Train Delay Evolution Model Using Continuous Time Markov Processes

İsmail Şahin (sahin@yildiz.edu.tr)[1*] and Mehmet Ş. Artan (msartan@hotmail.com)[1]

[1]Yildiz Technical University, Department of Civil Engineering, 34220 Istanbul, Turkey

*Corresponding author

Introduction

Investigation of train delay evolution and prediction is one of the hot topics in railway operations research. This is required for the effective and efficient realization as well as the quality improvement of train services. A comprehensive literature review on data-driven approaches for train delay prediction has recently been presented in Tiong et al. (2023). The authors prefer to develop a framework including design concept, modelling and evaluation. Furthermore, six modules are also defined to evaluate the existing literature and identify possible research gaps. Predictions are classified into short- and long-term with next station and multiple station coverage. The paper discusses the practicality of the models. Spanninger et al. (2022) review papers on train delay prediction and classify them as event- and data-driven approaches for multi- and single-step predictions in deterministic and stochastic modeling schemes. It is reported that their stand-alone applications can produce acceptable predictions while their combined implementations are more effective. However, the authors claim that the implementation and calibration of the approaches are difficult. The above-mentioned reviews cover a wide range of papers investigating the train delay evolution and prediction problems. The Markov chains is amongst the techniques used for this purpose. The interpretability, transparency and stability support its applicability in this respect. Artan and Şahin (2022) and Şahin (2017) investigated the train delay evolution, propagation and prediction problems by using the discrete-time discrete-space Markov chains. Having shown the workability of this discrete stochastic approach, in this paper we extended our research towards the continuous time Markov processes in order to investigate the stochastic evolution of train delays using the continuous probability functions of time. To the best of our knowledge, this modelling approach applied for train delay investigation is used for the first time in the literature.

Methodology

In the discrete time Markov chain case, transitions occur at steps with the corresponding probabilities defined in the one-step transition probability matrix representing a discrete sequence of random variables. In the continuous time Markov processes (or chains), however, the transition probabilities can be defined in time in a stochastic matrix sized in finite number of states because the transitions can occur at any instant. The transition probabilities can be obtained by matrix multiplications in the discrete-time case while they are obtained by solving a set of first order differential equations in the continuous-time case (Solberg, 2009 and Nelson, 1995).

The random variables indexed by a continuous variable t constitute a continuous time stochastic process having uncountably infinite set of variables $X(t)$, $t \geq 0$, which are related to each other. Therefore, the following equation representing the Markov assumption is used to simplify these inter relations in which for all state sequences $j_0, j_1, j_2, \ldots, j_n$ and all times $t_0 < t_1 < t_2, \ldots, < t_n$:

$$P(X(t_n) = j_n | X(t_{n-1}) = j_{n-1}, X(t_{n-2}) = j_{n-2}, \ldots, X(t_0) = j_0)$$

$$= \quad P(X(t_n) = j_n | X(t_{n-1}) = j_{n-1})$$

This simplifying assumption expresses that the probability of being in any state at any time depends only on the most recent state (or information), disregarding the previous states at earlier instants. This further assumes that the length of time the process has been in the current state is not influential (i.e., the process is memoryless). Although this simplifying assumption is strong, it conforms to the actual train operation processes (e.g., sectional running and station dwelling) represented by delays as state variables.

Another simplification is taken place for the actual train operation processes (e.g., sectional running and station dwelling), making use of the stationarity assumption represented by the following equation:

$$P(X(t+s) = j | X(s) = i) = P(X(t) = j | X(0) = i)$$

The stationarity assumption expresses that the probabilities for the same state transitions in the same time interval are the same during the entire processes.

