International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 07 – Issue 11 || November 2022 || PP. 01-05

Discrete Markov Chains with Embedded Regression for Simple Win-Loss Models

Joseph Kamau
Multi-media University of Kenya
Department of Mathematics

Anthony Karanjah
Multi-media University of Kenya
Department of Mathematics

Pius Kihara
Technical University of Kenya
Department of Mathematics

Abstract: Many situations and dynamic systems can be modeled as Discrete Markov Chains. They shift from one state to another according to their nature and can be described by evolving Markov Chain. However, major systems evolve continuously but can be discretized to evolve as discrete Markov chains (DTMC). This research looks at how to estimate the transition probabilities in a manner that incorporates a regression model whereby, covariates of the regression model likely influence the behavior of the state of the DTMC. The study looks at the interrelations of the states with a simple regression equation and how regression influences the generation of transition probabilities. The general objective of the research is to embed a regression model in DTMC that acts as a generator of transition probabilities(TP) and in particular apply it to simple win-loss models. The simple win-loss model will be represented as a Markov chain with two states.The results indicate that DTMC probabilities can be generated by a regression model. As the Markov chain evolves in time, the initial linear regression generating the probabilities changes to non-linear. This is because of the covariates of the model change with time.The research targets modeling systems with either a failure-success nature, win-loss nature, or death-life nature. If such a system can be improved to give a more accurate TP that indicates how the DTMC evolves as accurately as possible, then failures in such models can be reduced with a lot of success.
Key Words: Embed, Canonical form, Design matrix, partitioning, and Sequential matrix.

1. Introduction

1.1 Background of the Study

The key to improving systems and processes is examining and researching them to improve their performance. Observing how systems work is revolutionary in constructing model systems as close to the natural as possible. Scientists produced several equations that describe the deterministic totality of systems. After a large difference between the deterministic and adaptive models, the challenge grew increasingly difficult. To reduce error-related variations, statistical models were created to capture error-causing elements. Effective but not perfect. Over time, the problem has been to come up with models closest to the ideal-meaning that they can forecast from negligible error[1]. Random variables are used to represent systems that change randomly, suggesting that we can know the model's possible values but not its precise value at a certain time[2]. The implication requires models to characterize the behavior of random variables for different values, leading to probability distributions. Tracing random variables over time exacerbated the problem. Every random variable was realized. Joint distribution was necessary to handle several random variables, despite its limitations[3].

1.2 Statement of the Problem

Stochastic process studies entail developing a model of a physical process or condition, defining its parameters, and utilizing the model to forecast and regulate future activities. An ideal model has few mistakes from its hypothesized parameters and makes accurate predictions[4]. To accomplish this goal, the elements influencing the process should be collected and changed.

This research will construct a DTMC with regression-based probability. The regression model has time-varying covariates. This allows changing covariates as needed. Adding a regression model to DTMC will boost accuracy and reduce prediction mistakes. The research will also determine DTMC's distribution functions and equilibrium distributions. Finally, the equilibrium distribution model is compared to the proportion models.

2. Literature Review

Many natural processes are ongoing and massive, sufficiently distinct to model. Some continuous models can include a discrete model. Reference[5]describes embedding a time-discrete Markov chain into a continuous-time Markov chain on so many states. A continuous time Markov chain approximates the occurrence of each state at any point in time. When the states of DTMC are infinitely many, it implies that the states have an instantaneous rate of occurrence at a given time interval. The problem extends to the one-server waiting and renewal theory. The concept focuses on processes that appear to advance continuously but progress discretely, so the two models must be matched and joined.

