

A Bayesian-based study on the average waiting time of airport taxis

Chen Zhang

School of Mathematics and Statistics, Jiangxi Normal University, Nanchang, China

1411247556@qq.com

Abstract. In recent years, the incompatibility of airport taxis with passenger time, scale and other factors has led to the problem that airport taxi traffic is difficult to control. This topic starts from the perspective of the number of taxis in Zhengzhou Airport, uses Bayesian statistics and R implementation, applies Bayesian calculation to the analysis of the average waiting time of airport taxis, and establishes the Poisson distribution model of the actual airport taxi traffic flow and Exponential distribution model for waiting time. This topic clearly presents the influence of factors such as the number of passengers and the number of taxis on the average waiting time of taxis. The visual description of the airport taxi waiting time is realized by using R language. By analyzing the model from multiple angles, the model is solved and tested with data, and the model is evaluated and improved after the end. The results show that the taxi traffic flow Indeed, it obeys the Poisson distribution model with η as the parameter, and the average taxi waiting time obeys the exponential distribution model with λ as the parameter, and the posterior distributions of the two are gamma distribution and inverse gamma distribution respectively, which is the city taxi. Management provides basic theoretical tools.

Keywords: Poisson distribution; Exponential distribution; Determination of prior distribution; Estimation of hyperparameters.

1. Introduction

Mathematics is the foundation of all science and technology, and it is a science that studies the quantitative relationship and spatial form of the real world. With the development of society, the emergence and continuous improvement of electronic computers, mathematics is not only used in various fields of various disciplines of natural science, but also penetrates into various fields of economic management, social science and social activities. As we all know, to use mathematics to solve practical problems, we must first establish a mathematical model, and then we can analyze, calculate and study the actual problem on the basis of the model. Mathematical modeling activity is the whole process of discussing and solving practical problems, and it is a mathematical way of thinking. After the in-depth understanding of the Bayesian formula, the model suitable for the production and life of the formula after the Bayesian generalization will be broader.

In this paper, through the research on the number of taxis in Zhengzhou Airport and the waiting time in the storage tank, a mathematical model is constructed to more accurately use the Bayesian formula to solve the waiting time and number of taxis, so as to provide taxi companies and customers. some convenience. The process of mathematical modeling is briefly described below. The process of mathematical modeling is to extract mathematical models through simplification, assumption, and abstraction of real problems, and then apply mathematical methods and various computer tools to obtain mathematical solutions, and then feed them back to real problems to explain, analyze, and analyze. Check it out. If the test results are in line with or basically in line with reality, they can be used to guide practice. Otherwise, make assumptions, then abstract, modify, solve and apply. The construction of mathematical models includes model preparation, model assumptions, model establishment, model testing, model application.

Model preparation is to understand the background of the actual problem before modeling. The research on the number of taxis in Zhengzhou Airport and the waiting time in the storage tank in this paper is based on the situation that there are often mismatches between the number of customers and the number of taxis and the time of encounter at the airport. In the context of , it is hoped that through

the research results, the optimal departure time and quantity of taxis will bring convenience to taxis and customers. After clarifying the purpose, it is necessary to conduct in-depth research, find out the main contradiction, and collect the necessary data as required. We collected the number of taxis at Zhengzhou Airport on a random day and the waiting time in the storage tank.

Model assumption is that after the purpose is clear, on the basis of mastering the data, grasp the main contradiction of complex problems, discard some secondary factors, make appropriate assumptions about actual problems, so that complex problems can be simplified. This paper ignores the influence of external uncontrollable random factors, such as the weather and special time of holidays, for the number of taxis in Zhengzhou Airport and the waiting time in the storage tank. To establish a model is to determine the main variables according to the main contradiction, and then use appropriate mathematical tools to describe the relationship between the variables, thereby forming a simplified initial model of the mathematical model. In this paper, the main variables are determined as the taxi waiting time λ and the number of taxis μ , and then use the Bayesian prior to consider the rational selection of the optional prior family only with prior information, which can be estimated by various parameters. The method establishes the corresponding selection criteria [1], and forms a preliminary simple model. Model checking is to analyze the established model, use various methods to obtain mathematical results, and then return the obtained answer to the actual problem to test its rationality, and repeatedly modify the relevant content of the model to make it more realistic, which is more practical. The model and process solution will be introduced later. The final model uses the established model to analyze, explain existing phenomena, and predict future development trends, so as to provide reference for people's decision-making.