In the discrete case a transition could occur at the discrete point of time called epoch. The epochs, on the other hand, are the instants at which the transitions occur, in continuous time. The transitions are associated with the probabilities in the former while the transition between any distinct states i and j is associated with a transition rate λ_{ij} in the latter. The meaning of the transition rate λ_{ij} (≥ 0) is the speed of transition from current state i to next state j (or the number of transitions per unit time). The entire continuous time Markov process can be described by the transition rates between distinct states ($i \neq j$). For the diagonal elements ($i = j$), no transition occurs, and the rates are set as below:

$$\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$$

The transition rates can be easily established from the data. Let $T_{ij}(k)$ be the kth transition time observed from state i to state j and n_{ij} be the number of transition instances. The estimated mean time from state i to j can statistically be calculated as follows:

$$E(T_{ij}) \approx \frac{\sum_{k=1}^{n_{ij}} T_{ij}(k)}{n_{ij}}$$

Hence, the corresponding transition rate is formulated as below:

$$\lambda_{ij} = \frac{1}{E(T_{ij})}$$

The transition between any two distinct states without any observations has zero transition rate ($\lambda_{ij} = 0$). The formulation of the continuous time Markov process is completed once the transition rates matrix Λ is developed.

It is talked about a continuous function of time $p_{ij}(t)$ in the continuous case as opposed to the one-step p_{ij} or n-step $p_{ij}(n)$ conditional probabilities in the discrete case. Therefore, a transition probability function $p_{ij}(t)$ is defined as follows:

$$p_{ij}(t) = P(X(t) = j | X(0) = i)$$

Without knowing anything about intermediate transitions in between, a $p_{ij}(t)$ represents the conditional probability (between 0 and 1) that the process is in state j at time t known that it was in state i at time 0. Because the process cannot be in two distinct states at the same time, the initial conditions can be defined as follows:

$$p_{ij}(0) = 0 \text{ for } i \neq j \text{ and } p_{ii}(0) = 1 \text{ for } i = j$$

In matrix form, it is represented by $P(0) = I$, where I is the identity matrix. It is inferred that for any t the sum of the probability functions of every state j for a given initial state i must be 1:

$$\sum_{j} p_{ij}(t) = 1$$

For a stochastic process with K distinct states, the number of probability functions is K^2. These functions are obtained by the values in the transition rate matrix Λ, either analytically or numerically. The analytic probability functions are in the form of linear first-order differential equations (called Kolmogorov differential equations) as follows:

$$\frac{d}{dt} p_{ij}(t) = \sum_k p_{ik}(t) \lambda_{kj} \text{ or}$$

$$\frac{dP(t)}{dt} = P(t)\Lambda$$

Given the initial or boundary condition $P(0) = I$, $dP(t)/dt$ is the probability transition matrix with i, jth element $dp_{ij}(t)/dt$. The solution of the system of linear first-order differential equations is the set of probability functions based on the initial conditions. It can be noticed that the above equations are in the form $f'(t) = \lambda f(t)$ with constant coefficients λ_{ij}'s. Therefore, the solutions are the exponential functions.

Numerical Tests

The open train operations data of The Dutch railways covering weekdays between September 4, 2017 and December 8, 2017 are used to develop and test the model. Specifically, the total of 118 SPR (all stop local) trains travelling from Dordrecht (Ddr) to Den Haag Centraal (Gvc) on a 44.3-km line in approximately 56 minutes are considered. Train data was divided into training (70%) and test (30%) data set.

The entire trip of each SPR train is assumed to constitute two types of processes, namely sectional running and station dwelling. Accordingly, two separate transition rate matrices (Λ_R and Λ_D, respectively) were developed based on the state transition durations (T_{ij}) and numbers (n_{ij}) for trains in the training data set. In this extended abstract, only the sectional running between the control point (Rtst, 15.8 km) and station Rotterdam Zuid (Rtz, 16.9 km), and the station dwelling at Rotterdam Zuid. The complete model for the entire line will be developed in the full paper.

