



ASSESSMENT OF STATISTICAL PROBABILITY OF THE TECHNOLOGICAL TRANSPORTATION PROCESS

Adolfas Baublys

*Transport Research Institute, Vilnius Gediminas Technical University,
Plytinės g. 27, LT-2040, Lithuania. E-mail: abaublys@takas.lt*

Received 2002 05 20; accepted 2002 06 17

Abstract. In the course of freight transportation its technological process is influenced by lots of random factors. The classification of random factors in the whole hierarchical structure of the technological transportation process is presented together with the assessment of the dependence of separate random factors. Statistical information on freight and transport flows is renewed and replenished in the course of time. With the growth of information amount the cost of its storage increases as well. Therefore the relevant algorithms for obtaining required statistical assessment with the least statistical information are presented in the article. It is deduced that in the modelling of transport networks and freights as well as the flows of transport means in them, it is analytically proper to describe random factors by the non-parametric assessment.

Keywords: freight transportation; statistical information; random factors and their classification; minimal amount of stored statistical information.

1. Introduction

Aiming at successful freight transportation it is necessary to assess the factors influencing the transportation time, quality and the amount of the freight transported. At first sight it seems that various random factors constantly effect a strictly regulated process of transportation. Without the assessment of these factors with the help of determined models for planning and management of the transportation process its efficiency may decrease considerably.

Basing on the analysis of the world-wide known transport science magazines of the last 20 years, such as “*TRANSPORTATION SCIENCE*”, “*TRANSPORT THEORY AND STATISTICAL PHYSICS*”, “*TRANSPORTATION RESEARCH PART B - METHODOLOGICAL*” I have noticed, that in this field it is suggested to use statistical probability models. However, there are no articles about the random factors themselves, about their formation, classification, about the mathematical models necessary for the accumulation of statistical information in such a way that it should not be stored too long, so that having prior assessments of these factors and having supplemented them with new statistical information - automatically new assessment of the random factors would be attained.

2. Random Factors and Their Classification

Random factors have to be classified within the whole hierarchic structure of the technological transportation process. Such structure should comprise freight suppliers

and receivers, transport operators and forwarders, terminals with their technological processes, natural conditions, roads, i.e. everything that directly or indirectly influences the formation of random factors. Random factors may emerge because of subjective as well as objective reasons.

Objective reasons may be divided into three groups: technical, organisational and economical. The main technical reasons are - the condition of industrial basis of the transportation process (transport means, containers, lifting means, etc.) and the technological level of operations. Organisational reasons are: specialisation, scientific-technical information, means enhancing the quality of transportation, operation of customs and their subdivisions, moral responsibility for the quality of work, etc. Economical reasons are: tariffs, forms and levels of pay, work consumption, material incentives, etc.

The main subjective reasons include a professional, general and cultural level of employees, the emotional and psychological features of their characters, the adequacy of personal and public interests, their personal interest in the results of their work.

Let us discuss in detail the scheme of random factor impact on the transportation quality. A team and an operator should be interested in the quality of transportation: forming freight transport consignment, preparing documentation, packaging and handling freight during its transportation, etc. One of the regularities of random factor formation is the fact that the errors made in the earlier stage of transportation pass to the next stage. Effective and precise information in all stages facilitates the elimination of random factors.

The present classification of reasons of random factor occurrence does not pretend to be absolutely comprehensive, however, in our opinion, it demonstrates the complexity of the phenomena finally determining the efficiency of technological processes of transportation.

Formation and occurrence of random factors in the technological process of transportation is observed in the following stages: 1) scientific research work; 2) selection of the types of transportation; 3) formation of freight consignment; 4) determination of transportation routes; 5) freight handling/loading; 6) technological processes in terminals; 7) process of transportation (natural conditions, roads, etc.) 8) operation of customs. Reiterative random factors may occur because of various reasons, for instance one of them being an accident during transportation, the reasons of which may also vary: driver's fault, disorderly transport means, poor loading/handling, technical conditions of a road, meteorological conditions, etc.

Generally the technical-economical indices of a transportation process should be analysed as random factors, which in turn are random for every meaning of reason as well. The reasons are - the time or other parameters of the technological transportation process. Thus, the optimality criteria should also be considered as random, not determined.

The majority of indices defining the assessment of the technological transportation process is closely interconnected; therefore it has to be born in mind when these indices are used as the criteria of optimality. For the definition of the digital characteristics of technical-economical indices according to the statistical information, obtained in the technological transportation process, the assessment of mathematical probability, correlation and dispersion functions should be calculated.

3. Assessment of Interdependency of Different Random Factors

Different variable parameters at the beginning of the transportation process will be indicated by X_1, X_2, \dots, X_n , parameters determining the transportation process in progress will be indicated by Z_1, Z_2, \dots, Z_m , and variable parameters of the exit from the transportation process will be indicated by Y_1, Y_2, \dots, Y_s (Fig 1). Then initial random quantities marked by X_1, X_2, \dots, X_n will be analysed as components of the random vector X , the random quantities Z_1, Z_2, \dots, Z_m – as components of the random vector Z and the exit random quantities Y_1, Y_2, \dots, Y_s – as components of the random vector Y .

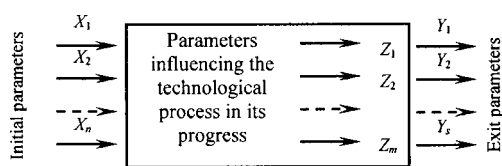


Fig 1. Scheme of the technological transportation process

It is not possible to limit ourselves only to the analysis of separate initial and variable factors characterising the quality of a transportation process in progress, because they are insufficient for obtaining comprehensive characteristics of the technical-economical indices of the transportation process as well as it is impossible to define the optimal variant of the transportation process management. The transportation process should be analysed as a multimeasurable process with a large number of initial parameters, the general assessment of which allows the complex evaluation of the efficiency of functioning of the technological transportation process.

