

Some Features of Formatting the Arrival Time Distribution

Vladimir Chebotarev[1], Boris Davydov[2], Kseniya Kablukova[1(✉)], and Vadim Gopkalo[2]

[1] CC FEB RAS, Kim Yu Chen Street 65, 680000 Khabarovsk, Russia
vladimir.ch@ccfebras.ru, kseniya0407@mail.ru
[2] FESTU, Seryshev Street 47, 680021 Khabarovsk, Russia
dbi@rambler.ru, vng@yandex.ru

Abstract. The paper discusses the basic stochastic two-train and multi-train models. These models describe the process of running and arrival time distributions formation. It is shown that the distribution of arrival times is determined by the cumulative effect of several independent factors. We investigate the process of occurring the asymmetric output probability density function. The main reason of distribution skewness is a train interaction. The study takes the real world data into account. In considered model the departure deviation and running time are input random variables. The types of input distributions are selected by using statistical data. The results of theoretical analysis are confirmed by statistical data obtained on the main railway lines of Russia.

Keywords: Stochastic model · Bimodal probability distribution · Arrival deviation · Deformation of the probability density function

1 Introduction

Stochastic models of train traffic are used to predict the situation progress in the presence of random disturbances. Traffic modeling allows you to detect intertrain conflicts and to consider their further distribution along the railway network. Most common are problems in which the cumulative value of delay all the trains is determined. This indicator reflects the result of appearing the set of knock-on delays in a wide spatial and temporal horizon. Obviously, with such a mapping of a complex multi-conflict situation, it is almost impossible to assess the impact of each individual deviation from the schedule.

In another group of papers, the process of delay propagation is studied, which is observed in a limited space-time zone when a single disturbance occurs. These researches are focused on solving local problems of train operational dispatching. When selecting a particular adjustment, the set of delays that appear in a restricted (local) area is a result of modeling. Such an adjustment is chosen as optimal, which reduces the number or duration of non-scheduled delays.

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11620, pp. 148–161, 2019.
https://doi.org/10.1007/978-3-030-24296-1_14

There is a gap between two approaches described. Models of the first group do not provide effective solutions to local problems. At the same time, the delay propagation process in the local area, which is considered in present article, doesn't allow to reflect some important points that must be taken into account when traffic macro-modeling is made at railway network. Many studies don't take into account the process of changing the scattering behavior of train arrival times under the influence of destabilizing factors. Ignoring this process leads to significant errors in deviations predicting from the schedule.

It should be subjected to comparative analyzing the results of delay calculations based on different imitating models of their propagation along the train chain. This will allow us to estimate the scale of the difficulties encountered in train delays predicting. The present paper attempts to perform the analysis of modeling results obtained from using some common scenarios which cause the process of delays propagation.

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. In one of the first papers considering this problem, an analytical method was proposed for delay determining by convolving the distributions of the initial (at the site entry) and newly arising deviations [1]. The analysis carried out is limited since the assumption of a uniform law of the secondary delays distribution is used. In reality, this assumption is not confirmed.

A productive approach is one in which the expected values of operating times are obtained from statistics of previous periods. Situational-heuristic method (SEMN) is used to determine these quantities, that allows you to take into account the peculiarity of train traffic conditions in the scheduled period while a prognosis being created [2]. A similar approach is used in the paper [3] in which modeling is performed 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 attempts are fraught with conservation of technological problems that exist on the site.

One of the fundamental papers is devoted to the problem of traffic stochastic modeling under consideration [4]. The authors describes a model of delay formation in a train set traffic using a probabilistic approach. Total run time of the train is considered as a sum of random intervals of movement along the elements of the section when analyzing the process of delay propagation. The paper [5] shows that calculation of the arrival time distribution at the terminal station is based on repeated use of convoluting the distributions of random variables characterizing each of the travel elements.

Below mentioned papers use this approach to analyze the process of delay propagation along the train chain. The authors of the study [6] solve the problem of interacting the two trains in a flow 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 develops in the papers [5,7,8]. The paper [7] represents trains traffic in a form occupying 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 from the sequence of activities which is determined by a stochastic event-graph.

Most fully the problem of predicting the random delay appearance and propagation across the large railway network is considered in [8]. Scattering of arrival moments 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 the study [5] which attempts to take the realistic distribution of operation time into account and choose the appropriate approximation. The approach proposed uses a probabilistic operational graph which considers transfer operation and conflict situations. The authors argue that mesoscopic modelling of traffic is the purposeful approach to compute the delay propagation. The purpose of the study is to reduce computing time. The authors apply the phase-type approximation function which is analytically converted to cdf. Such representation simplifies the data processing when searching the output distribution. The model created is a logically incomplete. The authors consider only station and crossing as the operating points where conflicts can occur.

The general characteristics of the schedule which is realized under conditions of random disturbances are considered to be its robustness and stability. Timetable robustness is understood as its ability to absorb small current deviations [9,10]. The compensatory mechanism begins to fail with increasing the intensity of traffic which leads to the multiplication of delays. A characteristic that shows the ability to eliminate the disturbances that occur is defined as timetable stability [11] and it correlates with the time interval that is needed to restore the normative traffic.

