AN OPTIMAL NUMBER DESIGN MODEL OF SERVICE WINDOWS FOR THE RSA FACILITIES ON EXPRESSWAYS

Kazuya AOKI, R&D Center, PASCO Corporation, Tokyo, JAPAN, E-mail: kiakzo6013@pasco.co.jp

Kouji YAMAMOTO, Shimizu Construction Office, Tokyo Regional Office, Central Nippon Expressway Co.Ltd., Shizuoka, JAPAN, E-mail:k.yamamoto.af@c-nexco.co.jp

Kiyoyuki KAITO, Assoc. Prof., Frontier Research Center, Osaka University, JAPAN, E-mail:kaito@ga.eng.osaka-u.ac.jp

Kiyoshi KOBAYASHI, Professor, Graduate School of Management, Kyoto University, JAPAN, E-mail: kkoba@psa.mbox.media.kyoto-u.ac.jp

ABSTRACT

This paper focuses upon the optimal number of service windows for Rest and Service Area (RSA) facilities on expressways. The arrival phenomena of the customers to the service facilities are formulated as mixed Poisson processes, and a novel methodology was proposed to decide the optimal number of service windows to satisfy the service level under a given condition by forecasting the queuing process. The arrival events of the customers, which are subject to the mixed Poisson processes, are represented as the probabilities of excesses over the threshold using log database of service time for each window. However, when a queue occurs in service facilities, the arrival rates cannot be directly measured from the log data. We developed a methodology to estimate the maximum likelihood rates of deviation from the average arrival probability applying Monte-Carlo simulation based upon the occupancy data of the facilities in busy periods. Furthermore, the simulation model is formulated to investigate the optimal number of windows, which is designed to keep the efficient service levels. Finally, the practical availability was investigated based on the case of the service windows for RSA facilities on the Tomei-Expressway in Japan.

Keywords: mixed Poisson processes, queue, optimal number design model, RSA facilities, Monte-Carlo simulation

INTRODUCTION

The importance of Rest and Service Area (RSA) facilities on expressways is recently rediscovered as a major transportation space for expressway customers, and renovation plans based on a renewed concept considering functionality and also for comfort and convenience has been carried out. For example, the number of existing service windows for RSA facilities was decided based on the standard which was designed according to the area characteristics and other factors such as the number of parking lots of each RSA. However, the number of customers who arrive at the facilities is different by each RSA facility. And the way a queue started due to concentrations of arrival customers is different from customers' characteristics. In order to provide services of equal level for customers, it is necessary to figure out the arrival phenomenon of each RSA facility accurately and then allocate service windows to reduce waiting time as much as possible.

The number of customers of RSA facilities is different from utilization characteristics of the RSA itself, thus it happens occasionally that customers arrive and concentrate in a facility at one time due to an arrival of a big tour groups. Furthermore, queuing phenomena is affected by when and how long the facilities are used by each customer. Therefore, in order to figure arrival phenomena of the customers accurately, log data about utility time of each customer is required. There are some ideas to acquire the log data, but the reality that it is impossible to observe hourly number of arrival customers, their utility time and initiation time of queuing simultaneously. Accordingly, a methodology to estimate arrival probabilities based upon limited information should be developed. In this research, a novel methodology is proposed to estimate the arrival probabilities by using the log data recording only at start and finish time of use of each customer at each service window such as resting rooms.

In the case of RSA facilities on expressways, arrival customers have different characteristics such as an individual or a party traveller, and these customers arrive randomly. And, their characteristics (an individual or a party traveller) are not recorded on the log data. In this paper, the phenomenon that heterogeneous customers arrive at RSA facilities by mixture is represented and adopted as probabilities of excesses over the threshold and the methodology to estimate approximate hourly number of arriving customers. And the simulation model is developed which analyses the degree of the impact of difference of the number of the service windows. The impact of service windows is estimated by changing the number of service windows and evaluated the queuing process by using the arriving probabilities. The second section of this paper briefly presents the fundamental idea of this research. In the third section, the basic idea of arriving probabilities is described. And the estimating method of arriving probabilities in the case that a queue occurs is proposed in the fourth section. The fifth section presents the procedure to find the optimal number of service windows using arriving probabilities. Finally, a case study using the methodologies proposed in this paper is considered in the sixth section.

RESEARCH FUNDAMENTALS

Outline of the research

A large number of studies have been conducted on mathematical models which analyze the probabilistic process of congestion situation such as a queue which occurs when customers receive a service, and some documents on queuing theories are published. The traditional researches on queuing analysis focused on analyses of the degree of the impact of waiting time or operation availability of facilities depending on the number of service facilities, under the condition of a certain distribution of arrival probabilities and service time. And the some studies were conducted on analyses of optimal number of service windows, service price and capacity by modelling stochastic processes using fundamental queuing theory. Furthermore, in the field of architectonics, studies about computing the optimal number of rest rooms in a station were reported. On the other hand, Kaito et al. indicated the relationship between hazard generation risk and road patrol frequency by formulation the generation processes of road hazard as the mixed Poisson processes.

In the present research, the congestion situation of arriving customers of RSA facilities on expressways is formulated. Since the data about how long the facilities are used by each customer is only observed, it is impossible to estimate the arriving probabilities directly. In the case of considering the complicated way that customers who have different arrival characteristics simultaneously, it is difficult to compute an optimal number of service windows analytically by using a traditional queuing theory. The mixing service system of heterogeneous arriving customers can be formulated as the mixed Poisson processes of which the events exceeding over a certain threshold follow the Poisson process. In this paper, a novel methodology is proposed to analyze the arriving phenomena follows the mixed Poisson processes.

Determination of optimal number of service windows

This paper focuses on how to decide the optimal number of service windows for service facilities on expressways. Now, the starting processes of queuing are simulated by using two parameters, (i) arriving probabilities and (ii) length of time used by each customer, and the probabilities that waiting time is started are computed. Arriving probabilities of each customer is different depending on period of time. In the case of RSA facilities on expressway in the section 6, it can be used anytime within 24 hours a day and the simulation analysis for computing the arriving probabilities and starting processes of queuing should be calculated according to the period of time. In this case, the optimal number of service windows is defined as the minimum number of service windows that satisfies a certain service level at the period of time those arrival probabilities reaches the maximum.

Also, in this research, the optimal number of service window is calculated directly by using starting processes of queuing (expected value of waiting time). In this case, construction cost

corresponding to number of windows, maintenance cost and external diseconomy because of waiting time are not considered.

Characteristics of observational data

In order to decide the number of service windows of a service facility, it is necessary to estimate arriving probabilities of customers using observational data about present usage situation. Usage situation of service facilities on expressways changes corresponding to period of time, a day of the week or season. It is impossible to observe such situation of customers constantly and acquire data of waiting time of each customer at crowded time.

An observational equipment to record the usage data of customers is installed in each RSA facility, and only start and finish time of usage are recorded. The number of service windows in use at certain times can be estimated by using these observational data. If there are sufficient numbers of service windows and queuing do not occur, it can be considered that a customer uses a facility at the same time as arriving and the number in use at a certain time can be considered as the arriving number.

On the other hand, in the case of all service windows are in use by arrival a big group, queuing are started. However, at this time, the only information that all facilities are in use is recorded and the information about queuing process cannot be observed. Therefore, the information about arriving number of customers during the occurrences of queuing cannot be figured out.

The observational data are accumulated as the time series data, in which all facilities are in use, and therefore continuance usage time are calculated. In a word, time from start of the starts of the queuing to finish can be calculated from the observational data. This paper proposes how to estimate the arriving number of customers based upon the information about continuous queuing starting time by using the observational data. In addition, the observational data that shows the using situation shown in the section 6 is basically observed and indicated to show the customers which booths are being used or not, not newly observed particularly for this study. This study does not use this data itself but secondarily processed data to estimate the number of arrivals.

ARRIVING PROBABILITY

Preconditions

Considering the continuous time axis starting from calendar time $t₀$, the point on the continuous time axis is called time point. The service facilities are composed of the unit facilities of N (windows), and each window provides service only for a single customer. When customers arrive at all service windows, a queuing is started. Customers who have different characteristics such as individual or party traveller arrive at service windows. The

size of customers who arrive at service windows is different depending on customer characteristics. And the log data about how long the service facilities are used cannot identify the characteristics of arriving customers. In this paper, the arriving probabilities, which are subject to the mixed Poisson processes, are represented as the probabilities of excesses over the threshold. In this section, the methodology to estimate arriving probabilities of customers by using the log data is proposed.