The observed delays are grouped into eight non-overlapping delay states between -1.5 and 6.5 minutes with 1-minute interval, disregarding the outliers (1.76% of all delay data set). The transition rate matrices and transition functions for initial delay state zero min are shown in Figure 1 and Figure 2 for the sectional running between Rtst and Rtz, and station dwelling at Rtz, respectively.

$$\Lambda_R = \begin{bmatrix} \mathbf{-0.8506} & 0.8506 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.9463 & \mathbf{-1.9979} & 0.8397 & 0 & 0 & 0.2119 & 0 & 0 \\ 0 & 1.0692 & \mathbf{-2.1523} & 0.6917 & 0.3914 & 0 & 0 & 0 \\ 0 & 0 & 1.1050 & \mathbf{-2.3716} & 0.6182 & 0.4175 & 0.2309 & 0 \\ 0 & 0 & 0 & 0.7246 & \mathbf{-1.6576} & 0.5954 & 0.3376 & 0 \\ 0 & 0 & 0 & 0 & 1.0526 & \mathbf{-2.1460} & 0.6767 & 0.4167 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{-0.5714} & 0.5714 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.0753 & \mathbf{-1.0753} \end{bmatrix}$$

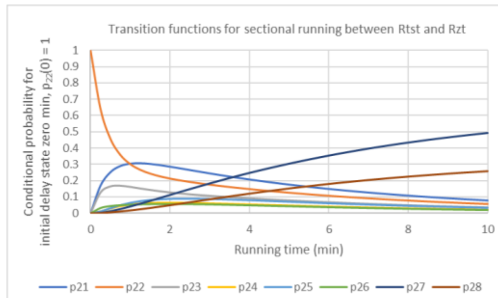

Figure 1. Transition rate matrix and functions between Rtst and Rtz sectional running

$$\Lambda_D = \begin{bmatrix} \mathbf{-0.6403} & 0.6403 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1.1712 & \mathbf{-1.9761} & 0.8049 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.1447 & \mathbf{-2.2215} & 0.7382 & 0.3386 & 0 & 0 & 0 \\ 0 & 0 & 1.1701 & \mathbf{-1.8699} & 0.6998 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.2022 & \mathbf{-1.9623} & 0.7601 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.1647 & \mathbf{-1.9557} & 0.7910 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.2658 & \mathbf{-2.0877} & 0.8219 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1.1111 & \mathbf{-1.1111} \end{bmatrix}$$

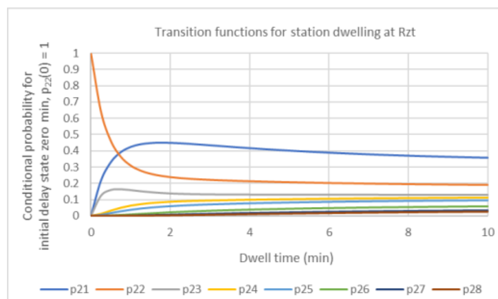

Figure 2. Transition rate matrix and functions at Rtz station dwelling

The transition functions can be used to estimate the long-term convergence to the state or steady-state delay. Figure 3 and Figure 4 show this convergence (partially in 10 min) for the sectional running between Rtst and Rtz and station dwelling at Rtz.