The results obtained in these papers do not allow analytically combining the results of calculating the probability distributions of unscheduled deviations after each regular operation. These distributions are obtained using various stochastic models.

3 Stochastic Model of Arrival Deviation Formation

Consider the following two-train model. Trains with numbers 1 and 2 depart one by one from a certain station and move in the same direction. Planned trajectories of their movement are the same, only the moments of departure are different. It is assumed that the condition of compliance with the minimum safe inter-train interval is fulfilled. When departing and in the process of movement

trains are influenced by random factors that lead to scattering of departure times and travel times. To describe the movement of a train with the number i ($i = 1, 2$) we introduce the following notations:

δ_i is the deviation from the scheduled departure time,

$F(t) = \mathbf{P}(\delta_i \le t)$ is the distribution function of the random deviation δ_i, $i = 1, 2$, of the departure moments from the schedule; it is assumed that the variables δ_1 and δ_2 are equally distributed;

$f(t)$ is the density function of δ_1,

ρ_i is the travel time between stations,

$$L_i(t) = \mathbf{P}(\rho_i < t), \tag{1}$$

$l_i(t)$ is the density function of ρ_i;

d_i is the departure time of the i-th train according to schedule;

a_i is the arrival time of the i-th train according to schedule;

ξ_i is the random deviation of the i-th train from the scheduled arrival time; then $a_i + \xi_i$ is actual arrival time of the i-th train;

$U_i(t) = P(\xi_i < t)$, $u_i(t)$ is the density function of ξ_i;

t_0 is the minimum permissible time interval between trains;

$s_0 = s_0(v) = v \cdot t_0$ is the minimum safe distance between trains, corresponding to an arbitrary speed v.

Figure 1 shows the formation of the arrival times $a_j + \xi_j$, $j = 1, 2$.

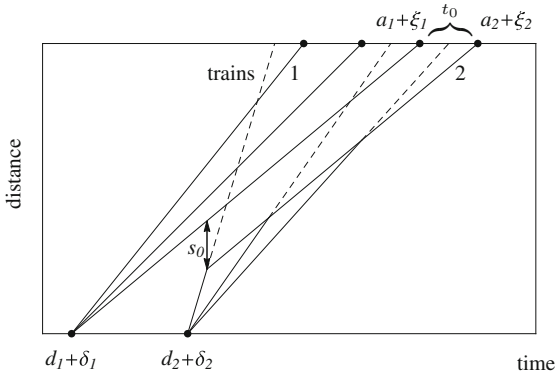


Fig. 1. The formation of the arrival times $a_j + \xi_j$, $j = 1, 2$.

Now we deduce formula

$$U_2(t) = \int_0^\infty F(t - d_1 + a_2 - t_0 - y) l_1(y)\, dy. \tag{2}$$

We have (see Fig. 1) $a_1 + \xi_1 - (d_1 + \delta_1) = \rho_1$, consequently, $a_1 + \xi_1 = \rho_1 + \delta_1 + d_1$. Next, since $a_2 + \xi_2 = a_1 + \xi_1 + t_0$, then

$$\xi_2 = \rho_1 + \delta_1 + d_1 - a_2 + t_0.$$

Hence,

$$U_2(t) = \mathbf{P}(\xi_2 < t) = \mathbf{P}(\rho_1 + \delta_1 < t - d_1 + a_2 - t_0). \tag{3}$$

Assume that ρ_1 and δ_1 are independent. Then (2) follows from (3).

Note that in [12] a different formula was deduced, although similar to (2). The difference is that in [12] the constant s_0 is taken as a given technological constant, but not t_0.

By (2) the density function of ξ_2 has the following form:

$$u_2(t) = \int_0^\infty f(t - d_1 + a_2 - t_0 - y) l_1(y)\, dy. \tag{4}$$

According to the collected statistics, the train travel time often obeys some bimodal distribution law. Studies show that in the case of relatively free traffic, the graph of the density function of the time travel has a longer right branch.

In [13] is shown that such distribution is well approximated by the gamma distribution. It can be assumed that the influence of several (in particular, two) factors on the train traffic is of a similar nature. Take the travel time density function $\tilde{l}(t)$ as a mixture of two gamma densities $\tilde{l}_1(t)$ and $\tilde{l}_2(t)$:

$$\tilde{l}(t) = \tilde{l}(t, p) = p\tilde{l}_1(t) + (1 - p)\tilde{l}_2(t), \quad 0 \le p \le 1, \tag{5}$$

where

$$\tilde{l}_1(t) = I(t > b_1) \frac{e^{-(t-b_1)/\beta}(t - b_1)^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}, \quad \tilde{l}_2(t) = \tilde{l}_1(t - b_2), \tag{6}$$

$\alpha > 0$, $\beta > 0$. Note that hereinafter, all the density functions with the *tilde* icon refer to the case when travel time density is a mixture of gamma distributions (5).

The parameters b_1, $b_2 > 0$ set the shift of the initial gamma densities $\tilde{l}_1(t)$ and $\tilde{l}_2(t)$ along the horizontal axis to the right, because travel time is a strictly positive random variable. In this case, it is necessary to ensure that the support of the density function $\tilde{l}(t)$ lies only on the positive semi-axis, and it is significantly separated from zero.

The bimodality of the density $\tilde{l}(t)$ is determined by the influence of two random destabilizing factors. To simulate the degree of influence of each of them, we use the weighted (probabilistic) parameter p, $0 \le p \le 1$. The more p, the stronger the first factor influences (the multiplier p before $\tilde{l}_1(t)$ in (5)) and the weaker the influence of the second factor, and vice versa.

Let us set the actual form and parameters of the input distributions in (4). Suppose that the departure deviation δ_1 obeys the exponential distribution law with a density function

$$f(t) = I(t > 0)\lambda e^{-\lambda t}, \quad \lambda > 0. \tag{7}$$

Let $\lambda = 0.3$. Such a value of the parameter λ corresponds to reality, because in this case, the mean departure deviation $\mathbf{E}\delta_1 = 1/\lambda \approx 3.3$ min.

Next, we set $\alpha = 16$, $\beta = 0.25$, $b_1 = 6$, $b_2 = 3$. The symbols $\tilde{\rho}_1$ and $\tilde{\rho}_2$ denote random variables with densities $\tilde{l}_1(t)$ and $\tilde{l}_2(t)$, respectively. By using the

properties of the gamma distribution, we get $\mathbf{E}\tilde{\rho}_1 = \alpha\beta + b_1 = 10$ min (under influence of factor 1) and $\mathbf{E}\tilde{\rho}_2 = \alpha\beta + b_1 + b_2 = 13$ min (under influence of factor 2). The standard deviation for both random variables is one: $\sqrt{\mathbf{D}\tilde{\rho}_1} = \sqrt{\alpha\beta^2} = 1$ min.

The following parameters are determined by the normative schedule:

$$d_1 = 0\,\text{min}, \quad a_2 = 15\,\text{min}, \quad t_0 = 3\,\text{min}. \tag{8}$$

By varying the parameter p, let us see the behavior of the density function $\tilde{l}(t) = \tilde{l}(t, p)$ from (5) and its impact to the corresponding density $\tilde{u}_2(t)$, defined by (4). The results of the calculation are collected in Table 1.

When the parameter p is close to zero (see Table 1, $p = 0.1$), the first factor has a negligible effect on the travel time, the influence of the second factor prevails. Due to its impact, the mean travel time increases, which is accompanied by a higher right-hand vertex of the density graph $\tilde{l}(t)$ compared to the left. In this case, the advance of the train traffic schedule is practically not observed because the graph of the density function $\tilde{u}_2(t)$ is concentrated on the positive semi-axis.

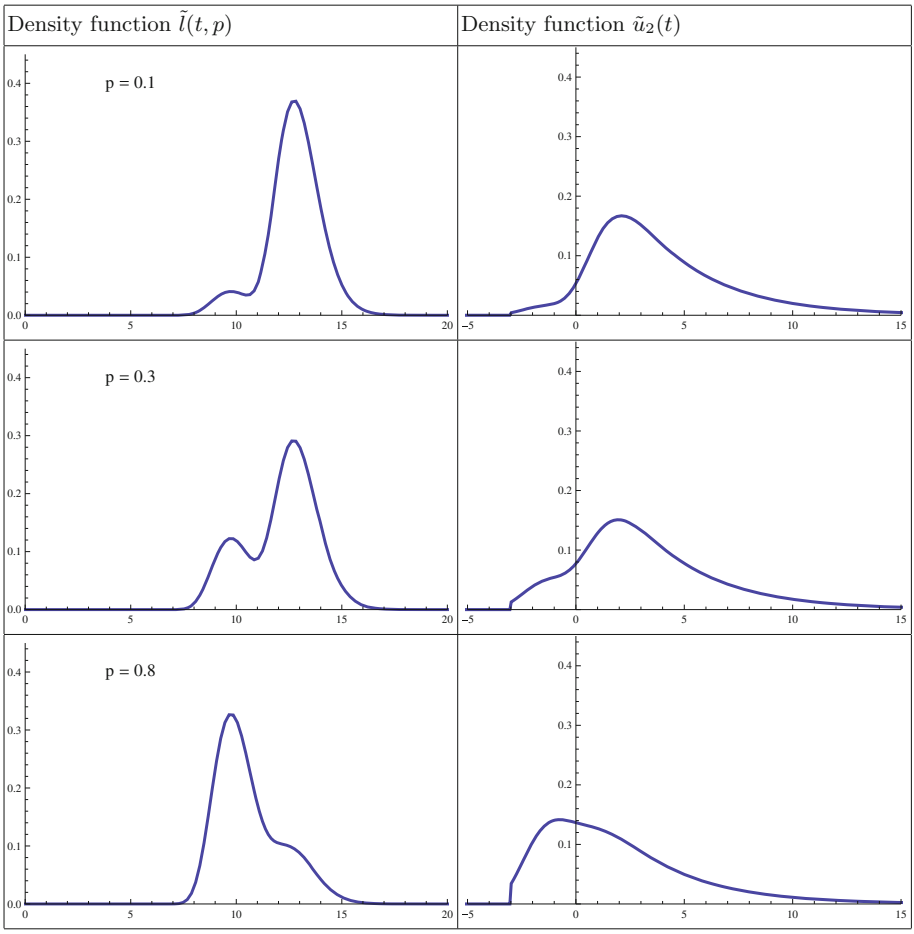
As the parameter p increases, the degree of each factor impact changes: the mean travel time gradually decreases, the bimodal density $\tilde{l}(t)$ (see Table 1, $p = 0.3$) is changed by the unimodal one (see Table 1, $p = 0.8$). The deformation of the density $\tilde{l}(t)$ entails the deformation of the density $\tilde{u}_2(t)$. This leads to increase in the probability to advance the train traffic schedule, namely the support of the density function $\tilde{u}_2(t)$ "captures" a segment of the negative semi-axis.

