RESEARCH ON THE OPTIMAL TRAIN TICKET OVERBOOKING STRATEGY FOR TRANSPORTATION DURING THE SPRING FESTIVAL

Zhi-Xin Liang\textsuperscript{1} Shuo Wen\textsuperscript{2} Feng-Wen Yang\textsuperscript{2}

\textsuperscript{1}School of Information, Beijing Wuzi University, Beijing, 101149, China
\textsuperscript{2} Beijing No.4 high school, Beijing, 100034, China
liangzhixin@sina.com, wenyan4002@sina.com, yfw@bhsf.cn

Keywords: overbooking strategy; expected revenue; mathematical model; maximum.

Abstract

This paper studies a train ticket overbooking rule and strategy for transportation during the Spring Festival. We determine the expected revenue of train operation by the probability theory. According to the principle which maximizes the expected revenue of the railway sector, we construct a mathematical model to determine the number of overbooking sleeper and seat tickets. Using the optimal solution of the model, the railway sector can find some tactics to reduce the waste of transport resources and improve the train operation efficiency.

1 Introduction

During the Spring Festival transportation period, people often find it hard to get a ticket, but there are some vacancies after the train starts, resulting in a waste of transport resources. On the other hand, some people can not get the ticket to go home for the Spring Festival. It has much to do with the ticket refund. During the Spring Festival, there is a long train ticket pre-sale period during which some people buy tickets and then change their plans and some other people repeat booking and have to refund. According to the report in the website www.people.com.cn on January 29th 2013, “As the data provided by the Transport Bureau of Ministry of Railway shows, the number of returned tickets had reached 460000 per day on average for the first three days after the Transport during the Spring Festival began. The Railway department said that the returned tickets would be sold online. But information about when the tickets will be refunded and which vehicles have returned tickets is asymmetry, and the railway sector can not get the information in advance, so there is no fixed rule to follow.” Those vacancies are caused by the refunded but have not been resold tickets.

Airlines often encounter the vacancy phenomenon. There are always some passengers who book tickets but fail to fly for various reasons. Therefore, overbooking, which means that they sell more tickets than the number of seats to reduce the possible vacancies, is a strategy adopted by many airlines. The problem with this strategy is that sometimes some passengers who get tickets can not travel on the flight because the aircraft is full, so airlines need to give them proper compensation. Airlines always make the best overbooking strategy to maximize their profit [1].

We can learn the practice from airlines, and adopt a train ticket overbooking rule and strategy for the transport during the Spring Festival. When vacancies appear on a train, the number of vacancies varies each day, but we can get the probability of sleeper and seat vacancy. Using the probability theory, we can determine the expected revenue of the train operation. According to the principle that maximize the expected revenue of the railway sector, determine the number of overbooking sleeper and seat tickets. After the sleeper and seat tickets are sold out, choose a right time to start the overbooking system and sell a certain number of overbooking tickets. The passengers who get overbooking tickets will exchange them for the relevant tickets one hour before the train starts. Selling a certain number of overbooking tickets will be able to reduce the waste of transport resources and improve the train operation efficiency.

2 Model Assumptions and Symbols

The overbooking rule is as follow: when the passengers who get overbooking tickets can not exchange the overbooking tickets for the relevant tickets because the berth or seat is full, they will be allowed to stand by. Their fare will be returned and they will receive some compensation. If the berth or seat is vacant after the train starts, the standing room ticket holders should be allowed to take it without adding fare.

Define the variables and symbols as follows:
\( n_{1} \): the number of berth tickets on a train.
\( a_{1} \): the price of a train berth ticket.
\( n_{2} \): the number of seat tickets on a train.
\( a_{2} \): the price of a train seat ticket.
\( n_{3} \): the number of standing tickets.
\( a_{3} \): the price of a standing ticket.
\( x_{1} \): the number of overbooking sleeper tickets, which \( 0 \leq x_{1} \leq n_{1} \).
\( x_{2} \): the number of overbooking seat tickets, which \( 0 \leq x_{2} \leq n_{2} \).
$M_1$: the practical number of passengers who book the sleeper tickets.

$M_2$: the practical number of passengers who book the seat tickets.

$\lambda$: compensation coefficient, the overbooking sleeper ticket holders will gain $\lambda_1$ times of the fare and the overbooking seat ticket holders will gain $\lambda_2$ times of the fare when they can not exchange the overbooking tickets for the relevant tickets.

$p$: attendance, the probability of a ticket holder becomes a passenger, attendance of sleeper is $p_1$ and attendance of seat is $p_2$.

$E(x_1, x_2)$: the expected revenue of train operation.

Assume as follows:

(1) It is independent whether a passenger with sleeper or seat ticket onboard the train.

(2) All the overbooking ticket holders will become passengers.

3. The Mathematical Model for Optimal Overbooking Strategy

The practical number of passengers who book the sleeper or seat tickets is different every day, so $M_1$ and $M_2$ are random variables, then

$$E(x_1, x_2) = \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} P(M_1 = k_1, M_2 = k_2) \cdot f(k_1, k_2),$$

Where $f(k_1, k_2)$ is the profit when $M_1 = k_1, M_2 = k_2$, $P(M_1 = k_1, M_2 = k_2)$ is the probability when $M_1 = k_1, M_2 = k_2$.

The optimal overbooking strategy Model is

$$\max E(x_1, x_2)$$

s.t.

$$0 \leq x_1 \leq n_1,$$

$$0 \leq x_2 \leq n_2.$$

4. Solving Method

From the model assumptions, we know $M_1$ and $M_2$ are independent random variables, where $M_1 \sim B(n_1, p_1)$ and $M_2 \sim B(n_2, p_2)$. Therefore

$$P(M_1 = k_1, M_2 = k_2) = P(M_1 = k_1) \cdot P(M_2 = k_2) = C^n_{n_1} p_1^{k_1} (1-p_1)^{n_1-k_1} \cdot C^n_{n_2} p_2^{k_2} (1-p_2)^{n_2-k_2}.$$  

$f(k_1, k_2)$ is the profit when $M_1 = k_1, M_2 = k_2$, it is depended on the value of $x_1, x_2$.

If $0 \leq n_1 - k_1 < x_1, 0 \leq n_2 - k_2 < x_2,$

$$f(k_1, k_2) = \begin{cases} n_1 a_1 + n_2 a_2 + (n_1 - k_1) a_1 + (n_2 - k_2) a_2, & \text{if } 0 \leq n_1 - k_1 < x_1, n_2 - k_2 \geq x_2, \\ n_1 a_1 + n_2 a_2 + (n_1 - k_1) a_1 + (n_2 - k_2) a_2 - (n_1 - k_1) \lambda a_1 - (n_2 - k_2) \lambda a_2, & \text{if } n_1 - k_1 \geq x_1, 0 \leq n_2 - k_2 < x_2, \\ n_1 a_1 + n_2 a_2 + (n_1 - k_1) a_1 + (n_2 - k_2) a_2 - (n_1 - k_1) \lambda a_1, & \text{if } n_1 - k_1 \geq x_1, n_2 - k_2 \geq x_2, \\ n_1 a_1 + n_2 a_2 + (n_1 - k_1) a_1 + (n_2 - k_2) a_2, & \text{if } n_1 - k_1 \geq x_1, n_2 - k_2 < x_2. \\ \end{cases}$$

$$E(x_1, x_2) = n_1 a_1 + \lambda_1 \sum_{k=0}^{n_1} P(M_1 = k) - \lambda_2 a_1,$$

$$+ n_2 a_2 + \lambda_1 \sum_{k=0}^{n_2} P(M_2 = k) - \lambda_2 a_2,$$

$$+ a_1(1+\lambda_1) x_1 \sum_{k=0}^{n_1} P(M_1 = k) - \lambda_2 a_1 n_1 p_1,$$

$$+ a_2(1+\lambda_2) x_2 \sum_{k=0}^{n_2} P(M_2 = k) - \lambda_2 a_2 n_2 p_2,$$

$$+ a_1(1+\lambda_1) \sum_{k=1}^{n_1} k P(M_1 = k) + a_2(1+\lambda_2) \sum_{k=1}^{n_2} k P(M_2 = k).$$

