Mechanics Transport Communications

Academic journal

ISSN 1312-3823

volume 12, issue 3/1, 2014

article № 0962

http://www.mtc-aj.com

ANALYSIS OF DELAYS OF FREIGHT AND PASSENGER TRAINS ON THE RAILWAY NETWORK OF BULGARIA

Kiril Karagyozov, Mirena Todorova
kkaragyozov@yahoo.com, mirena_todorova@abv.bg

Todor Kableshkov University of Transport
158 Geo Milev Street, 1574 Sofa,
BULGARIA

Key words: *probability distribution, sum of random variable, data analysis of delays*

Abstract: *The analysis of random variables – train delays due to various reasons (infrastructure, carrier, external influences) is based on processing of statisticsdata available on the number and duration of failures. The statistics data for these reasons are available in an aggregate form, making it necessary to develop a model for the evaluation of the failure duration. The probability dependencies (well-known in the Theory of Probabilities) of central moments of a random variable, which is a sum of random number of independently distributed random variables, are used. This approach makes it possible to find the mean value and variance of the duration of failure /delay/ and these two features can be used to approximate the unknown distribution of the duration of failure by Gamma distribution.*

Train delays can be due to various reasons such as reasons resulting from infrastructure fault, carrier's fault or external influences. The report on the performance of train schedules is made by the Traffic Control Center with the NRIC for a certain period of time by categories of trains, trains and more by certain reasons. Based on Article 22, para. 3 of the Contract on Access to and Study on Railway Infrastructure signed between the infrastructure manager and the railway undertaking, every month a bilateral protocol is signed, which reports by each category in total how many trains /number/ have been late /minutes/ [2]. In order to use these aggregated statistical data it is necessary to develop an evaluation model of the delay duration including the following steps:

1. If we begin with the mean E[X], then

(1) $$E[X]=\sum_{n=0}^{\infty}E[X|N=n]\,p_N(n)$$

$$=\sum_{n=0}^{\infty}E[\xi_1+..+\xi_N|N=n]\,p_N(n) \qquad \text{definition of X}$$

$$=\sum_{n=0}^{\infty}E[\xi_1+..+\xi_n|N=n]\,p_N(n)$$

$$= \sum_{n=0}^{\infty} E[\xi_1 + .. + \xi_n]\, p_N(n)$$

$$= \mu \sum_{n=0}^{\infty} n\, p_N(n) = \mu\nu$$

Sums of the form $X = \xi_1 + .. + \xi_N$, where N is random, arise frequently and in varied contexts [1,3,4]. Our study on random sums begins with a crisp definition and a precise statement of the assumptions effective in this section, followed by some quick examples.

We postulate a sequence $\xi_1, \xi_2 ....$ of independent and identically distributed random variables. Let N be a discrete random variable, independent of $\xi_1, \xi_2 ....$ and having the probability mass function $p_N(n) = \Pr\{N = n\}$ for $n = 0,1,...$ Define the random sum X by

(2) $$X = \begin{cases} 0 & N = 0, \\ \xi_1 + .. + \xi_N \quad \text{if} & N > 0. \end{cases}$$

We save space by abbreviating (2) to simplify $X = \xi_1 + .. + \xi_N$ understanding that X=0 whenever N=0.

2. Let *N* be a random variable assuming positive integer values 1*,* 2*,* 3.... Let *Xi* be a sequence of independent random variables which are also independent of *N* and with *E*[*Xi*] = *E*[*X*] the same for all *i*. Then

(3) $$E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X]$$

3. Let *N* be a random variable assuming positive integer values 1*,* 2*,* 3.... Let *Xi* be a sequence of independent random variables, which are also independent of *N* with Var[*Xi*] the same for *i*. Then

(4) $$Var\left[\sum_{i=1}^{N} X_i\right] = E[N]Var[X] + \left(E[X]\right)^2 Var[N]$$

Because of the small set of data, it is looked for evaluation of the mean number of delayed trains, mean delay and variances corresponding to them.

Let look the random conditional variable Z –the total delay if there are Y delays/month

[5] $$Z(y = Y) = \sum_{i=1}^{Y} Xi$$

Y – random number of delays per month;

Xi – unknown random variable of one delay;

Z (y=Y) - random variable of total delays with Y delays/month.

