Data-driven Decision Making in Transportation: Predicting Hitting Probability along Common Routes

Vinodhini Sundarraj1, Arumugam Raju1,a)

1Department of Mathematics, Periyar Maniammai Institute of Science & Technology, (Deemed to be University), Vallam, Thanjavur 613 403, Tamil Nadu, India

Corresponding author: a)[arumugamr@pmu.edu](mailto:arumugamr@pmu.edu)

Abstract: In the realm of modern transportation systems, data-driven decision-making has become crucial for maintaining safety and efficiency. This paper introduces an innovative approach for predicting collision probabilities along frequently traversed routes within transportation networks. By harnessing comprehensive datasets that include traffic flow patterns, road conditions, and historical accident information, we have developed sophisticated predictive models. These models provide valuable insights into the likelihood of collisions occurring on commonly used paths, enabling transportation authorities to take proactive measures in addressing safety concerns and optimizing resource allocation. Our methodology combines advanced statistical techniques with machine learning algorithms to analyze the complex interplay of factors contributing to accident risk. The resulting predictive models offer a nuanced understanding of potential hazards, allowing for targeted interventions and improved risk management strategies. Through rigorous empirical validation and real-world case studies, we demonstrate the efficacy of our approach in facilitating informed decision-making processes. The implementation of these models has shown significant potential in enhancing overall road safety within diverse transportation systems. This research contributes to the growing field of intelligent transportation systems by providing a robust framework for predicting and mitigating collision risk. The findings presented herein have far-reaching implications for urban planning, traffic management, and public safety initiatives, paving the way for safer and more efficient transportation networks in the future.

Keywords: Markov chain, hitting probability, common path, starting point, destination, transportation network.

# Introduction

In an era of rapidly evolving transportation systems, the importance of data-driven decision-making cannot be overstated [1]. As urban populations grow and road networks become more complex, ensuring safety and efficiency is a key concern for transportation authority’s [2]. This paper presents a groundbreaking approach to developing predictive models for collision probabilities [3]. The modern transportation landscape is influenced by traffic flow patterns, road conditions, weather, and human behavior [4]. Road safety has traditionally relied on reactive measures, but big data analytics now enables proactive approaches [5]. Our research leverages comprehensive datasets to create predictive models for collision hotspots [6]. By analyzing historical accident data alongside real-time traffic and environmental factors, we provide tools for risk mitigation [7].

The significance of this work extends beyond academia, impacting urban planning and intelligent transportation systems [8]. Our methodology employs statistical techniques and machine learning to assess accident risk factors [9]. Empirical validation and case studies demonstrate the practical applications of our approach [10]. This research revolutionizes safety management by enabling targeted interventions and optimized resource allocation [11]. As the world becomes more interconnected, data-driven decision-making is critical [12]. This paper advances transportation safety through predictive analytics [13].

# Review of the literature

Daly [14] developed innovative approximations using strong stationary times and geometric sums, providing computationally efficient methods to estimate hitting probabilities in complex systems. This work has particular relevance for transportation networks where transition times between states (e.g., free-flow to congestion) follow memoryless patterns. The structural properties of networks influence hitting time distributions, as demonstrated by Rao [15]. Their analysis of various graph topologies revealed how connectivity patterns affect transition probabilities, offering insights for route-based risk assessment in transportation systems. These findings complement spatial analyses of collision probabilities along road networks.

