Analysis of Travel Time Distribution for Varying Length of Time Interval

Alireza Ganjkhanloo¹, Afshin Shariat-Mohaymany²*, Mojtaba Rajabi-Bahaabadi³, Amin Sayyad⁴

¹ M.Sc. Graduate, School of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran
² Associate Professor, School of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran
³ Ph.D. candidate, School of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran
⁴ M.Sc. Graduate, School of Civil Engineering, Islamic Azad University Tehran Science and Research Branch, Tehran, Iran

ABSTRACT: This study intends to determine the most appropriate distribution for modeling travel time variability. It also aims to explore the effects of the time of day and the length of the analysis time interval on the type of the best-fit probability distribution function. To this end, four analysis time intervals of different lengths ranging from five minutes to three hours are considered. Subsequently, for each analysis time interval, travel time data collected at different times of day are fitted to 12 common probability distribution functions. The Akaike Information Criterion is then used to evaluate the goodness of fitting and to rank the probability distribution functions. The results of this study indicate that the Gaussian mixture distributions are superior to single distributions to represent travel time distribution. In addition, single probability distribution functions can model the distribution of travel time observations when the length of the time interval is short. Among single probability distribution functions, the Burr distribution provides the best fit to the travel time data. The results of this research also show that the type of the best-fit probability distribution function does not change significantly over the time of day.

1. INTRODUCTION

The travel time is considered as one of the most important traffic information to travelers [1]. Accurate travel times help travelers to navigate to their destination more effectively and help traffic operators to make real-time decisions more appropriately [2]. It also plays a key role in travelers’ decisions [3]. As a result, travel time has been widely used as an important factor in project appraisal [4].

Travel time on a road is not constant. It varies from day-to-day and across different times of day [5]. Travel time is subject to random variations owing to demand fluctuations and supply uncertainty [6]. Travel demand uncertainty is mainly caused by day-to-day demand variation and travelers’ behavior uncertainty [7]. The supply uncertainty is due to different disturbances on the road (e.g., traffic incidents, work zone activities, adverse weather conditions) affecting the road capacity [8].

It is increasingly recognized that travelers consider not only travel time but also its variability in their choice situations such as mode, departure time and route choice to arrive on-time at their destinations [9]. Several studies such as Li et al. (2010) [10] and Asensio and Matas (2008) [11] reported that travelers are willing to pay not only for the reduction in travel time but also for the reduction in travel time variability.

Furthermore, travel time variability has important practical applications in transportation engineering. It is commonly considered as a major indicator of roadway performance and service quality [12]. Travel time variability also serves as a major factor in dynamic congestion pricing, where the time-varying toll is determined based on the traffic congestion level [13]. To ensure on-time arrival, the variability in travel time is also considered in the design of route guidance systems, transit networks and logistics systems [14, 15].

Owing to broad real applications of travel time variability, a considerable number of studies have been conducted to model travel time variability. Modeling travel time variability has been, and remains, a primary concern to researchers and practitioners in the field of transportation engineering. One of the most common approaches for modeling travel time variability is to describe travel time as a random variable. Modeling travel time as a random variable requires the determination of the probability distribution function (PDF) of travel time and estimation of its parameters.

To date, several studies have been conducted to determine the distribution of travel time using real data. For example, Herman and Lam (1974) [16] fitted several continuous probability distributions to real travel time data collected in Detroit. They suggested the Gamma distribution for representing travel time variability. Enam and Al-Deek...
(2006) [17] indicated that the lognormal distribution can be a reasonable representation of an empirical travel time distribution. The lognormal distribution was also recommended by Rakha et al. (2006) [18], Faouzi and Maurin (2007) [19], Lu and Dong (2018) [20], Rahman et al. (2018) [21], Chen et al. (2018) [22], and Rajabi-Bahaabadi et al. (2019) [23] for modeling travel time variability. WeiFeng et al. (2013) [24] collected travel times of several expressways in Shanghai, China. They found that travel time variability can be represented by a Beta distribution. Susilawati et al. (2013) [25] and Taylor (2017) [26] proposed the Burr distribution for modeling travel time variability. Recent studies showed that mixture distributions provide a better fit to travel time observations compared to single probability distributions [27–29]. For example, Guo et al. (2010) [27] fitted travel time data to several single and mixture distributions. They indicated that the lognormal mixture distribution provides the best fit to travel time data under moderate to congested traffic conditions. Ma et al. (2016) [28] found that the normal mixture models are superior to single probability distributions to model travel time variability. Recently, Yang and Wu (2016) [29] applied mixture distributions to model travel time variability. The results of their study showed that mixture models outperformed the single alternatives in terms of goodness of fit.

The above-mentioned studies on modeling the travel time distribution tend to yield inconsistent overall results. Sated more explicitly, different probability distributions have been proposed to model travel time variability. The difference between the types of proposed distributions may be due to several reasons such as time of day, road type and the length of the analysis time interval. To the best of our knowledge, no empirical study has been carried out to identify how the length of the analysis time interval might change the travel time distribution. Furthermore, few attempts have been made to examine the effect of the time of day on the travel time distribution. Accordingly, the present study intends to explore the effects of the time of day and the length of the analysis time interval on the type of the best-fit probability distribution function.

The remainder of the paper is organized as follows. In the next section, the methodology of the research is presented. Subsequently, the real travel time dataset is described in Section 3. The results of fitting travel time data to several common distributions are presented in Section 4. Finally, the paper is concluded with some suggestions for future research.

2. METHODOLOGY

The Statistical distributions can be divided into two general groups: 1) single probability distributions and 2) mixture probability distributions. Single probability distributions are unimodal distributions such as the normal distribution while mixture distributions are multimodal statistical distributions that are created from the convex combination of several single distributions. Suppose that $p_k(x)$ represents a single statistical distribution. The probability density function of a mixture distribution with $k$ single distribution is defined as

$$f(x) = \sum_{k=1}^{K} w_k p_k(x). \quad (1)$$

In the above equation, $w_k$ is a positive value representing the weight of $k$th distribution. As the mixture distribution is the convex combination of several single distributions, the sum of the weights is equal to one. In other words

$$\sum_{k=1}^{K} w_k = 1. \quad (2)$$

Based on the above equations, it can be concluded that a single distribution such as the normal distribution is a special case of a mixture distribution in which $k$ is equal to one.

In this study, 11 single distributions including birnbaum–saunders (BS), burr, rician, gamma, lognormal (LN), generalized extreme value (GEV), t-location-scale (TLS), nakagami (NK), normal and weibull are considered. The parameters of single distributions are estimated by the well-known maximum likelihood estimation (MLE) method. Furthermore, Gaussian mixture distribution is considered as a special case of mixture distributions. The Gaussian mixture distribution is composed of several normal distributions. To estimate parameters of the Gaussian mixture distribution, the expectation–maximization (EM) algorithm is employed [31].

To select the most appropriate distribution, the Akaike Information Criterion (AIC) is used. The AIC is a measure of the relative goodness of fit of a statistical model. It means that the change in the AIC value between two models estimated with the same data set can be used to identify the model that better fits the data [32]. The smaller the value of AIC, the better the model. The AIC is defined as follows:

$$AIC = 2n - 2LL, \quad (3)$$

where $n$ is the number of estimated parameters in the model and $LL$ represents the maximized log-likelihood for the estimated model.

As the AIC is only a measure of the relative goodness of fit, the Kolmogorov–Smirnov test (K-S test) is used to test whether travel time observations come from a hypothesized distribution. If the $p$-value of the test if greater than 0.01, then it can be concluded that travel time observations come from the hypothesized distribution at 99 % confidence level.

As mentioned earlier, the Statistical distributions can be divided into two general groups: 1) single probability distributions and 2) mixture probability distributions. Single distributions are unimodal while mixture distributions are multimodal. If the distribution of travel time observations is unimodal, then single distributions are more likely to fit travel time observations. To test the unimodality of the distribution of travel time observations, Hartigan’s dip test [33] was used. The null hypothesis $H_0$ for the dip test is that the distribution is unimodal. If the $p$-value is greater than 0.01, the null hypothesis is not rejected at 99 % confidence level.
3. DESCRIPTION OF THE STUDY AREA

The case study site is a freeway section that connects Karaj to Tehran, two major cities in Iran. The length of the section is about 20 km. The freeway is a major commuting corridor in Iran. Its average annual daily traffic (AADT) is approximately 100,000 vehicles. It is instrumented with Bluetooth sensors that can record time stamps and the media access control (MAC) addresses of Bluetooth-enabled devices. Vehicle travel time can be estimated by matching the MAC addresses recorded by the sensors. In this paper, a 1-min aggregation level was used for travel times. Travel time data used in this study were collected from January 1, 2016 to June 21, 2017. The workday data from 5:00 AM to 11:00 AM were used because traffic patterns are similar during workdays.