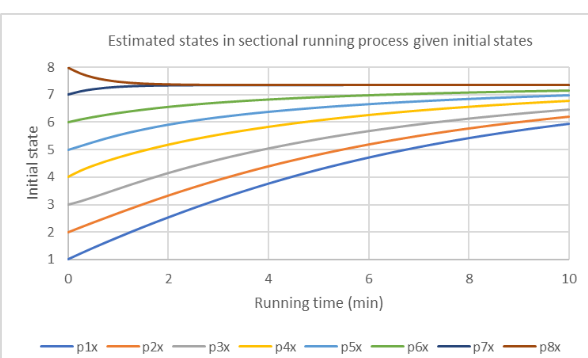

Figure 3. Estimated states in sectional running process

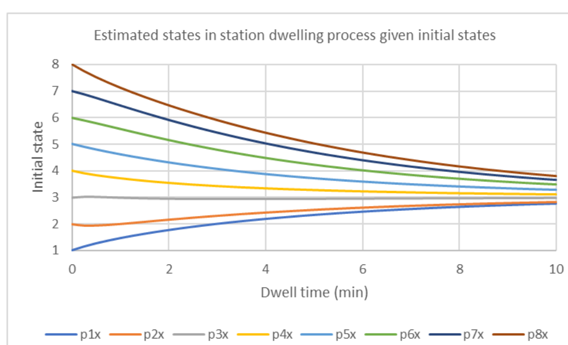

Figure 4. Estimated states in station dwelling process

Discussions and Conclusions

Once the transition rate matrices are developed for the sectional running and station dwelling processes of SPR trains, it is possible to calculate various measures such as delay prediction of trains at the next station or control point (Rtst and Rtz in this case). It is apparent that the characteristics of the two exampled sectional running and station dwelling transition functions are different as shown in Figure 1 and Figure 2. For instance, the on-time trains with initial zero delay ($p_{22}(0) = 1$), the proportion of on-time performance decreases fast approx. within the first two minutes in both processes. Thereafter, these trains continue deteriorating in the sectional running process (approx. only 6% on-time in 10 min) while keeping their on-time performance in dwelling process at a certain proportion (approx. 19% on-time in 10 min). Furthermore, the proportions of trains delayed 5 min or more ($p_{27}+p_{28}$) sums to 75% in the running process while the proportion of early trains (p_{21}) gets increased fast in the first two minutes of dwell and converges to approx. 36% in 10 min. Similar more outcomes can be drawn in the same manner. Because the prediction can be a function of the running or dwelling process time, the prediction horizon can cover multiple sectional running and station dwelling; i.e., longer prediction horizon.

It is apparent in Figure 3 and Figure 4 that the estimated states (weighted by the transition probabilities at a particular time) given the initial states for the sectional running and station dwelling processes of the respective section and station give different results, respectively. SPR trains running from Rtst to Rtz converge to approx. state 6 or 4 min delay up to 10 min sectional running time independent of the initial state. This implies that the time supplement in the running times of trains in this section (if any) is not effective to make up the initial delays (Figure 3). SPR trains dwelling at station Rtz converge to approx. slightly over state 3 or 1 min up to 10 min dwell time independent of the initial state. This implies that the buffer time added to the dwell time at this station is effective in reducing the initial delays (Figure 4).

The functional form of transition probabilities also allows to predict the delay at the end of a time horizon given the delay at the beginning of the horizon (i.e., initial delay) depending on the process type. For instance, the predicted delays at the end of the sectional running between Rtst and Rtz and at the end of station Rtz dwell both fit to the actual delays with the coefficient of determination value of 0.90.

The full model of the entire line is going to be presented in the full paper including comparisons. The characteristics of the Markov models including the continuous time Markov process are proved to be promising alternative for train delay analysis, evolution and prediction. It deserves more elaboration and the extended numerical tests.

Acknowledgement

We appreciate INFORMS Railway Applications Section (RAS) for sharing train operations data used to develop the models in this paper. This research is supported by the Coordinatorships of Scientific Research Projects, Yildiz Technical University under the doctoral dissertation grant with project # FDK-2024-6153.

References

Artan, M. Ş. and Şahin, İ., 2022. Exploring patterns of train delay evolution and timetable robustness. IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 11205–11214, doi: 10.1109/TITS.2021.3101530.

Tiong, K. Y., Mab, Z., Palmqvist, C.-W., 2023. A review of data-driven approaches to predict train delays. Transportation Research Part C 148, 104027.

Spanninger, T., Trivella, A., Büchel, B., Corman, F., 2022. A review of train delay prediction approaches. Journal of Rail Transport Planning and Management, Vol. 22, 100312.

Şahin, İ., 2017. Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances. Journal of Rail Transport Planning and Management, vol. 7, no. 3, pp. 101–113, doi: 10.1016/j.

Solberg, J. J., 2009. Modeling random processes for engineers and managers. John Wiley & Sons, Inc.

Nelson, B. L., 1995. Stochastic Modeling. McGraw Hills, Inc.