Thus, the choice of the parameter p allows adjusting the degree of influence of each factor depending on the specific situation. In this case the following rule which is in good agreement with reality is observed: the smaller the value of the mean travel time, the smaller the mean of the delay. In addition, in this case negative arrival deviations are more often observed, that is an advance of the schedule.

4 An Example of the Joint Influence of Random Factors on the Arrival Deviation Distribution

One of the independent factors that are involved in the formation of travel time as a random variable is the mutual influence of trains. A stochastic model that reflects this influence was proposed in [6]. The paper studies the case when a faster train catches up with a slow leading train in the middle of the section, and later both trains move with a fixed time interval t_0 (see the diagram in Fig. 2). In the simulation, it is assumed that the train speed fluctuates, as it is subject to random influences. Obviously, the scattering rate can be considered as one of the factors that we considered in Sect. 3 in constructing a model for the formation of arrival deviations. Thus, the multimodality of the run time is determined by the presence of two random variables, namely, the departure time and train speed. In the following, we will assume that slow trains, which move in a flow, belong to type 1, fast – to type 2.

Table 1. The behavior of the functions $\tilde{l}(t,p)$ and $\tilde{u}_2(t)$ for different values of the weight parameter p.



In Sect. 3, we considered the first important particular case of the proposed model, when train travel time obeys a mixture of two gamma distributions (5). As already mentioned, the multimodality of the run time density can be determined not only by the influence of several random factors, but also by the presence in the flow on one track the trains of different types. The train type is determined by its speed [6].

The article [6] deals with the general case of finding the distribution of the actual travel time on open tracks that takes $K > 0$ types of trains into consideration. Each of these types is characterized by values of free running time on a section F_n, where n denotes a train number in the flow. The cumulative

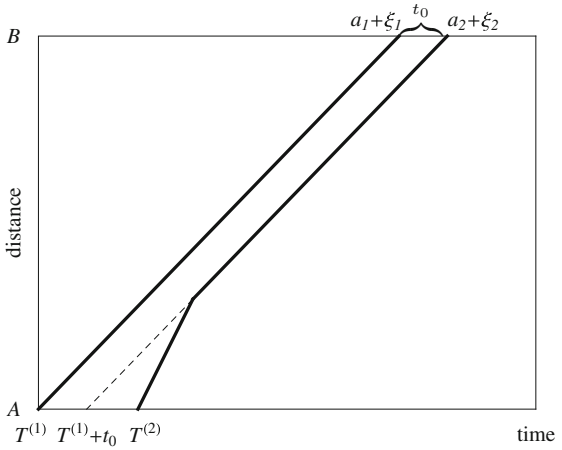


Fig. 2. Train interaction scenario: fast train catches up with slow one.

distribution function of F_n (provided that train n is of type k, i.e. $Q_n = k$) is determined as follows:

$$V_k(t) = \mathbf{P}(F_n \le t | Q_n = k). \tag{9}$$

Random variable R_n denotes actual travel time of the train n. Suppose that the number of types $K = 2$. Let p be a probability that train n is of type 1. Further, we use a formula (16) from the article [6] with new notation. The cumulative distribution function of the actual travel time of the train type j, $j = 1, 2$, is determined as follows for $t \ge 0$:

$$\begin{aligned}\hat{L}_j(t) \equiv \hat{L}_j(t,p) &= \lim_{n\to\infty} \mathbf{P}(R_n \le t | Q_n = j) \\ &= V_j(t) \exp\Big(-\hat{\lambda} \int_t^\infty \Big(1 - p\,V_1(x) - (1-p)\,V_2(x)\Big)\,dx\Big), \end{aligned} \tag{10}$$

where $\hat{\lambda}^{-1}$ denotes the mean departure intertrain interval provided that it has an exponential distribution with the parameter $\hat{\lambda} > 0$.

The parameter $\hat{\lambda}$ can be found from the following relation:

$$\hat{\lambda}^{-1} = \mathbf{E}H - t_0, \tag{11}$$

where $\mathbf{E}H$ denotes an average intertrain interval by the train flow (in minutes).

Denote a probability density function of the train travel time (train has a type j, $j = 1, 2$) as

$$\begin{aligned}\hat{l}_j(y) \equiv \hat{l}_j(y,p) = \hat{L}_j'(y) = \exp\Big\{ -\hat{\lambda} \int_y^\infty \Big(1 - pV_1(x) - (1-p)V_2(x)\Big)\,dx\Big\} \\ \times\Big(v_j(y) + \hat{\lambda} V_j(y)\Big(1 - pV_1(y) - (1-p)V_2(y)\Big)\Big), \end{aligned} \tag{12}$$

where $v_j(y) = V_j'(y)$.

Consider a mixture of densities

$$\hat{l}(y) \equiv \hat{l}(y,p) = p\hat{l}_1(y,p) + (1-p)\hat{l}_2(y,p), \tag{13}$$

with $0 \le p \le 1$. The probability density function $\hat{l}(y)$ is used for simulation of arrival deviations with the following types of input distributions:

$$V_1(y) = \frac{I(y > b_1)}{\Gamma(\alpha_1)} \int_0^{(y-b_1)/\beta_1} x^{\alpha_1 - 1} e^{-x}\, dx, \tag{14}$$

$$V_2(y) = \frac{I(y > b_2)}{\Gamma(\alpha_2)} \int_0^{(y-b_2)/\beta_2} x^{\alpha_2 - 1} e^{-x}\, dx, \tag{15}$$

$$v_1(y) = \frac{dV_1(y)}{dy} = I(y > b_1) \frac{e^{-(y-b_1)/\beta_1}(y-b_1)^{\alpha_1 - 1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}}, \tag{16}$$

$$v_2(y) = \frac{dV_2(y)}{dy} = I(y > b_2) \frac{e^{-(y-b_2)/\beta_2}(y-b_2)^{\alpha_2 - 1}}{\Gamma(\alpha_2)\beta_2^{\alpha_2}}, \tag{17}$$

where $\Gamma(\cdot)$ implies the Euler gamma-function.

According to (14)–(17), a free running time of the train type j, $j = 1, 2$, obeys a gamma-law with the parameters $\alpha_j > 0$, $\beta_j > 0$ and shift $b_j > 0$.

Note that hereinafter, all density functions with the *cover* icon refer to the case when density of the travel time is based on the paper [6].

We assume the parameter $\hat{\lambda} = 0.5$. This value describes the situation when $\mathbf{E}H = 5$ min, $t_0 = 3$ min (see (11)).

Let us set the real parameters of the input distributions $v_1(y)$ and $v_2(y)$ estimated on the basis of statistical data:

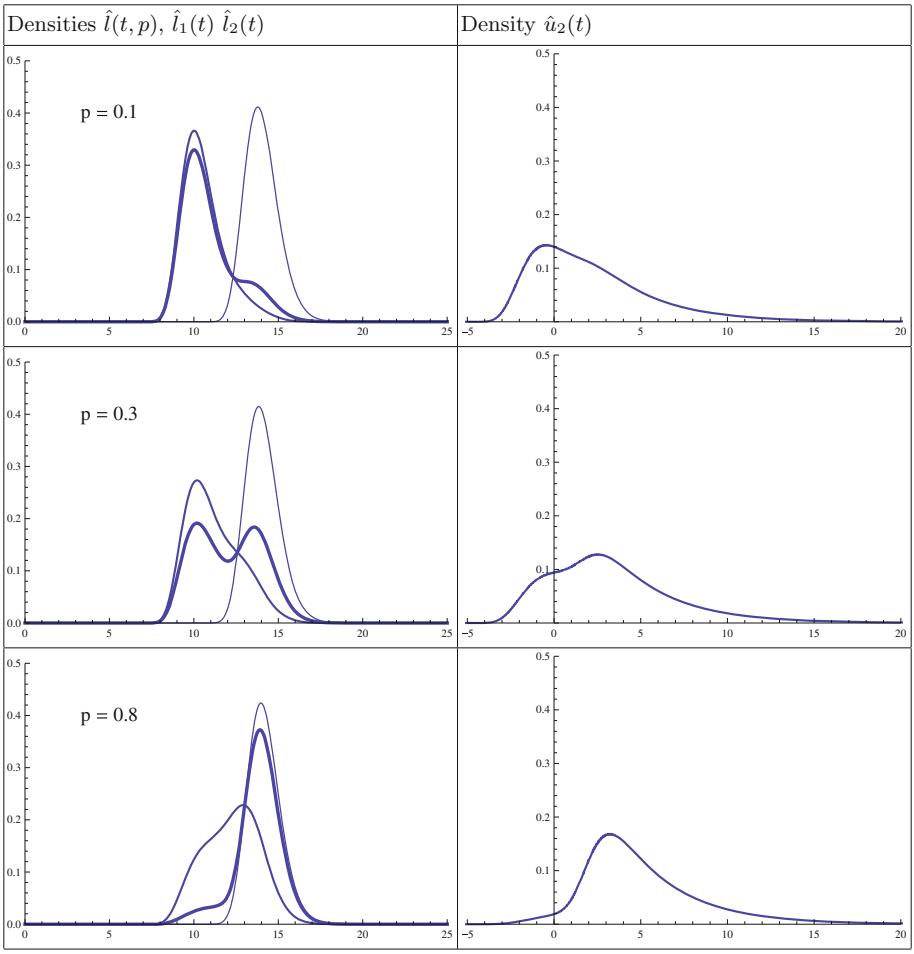
$$\alpha_1 = 16, \quad \beta_1 = 0.25, \quad b_1 = 10,$$
$$\alpha_2 = 9, \quad \beta_2 = 1/3, \quad b_2 = 7.$$

In this scenario, train type 1 characterizes slow trains with the mean travel time $\alpha_1\beta_1 + b_1 = 14$ min, type 2 – fast trains with the mean $\alpha_2\beta_2 + b_2 = 10$ min. Dispersions of both distributions are equal to 1: $\alpha_1\beta_1^2 = \alpha_2\beta_2^2 = 1$ min^2.

We explore how the behavior of a density $\hat{l}(t) = \hat{l}(t,p)$ changes being determined by the equality (13), as well as a density $\hat{u}_2(t)$ defined by the equality (4), by changing the parameter p. In Table 2, parameter p is fixed. The thinner and bold lines correspond to the graphs of $\hat{l}_1(t)$ and $\hat{l}_2(t)$ respectively, the boldest line represents the graph of $\hat{l}(t)$ (for all left sided figures). In the left column of Table 2, the graphs of densities $\hat{l}_1(t)$ and $\hat{l}_2(t)$ defined by the equality (12) with $j = 1, 2$, and their mixtures $\hat{l}(t,p) = p\hat{l}_1(t) + (1-p)\hat{l}_2(t)$ are depicted. Corresponding densities $\hat{u}_2(t)$ are shown at the right column.

The Table 2 reflects the form of densities $\hat{l}(t)$ and $\hat{u}_2(t)$ changes with the growth of the parameter p. This fact can be explained as follows. If value of p is close to 1 then slow trains (with a density $\hat{l}_1(t)$) prevail in the common flow. Therefore, the function $\hat{l}(t)$ is close to unimodal density $\hat{l}_1(t)$. Then $\hat{u}_2(t)$ is also unimodal (see Table 2, $p = 0.8$). At the same time, the number of fast trains

Table 2. Behavior of the travel time density function $\hat{l}(t,p)$ and arrival deviation density function $\hat{u}_2(t)$ with a different ratio of slow and fast trains.



which are influenced by slow ones is less than the number of slow trains. As a result, the frequency of delays is small.

If the parameter p is close to zero (but not equal to 0) then the share of fast trains in the flow is greater than the share of slow ones. If there is a delay of a preceding train, the knock-on effect of secondary delays will be observed more often. Hence, the density $\hat{l}(t)$ is close to the density $\hat{l}_2(t)$, which is unimodal in the scenario under consideration. Then $\hat{u}_2(t)$ is also unimodal (see Table 2, $p = 0.1$).

If $p = 0.3$, then the density $\hat{l}(t)$ takes into account the properties of both densities $\hat{l}_1(t)$ and $\hat{l}_2(t)$, which leads to the bimodality of $\hat{l}(t)$. This in turn leads

to a deformation of $\hat{u}_2(t)$ whose form can be bimodal or close to bimodal (see Table 2, $p = 0.3$).

The density form of fast train actual travel times $\hat{l}_2(t)$ varies as the parameter p changes. The less p, the more fast trains in the common flow and the more correct (symmetric) form of the graph $\hat{l}_2(t)$ is observed. Note the support of function $\hat{l}_2(t)$ for $p = 0$ is the "narrowest" since there is no impact of slow train traffic. Increasing the parameter p entails support expansion of $\hat{l}_2(t)$ and a gradual change in the skewness from positive (for $p \to 0$) to negative (for $p \to 1$). Negative asymmetry is acquired due to the fact that fast trains are forced to adjust to slow ones reducing its speed. Consequently, the support of $\hat{l}_2(t)$ completely "absorbs" the support of $\hat{l}_1(t)$ and running times of fast trains begin to coincide with running times of slow ones.

5 Validation of the Model by Using Statistical Data

This section is devoted to the analysis of research results by means of real statistical data of travel time and arrival deviations. The observed features are discussed taking into account the results of a theoretical analysis conducted in Sects. 3 and 4.

We explore the statistics collected for the trains of the suburban radial railway line Moscow-Tver. It turned out that bimodal travel time densities are not uncommon, but appear quite often. Table 3 contains the examples of histograms plotted from samples of travel time and arrival deviations distributions in the morning and evening rush hours. Note that for the convenience of displaying in travel time histograms, the vertical axis intersects the horizontal one not at the origin, but at the point (18, 0) or (5, 0).

The given examples show that the appearance of a bimodal travel time distribution leads to a rise in the left tail of the arrival deviation density, that is, it creates a left-sided asymmetry. If the influence of random factors on the train traffic is comparable, then it leads to the presence of two modes in the output density function. Thus, the nature of the histograms constructed on the basis of experimental data agrees well with the result of theoretical constructions for both special cases of the travel time distribution (5) and (13), which are given in Sects. 3 and 4 of this article.

