

Estimating Suitable Probability Distribution Function of Risk for Different Traffic States

Qian Li

Institute of systems science, School of Traffic& Transportation, Beijing Jiaotong University, Beijing, 100044 China

Email:14120759@bjtu.edu.cn

Key words: Time to collision; MLE; Normal distribution; Generalized extreme value distribution; Chi-squared test

Abstract: Generally, time to collision (TTC) is used to quantify traffic risk. In this paper, we aimed at studying suitable probability distribution of TTC under different traffic states. By using maximum likelihood estimation(MLE), the parameters of different states are found. Normal distribution and generalized extreme value distribution are used to assume probability distribution function, and chi-squared test is used to test the results. It is found that the normal distribution is more suitable for the data.

Introduction

With the increasing car-ownership, scientists have paid more attention to forecast the risk of avoiding accidents[1~3]. Over the past decade, numerous crash prediction models have been developed. Ma et al. offered a multivariate Poisson-lognormal method which can simulate crash counts from the perspective of injury severity[4]. Pirdavani et al. combined crash counts with character of traffic flow, and set up a real-time crash risk prediction model to predict the probability of accidents[5]. Xu et al. tried to predict crash likelihood by analyzing the severity on free ways, and a logit model was developed. And Chen et al. combined the multinomial logit model with the Bayesian network in rear-end crashes[6]. Most researchers are focusing on the crash counts, and most of the researches are related to the predictions which are after the collision accidents. But if there were no accidents happened, drivers just depended on the traffic states to change behaviors to avoid risk was not mentioned. But there's one thing that is not mentioned which is if there are no collision accidents happened, what should the drivers do to avoid risks according the different traffic states. So what the drivers should do at different states to avoid accidents is very important.

Study sites and data

This study employs vehicle trajectory data available from FHWA's Next Generation Simulation (NGSIM) program [7]. These data were extracted from the northbound traffic on I-80 in Emeryville, California. The length of I-80 is approximately 500 m, as shown in Fig. 1. On this study segment, vehicle positions were recorded every 0.1 s from 5:15 p.m. to 5:30 p.m. on April 13 2005. Speed is estimated based on vehicle positions every 0.1 s. Lane numbers are increased from left to right.

Methodology

Time to collision(TTC) is one of the indicators that are used to estimate the safety of traffic situations[8~11].It indicates the time span left before two vehicles collide, if nobody takes evasive action. In these researches, TTC has often been used as a safety indicator for traffic risk by finding the minimum TTC[12].

TTC is computed as Eq.1:

$$TTC = \frac{D}{V_l - V_f}, \forall V_l > V_f \quad (1)$$

Where V_l is the speed of leading vehicle, V_f is the speed of following vehicle, D is the spacing

between the leader and follower.

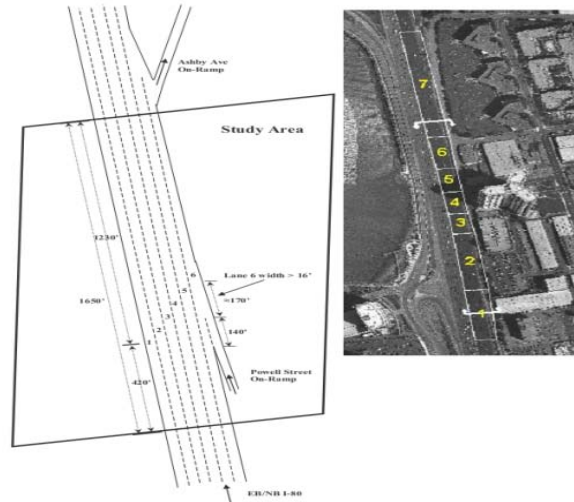


Fig.1. the distraction of I-80

As the Eq.1 shown, only when the speed of following vehicle is higher than the leading vehicle one, the TTC is defined. And smaller TTC values indicate higher accident risks and vice versa.

As known, different traffic states have different risk levels. And it is very important to fit the distributions. In this paper, we divide spacing into different 10 states, which can represent different traffic states, and try to find the most appropriate distribution.

There are various types of probability distribution functions may fit the NGSIM data. According to the shape of the risk indicator's histogram, we choose the normal distribution and generalized extreme value distribution as the candidates. The Eq. 2 and Eq. 3 show the probability density function(pdf) for these two distributions.

