Estimation and Analysis of Minimum Traveling Distance in Self-driving Vehicle to Prove Their Safety on Road Test

Ruxin Li
Traffic management engineering, People’s Public Security University of China, Beijing, 102623, China

Abstract. Self-driving cars as a new technology need time to test their safety. How to supervise the burgeoning autonomous vehicles is a new challenge for the countries that support this technology. How to effectively manage the development of the automatic driving and ensure traffic safety will certainly be one of the important tasks on traffic management in the future. At present, there is no complete assessment system for the evaluation of the standard and safety performance of autonomous vehicles. Under these circumstances, what kind of requirements does the self-driving vehicles on road test need to achieve can be authorized to formal use for the public is a serious question now. This paper uses the method of mathematical statistics to estimate the lower limit of how many distances the Chinese autonomous vehicles need to drive on the road test to provide relevant reference for the traffic management department.

1. Introduction
According to the statistics of China’s traffic accident statistics, about 60,000 people were killed in traffic accidents every year in China. And almost 220,000 people were injured in traffic accidents. The property and other losses were resulted from the accident are also difficult to calculate. There were more data indicates that 90% of traffic accidents occur due to human error [1], and the development of self-driving vehicles may be greatly reduced the accident that were caused by operational errors, but autonomous driving technology is still in its infancy now, and the management of autonomous driving is also imperative.

Domestic standards and specifications for autonomous vehicles have not yet been formed[2]. Meanwhile the safety of this new technology is still under study and discussion. Although Chinese scholars have done lots of research and discussion on the safety aspects of autonomous vehicles, but most of them were evaluation and experimental design based on experimental scenarios. So it is a controversial problem to testify the autonomous vehicles is practical and reliable to driving for everyone through driving test or driving assessment. Whether the self-driving score obtained by the experimental driving evaluation has obtained the clear intelligent level of the self-driving vehicles, and whether it represents the standard level of the road that can be used and put into use needs pondering. Analysis a large of data through the road test is a more reliable and intuitive method to measure it’s safe. But the self-driving vehicles has just started the road tests, how many numbers of kilometers required for automatic driving on the road can meet the safety under certain degrees. This paper uses interval estimation as a statistical method to estimate the minimum driving kilometers required under the certain confidence and certain overall effect.

2. Statistical method
In the research on the safety of self-driving cars, the object should be the whole of the entire autonomous vehicle. Under the certain accident rate and mortality, our work is to find out how many kilometers the whole automatic vehicles need to travel to meet the test requirements of the overall self-driving vehicle. The purpose of statistical analysis is to infer the overall distribution from the sample distribution, and now we have to calculate how much sample size is needed at least.

The basic problems of statistical inference can be roughly divided into two categories: hypothesis testing and parameter estimation[3]. Hypothesis testing is made the assumptions based on the sample: whether it is accepted or rejected. In order to infer certain properties of the population, some assumptions about the population are proposed where the overall distribution function is completely unknown or only knows its form and its parameters are not known. For example, the assumption that the overall obeys the Poisson distribution is proposed. Parameter estimation estimate the overall parameters using the sample data, which can be divided into two types: point estimation and interval estimation. The point estimate is to use the sample data to estimate the value of the population parameter, in other words, to use the sample data to calculate a single estimate value. Interval estimation is that determine the interval in which the overall parameters are located using the sample data. Meanwhile it ensures that the overall parameters are within this interval with a certain probability. It can be seen that the point estimate is only an approximation of the unknown parameter. But it even an accuracy or error range of this approximation were not given. So that is the defect of the point estimate. And we use the interval estimate for the statistical work to cover the defect of the point estimate.

The minimum sample size estimation of driving test is that seeking the correspondence between sample action probability and overall action probability based on mathematical statistics. Finally it can estimate the minimum sample size required to meet certain confidence and certain overall action probability. We are most concerned about 95% of which The above overall probability of action. Obviously, it is more appropriate to use the statistical inference method of interval estimation to carry out the minimum sample size estimation analysis of the autonomous vehicle test.

3. No-failure rate minimum mileage estimation

3.1 Analysis process

We regard the failure rate of self-driving vehicles as the failure rate of products in industrial production. The probability of accidents occurring every 1km of self-driving vehicles is expressed by the mortality rate. Generally, it is necessary to estimate the lower bound of the probability confidence, so we can get the equation \( P(p \leq p') \geq C \), where \( P \) is the probability of the event \( p \leq p' \), for \( p \) is the probability of the overall action, \( p' \) is the is the lower bound of the overall action probability, and \( C \) is the confidence[4]. This formula indicates that the overall probability that the overall probability of action \( p \) is not less than \( p' \) is not lower than \( C \).