It is obvious that according to the characteristics of the multimeasurable process, the meanings of every exit variable Y should be defined. Whereas the exit variables of the multimeasurable process may be independent, correlated or connected by the functional interdependence. Practically the characteristics of the transportation process should be analysed in each of the aforesaid cases. At the beginning we shall analyse the characteristics of the transportation process, when the initial variable quantities are interdependent Y_1, Y_2, \dots, Y_s .

Let us presume that general probability density of the vectored random quantities X, Z and Y is normal

$$\Phi_t(U_1, \dots, U_t) = \frac{1}{\sigma_{u_1}, \dots, \sigma_{u_t} \sqrt{(2\pi)^t E}} \times \exp \left\{ -\frac{1}{2E} \sum_{\mu, \nu=1}^t E_{\mu, \nu} \left(\frac{U_\mu - M\{U_\mu\}}{\sigma_{u_\mu}} \right) \left(\frac{U_\nu - M\{U_\nu\}}{\sigma_{u_\nu}} \right) \right\}, \quad (1)$$

here, for the sake of abridgement the random vector quantity U is introduced which is made of $X_1, \dots, X_n, Z_1, \dots, Z_m, Y_1, \dots, Y_s$; $t = n + m + s$; E - determinant of series t .

$$E = \begin{vmatrix} 1 & \rho_{u_1 u_2} & \dots & \rho_{u_1 u_t} \\ \rho_{u_2 u_1} & 1 & \dots & \rho_{u_2 u_t} \\ \dots & \dots & \dots & \dots \\ \rho_{u+u_1} & \rho_{u+u_2} & \dots & 1 \end{vmatrix} \quad (2)$$

or X, Z, Y in the variables marking

$$E = \begin{vmatrix} 1 & \rho_{x_1 x_2} & \dots & \rho_{x_1 x_n} & \rho_{x_1 z_1} & \dots & \rho_{x_1 z_m} & \rho_{x_1 y_1} & \dots & \rho_{x_1 y_s} \\ \rho_{x_2 x_1} & \rho_{x_2 x_2} & \dots & 1 & \rho_{x_2 z_1} & \dots & \rho_{x_2 z_m} & \rho_{x_2 y_1} & \dots & \rho_{x_2 y_s} \\ \rho_{z_1 x_1} & \rho_{z_1 x_2} & \dots & \rho_{z_1 x_n} & 1 & \dots & \rho_{z_1 z_m} & \rho_{z_1 y_1} & \dots & \rho_{z_1 y_s} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{z_m x_1} & \rho_{z_m x_2} & \dots & \rho_{z_m x_n} & \rho_{z_m z_1} & \dots & 1 & \rho_{z_m y_1} & \dots & \rho_{z_m y_s} \\ \rho_{y_1 x_1} & \rho_{y_1 x_2} & \dots & \rho_{y_1 x_n} & \rho_{y_1 z_1} & \dots & \rho_{y_1 z_m} & 1 & \dots & \rho_{y_1 y_s} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{y_s x_1} & \rho_{y_s x_2} & \dots & \rho_{y_s x_n} & \rho_{y_s z_1} & \dots & \rho_{y_s z_m} & \rho_{y_s y_1} & \dots & 1 \end{vmatrix}$$

$E_{\mu\nu}$ is an algebraic supplement $\rho_{\mu\nu}$ in the determinant (2).

For the analysis of the case when the exit variables are independent it is necessary to determine the characteristics of every variable Y_k ($k = 1, 2, \dots, s$) as well as the influence exercised on them by the initial variables X and the variables Z , characterising the inner state of the process. Let us indicate the general probabilities' X, Z and Y density $\varphi_{n+m+1}(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m)$, whereas the random vectored quantities' X and Z by the general probability density $\varphi_{n+m}(X_1, \dots, X_n, Z_1, \dots, Z_m)$. The probability densities $\varphi_{n+m+1}(\varphi_k, X_1, \dots, X_n, Z_1, \dots, Z_m)$ and $\varphi_{n+m}(X_1, \dots, X_n, Z_1, \dots, Z_m)$ are not zero ones and they correspond to the equation (1) with determinant σ and F meanings correspondingly of $(n+m+1)$ and $(n+m)$ series

$$\sigma = \begin{vmatrix} 1 & \rho_{x_1x_2} & \dots & \rho_{x_1z_m} & \rho_{x_1y_k} \\ \rho_{x_2x_1} & 1 & \dots & \rho_{x_2z_m} & \rho_{x_2y_k} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{z_mx_1} & \rho_{z_mx_2} & \dots & 1 & \rho_{z_my_k} \\ \rho_{y_kx_1} & \rho_{y_kx_2} & \dots & \rho_{y_kz_m} & 1 \end{vmatrix}; \quad (3)$$

$$F = \begin{vmatrix} 1 & \rho_{x_1x_2} & \dots & \rho_{x_1x_n} & \rho_{x_1z_1} & \dots & \rho_{x_1z_m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{x_nx_1} & \rho_{x_nx_2} & \dots & 1 & \rho_{x_nz_1} & \dots & \rho_{x_nz_m} \\ \rho_{z_1x_1} & \rho_{z_1x_2} & \dots & \rho_{z_1x_n} & 1 & \dots & \rho_{z_1z_m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{z_mx_1} & \rho_{z_mx_2} & \dots & \rho_{z_mx_n} & \rho_{z_mz_1} & \dots & 1 \end{vmatrix}. \quad (4)$$

General characteristic of the technological transportation process is a conditional probability density $\varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m))$, according to which the meaning may be defined by the general characteristics of the prior variables X and the inner state variables Z by transformation of the distribution laws of these random quantities.