Continuous Markov chains contain constant transition probabilities. Some probability changes over time and with new knowledge. Such information change can be captured by a regression model, leading to the calculation of transition probabilities. Macro data are used, but microdata can be approximated locally without a model[6]. The challenge involves estimating non-stationary probability using a dynamic model. Transition probabilities evolve in the order of indexed parameters in discrete Markov chains[7]. Each parameter level yields a discrete random variable. The Markov chain comprises all possible evolution routes for a random variable.Reference [8] states that Markov chain sample pathways can approximate autoregression models with good accuracy. Reverse autoregression models can be used to approximate a Markov chain and get transition probabilities. It's used in econometrics, finance, and economics.

DNA sequences can be analyzed as discrete Markov chains. Log-linear models are used to introduce the model. Log-linear models can yield contingency tables with the same margins as Markov chains because the observations depend on each other. The table's odd structure can be compared to a multinomial distribution[9]. Higher-dimensional sequences require a higher-order Markov chain model, which reduces power and computation. S-plus analysis was consistent and inherited multinomial distribution features appropriately. In the periodicity context of the Markov chain and its variation, degrees of freedom are defended.

Estimating the probability transition matrix describing chronic illness evolution is required when discrete Markov chains are employed to describe disease progression. Estimating the chain's transition probabilities is difficult due to a complex relationship. Reference [10]the transition matrix can be computed via maximum likelihood methods and variations to obtain maximum likelihood probabilities. Methods operate when the model cycle length matches with the observation interval, when it doesn't, and when cycle lengths are uneven. They propose a bootstrap method for assessing the accuracy and uncertainty of maximum likelihood probability estimates and for making model inferences like confidence intervals and forecasts.

A discrete Markov chain models disease transmission when affected people interact with the public. The model combines infection and recovery rates to reach equilibrium in a disease-infected population. Infected illness cases are fitted. People-to-people transmission is quantified via a discrete Markov chain model, and the model's time series are reevaluated to get more precise numbers[11]. The Monte Carlo approach estimates missing, unknown, unreported infected, recovered, and other population factors. This is the stochastic discrete model for infected-recovering time series. Markov chains and stochastic processes have first-hit times properties. The understudy's hitting time depends on the process's status. The endpoint of an individual reaches its critical point when it hits the worst threshold condition for the first time. Threshold regression can estimate the first-hitting time[12]. This regression has covariate-friendly structures. Covariates can affect process parameters, hitting times, and time variables.

An epidemic disease model was proposed by[13]. A discrete-time Markov chain on how the disease spread is implemented. A family of models which are parametrized by several disease-infected subjects per unit in every stochastic evolution is solved to obtain the transition probabilities for the contact process model. Reference [13] research focuses on how individuals get infected in each node and their accompanying probabilities. Finally, different infection models are constructed for a whole dimension of models and their crucial properties are determined. Reference [14]devised a buffer occupancy model to analyze computer buffer time and refreshment. The model is a discrete Markov chain that uses probability to measure the buffer's position and playback continuity. They devised the probability approximation using a probability distribution with a buffer. The model's convergence and asymptotic behavior were simulated to test its accuracy.

Mixed and mixture regression models model responses using Beta distribution. Beta distributions create transition probabilities for Markov chains. Markov chains model discontinuous structures for bounded scale scores[15]. Deterministic marginal maximum likelihood and Monte Carlo approaches are used to measure the model's efficiency. The assessment shows that modified mixture regression conforms to the probability transition model estimation. Complex deterioration separates soft tissues as the human body undergoes organized stages of breakdown. When the source of data is only conditionally monitored from physical facts, prognostics are used. According to reference [16], they implement a stochastic damage model for remaining degradation life and embed a Markov chain in the estimating scope. The embedded model determines time estimation reliability. The model is tested using fatigue crack propagation. Significant predictions are made, and revised equations are bound by the Markov chain's remaining life and time dependence.

Predicting ego-vehicle speed improves vehicle safety and fuel economy. An accurate prediction model must be estimated to maximize these qualities. Reference [17]examines six velocity prediction models using a Markov chain or Recurrent Neural Network. The two models evaluate velocity models' accuracy and execution time. Input parameters are fed into an embedded model, which provides an ego-vehicle profile. The two conjoined models' parameters are optimized using a Radial basis. The embedded Markov chain model is less accurate but faster in its execution.

