DOI: https://doi.org/10.3126/jori.v12i1.84834

UNDERSTANDING TRANSITION MATRICES AND THEIR ROLE IN ANALYZING VIRUS MUTATION

Nand Kishor Kumar[1], *Dipendra Prasad Yadav*[2]

[1] *Trichandra Campus, Tribhuvan University, Nepal, nandkishorkumar2025@gmail.com*

[2]*Thakur Ram Multiple Campus, Nepal, dipendra.yadav2032@gmail.com*

Corresponding Author:

Dipendra Prasad Yadav; Thakur Ram Multiple Campus, Nepal, dipendra.yadav2032@

ABSTRACT

Transition matrices are a valuable tool for understanding the mutation and evolution of viruses. They provide insights into the probabilities of genetic changes, helping researchers make informed decisions in the development of vaccines, antiviral treatments, and public health strategies. As technology and data collection methods continue to advance, transition matrices will play an increasingly significant role in our ongoing battle against viral diseases, helping us stay one step ahead of the ever-changing world of viruses.

Keywords: *transition matrix, virus, viral infection, health services, public health*

INTRODUCTION

The concept of transition matrices is a fundamental tool in various fields of science and mathematics. In biology and virology, transition matrices play a crucial role in understanding the evolution and mutation of viruses (Koonin, 2016; Sole et al., 2021). Viruses are known for their ability to adapt and evolve rapidly, and transition matrices provide a powerful method for analysing and predicting these mutations.

Multiple factors, such as polymerase fidelity, sequence context, template secondary structure, cellular milieu, replication processes, proofreading, and post-replicative repair, influence viral mutation rates. Certain diversity-generating elements encoded in viruses and host-encoded cytidine/adenine deaminases can introduce large numbers of mutations (Sanjuan & Calap, 2016). A well-established framework for evaluating the fitness costs and advantages of such features is provided by Otero and Sanjuan's (2022) evolutionary perspective.

Great mathematicians and scientists like Hanssen et al. (2010), XU et al. (2017), De Almeida et al. (2020), Rodriguez et al. (2019 & 2020), De Almeida et al. (2022) and Garcia-Estrada et al. (2022) have published numerous papers on the issue of viruses, including their breeding, dispersion, evolution, and mechanisms.

This article explores the fundamentals of transition matrices and their applications in studying virus mutation, beginning with a historical overview of viruses.

HISTORY OF VIRUSES

Early References to Viruses

One of the earliest references to viral disease is found in Homer's mention of "rabid dogs," referring to rabies. Mesopotamians were also aware of rabies. Polio, caused by a virus, was depicted in ancient Egyptian artwork as lower limb paralysis (Mehndiratta et al., 2014). In South and Central America, smallpox epidemics devastated populations, a virus now eradicated worldwide (Henderson, 2011).

The Discovery of Viruses

Between 1886 and 1903, Ivanovsky first noted bacteria-like substances, and in 1898, Beijerinck demonstrated the filterable nature of viruses, proving them to be obligatory parasites (Ananthanarayan, 2006). Viruses cannot survive independently, and their genetic makeup can change over time, sometimes increasing or reducing virulence in the human host. Errors during genome replication make viruses prone to mutation, although complex organisms possess efficient proofreading mechanisms (Mandal, 2023).

The Origin of Viruses

Despite extensive study, the origins of viruses remain uncertain due to the absence of viral fossils. Their delicate structures prevent fossilisation or long-term preservation (Emerman & Malik, 2010). Research, therefore, relies on isolated viruses or recent material. Viruses likely originated in aquatic environments alongside hosts and later diversified as life spread onto land (Holland & Domingo, 1998; Emerman & Malik, 2010). Today, nearly all life forms, from fungi and bacteria to plants and archaea, host viruses (Forterre, 2010; Leonard & Toro, 2023).

VIRUS MUTATION

A mutation is a change in the genetic material, either DNA or RNA, from the original normal or "wild type" version of the genome of a particular organism or biological entity (Baake & Gabriel, 2000). The mutation may have occurred previously, or it may be entirely new. The natural process of creatures evolving and changing is called evolution. Geographic segregation is one process that can lead to the emergence of distinct lineages (strains or variants, depending on the preferred term). The presence of mutations in the SARS-CoV-2 coronavirus was neither unexpected nor surprising (Callway, 2020; Mandal, 2023).