Generally, the conditional density

$\varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m))$ is determined by the general probability densities

$$\varphi_{n+m+1}(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m) \text{ and}$$

$$\varphi_{n+m}(X_1, \dots, X_n, Z_1, \dots, Z_m):$$

$$\varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)) =$$

$$\frac{\varphi_{n+m+1}(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m)}{\varphi_{n+m}(X_1, \dots, X_n, Z_1, \dots, Z_m)}. \quad (5)$$

If common probability densities

$$\varphi_{n+m}(X_1, \dots, X_n, Z_1, \dots, Z_m) \text{ and}$$

$\varphi_{n+m+1}(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m)$ are normal, then for the analysed case the conditional probability density is:

$$\begin{aligned} \varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)) = \\ \frac{1}{\sqrt{2\pi}\sigma_{y_k}} \sqrt{\frac{F}{\sigma}} \exp \left\{ \frac{1}{2F} \sum_{k,l=1}^{n+m} F_{kl} \times \right. \\ \left. \left(\frac{v_k - M\{v_k\}}{\sigma_{v_k}} \right) \left(\frac{v_l - M\{v_l\}}{\sigma_{v_l}} \right) - \frac{1}{2\sigma} \sum_{m,v=1}^{n+m+1} \sigma_{\mu\nu} \times \right. \\ \left. \left(\frac{U_\mu - M\{U_\mu\}}{\sigma_{\mu\mu}} \right) \left(\frac{U_\nu - M\{U_\nu\}}{\sigma_{\nu\nu}} \right) \right\}, \end{aligned}$$

here $F_{k,l}$ – of element $\rho_{k,l}$ algebraic supplement in the determinant (4); $\sigma_{\mu,\nu}$ – element $\rho_{\mu,\nu}$ algebraic supplement in the determinant (3) $v(X_1, \dots, X_n, Z_1, \dots, Z_m)$, $U(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m)$. Given the conditional density $\varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m))$ for the definition of characteristics X and Z general characteristics of the exit variable Y , i.e. the probability density $\varphi_{y_k}(Y_k)$ is defined with the help of the following integral equation:

$$\begin{aligned} \varphi_{y_k}(Y_k) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)) \varphi_{n+m} \times \\ (X_1, \dots, X_n, Z_1, \dots, Z_m) dX_1, \dots, dX_n, dZ_1, \dots, dZ_m. \end{aligned}$$

Knowing general characteristics $\varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m))$ we may also define the main digital characteristics of X and Z in the aspect of the exit variable Y_k , i.e. the conditional mathematical probability and conditional dispersion:

$$M\{Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)\} =$$

$$\frac{\int_{-\infty}^{\infty} Y_k \varphi_{n+m+1}(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m) dy_k}{\int_{-\infty}^{\infty} \varphi_{n+m+1}(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m) dy_k} =$$

$$\int_{-\infty}^{\infty} Y_k \frac{\varphi_{n+m+1}(Y_k, X_1, \dots, X_n, Z_1, \dots, Z_m)}{\varphi_{n+m}(X_1, \dots, X_n, Z_1, \dots, Z_m)};$$

thus we shall assess the equation (5) and shall obtain

$$M\{Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)\} =$$

$$\int_{-\infty}^{\infty} Y_k \varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)) dy_k.$$

Whereas the conditional dispersion Y_k in the aspect of X and Z

$$D\{Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)\} =$$

$$\int_{-\infty}^{\infty} [Y_k - M\{Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)\}]^2 \times$$

$$\varphi(Y_k / (X_1, \dots, X_n, Z_1, \dots, Z_m)) dy_k .$$

Thus, defining the characteristics of the technological transportation process, when the random factors are assessed, it is necessary to determine the unconditional and conditional dispersion laws. For this reason the constant accumulation of statistic information is necessary.

For practical purposes it is often expedient to use digital characteristics instead of the random factor distribution laws. Although the digital characteristics give insufficient information on random factors, however, for the solution of certain transportation issues they fully suffice and their determination is far easier. A complete analysis and synthesis of the characteristics of the transportation processes is carried out according to the general characteristics, i.e. according to the unconditional and conditional distribution laws. These laws may be employed for the definition of different characteristics of the technological transportation processes. However, as the presented formula shows, the characteristics of the technological transportation processes may be determined in such cases when the general laws of random quantity distribution are known. This condition must be observed in the designing of systems for the accumulation and processing of statistical information.

If the exit variables are correlated, for the management of the technological transportation process the characteristics of one or several exit variables may be used.

4. Accumulation of Statistical Information on the Transportation Process

4.1. Non-parametrical Assessment of Distribution Density

In the modelling of transport networks and the flows of freight and transport means the analytical description of the density distribution of their random factors is carried out by parametric assessment. In certain issues the distribution type of analysed factor density in the course of time is altering (new nomenclature of consignment, shifting to other working conditions, etc.), therefore the distribution description by the law of a selected type is inaccurate and absolutely impossible (from the point of view of calculation). Besides, the approximation of empirical distributions using parametric assessments is very sensitive to data errors. Even one mistake made during the introduction of the prior information x_1, \dots, x_N into the computer may exercise such a great effect on the meanings of sampling moments, that the analytical description of the curves will be very inexact. Unlike the parametric - the non-parametric assessment of density distribution has a number of merits. Firstly, when in the course of time the type of flows distribution density changes, it is possible

to approximate by the exactness sufficient enough for practical use. Secondly, this assessment is far less sensitive to data errors.

In the case of non-parametric assessment it may be written:

$$f(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi\lambda}} e^{-(x-x_i)^2 / 2\lambda^2} , \tag{6}$$

here λ is the coefficient of equalisation.

Since in the given assessment we have to remember all sample realisations: x_1, \dots, x_N , we may use the following assessment:

$$f(x) = - \frac{1}{\sqrt{2\pi\lambda}} e^{-y^2 / 2\lambda^2} \sum_{j=1}^t C_j y_j , \tag{7}$$

here the coefficients C_j are defined by the formula:

$$C_j = \frac{1}{j!(\lambda)^{2j}} \frac{1}{N} \sum_{i=1}^N y_i^j e^{-y_i^2 / 2\lambda^2} , \tag{8}$$

here $y = x - \Delta$; $y_i = x_i - \Delta$, and the coefficient Δ equals approximately to the mathematical probability of distribution.