The histograms of the travel time given in Table 3 are fairly well approximated by a mixture of gamma distributions (5) with an approximately same degree of each factor influence on the travel time. In Tables 1, 2 this variant corresponds to the figures, which illustrate the densities of the considered random variables in the case when the weight parameter p is equal to 0.3. For example, the following calculated parameters of the resulting arrival deviation density $\tilde{u}_2(t)$ are characterize morning peak hours:

$$p = 0.3, \quad \alpha = 16, \quad \beta = 0.25, \quad b_1 = 17.5, \quad b_2 = 2.5, \quad \lambda = 1, \qquad (18)$$

the parameters of the arrival deviation density $\hat{u}_2(t)$:

$$p = 0.3,\ \alpha_1 = 16,\ \beta_1 = 0.25,\ b_1 = 21,$$
$$\alpha_2 = 9,\ \beta_2 = 1/3,\ b_2 = 19,\ \hat{\lambda} = 0.5,\ \lambda = 1. \quad (19)$$

Moreover, the normative parameters for a given track are following:

$$d_1 = 0\,\text{min}, \quad a_2 = 27\,\text{min}, \quad t_0 = 3\,\text{min}. \quad (20)$$

Table 4 contains a histogram of arrival deviations for morning rush hours from Table 3 with two approximating density curves: (i) graph of the density

Table 3. Histograms of the travel time and arrival deviation distribution.

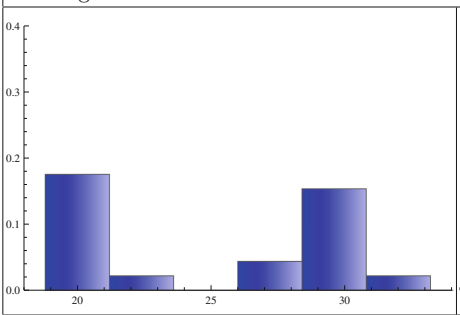
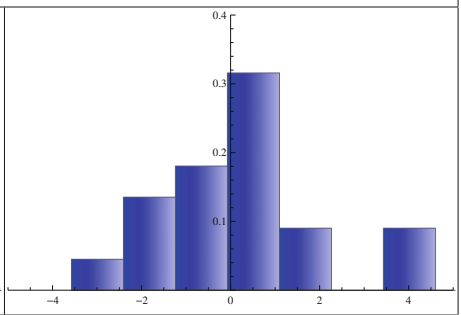
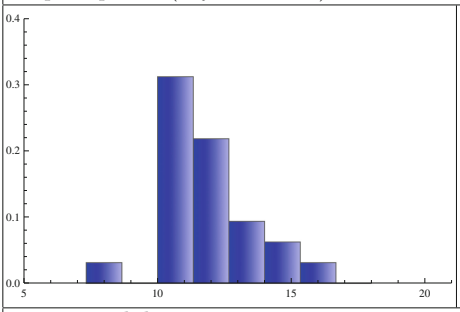
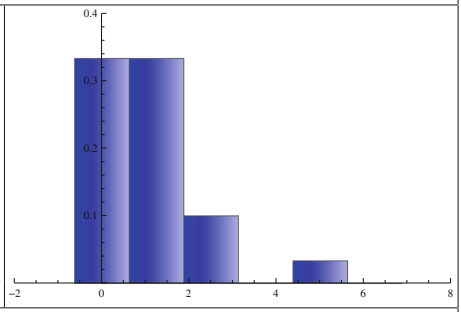
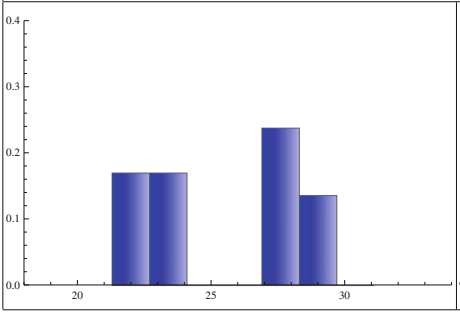
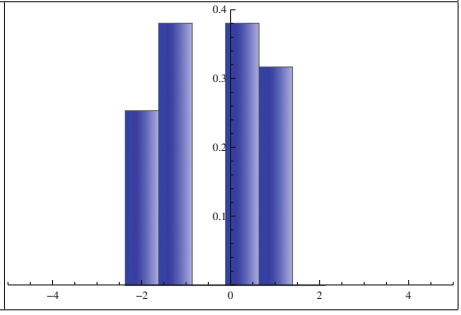
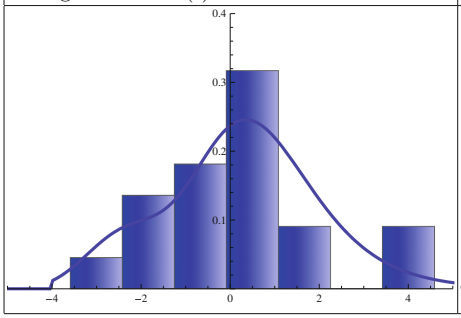
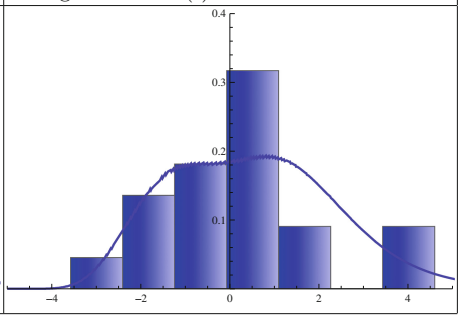
Travel time	Arrival deviations
Morning rush hours	
	
Off-peak period (daytime hours)	
	
Evening rush hours	
	

Table 4. Agreement of $\tilde{u}_2(t)$ and $\hat{u}_2(t)$ with statistical data.