4. RESULTS

The main purpose of this section is to determine the most appropriate distributions for modeling travel time variability at different times of the day. As mentioned earlier, the second purpose of the study is to examine how the length of the analysis time interval may affect the type of travel time distribution. In this regard, in order to obtain a better view of the travel time data, a descriptive analysis of the data is first provided. Subsequently, four analysis time intervals including 1) three-hour interval, 2) one-hour interval 3) fifteen-min interval and 4) five-min interval are considered. At different times of day and for each analysis time interval, twelve probability distribution functions are initially fitted to travel time data. Finally, the candidate distributions are then ranked based on the AIC values.

Fig. 1 shows a box plot of the travel time data at different hours during the study period (5:00-11:00 AM). From Fig. 1, it can be seen that the mean and the standard deviation of travel time at different hours are not constant. In this study, the Kruskal-Wallis test [34] was used to examine the hypothesis of equality of average travel times. Also, Levene's test [35] was used to examine the hypothesis of equality of standard deviations. The results of these two tests show that at 95% confidence level, the means of travel times, as well as travel time variances, are not equal at different intervals. This implies the stochastic and time-dependent nature of travel time. Furthermore, according to Fig. 1, the peak period occurs during 7:00-8:00 AM. This may be due to the fact that commuters depart their origin (Karaj) for a work trip over this period.

In order to find the most appropriate distribution for modeling travel time variability, the study period (5:00 AM to 11:00 AM) was divided into several analysis time intervals with varying length. More explicitly, the study period is divided into two analysis intervals with the length of 3 hours, six 1-hour intervals, twenty-four 15-minute intervals and seventy-two 5-minute intervals. Then, for each analysis time interval, candidate distributions were fitted to travel time data collected during each interval. The value of the Akaike Information criterion was then used to rank probability distributions. The smaller the value of AIC, the better the fit of the distribution to the data. Therefore, for each time interval, the distribution with the smallest value of AIC is given the rank 1, the one with the second smallest value of AIC is given the rank 2, and so on up to the distribution with the largest value of AIC, which is assigned the rank 12. Fig. 2 illustrates the average rank of the candidate probability distributions for three-hour, one-hour, fifteen-min and five-min intervals. As can be seen in Fig. 2, on average, the Gaussian mixture distribution has the best fit to the travel time data for all analysis time intervals. Among single distributions, the burr distribution has the best fit to travel time data.

To answer the question whether travel time observations come from the fitted distributions, the Kolmogorov-Smirnov test (K-S test) was used. Here, the null hypothesis is that the data comes from a given distribution. In contrast, the alternative hypothesis is that the data does not follow a given distribution. In this study, the null hypothesis is accepted if the p-value is greater than 0.01. For each interval length, Table
Table 1. The percentage of intervals that fitted distributions passed the K-S test

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Interval length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 min</td>
</tr>
<tr>
<td>BS</td>
<td>12.5%</td>
</tr>
<tr>
<td>Burr</td>
<td>37.5%</td>
</tr>
<tr>
<td>Gamma</td>
<td>15.3%</td>
</tr>
<tr>
<td>GEV</td>
<td>18.1%</td>
</tr>
<tr>
<td>Logistic</td>
<td>12.5%</td>
</tr>
<tr>
<td>Lognormal</td>
<td>22.2%</td>
</tr>
<tr>
<td>Nakagami</td>
<td>11.1%</td>
</tr>
<tr>
<td>Normal</td>
<td>12.5%</td>
</tr>
<tr>
<td>Rician</td>
<td>12.5%</td>
</tr>
<tr>
<td>TLS</td>
<td>26.45</td>
</tr>
<tr>
<td>Weibull</td>
<td>0%</td>
</tr>
<tr>
<td>Mixture</td>
<td>98.6%</td>
</tr>
</tbody>
</table>

Table 2. The results of Hartigan’s dip test

<table>
<thead>
<tr>
<th>Parameter</th>
<th>5-min</th>
<th>15-min</th>
<th>1-hour</th>
<th>3-hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptance of H1 (%)</td>
<td>52.7%</td>
<td>20.8%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Fig. 2. The average rank of probability distributions for different analysis time intervals

Fig. 3 compares the empirical distribution of travel time data at one-hour intervals with the estimated mixture distributions.