The unconditional random variable has the following numerical characteristics:

4. Mean number of delays per month is estimated by descriptive statistics for mean

[6] $$m_Y = \frac{\sum_{j=1}^{N} Yj}{\mathbf{N}} ,$$

where N- number of months;

5. Variance of random number y is determined by descriptive statistics for variance:

[7] $$Var(y) = \sigma(y)^2 = \frac{\sum_{j=1}^{N}(y_j - m_y)^2}{N-1} ;$$

Where σ_y is standard deviation of random variable y estimated by data set

6. The mean of one delay is the simple average of total sum of delay for N months divided of total number of delays for N months is estimated by data set:

[8] $$m_x = \sum_{j=1}^{N} Z_j \Big/ \sum_{j=1}^{N} Y$$

It is known from probability theory that the mean and the variation of unconditional random variable Z- the sum of random number random variables x_i -independents and equally distributed are:

[9] $$M(Z) = m_x m_{y_x} ;$$

[10] $$Var(Z) = \sigma_x{}^2 * m_Y + \sigma(y)^2 * m_x^2$$

From existing data set we can estimate the simple mean of random variable x, y and Z and standard deviation of random variable y and Z. If we assume the existence of (5) and (6) the only unknown is the standard deviation of random variable x.

From (6) we can obtain the unknown variation of random variable x – each train delay per reason.

[11] $$\therefore \sigma(x)^2 = \frac{Var(Z) - \sigma(y)^2 * m_x^2}{m_Y}$$

The variance of Z is estimated by existing data set of total delay Zj for month- j. j=1,..N as:

[12] $$Var(Z)^* = \frac{\sum_{j=1}^{N}(Zj - Mz)^2}{N-1}$$

It is assumed that the probability distribution of random variable duration of delay Xi is Gamma distribution with mean M(x) and variance $\boldsymbol{\sigma(x)}$. The equation for the gamma probability density function is:

(13) $$f(x) = \frac{\beta^{-\alpha} x^{\alpha-1} e^{\frac{-x}{\beta}}}{\Gamma(\alpha)} ;$$

where $\Gamma(\alpha)$ is the Gamma function $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$

The mean and the variance of Gamma probability distribution of random variable x are:

[14] $m(x)=\alpha\beta;\ Var(x)=\sigma^2(x)=\alpha\beta^2;$

The unknown parameters При оценени от статистистическите данни средна и дисперсия за параметрите на гама разпределението се получават:

[15] $$\alpha = \frac{M(x)^2}{\sigma^2}$$

[16] $$\beta = \frac{M(x)}{\alpha}$$

Based on the statistics provided by the NRIC about the delays of freight trains throughout the network attributable to the transport infrastructure and carrier respectively that represent an annual sample per months, the following parameters are determined:

A. For delays attributable to infrastructure, it is obtained that

Table 1.

mean number of failures per month	3,84
mean delay per failure	74,75
mean total delay	286,88
Standard deviation of total delay	334,45
Standard deviation of delay per failure	114,76
Standard deviation of number of failures	3,31
α- shape parameter of Gamma probability distribution	0,42
β – scale parameter of Gamma probability distribution	176,2

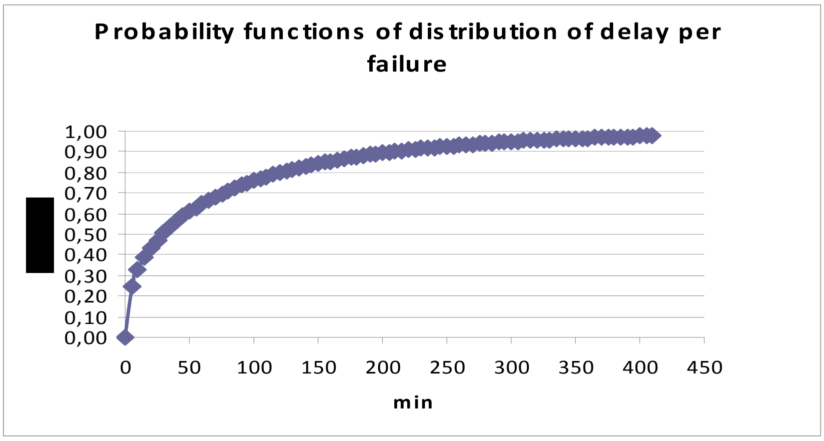

Fig.1

B. For delays attributable to the carrier, it is obtained that:

Table 2.

mean number of failures per month	20,49
mean delay per failure	77,22
mean total delay	1582,03
Standard deviation of total delay	1008,62
Standard deviation of delay per failure	126,2
Standard deviation of number of failures	10,77
α- shape parameter of Gamma probability distribution	0,37
β – scale parameter of Gamma probability distribution	206,24

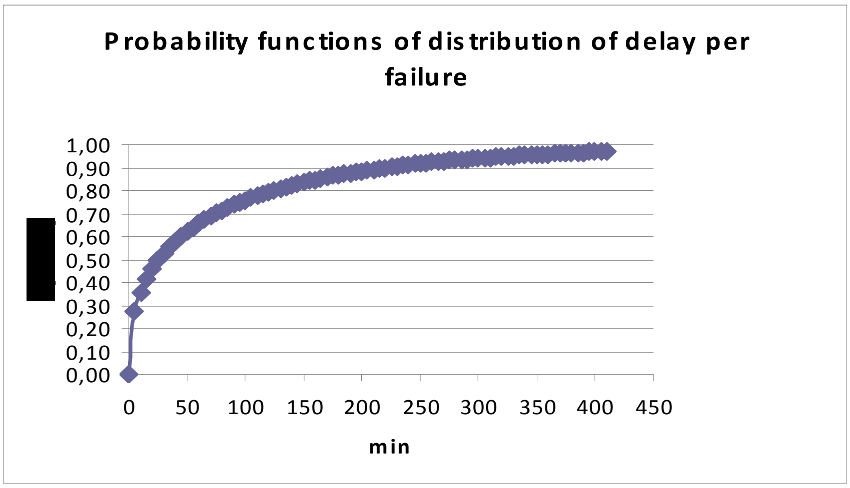

Fig.2

The data refer to the whole railway network but according to the information given by BDZ 30% of all train-kilometer work is on the line studied in the project.

The analysis of delays of passenger trains is based on more detailed information as the trains being surveyed pass along the line examined (with the exception of international trains, which are only a few in number and often arrive late at the border stations). The reasons that lead to train delays are divided into: infrastructure, traffic, signaling and telecommunications, catenary and energy, locomotives, wagons, traffic management and control, etc. The first four reasons are attributable to the infrastructure, next three ones are attributable to the carrier, others include all reasons that are not mentioned as well as bad weather. The diagram shows the relative shares of delays by reasons.

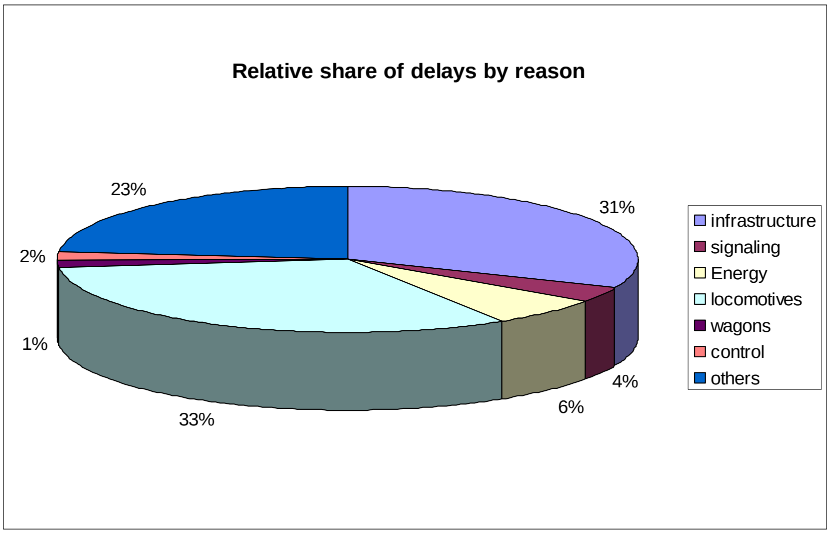

Fig.3

The parameters of delays attributable to the NRIC are given in the table and diagram below .

Table 3.

	Infrastructure	Signaling	Energy
mean number of failures per month	5,57	4,71	3,71
mean delay per failure	32,41	4,33	9,73
mean total delay	180,57	20,43	36,14
Standard deviation of total delay	185,25	28,06	69,02
Standard deviation of delay per failure	32,83	5,29	15,76
Standard deviation of number of failures	5,19	5,91	6,37
α- shape parameter of Gamma probability distribution	0,97	0,67	0,38
β – scale parameter of Gamma probability distribution	33,25	6,45	25,52

Probability distribution function of random variable –delay per failure

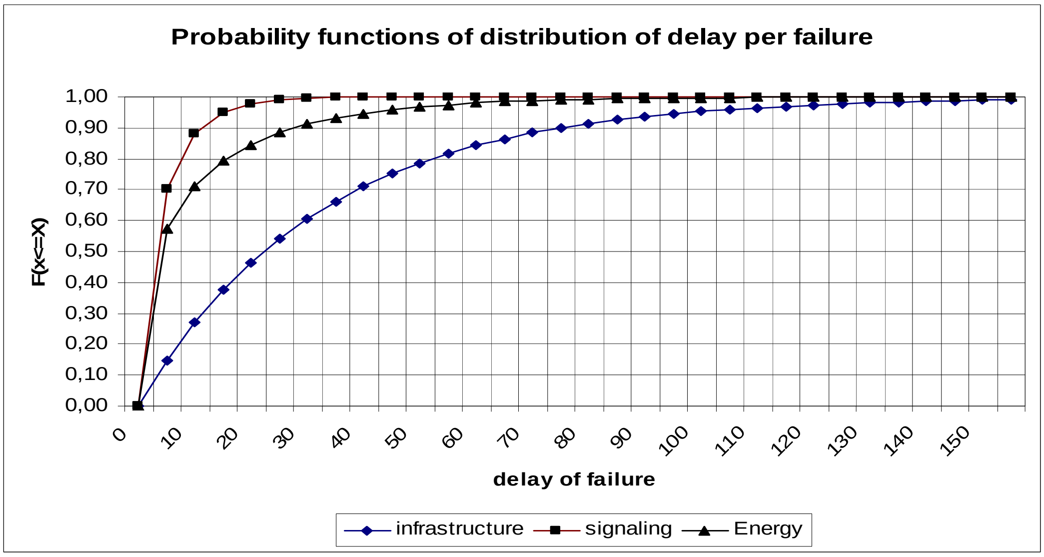

Fig.4

The parameters of delays attributable to the NRIC are given in the table below and it has been made for the other reasons.

Table 4.

	Locomotives	Wagons	Control	Others
mean number of failures per month	5,57	1,43	3,14	15,07
mean delay per failure	33,38	5,6	3,64	8,97
mean total delay	186	8	11,43	135,14
Standard deviation of total delay	182,66	10,72	16	130,38
Standard deviation of delay per failure	24,43	1,03	2,81	14,54
Standard deviation of number of failures	5,19	1,9	4,18	13,11
α- shape parameter of Gamma probability distribution	1,87	29,73	1,67	0,38
β – scale parameter of Gamma probability distribution	17,88	0,19	2,18	23,57

Due to the varying length of delays, in the diagrams they are grouped as delays attributable to the locomotive fleet and other.

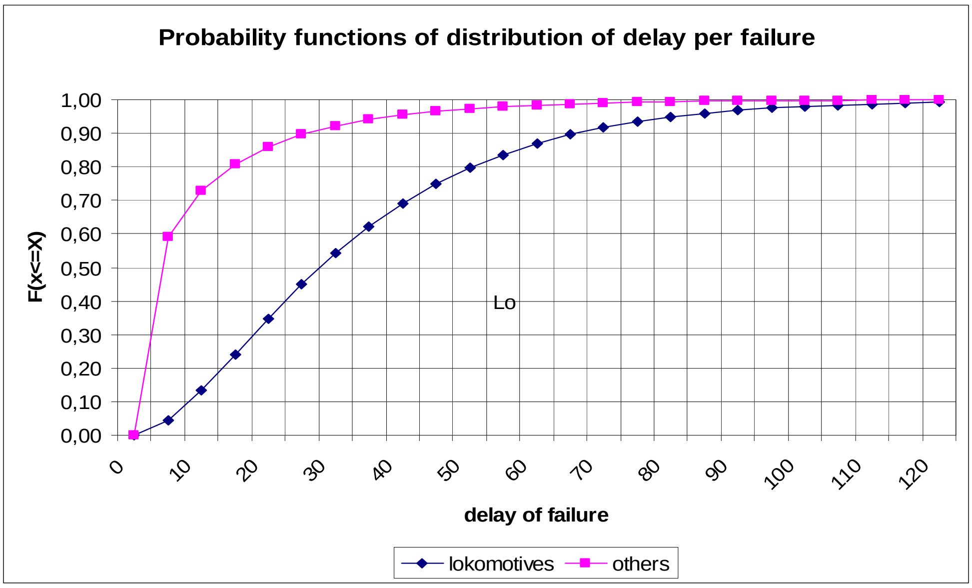

Fig.5

In the this diagrams are grouped as delays attributable to the wagon fleet and traffic management and control.

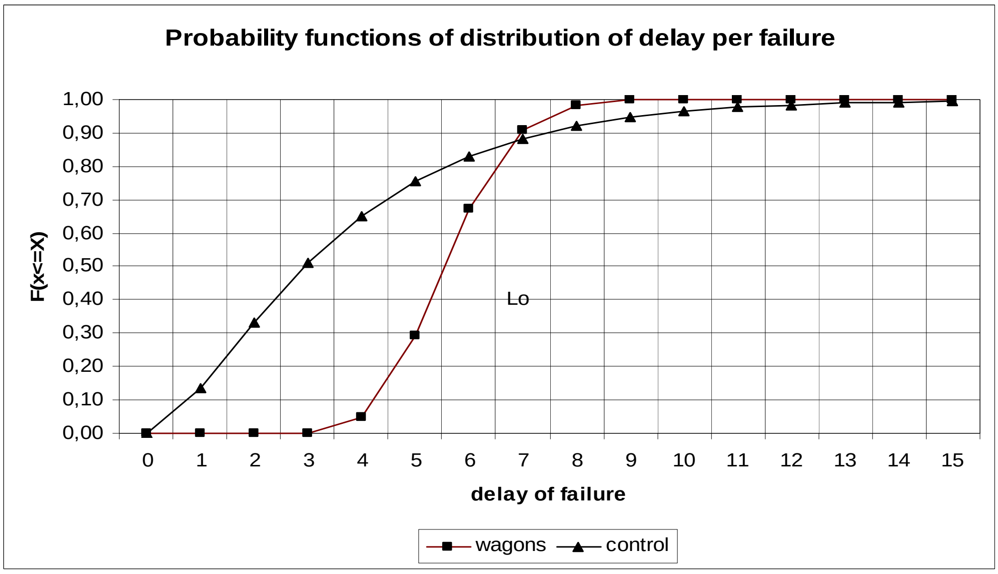

Fig.6

The results obtained for the parameters and distribution of delays duration by reasons are used in quantifying the life cycle costs (LCC) and analysis of reliability, availability, maintainability and safety (RAMS) of a newly designed car. That is part of the work program FP7-SST-2010-RTD-1 Grant agreement no: 265740 – SUSTRAIL " The sustainable freight railway: Designing the freight vehicle – track system for higher delivered tonnage with improved availability at reduced cost "aimed at analyzing the use of new design solutions with new rolling stock.

References

[1] Вентцель Е.С., Овчаров, задачи и упражнения по теории вероятностей, Moskwa,2003

[2] Двустранен протокол между Управителя на железопътната инфраструктура и Превозвача БДЖ, 2012 г.

[3] Taylor H.,Karlin S., An introduction to stochastics modeling 3-rd edition

[4] Tijms H.C., A first course in stochastic models, WILEY, 2003;

АНАЛИЗ НА ЗАКЪСНЕНИЯТА НА ТОВАРНИТЕ И ПЪТНИЧЕСКИ ВЛАКОВЕ ПО ЖЕЛЕЗОПЪТНАТА МРЕЖА НА РЕПУБЛИКА БЪЛГАРИЯ

Кирил Карагъозов, Мирена Тодорова
kkaragyozov@yahoo.com, mirena_todorova@abv.bg

ВТУ „Тодор Каблешков”, катедра “ТОУТ”
ул. Гео Милев 158, 1574 София,
БЪЛГАРИЯ

Ключови думи: *сума на случайни променливи, вероятностно разпределение, сумата на случайна променлива, анализ на данни от закъснения*

Резюме: *Анализът на случайните величини -закъснения на влаковете в следствие на различни причини (инфраструктура, превозвач, външни влияния) се основава на обработката на наличните статистически данни за брой и продължителността на отказите. Статистическите данни за тези причини са налични в агрегиран вид, което налага да се разработи модел за оценка на отделната продължителност на отказа. Използват се известните в Теория на вероятностите зависимости на централните моменти на случайна величина, която е сума на случаен брой независимо разпределени случайни величини. Този подход позволява да се намерят средната стойност и дисперсията на продължителността на отказа /закъснението/ и тези две характеристики да се използват за апроксимация на неизвестното разпределение на продължителността на отказа с Гама разпределение.*