

ON THE DISTRIBUTION OF THE RANDOM ARRIVAL TIME

BORIS DAVYDOV¹, VLADIMIR CHEBOTAREV² & KSENIYA KABLUKOVA²

¹FESTU, Russia

²CC FEB RAS, Russia

ABSTRACT

This paper proposes the generalised model for formation the patterns of train arrival time distribution. The model provides that the arrival time has a random component. A feature of the model is the allowance for the stochastic dependence of travel interval on train departure time deviation. Within the framework of the model, a formula for the distribution function of the arrival time is obtained. The paper describes the method for checking the degree of stochastic relationship of the above random variables. Theoretical formulations are confirmed by analysis of statistical data reflected by the passenger train traffic on Russian railways.

Keywords: train traffic, stochastic model, operation time distribution, statistical dependency.

1 INTRODUCTION

Stochastic models are used for probabilistic forecasting of the traffic breakdowns when random disturbances occur (see, for example, [1] and [2]). They make it possible to identify intertrain conflicts and predict their further propagation over the railway network. The models proposed allow you to eliminate the delays or to reduce their duration. Distributions of the following random variables such as departure moment, travel time, headway, etc. can be considered as input data of such a model. Stochastic simulation results in the arrival delay probability distribution such as functional relation from the input random variables. In turn, knowledge of the delay distribution is the basis for predicting the delay duration as well as a set of measures to eliminate it.

It is usually assumed that the type of input distributions is standard when arrival delay function calculating. As a rule there are used an exponential distribution function or its modifications in a majority of papers. Such the model describes only particular cases and do not cover other possible variants of input distributions observed in reality. Our study shows that a more general approach is to represent the input random variables using the more comprehensive class of probability distribution.

This article proposes the stochastic model of the train run along some railway section which includes several stations. The input data are the distributions of departure moments and travel times. The result of simulation process is the set of distributions the arrival time deviation which is obtained for each station. The analytical method assumes that there are no restrictions on the form and the type of input distributions.

The statement of the problem analysis arose as a result of studying the real statistics of train traffic on the Russian railways. The research has shown the need to consider two cases of formation the output densities. The first case relies on the situation when the input random variables (departure time and travel time) are independent. In this case, the output distribution is a convolution of the input distributions. In the second case, a dependence of input random variables is observed (the travel time depends on the departure time). Based on the available statistics we create the output distribution only for those stations at which the input random variables are linearly dependent. The analysis is shown how well the output distribution of arrival times is consistent with the real statistics.



An ability to use arbitrary input distributions expands the possibilities of applying the model to various situations with different types of input random variables behave differently.

2 LITERATURE REVIEW

Modeling the process of random delays occurrence and their propagation along the train chain makes it possible to assign rational traffic adjustments and predict their likely consequences. The analytical method which uses convolutions of distributions of the initial (at the site entry) and newly arising deviations to obtain the delay density is proposed in Muhlhans [3]. Applications of this approach are limited since only the uniform distribution the secondary delays is used. In reality, this assumption is rarely confirmed. One effective approach which allows us to predict statistical characteristics of the train traffic is Situational-Heuristic Method (SEMN) [4]. The expected values of operating times are obtained from statistics of previous periods taking the current on-site situation into account. A similar approach is used in Karetnikov [5] which models the local traffic using average values of train processing time. It should be noted that there are attempts to derive some average train schedule based on the emerging situation. Such the attempts are fraught with conservation of technological problems that exist on the site.

One of the fundamental papers is devoted to the generic problem of traffic stochastic modeling under consideration [1]. The authors describe a model of delay formation in a train set using a probabilistic approach. Total run time of the train is considered as a sum of random running intervals along the section when analyzing the process of delay propagation. Buker and Seybold [6] show that calculation of the arrival time distribution at the terminal station is based on repeated use of convoluting the distributions of random variables characterizing the elementary operations.

Below mentioned papers use this approach to analyze the process of delay propagation along the train chain. Boucherie and Huisman [7] solve the problem of interacting the two trains that have different speeds and are influenced by random impacts. The result of analysis is used to model the train traffic in a dense heterogeneous train flow. This idea is developed further in the papers [2], [6], [8]. Description of the traffic process [2] occupies an intermediate position between the macro-models and models with very detailed description of the process, i.e. micro-models. Approximation method is proposed for exact representation of delay distributions. Cumulative distribution is calculated as the result of sequence the elementary activities; the activity chain is determined by a stochastic event-graph.

The problem of predicting the random delay appearance and propagation across the large railway network most fully is considered in Berger et al. [8]. Scattering of arrivals is treated as a random variable generated with joint accounting the departure times and the running intervals. The main subject of research is the process of adjusting the schedule at transfer stations in the presence of multiple delays. The model operates with discrete distributions of travel, dwell times and deviations from the schedule. The authors use a simplified description of input random variables so that they receive a small computing time.