Using

$$\sum_{k=0}^{n_1} C^n_{n_1} p_1^k (1-p_1)^{n_1-k} = np_1 \sum_{k=0}^{n_1} C^n_{n_1} p_1^k (1-p_1)^{n_1-k}$$

and the central limit theorem, replacing binomial distribution by normal distribution[2],

$$E(x_1, x_2) = \int_{-\infty}^{\infty} (n_1 a_1 - \lambda_1 a_1 x_1 - \lambda_2 a_2 x_2) f(x_1) f(x_2) dx_1 dx_2$$

$$+ a_1(1+\lambda_1) x_1 \int_{-\infty}^{\infty} f(x_1) f(x_2) dx_1 dx_2$$

$$+ a_2(1+\lambda_2) x_2 \int_{-\infty}^{\infty} f(x_1) f(x_2) dx_1 dx_2$$

$$+ a_1(1+\lambda_1) \int_{-\infty}^{\infty} x_1 f(x_1) f(x_2) dx_1 dx_2$$

where

$$f_1(x) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{p_1(1-p_1)}} e^{-\frac{1}{2} x^2 (1-p_1)},$$

$$f_2(x) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{p_2(1-p_2)}} e^{-\frac{1}{2} x^2 (1-p_2)}.$$
\[ f'_1(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{(n_1-1)p_1(1-p_1)}} e^{-\frac{(x-n_1+p_1)^2}{2(n_1-1)p_1(1-p_1)}}. \]
\[ f'_2(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{(n_2-1)p_2(1-p_2)}} e^{-\frac{(x-n_2+p_2)^2}{2(n_2-1)p_2(1-p_2)}}. \]

From the practical significance of \( E(x_1, x_2) \), we know there must be \( x_1, x_2 \), \( 0 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2 \) which maximize the expected revenue \( E(x_1, x_2) \). The point \( (x_1, x_2) \) is the extreme point of \( E(x_1, x_2) \) when \( 0 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2 \). By the necessary condition of extreme value existence, we know
\[
\begin{align*}
E'_1(x_1, x_2) &= 0, \\
E'_2(x_1, x_2) &= 0.
\end{align*}
\]

Calculate the partial derivative,
\[
E'_1(x_1, x_2) = -\lambda_1 a_1 + [n_1 a_1 (1 + \lambda_1) - a_1 (1 + \lambda_1)] f_1(n_1 - x_1 + 0.5)
+ a_1 (1 + \lambda_1) \int_{x_1 - 0.5}^{x_1 + 0.5} f_1(x) dx - a_1 (1 + \lambda_1) f'_1(n_1 - x_1 - 0.5),
\]
\[
E'_2(x_1, x_2) = -\lambda_2 a_2 + [n_2 a_2 (1 + \lambda_2) - a_2 (1 + \lambda_2)] f_2(n_2 - x_2 + 0.5)
+ a_2 (1 + \lambda_2) \int_{x_2 - 0.5}^{x_2 + 0.5} f_2(x) dx - a_2 (1 + \lambda_2) f'_2(n_2 - x_2 - 0.5).
\]

Using
\[
\int_a^b f(x) dx = \Phi(\frac{b-n_1 p_1}{\sqrt{n_1 p_1 (1-p_1)}) - \Phi(\frac{a-n_1 p_1}{\sqrt{n_1 p_1 (1-p_1)})}, i = 1, 2
\]
where \( \Phi(x) \) is the distribution function of normal distribution, solve the equation set
\[
\begin{cases}
E'_1(x_1, x_2) = 0, \\
E'_2(x_1, x_2) = 0.
\end{cases}
\]

If the solution is unique, it is the maximum value point of \( E(x_1, x_2) \) when \( 0 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2 \). Otherwise, calculate all the \( E(x_1, x_2) \), \( (x_1, x_2) \) which value of \( E(x_1, x_2) \) is largest is the optimal solution.

5 Example and Results Analysis

Assume that the number of berths on a train is 480, the price is 250 yuan per berth, the attendance is 0.9; the number of seats is 1200, the price is 180 per seat, the attendance is 0.8; the number of standing tickets is 200, the price is 100 per ticket, namely
\[ n_1 = 480, n_2 = 1200, n_3 = 200, a_1 = 250, \]
\[ a_2 = 180, a_3 = 100, p_1 = 0.9, p_2 = 0.8. \]

Set the compensation coefficient \( \lambda_1 = 0.4, \lambda_2 = 0.2 \).

Calculating by above-mentioned model, then
\[ E'_1(x_1, x_2) = (168000 - 350 x_1) \frac{1}{\sqrt{86.4\pi}} e^{-\frac{(48.5-x_1)^2}{86.4}} - 100 \]
\[ + 350 \Phi(\frac{48.5-x_1}{\sqrt{43.2}}) - 350 \Phi(\frac{-432.5}{\sqrt{43.2}}) - 350 \times \frac{1}{\sqrt{86.22\pi}} e^{-\frac{(48.4-x_1)^2}{86.22}}, \]
\[ E'_2(x_1, x_2) = (259200 - 216 x_2) \frac{1}{\sqrt{384\pi}} e^{-\frac{(240.5-x_2)^2}{384}} - 36 \]
\[ + 216 \Phi(\frac{240.5-x_2}{\sqrt{92}}) - 216 \Phi(\frac{-960.5}{\sqrt{92}}) - 216 \times \frac{1}{\sqrt{383.68\pi}} e^{-\frac{(240.3-x_2)^2}{383.68}}. \]

Set \( \begin{cases}
E'_1(x_1, x_2) = 0, \\
E'_2(x_1, x_2) = 0,
\end{cases} \) solve the equation set to obtain \( (x_1, x_2) \).

Since the equations are very complex, we use MATLAB to obtain the numerical solution of the equations, it is \( x_1 = 68.16120 \), \( x_2 = 284.6094 \). Using rounding-off method, we obtain the integral solution \( x_1 = 68, x_2 = 285 \).

So when the number of overbooking sleeper tickets and overbooking seat tickets are 68 and 285 respectively, the railway sector will obtain maximum expected revenue \( E(68, 285) = 352030 \).

In fact, the number and price of sleeper and seat tickets are determined by the railway sector, so the expected revenue will change with the attendance and the compensation coefficient. The attendance is objective, we can obtain it through statistics. But the compensation coefficient, which is different from the attendance, must be determined by the railway sector. When \( p_1 = 0.9, p_2 = 0.8 \), we took different sets of the compensation coefficient to calculate, the results are listed in the table 1.

<table>
<thead>
<tr>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( x_1 )</th>
<th>( x_2 )</th>
<th>( E(x_1, x_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.01</td>
<td>74</td>
<td>295</td>
<td>355820</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>72</td>
<td>290</td>
<td>354980</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>68</td>
<td>285</td>
<td>352030</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>67</td>
<td>280</td>
<td>342800</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>66</td>
<td>278</td>
<td>330430</td>
</tr>
</tbody>
</table>

As can be seen, with the increase of the compensation coefficient, the number of overbooking tickets will be reduced, and the expected revenue will be reduced. With
the increase of the compensation coefficient, the number of overbooking tickets does not decrease much but the expected revenue decrease greatly. At the same time, we hope more people can go home for the Spring Festival. So the compensation coefficient can be set smaller.

However, the compensation coefficient can not be too small. Otherwise, the compensation is too low to attract people to choose the overbooking system. For example, compare $\lambda_1 = 0.4, \lambda_2 = 0.2$ with $\lambda_1 = \lambda_2 = 0.01$, cash indemnity is from 100 yuan and 36 yuan down to 2.5 yuan and 1.8 yuan. It is not conducive to attract people to choose the overbooking system.

6 Conclusion

In short, to resolve the contradiction about booking hard and the vacancy phenomenon, the railway sector can establish train ticket overbooking system. By establishing and solving the optimal model, determine the number of overbooking tickets. The railway sector can reduce the waste of transport resources and improve the train operation efficiency by selling appropriate quantity of overbooking tickets, and at the same time, the interests of passengers will be protected better.

In the model assumption, not only the ticket price but also the attendance of the whole course is single, actual problems are more complex. For example, some of the trips are not direct, the railway sectors judge the value of ticket piecewise, the attendance varies from station to station, and there are upper, middle and lower berths. Furthermore, we add ‘when $n_1 > M_1$ or $n_2 > M_2$, giving priority to overbooking sleeper ticket holders’ when we formulate the overbooking rule. If we consider these cases, we can set up the mathematical model by similar principles, but the solving method may be more complex, which is left for further study.

References