2. Conditional assumptions

- 1). Assume that the airport taxi traffic flow follows a Poisson distribution with parameter η .
- 2). Assume that the average waiting time of airport taxis obeys an exponential distribution with parameter λ .
- 3). Assume that each recorded taxi waiting time is independent of each other.
- 4). Assume that the data of this day can represent the general situation of Zhengzhou Airport every day, ignoring uncontrollable factors such as weather

3. Symbol description

Table 1. Symbol description

Symbol	illustrate	unit
T	taxi waiting time	min
λ	Average taxi waiting time	min
η	Reciprocal of the average number of taxis in the storage tank	/vehicle
n	Number of samples	vehicle
$\bar{x}$	sample mean	vehicle
α	Hyperparameters	/
β	Hyperparameters	/
x	sample value	min
t_i	Departure time of the i-th taxi	min

4. Problem Analysis

For the exponential distribution model of the average waiting time of airport taxis in the storage tank, the average waiting time T of airport taxis obeys the exponential distribution. In this paper, we need to estimate the parameters of the exponential distribution. In this paper, we know that the inverse gamma distribution and the uninformed prior are the Prior distribution, at the same time, this paper verifies that the inverse gamma distribution is still its conjugate prior distribution, as well as its

posterior distribution and posterior expectation estimation (the process has been omitted). Now select n vehicles for the censored life test, extract the rental car, record its waiting time, and calculate the joint density function of the censored sample. Then this paper determines the prior distribution of its parameters and the hyperparameters for determining the prior distribution, that is, the prior distribution. Use prior moments and prior quantiles to determine the hyperparameters of the prior distribution when the type is known. In this way, we can get the prior distribution density of λ and the posterior distribution density. Therefore, this paper randomly selects 100 taxis, and continuously measures the waiting time for 30 minutes under the specified conditions, that is, the posterior distribution is obtained by updating the prior distribution according to the sample, and the average waiting time is obtained by estimating the expected parameters.

5. Model establishment

5.1 The exponential distribution model of the average waiting time of taxis in the airport storage tank

The waiting time of taxis in the storage tank is different in different time periods. There are 24 hours from 9:00 pm on the first day to 9:00 pm on the next day. The waiting time of each taxi is as shown in Figure 1.

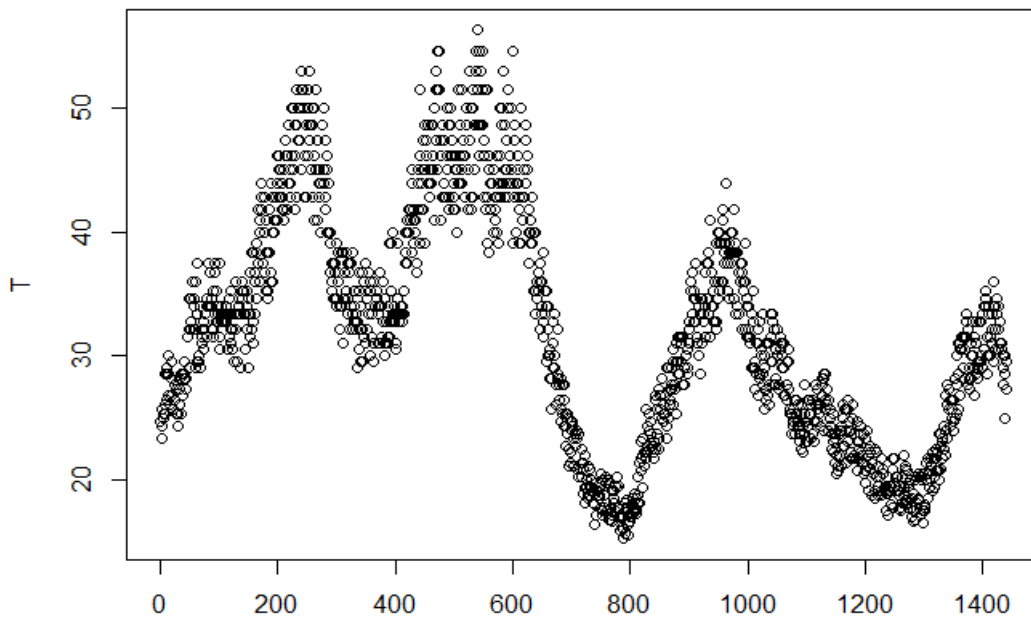


Figure 1. The waiting time of each taxi

5.1.1 Estimation method of exponential distribution.

In this paper, the inverse gamma distribution will be used here, and its density function is

$$p(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\left(-\frac{\beta}{y}\right), y > 0$$