The proposed approach was further developed in Buker and Seybold [6] which attempts to take the realistic distribution of operation time into account and choose the appropriate approximation. The framework proposed uses a probabilistic operational graph which considers passenger transfer operations and conflict situations. The authors argue that mesoscopic modelling of traffic is the purposeful approach to compute the delay propagation. The model considers only station and crossing as the operating points where conflicts can occur. The purpose of the study is to reduce computing time.

The general characteristic of the schedule is the ability to absorb small current deviations [9], [10]. The process of random deviations calculation is the basis of the methodology for



determining corrective actions that reduce the time required to restore the normative schedule [11].

3 A GENERALIZED MODEL FOR THE FORMATION OF THE DISTRIBUTION OF THE ARRIVAL TIME

In this section we propose a model for the formation of the arrival times taking into account that the departure moments, travel times and possible delays can have the random nature. Random trajectories of train run are depicted in Fig. 1 as “beams” of segments emanating from the stations S_1 and S_2 .

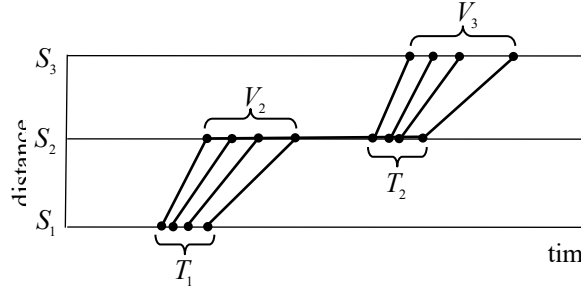


Figure 1: Scheme of the formation of arrival moments.

In Fig. 1 symbols T_i , $i = 1, 2$, denote the departure times from stations S_i respectively, and symbols V_k , $k = 2, 3$, the arrival times at stations S_k respectively. For each $i = 1, 2$, the values of T_i are scattered, and to emphasize this, some possible values of T_i are indicated by means of a brace. Similarly, for each $k = 2, 3$, some possible values of V_k are indicated by means of a brace too.

To describe the proposed model, we need the following notations:

d_i is the scheduled departure time from the station S_i , $i = \overline{1, N-1}$, where N is the total number of stations; $\delta_i = T_i - d_i$;

$\rho_{i,j+1}$ is the travel time from station S_i to station S_{i+1} ;

a_i is the scheduled arrival time at the station S_i , $i = \overline{2, N}$; $\xi_i = V_i - a_i$;

τ_i is the stop duration at station S_i , $i = \overline{2, N-1}$.

Distribution functions of random variables T_i , $\rho_{i,j+1}$ and V_i are denoted by $F_i(t)$, $L_{i,j+1}(t)$ and $G_i(t)$, respectively, and are defined by the following equalities: $F_i(t) = P(T_i < t)$, $L_{i,j+1}(t) = P(\rho_{i,j+1} < t)$ and $G_i(t) = P(V_i < t)$.

Corresponding density functions, if they exist, will be denoted by the same letters as distribution functions, only lowercase ones: $f_i(t)$, $l_{i,j+1}(t)$ and $g_i(t)$. Fig. 2 shows a scheme of the formation of the following random variables: V_i , T_i and V_{i+1} .

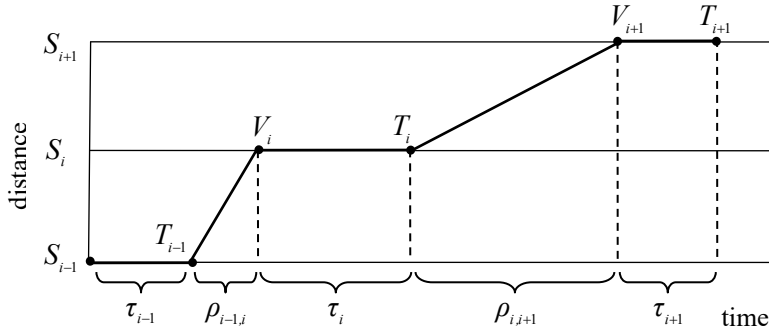


Figure 2: Scheme of the formation of the random variables: V_i , T_i and V_{i+1} .

According to this scheme, in particular, we have,

$$V_{i+1} = T_i + \rho_{i,i+1}. \quad (1)$$

Now we will show that the following equation is valid for the distribution function of V_{i+1} :

$$G_{i+1}(t) = \int_{-\infty}^{\infty} P(\rho_{i,i+1} < t - x | T_i = x) dF_i(x). \quad (2)$$

We will use nonstrict reasoning for the derivation of (2) not to clutter up the paper.