Poisson arrival probability

It can be assumed that arrival process of customers at the RSA facilities counting processes of continuous time type which shows starting processes of random phenomena. When assuming that arriving processes are mutually independent according to time each zone, it can be defined as the Poisson arrival process.

Here, an infinitely continuous time axis starting from initial time is used. Accumulated number of arriving customers at service facilities $i(i=1,\dots,I)$ at time range $[0,t](t\geq0)$ is represented by state variable $n_i(t)$. Arriving processes can be described as counting process $n(t)$ $(t \ge 0)$ which jump is generated at arriving time. That is, when arriving time of customer is represented as $\tau_k (k = 1, 2, \cdots)$, counting process is defined as;

$$
n(t) = \sum_{k=1}^{\infty} I_{\tau_k \leq t} \tag{1}
$$

where the following notation rule is given.

$$
I_{\tau_k \le t} = \begin{cases} 1 & \text{(when } \tau_k \le t) \\ 0 & \text{(others)} \end{cases}
$$
 (2)

Here, counting process is right-continuous which jump is generated at arriving time. Also, when counting process $n(t)$ satisfies the following conditions, (i) $n(0) = 0$ is almost and certainly accepted, (ii) An independent increment is possessed, (iii) *ⁿ*(*t*) is assumed to be the Poisson distribution $P_{\rho}(\lambda t)$ to arbitrary t, $n(t)$ is called homogeneous Poisson process. When arriving process of customer follows homogeneous Poisson process, the following equation is defined to arbitrary $0 < u < t$;

$$
P(n(v) - n(u) = n) = P(n(v - u) = n)
$$

$$
= \frac{\{\lambda(v - u)\}^n}{n!} \exp\{-\lambda(v - u)\}\tag{3}
$$

where, $n(v) - n(u)$ shows the number of arriving customers at time range $[u, v)$.

$$
12^{th}
$$
 WCTR, July 11-15, 2010 – Lisbon, Portugal

Probability of excesses over the threshold

An arriving process of customer of RSA facilities on expressway, as mentioned since, cannot be formulated by homogeneous Poisson process because the size of customers who arrive is distributed in the probability. In this paper, the phenomenon that customers who have different characteristics arrive at service facilities by a mixed manner is reported as the mixed Poisson process and arriving probabilities according to the number of arriving customers is calculated approximately by the difference of the probabilities of excesses over the threshold.

When the occurrence probability of the event that n_i customers or more arrive at unit time to the service facility *i* follows homogeneous Poisson process, that probability can be expressed by;

$$
P_{N=n_i} = \lambda_i \exp(-\lambda_i) \tag{4}
$$

Also, the probability $p(n_i)$ that n_i customer arrives at unit time to the service facility *i* can be computed approximately by using following equation;

$$
p(n_i) = P_{N=n_i} - P_{N=n_i+1}
$$
\n(5)

Arriving probabilities of time zone that queuing is not started in service facilities can be estimated directly by using equation (4) and (5) with observational data.

ARRIVING PROBABILITY IN QUEUING

Basic idea of arriving probabilities

The observational data of time zone that queuing is started in service facilities have only information that all facilities are in usage, and the number of arriving customers cannot be observed directly. However, information on continuous time that the queuing is started can be obtained based on information that all service facilities are in usage. The methodology is proposed to estimate the arriving probabilities of time zone that queuing is started based on archived data before the start of the queuing.

The arriving probability of time zone that queuing is not started is expressed by $p(n_i)$. This $p(n_i)$ is defined as the bench-mark case, and difference between bench-mark case and the number of arriving customers of time zone that queuing is started is expressed by compensation coefficient β . That is, it is assumed that $n_i \cdot \beta$ customers arrive at the same probability as the bench-mark case at time zone that the queuing has been started. At this time, the following equation can be defined.

$$
12^{th}
$$
 WCTR, July 11-15, 2010 – Lisbon, Portugal

$$
q(n_i \cdot \beta) = p(n_i) \tag{6}
$$

Estimation method

In order to express arriving processes of customers at the time in queuing using equation (6), compensation coefficient β should be estimated with observational data. Arriving probabilities of customers are defined as the difference of the probabilities of excesses over the threshold in this paper, thus it is difficult to compute the compensation coefficient β analytically. Hence, the simulation that assumes multiple arriving sizes of customers while changing the compensation coefficient is executed. And the methodology is described to estimate the compensation coefficient β by the maximum likelihood estimation based on the continuous starting time of queuing under given conditions.

The compensation coefficient of a service facility that shows the difference from bench-mark case is expressed by β and the probability that queuing for t minutes is continuously started is expressed by $n(t, \beta)$. And it is assumed that the observational data concerning K was acquired. The observational data show the situation of queuing of service facilities for each time zone and dummy variable τ^k is defined as;

$$
\tau^{k} = \begin{cases} t & \text{(when queuing : } t \text{ minutescontinuously)} \\ 0 & \text{(when not queuing)} \end{cases}
$$
 (7)

where index k shows the sample number of observational data. Here, log likelihood function of the simultaneously occurrence probability that generates the observational data of *K* is formulated as;

$$
\ln\{L(\beta)\} = \ln \prod_{t=0}^{T} \prod_{k=1}^{K} \{n(t,\beta)\}^{\delta^k}
$$

$$
= \sum_{t=0}^{T} \sum_{k=1}^{K} \delta^k \ln\{n(t,\beta)\}
$$
(8)

where T shows the maximum value of continuous waiting time and dummy variable δ^k is defined by;

$$
\delta^k = \begin{cases} 1 & (\text{when } \tau^k = t) \\ 0 & (\text{others}) \end{cases}
$$
 (9)

The compensation coefficient β can be computed as the parameter when log likelihood function (8) is maximum value.

$$
12^{th}
$$
 WCTR, July 11-15, 2010 – Lisbon, Portugal

The probability $n(t, \beta)$ is computed by numerical simulation. One deterministic phenomenon that can happen as an arrival process is generated by using the Monte Carlo simulation, and an innumerable sample set by the trial quite a times is calculated.

OPTIMAL NUMBER DESIGN MODEL

Optimal number of service windows and service level

There is a trade-off relationship between the number of service windows and waiting time of customers. If there are a lot of service windows, waiting time decrease. In this paper, service level is defined by the maximum waiting time of customers in queuing. Hence, the target service level of a facility is set as the acceptable waiting time, and the optimal number of the facility is defined by the minimum number of service windows that satisfies that service level. The starting process of queuing is computed by the Monte Carlo simulation, hence, the frequency that satisfies the service level within the trial number of simulation can be expressed by using probability. Also, the simulation model applied here can use same model explained before.

Naturally, the optimal number of service windows defined in this paper is the optimal solution with conditions given the service level and it is needless to say to that there are many issues for optimization of the entire service facility including the method to decide the optimal service level.

Methods of analysis

The procedure to analyze the optimal number of service windows is shown in the Fig-1. The procedure from collection of the log data to find the optimal number of service windows is classified into three phases.

In Phase-1, the situation of queuing according to time zone (24 hours) is investigated using the log data recorded start and finish time of each customer. The arriving probabilities of time zone that queuing is not started are estimated directly by using log data. On the other hand, in the case that the queuing is started, the frequency value of continuous waiting time is counted.

In Phase-2, the compensation coefficient at the time zone where the queuing is started is estimated by the maximum likelihood method. At first, the arriving probability of the time zone that queuing is not started, which is computed in Phase-1, is set as the bench mark case. And the compensation coefficient of arriving probabilities of each time zone that queuing is started is computed given the frequency of continuous waiting time as input conditions.

Figure 1 – Analysis procedure

Finally, Phase-3 simulates the process of starting of queuing for each time zone with some conditions, estimated arriving probabilities of each time zone, use time of a person and target service level. The transition of expected value of (i) arriving number of customers, (ii) waiting number and (iii) waiting number of customers who are waiting more than service level are computed by using the Monte Carlo simulation. Also the number of samples that exceeds the service level (time) is counted. It is assumed that the time usage of customer follows the normal distribution with mean and variance which are calculated by using log data, and the time usage of each customer is computed by generating random number from normal distribution.