Now, let be a sample from an exponential distribution with a density of:

$$p(x|\lambda) = \lambda^{-1} e^{-\lambda^{-1}x}, x > 0$$

1) Now verify that the inverse gamma distribution is the conjugate prior distribution of the parameter λ

Assuming that the parameter λ obeys the distribution, its density function is

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{\alpha+1} \cdot \exp\left\{-\frac{\beta}{\lambda}\right\}, \lambda > 0$$

$$\begin{aligned}\pi(\lambda|x_1, x_2, \dots, x_n) &= \frac{p(x_1, x_2, \dots, x_n|\lambda) \cdot \pi(\lambda)}{\int_0^{+\infty} p(x_1, x_2, \dots, x_n|\lambda) \cdot \pi(\lambda) d\lambda} \\ &= \frac{\lambda^n \cdot e^{-\sum_{i=1}^n x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{-(\alpha+1)} \cdot e^{-\beta/\lambda}}{\int_0^{+\infty} \lambda^n \cdot e^{-\sum_{i=1}^n x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{-(\alpha+1)} \cdot e^{-\beta/\lambda} d\lambda} \\ &= \frac{\lambda^{(n+\alpha-1)} \cdot \exp\left\{-\left(\beta/\lambda - \lambda \sum_{i=1}^n x_i\right)\right\}}{\int_0^{+\infty} \lambda^{(n+\alpha-1)} \cdot \exp\left\{-\left(\beta/\lambda - \lambda \sum_{i=1}^n x_i\right)\right\} d\lambda} \\ &= \frac{(\beta + \sum_i x_i)^{\alpha+n}}{\Gamma(\alpha+n)} \lambda^{-(\alpha+n)} \exp\left\{-\frac{(\beta + \sum_i x_i)}{\lambda}\right\}\end{aligned}$$

Obviously, the posterior distribution of the parameter at this time is

$$\pi(\lambda|x) \propto p(x|\lambda)\pi(\lambda) \propto \lambda^{-n} e^{-n\bar{x}\lambda^{-1}} \lambda^{-(\alpha+1)} e^{-\beta\lambda^{-1}} = \lambda^{-(\alpha+n+1)} e^{-(\beta+n\bar{x})\lambda^{-1}}$$

and has the form $IGamma(\alpha + n, \beta + n\bar{x})$. Therefore the inverse gamma distribution is the conjugate prior distribution of the exponential distribution parameter λ .

Where $\bar{x}$ is the sample mean. Therefore, the posterior expected distribution of the parameter λ at this time is

$$\hat{\lambda}_B = E(\lambda|x) = \frac{\beta + n\bar{x}}{a + n - 1} = \frac{\alpha - 1}{\alpha + n - 1} \times \frac{\beta}{\alpha - 1} + \frac{n}{\alpha + n - 1} \times \bar{x}$$

The posterior distribution is the inverse gamma distribution. At this point, the posterior distribution is $\pi(\lambda|x) \propto p(x|\lambda)\pi(\lambda) \propto \lambda^{-n} e^{-n\bar{x}\lambda^{-1}} \lambda^{-(\alpha+1)} e^{-\beta\lambda^{-1}} = \lambda^{-(\alpha+n+1)} e^{-(\beta+n\bar{x})\lambda^{-1}}$

where $\bar{x}$ is the sample mean. Therefore, the posterior expectation estimate of the parameter λ at this time is $\hat{\lambda}_B = E(\lambda|x) = \frac{\beta+n\bar{x}}{a+n-1} = \frac{\alpha-1}{\alpha+n-1} \times \frac{\beta}{\alpha-1} + \frac{n}{\alpha+n-1} \times \bar{x}$

2) When putting the density of the exponential distribution $p(x|\lambda) = \lambda^{-1} e^{-\lambda^{-1}x}, x > 0$