Results of this test confirm that the distributions of travel time observations at all three-hour and all one-hour intervals are multimodal at 99% confidence level. For 15-min intervals, the travel time distribution is unimodal in 5 out of 24 intervals (20.8%). Furthermore, the proportion of unimodality cases is 52.7% for 5-min intervals. According to the results of the Hartigan’s dip test, it can be concluded that travel time distributions are usually multimodal; however, as the length of the analysis time interval becomes shorter, travel time distributions are more likely to be unimodal. Therefore, single probability distributions (unimodal distributions) such as the Burr distribution can model the distribution of travel time observations only when the length of time interval is short.
As shown in Fig. 3, travel time distributions at each of the six analysis intervals were best described by Gaussian mixture distributions because they are able to model multimodality of travel time data by combining several probability distributions. As mentioned earlier in the introduction section, this study also attempts to answer the question whether the type of best fitting distribution to travel time data changes at different times of the day. To this end, several probability distributions were fitted to travel time data at different times of the day. The results of our analysis showed that the type of the best-fit probability distribution does not change at different times of the day. As an example, Fig. 4 shows the rank of each probability distribution at different times of the day. The length of the analysis time interval is equal to 15 min in Fig. 4. As can be seen in Fig. 4, the Gaussian mixture distribution has the first rank at all 15-min intervals of the study period. Furthermore, at most times of day, the burr distribution has the second and the third rank.
5. DISCUSSION AND PRACTICAL APPLICATIONS

Transportation systems such as dial-a-ride and delivery systems characterized by unreliable services lose their customers over time because such systems impose significant penalty costs to their customers. Several attempts have been made over the last two decades to provide customers and travelers with reliable services and to incorporate travel time variability into the design of transportation systems. The determination of travel time distribution is the first step for considering travel time variability. Although mixture distributions are leading candidates for representing travel time variability, almost all previous studies considered that travel time follows a single distribution to incorporate travel time variability into the design of transportation systems. This is due to the fact that some single probability distributions enjoy desirable properties that make incorporating travel time variability into the design of transportation systems easier and tractable. As a result, for designing transportation systems under travel time uncertainty, this paper suggests that the length of time intervals should be short (e.g., 5 minutes) because single distributions fail to represent travel time distributions for longer time intervals.

As mentioned in the introduction section, some previous studies proposed single distributions for representing travel time distribution. However, according to the results of this study, single distributions such as the Burr distribution may not fit accurately to travel time observations for long time intervals; this may bias the estimation of travel time variability measures. As a result, for long intervals, it is recommended to use mixture probability distributions for the analysis of travel time variability.

6. CONCLUSION

Modeling travel time distribution is a preliminary step for considering travel time variability. In this paper, the effectiveness of a wide range of probability distributions for modeling travel time variability was investigated. Furthermore, the effects of time of day and the length of analysis interval on the distribution of travel time were studied.

The results of this study showed that distributions of travel time observations at three-hour and one-hour intervals are multimodal at 99% confidence level. The proportion of multimodality cases is 47.3% and 79.2% for 5-min and 15-min intervals, respectively. As the travel time distribution is usually multimodal, single probability distributions fail to describe travel time variability especially for long time-intervals. Furthermore, based on the results of this study, the Gaussian mixture distribution provides a reasonable representation of observed travel time distributions. More specifically, the Gaussian mixture distribution passed the K-S test in 98.6%, 83.3%, 83.3% and 50% of cases when the length of time interval is five minutes, fifteen minutes, one hour and three hours, respectively. It worth mentioning that the Burr distribution had the best fit to travel time data among single distributions. Finally, the results of this study showed that the best-fit probability distribution function does not change significantly over the time of day.

In this study, the fitting performance of the Gaussian mixture distribution, which formed from several normally distributed components, to travel time observations was examined. For future research, it is recommended to examine the ability of additional mixture models with non-normal components. Furthermore, future research can focus on testing the ability of additional single probability distributions for modeling the travel time distribution.

REFERENCES

[16] R. Herman, T. Lam, Trip time characteristics of journey to and from work, Transportation and traffic theory, 6 (1974) 57-85.
[18] H.A. Rakha, I.E.-. Shavarby, M. Arafah, F. Dion, Estimating Path Travel-