Histogram and $\tilde{u}_2(t)$	Histogram and $\hat{u}_2(t)$
	

$\tilde{u}_2(t)$ with the parameters from (18), (20) (left sided), (ii) graph of the density of $\hat{u}_2(t)$ with the parameters from (19), (20) (right sided). Parameters correspond to the considered scenario. From the figures in Table 4 it can be seen that the theoretical density from (4) in both cases is consistent with the experimental data, but in the first case (left) the agreement is better. It should be noted that the degree of agreement can be increased by using the "cutting" of the density $u_2(t)$ ($\tilde{u}_2(t)$ or $\hat{u}_2(t)$) into the finite segment (the histogram support).

Results similar to passenger train traffic were obtained when considering statistical data on the freight train traffic through Trans-Siberian main line. The analysis shows that during periods of increase the train flow intensity, the number of sections in which a bimodal distribution of travel time is observed reaches 20%. Accordingly, the shape of the arrival deviations density curve becomes symmetrical or has left-sided skewness.

6 Conclusions and Future Research

The stochastic model of the arrival deviation formation considered in this paper allows us to substantiate the assumption that the presence of independent influencing factors leads to a change in the shape and symmetry of arrival deviation distribution. The flexibility of the model is that the resulting analytical formula allows using arbitrary theoretical distributions that adequately reflect the influence of these factors in various real-world scenarios.

Experimental data collected on Russian railways show that, in practice, bimodal travel time density is not uncommon and is observed quite often. Such nature of the travel time distribution also causes the deformation of the output arrival deviation distribution that is often observed in reality, that is, it acquires a negative skewness. The factor causing the bimodality of travel time density is often determined by mutual influence of the trains in flow due to the difference in their speeds. This fact allows us to build a more adequate train traffic model and increase the accuracy of delay prediction.

Subsequently, the authors plan to carry out an in-depth analysis of the arrival deviation formation process and apply the regularities identified in this research to a number of existing stochastic models of train traffic.

References

1. Muhlhans, E.: Berechnung der Verspatungsentwicklung bei Zugfahrten/ Eisenbahntechn. Rundschau ETR **39**(7/8), 465–468 (1990)
2. Shapkin, I.N., Yusipov, R.A., Kozhanov, E.M.: Simulation of train functioning on the basis of multivariate regulation of technological operations. Bull. VNIIZhT **4**, 30–36 (2006). (in Russian)
3. Karetnikov, A.D., Vorobyev, N.A.: Schedule of train traffic. Transport, Moscow, 301 p. (1979). (in Russian)
4. Carey, M., Kwiecinski, A.: Stochastic approximation to the effects of headways on knock-on delays of trains. Transp. Res. Part B **28**(4), 251–267 (1994)
5. Buker, T., Seybold, B.: Stochastic modelling of delay propagation in large networks. J. Rail Transp. Plan. Manag. **2**(12), 34–50 (2012)
6. Boucherie, R.J., Huisman, T.: Running times on railway sections with heterogeneous train traffic. Transp. Res. Part B Methodol. **35**(3), 271–292 (2001)
7. Meester, L.E., Muns, S.: Stochastic delay propagation in railway networks and phase-type distributions. Transp. Res. Part B **41**, 218–230 (2007)
8. Berger, A., Gebhardt, A., Muller-Hannemann, M., Ostrowski, M.: Stochastic delay prediction in large train networks. In: Proceedings of the 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2011), Saarbrucken, Germany, 8 September 2011, pp. 100–111 (2011)
9. Salido, M.A.: Robustness in railway transportation scheduling. In: Salido, M.A., Barber, F., Ingolotti, L. (eds.) 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. Ph.D. thesis, Universiteit van Amsterdam, 335 p. (2010)
11. Goverde, R.M.P.: Punctuality of railway operations and time-table stability analysis. Ph.D. thesis, Technical University of Delft, 165 p. (2005)
12. Chebotarev, V., Davydov, B., Kablukova, K.: Random delays forming in the dense train flow. In: World Conference on Transport Research (WCTR 2019), Mumbai, 26–31 May 2019, 13 p. (2019, in Publication)
13. Chebotarev, V., Davydov, B., Kablukova, K.: Probabilistic model of delay propagation along the train flow. In: Kostogryzov, A. (ed.) Probabilistic Modeling in System Engineering, IntechOpen, pp. 171–193 (2018)