TRANSITION MATRIX

Definition

A transition matrix is a square matrix that describes the probabilities of transitioning from one state to another in a Markov process. In the context of virus mutation, these states represent different genetic variants of the virus. Transition matrices can be used to model and quantify the rates of genetic changes, the emergence of new strains, and the likelihood of specific mutations occurring over time (Otero & Sanjuan, 2022).

Stochastic matrix

A transition matrix is a stochastic matrix whose (i, j) entry specifies the probability that an element will move from state s_i to state s_j during the next phase of the process. The probability is denoted by $P_{ij,}$ and it is independent of which states the chain was in prior

to the current state. Transition probabilities are the probability P_{ij}.

$(P_t)_{i,j} = \mathbb{P}(X_{t+1} = j | X_t = i)$.

P= $\begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,j} & \cdots \\ \vdots & & \ddots & \vdots & \\ P_{i,1} & P_{i,2} & \cdots & P_{i,j} & \cdots \end{bmatrix}$

This shows each row of the matrices is a probability vector and the sum of its entries is 1. The probability for going from state I to j is now calculated to be the following (Satorra, 2016):

$P_r(j|i) = P_{i,j}$

One important remark is that the row probability sums up to 1, since the sum of the probabilities of going from state i to any state is 1 (100 %). The probability of going from any state to another in k steps for a Markov chain can be calculated by using the matrix P. The probability is then given by P^k (Satorra, 2016).

N-step transition matrix

This higher-order transition matrix and find the chance of that transition occurring over multiple steps. The N-step matrix is calculated by raising the transition matrix to the power of N.

Properties of the transition matrix

The product of subsequent ones describes a transition along the time interval spanned by the transition matrices.

$P_0 . P_1$ has its $(i,j)^{th}$ position, the probability $X_2 = j$ given that $X_0 = 1$. In general, $(i,j)^{th}$ position of the $P_t . P_{t+1}$ is the probability $\mathbb{P}(X_{t+k+1} = j | X_t = i)$.

Theorem 1. Prove that for any natural numbers t and states i,j$\in S$, the matrix entry $(P_t . P_{t+1})_{i,j} = \mathbb{P}(X_{t+2} = j = i)$.

Proof. Let M = P_{t}.., for matrix multiplication,

$M_{i,j} = \sum_{k=1}^{n} (P_t)_{i,k} (P_{t+1})_{k,j} = \sum_{k=1}^{n} \mathbb{P}(X_{t+1} = k | X_t = i) \mathbb{P}(X_{t+2} = j | X_{t+1} = k)$

$= \mathbb{P}(X_{t+2} = j | X_t = i)$ is the expression of conditional probability.

The $\mathcal{K}$-step transition matrix is

$P_t^k = P_t . P_{t+1} \cdots\cdots\cdots P_{t+k+1}$ and by the above expression, satisfies,

$P_t^k = \begin{pmatrix} \mathbb{P}(X_{t+k} = 1 | X_t = 1) & \mathbb{P}(X_{t+k} = 2 | X_t = 1) & \mathbb{P}(X_{t+k} = n | X_t = 1) \\ \mathbb{P}(X_{t+k} = 1 | X_t = 2) & \mathbb{P}(X_{t+k} = 2 | X_t = 2) & \mathbb{P}(X_{t+k} = n | X_t = 2) \\ \vdots & \vdots & \ddots \quad \vdots \\ \mathbb{P}(X_{t+k} = 1 | X_t = n) & \mathbb{P}(X_{t+k} = 2 | X_t = n) & \mathbb{P}(X_{t+k} = n | X_t = n) \end{pmatrix}$

Example. For the time-independent Markov chain described by the picture below. What is its 2-step transition matrix?

The matrix is P = $\begin{pmatrix} .4 & .6 \\ .8 & .2 \end{pmatrix}$

Diagram 1.

Now, from- two-step transition matrix is

$P^2 = \begin{pmatrix} .4 & .6 \\ .8 & .2 \end{pmatrix} * \begin{pmatrix} .4 & .6 \\ .8 & .2 \end{pmatrix}$ where * is multiplication.