CASE STUDY

Case study overview

The above-mentioned optimal number design model was applied to the RSA facilities of six SA and PA on the Tomei-expressway managed by Central Nippon Expressway Company Limited. The arriving probabilities of each facility were estimated and the waiting time in current situation is forecasted by using log data which was observed from August, 2007 to April, 2009. Furthermore, the simulation when the number of service windows was changed was carried out and the optimal number of service windows that satisfies given service level was computed. In this study, it was assumed that the customer's usage behavior was independent. The acceptable waiting time which is set as the service level is for 2 minutes,

Table 1 – Conditions of analysis

Table 2 – Estimation results of compensation coefficients (Kohoku-PA/down/Female)

which is limit that number of customers who point out dissatisfaction begins to increase, based on the field survey data in 2007. Also, usage time that a customer uses a facility at one time is set as shown in Table-1 indicating the conditions of simulation to forecast of the waiting time are included.

Figure 3 – Estimation results of arriving number and waiting time (Kohoku-PA/down/Female)

Forecasting of arriving probabilities and waiting time

At first, the arriving probabilities of time zone that queuing is not started were computed using observational data (Phase-1). The relation between arriving numbers and arriving probabilities in Kohoku-PA/down/Female were shown in Fig-2. At time zone 1am to 7am and from 9pm to 24pm, queuing is not started. Hence arriving probabilities can be estimated by using the difference of the probabilities of excesses over the threshold.

On the other hand, in order to compute the arriving probabilities of the time zone that queuing is started, the compensation coefficient was computed based on arriving probabilities of time zone that queuing is not started (Phase-2). The trial number of the Monte Carlo Simulation to compute the compensation coefficient by the maximum likelihood method was set to 1,000 times for each time zone, and the search range of the compensation coefficient was set from 1.0 to 10.0 with 0.1 increments in between. And the compensation coefficient that likelihood function becomes the maximum was computed. The arriving probabilities of the time zone from 1am to 2am that queuing is not started were set as the bench mark case to estimate the compensation coefficient. The estimation results of the compensation coefficient of time zone that queuing is started is shown in Table-2. In the case of this facility, customer's arriving concentrates from 9am to 1pm.

Next, the transition of queuing situation was simulated by using estimated arriving probabilities and compensation coefficients (Phase-3). The simulation results of transition of arriving probabilities and waiting time in the case of present number of service windows is shown in Fig-3. The horizontal axis indicates time zone for 24 hours. In the case of this facility, customer's arriving concentrates from 9am to 1pm and the number of arriving is about 7 customers on the average per minute. At this time zone, queuing is started and expected number of waiting customers exceeds 7 or less on the average per minute. Furthermore, of these customers, the maximum number of customers that are waiting for 2 minutes or more, which defined as service level, is over 3. The present number of service

Figure 4 – Relation between probability of waiting over 2 minute and number of service window (Kohoku-PA/down/Female)

windows of this facility is 33. This results show that queuing is started and service level cannot be satisfied at the crowded time zone.

Computing of optimal number of service windows

Next, the number of service windows needed to satisfy the service level was computed by changing the number of service windows. Table-3 shows the present number of service windows in the 23 facilities of six SA/PA, estimated the optimal number of service windows by the simulation and the difference between present and optimal number of service windows. Also, the relation between the probabilities that waiting time becomes 2 minutes longer and the number of service windows in Kohoku-PA/down/Female is shown in Fig-4. This figure clearly indicates that the probability that queuing is started decreases when there are a lot of service windows. In the case of this facility, the probability that queuing is started is about 1% under the condition of present number of service windows. And it shows that the optimal number of service windows to satisfy the service level is 67 or more. In addition, the number of windows when the probability of waiting 2 minutes over satisfies less than 0.1% (significance level 99.9% or more) is defined as optimal number of service windows.

The optimal number of service windows where queuing is started under present number of service windows is more than present number. For example, the optimal number of service windows in Ebina-SA/up/Female/right is 96 while present number is 50 service windows. Accordingly its facility needs increase of 46 (92%). On the other hand, in facilities that the probability of queuing is not high, it can be assumed that service level is satisfied. Hence, the estimation results of optimal number of service windows are less than present number. Also in Table-3, the results of maximum waiting time of customers under the condition of present number of service windows are described in parallel. For example in Ashigara-SA/down /Female/east, the maximum waiting time is 15 minute. The service level is greatly exceeded.