3. Modeling win-loss by embedding a Regression model to a DTMC model

Modeling win-loss by embedding a regression model to a DTMC model will occur in four stages. The first stage is the partitioning of the regression model as described below;

The simple regression model of the win-loss modelis given by;$z_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

The response variable z_i is partitioned into $z_{1i}$for the responses in favor of the win situation and$z_{2i}$for theresponses in favor of the loss situation.

The independent variable$x_i$is also partitioned into the x_{1i}independent variable corresponding to the response variable $z_{1i}$and$x_{2i}$for the independent variable corresponding to the response variable z_{2i}.

The following table summarizes the partition and the partitioning criteria for the win-loss model:

Variable i	Outcomes; win=1, loss=0	Y_i partition; z_{1i}, z_{2i}	x_i partition; x_{1i}, x_{2i}
1	1	z_{11}	x_{11}
2	1	z_{12}	x_{12}
. . .	. . .	. . .	. . .
N	1	Z_{1n}	x_{1n}
n+1	0	$z_{2(n+1)}$	$x_{2(n+1)}$
. . .	. . .	. . .	. . .
N	0	z_{2N}	x_{2N}

Table 1

The first column shows the index of the responses collected from 1 to N, the second column shows the outcome Ifith index of responses resulting in a win or a loss, and the third column shows the grouping of the responses into two groups either a win which runs fromz_{11} to and z_{1n}or a loss from $z_{2(n+1)}$to $z_{2N}$and the last column shows the grouping of the independent variable x.

Given the regression partition and the threshold of the independent variable, the win-loss model for the different transitions between states can be written as follows;

The transition from a win to a win state:

$$z_{1(ww)i} = \beta_0 + \beta_{1x_{1(ww)i}}. \quad (1)$$

The transition from a win to a loss state is

$$z_{\{1(wl)i\}} = (\beta_1) * x_{\{(wl)i\}}. \quad (2)$$

And

$$z_{\{1(ww)i\}} + z_{\{1(wl)i\}} = z_{\{1i\}} = \beta_0 + \beta_{1x_{\{1i\}}}. \quad (3)$$

The transition from a loss to a loss state is

$$z_{\{2(ll)i\}} = b_0 + b_{1x_{\{2(ll)i\}}}. \quad (4)$$

The transition from a loss to a win state:

$$z_{\{2(lw)i\}} = b_{1x_{\{2(lw)i\}}}. \quad (5)$$

And

$$z_{\{2(ll)i\}} + z_{\{2(lw)i\}} = z_{\{2i\}} = b_0 + b_{1x_{\{2i\}}}. \quad (6)$$

In the next step, we create the DTMC and evaluate its validity.
Using the above equations, the following DTMC for the win loss model can be defined;

$$P = \begin{bmatrix} \frac{z_{1(ww)}}{z_1} & \frac{z_{\{1(wl)\}}}{z_1} \\ \frac{z_{2(lw)}}{z_2} & \frac{z_{2(ll)}}{z_2} \end{bmatrix} \quad (7)$$

Which is equivalent to:

$$P = \begin{bmatrix} \frac{\beta_0+\beta_{1x_{1(ww)i}}}{\beta_0+\beta_{1x_{1i}}} & \frac{(\beta_1)*x_{\{(wl)i\}}}{\beta_0+\beta_{1x_{1i}}} \\ \frac{b_0+b_{1x_{\{2(ll)i\}}}}{b_0+b_{1x_{\{2i\}}}} & \frac{b_{1x_{\{2(lw)i\}}}}{b_0+b_{1x_{\{2i\}}}} \end{bmatrix} \quad (8)$$