$\begin{pmatrix} .4\times.4+.8\times.6 & .4\times.6+.4\times.2 \\ .8\times.4+.2\times.8 & .8\times.6+.2\times.2 \end{pmatrix} =$

$\begin{pmatrix} .64 & .32 \\ .48 & .52 \end{pmatrix}$ process changing from one state to another state.

Diagram 1 denotes a two-state Markov process. Here, the arrows originate from the current state and point to be future state, and numbers associated with the arrows indicate the probability of the Markov process E changing from one state to another state.

ROLE OF TRANSITION MATRICES IN VIRUS MUTATION

Transition matrices are used to analyse the evolutionary pathways of viruses. By studying the probabilities of genetic changes occurring from one viral strain to another, researchers can construct a picture of how the virus mutates and adapts. This information is invaluable for understanding the origins and spread of viral outbreaks. Transition matrices allow researchers to predict the future mutational patterns of a virus. By analysing historical data, one can estimate the likelihood of specific mutations occurring in the future. This predictive capability is essential for designing effective vaccines and treatments. Transition matrices are employed in phylogenetic studies to build evolutionary trees of viral strains. By calculating the probabilities of mutations, researchers can determine the relationships between different viral variants and their common ancestors.

In the case of viruses like HIV or influenza, transition matrices help analyse the development of drug resistance. They provide insights into how a virus may evolve to overcome antiviral treatments, enabling healthcare professionals to adapt treatment strategies accordingly. Transition matrices can guide vaccine development by identifying stable viral strains and predicting the likelihood of specific mutations that could impact vaccine efficacy. This information aids in the design of vaccines that offer broader and longer-lasting protection.

MATERIALS AND METHODS FOR TRANSITION MATRIX

The construction of a transition matrix for virus mutation typically involves gathering genetic sequence data from a population of the virus over time. By comparing these sequences, researchers can identify specific mutations that have occurred and determine their probabilities. These probabilities are then organised into a matrix format, with rows and columns representing different genetic states and the entries indicating the probabili-

ties of transitioning from one state to another.

CHALLENGES

The use of transition matrices in virus mutation analysis is not without challenges. Gathering accurate genetic data and establishing mutation probabilities can be complex, especially for rapidly evolving viruses. Additionally, the accuracy of predictions depends on the quality and quantity of data available. Let a virus exist in N different strains, and in each generation, either stays the same or, with probability α, mutates to another strain, which is chosen at random.

The probability of the strain is calculated as an N-state chain with $N \times N$ transition matrix P given by

$P_{ii} = 1-\alpha$, $P_{ij} = \alpha/(N-1)$ for $I \neq j$ $\rightarrow$(1)

From this N-step transition matrix model, we can find the probability of the strain in other generation and nth generation. This will give information about all generations and can answer whether same or different. This information would be found by computing $P_{11}^{(n)}$. In fact, there is a much simpler approach that depends on exploiting the symmetry present in the mutation rules.

At any time, a transition is made from the initial state to another with probability α and a transition from another state to the initial state with probability $\alpha/(N-1)$.

So, we have a two-state chain diagram, and by putting $\beta/(N-1)$ in, we show the relation.

$P^{n+1} = P^n$ P and suppose initial state 1 and other state 2, then I = 1, j = 2 gives

$P_{11}^{(n+1)} = P_{12}^{(n)} + P_{11}^{(n)}(1-\alpha)$ $\rightarrow$ (2)

We can also simplify that

$P_{11}^{(n)} + P_{12}^{(n)} + P_1(X_n = 1 \text{ or } 2)$ eliminating P_{12} so new relation

$P_{11}^{(n+1)} = (1-\alpha-\beta)P_{11}^{(n)} + \beta$ and $P_{11}^{(0)} = 1$

(3)

$$P_{11}^{(n)} = \begin{cases} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & ; \alpha+\beta > 0 \\ 1 & ; \alpha+\beta = 0 \end{cases}$$

Now putting $\beta/(N-1)$ then

$P_{11}^{(n)} = \frac{\alpha/(N-1)}{\alpha+\alpha/N-1} + \frac{\alpha}{\alpha+\frac{\alpha}{(N-1)}}$ $\rightarrow$(3)

On solving, we get the probability

$\frac{1}{N} + \left(1-\frac{1}{N}\right)\left(1-\frac{\alpha N}{N-1}\right)^n$ $\rightarrow$(4)

With the help of Equation (4), we can study the virus mutation.