First, suppose that for some constants $a < b$ the equality $P(a \leq T_i \leq b) = 1$ holds. Divide the segment $[a, b]$ into n equal parts by points: $a = x_0 < x_1 < \dots < x_n = b$. Using the well-known total probability equation we obtain from (1) that for each $i = 1, \overline{N-1}$,

$$\begin{aligned} P(V_{i+1} < t) &= P(T_i + \rho_{i,i+1} < t) \\ &= \sum_{\{x_k\}} P(x_k \leq T_i < x_{k+1}) P(\rho_{i,i+1} < t - T_i | x_k \leq T_i < x_{k+1}), \end{aligned} \quad (3)$$

where $P(\rho_{i,i+1} < t - T_i | x_k \leq T_i < x_{k+1})$ is the conditional probability of event $\{\rho_{i,i+1} < t - T_i\}$, provided that event $\{x_k \leq T_i < x_{k+1}\}$ has occurred. Let us consider in eqn (3) the following term:

$$P(x_k \leq T_i < x_{k+1}) P(\rho_{i,i+1} < t - T_i | x_k \leq T_i < x_{k+1}).$$

At first, $P(x_k \leq T_i < x_{k+1}) = F_i(x_{k+1}) - F_i(x_k)$. In addition, assuming that the difference $x_{k+1} - x_k$ is small enough, we have the approximate equality $T_i \approx x_k$, since $x_k \leq T_i < x_{k+1}$. Hence, rearranging the factors in (3), we obtain the equality

$$G_{i+1}(t) \approx \sum_{\{x_k\}} P(\rho_{i,j+1} < t - x_k \mid x_k \leq T_i < x_{k+1}) [F_i(x_{k+1}) - F_i(x_k)],$$

the right hand-side of which is the Stieltjes sum. When infinitely grinding the segment $[a, b]$ this sum converges to the Stieltjes integral

$$\int_a^b P(\rho_{i,j+1} < t - x \mid T_i = x) dF_i(x).$$

Letting that a tends to $-\infty$ and b tends to $+\infty$, we obtain eqn (2).

If density function $f_i(x)$ exists, then

$$G_{i+1}(t) = \int_{-\infty}^{\infty} P(\rho_{i,j+1} < t - x \mid T_i = x) f_i(x) dx.$$

In addition, if the conditional density $\frac{d}{dt} P(\rho_{i,j+1} < t - x \mid T_i = x)$ exists, then density $g_{i+1}(t)$ exists and is equal to

$$g_{i+1}(t) = \int_{-\infty}^{\infty} \frac{d}{dt} P(\rho_{i,j+1} < t - x \mid T_i = x) f_i(x) dx.$$

The conditional distribution function $P(\rho_{i,j+1} < t - x \mid T_i = x)$ is sometimes written differently. Let $f_i(x, y)$ be the density of the joint distribution of T_i and $\rho_{i,j+1}$. Let x be such that $f_i(x) \neq 0$. For every such x the conditional distribution function of the random variable $\rho_{i,j+1}$ relative to the random variable T_i is defined as

$$L_{i,j+1}(y \mid x) := P(\rho_{i,j+1} < y \mid T_i = x) = \frac{1}{f_i(x)} \int_{-\infty}^y f_i(x, u) du.$$

Thus,

$$G_{i+1}(t) = \int_{-\infty}^{\infty} L_{i,j+1}(t - x \mid x) f_i(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{t-x} f_i(x, u) du dx. \quad (4)$$

Note that $l_{i,j+1}(u \mid x) := \frac{f_i(x, u)}{f_i(x)}$ is called the conditional density function of $\rho_{i,j+1}$

relatively to T_i .

Eqns (2) and (4) contain the conditional distribution that allows one to take into account the dependence of $\rho_{i,j+1}$ on T_i . These eqns are universal, but their application requires additional study of the joint distribution of random variables T_i and $\rho_{i,j+1}$.

If random values T_i and $\rho_{i,j+1}$ are independent, then (2) takes the form of convolution:

$$G_{i+1}(t) = (F_i * L_{i,j+1})(t) \equiv \int_{-\infty}^{\infty} L_{i,j+1}(t - x) dF_i(x) = \int_{-\infty}^{\infty} L_{i,j+1}(t - x) f_i(x) dx. \quad (5)$$

To obtain equalities for density $g_{i+1}(t)$, it is necessary to replace the distribution functions in (5) with the corresponding densities. Note that the eqn for the moments of

arrival/departure in the discrete case is presented in Berger et al. [8]. This eqn, although in other notation, is a special case of equality (2).

Remark: let us say a few words about question of the existence of the Stieltjes integral in (2). According to Fichtenholz [12, p. 92] this integral exists if the conditional distribution function $P(\rho_{i,j+1} < t - x | T_i = x)$ is continuous with respect to the argument x .