When the whole is regarded as a scale parameter family, the uninformative prior of the parameter λ is $\pi(\lambda) = \lambda^{-1}$, and the posterior distribution is $IGamma(\alpha + n, \beta + n\bar{x})$, so we can get the posterior expectation estimate of the parameter λ as:

$$\hat{\lambda}_B = E(\lambda|x) = \frac{n\bar{x}}{n-1} = \frac{1}{n-1} \sum_{i=1}^n X_i$$

3) And the Jeffreys uninformative prior of parameter λ is $\pi(\lambda) = \lambda^{-1}$. So this prior is consistent with the prior of case 2). In that case, the posterior expectation estimate of parameter λ at this time is consistent with the result of case 2)

5.1.2 Model Construction

It is known from the early experience that the waiting time of taxis generally obeys an exponential distribution. We now assume that its density function and distribution function are:

$$p(t|\lambda) = \lambda^{-1} e^{-t/\lambda}, t > 0, F(t|\lambda) = 1 - e^{-t/\lambda}$$

Among them $E(T) = \lambda$, and λ is the average waiting time.

Randomly select n rental cars at different times, record their waiting time, record until the r ($r \leq n$) taxi leaves, record their departure time as, and the other $n-r$ taxis have no passengers until the

experiment stops. ride. The sample that obtains the waiting time is , and the joint density function of this censored sample is

$$p(t|\lambda) \propto \left[\prod_{i=1}^r p(t_i|\lambda) \right] [1 - F(t_r|\lambda)]^{n-r} = \lambda^{-r} \exp\{-s_r/\lambda\}$$

Where $s_r = t_1 + t_2 + \dots + t_r + (n-r)t_r$ is called the total experiment time.

Now we process the collected data, randomly sample the data without replacement, draw 200 sets of data, calculate the mean, median, and 1/4 and 3/4 quantiles,

In order to find the Bayesian estimate of the average taxi waiting time λ , we must first determine the prior distribution of λ . We tested the uninformative prior $\pi(\lambda) = \lambda^{-\frac{1}{2}}$ and the inverse gamma $IGamma(\alpha, \beta)$ respectively, and found that the inverse gamma distribution $IGamma(\alpha, \beta)$ was chosen as the prior distribution $\pi(\lambda)$ of λ . is feasible, so the posterior density of the parameter λ can be written as:

$$\begin{aligned} \pi(\lambda|t) &\propto p(t|\lambda)\pi(\lambda) \propto \lambda^{-(\alpha+r+1)} e^{-(\beta+s_r)/\lambda} \\ \pi(\theta|t) &= \pi(\theta) \cdot p(t|\theta) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{-(\alpha+1)} \cdot \exp\left\{-\frac{\beta}{\theta}\right\} \cdot \theta^{-r} \cdot \exp\left\{-\frac{s_r}{\theta}\right\} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{-(\alpha+r+1)} \cdot \exp\left\{-\frac{\beta + s_r}{\theta}\right\} \end{aligned}$$

The kernel of the inverse gamma distribution is obtained, so the posterior distribution of λ is $IGamma(\alpha + r, \beta + s_r)$. If the posterior mean is taken as the Bayesian estimate of λ , there are $\hat{\theta} = E(\theta|t) = \frac{\beta+s_r}{\alpha+r-1}$.

Then, we need to determine the values of the hyperparameters α and β . We use R software to process the experimental data. The longest waiting time is not more than 55min, and the shortest waiting time is not less than 15min. After processing such a large amount of prior information, it is confirmed that the average waiting time of taxis in the Zhengzhou Airport storage tank is not less than 30 minutes, its 1/4 quantile is about 24 minutes, and its 3/4 quantile is about 38 minutes. This data is in line with the actual situation of the average waiting time of taxis. According to the properties of the inverse gamma distribution, the following equation can be listed:

$$\begin{cases} \frac{\beta}{\alpha-1} = 30 \\ \int_0^{24.32} \pi(\lambda) d\lambda = 0.25 \end{cases}$$

The prior $\pi(\lambda)$ is the density function of the inverse gamma distribution, and its mathematical expectation is $E(\lambda) = \beta/(\alpha-1)$. Using R to solve this system of equations, we can get

$$\begin{cases} \alpha = 17.93 \\ \beta = 507.85 \end{cases}$$

In this way, we can get the prior distribution density of λ is $IGamma(17.93, 507.85)$ and the posterior distribution density is $IGamma(17.93 + r, 507.85 + s_r)$

Now randomly select 100 taxis, and continuously measure the waiting time for 1.5 minutes under the specified conditions. At this time, the total experimental time is

$$s_r = 1.5 \times 30 = 45$$

Thus, the posterior distribution density of λ is $IGamma(17.93, 512.35)$, according to which, the Bayesian estimate of the average waiting time of taxis is: $\hat{\theta} = \frac{512.35}{17.93-1} = 30.26$.