We express the failure rate of the vehicle with the fatalities rate \( F \) per kilometers, then the reliability \( p \) is \( 1-F \)[5]. It is can be obtained by the distribution law of the binomial distribution[6]. We can get the minimum driving kilometers with overall probability \( p \) and confidence \( C \)(see Eq.(1)).

\[
\sum_{i=0}^{n} C_n^i p^{n-i} (1-p)^i = 1 - C \tag{1}
\]

In the equation that has given above, \( n \) is the total number of miles tested, \( C \) is the confidence level, \( p \) is the reliability of the product, and \( f \) is the number of deaths during the test. The equation (1) is converted to equation (2) where the question that how many kilometers of self-driving vehicle during the test needed to travel without any casualties can prove the reliability of today's ordinary vehicles under no failure accident:

\[
C = 1 - p^n \tag{2}
\]

And then we can converted equation (2) into equation (3):
3.2 Million kilometre mortality rate of a vehicle

The safety of human drivers is a key indicator to compare the safety of autonomous vehicles. The number of crashes, injuries and deaths caused by human drivers is huge in China each year, but the incidence of these accidents is relatively low compared to the kilometers traveled by human. Internationally, the death rate of traffic accidents is usually expressed in terms of the mortality rate of 100 million kilometers. For example, the mortality rates of traffic accidents in the United States, the former Federal Republic of Germany, and France in the 1980s were 2.21, 3.8, and 4.6 respectively[3].

Based on the mortality rate of millions of vehicle-kilometers on the highways in China today, We can calculate how many kilometers the autonomous vehicles need to test where we considered that it’s safety have ran up to the safety of today's ordinary vehicles. We use the million-kilometer mortality rate as an indicator to evaluate the driving reliability of the vehicle, that is $R = p = 1 - F$. According to the statistics of China's traffic accident statistics and the statistical bulletin of the development of the transportation industry, the mortality rate of millions of kilometers in China's highways from 2008 to 2015 can be obtained, as shown in Table 1 below:

| Year  | Kilometers one year people drive (10^4 km) | Accident (10^3 crashes) | Injured (10^3 injured) | Mortality rate per million vehicle kilometers |
|-------|-------------------------------------------|--------------------------|------------------------|----------------------------------------------|
| 2015  | 37019030                                  | 82                       | 54                     | 1.48%                                        |
| 2014  | 43761310                                  | 85                       | 57                     | 1.30%                                        |
| 2013  | 83736840                                  | 87                       | 58                     | 0.69%                                        |
| 2012  | 74721705                                  | 89                       | 61                     | 0.82%                                        |
| 2011  | 42973640                                  | 96                       | 64                     | 1.49%                                        |
| 2010  | 38451290                                  | 97                       | 63                     | 1.64%                                        |
| 2009  | 39990415                                  | 91                       | 60                     | 1.50%                                        |
| 2008  | 35811245                                  | 108                      | 60                     | 1.68%                                        |

3.3 Million kilometre the autonomous vehicles needed to test with no failure

Therefore, the average mortality rate is 1.325 per 100 million kilometers from the number of the mortality rate of millions of kilometers in China from 2008 to 15 years. We use the average calculation value as the failure rate of the current vehicle’s safety on the road in China now.

As we all know, confidence and confidence intervals are interrelated. The higher the confidence, the wider the confidence interval. Conversely, the lower the confidence, the narrower the confidence interval. Therefore, the choice of confidence must base on the principle of small probability (the principle of the actual impossibility of small probability events) and the actual needs of the professional field, that is, the level of confidence should be appropriate. In the field of mathematical statistical analysis, the confidence is usually 0.99, 0.95 or even 0.99[3], and the typical value is 0.95. In the reliability estimation of success or failure products, the confidence is usually taken as 0.90. But the degree is usually taken as 0.8 which is believed in the fields of engineering design and medicine. In this paper, the following results are calculated according to different confidence levels in equation:

$$n = \frac{\ln(1 - C)}{\ln(p)}$$

(3)
(3), as shown in Table 2 below:

| Confidence | Failure rate | Reliability | Million kilometre the autonomous vehicles needed to test |
|------------|--------------|-------------|------------------------------------------------------|
| C=80%      | 1.325*10^-6%| 99.999998675% | 121467011.6                                          |
| C=90%      | 1.325*10^-6%| 99.999998675% | 173780006                                             |
| C=95%      | 1.325*10^-6%| 99.999998675% | 226093000.5                                           |
| C=99%      | 1.325*10^-6%| 99.999998675% | 347560012.1                                           |

It can be seen from the table that at 90% confidence, the number of test kilometers in a fault-free situation where autonomously driven vehicles need to travel nearly 200 million kilometers in that we statistically believe the ability that autonomous vehicles have had reached the level which that the ordinary vehicles human drive.