Generalized extreme distribution is given as bellows,

$$f_{gev}(x|\mu, \sigma, \kappa) = \frac{1}{\sigma} g(x)^{\kappa+1} e^{-g(x)}$$

$$g(x) = \begin{cases} \left(1 + \left(\frac{x-\mu}{\sigma}\right)\kappa\right)^{\frac{1}{\kappa}} & \kappa \neq 0 \\ e^{\frac{x-\mu}{\sigma}} & \kappa = 0 \end{cases} \quad (2)$$

Where κ means the shape parameter of pdf, μ means the location parameter of pdf, and σ means the scale parameter of pdf.

Normal distribution is given as bellows,

$$f_{norm}(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

Where μ and σ stand for the mean and standard deviation of normal distribution.

In order to estimate the parameters of the distributions, the maximum likelihood estimation(MLE) is applied. To test goodness of fit, we select chi-squared(χ^2) test which is the most popular test. The spacing data which to be estimated is $X_i (i = 1, 2, \dots, n)$, n is the total number of test data. First, divide the range of estimated X_i into k interval, i.e., $[a_0, a_1][a_1, a_2) \dots [a_{k-1}, a_k)$, and then calculate each value, $N_j (j = 1, 2, \dots, k)$ for each interval. Where N_j means the number of X_i at j interval, the test statistics is drawn as Eq. 4:

$$\chi^2 = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j} \quad (4)$$

According to the results of χ^2 test, we find that normal distribution is the best model.

