ISSN 1064-2307, Journal of Computer and Systems Sciences International, 2021, Vol. 60, No. 5, pp. 770–775. © Pleiades Publishing, Ltd., 2021.
Russian Text © The Author(s), 2021, published in Izvestiya Akademii Nauk, Teoriya i Sistemy Upravleniya, 2021, No. 5, pp. 105–110.

MATHEMATICAL SIMULATION

A Method for Calculating the Route Correspondence of Passenger Flows by the Entry and Exit Data

V. N. Embulaev

Vladivostok State University of Economics and Service, Vladivostok, Russia

e-mail: Vladimir.Embulaev@vvsu.ru

Received October 26, 2020; revised March 24, 2021; accepted May 31, 2021

Abstract—A method for calculating the route correspondence of passenger flows by entry and exit data is developed. As shown below, the number of passengers transiting between two given stations is a discrete random variable. A probability is assigned to each value of this variable. The number of passengers who pass through is defined as the value of a random variable with the highest probability.

DOI: 10.1134/S1064230721050075

INTRODUCTION

Since the transport system of a large city belongs to the class of large-scale systems, there have always been problems with organizing passenger transportation along the city routes [1, 2]. According to experts, in large cities, this problem can be solved through the most rational use of available and ready-to-use vehicles based on the coordinated management of all types of urban passenger transport [3, 4].

The statement and solution of the coordinated management problem for the transport system of a large city indicated the following [5]: in the long-term, current, and operational management, the entire information base is divided into three main arrays: conditionally constant (data on the route network and the composition of the vehicles), normative (economic indices), and variable (data on passenger flows and the related quantitative and qualitative indices of the transport system). As noted, the main problem in the coordinated management of the entire transport system of a large city is the availability of information about the route correspondence of passenger flows (RCPFs), which differ in capacity on all city routes and change within 24 hours on each city route. Therefore, the process of acquiring such information should be computerized.

As is known, the most efficient approach to computerizing the acquisition of information about the RCPFs is processing the entry and exit data: the numbers of passengers entering (a_i, $i = \overline{1,n}$) and exiting (b_j, $j = \overline{1,n}$) the vehicle at each station, where n denotes the number of stations on the route [6]. Among the underinvestigated problems, we mention calculating the distribution of passenger flows along the route, i.e., how many passengers x_{ij} transited between any two stations i and j ($i \le j$) along the route during one run.

1. PROBLEM STATEMENT

For the effective management and organization of passenger flows along the routes of a large city, the availability of information about the RCPFs is important. The developers were forced to determine such information by processing the entry and exit data (the number of passengers entering and exiting the vehicle, respectively) captured for each station of the route.

Note that when capturing the general entry and exit data for each station of the route, the information about the transit of any individual passenger is not captured. His motives for choosing the transit between stations i and j along the route remain unknown. In other words, his entry and exit stations remain unknown as well. Moreover, his transit along the route does not depend on the transits chosen by other passengers. Hence, the transit of any passenger along the route can be considered a random process independent of the transits chosen by other passengers. For each passenger in the vehicle's cabin, the entry and exit stations (i and j) are supposed unknown. Hence, the following assumption is admissible: for any passenger, the events to exit at this station or go further are equally probable.

Thus, the route correspondence of passenger flows can be found by entry and exit data within a probabilistic interpretation.

While the vehicle stops at station j, passenger exchange occurs: of those who arrived from the previous station (denoted by Q_{j-1}), first, a group of b_j passengers exits the cabin, and then a group of a_j passengers enters the cabin. When the vehicle departs from station j, there will be Q_j passengers in the cabin:

$$Q_j = (Q_{j-1} - b_j) + a_j = \sum_{r=1}^{j}(a_r - b_r).$$

Some of the a_i passengers who entered at station i may exit at station $(i + 1)$. They are the passengers transiting between stations i and $(i + 1)$, denoted by $x_{i,i+1}$. Subtracting this number from the passengers who entered at station i, we obtain the remaining passengers among a_i who will continue their transits along the route. We denote this number by $a_{i,i+1}$: $a_{i,i+1} = a_i - x_{i,i+1}$.

Upon the vehicle's arrival at station $(i + 2)$, together with the passengers who exited (b_{i+2}), those who entered at station i can also exit. They are the passengers transiting between stations i and $(i + 2)$ along the route, denoted by $x_{i,i+2}$. Subtracting this number from $a_{i,i+1}$, we obtain the passengers who entered at station i and will continue their transits along the route after station $(i + 2)$

$$a_{i,i+2} = a_{i,i+1} - x_{i,i+2} = a_i - x_{i,i+1} - x_{i,i+2} = a_i - \sum_{r=i+1}^{i+2} x_{ir}.$$

These considerations are easily repeated further.

For any i and j, this number is calculated by the formula

$$a_{ij} = a_i - \sum_{r=i+1}^{j} x_{ir}.$$

In other words, a_{ij} is the number of remaining passengers among those who entered at station i and will continue their transits along the route after station j. It is determined by subtracting from a_i those passengers who have already passed the route segments between stations i to $(i + 1)$ $(x_{i,i+1})$, stations i and $(i + 2)$ $(x_{i,i+2})$, ..., stations i and $(j - 1)$ $(x_{i,j-1})$.

In the transportation process along the route, when the vehicle stopped at station j, there were Q_{j-1} passengers inside the cabin, among whom there were those who entered at station i (a_{ij}). During passenger exchange at the station, those passengers who entered the cabin at station i could exit along with b_j. These passengers are from the group a_{ij}. In this case, the number x_{ij} that will simultaneously belong to a_{ij} and b_j is the required value.

Let the transits of passengers between any two stations on the route be determined in this way. Then they can be described by a table of the route correspondence of passenger flows (Table 1). (By a natural assumption, $x_{ii} = 0$: no passengers start and end their transits at the same station i.) The other notations are the following: n is the number of stations on the route; a_i is the number of passengers who entered the vehicle at station i; b_j is the number of passengers who exited the vehicle at station j; and x_{ij} is the number of corresponding passengers between stations i and j, $i \le j$.

According to Table 1, the problem of determining x_{ij} by the entry a_i and exit b_j data has the following statement:

$$\sum_{j=i}^{n} x_{ij} = a_i; \quad \sum_{i=1}^{j} x_{ij} = b_j; \quad x_{ij} \ge 0; \quad i = \overline{1,n}; \quad j = \overline{1,n}, \tag{1.1}$$

where

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} b_j \quad \left(\text{or } \sum_{i=1}^{n}\sum_{j=i}^{n} x_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{j} x_{ij}\right). \tag{1.2}$$

The system (1.1) consists of $2n$ linear algebraic equations with $n(n+1)/2$ unknowns. A unique solution exists if $2n = n(n+1)/2$. Consequently, a unique solution is possible for $n \le 3$. In the case $n > 3$, there are,

Table 1. The matrix of route correspondence of traffic flows

Entry station no.	Exit station no.							Passengers who entered
	1	2	3	4	5	…	n	
1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	…	x_{1n}	a_1
2		x_{22}	x_{23}	x_{24}	x_{25}	…	x_{2n}	a_2
3			x_{33}	x_{34}	x_{35}	…	x_{3n}	a_3
4				x_{44}	x_{45}	…	x_{4n}	a_4
5					x_{55}	…	x_{5n}	a_5
.						…	.	.
.						…	.	.
.						…	.	.
n							x_{nn}	a_n
Passengers who exited	b_1	b_2	b_3	b_4	b_5	…	b_n	

in principle, infinitely many solutions. Therefore, a unique solution cannot be obtained without additional assumptions about the distribution of passenger flows between the route stations.

2. A PROBABILISTIC METHOD FOR SOLVING THE PROBLEM

When the vehicle stopped at station j, there were Q_{j-1} passengers in the cabin, from which b_j exited. Recall that for each passenger in the vehicle's cabin, the events to exit at this station or continue further are assumed equiprobable. Hence, the exits of different groups of b_j passengers from all who arrived form a system of equiprobable and incompatible events. The total number of such groups in various combinations is equal to $C_{Q_{j-1}}^{b_j}$ (the number of b_j combinations from Q_{j-1} elements).

All passengers who arrived at station j (Q_{j-1}) can be conditionally divided into two groups: one group consists of those passengers who entered the vehicle's cabin at station i and continued their transits up to station j (a_{ij}), and the second group includes the others ($Q_{j-1} - a_{ij}$). Since b_j passengers exited at station j, among them, there may be those belonging to the group a_{ij}. Let us denote them by λ_{ij}. Obviously, λ_{ij} is a random variable taking integer values. And the group λ_{ij} can be combined from a_{ij} in $C_{a_{ij}}^{\lambda_{ij}}$ different ways. For each specific group λ_{ij}, the remaining passengers ($b_j - \lambda_{ij}$) can be chosen in $C_{Q_{j-1}-a_{ij}}^{b_j-\lambda_{ij}}$ different ways. The total number of successful outcomes will be equal to $C_{a_{ij}}^{\lambda_{ij}} C_{Q_{j-1}-a_{ij}}^{b_j-\lambda_{ij}}$.

According to the classical approach [7], the probability that among the exiting passengers b_j exactly λ_{ij} passengers will belong to the group a_{ij} is given by

$$P_{b_j}(\lambda_{ij}) = \frac{C_{a_{ij}}^{\lambda_{ij}} C_{Q_{j-1}-a_{ij}}^{b_j-\lambda_{ij}}}{C_{Q_{j-1}}^{b_j}}. \tag{2.1}$$

In this description of the problem, two cases are possible: $a_{ij} \geq b_j$ and $a_{ij} \leq b_j$.

For $a_{ij} \geq b_j$, the random variable λ_{ij} can take integer values from 0 to b_j; for $a_{ij} \leq b_j$, from 0 to a_{ij}. Generally, it can take values on the interval

$$0 \leq \lambda_{ij} \leq \min[a_{ij}, b_j]. \tag{2.2}$$

Table 2. The values and probabilities of the random variable

Values of λ_{ij}	0	1	2	...	k
Probabilities	p_0	p_1	p_2	...	p_k

In problem (1.1) and (1.2), the random variable λ_{ij} (passenger transits between specific stations i and j on the route) can take nonnegative integer values on interval (2.2) only. Any value outside this interval cannot be taken as a solution of (1.1) and (1.2).

The random variable λ_{ij} is a discrete variable that can take any integer value on the interval (2.2). For each value of this variable, we assign a corresponding probability by formula (2.1); see Table 2, where $k = \min[a_{ij}, b_j]$, and the probabilities p_l, $l = \overline{0,k}$, are given by (2.1) for a specific value of λ_{ij}.

Of all the integers λ_{ij} on the interval (2.2), as the only solution, we select the value for which the probability reaches maximum with respect to the argument λ_{ij}:

$$x_{ij} = \operatorname{argmax} P_{b_j}(\lambda_{ij}).$$

As x_{ij} is the most probable value for the random variable λ_{ij}, the two adjacent numbers $(x_{ij} - 1)$ and $(x_{ij} + 1)$ satisfy the following inequalities [7]:

$$P_{b_j}(x_{ij} - 1)/P_{b_j}(x_{ij}) \le 1 \quad \text{and} \quad P_{b_j}(x_{ij})/P_{b_j}(x_{ij} + 1) \ge 1.$$

Expanding these inequalities, we obtain:

$$\frac{P_{b_j}(x_{ij} - 1)}{P_{b_j}(x_{ij})} = \frac{x_{ij}(Q_{j-1} - a_{ij} - b_j + x_{ij})}{(a_{ij} - x_{ij} + 1)(b_j - x_{ij} + 1)} \le 1,$$

$$\frac{P_{b_j}(x_{ij})}{P_{b_j}(x_{ij} + 1)} = \frac{(x_{ij} + 1)(Q_{j-1} - a_{ij} - b_j + x_{ij} + 1)}{(a_{ij} - x_{ij})(b_j - x_{ij})} \ge 1.$$

Solving them for x_{ij} gives

$$\frac{(a_{ij} + 1)(b_j + 1)}{Q_{j-1} + 2} - 1 \le x_{ij} \le \frac{(a_{ij} + 1)(b_j + 1)}{Q_{j-1} + 2}.$$

The formulas $(a_{ij} + 1)(b_j + 1)/(Q_{J-1} + 2)$ and $a_{ij}b_j/Q_{j-1}$ yield fractional numbers located on the numerical axis between the same integers, and the difference between them is negligible. In addition, the resulting values x_{ij} are rounded to the nearest integer. Hence, we write

$$\frac{a_{ij}b_j}{Q_{j-1}} - 1 \le x_{ij} \le \frac{a_{ij}b_j}{Q_{j-1}}.$$

According to this two-sided inequality,

$$x_{ij} = \frac{a_{ij}b_j}{Q_{j-1}}.$$

In this formula, the values a_{ij} and Q_{j-1} are unique, and b_j is the initial value. Therefore, the values x_{ij} will be unique for any i and j. Consequently, this formula provides a unique solution of the problem; see Table 1.

However, note that when calculating the elements x_{ij}, their rounded values to the nearest integer are taken as the only solution. In some rows (or columns) of the table, their sums may not coincide with the original entry (or exit) data. To maintain a balance in each row and column of the table, we propose the following calculation method. If $j = i$, then $x_{ij} = 0$, and the elements are located on the main diagonal in the table. If $j = n$, then the elements are located in the last column of the table: $x_{1,n}$, $x_{2,n}$, $x_{3,n}$, ..., $x_{n-1,n}$. In this case, they are calculated as follows: the sum of all elements is found, e.g., in *i*th row to the desired one, and then the desired one is determined as the difference between a_i and the sum:

$$x_{in} = a_i - \sum_{r=i}^{n-1} a_{ir}.$$

If $j = i + 1$, then the elements are located first above the diagonal in each column of the table: x_{12}, x_{23}, x_{34}, ..., $x_{n-1,n}$. In this case, they are calculated as follows: the sum of all elements is found, e.g., in *j*th column to the desired one, and then the desired one is determined as the difference between b_j and the sum:

$$x_{ij} = b_j - \sum_{r=1}^{j-2} x_{rj}.$$

With such modifications, the passenger transits along the route in the problem (1.1) and (1.2) are calculated by the general formula

$$x_{ij} = \begin{cases} 0 \quad \text{if} \quad i = j; \\ b_j - \sum_{r=1}^{j-2} x_{rj} \quad \text{if} \quad j = i + 1; \\ a_i - \sum_{r=i}^{j-1} x_{ir} \quad \text{if} \quad j = n; \\ \dfrac{a_{ij} b_j}{Q_{j-1}} \quad \text{for other} \quad i \quad \text{and} \quad j. \end{cases}$$

CONCLUSIONS

To check the calculation method proposed for the elements x_{ij}, we used the existing tables of the route correspondence of passenger flows by runs and hours. We took the entry and exit data from the available tables, found the route correspondence of passenger flows based on these tables using the method, and then compared with the tabular ones.

As indicated by the comparative analysis, the deviations between x_{ij} and the route correspondence of the passenger flows available from the tables have the following average fluctuations: (a) for the run elements x_{ij}, within the range 1.1–5.3%, with an average value of 3.2%; (b) for the hourly elements x_{ij}, within the range 5.8–11.6%, with an average value of 8.7%. Therefore, the accuracy of calculating the elements x_{ij} is higher for small values of the entry and exit data, which characterize weak passenger flows.

The importance of this conclusion must be considered when developing a computerized system of models and algorithms for processing transport survey data. For example, the technology of capturing the entry and exit data allows first acquiring information about the transits of passengers for each run. However, current and long-term planning mainly requires the information specified by hours of the day and the general daily information. In this case, the hourly elements x_{ij} can be determined in two ways:

(1) sum the entry and exit data of each run over the hour-long intervals and then calculate the elements x_{ij};

(2) calculate the elements x_{ij} based on each run's entry and exit data and then sum them up over the hour-long intervals.

The daily elements x_{ij} can be obtained in the same two ways.

Calculations by the first and second methods give different results, and the second one is more accurate. This fact can be explained as follows. First, each run's entry and exit data separately characterize weaker passenger flows than the initial hourly data used for determining the elements x_{ij}. Second, the second method gives more stable results than the first one under the assumption that any passenger will exit the vehicle's cabin with an equal probability regardless of his route.

Finally, the second method is preferable due to the misbalance between the entry and exit data in each hour. (Quite often, the beginning and end of the passenger's trip belong to different hour-long intervals.)

REFERENCES

1. M. R. Yakimov and O. V. Evseev, "Simulation models in forming efficient transport systems," Transp. Ros. Fed., No. 1 (79), 56–60 (2019).
2. K. P. Andreev and V. V. Terent'ev, "Passenger transportation and optimization of the urban route network," Mir Transp., No. 2, 159–161 (2017).
3. A. L. Nosov, "Quality control of urban passenger transport using a transport model," Logist. Segodnya, No. 1, 38–47 (2015).
4. Kh. A. Faskhiev, "An efficient urban bus: How to make a choice?," Logist. Segodnya, No. 4, 274–288 (2017).
5. V. N. Embulaev, O. G. Degtyareva, and N. P. Belozertseva, *A Systematic Approach in the Theory and Practice of Organizing Urban Passenger Transportation* (VGUES, Vladivostok, 2013) [in Russian].
6. K. P. Andreev, V. V. Terent'ev, and S. N. Kulik, "Survey of passenger traffic on city bus routes," Nov. Nauka: Probl. Perspekt., No. 2, 159–161 (2016).
7. V. E. Gmurman, *Fundamentals of Probability Theory and Mathematical Statistics* (Yurait, Moscow, 2016; Am. Elsevier, New York, 1968).

Translated by A. Mazurov