4 A METHOD FOR CHECKING THE DEGREE OF STOCHASTIC DEPENDENCE BETWEEN INPUT VARIABLES

The stochastic model described in Section 3 operates with random variables T_i , $\rho_{i,j+1}$ and V_{i+1} for which equality (1) holds. Recall the following property of variances: two random variables are uncorrelated if and only if the variance of their sum is equal to the sum of the variances. Using this property and equality (1), one can check whether random variables T_i and $\rho_{i,j+1}$ are correlated or not.

Suppose that for each station S_i we have a three-dimensional sample of volume n , from the distribution of the three-dimensional random vector $(T_i, \rho_{i,j+1}, V_{i+1})$. We will denote by symbol $D^*[X]$ the sample variance of a random variable X (there is mean the asymptotically unbiased estimate):

$$D^*[X] = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2,$$

where x_i are the sample values, \bar{X} is the sample mean. For each fixed $1 \leq i \leq N$, we will check the equality

$$D^*[V_{i+1}] = D^*[T_i] + D^*[\rho_{i,j+1}], \quad (6)$$

admitting an approximate equality in (6). In the case when equality (6) is fulfilled at least approximately, we will test the hypothesis that the correlation coefficient for the random variables T_i and $\rho_{i,j+1}$ is equal to zero, using Student's t -test. To this end, we calculate the value of the following statistic:

$$T^* = \frac{r^* \sqrt{n-2}}{\sqrt{1-r^{*2}}}, \quad (7)$$

where r^* is the sample correlation coefficient, i.e.

$$r^* = \frac{\text{cov}^*[T_i, \rho_{i,j+1}]}{\sqrt{D^*[T_i] D^*[\rho_{i,j+1}]}} = \frac{\sum_{j=1}^n (T_{ij} - \bar{T}_i)(\rho_{(i,j+1)j} - \bar{\rho}_{i,j+1})}{\sqrt{\sum_{j=1}^n (T_{ij} - \bar{T}_i)^2 \sum_{j=1}^n (\rho_{(i,j+1)j} - \bar{\rho}_{i,j+1})^2}}, \quad (8)$$

where $\text{cov}^*[\cdot, \cdot]$ denotes the sample covariance, \bar{T}_i and $\bar{\rho}_{i,i+1}$ are the sample means, T_{ij} denote sample values of T_i , and $\rho_{(i,i+1)j}$ sample values of $\rho_{(i,i+1)}$, $1 \leq j \leq n$; $1 \leq i \leq N$. Note that for Student's statistic we use the traditional notation T^* .

Let r be the Pearson correlation coefficient. The hypothesis $r = 0$ is considered as not contradict to experimental data with a significance level α if $|T^*| < t_{\text{crit}}(\alpha, n-2)$. The critical value $t_{\text{crit}}(\alpha, n-2)$ is taken from the Student's distribution table.

If the hypothesis $r = 0$ is not rejected for the random variables T_i and $\rho_{i,i+1}$, then we can assume their independence (although it is well known that the non-correlation of random variables does not mean their independence). In this case, to find the distribution function (and, as a consequence, the density function) of the arrival times, one should use eqn (5) containing the convolution operation.

In the opposite case, if the hypothesis $r = 0$ is rejected, and the equality of the sample variances (6) is not satisfied, we make an assumption about the stochastic dependence of $\rho_{i,i+1}$ on T_i . If the modulus of the sample coefficient of correlation is close to 1, then we can consider that random variables $\rho_{i,i+1}$ and T_i are linearly dependent, i.e.

$$\rho_{i,i+1} = aT_i + b, \quad (9)$$

where a and b are the coefficients of linear regression and are calculated by the known eqns:

$$a = r^* \frac{\sqrt{D^*[\rho_{i,i+1}]}}{\sqrt{D^*[T_i]}} = \frac{\sum_{j=1}^n (T_{ij} - \bar{T}_i)(\rho_{(i,i+1)j} - \bar{\rho}_{i,i+1})}{\sum_{j=1}^n (T_{ij} - \bar{T}_i)^2}, \quad b = \bar{\rho}_{i,i+1} - a\bar{T}_i. \quad (10)$$

We proceed to check the dependence of the random variables $\rho_{i,i+1}$ and T_i on the example of a specific section of the railway.

5 CHECKING THE DEPENDENCE OF INPUT RANDOM VARIABLES BY USING STATISTICAL DATA

Consider the real section of the "Moscow-Tver" suburban line with intense train traffic. The line is limited by the terminals "Moscow" (point 1) and "Tver" (point 13) and includes 11 intermediate stations. The average travel time at the line is about three hours. Statistics are collected for departure, travel and arrival times for morning rush hours. Immediately, we note that all sample values are expressed in minutes. In addition, samples for departure and arrival times are shifted to the left by the minimum observed sample value of departure time from station 1. This is done in order to take the zero time point as the departure time from station 1.

Let us find out whether equality (6) holds for some stations of the considered railway. Table 1 contains the calculated values of the sample variances $D^*[V_{i+1}]$, $D^*[T_i]$ and



$D^*[\rho_{i,i+1}]$, the sum $D^*[T_i] + D^*[\rho_{i,i+1}]$, as well as the values of the sample correlation coefficient r^* for monitoring calculations. Using data provided in Table 1 one can see that the independence of the quantities $\rho_{i,i+1}$ and T_i is possible for $i = 3, 4, 12$.

Next, we test the hypothesis $r = 0$ by applying Student's t -test. To this end, we calculate the sample correlation coefficients (8) and statistic (7). The hypothesis is tested with two significance levels, $\alpha_1 = 0.05$ and $\alpha_2 = 0.1$. The calculation results are presented in Table 2.

Table 1: Verification of statistical dependence the input variables (eqn (6)).

i	$D^*[V_{i+1}]$	$D^*[T_i]$	$D^*[\rho_{i,i+1}]$	$D^*[T_i] + D^*[\rho_{i,i+1}]$	r^*
1	23.9097	25.0556	0.6875	25.7431	-0.2209
2	29.7431	23.9097	0.5556	24.4653	0.7241
3	31.0764	30.0556	0.2431	30.2986	0.1439
4	28.6667	28.0764	0.5764	28.6528	0.0017
5	17.5764	24.8056	0.8542	25.6597	-0.8780
6	35.6389	17.2500	4.7222	21.9722	0.7571
7	34.5208	35.6389	0.2431	35.8819	-0.2312
8	41.0000	33.7431	2.2431	35.9861	0.2882
9	24.7934	31.7025	0.6281	32.3306	-0.8445
10	0.5600	0.8100	0.2900	1.1000	-0.5571
11	0.2400	0.4900	0.2100	0.7000	-0.7170
12	1.6100	0.5600	1.2500	1.8100	-0.1195

Table 2: The results obtained by applying Student's t -test.

i	r^*	n	T^*	$t_{\text{crit}}(\alpha_1, n-2)$	Conclusion with $\alpha_1 = 0.05$	$t_{\text{crit}}(\alpha_2, n-2)$	Conclusion with $\alpha_2 = 0.1$
1	-0.2209	12	-0.7161	2.2281	1: accepted	1.8125	1
2	0.7241	12	3.3196	2.2281	0: rejected	1.8125	0
3	0.1439	12	0.4598	2.2281	1	1.8125	1
4	0.0017	12	0.0055	2.2281	1	1.8125	1
5	-0.8780	12	-5.8017	2.2281	0	1.8125	0
6	0.7571	12	3.6650	2.2281	0	1.8125	0
7	-0.2312	12	-0.7516	2.2281	1	1.8125	1
8	0.2882	12	0.9516	2.2281	1	1.8125	1
9	-0.8445	11	-4.7313	2.2622	0	1.8331	0
10	-0.5571	10	-1.8974	2.3060	1	1.8595	0
11	-0.7170	10	-2.9093	2.3060	0	1.8595	0
12	-0.1195	10	-0.3405	2.3060	1	1.8595	1



According to presented data, it can be assumed that at stations 4 and 12, the independence of random variables $\rho_{i,i+1}$ and T_i ($i = 4, 12$) takes place. To find the distribution density function $g_{i+1}(t)$, it is required to use (5).

We can see in Table 2 that for $i = 5$ and 9, the coefficient r^* is close to (-1). In these cases, we find out the linear dependence (9). Then the distribution function of the arrival time can be written as follows:

$$G_{i+1}(t) = P(T_i + aT_i + b < t) = P\left(T_i < \frac{t-b}{1+a}\right) = F_i\left(\frac{t-b}{1+a}\right). \quad (11)$$

In the next section we will obtain the density functions for the arrival times and check the degree of their consistency with real-world data.

6 COMPARISON OF THE MODEL WITH REAL STATISTICS EXAMPLES

Let us verify the consistency of the model with real data. We begin the verification with the case when $\rho_{i,i+1}$ and T_i are independent. In section 5, we found out that there is a reason that for $i = 4, 12$ random variables $\rho_{i,i+1}$ and T_i are independent. Therefore, for these stations we will check how well the result obtained after applying the equality, similar to (5), for density functions will be consistent with real data. The results of the verification are presented in Table 3, which contains histograms based on statistics for departure times T_4 , T_{12} (row 1) and travel times $\rho_{4,5}$, $\rho_{12,13}$ (row 2). We approximate these histograms by the following density functions:

$$f_4(t) = I(t \geq 32) 0.4 e^{-0.4(t-32)}, \quad f_{12}(t) = I(t > 143) 10 e^{-2.9(t-143)} (t-143)^{1.4},$$

$$l_{4,5}(t) = I(t > 11.7) 35.3 e^{-3.8(t-11.7)} (t-11.7)^{2.75} \text{ and}$$

$$l_{12,13}(t) = I(27.5 \leq t \leq 32.5) 0.3564 e^{-0.36(t-29.5)^2}, \text{ respectively.}$$

Let us explain what $I(\cdot)$ means. The multiplier $I(\cdot)$ is an indicator, i.e.

$$I(x \in A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases}$$

where A is an arbitrary set. The functions $g_5(t)$ and $g_{13}(t)$ are determined by using an analog of eqn (5) for densities, i.e.

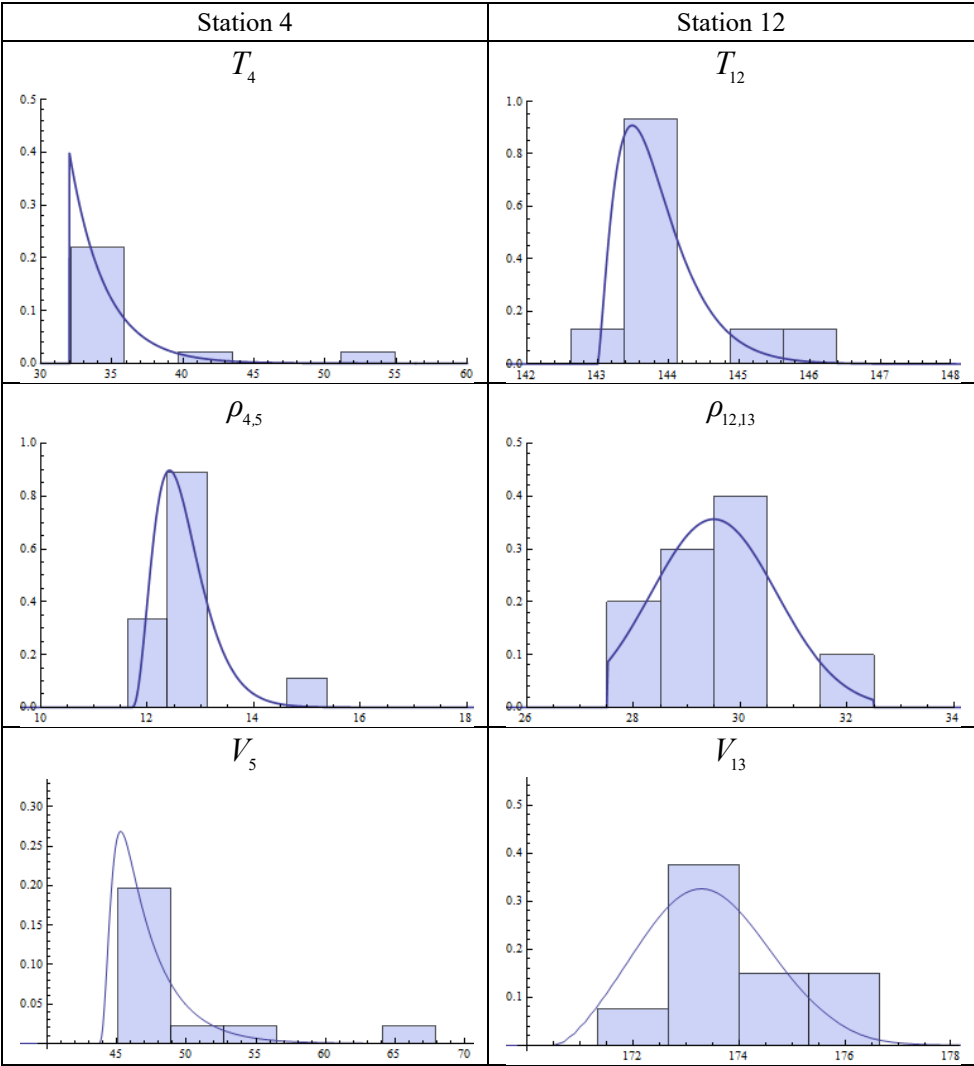
$$\begin{aligned} g_5(t) &= \int_{-\infty}^{\infty} l_{4,5}(t-x) f_4(x) dx \\ &= I(t > 43.7) 14.12 \int_{32}^{t-11.7} e^{-3.8(t-x-11.7)-0.4(x-32)} (t-x-11.7)^{2.75} dx, \end{aligned}$$

$$\begin{aligned}
 g_{13}(t) &= \int_{-\infty}^{\infty} l_{12,13}(t-x) f_{12}(x) dx \\
 &= 3.564 \int_{143}^{\infty} I(27.5 \leq t-x \leq 32.5) e^{-0.36(t-x-29.5)^2 - 2.9(x-143)} (x-143)^{1.4} dx.
 \end{aligned}$$

Calculations using the latest eqns are performed by means the technical computing system Wolfram Mathematica.