In the phenomenon (7) the member number t is found in an experimental way. Since even because of very “curved” distributions and very inaccurate data the densities calculated according to the (7) using four members coincide, so further we shall presume that because of the (7) $t = 4$. Non-parametric assessment of such type requires remembering only of the coefficients $\lambda, \Delta, C_0, C_1, \dots, C_t$, i.e. the $t + 3$ parameters, that simplify the solution of the work on the problem.

Selection of equalisation coefficients. Non-parametric dispersion assessment includes the equalisation coefficient λ , the optimal meaning of which is determined by the distribution dispersion and sample size: the larger dispersion and the smaller is the sample, the larger is λ . If λ is very small, then even in case of a very large sample the non-parametric assessment will be very sensitive and consequently imprecise, otherwise the unevenness of the distribution density profile will not be assessed.

Let us analyse the case when the distribution density

$$f(x) = \sum_{i=1}^k P_i \frac{1}{\sqrt{2\pi\sigma_i}} e^{-(x-\sigma_i)^2 / 2\sigma_i^2} . \tag{9}$$

The latter formula practically allows good approximation of the most observed distributions in freight flows research.

For the selection of an optimal equalisation parameter we shall apply the approximation quality criterion.

$$U = \frac{1}{Q} \int_{-\infty}^{\infty} E_x [f_N(x) - f(x)]^2 dx \tag{10}$$

here the E_x – is the mathematical probability of the quantity N , $Q = \int_{-\infty}^{\infty} f^2(x)dx$. Having indicated

$$\frac{1}{\sqrt{2\pi a}} e^{-(x_i-b)^2/2a^2} = N(x, b, a^2), \tag{11}$$

the expressions (6) and (9) will be written

$$f_N(x) = \frac{1}{N} \sum_{i=1}^N (x, x_i, \lambda^2), \tag{12}$$

$$f(x) = \sum_{j=1}^k P_j N(x, \mu_j, \sigma_j^2). \tag{13}$$

Calculating the (10) we shall bear in mind, that

$$\int_{-\infty}^{\infty} N(x, a_1 b_1) N(x, a_2, b_2) dx = N(a_1, a_2, b_1 + b_2), \tag{14}$$

$$\int_{-\infty}^{\infty} [N^2(x, a_1, b_1)]^2 N(x, a_2, b_2) dx = \sqrt{\frac{2b_2 + b_1}{a}} [N(a_1, a_2, 2b_2 + b_1)]. \tag{15}$$

Thus we obtain

$$E_x [f_N(x) - f(x)]^2 = E_x [f_N(x)]^2 - 2E_x f_N(x) f(x) + f^2(x), \tag{16}$$

$$\frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} N(x, x_i, \lambda^2) \prod_{i=1}^N f(x_i) dx_1, \dots, dx_N. \tag{17}$$

Since

$$\int_{-\infty}^{\infty} f(x_i) dx_i = 1,$$

the expression (17) may be written in the following way:

$$E_x f_N(x) = \sum_{s=1}^k P_s N(x, \mu_s, \lambda^2 + \sigma_s^2). \tag{18}$$

Further we calculate:

$$E_x [f_N(x)]^2 = \frac{1}{N} \sum_{s=1}^k P_s \sqrt{\frac{2\sigma_s^2 + \lambda^2}{\lambda^2}} N(x, \mu_s; 2\sigma_s^2 + \lambda^2) +$$

$$\left(1 - \frac{1}{N}\right) \sum_{s=1}^k P_s N(x, \mu_s, \sigma_s^2 + \lambda^2)^2. \tag{19}$$

After putting the (19) and (18) into (16) and (10), and afterwards applying the equation (14) we shall obtain

$$U = \left[\frac{1}{N} \sum_{s=1}^k P_s \sqrt{\frac{\lambda^2 + 2\sigma_s^2}{\lambda}} N(\mu_s, \mu, 4\sigma_s^2 + 2\lambda)^2 + \left(1 - \frac{1}{N}\right) \sum_{s,l=1}^k P_s P_l N(\mu_s, \mu_l, \sigma_s^2 + \sigma_l^2 + 2\lambda^2) - 2 \sum_{s,l=1}^k P_s P_l N(\mu_s, \mu_l, \sigma_s^2 + \sigma_l^2 + \lambda^2) + \sum_{s,l=1}^k P_s P_l N(\mu_s, \mu_l, \sigma_s^2 + \sigma_l^2) \right] / Q \tag{20}$$

here

$$Q = \sum_{s,l=1}^k P_s P_l N(\mu_s, \mu_l, \sigma_s^2 + \sigma_l^2).$$

The latter (20) expression enables the finding of optimal meaning μ for the distribution described in the formula (9). Unknown parameters are included in this formula.

Presuming that k is set (practically good results are obtained when $k = 3 \div 5$), the parameters P_i, μ_i, σ_i may be found by the way of the iterative procedures.

The algorithm of assessment finding is formed by the following steps:

1) we set $P_1^{(1)}, \dots, P_k^{(1)}, \mu_1^{(1)}, \dots, \mu_k^{(1)}, \sigma_1^{(1)}, \dots, \sigma_k^{(1)}$;

2) we calculate

$$\alpha_{ij}^{(t)} = \frac{P_j^{(t-1)} N(x_i, \mu_j, \sigma_j^{(t-1)})}{\sum_{j=1}^k P_j^{(t-1)} N(x_i, \mu_j^{(t-1)}, \sigma_j^{(t-1)})}, i, j=1, \dots, N;$$

3) we obtain new meanings $P_1^{(t)}, \dots, P_k^{(t)}, \mu_1^{(t)}, \dots,$

$\mu_k^{(t)}, \sigma_1^{(t)}, \dots, \sigma_k^{(t)}$ according to the formulae

$$P_j^{(t)} = \frac{\sum_{i=1}^N \alpha_{ij}^{(t)}}{N},$$

$$\mu_j^{(t)} = \frac{\sum_{i=1}^N \alpha_{ij}^{(t)} x_i}{\sum_{i=1}^N \alpha_{ij}^{(t)}} = \frac{\sum_{i=1}^N \alpha_{ij}^{(t)} x_i}{P_j^{(t)} N},$$

$$\sigma_j^{(t)} = \frac{\sum_{i=1}^N \alpha_{ij}^{(t)} (x_i - \mu_j^{(t)})^2}{P_j^{(t)} N}.$$

We repeat the process starting from the second step until the meanings P_i, μ_i, σ_i will vary within narrow limits, i.e. until we reach the desired precision.