At the same time, we can also calculate the credible lower bound of the posterior expected distribution through the posterior distribution $IGamma(17.93, 512.35)$, and it is now proved that:

From the prior distribution of θ obeying $IGamma(\alpha, \beta)$, the prior distribution of $\hat{\theta}^{-1} \sim Gamma(\alpha, \beta)$:

$$\begin{aligned}\pi(\theta|x_1, \dots, x_n) &= \frac{\pi(\theta) p(x_1, \dots, x_n|\theta)}{\int_0^{+\infty} \pi(\theta) p(x_1, \dots, x_n|\theta) d\theta} \\ &= \frac{\left(\theta^n e^{-\theta \sum_i x_i}\right) \cdot \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} e^{-\beta\theta}\right)}{\int_0^{+\infty} \left(\theta^n e^{-\theta \sum_i x_i}\right) \cdot \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} e^{-\beta\theta}\right) d\theta} \\ &= \frac{\theta^{n+\alpha-1} e^{-(\beta + \sum_i x_i)\theta}}{\int_0^{+\infty} \theta^{n+\alpha-1} e^{-(\beta + \sum_i x_i)\theta} d\theta} \\ &= \frac{\left(\beta + \sum_i x_i\right)^{n+\alpha}}{\Gamma(n+\alpha)} \cdot \theta^{n+\alpha-1} e^{-(\beta + \sum_i x_i)\theta}\end{aligned}$$

That is to obey the posterior distribution of $\hat{\theta}^{-1} \sim Gamma(\alpha + r, \beta + s_r)$, now we can find the lower credible limit of the confidence level of $1 - \gamma = 0.9$, and set θ_L as the 0.9 credible lower limit of θ , then we have: $p(\theta \geq \theta_L|t) = 0.9 \Leftrightarrow p(\theta^{-1} \leq \theta_L^{-1}|t) = 0.9$

So θ_L can be obtained as the 0.9 quantile of $Gamma(17.93, 512.35)$, we use R software to calculate the θ_L^{-1} .

```
> qqgamma(0.9, 17.93, 512.35)
[1] 0.04591669
```

$$\hat{\theta}_L = 1 / 0.04591669 = 21.77857(\text{min})$$

So it is close to the 0.9 quantile obtained by the sample, which is in line with the actual situation.

5.2 Poisson distribution model of the number of taxis in the airport storage tank

Here, taxis and the number of passengers has been seen as a steady stream in this paper, and the taxis are regarded as a Poisson flow, and the number of taxis obeys the Poisson distribution with the parameter η . Its probability function is $p(x|\eta) = \frac{\eta^x}{x!} e^{-\eta}, x = 0, 1, 2, \dots$

Considering the reasonable selection of the optional prior family only with prior information, the corresponding selection criteria can be established with the help of various methods of parameter estimation [1]

(1) When the mean (variance) η of the Poisson distribution takes the gamma distribution as the prior, the gamma distribution is its conjugate prior distribution, and the following proves that the gamma distribution is its conjugate prior distribution:

Let the parameters of the Poisson distribution λ obey the prior distribution $Gamma(\alpha, \beta)$ and its density function is:

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{\alpha-1} \cdot \exp\left\{-\frac{\beta}{\lambda}\right\}, \lambda > 0$$

The sample $x = (x_1, x_2, \dots, x_n)$ is taken from a Poisson distribution, and its joint density function is: $p(x|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! x_2! \dots x_n!} e^{-n\lambda}, -\infty < x_1, \dots, x_n < +\infty$.

$$\begin{aligned}\pi(\lambda|x_1, \dots, x_n) &= p(x_1, \dots, x_n|\lambda) \cdot \pi(\lambda) \\ &= \frac{\frac{1}{\prod_{i=1}^n \Gamma(x_i)} \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_0^{+\infty} \frac{1}{\prod_{i=1}^n \Gamma(x_i)} \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda} \\ &= \frac{\lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-(n+\beta)\lambda}}{\int_0^{+\infty} \lambda^{\alpha + \sum_{i=1}^n x_i - 1} e^{-(n+\beta)\lambda} d\lambda} \\ &= \frac{(n+\beta)^{\alpha+n\bar{x}}}{\Gamma(\alpha+n\bar{x})} \cdot \lambda^{\alpha+n\bar{x}-1} \cdot e^{-(n+\beta)\lambda}\end{aligned}$$