4. Kilometre the autonomous vehicles needed to test with certain failure rate

4.1 Analysis process

The Poisson distribution can be used as an approximation of the binomial distribution with the parameter \( \lambda \) where \( n \) of the binomial distribution is huge and \( p \) is small[7], which means \( x \) deaths occur in a given \( n \) km. And we can know that \( \hat{\lambda} = \frac{x}{n} \), the confidence interval of \( 1 - \alpha \) is

\[
\left( \frac{x - z_{\alpha/2} \sqrt{x}}{n}, \frac{x + z_{\alpha/2} \sqrt{x}}{n} \right),
\]

so the half of the confidence interval is \( \frac{z_{\alpha/2} \sqrt{x}}{n} \), we can interpreted into the equation (4) in different degree of precision \( \delta \):

\[
1 - \delta = \frac{n \sqrt{x}}{\hat{\lambda}}.
\]

The failure rate is estimated as \( \hat{\lambda} = \frac{x}{n} \), and we can converted the equation (4) into the equation (5)

\[
x = \left( \frac{z_{\alpha/2}}{1 - \delta} \right)^2.
\]

We finally can calculate the kilometre the autonomous vehicles needed to test according the data that we count in section 3.2.

That is

\[
n = \frac{x}{F}.
\]

Therefore, the above formula can be used to calculate kilometre of the self-driving car needed to travel in different precision \( \delta \) at the current level. We calculate according to the 95% confidence interval, as shown in Table 3 below:

| Precision \( \delta \) | Failure rate | Kilometre the autonomous vehicles needed to test |
|-------------------------|-------------|-----------------------------------------------|
|                         |             |                                               |

Table 3. Kilometre the autonomous vehicles needed to test at least in different precision \( \delta \)
δ=80%  1.325*10^(-6)%  7248301886
δ=90%  1.325*10^(-6)%  28993207547
δ=95%  1.325*10^(-6)%  115972830188

It can be seen from the calculation that in the range of 80% accuracy and 95% confidence interval, the self-driving car test mileage needs to reach at least 7.2 billion kilometers. And the higher the accuracy, the greater the mileage the autonomous vehicles need to test. The test will take decades or even longer to complete.

5. Conclusion
In this paper, the confidence interval of the distribution parameters is derived by the statistical analysis theory. The statistical inference method of the interval estimation is used to estimate the lower limit of how many distances the driving test of the autonomous vehicle in China needs to travel, and the minimum test under the no-failure rate and the certain failure rate. The calculations show that the development of the China's autonomous driving still needs a long way to go. Just take the driving testing for example, the autonomous vehicles still needs billions of kilometers of it to achieve a certain level of safety in a statistically significant manner. However, for the selection of indicators, this paper only selects the death rate of millions of car kilometers, the index is still a single, and there is a need for further in-depth research and analysis.

References
[1] Nynke E. Vellinga (2017) From the testing to the deployment of self-driving cars: Legal challenges to policymakers on the road ahead. Computer Law & Security Review, The International Journal of Technology Law and Practice: 847-863.
[2] Weng, Y. X, Dominic H (2014) The road of intelligent car: the regulatory research of smart car, self-driving vehicles safety (04): 632-655.
[3] Sheng Z, Xie S. Q, Pan S. Y (1989) Probability and mathematical statistics. Higher education press, Beijing.
[4] Li K, Ma H. K, Zhou H. J (2009) Estimating analysis of least sample of microwave effect experiment. High power laser and particle beams, 21(04): 541-544
[5] Zhao Y, Yang J, Ma X. B (2009) Reliability data analysis tutorial. Beihang University Press, Beijing.
[6] Nidhi K, Susan M. Paddock (2016) Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability. Transportation Research Part A 94: 182-193.
[7] Meng Z. W (1995) Confidence interval of parameter of bernoulli distribution. Engineering Science Mathematics 14(11): 169-171.
[8] Li H (2010) Research on estimation of parameter in binomial distribution. Acta mathematicae applicatae sinica 33(03): 385-394.