Results of experimental calculations for the coefficient of optimal equalisation

Sample number	Quantity of the sample N	Optimisation according to the formula (20)	
		λ_{opt}	U_{min}
1	40	0,844	0,0472
2	140	0,623	0,0199
3	40	0,858	0,04655
4	40	0,709	0,0623
5	113	0,551	0,0297
6	40	1,293	0,0679
7	181	0,894	0,0230
8	40	0,867	0,0586
9	131	0,656	0,0258

Experimental calculations were carried out, i.e. according to the above formulae the coefficients $\mu_1, \dots, \mu_k, \sigma_1, \dots, \sigma_k, P_1, \dots, P_k$ were calculated included into (9) and with the help of minimisation of (20) the optimal meanings of the equalisation coefficient λ_{opt} were obtained, as well as the meaning of U_{min} was defined corresponding to λ_{opt} . The results of the experiment are presented in Table, which shows that using the formula obtained by us (9) the accurateness of approximation of density of random quantities is significantly more precise, the equalisation parameter defined according to (20) is 1,2–2 times less than in the case of normal distribution. That is why it is expedient to use the formula (9) for the optimisation of the parameter λ and for the description statistical data of freight flows.

4.2. Mathematical Models in Accumulation of Statistical Information

Statistical information on the freight and transport flows is renewed and replenished in due course. With the growth of the amount of the information the costs of its storage are increasing too. Therefore, our aim is to obtain necessary statistical assessment with the minimal amount of statistical information.

The assessment of distribution of random factors by the Bayesian method. Presumably we have a sample X , made of N meanings x_1, \dots, x_N . Considering the meanings x as random quantities we shall assess distribution function of the random quantity X . For the analytical description of the distribution X , relevant prior information on the distribution type is necessary. Let us presume that it is known, i.e. the parametric probabilities density $f(X/\theta)$ is known here

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_r \end{pmatrix} - \text{vector of the parameters describing the}$$

distribution of the random quantity x ; r – is the number of parameters and presumably known the prior distribution of the vector parameters $\theta f_{apr.}(\theta)$, the latter may be defined by experimental assessment. Then X distribution may be assessed with the help of the Bayesian formula [1].

At the beginning we shall obtain the aposterioric density of distribution of the vector parameters. Considering that observations in the sample x_1, \dots, x_N are independent, according to the Bayesian formula it is as follows:

$$f(\theta / x_1, \dots, x_N) = \frac{\prod_{i=1}^N f(x_i / \theta) f_{apr.}(\theta)}{\int_{\Omega_\theta} \prod_{i=1}^N f(x_i / \theta) f_{apr.}(\theta) d\theta}, \quad (21)$$

here the integration is in r – measurable area Ω_θ by changing the parameter θ . The aposterioric distribution of X is obtained integrating

$$f(X / x_1, \dots, x_N) = \int_{\Omega_\theta} f(X / \theta) f(\theta / x_1, \dots, x_N) d\theta. \quad (22)$$

Putting (21) into (22) we obtain:

$$f(X / x_1, \dots, x_N) = \frac{\int_{\Omega_\theta} f(X / \theta) \prod_{i=1}^N f(x_i / \theta) f_{apr.}(\theta) d\theta}{\int_{\Omega_\theta} \prod_{i=1}^N f(x_i / \theta) f_{apr.}(\theta) d\theta}. \quad (23)$$

After the new sample of the phenomena x_{n+1}, \dots, x_{n+m} is obtained and it is necessary to specify the distribution $f(X)$ using new data, then instead of the prior distribution $f_{apr.}(\theta)$ we shall use the former prior one $f(\theta / x_1, \dots, x_N)$. Let us notice that such method does not require the storage of all the meanings of a sample; only the meanings of the aposterioric distribution $f(\theta / x_1, \dots, x_N)$ are subject to storage, namely – only a certain number of coefficients characterising it. At the beginning of the data accumulation process we may not have $f_{apr.}$ on the whole, then the Bayesian postulate is applied and the prior distribution is considered uniform [1].

The presented formulae would lose their practical importance if they do not have concrete distribution types $f(X/\theta)$ and $f_{apr.}(\theta)$. Besides, it is preferable that the analytical type of the prior and aposterioric density would be uniform, because then the calculation algorithms and programmes become simpler.

Normal distribution. Let us analyse the case when random qualities are distributed according to the normal law, the density of which is

$$f(X/\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right].$$

In the given case the parameter vector $\theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$, and

$$f(x_1, \dots, x_N/\mu, \sigma) = C_0 \sigma^{-N} \exp\left(-\frac{1}{2} \frac{S(N-1)}{\sigma^2}\right) \exp\left(-\frac{1}{2} \frac{N(\bar{x}-\mu^2)}{\sigma^2}\right), \quad (24)$$

here $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$; $S = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$ - the sample medium and dispersion accordingly $C_0 = (2\pi)^{-N/2}$.

The prior distribution of parameters μ and σ is selected so that its density would be analogous to the density of the conditional density

$$f(\mu, \sigma) = C_1 \sigma^{-n} \exp\left[-\frac{1}{2} \frac{a(n-1) + n(\mu + \mu_0)^2}{\sigma^2}\right]. \quad (25)$$

The coefficient C_1 depends on n and may be obtained from the rationing condition. The coefficient C_1 is not calculated here, because it does not depend on μ and σ . After putting (25) into (21), and after corresponding rearrangement, we shall obtain

$$f(\sigma, \mu/x_1, \dots, x_N) = \frac{1}{C_3} \sigma^{-N^*} \exp\left[-\frac{1}{2} \frac{(N^*-1)a^* + N^*(\mu - \mu^*)^2}{\sigma^2}\right], \quad (26)$$

here

C_3 is obtained from the rationing condition;

$$N^* = N + n;$$

$$a^* = S \frac{N-1}{N+n-1} + a \frac{n-1}{N+n-1} + \frac{Nn(\mu_0 - \bar{x})^2}{(N+n)[(N+n)-1]}; \quad (27);$$

$$\mu^* = \bar{x} \frac{N}{N+n} + \mu_0 \frac{n}{N+n}. \quad (28)$$

Now let us calculate the coefficient C_3 . Having noticed, that

$$\frac{1}{\sqrt{2\pi}\sigma/\sqrt{N^*}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \frac{(\mu - \mu^*)^2}{\sigma^2/N^*}\right) d\mu = 1,$$

after the relevant rearrangements

$$C_3 = \frac{1}{2} \sqrt{\frac{2\pi}{N^*}} \left[\frac{a^*(N^*-1)}{2} \right]^{(N^*-2)/2} \Gamma\left(\frac{N^*-2}{2}\right).$$

Thus the aposterioric distribution of the parameters σ and μ

$$\sigma^{-N^*} \exp\left[-\frac{1}{2} \frac{(N^*-1)a^* + N^*(\mu - \mu^*)^2}{\sigma^2}\right] = \frac{\sigma^{-N^*} \exp\left[-\frac{1}{2} \frac{(N^*-1)a^* + N^*(\mu - \mu^*)^2}{\sigma^2}\right]}{\sqrt{\frac{\pi}{2N^*}} \left[\frac{a^*(N^*-1)}{2} \right]^{(N^*-2)/2} \Gamma\left(\frac{N^*-2}{2}\right)}. \quad (29)$$

The aposterioric density of x distribution

$$f(X/x_1, \dots, x_N) = \frac{\Gamma\left(\frac{N^*-1}{2}\right)}{\sqrt{2\pi} \Gamma\left(\frac{N^*-2}{2}\right) \sqrt{\frac{a^*(N^*-1)}{2}}} \times \left[1 + \frac{N^*(X - \mu^*)^2}{(N^*-1)a^*} \right]^{-(N^*-1)/2}. \quad (30)$$

The aposterioric density may be approximated by the normal one (23) putting instead of σ and μ more probable meanings obtained from (29). By differentiation we obtain the following more expected meanings of the parameters $\mu = \hat{\mu}$, $\hat{\sigma}^2 = a^*$. Then the aposterioric density is approximated by the phenomenon

$$f(X/\hat{\mu}, \hat{\sigma}) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left[-\frac{1}{2} \frac{(X - \hat{\mu})^2}{\hat{\sigma}^2}\right]. \quad (31)$$

In the presence of large N (31), (30) approximates the density with great precision. As (30) and (31) include only parameters $N^* = N + n$, a^* (or $\hat{\sigma}$), μ^* (or $\hat{\mu}$), then every aposterioric distribution may be very simply used as an aprioric one. For this purpose only the parameters are recalculated

$$N_i = N_{i-1} + N,$$

$$\hat{\mu}_i = \mu_{i-1} \frac{N_{i-1}}{N_{i-1} + N} + \frac{N}{N_{i-1} + N} \bar{x},$$

$$\hat{\sigma}_i^2 = \hat{\sigma}_{i-1}^2 \frac{N_{i-1} - 1}{N + N_{i-1} - 1} + S \frac{N}{N + N_{i-1} - 1} +$$

$$\frac{NN_i(\mu_i - \bar{x})^2}{N + N_i(N + N_{i-1})}.$$

Consequently suffice it to remember only three coef-

ficients: N_{i-1} , μ_{i-1} and σ_{i-1} . To demonstrate under which given N it is possible to use the simple (31) instead of the complicated formula (30) let us analyse the following example.

The sample, having 100 random quantity meanings, characterising the quantity alteration of freight consignment brought to the terminal, was divided into 20 portions with 5 realisations each. The meanings of the parameters of the aprioric distribution are: $\mu_0 = 300$, $\sigma_0 = 10$, $N_0 = 1$. According to the first sample we calculated x and s^* , and further according to (27) and (28) a^* and μ^* . According to these parameters we have calculated the meanings of the aposterioric density by the precise formula (30) and the approximate formula (31). Further on we use the meanings a^* and μ^* as the parameters of aprioric distribution and we repeat the calculation procedure for the second, the third, etc., samples. The diagrams of the distribution density are presented in Fig 2, where the continuous line shows the distribution obtained according to (30), and the dotted line – those according to (31).

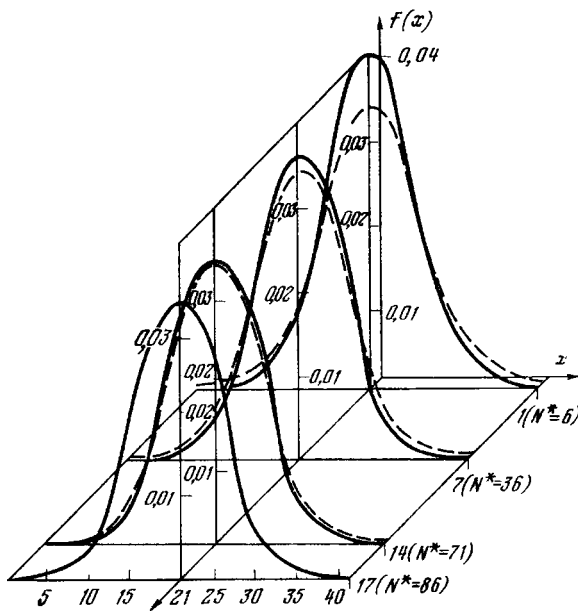


Fig 2. The diagrams of the distribution density

In the seventeenth interval the curves of the distribution density practically coincide completely. Already in the presence of $N = 70 - 85$ both density curves practically coincide and when the sample exceeds 100 it is possible to use much more simple (normal) density.

Logarithmic-normal distribution. As experiments have proved, using logarithmic normal distribution

$$f(x) = \frac{1}{x\beta\sqrt{2\pi}} \exp\left(-\frac{\ln(x-a)^2}{2\beta^2}\right)$$

it is possible to introduce simplification without losing precision (sufficient for practice). Namely, if indicated

$y = \ln x$, then random quantity y is distributed according to the normal density:

$$f(y) = \frac{1}{\sqrt{2\pi}\beta} \exp\left(-\frac{(y-a)^2}{2\beta^2}\right) \tag{32}$$

There may occur difficulties in the interpretation of the results with (32), however during the calculation the simplicity of the formula (32) is very clear. For example, we need to calculate probability limits for x . Since y – monotharic function x , it is possible to define probability limits y : y_1 and y_2 (probability limits x : $x_1 = e^{y_1}$ and $x_2 = e^{y_2}$).

The parameters a and β are correspondingly the medium and dispersion of the rearranged random quantity $y = \ln x$ and they are defined

$$a = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \sum_{i=1}^n \ln x_i, \tag{33}$$

$$\beta^2 = \frac{1}{N-1} \left(\sum_{i=1}^n (\ln x_i)^2 - a^2 \right) \tag{34}$$

Thus, the Bayesian model of normal distribution may be used for logarithmic normal distribution as well, after its relevant rearrangement.

Exponential distribution, characterised by one parameter

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^N x_i,$$

which is assessed according to a sample.

After obtaining the new data the meaning was corrected

$$\hat{\lambda}_n = \frac{N\lambda_{c_i} + \sum_{i=1}^{N_N} x_i}{N_N + N},$$

here N – sample quantity according to which there were assessed earlier λ meanings; N_N – the amount of realisations obtained.

Non-parametric distribution. It was demonstrated earlier that the optimal meaning of the equalisation coefficient depends upon the sample quantity N . Therefore the coefficients C_j are the main function λ , then, having obtained new information, the coefficient λ has to be changed and it is not possible to define the meanings of C_j by the formula (8). Therefore it has to be redone.

Since

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \dots, \tag{35}$$

then (8) may be written in this way:

$$C_j = \frac{1}{j!} \frac{1}{\lambda^2 j N} \times$$

$$\sum_{i=1}^N y_i^j \left(1 - \frac{y_i^2}{2\lambda^2} + \frac{y_i^4}{2(2\lambda^2)^2} - \frac{y_i^6}{6(2\lambda^2)^3} + \dots \right). \quad (36)$$

Limited by the first members of the line, we have:

$$C_j = \frac{1}{j! \lambda^{2j}} \frac{1}{N} y_i^j \sum_{s=0}^k \frac{(y_i)^{2s} (-1)^s}{s! 2^s \lambda^{2s}}. \quad (37)$$

After changing the order of summing, we shall obtain:

$$C_j = \frac{1}{j!} \sum_{s=0}^k \frac{(-1)^s}{s! 2^s \lambda^{2(s+j)}} b_{2s+j}, \quad (38)$$

here

$$b_t = \frac{1}{N} \sum_{i=1}^N (y_i)^t.$$

Thus, for the purpose of data accumulation (rehabilitation of the distribution densities) it is necessary to store $1+2k$ the coefficient b_t . After receiving new information it is necessary to obtain by (20) the meaning of the optimal equalisation coefficient λ_{opt} for the new sample quantity, afterwards to recalculate the meaning b_t by the formula

$$b_{t_n} = \frac{N_s}{N_s + N_g} b_{t_s} + \frac{1}{N_s + N_g} \sum_{i=1}^{N_g} y_i^t,$$

here b_{t_s} – the old meaning of the coefficient b_t . And N_s – sample quantity before the obtaining of new data; N_g – quantity of obtained data.

According to the calculated meanings b_1, \dots, b_{t+2k} and formulae (38) and (7) it is possible to obtain the assessment of distribution density.

In the course of calculation it is necessary to determine the degree of the polynomial (35). As it is apparent from (8), the sets to which $y_i / 2\lambda^2 > 8 \div 10$, practically will have no influence on the meaning of the coefficient C_j . Therefore the members, which $y_i > 20\lambda^2$, in calculation of the coefficients b_t , should be dismissed. Then in the line (36) the maximum member $\left[\frac{y_i^2}{2\lambda^2} \right]$ will not exceed 10. Thus k is obtained on condition that the member remaining in the line (36) is far less than one, i.e.

$$\frac{\left\{ \left[\frac{y_i^2}{2\lambda^2} \right]_{\max} \right\}^k}{K!} \ll 1.$$

In reality with $10^k / 10! \leq 0,001$ we obtain $k \approx 30$.

The Weibull distribution is characterised by two parameters and its density [2]

$$f(x) = \frac{m t^{m-1}}{x_0} \exp \left[- \left(\frac{t}{x_0} \right)^m \right], t \geq 0, m > 0, x_0 > 0,$$

here m – the parameter of form; x_0 – the parameter of scale. This distribution is used in the research of reliability of transport and loading means, in description of failure probability (breakdown).

Let us presume that during certain time we analyse n amount of transport means, and during the time $r, 1 \leq r \leq n$, transport means failed – broke down, got out of order. The time of failure will be defined t_1, \dots, t_r . This data may be used in calculation of the Weibull distribution parameters m and x_0 .

The latter parameters may also be defined by the method of maximum similarity; for this purpose it is necessary to solve the non-linear equations

$$x = \frac{1}{r} \left[\sum_{i=1}^r x_i^m + (n-r)x_r^m \right],$$

$$x = \frac{\sum_{i=1}^r x_i^m \ln x_i + (n-r)x_r^m \ln x_r}{\frac{r}{m} + \sum_{i=1}^r \ln x_i}.$$

Since into the statistics $\sum t_i^m, \sum t_i^m \ln t_i$ an unknown parameter m is included, the whole sample should be used in the calculations, for which cause the storage of not only certain, but of all realisations is necessary.