Table 3 – Estimation results of optimal number of service window and maximum waiting time

As described above, some useful information can be provided for renewal planning of RSA facilities by estimating arriving probabilities, the optimal number of service windows and the maximum waiting time using the optimal number design model proposed in this study. However the log data used in this study includes all data in investigated period without consideration of seasonable variable of customer's behaviour on expressway. In order to make a detailed renewal plan, it is necessary to estimate according to busy season, common season, low season and a day of the week. In this case, it is possible to analyze it with same model by dividing the log data. Also, comprehensive countermeasure such as not only layout planning of service windows but operational planning with present facilities is highly important to reduce the maintenance cost.

CONCLUSIONS

This study proposed a novel methodology to decide the optimal number of service windows of RAS facilities on expressways. At first, the arriving probabilities of customers who arrive service facilities randomly were estimated using observational data. At the time, when arriving number of customers cannot be observed directly due to starting of queuing, the methodology to estimate the compensation coefficient parameter indicating difference from

average arriving probabilities at time zone where queuing is not started based on observational data. Furthermore, the simulation model to forecast the queuing situation using the estimated arriving probabilities for each facility is developed and the methodology is proposed to compute the optimal number of service windows needed to satisfy the service level which is given at a maximum waiting time. Finally, the practical availability of these methodologies is investigated by the case study for RSA facilities on Tomei-Expressway. In the result, it indicated that the optimal number of service windows can be computed objectively according to utilization characteristic of each facility. By trying further analysis using this model and adopting these results in renewal plan of RSA facilities, it became possible to compute the optimal number of service windows and provide customers with comfortable service. There is a possibility that this model can be widely applicable to other service facilities of which observational data about customer's behaviour can be acquired. It is necessary to study further applications in the future.

The more elaboration might be necessary for the methodology in the scene of decision making about the optimal number of service windows using this model. In order to simulate more realistic situation, the simple simulation model proposed in this paper is enhancement to multi-agent type model which can consider multiple characteristics of each customer or facility. The model's enhancing according to the situation of the targeted queue mechanism will be the future tasks. On the other hand, in this paper, the service level of waiting time (acceptable waiting time) was given arbitrarily. This proposed method makes it possible to simulate to arbitrary service level. However, in order to secure the validity of the service level, it is necessary to verify it by continuous survey of customer satisfaction and make a logic model to improve it continuously according to PDCA cycle. Also, in the case of making a plan about optimal service windows for existing facilities, it is necessary to create a plan subject to continue to use existing facility. Consequently, it may be considerable that the optimal number of service level cannot be kept only by constructing of new facilities depending on the situation of existing facilities. In this case, comprehensive countermeasures including not only construction of new facilities but reduction in time usage, the optimal distribution by division operation depending on congestion situation will be the matters of proprieties.

REFERENCES

- R.Nelson (1995). Probability, Stochastic Processes, and Queuing Theory, Springer-Verlag, New York.
- Bell, C.E. and S. Stidham (1983). Individual versus Social Optimization in the Allocation of Customers to Alternative Servers, Management Science, Vol.29, No.7, pp.831-839.
- Igaki, N., Yamakawa, S. and Sumita, U. (2000). Social Benefit Analysis of Congestion Systems with Heterogeneous Users : Mathematica Japonicae, Vol.51, No.2, pp.279- 291.
- Stidham, S. (1992). Pricing and capacity decisions for a service facility : Stability and multiple local optima, Management Science, Vol.38, No.8, pp.1121-1139.
- Yechiali, U. (1972). Customers' optimal joining rules for the GI/ M/ s queue, Management Science, Vol.18, No.7, pp.434-443.

- Kiyoyuki KAITO, Kiyoshi KOBAYASHI, Toshiaki KATOH and Noriko IKUTA (2007). ROAD PATROL FREQUENCY AND HAZARDS GENERATION RISKS, JSCE Journals F, Vol. 63, No. 1, pp.16-34 .
- Pickands, J. (1971). The two-dimensional process and extremal processes, Journal of Applied Probabilities, Vol.8, pp.745-756, 1971.