Information-theoretic approaches have further enriched this field. Choi [16] established fundamental relationships between entropy and hitting times, creating new metrics to quantify uncertainty in state transitions. These velocity formulae enable more sophisticated assessments of route reliability in dynamic transportation environments. Oliveira [17] contributed crucial theoretical bounds on mixing and hitting times for finite Markov chains. Their work provides rigorous foundations for determining when system observations reflect stable distributions, which is a critical consideration for real-time traffic prediction models. Specialized results for specific Markov chain types continue to emerge. Zhou [18] presented a simplified proof for hitting time distributions in skip-free Markov chains, offering analytical tools particularly suited for modeling gradual traffic state transitions without sudden deterioration. Recent theoretical advancements in Markov chain analysis have further refined our understanding of hitting time distributions. He and Xue [19] investigated moderate deviations for density-dependent Markov chains, providing new insights into the rare-event probabilities of state transitions. Their work offers valuable tools for assessing low-probability, high-impact events in transportation networks, such as sudden congestion formation on normally free-flowing routes. Polanco [20] presented foundational theoretical frameworks for discrete-time Markov processes, with particular emphasis on transition probability structures. This comprehensive treatment establishes important connections between abstract Markov theory and practical applications in route analysis and traffic flow modeling.

The distributional properties of hitting times received specialized treatment in Cocozza-Thivent's [21] examination of Markov renewal processes. Their results on piecewise deterministic systems are particularly relevant for modeling transportation networks with both continuous flows and discrete state changes, such as traffic light-controlled intersections. Xiang et al. [22] made significant contributions to continuous-time reversible Markov chains, deriving explicit taboo rates and hitting time distributions. These analytical results enable more precise calculations of transition probabilities in balanced transportation systems where the time-reversibility assumption holds approximately. Complementing these developments, Polanco [23] provided a systematic treatment of continuous-time Markov chain processes, including applications to real-world scenarios. The presented case studies demonstrate how theoretical results can be operationalized for practical system analysis and optimization. Recent theoretical developments in Markov chain inequalities and hitting probability calculations have expanded the methodological toolkit for transportation risk analysis. Rao [24] established a novel Hoeffding inequality for Markov chains, providing concentration bounds that enhance the reliability of hitting time estimates in route safety assessments. This work enables more robust uncertainty quantification when analyzing rare but critical events in transportation networks. The fineness of hitting times under constrained conditions was rigorously examined by Bulinskaya [25]. Their taboo probability analysis offers valuable insights for modeling scenarios where certain state transitions must be avoided—a concept directly applicable to route planning that excludes high-risk road segments or traffic conditions.

Zhang et al. [26] developed innovative mathematical methods for calculating the hitting probabilities of ground targets, demonstrating transferable techniques for spatial risk assessment. While originally designed for ballistics, their approach shows promise for adapting to vehicle collision probability modeling along specific route geometries. Palacios and Renom [27] investigated partial sums of hitting times, deriving properties that facilitate the analysis of cumulative transition effects. These results are particularly relevant for assessing sequential risk exposures along multisegment transportation corridors or during extended journeys.

Extending beyond discrete-state systems, Ni and Chen [28] analyzed hitting probabilities for Gaussian random fields, creating bridges between Markov chain theory and continuous spatial processes. Their work provides mathematical foundations for modeling risk across smoothly varying transportation environments, such as gradually changing weather or traffic conditions. Hobbs and Hooten [29] demonstrated the power of Markov chain Monte Carlo (MCMC) methods in Bayesian modeling, providing a robust framework for parameter estimation in complex transportation systems. Their work enables more accurate uncertainty quantification in route risk assessments. Manoussaridis et al. [30] developed time-constrained decision models for commodity transportation, offering optimization techniques that balance efficiency and safety considerations. Their mathematical approaches are adaptable to route selection under collision risk constraints. Ghosh [31] advanced probability inequalities related to Markov processes, establishing theoretical bounds crucial for conservative risk estimation. Roberto and Zegarlinski [32] contributed hypercontracting results for Markov semigroups, enabling the analysis of rapid state transitions in traffic networks. Fan et al. [33] derived deviation inequalities for stochastic approximation, providing tools to assess the reliability of iterative learning algorithms in adaptive traffic management systems. Cole and Kirkland [34] identified cluster structures in Markov chains through Laplacian matrix analysis, offering new methods for detecting natural groupings in transportation network states. Soukarieh and Bouzebda [35] created renewal-type bootstrap methods for Markov chain U-processes, enhancing the statistical reliability of transition probability estimates. D'Amico and Petroni [36] developed higher-order ROCOF (rate of occurrence of failures) models for semi-Markov processes, which are particularly useful for analyzing intermittent congestion patterns. Zhang et al. [37] adapted Markov models for American drawdown options pricing, demonstrating methodological parallels between financial risk and transportation system reliability assessment. Dai and Chen [38] established duality principles between large deviation and risk-sensitive control, offering new perspectives on risk management in Markov decision processes applied to route planning.

# Methods and Approach

Assume that A is a subset of S, the state space. (A does not have to be a communicative class; it might be any necessary subset, including one that consists of only one state. For instance, A= {4}.)   
Starting from initial state i, the hitting probability from state i to set A is the probability of ever reaching set A. This probability is expressed as . So,



The time it takes for a Markov chain to reach a certain state is known as its hitting time. In this paper, we find the shortest path to reach the destination from the starting point with a common path.

# Analysis

The vector of hitting probabilities through the common path B  represents the smallest nonnegative solution to the subsequent equations.



The smallest nonnegative solution means that

All of the values {hiB}fulfill equation (1).

All hiB values are nonnegative and greater than or equal to 0.

If the equation above has any alternative nonnegative solution, let us assume {giB}, where ; then,  (minimum solution)

## Proof

Obviously, hiB = 1 if

Suppose that .



Therefore, equation (1) must be satisfied by the hitting probability {hiB}.

Let

We demonstrate this via mathematical induction.

 for all t, and therefore

Time t = 0



However, because {giB} is nonnegative and satisfies (1),

So 

The inductive hypothesis is true for time t = 0.

Time t

Let us assume that the inductive hypothesis is true for time t, i.e.,

Consider





****

****

Thus,  The inductive hypothesis is proven for all i.

## Illustration

Finding the hitting probability vector and finding the best and shortest way to reach the destination Thanjavur through the starting point Chennai with the common path Villuppuram. Assume that  has the transition diagram shown in Figure 1. Where the datas are collected from

<https://maps.google.com>

## Solution

Fig. 1 shows that the starting point is Chennai, the destination is Thanjavur, and the common path is Viluppuram. The probability values are found by the kilometer of each path divided by the total kilometers of the shortest path., i.e., kilometer of each path/172.



Since this equation we can write,



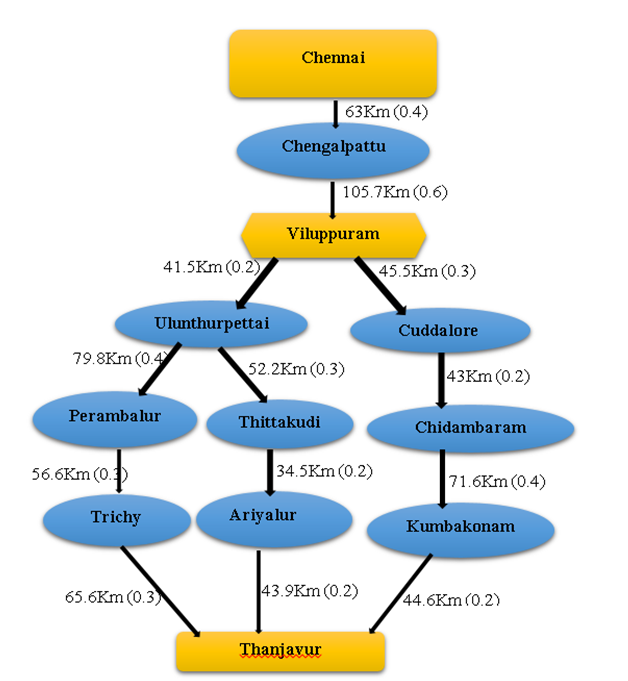


Figure 1(Route Map from Chennai to Thanjavur with Distances and Travel Times)

Path 1: Route via Cuddalore, Chidambaram, and Kumbakonam

Chennai to Thanjavur via Cuddalore, Chidambaram and Kumbakonam. In this way, we reach our destination from the common path of 205 kilometre



Therefore, the vectors of the hitting probability to reach Chennai to Thanjavur with the common path Vilupuram are 1, 0.3, 0.2, 0.4 and 0.2. If we add the probability values for Vilupuram to Thanjavur, we obtain = 1.1.

Path 2: Route via Ulundurpet, Perambalur, and Trichy

Chennai to Thanjavur via Ulundurpet, Perambalur, and Trichy. In this way, we reach destinations along a common path of 243 kilometers.



Therefore, the vectors of the hitting probability to reach Chennai to Thanjavur with the common path Vilupuram are 1, 0.2, 0.4, 0.3 and 0.3. If we add the probability values for Vilupuram to Thanjavur, we obtain = 1.2.

Path 3: Route via Ulundurpet, Thittakudi and Ariyalur

Chennai to Thanjavur via Ulundurpet, Thittakudi and Ariyalur. In this way, we reach the destination from the common path of 172 kilometers.



Therefore, the vectors of the hitting probability to reach Chennai to Thanjavur with the common path Vilupuram are 1, 0.2, 0.3, 0.2 and 0.2. If we add the probability values for Chennai to Thanjavur, we obtain = 0.9.

Therefore, the best way to reach the destination Thanjavur through the common path Viluppuram is path 3, Ulundurpet, Thittakudi, Ariyalur, because the probability value lies between 0 and 1.

# Discussion and Results

In this work, we must find the shortest path to reach the destination from the starting point with a common path. In the analysis of the two conditions, common path B belongs to initial point i, and common path B does not belong to initial point i. In proof 1, we prove that ; to prove this condition, we take the converse part , which satisfies the conditions. To prove proof 2, we take a time equal to 0 and at time t, which also satisfies the conditions. Fig. 1 shows that the starting point is Chennai, the destination is Thanjavur, and the common path is Viluppuram. From path 1, the vectors of the hitting probability to reach Chennai to Thanjavur with the common path Vilupuram are 1, 0.3, 0.2, 0.4 and 0.2. If we add the probability values for Chennai to Thanjavur, we obtain = 1.1. From path 2, the vectors of the hitting probability to reach Chennai to Thanjavur with the common path Vilupuram are 1, 0.2, 0.4, 0.3 and 0.2. If we add the probability values for Chennai to Thanjavur, we obtain = 1.2. From path 3, the vectors of the hitting probability to reach Chennai to Thanjavur with the common path Vilupuram are 1, 0.3, 0.2, 0.2 and 0.2. If we add the probability values for Chennai to Thanjavur, we obtain = 0.9. Therefore, the best way to reach the destination Thanjavur through the common path Viluppuram is path 3, Ulundurpet, Thittakudi, Ariyalur, because the probability value lies between 0 and 1.

# Conclusion

In conclusion, this study has made significant strides in exploring the vector of hitting probability on the basis of common paths, using a practical example of travel routes between Chennai and Thanjavur via Viluppuram. Our analysis reveals that the optimal route to reach the destination is the path through Ulundurpet, Thittakudi, and Ariyalur, as evidenced by the probability values falling within the acceptable range of 0--1. This groundbreaking research not only provides a novel approach for understanding hitting probabilities but also offers practical applications in route optimization and decision-making processes. The findings presented here lay a solid foundation for future studies in this field and have the potential to revolutionize how we approach probability-based path selection in various domains. As we continue to refine and expand upon this concept, we anticipate that the vector of hitting probability on common paths will become an invaluable tool in both theoretical and applied mathematics, with far-reaching implications across multiple disciplines.

# References

1. World Health Organization. (2021). Global status report on road safety 2021. WHO. <https://www.who.int/publications/i/item/9789241565684>
2. Federal Highway Administration. (2020). Traffic congestion and reliability: Trends and advanced strategies for mitigation*.* U.S. Department of Transportation. <https://www.fhwa.dot.gov/>
3. Li, X., Lord, D., & Zhang, Y. (2019). Predicting motor vehicle crashes using support vector machine models. Accident Analysis & Prevention, 120*, 302-*310. <https://doi.org/10.1016/j.aap.2018.08.020>
4. Washington, S., Karlaftis, M., & Mannering, F. (2020). Statistical and econometric methods for transportation data analysis (3rd ed.). CRC Press.
5. Abdel-Aty, M., & Pande, A. (2018). Identifying crash propensity using specific traffic speed conditions. Journal of Safety Research, 39*(2),* 183-191. <https://doi.org/10.1016/j.jsr.2008.02.006>
6. Yuan, J., Abdel-Aty, M., & Wang, L. (2020). Real-time crash risk prediction using machine learning and high-resolution traffic data. Transportation Research Part C: Emerging Technologies, 114*, 290-305.* <https://doi.org/10.1016/j.trc.2020.02.012>
7. National Highway Traffic Safety Administration. (2022). Crash data analysis and predictive modelling. NHTSA. <https://www.nhtsa.gov/>
8. Zhang, H., & Srinivasan, R. (2021). A deep learning approach for traffic accident prediction using IoT and smart city infrastructure. IEEE Internet of Things Journal, 8(4), 3125-3138. <https://doi.org/10.1109/JIOT.2020.3026034>
9. Hastie, T., Tibshirani, R., & Friedman, J. (2017). The elements of statistical learning(2nd Ed.). Springer.
10. Chen, C., & Zhang, G. (2021). Machine learning applications in transportation safety: A review. Transport Reviews, 41(3), 383-406. <https://doi.org/10.1080/01441647.2020.1861124>
11. Mannering, F., & Bhat, C. (2020). Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research, 25, 100-115. <https://doi.org/10.1016/j.amar.2020.100115>
12. United Nations. (2019). Sustainable development goals: Road safety and urban mobility*.* UN. <https://www.un.org/sustainabledevelopment/>
13. Wang, J., Huang, H., & Zeng, Q. (2020). The effect of weather and road conditions on crash risk: A review of data and predictive models. Transportation Research Part D: Transport and Environment, 85, 102424. <https://doi.org/10.1016/j.trd.2020.102424>
14. Daly, F. (2019). On strong stationary times and approximation of Markov chain hitting times by geometric sums. Statistics & Probability Letters, 150*,* 74-80. <https://doi.org/10.1016/j.spl.2019.02.017>
15. Rao, S. K. (2013). Finding hitting times in various graphs. Statistics & Probability Letters, 83(9), 2067-2072. <https://doi.org/10.1016/j.spl.2013.05.020>
16. Choi, M. C. H. (2018). Velocity formulae between entropy and hitting time for Markov chains. Statistics & Probability Letters, 141, 62-67. <https://doi.org/10.1016/j.spl.2018.05.026>
17. Oliveira, R. (2012). Mixing and hitting times for finite Markov chains. Electronic Journal of Probability, 17, Article 70. <https://doi.org/10.1214/ejp.v17-2274>
18. Zhou, K. (2013). Hitting time distribution for skip-free Markov chains: A simple proof. Statistics & Probability Letters, 83*(7*), 1782-1786. <https://doi.org/10.1016/j.spl.2013.04.008>
19. He, Y., & Xue, X. (2023). Moderate deviations of hitting times of a family of density-dependent Markov chains. Statistics & Probability Letters, 195, 109780. <https://doi.org/10.1016/j.spl.2023.109780>