Accumulation of data. As it was said before, the storage of all data is not convenient and not possible, however a precise method of defining the parameters m and x_0 does not exist. Therefore we shall make a compromise: we shall assess the parameters by the method of moments.

Let us presume that the first two moments of the sample are known

$$m_1 = \frac{1}{r} \sum_{i=1}^r t_i, m_2 = \frac{1}{r} \sum_{i=1}^r t_i^2. \quad (39)$$

Precise meanings of the prior moments of the Weibull distribution [3].

$$m_1 = x_0^{1/m} \Gamma \left(1 + \frac{1}{m} \right), m_2 = x_0^{2/m} \Gamma \left(1 + \frac{2}{m} \right). \quad (40)$$

Having equalised the meanings of sample parameters to their analytical expressions, we shall obtain the system of equations enabling to find x_0 and m .

$$x_0^{1/m} \Gamma \left(1 + \frac{1}{m} \right) = m_1, x_0^{m/2} \Gamma \left(1 + \frac{2}{m} \right) = m_2.$$

After the rearrangement of the system of non-linear equations, i.e. after division of the second one from the first square

$$\frac{x_0^{2/m} \Gamma \left(1 + \frac{2}{m} \right)}{x_0^{2/m} \Gamma \left(1 + \frac{1}{m} \right) \Gamma \left(1 + \frac{1}{m} \right)} = \frac{\hat{m}_2}{\hat{m}_1^2}.$$

Since $\Gamma(a+1) = \Gamma(a) \cdot a$, then

$$\frac{\Gamma\left(\frac{2}{m}\right)\frac{2}{m}}{\Gamma\left(\frac{1}{m}\right)\Gamma\left(\frac{1}{m}\right)\frac{1}{m^2}} = \frac{2m\Gamma\left(\frac{2}{m}\right)}{\Gamma\left(\frac{1}{m}\right)\Gamma\left(\frac{1}{m}\right)} = \frac{\hat{m}_2}{\hat{m}_1^2}. \quad (41)$$

By the application of the LaGrange's formula for the gamma function argument of equation [3]

$$\Gamma(2z) = 2^{2z-1} \pi^{-1/2} \Gamma(z) \Gamma\left(z + \frac{1}{2}\right),$$

from (41) we obtain much more convenient phenomenon for obtaining m .

$$\hat{m}_1^2 2^{2/m} m \Gamma\left(\frac{1}{m} + \frac{1}{2}\right) - \hat{m}_2 \sqrt{\pi} \Gamma\left(\frac{1}{m}\right) = 0. \quad (42)$$

The quantity m is obtained solving the equation (42), and x_0 – by the obtained meaning and by one of (40) equations, for example

$$x_0 = \left[\frac{m_1}{\Gamma\left(1 + \frac{1}{m}\right)} \right]^m. \quad (43)$$

Searching for the parameters x_0 and m by (42), (43) and (39), stored are only the statistics

$$\hat{m}_1 = \left(\frac{1}{w}\right) \sum t_i, \quad \hat{m}_2 = \left(\frac{1}{w}\right) \sum t_i^2.$$

The method of moments gives non-optimal assessment of the meanings m and x_0 . The quantities of small samples m and x_0 , are calculated by the method of moments and they may differ from the calculated ones by the method of maximum simplicity, therefore, given a small n it is purposeful to calculate m and x_0 by the latter method and for this it is necessary to remember all the meanings t_1, \dots, t_r . For a large n it is better to use the method of moments [4].

Finally it should be mentioned that stochastic models have to reflect the main regularities of the investigated object. The degree of adequacy in the given case depends on how precisely in the models is assessed the interdependency between incoming and outgoing parameters, interactive system and environment, ability to correct by a model the decisions and finally – the application of stochastic methods in obtaining the optimal behaviour scheme of the investigated system.

5. Conclusions

1. Generally the technical-economic indices of transportation should be analysed as random factors which are

random to each meaning of the argument as well. Arguments mean the time and other parameters of the technological transportation process. Thus the optimum criteria should be considered not as determined, but as random ones.

2. Most indices according to which the technological transportation process is assessed, are interconnected, therefore it has to be born in mind that these indices are used as criteria of optimality. For the determination of digital characteristics of the technical-economical indices according to the statistical information received during the technological transportation process, the mathematical probability assessment as well as correlation and dispersion functions should be calculated.

3. For the modelling of transport networks and the flows of freight and transport means in them, it is analytically purposeful to describe the random factors by the non-parametric assessment. Differently from parametric, the non-parametric assessment of distribution density has a number of merits. Firstly, when in the course of time the distribution density type of flows changes, it is possible to approximate with precision sufficient for practical use. Secondly, this assessment is significantly less sensitive to data errors.

4. Statistical information on freight and transport flows is renewed and replenished in the course of time. With the growth of information amount the cost of its storage increases too, therefore it is necessary to obtain the required statistical assessment with the least information amount. For this purpose the mathematical models necessary for the accumulation of statistics are created and verified. The models for the following cases are created, when the density of distribution of random factors is assessed by: a) the Bayesian analysis; b) the normal distribution; c) the logarithmic-normal; d) the non-parametric distribution; e) the Weibull distribution. For all these cases the algorithms for the accumulation of minimum statistical data are created.

References

1. Candle L. M., Stewart A. The theory of distribution (Теория распределения). М.: Nauka, 1966. 584 p (in Russian).
2. Baytman P., Erdain A. The higher transcendental functions (Высшие трансцендентные функции). М.: Nauka, 1965. 343 p (in Russian).
3. Winkler R. L. The Assessment of prior Distribution in Bayesian Analysis. J. Amer. statis. assoc., 1967, Vol 62, No 319.- p 320 – 382.
4. Baublys A. Statistical models for vehicle fleet management. *Transport Engineering (Transportas)*, Vol 2(11). Vilnius: Technika, 1995, p 5 – 29.