Obviously, the conjugate prior distribution of the parameter λ is $Gamma(\alpha, \beta)$, and its posterior distribution function has the form $Gamma(\alpha + n\bar{x}, n + \beta)$

In this paper, the posterior expectation of η is obtained as

$$E(\eta|x) = \frac{n\bar{x} + \alpha}{\beta + n} = \frac{n}{\beta + n} * \bar{x} + \frac{\beta}{\beta + n} * \frac{\alpha}{\beta}$$

Where $\bar{x}$ is the sample mean, so the posterior expectation of the mean (variance) η of the Poisson distribution at this time is estimated as $\hat{\lambda}_B = E(\lambda|x) = \frac{n}{\beta+n} + \bar{x} + \frac{\beta}{\beta+n} * \frac{\alpha}{\beta}$, which is a weighted average between the sample mean and the prior mean.

(2) When the prior distribution is taken as Jeffreys uninformative prior $\pi(\eta) = \eta^{-\frac{1}{2}}$, the posterior distribution is gamma distribution $Gamma(n\bar{x} + \frac{1}{2}, n)$. Therefore, the posterior expectation estimate of the mean (variance) η of the Poisson distribution at this time is [6] $\hat{\eta}_B = E(\eta|x) = \frac{n\bar{x} + \frac{1}{2}}{n} = \bar{x} + \frac{1}{2n}$

In principle, the decision of the prior distribution should be based on knowledge of the problem in question, including empirically accumulated knowledge [3]. According to the prior information, now take $Gamma(\alpha, \beta)$ as the prior distribution of θ_1 and θ_2 , and then we can get their posterior distribution as $Gamma(\alpha + n\bar{x}, n + \beta)$.

6. Model Evaluation and Improvement

6.1. Advantages:

- 1). The model transforms complex practical problems into simple and understandable Bayesian application problems for easy understanding.
- 2). The model is solved with R software, which makes the solution of the problem intuitive, which is beneficial to the reader's understanding and analysis.
- 3). The model has strong applicability and is suitable for solving many models, which is relatively classic.

6.2. Disadvantages:

- 1). Although the model is applicable and relatively simple, the assumption is relatively ideal. In reality, whether it is the time to arrive at the airport, the weather, or the flow of people, it will affect the waiting time. It can be properly optimized and converted, and variables can be added to make the model become relatively complex and closer to reality.
- 2). The amount of calculation is large. In the process of calculation, it is necessary to use R software for many times to obtain the solution of the model, and a large amount of data is required.

References

- [1] Li Yong. Research on prior distribution selection method based on prior information [J]. Journal of Southwest University (Natural Science Edition), 2009,31(01):15-18.
- [2] He Zongke, Zhong Zhong, Kuang Zhili, Hu Xianghong, Wang Wu. Research on Bayesian Statistical Experiment Design under Exponential Distribution [J]. Electronic Product Reliability and Environmental Testing, 2020, 38(S1): 1-5
- [3] Huang Yuelan. Research on the Estimation of Change Points of Single Parameter Exponential Distribution[J]. Journal of Guangxi Normal University for Nationalities, 2014,31(03):1-3.DOI:10.19488/j.cnki.45-1378/g4.2014.03. 002.
- [4] He Chaobing, Du Baojian, Liu Huawen. Point Estimation of Mixed Poisson Distribution Parameters under Random Censored Experiments with Incomplete Information [J]. Mathematical Practice and Understanding, 2014, 44(12): 187-195.
- [5] Wu Qiguang, Li Guoying, Zhao Yonghui. A new class of prior distributions for exponential reliability growth model [J]. Acta Mathematica Phys., 2003(04): 474-484.
- [6] Mao Shisong. Bayesian Statistics [M]. Beijing: China Statistics Press, 1999
- [7] Chen Jiading, Zhang Pin. Assurance test in MIL-HDBK-781 [J]. Mathematical Statistics and Management, 2000, 19(2): 39-45.