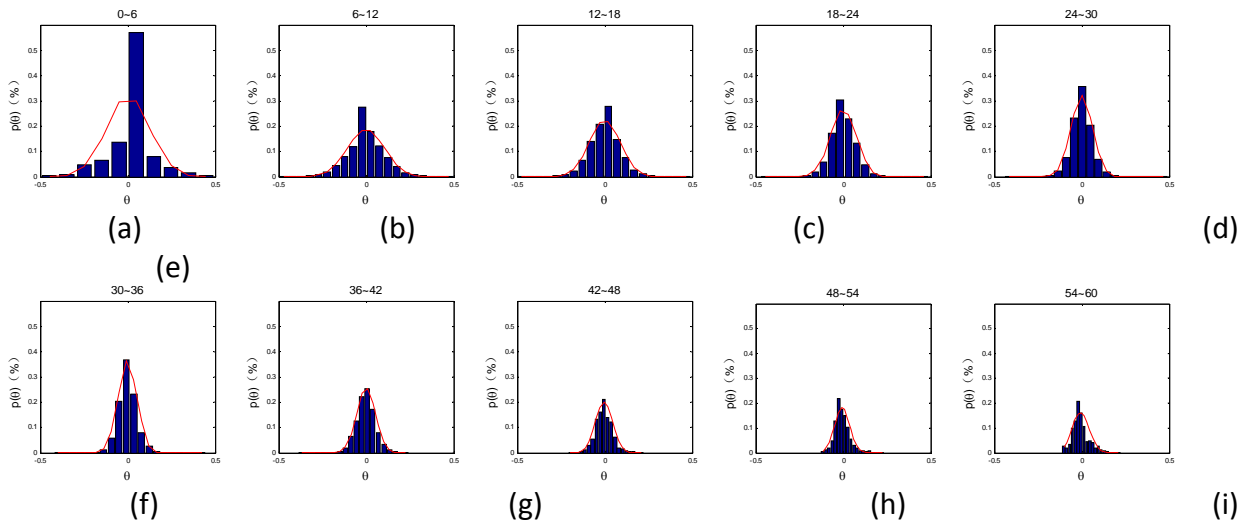


Fig.2. normal distribution

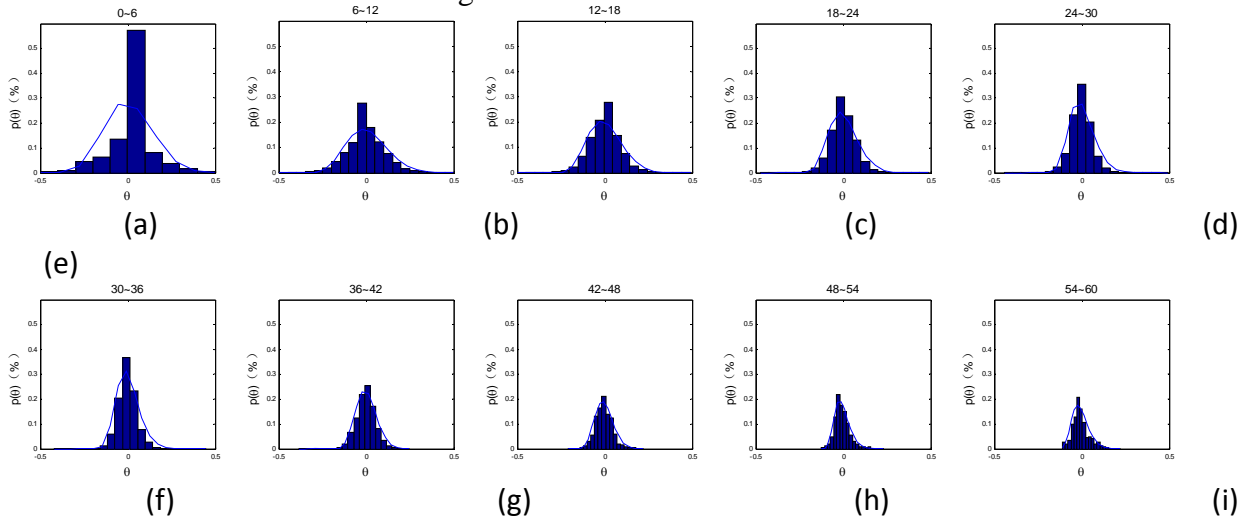


Fig.3. generalized extreme value distribution

Results

The results of the distribution fitting are as Fig.2 and Fig.3. Fig.2 shows normal distribution, and Fig.3 shows generalized extreme value distribution, and the χ^2 test is listed in table1. As the Fig.2 and Fig.3 shown, the θ denotes the TTC, and $p(\theta)$ denotes the probability of TTC. The shape of different traffic states looks different. And the θ near the number of 0 has a very large number, it shows that the following vehicle is trying best to contain the same speed with leading one. With the increase of spacing, the percentage of θ at 0 is getting smaller and smaller. According to the comparison of Fig.2 and Fig.3, the fitting effect of distributions is better when distance is greater.

As the table1 shown, the normal distribution is better fit than generalized extreme value distribution. But at the ninth and tenth states, the result of χ^2 test of normal is higher, it shows that when the spacing is very large, the distribution of θ is more like the generalize extreme value.

Table 1 the results of χ^2 test

distribution 10 states	generalized extreme value	normal
	1	0.6239
2	0.0819	0.056
3	0.061	7 0.021
4	0.0473	2 0.013
5	0.0692	3 0.015
6	0.0698	1 0.024
7	0.0268	0.010
8	0.0489	3 0.041
9	0.0758	1 0.582
10	0.0089	1 0.145
		8

Conclusions

Different traffic states have different risk. And TTC is a widely used indicator to quantify the risk. In this study, we divided the spacing into 10 parts, and every part represents a type of traffic state. And we found the normal distribution is the best fitted distribution. Normal distribution is broadly used to assume probability distribution function. But when the distance is large, the generalized extreme value is much better.

References

- [1]Othman S, Thomson R, Lannér G:Safety analysis of horizontal curves using real traffic data[J]. Journal of transportation engineering, 140(4): 04014005. (2013)
- [2]Chen Z, Wu C, Zhong M, et al:Identification of common features of vehicle motion under drowsy/distracted driving: A case study in Wuhan, China[J]. Accident Analysis & Prevention, 81:251-259.(2015).
- [3]Qu W, Ge Y, Xiong Y, et al:The relationship between mind wandering and dangerous driving behavior among Chinese drivers[J]. Safety Science, 78: 41-48. (2015)
- [4]Pirdavani A, De Pauw E, Brijs T, et al:Application of a rule-based approach in real-time crash risk prediction model development using loop detector data[J]. Traffic injury prevention, (2015). <http://dx.doi.org/10.1080/15389588.2015.1017572>
- [5]Xu C, Tarko A P, Wang W, et al:Predicting crash likelihood and severity on freeways with real-time loop detector data[J]. Accident Analysis & Prevention, 57: 30-39. (2013)
- [6]Chen C, Zhang G, Tarefder R, et al:A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes[J]. Accident Analysis & Prevention, ,80: 76-88. (2015)
- [7]FHWA, 2008.The Next Generation Simulation (NGSIM).<<http://www.ngsim.fhwa.dot.gov/>> (05.08.08).

- [8]Li Y, Bai Y:Effectiveness of temporary traffic control measures in highway work zones. *SafetyScience* 47:453–458. (2009)
- [9]Meng Q, Qu X:Estimation of rear-end vehicle crash frequencies in urban road tunnels. *AccidentAnalysis and Prevention* 48: 254–263. (2012)
- [10]Lu G, Cheng B, Lin Q, Wang Y:Quantitative indicator of homeostatic risk perception in car following. *Safety Science* 50:1898–1905. (2012)
- [12]vanWinsum, W., Heino, A:Choice of time-headway incar-following and the role of time-to-collision information in braking. *Ergonomics* 39 (4), 579–592. (1996)