It can be seen from row 3 of Table 3, the resulting densities visually are in good agreement with the statistics of arrival times.

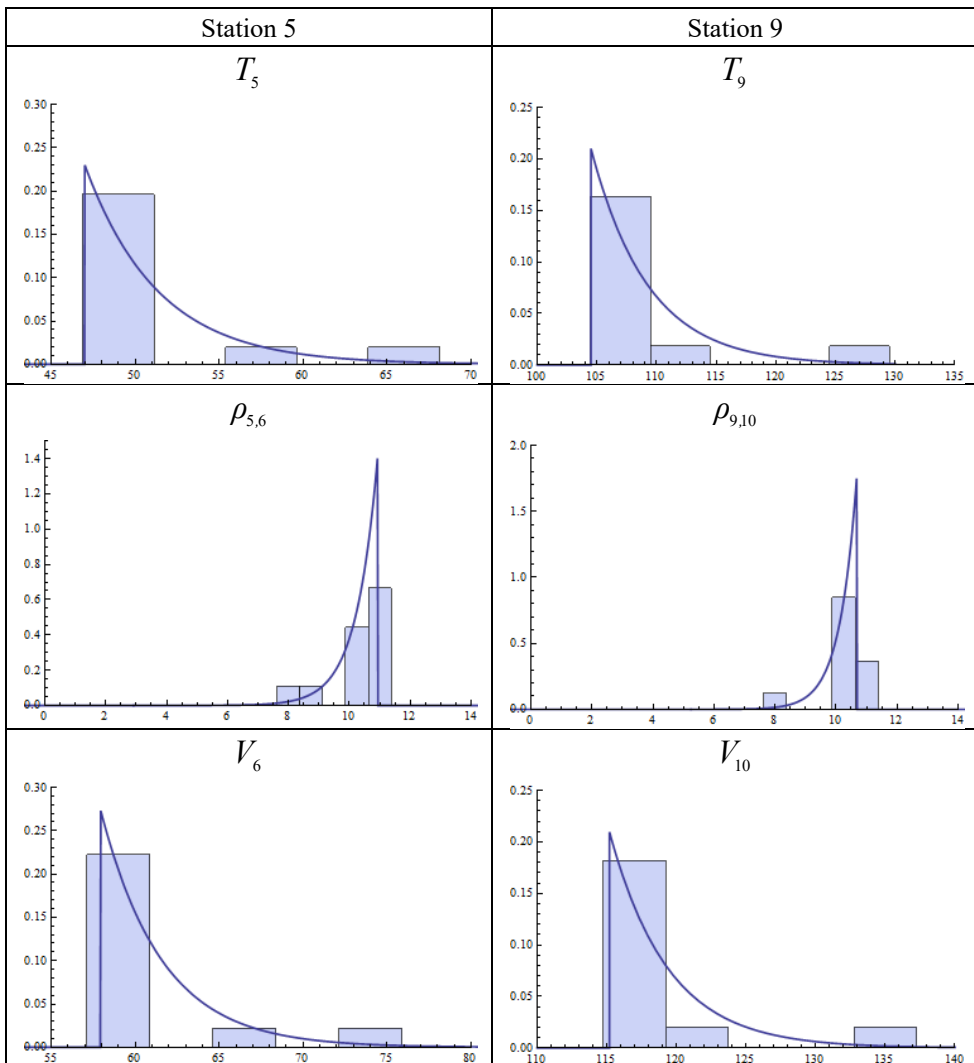
Table 3: Histograms and approximating densities when $\rho_{i,i+1}$ and T_i are independent.



Now we check the consistency of the model with real data in the case of dependence $\rho_{i,i+1}$ and T_i . We restrict ourselves by the linear dependence of these values at stations 5 and 9, identified in Section 5 (see Table 1). The results of the verification are given in Table 4. Travel time (row 2) depends on the departure time (row 1) in accordance with (9). The histograms in row 1 are approximated by the following density functions:

$$f_5(t) = I(t \geq 47)0.23e^{-0.23(t-47)}, \quad f_9(t) = I(t \geq 104.5)0.21e^{-0.21(t-104.5)}.$$

Table 4: Histograms and approximating densities when $\rho_{i,i+1}$ and T_i are dependent.