We now check the following conditions that a DTMC transition matrix must satisfy;

$$Y_n \in S, for \; \forall \; n \geq 0 \quad (9)$$

where Y_n is the process whose underlying transition matrix is P and S is the state space of Y_n.
The state space $S = \{1,0\}$ is a discrete state space, at any discrete time $n \geq 0$, we have the following collection of random variables;

$$\{Y_0, Y_1, Y_2, \dots, Y_n\}$$

This implies that

$$Y_0 \in S, Y_1 \in S \dots Y_n \in S$$

hence

$$Y_n \in S \text{ for all } n \geq 0$$

Also, we check if the summation of components in a row of the DTMC adds up to 1;

$$\sum_{j\in S} p_{ij} = 1 \quad (10)$$

For the first row we have

$$p_{11} + p_{10} = \frac{\beta_0+\beta_1 x_{1(ww)i}}{\beta_0+\beta_1 x_{1i}} + \frac{\beta_1 x_{1(wl)i}}{\beta_0+\beta_1 x_{1i_0}} \quad (11)$$

$$\Leftrightarrow \frac{\beta_0+\beta_1(x_{1(wl)i}+x_{\{1(wl)i)\}}}{\beta_0+\beta_1 x_{1i}} \quad (12)$$

But

$$x_{1(wl)i} + x_{1(wl)i} = x_{1i} \quad (13)$$

$$\Rightarrow \frac{\beta_0+\beta_1(x_{1(wl)i}+x_{\{1(wl)i)\}}}{\beta_0+\beta_1 x_{1i}} \quad (14)$$

$$=(\beta_0 + \beta_1 x_{1i})/(\beta_0 + \beta_1 x_{1i}) \quad (15)$$

$$=1 \quad (16)$$

Similarly;

$$p_{01} + p_{00} = \frac{b_1 x_{2(lw)i}}{b_0+b_1 x_{2i}} + \frac{b_0+b_1 x_{2(ll)i}}{b_0+b_1 x_{2i}} \quad (17)$$

$$\Rightarrow \frac{b_0+b_1(x_{2(lw)i}+x_{\{2(ll)i)\}}}{b_0+b_1 x_{2i}} \quad (18)$$

but

$$x_{2(lw)i} + x_{2(ll)i} = x_{2i} \quad (19)$$

$$=(b_0 + b_1 x_{2i})/(b_0 + b_1 x_{2i}) \quad (20)$$

$$=1 \quad (21)$$

4. Conclusion

The research focused on creating a DTMC with embedded regression for the win-loss model. From the research discussed above, the regression was a simple regression model, and the states of the DTMC partially depended on the covariate. The states probabilities for the transition probability matrix are written in the form regression model. The regression is written with the response being the percentage probability of transition from one state to another and the covariates of the regression model being factors influencing how the model shift from one state to another. The regression model generates the probability and the DTMC is used for predicting the states in long run.

The model is valid since it satisfies all the equations that a transition probability matrix should satisfy. Themodelis an improvement of a proportional model where covariates generate extra information for state transition probabilities.

Acknowledgment

My family's help is appreciated. They helped me financially and kept me motivated. Evelyne's patience, attention, input, and efforts inspired me every day to produce something positive in this study project without giving up. My gratitude also goes to my supervisors and the Mathematics Department for their effort, corrections, direction, and sacrifice.

References

[1]. J. Shin, K. Yeon, S. Kim, M. Sunwoo, M. H.-I. Access, and undefined 2021, "Comparative study of Markov chain with recurrent neural network for short term velocity prediction implemented on an embedded system," *ieeexplore.ieee.org*, Accessed: Oct. 31, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9345689/

[2]. S. R.-I. to P. Models and undefined 2014, "Random variables," *fatihcavdur.home.uludag.edu.tr*, Accessed: Oct. 31, 2022. [Online]. Available: http://fatihcavdur.home.uludag.edu.tr/docs/end5155/lec02.pdf

[3]. M. L. Howard, "Distribution functions and the Daniell, Kolmogorov theorem," 1972, Accessed: Oct. 31, 2022. [Online]. Available: https://uh-ir.tdl.org/handle/10657/8627

[4]. U. Bhat and G. Miller, *Elements of applied stochastic processes*. 2002. Accessed: Oct. 31, 2022. [Online]. Available: https://mubert.marshall.edu/bert/syllabi/307320170114716368193783.pdf

[5]. J. Runnenburg and C. Scheffer, *On Elfving's problem of imbedding a time-discrete Markov chain in a time-continuous one for finitely many states*. 1962. doi: 10.1016/S1385-7258(62)50052-8.

[6]. J. Wu, "End-to-end seismic risk analysis framework for the identification of infrastructure network retrofits," 2017, Accessed: Oct. 31, 2022. [Online]. Available: https://search.proquest.com/openview/55c72f52b0b6be03163a9239cee1c65e/1?pq-origsite=gscholar&cbl=18750&diss=y

[7]. J. R.-F. E. J. Theor. Stat and undefined 2001, "A review of asymptotic convergence for general state space Markov chains," *probability.ca*, Accessed: Oct. 31, 2022. [Online]. Available: http://probability.ca/jeff/ftpdir/gentheory.pdf

[8]. G. T.-E. letters and undefined 1986, "Finite state markov-chain approximations to univariate and vector autoregressions," *Elsevier*, vol. 20, pp. 177–181, 1986, Accessed: Oct. 31, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0165176586901680

[9]. P. J. Avery and D. A. Henderson, "Fitting Markov chain models to discrete state series such as DNA sequences," *J R Stat Soc Ser C Appl Stat*, vol. 48, no. 1, pp. 53–61, 1999, doi: 10.1111/1467-9876.00139.

[10]. B. A. Craig and P. P. Sendi, "Estimation of the transition matrix of a discrete-time Markov chain," *Health Econ*, vol. 11, no. 1, pp. 33–42, 2002, doi: 10.1002/HEC.654.

[11]. J. Chodosh *et al.*, "Meta-analysis: Chronic disease self-management programs for older adults," *Ann Intern Med*, vol. 143, no. 6, Sep. 2005, doi: 10.7326/0003-4819-143-6-200509200-00007.

[12]. M. Lee, G. W.-S. Science, and undefined 2006, "Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary," *projecteuclid.org*, vol. 21, no. 4, pp. 501–513, 2006, doi: 10.1214/088342306000000330.

[13]. S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, "Discrete-time Markov chain approach to contact-based disease spreading in complex networks," *iopscience.iop.org*, 2010, Accessed: Oct. 31, 2022. [Online]. Available: https://iopscience.iop.org/article/10.1209/0295-5075/89/38009/meta

[14]. A. Adamu, Y. Gaidamaka, and A. Samuylov, "Discrete Markov chain model for analyzing probability measures of P2P streaming network," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6869 LNCS, pp. 428–439, 2011, doi: 10.1007/978-3-642-22875-9_39.

[15]. J. Verkuilen, … M. S. E. and B., and undefined 2012, "Mixed and mixture regression models for continuous bounded responses using the beta distribution," *journals.sagepub.com*, vol. 37, no. 1, pp. 82–113, 2012, doi: 10.3102/1076998610396895.

[16]. J. Chiachio, M. Chiachio, J. Chiachío, M. L. Jalón, M. Chiachío, and A. Kolios, "A Markov chains prognostics framework for complex degradation processes," *Elsevier*, doi: 10.1016/j.ress.2019.106621.

[17]. J. Shin, K. Yeon, S. Kim, M. Sunwoo, M. H.-I. Access, and undefined 2021, "Comparative study of Markov chain with recurrent neural network for short term velocity prediction implemented on an embedded system," *ieeexplore.ieee.org*, Accessed: Oct. 31, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9345689/