.

Furthermore, the notion of 'bisimilarity' can be introduced after [4]. The Kantorovich metric [15] can be applied to the probability functions. The aim of the use of the Kantorovich metric is the minimization of the reward. There exists, nevertheless, a notion of distances that do not define a metric, such as the Kullbach-Leibler distance [16].

Fixed-point metrics are used for chains that are not Markovian; the Kantorovich metric can be applied in this case.

Function bounds can be evaluated.

Error bounds of MDP follow.

The nodes are here understood as the loci of the allele manifestations.

The use is here newly proposed of the Kantorovich metric, after which the (number of) copies of the nodes allows one to select both the allele and the locus of its manifestation. The prospect is indicated for the first-order Markov process of the mover-stayer scheme. The notion of bisimilarity is indicated in the study of the committee found in [13].

More in particular, the expectation value of the reward Γ is here specified after the Kantorovich measure. Given g an application

$$G: w \rightarrow [0, \infty); \quad (14)$$

The reward Γ is calculated as

$$\Gamma = \int_0^{\tau} g_{w(t)} dt. \quad (15)$$

In the case the state w is beginning to be populated at the time τ_0 , the time τ_0 is defined as the 'first hitting time, and the reward Γ_{τ_0} is newly specified as

$$\Gamma = \int_{\tau_0}^{\tau} g_{w(t)} dt. \quad (16)$$

in this case, g is at least non-decreasing.

As an easy example, the choice $g \equiv 1$ can be taken for cancer states, while $g = 0$ can be assigned for non-cancer states.

Outlook

In the present paper, the allele manifestation in cancer is newly studied as a time-continuous chain whose states are in an opportune Borel (sub-)set; the new chain is explicitly analytically written. The allele manifestation is newly expressed as a Markov Decision Process. The corresponding first-order Markov process is analyzed.

The allele manifestation in time is analytically newly studied as directed graphs on the opportune manifold: after the use of the groupoid technique, the measure on the opportune manifold is one issued from a σ -algebra; the measure is identified with the Kantorovich measure for Markov Decision Processes. The algebra leads to the definition of the committer for allele manifestation. After the definition of the allele loci, the analysis from [17] can be addressed to find the probabilities of the copy number aberrations.

The reward is calculated in the examples of 'non-cancer states' and of 'cancer states'.

The first-order Markov process is proven to descend from a time-continuous chain within the framework of a particular stayer-mover model, from which the choice of the representation of the probability matrix is taken.

No use of estimators is made.

The model is apt to build Markov states models, hidden Markov Models, sub-hidden Markov models, and further schemes.

After the definition of the committer, the sojourn times, and the hitting times can be calculated. Therefore, the time evolution of the eigenvalues can be analytically calculated as well. The time evolution of the errors in discretization schemes is thus possible as well.

Conclusion

The allele manifestation in cancer can be studied as a graph (of which the nodes correspond to the manifestation) on the opportune manifold. The paths on the graph correspond to a Markov Decision Model. The chain of a particular mover-stayer model is newly written: the chain is newly found to admit a Kantorovich metric.

As a result, the study of the transition probabilities allows one to define the allelemanifestations within the experimental error propagation. The technique here is newly developed and is therefore apt to describe the process of cancer formation. Furthermore, it is possible to confront the new theory with the experimental data of cancerous aggregation.

Moreover, if the choice to implement time discretization is taken, the study of clinical screens within the corresponding available data is newly favored, i.e. after the long-standing problem proposed.

References

1. F. Vermolen, I. Poeloenen. (2020). Uncertainty quantification on a spatial Markovchain model for the progression of skin cancer, *J. Math. Biol.* 80(3): 545-573.
2. T.I. Gossman, D. Waxman. (2022). Correcting Bias in Allele Frequency Estimates Due to an Observation Threshold: A Markov Chain Analysis, *Genome Biol. Evol.* 14(4): 047.
3. M. Bernardo, M. Bravetti. (2003). Performance measure sensitive congruences for Markovian process algebras, *Theoretical Computer Science.* 290(1): 117-160.
4. N. Ferns, P. Panangaden, D. Precup. (2004). Metrics for finite Markov decision processes, *UAI '04: Proceedings of the 20th conference on Uncertainty in artificial intelligence.* 162-169.
5. L.R. Yates, P.J. Campbell. (2012). Evolution of the cancer genome, *Nature review genetics.* 13(11): 795-806.
6. T. Brugère, Z. Wan, Y. Wang. (2024). Distances for Markov Chains, and Their Differentiation, in *Proceedings of Machine Learning Research.* 237:1-55.
7. F. Harary. (1965). *Structural models. An introduction to the theory of directed graphs*, Wiley.
8. I. Cho, *Measure Theory on Graphs*, e-print arXiv: math/0602025
9. H.H. Chen, S.W. Duffy, L. Tabar. (1997). A mover-stayer mixture of Markov chain models for the assessment of dedifferentiation and tumour progression in breast cancer, *Journal of Applied Statistics.* 24(3): 265-278.
10. Blumen, Kogan, and McCarthy. (1955). *The Industrial Mobility of Labor as a Probability Process*, Ithaca, New York: Cornell University Press.
11. A.L. Gibbs, F.E. Su. (2002). On Choosing and Bounding Probability Metrics, *International Statistical Review.* 70(3): 419-435.
12. P. Diaconis. (1988). *Group Representations in Probability and Statistics*, Hayward, CA: Institute of Mathematical Statistics. 34.

13. O.M. Lecian. (2024). Markov Models of Genomic Events, Global Journal of Medical and Clinical Case Reports. 11: 018-020.
14. M.L. Puterman. (1994). Markov decision processes: Discrete stochastic dynamic programming, Sons, Inc.
15. L.V. Kantorovich. (1939). Mathematical Methods of Organizing and Planning Production, Management Science. 6: 366-422.
16. S. Kullback, R.A. Leibler. (1951). On information and sufficiency, Annals of Mathematical Statistics. 22(1): 79-86.
17. C. Ma, M. Balaban, J. Liu, S. Chen, L. Ding, B.J. Raphael. Inferring allelespecific copy number aberrations and tumor phylogeography from spatially resolved transcriptomics.