

Forecasting of the Specialities Demand in the Sector of Mineral Resources of the Labor Market

A. N. Ilyushina

Saint-Petersburg College of Management and Commerce
St. Petersburg, Russia
bdbyu@mail.ru

P. M. Afanasyev

Admiral Makarov State University of Maritime
and Inland Shipping
St. Petersburg, Russia
paulafanasiev@gmail.com

I. M. Novozhilov¹, A. S. Voznesensky²

Saint-Petersburg Electrotechnical University "LETI"
St. Petersburg, Russia

¹novozhilovim@list.ru, ²A-Voznesensky@yandex.ru

Abstract— The article observes the issue of forecasting and creating a predictive model based on data obtained as a result of expert evaluation. Authors introduce a calculation technique and implement an experimental calculation of the demand for one factor (specialty or group of specialty) within the framework of a general request of the labor market.

Key words— *staff training; employment; graduate*

I. FIELD OF THE RESEARCH

We often face in practice with the task of forecasting random variables, whether it is a sharp drop in the demand for specialists or the opening of a new educational institution. This fact is a prerequisite for the application of probabilistic models. Probabilistic models allow us to calculate the probability that the future value of the parameter of the predicted process will be less than a certain number, for example, the probability that $y < y_i$:

$$P_i = P(y < y_i).$$

The value y can lie within the limits $y_0 \leq y \leq y_n$ because $P(y < y_0) = 0$ and $P(y < y_n) = 1$

The distribution curve of the continuous random variable y shown in the figure is a graph of the distribution function $P(y)$. The distribution function exists for both continuous and discrete random variables and is a universal characteristic of random variables.

Knowing the distribution function, we can find the probability of a random variable falling on a given section by the equation $y_0 \leq y \leq y_n$:

$$P(y_0 \leq y \leq y_n) = 1 - P_i.$$

For continuous random variables, is often considered the derivative of the distribution function of the form

$$f(y) = \frac{dP(y)}{dy},$$

or the distribution density of a continuous random variable y . The probability of a random variable y entering a certain section is shown by the equation

$$P(y_1 \leq y \leq y_2) = \int_{y_1}^{y_2} f(y) dy.$$

Thus, we can carry out prediction of the probability of an event when predicting the distribution functions considered. Moreover, in many practical cases, there is no need to characterize the random variable completely, and it is enough to predict only some distribution parameters (for example, mathematical expectation and variance).

In some cases, we can describe the data obtained during the process of observations of the predicted process by well-known distributions of continuous and discrete random variables, including normal distribution, uniform distribution, exponential distribution, Poisson distribution, and some others.

If the type and parameters of the named distributions do not change in time and a sufficient number of observations is available, then solving the forecasting problem does not cause any particular difficulties. We can construct an empirical distribution, solve the task of the theoretical distribution curve for a given empirical distribution choice, and carry out forecasting using the required accuracy. However, in practice, as a rule, the researcher has limited information about the process and, in addition, it is not always possible to guarantee the invariance of the form and distribution parameters. These conditions predetermine the use of more complex probabilistic models based on the latest achievements of probability theory. These most intensively developed areas of probability theory include, in particular, the theory of small samples and the theory of summing a random number of independent random variables.

II. FORMULATION OF THE PROBLEM

The analysis of the forecasting object and the forecast background at the retrospection period (forecast base period) established that the system development process can be represented by a stepwise process (a sequence of jumps made at random times). The magnitude y_i is a random variable which behavior is described by the distribution law $f(y)$. The number of jumps n at the forecast lead-time is random, distributed according to the law P_n . It is required to determine the distribution function of the output parameter of the system y .

III. A NUMERICAL EXAMPLE

The traditional (main) analytical apparatus of probability theory and mathematical statistics is the apparatus of characteristic functions. It is known that if Y is a real random variable, then there exists a complex random variable $X = \exp\{itY\}$ (where $i = \sqrt{-1}$ – is an imaginary unit, and t is a real number).

The function of a view

$$\phi(t) = EX = Ee^{itY},$$

Where E is the symbol of mathematical expectation, and is called the characteristic function of a random variable Y . That means that the characteristic function of a random variable Y is the mathematical expectation of a complex random variable X .

The characteristic function is dimensionless, and the parameter t has a dimension inverse to the dimension of a random variable Y .

We use the main properties of characteristic functions to solve the problem, from the conditions of the solution of which we know that the output parameter of the system y depends both on the random number of jumps n in the lead-time period and on the random value y_i of each jump. Moreover, the random variables $y_1, y_2, \dots, y_r, \dots, y_n$ are independent, equally distributed and do not depend on the random variable n .

We assume that the Poisson's law can determine the number of jumps in the forecast lead-time

$$P_n(r=n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad n = 0, 1, 2, \dots$$

With a parameter λ , the relation $\lambda = \bar{n}$ is valid for the Poisson distribution.

The random variable y (the value of the jump) has a standard normal distribution $N(0,1)$ with parameters $m_y = 0$, $\sigma^2 = 1$ and probability density

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

Thus, in order to obtain the distribution law of the output parameter, it is necessary to consider the distribution of the sum of the Poisson number of standard normal quantities.

Based on the multiplicative property of the characteristic function, the characteristic function of the sum of independent random variables is equal to the product of the characteristic functions of random variables, what means that if $y = \sum_{i=1}^n y_i$, then

$$\phi(t) = Ee^{ity} = \prod_{i=1}^n \phi_i(t),$$

We can write that the cumulative distribution function $F(y)$ of the sum of a random number n of random variables y_i is determined by the characteristic function

$$\phi(t) = \sum_{n=0(1)}^{\infty} P_n \phi_i^n(t),$$

Where $\phi_i(t)$ is the characteristic function of a random variable y_i .

Below we consider the characteristic function of a standard normal distribution:

$$\begin{aligned} \phi_i(t) &= \int_{-\infty}^{\infty} e^{ity} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-it)^2}{2} - \frac{t^2}{2}} dy = \\ &= e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-it)^2}{2}} dy. \end{aligned}$$

Since the integral looks like $\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \sqrt{2\pi}$, then

$$\phi(t) = e^{-\frac{t^2}{2}}.$$

Hence, the characteristic function of the sum of the Poisson number of standard normal quantities has the form of

$$\phi(t) = \sum_{n=0(1)}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} e^{-\frac{t^2}{2}n} = \sum_{n=0(1)}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda - \frac{t^2}{2}n}.$$

For definiteness, we exclude the case when $n = 0$ from our consideration. Then

$$\phi(t) = \sum_{n=1}^{\infty} \frac{\lambda^n}{n!(1-e^{-\lambda})} e^{-\lambda - \frac{t^2}{2}n}.$$

Based on the inversion formula

$$f(y) = \frac{dF(y)}{dy};$$

$$f(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \sum_{n=1}^{\infty} P_n \phi_i^n(t),$$

then

$$f(y) = \frac{1}{2\pi} \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!(1-e^{-\lambda})} \int_{-\infty}^{\infty} e^{-ity - \frac{t^2}{2}n} dt.$$

After the integration, we obtain the desired distribution density:

$$f(y) = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!(1-e^{-\lambda})} \frac{1}{\sqrt{2\pi n}} e^{-\frac{y^2}{2n}}$$

IV. CONCLUSION

As we have already noted, the methods of mathematical modeling are one the most general and at the same time quite strict forecasting methods. However, when we use them, we cannot predict changes in the nature of technology over time. The characteristics are defined as the result of optimization. However, we can predict the input data necessary for the functioning of the general mathematical model. We mean the prediction of exogenous variables, that is, those that are determined by dependencies, which are not a part of the main model, in contrast to endogenous variables (in particular, performance indicators of technology), which are the desired (output) variables of the main mathematical model.

We can make a prediction of input data to optimize the characteristics of the equipment, and sometimes the characteristics themselves, by one of the following methods, which differ in the scientific basis and the reliability of the forecasting results:

- compilation, according to well-known laws of nature of theoretical cause-and-effect mathematical models, which determine the predicted development processes and evaluate the parameters of these processes according to the history list and current state;
- manufacturing and testing of prototypes and experimental samples, as well as physical modeling;
- compilation of empirical dependencies according to the statistics of the history list and the present data, what is called the regression analysis and extrapolation.

Sometimes researchers also distinguish the expert assessment methods, but in most cases, experts themselves should use the above mentioned methods.

REFERENCES

[1] Kukharova T.V., Pershin I.M. Conditions of Application of Distributed Systems Synthesis Methods to Multidimensional Object. 2018

International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon 2018, 2018, October 2-4, