ISSN 2070-0482, Mathematical Models and Computer Simulations, 2020, Vol. 12, No. 4, pp. 590–596. © Pleiades Publishing, Ltd., 2020.
Russian Text © The Author(s), 2019, published in Matematicheskoe Modelirovanie, 2019, Vol. 31, No. 12, pp. 86–96.

Development of a Mathematical Model to Calculate Passenger Journeys on a Route According to the Entry and Exit Data

V. N. Embulaev*

Vladivostok State University of Economics and Service, Vladivostok, 690014 Russia

**e-mail: Vladimir.Embulaev@vvsu.ru*

Received April 15, 2019; revised April 15, 2019; accepted July 1, 2019

Abstract—The random variable of passenger journeys between any two stopping points on a route (passenger correspondence) represents a discrete variable. For each value of the discrete random variable, its probability is calculated. As a solution to the problem of determining the number of corresponding passengers between two stopping points on a route, we take the most probable value of the random variable.

Keywords: passenger flows, random variable, probability of event, law of distribution, most probable number

DOI: 10.1134/S2070048220040080

INTRODUCTION

Organization and administration of passenger transport on the routes of a major city are always difficult problems [1, 2]. However, according to experts, these problems can be solved by the most rational use of the available and ready-to-use transportation vehicles (TVs) based on the coordinated administration of all types of urban passenger transport in a major city [3, 4].

The analysis of the task of the coordinated administration of the transport system of a major city shows that various transport tasks of future, current, and operational administration are solved in the administration process [5]. Here, the entire information base is divided into three main arrays: conditional-permanent (data on the route network and the composition of TVs), regulatory (economic indicators), and variable (data on passenger flows and the quantitative and quality performance of the transport system determined from them). The main problem with the coordinated administration of the entire transport system of a major city is the process of receiving information on passenger flows (which change both in time and space) and therefore the reception of this information needs to be automatized and mechanized.

It is known that the greatest advances in automating and mechanizing the receipt of information on passenger flows are associated with technologically recording the data on the entry ($a_i, 1 \le i \le n$) and exit ($b_j, 1 \le j \le n$) of passengers at each stopping point (SP) on a route (where n is the number of SPs on the route) [6]. The task of entry–exit data processing with the receipt of information on the distribution of passenger flows on the route ($x_{ij}, i \le j$), i.e., determining the number of passengers that have travelled between each pair of SPs (i, j) on the route during, e.g., one run, has not been sufficiently studied. This defines the task of developing a mathematical model for calculating passenger travels between each pair of SPs on the route according to the entry and exit data.

2. FORMULATION OF THE PROBLEM

When collecting the data on the entry and exit at SPs of a route, the information of each individual passenger on the route remains unknown; therefore, such movements can be taken as random and independent of the choice of routes of other passengers. This allows us to make the following assumption: the choice made by each passenger on the path of travel (i, j) on the route (i.e., the passenger's entry into the interior of the TV at the ith SP and the exit at the jth SP) is random and does not depend on the choice of other passengers of their path of travel on the route. With this assumption we can assume that when a TV stops at an SP of the route, for each passenger, the event of getting off the TV at this SP or remain on it can be considered equally probable.

Table 1. Matrix of elements of passenger-flow correspondence

Entry SPs (nos.)	Exit SPs (nos.)							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
Got off	b_1	b_2	b_3	b_4	b_5	...	b_n	

Thus, solving the problem of determining elements of the route correspondence of passenger flows according to the entry–exit data requires a probabilistic interpretation. When the TV stops at the jth SP, first of all the Q_{j-1} passengers that have arrived from the preceding (j–1)th SP, b_j passengers get off the TV, and then a_j passengers enter the TV. When the TV departs from the jth SP, we have Q_j passengers in the interior; here,

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

When the TV stops at the ith SP, a_i passengers enter it; it is possible that some of them can exit at the (i + 1)th SP, and this corresponds to the number of passengers that travel from the ith to the (i + 1)th SP. Denote this number by $x_{i,i+1}$. If we subtract it from the number of passengers that enter at the ith SP, we obtain the number of the remaining passengers, which continue their travel on the route (i.e., potentially corresponding from the ith SP). Denote this number by $a_{i,i+1}$; it is determined as follows: $a_{i,i+1} = a_i - x_{i,i+1}$.

In general, for any i and j this number is calculated by the formula

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

here, for $j = i, a_{ij} = a_{ii} = a_i$.

In other words, a_{ij} is the number of remaining passengers from all those that entered the TV at the ith SP. It is determined by subtracting from a_i the number of passengers that have already completed their travel on the route $(i, i + 1) - x_{i,i+1}$, $(i, i + 2) - x_{i,i+2}$, ..., $(i, j - 1) - x_{i,j-1}$.

In the transportation process on the route, when the TV stops at the jth SP, we have Q_{j-1} passengers inside, also including persons that entered at the ith SP (group a_{ij}). As a result of the passenger exchange at an SP, together with b_j, also those passengers that entered the TV at the ith SP, i.e., from the group a_{ij}, could exit. In this case, the number x_{ij}, which belongs to a_{ij} and b_j at the same time, is specifically unknown.

If we determine in such a way all the journeys of passengers between any two SPs on a route, then these travels can be presented as a table of elements of the passengers' route-correspondence (see Table 1).

Here, n is the number of SPs on a route, a_i is the number of passengers that enter the TV at the ith SP, b_j is the number of passengers that get off the TV at the jth SP, and x_{ij} is the number of corresponding passengers from the ith to the jth SP ($i \le j$).

Table 1 shows that the summation of elements by rows corresponds to the entry data and the summation by columns corresponds to the exit data. In this respect, the mathematical formalization of the problem of determining x_{ij} by the given a_i and b_j consists of the following points:

we have the system of linear algebraic equations

$$\sum_{j=i}^{n} x_{ij} = a_i,\ \ \sum_{i=1}^{j} x_{ij} = b_j,\ \ x_{ij} \geq 0,\ \ 1 \leq i \leq n,\ \ 1 \leq j \leq n. \tag{1}$$

moreover, the condition

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} b_j \ \left(\text{or } \sum_{i=1}^{n}\sum_{j=i}^{n} x_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{j} x_{ij}\right) \tag{2}$$

is met.

Because system (1) consists of $2n$ equations with $n(n+1)/2$ unknowns, a unique solution is possible only if the condition $2n = n(n+1)/2$ is met. Hence, the unique solution is possible when $n \leq 3$, while when $n > 3$ there exist, in general, countless solutions, and therefore it is impossible to obtain an unambiguous solution to this problem without additional assumptions about the distribution of passenger flows between the SPs of a route.

3. PROBABILITY METHOD FOR SOLVING THE PROBLEM

Taking into account the accepted assumption relative to the behavior of passengers when choosing the ways of travel on the route, it is easy to notice that during the stop of the TV at the jth SP, there are Q_{j-1} passengers in the interior, and from among them b_j persons get off the TV. Since the exit of any group of b_j passengers (of all those that have arrived) from the TV represents an equiprobable and disjoint event, the total number of such groups in various combinations is equal to the number of combinations of Q_{j-1} elements taken b_j at a time; i.e., $C_{Q_{j-1}}^{b_j}$.

All the passengers who have arrived at the jth SP (Q_{j-1}) can be divided into two groups: the first group consists of passengers that enter the interior of the TV at the ith SP and do not exit before the jth SP (a_{ij}) and the second group consists of all the other passengers ($Q_{j-1} - a_{ij}$). Since at the jth SP, b_j passengers exit, they can include persons that belong to group a_{ij}; denote them by λ_{ij}. It is obvious that λ_{ij} is a random variable, which takes integer values. Here, the group of λ_{ij} persons from a_{ij} can be taken in $C_{a_{ij}}^{\lambda_{ij}}$ ways in various combinations. However, for each certain group of λ_{ij}, the remaining passengers ($b_j - \lambda_{ij}$) here can be taken in $C_{Q_{j-1}-a_{ij}}^{b_j-\lambda_{ij}}$ different ways. Then the total number of conducive cases is $C_{a_{ij}}^{\lambda_{ij}} C_{Q_{j-1}-a_{ij}}^{b_j-\lambda_{ij}}$.

Based on the above, it is easy to determine the probability that of the b_j passengers that got off, exactly λ_{ij} persons belong to group a_{ij}:

$$P_{b_j}(\lambda_{ij}) = \frac{C_{a_{ij}}^{\lambda_{ij}} C_{Q_{j-1}-a_{ij}}^{b_j-\lambda_{ij}}}{C_{Q_{j-1}}^{b_j}}. \tag{3}$$

Here, we have two possible cases: (1) where $a_{ij} \geq b_j$ and (2) where $a_{ij} \leq b_j$.

In the case $a_{ij} \geq b_j$, the following situations are possible (and the corresponding number of their combinations):

$$\begin{array}{llll} b_j \in a_{ij} & \text{and } 0 \in (Q_{j-1} - a_{ij}), & C_{a_{ij}}^{b_j} & \text{and } C_{Q_{j-1}-a_{ij}}^{0}, \\ (b_j - 1) \in a_{ij} & \text{and } 1 \in (Q_{j-1} - a_{ij}), & C_{a_{ij}}^{b_j-1} & \text{and } C_{Q_{j-1}-a_{ij}}^{1}, \\ & \vdots & & \vdots \\ (b_j - k) \in a_{ij} & \text{and } k \in (Q_{j-1} - a_{ij}), & C_{a_{ij}}^{b_j-k} & \text{and } C_{Q_{j-1}-a_{ij}}^{k}, \end{array}$$

where $k = \min[b_j, (Q_{j-1} - a_{ij})]$.

In this case, λ_{ij} can be integer over the interval $(b_j - k) \le \lambda_{ij} \le b_j$. Since

$$(b_j - k) = (b_j - \min[b_j, (Q_{j-1} - a_{ij})]) = \max[(b_j - b_j), (b_j - Q_{j-1} + a_{ij})] = \max[0, (b_j + a_{ij} - Q_{j-1})],$$

the variation interval of the random variable λ_{ij} for $a_{ij} \ge b_j$ is

$$\max[0, (b_j + a_{ij} - Q_{j-1})] \le \lambda_{ij} \le b_j. \tag{4}$$

In the case $a_{ij} \le b_j$, the following situations are possible (and the corresponding number of their combinations):

$$\begin{array}{llll}
a_{ij} \in a_{ij} & \text{and} \quad (b_j - a_{ij}) \in (Q_{j-1} - a_{ij}), & C_{a_{ij}}^{a_{ij}} & \text{and} \ C_{Q_{j-1}-a_{ij}}^{b_j - a_{ij}}, \\
(a_{ij} - 1) \in a_{ij} & \text{and} \quad (b_j - a_{ij} + 1) \in (Q_{j-1} - a_{ij}), & C_{a_{ij}}^{a_{ij}-1} & \text{and} \ C_{Q_{j-1}-a_{ij}}^{b_j - a_{ij}+1}, \\
\vdots & & \vdots & \\
(a_{ij} - k) \in a_{ij} & \text{and} \quad (b_j - a_{ij} + k) \in (Q_{j-1} - a_{ij}), & C_{a_{ij}}^{a_{ij}-k} & \text{and} \ C_{Q_{j-1}-a_{ij}}^{b_j - a_{ij}+k},
\end{array}$$

where $k = \min[a_{ij}, (Q_{j-1} - a_{ij}) - (b_j - a_{ij})] = \min[a_{ij}, (Q_{j-1} - b_j)]$.

In this case, λ_{ij} can be integer over the interval $(a_{ij} - k) \le \lambda_{ij} \le a_{ij}$. Since

$$(a_{ij} - k) = (a_{ij} - \min[a_{ij}, (Q_{j-1} - b_j)]) = \max[(a_{ij} - a_{ij}), (a_{ij} - Q_{j-1} + b_j)] = \max[0, (b_j + a_{ij} - Q_{j-1})],$$

the variation interval of the random variable λ_{ij} for $a_{ij} \le b_j$ is

$$\max[0, (b_j + a_{ij} - Q_{j-1})] \le \lambda_{ij} \le a_{ij}. \tag{5}$$

Expressions (4) and (5) can be written in the general form

$$\max[0, (b_j + a_{ij} - Q_{j-1})] \le \lambda_{ij} \le \min[a_{ij}, b_j]. \tag{6}$$

In this case, of all integers λ_{ij} in segment (6), as the unique unknown x_{ij}, we can take a value for which the probability calculated by formula (3) reaches its maximum value in the argument λ_{ij}:

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

Because x_{ij} is the most probable number for the random variable λ_{ij}, the following inequalities must hold for two standing numbers $(x_{ij} - 1)$ and $(x_{ij} + 1)$ [7]:

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

We expand these inequalities and obtain the expressions

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

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

The solution of these inequalities for x_{ij} yields the two-sided inequality

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

For large a_{ij}, b_j, and Q_{j-1}, we can omit 1 and 2 in these inequalities:

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

From this double inequality it follows that in order to calculate the elements x_{ij}, the following formula can be used (rounding the values to an integer):

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

In general, the mathematical model of calculating x_{ij}, with allowance for the specifics of generating a table of elements of the route correspondence of passenger flows (see Table 1), is presented as follows:

$$x_{ij} = \begin{cases} 0, \text{ if } i = j, \\ b_j - \sum_{r=1}^{j-2} x_{rj}, \text{ if } j = i+1, \\ a_i - \sum_{r=i}^{j-1} x_{ir}, \text{ if } j = n, \\ (a_{ij}b_j)/Q_{j-1} \text{ for all others } i \text{ and } j. \end{cases}$$

4. SOLUTION USING THE BERNOULLI SCHEME

From the accepted assumption associated with the behavior of passengers in choosing the travel path on the route, it follows that when the TV approaches the jth SP, for each person of the passengers onboard, two events (get off the TV at this SP, or move on) are taken to be equally probable. Hence, for each passenger, the probability of the event of getting off the TV at the jth SP is $p = 1/2$ (and, respectively, the probability of the event of not getting off, i.e., travelling further is $q = 1/2$). Then, using the Bernoulli formula, we can determine the probability that at a particular SP, λ_{ij} passengers among the total number of passengers onboard will get off the TV:

$$P_m(\lambda_{ij}) = C_m^{\lambda_{ij}} p^{\lambda_{ij}} q^{m-\lambda_{ij}}, \tag{7}$$

where m is the number of integers in segment (6); this number can be determined by calculating the length of the given segment, i.e.,

$$m = \min[a_{ij},\ b_j] - \max[0, (a_{ij} + b_j - Q_{j-1})]. \tag{8}$$

Because in this case $p = q$, formula (7) can be written as follows: $P_m(\lambda_{ij}) = C_m^{\lambda_{ij}} p^{\lambda_{ij}} p^{m-\lambda_{ij}} = C_m^{\lambda_{ij}} p^m$. This means that the numerical probability of the random variable λ_{ij} in segment (6) depends mainly on the binomial factor $C_m^{\lambda_{ij}}$. Hence, the value of λ_{ij} for which the binomial factor $C_m^{\lambda_{ij}}$ is the largest in the argument λ_{ij} can be taken as the solution, i.e.,

$$x_{ij} = \max_{\lambda_{ij}} \frac{m!}{\lambda_{ij}!(m-\lambda_{ij})!}.$$

According to the theory of combinatorics, this value for $p = q$ reaches its maximum in the middle of segment (6):

$$x_{ij} = (\min[a_{ij}, b_j] + \max[0, (a_{ij} + b_j - Q_{j-1})])/2. \tag{9}$$

However, it is known that the most probable number λ_{ij} in solving problems using the Bernoulli scheme can be determined from the double inequality $mp - p \le \lambda_{ij} \le mp + q$ [7]. Since $p = q$, this double inequality can be presented as $mp - p \le \lambda_{ij} \le mp + p$ or $|\lambda_{ij} - mp| \le p$; i.e., the difference between λ_{ij} and mp must not exceed $1/2$ in absolute magnitude. This means that the most probable number λ_{ij} in segment (6) can be determined using the expression of mp, which is rounded to the nearest integer. With allowance for formula (8) and the relation $p = 1/2$, we have

$$x_{ij} = mp = (\min[a_{ij}, b_j] - \max[0, (a_{ij} + b_j - Q_{j-1})])/2. \tag{10}$$

From the analysis of expressions (9) and (10), it follows that their right sides can be equal if and only if the following equality holds: $\max[0, (a_{ij} + b_j - Q_{j-1})] = 0$; i.e., $(a_{ij} + b_j - Q_{j-1})$ must not be greater than

Table 2. Matrix of elements of fixed-route passenger-flow correspondence (values of x_{ij})

Entry SPs (nos.)	Exit SPs (nos.)										Entered
	1	2	3	4	5	6	7	8	9	10	
1	0	9	6	4	5	3	10	5	2	2	46
2		0	5	3	5	2	4	2	1	0	22
3			0	0	0	0	1	1	0	0	2
4				0	0	0	3	2	1	0	6
5					0	0	2	1	1	0	4
6						0	1	7	3	3	14
7							0	10	4	1	15
8								0	13	0	13
9									0	11	11
10										0	0
Got off	0	9	11	7	10	5	21	28	25	17	133

Table 3. Matrix of elements of potentially corresponding passengers between two SPs on a route (values of a_{ij})

Entry SPs (nos.)	Exit SPs (nos.)									
	1	2	3	4	5	6	7	8	9	10
1	46	37	31	27	22	19	9	4	2	0
2		22	17	14	9	7	3	1	0	0
3			2	2	2	2	1	0	0	0
4				6	6	6	3	1	0	0
5					4	4	2	1	0	0
6						14	13	6	3	0
7							15	5	1	0
8								13	0	0
9									11	0
10										0

Table 4. Matrix of values determined by calculating interval of possible solutions (values of z_{ij})

Entry SPs (nos.)	Exit SPs (nos.)									
	1	2	3	4	5	6	7	8	9	10
1	–	–	−11	−12	−12	−16	−12	−9	−2	–
2		–	–	−26	−25	−29	−24	−16	−5	–
3			–	–	−37	−36	−29	−19	−6	–
4				–	–	−32	−25	−15	−5	–
5					–	–	−27	−16	−5	–
6						–	–	−5	0	–
7							–	–	−1	–
8								–	–	–
9									–	–
10										–

zero. If that is so, then the segment of possible solutions (6) of problem (1) and (2) is presented as $0 \le \lambda_{ij} \le \min[a_{ij}, b_j]$.

CONCLUSIONS

To make sure that the left boundary of the segment of possible solutions (6) of problem (1) and (2) is zero, consider the following example. Assume that as a result of fixing the entry–exit data on a route with ten SPs, the following source data is obtained: $a_i = \{46; 22; 2; 6; 4; 14; 15; 13; 11; 0\}$ and $b_j = \{0; 9; 11; 7; 10; 5; 21; 28; 25; 17\}$, respectively.

Since any integer from interval (6) is a solution to problem (1) and (2), in order to verify the validity of the result, we first calculate values of x_{ij} using the mathematical model proposed in this paper. In the end, we obtain the numerical values of the elements of the route passenger-flow correspondence; see Table 2.

Using the formula $a_{ij} = a_i - \sum_{r=i+1}^{j-1} x_{ir}$, we determine the numerical values of the elements of potentially corresponding passengers between any two SPs on the route; see Table 3.

Finally, calculate the values of $z_{ij} = a_{ij} + b_j - Q_{j-1}$ only for those i and j for which the calculation of elements of passenger correspondence according to the algorithm described in this paper is made by the formula $x_{ij} = (a_{ij} b_j)/Q_{j-1}$. These values are presented in Table 4, which shows that all $z_{ij} \le 0$, which is what we needed to confirm.

REFERENCES

1. V. A. Fedorov, "Modern tasks and problems of field surveys of passenger flows (on the example of St. Petersburg)," Molod. Uchen., No. 2, 332–342 (2015).
2. K. P. Andreev and V. V. Terentev, "Passenger transportation and optimization of the urban route network," Mir Transp., No. 2, 159–161 (2017).
3. A. P. Nosov, "Managing the quality of urban passenger transport using a transport model," Logist. Segodnya, No. **1**, 38–47 (2015).
4. Kh. A. Faskhiev, "An efficient bus for urban transport: 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 to the Theory and Practice of Organizing Urban Passenger Traffic* (VGUES, Vladivostok, 2013) [in Russian].
6. K. P. Andreev, V. V. Terentev, and S. N. Kulik, "Passenger traffic survey on city bus routes," Nov. Nauka: Probl. Perspekt., No. 2, 159–161 (2016).
7. V. E. Gmurman, *Theory of Probability and Mathematical Statistics* (Yurait, Moscow, 2019) [in Russian].

Translated by L. Kartvelishvili

SPELL: 1. OK