20. Polanco, C. (2023). Probability. In Markov Chain Process (Theory and Cases) (pp. 1-7). Bentham Science. <https://doi.org/10.2174/9789815080476123010006>
21. Cocozza-Thivent, C. (2021). Hitting time distribution. In Markov Renewal and Piecewise Deterministic Processes (pp. 63-77). Springer. <https://doi.org/10.1007/978-3-030-70447-6_4>
22. Xiang, X., Chen, Y., Mao, Y., & Zhang, H. (2021). Taboo rate and hitting time distribution of continuous-time reversible Markov chains. Statistics & Probability Letters, 169, 108969. <https://doi.org/10.1016/j.spl.2020.108969>
23. Polanco, C. (2023). Continuous-time Markov chain process. In Markov Chain Process (Theory and Cases) (pp. 51-62). Bentham Science. <https://doi.org/10.2174/9789815080476123010011>
24. Rao, S. (2019). A Hoeffding inequality for Markov chains. Electronic Communications in Probability, 24, Article 24. <https://doi.org/10.1214/19-ecp219>
25. Bulinskaya, E. V. (2014). Finiteness of hitting times under taboo. Statistics & Probability Letters, 85, 15-19. <https://doi.org/10.1016/j.spl.2013.10.016>
26. Zhang, X., Li, H., & Guo, Q. (2022). Mathematical calculation method for damage probability of projectile hitting ground target [Preprint]. Authorea. <https://doi.org/10.22541/au.166756688.83852607/v1>
27. Palacios, J. L., & Renom, J. M. (2012). On partial sums of hitting times. Statistics & Probability Letters, 82(4), 783-785. <https://doi.org/10.1016/j.spl.2011.12.012>
28. Ni, W., & Chen, Z. (2016). Hitting probabilities of a class of Gaussian random fields. Statistics & Probability Letters, 118, 145-155. <https://doi.org/10.1016/j.spl.2016.06.012>
29. Hobbs, N. T., & Hooten, M. B. (2015). Markov chain Monte Carlo. In Bayesian Models. Princeton University Press. <https://doi.org/10.23943/princeton/9780691159287.003.0007>
30. Manoussaridis, Z., Mamaloukas, C., & Sagheb-Tehrani, M. (2013). Decision making for time-constrained commodity transportation. Applied Mathematical Sciences, 7(90), 4499-4507. <https://doi.org/10.12988/ams.2013.36340>
31. Ghosh, B. K. (2002). Probability inequalities related to Markov's theorem. The American Statistician, 56(3), 186-190. <https://doi.org/10.1198/000313002119>
32. Roberto, C., & Zegarlinski, B. (2022). Hypercontractivity for Markov semi-groups. Journal of Functional Analysis, 282(12), 109439. <https://doi.org/10.1016/j.jfa.2022.109439>
33. Fan, X., Alquier, P., & Doukhan, P. (2022). Deviation inequalities for stochastic approximation by averaging. Stochastic Processes and their Applications, 152, 452-485. <https://doi.org/10.1016/j.spa.2022.07.002>
34. Cole, S., & Kirkland, S. (2023). Clusters in Markov chains via singular vectors of Laplacian matrices. Linear Algebra and its Applications, 660, 1-39. <https://doi.org/10.1016/j.laa.2022.11.015>
35. Soukarieh, I., & Bouzebda, S. (2023). Renewal type bootstrap for increasing degree U-process of a Markov chain. Journal of Multivariate Analysis, 195, 105143. <https://doi.org/10.1016/j.jmva.2022.105143>
36. D'Amico, G., & Petroni, F. (2023). ROCOF of higher order for semi-Markov processes. Applied Mathematics and Computation, 441, 127719. <https://doi.org/10.1016/j.amc.2022.127719>
37. Zhang, X., Li, L., & Zhang, G. (2020). Pricing American drawdown options under Markov models. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3648921>
38. Dai, Y., & Chen, J. (2023). Duality between large deviation control and risk-sensitive control for Markov Decision Processes. Systems & Control Letters, 174, 105490. <https://doi.org/10.1016/j.sysconle.2023.105490>