CONCLUSION

A useful mathematical tool for modelling, analysing, and interpreting transitions between several states is a transition matrix. Its uses cut across many different domains and spe-

cialities, supporting decision-making, forecasting results, and helping to understand system dynamics. Unlocking a transition matrix's potential and utilising its analytical power requires an understanding of its fundamentals, mathematical underpinnings, construction procedure, and interpretation techniques.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

FUNDING

None

REFERENCES

Ananthanarayan, R. (2006). *Ananthanarayan and Paniker's textbook of microbiology*. Orient Blackswan.

Baake, E., & Gabriel, W. (2000). Biological evolution through mutation. *Annual reviews of computational physics VII*, *7*, 203-264.

Callaway, E. (2020). The coronavirus is mutating—does it matter?.

De Almeida G.Q., Silva J.O., Copati M.G.F., Dias F.O., Santos M.C. (2020). Tomato breeding for disease resistance. *Multi-Sci. J. 3*, 8–16. doi: 10.33837/msj.v3i3.1287.

Emerman, M., & Malik, H. S. (2010). Paleovirology—modern consequences of ancient viruses. *PLoS biology*, *8*(2), e1000301.

Forterre, P. (2010). Defining life: the virus viewpoint. *Origins of Life and Evolution of Biospheres*, *40*(2), 151-160.

Garcia-Estrada R.S., Diaz-Lara A., Aguilar-Molina V.H., Tovar-Pedraza J. M. (2022). Viruses of Economic Impact on Tomato Crops in Mexico: From Diagnosis to Management-A Review. *Viruses. 14*, 1251. doi: 10.3390/v14061251.

Hanssen I.M., Lapidot M., Thomma, (2010). B.P. Emerging viral diseases of tomato crops. *Mol. Plant-Microbe Interact, 23*, 539–548. doi: 10.1094/MPMI-23-5-0539.

Henderson, D. A. (2011). The eradication of smallpox–an overview of the past, present, and future. *Vaccine*, *29*, D7-D9.

Holland, J., & Domingo, E. (1998). Origin and evolution of viruses. *Virus genes*, *16*(1), 13-21.

Leonard, J. M., & Toro, D. D. (2023). Defining the microbiome components (bacteria, viruses, fungi) and microbiome geodiversity. *Surgical infections*, *24*(3), 208-212.

Koonin, E. V. (2016). Viruses and mobile elements as drivers of evolutionary transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1701), 20150442.

Mandal, A. (2023). Virus history. *News-Medical*. https://www.news-medical.net/health/Virus-History.aspx.

Mehndiratta, M. M., Mehndiratta, P., & Pande, R. (2014). Poliomyelitis: historical facts, epidemiology, and current challenges in eradication. *The Neurohospitalist*, *4*(4), 223-229.

Otero, E.S., & Sanjuan, R. (2022). Cooperative virus-virus interactions: An evolutionary perspective. *PMC Journal*.

Rodriguez E., Téllez M.M., Janssen D. (2019). Whitefly Control Strategies against Tomato Leaf Curl New Delhi Virus in Greenhouse Zucchini. *Int. J. Environ. Res. Public Health*. *16*, 2673. doi: 10.3390/

ijerph16152673.

Rodriguez, V., Lagares, A., Arteaga, H., Mattar, S., & Ruiz, L. C. (2020). Viral emerging pathogen evolution. In *Emerging and reemerging viral pathogens* (pp. 35-51). Academic Press.

Sanjuan, R. & Calap, P. D. (2016). Mechanisms of viral mutation. *Cell Mol Life Science*,*72*(23), 4433-4448.

Satorra, P., & Tebé, C. (2024). Bayesian spatio-temporal analysis of the COVID-19 pandemic in Catalonia. *Scientific Reports*, *14*(1), 4220.

Solé, R., Sardanyés, J., & Elena, S. F. (2021). Phase transitions in virology. *Reports on Progress in Physics*, *84*(11), 115901.

Xu C.X., Sun X.P., Taylor A., Jiao C., Xu Y.M., Cai X.F., Wang X.L., Ge C.H., Pan G.H., Wang Q.X. (2017). Diversity, distribution, and evolution of tomato viruses in China uncovered by small RNA Sequencing. *J. Virol, 91,* e00173–17. doi: 10.1128/JVI.00173-17.