Density functions in row 2 of Table 4 are obtained from $f_5(t)$ and $f_9(t)$ as follows:

$$l_{5,6}(t) = I\left(\frac{t-b_1}{a_1} \geq 47\right) \frac{0.23}{|a_1|} e^{-0.23\left(\frac{t-b_1}{a_1}-47\right)}, \quad l_{9,10}(t) = I\left(\frac{t-b_2}{a_2} \geq 104.5\right) \frac{0.21}{|a_2|} e^{-0.21\left(\frac{t-b_2}{a_2}-104.5\right)}.$$

The parameters a_1, b_1, a_2, b_2 are determined from the real data according to (10):

$$a_1 = -0.1629, \quad b_1 = 18.5868, \quad a_2 = -0.1189, \quad b_2 = 23.1022.$$

The densities $g_6(t)$ and $g_{10}(t)$ are determined by using eqn (11), i.e.

$$g_6(t) = \frac{dF_5(y)}{dt} = \frac{1}{1+a} f_5(y), \quad g_{10}(t) = \frac{dF_9(y)}{dt} = \frac{1}{1+a} f_9(y), \quad y = \frac{t-b}{1+a}.$$

It can be observed from row 3 of Table 4, the output densities are in good agreement with the arrival time statistics.

7 CONCLUSIONS

The arrival time of a train can be described using two input random variables: departure time and travel time. The form of the arrival time distribution law is determined by whether the input random variables are stochastically dependent or independent. If the values are independent, then the output distribution is a convolution of two input distributions. In the opposite case, it is necessary to describe the relationship between input random variables: the dependence of the travel time on the departure time. And the output distribution will be like a distribution of the departure time converted by a certain way. In practice, on one section of the rail line, consisting of several stations, both dependent and independent input random variables are observed. It is shown in the work that in the case of input data independence, the density function obtained by the eqn containing convolution integral gives a good approximation to real data of arrival times.

In the present paper, only a linear dependence of the travel time on the arrival time is described. We specially selected examples when the modulus of the sample correlation coefficient is close to unity. In these examples, it is shown that the output density function obtained as a transformation of the departure time density function also gives a good approximation to real data of the arrival times.

It should be noted that cases when the modulus of the sample correlation coefficient takes intermediate values between 0 and 1 will be considered in a further study of the problem.

ACKNOWLEDGEMENT

This work is partially supported by the RF Ministry of Transport according to the research work plan of scientific and educational organizations for 2020, project “Improving the operations management of the freight train traffic”.

REFERENCES

- [1] Carey, M. & Kwiecinski, A., Stochastic approximation to the effects of headways on knock-on delays of trains. *Transportation Research Part B*, **28**(4), pp. 251–267, 1994.
- [2] Meester, L.E. & Muns, S., Stochastic delay propagation in railway networks and phase-type distributions. *Transportation Research Part B*, **41**, pp. 218–230, 2007.
- [3] Muhlhans, E., Berechnung der Verspätungsentwicklung bei Zugfahrten/ Eisenbahntechn. Rundschau ETR, **39**(7/8), pp. 465–468, 1990.



- [4] Shapkin, I.N., Yusipov, R.A. & Kozhanov, E.M., Simulation of train functioning on the basis of multivariate regulation of technological operations. *Bulletin VNIIZhT*, **4**, pp. 30–36, 2006. (In Russian.)
- [5] Karetnikov, A.D., *Schedule of Train Traffic*. Transport: Moscow, p. 301, 1979. (In Russian.)
- [6] Buker, T. & Seybold, B., Stochastic modelling of delay propagation in large networks. *Journal of Rail Transport Planning and Management*, **2**(12), pp. 34–50, 2012.
- [7] Boucherie, R.J. & Huisman, T., Running times on railway sections with heterogeneous train traffic. *Transportation Research Part B: Methodological*, **35**(3), pp. 271–292, 2001.
- [8] Berger, A., Gebhardt, A., Muller-Hannemann, M. & Ostrowski, M., Stochastic delay prediction in large train networks. *Proceedings of the 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS'11)*, Saarbrücken, Germany, pp. 100–111, 2011.
- [9] Salido, M.A., Robustness in railway transportation scheduling. *Proceedings of the 7th World Congress on Intelligent Control and Automation (WCICA 2008)*, Chongqing, China, pp. 2833–2837, 2008.
- [10] Al-Ibrahim, A., Dynamic delay management at railways. A semi-Markovian decision approach, PhD thesis, Universiteit van Amsterdam, 2010.
- [11] Goverde, R.M.P., Punctuality of railway operations and time-table stability analysis, PhD thesis, Technical University of Delft, 2005.
- [12] Fichtenholtz, G.M., *The Course of Differential and Integral Calculus*. Vol. 3. Moscow, 2001.

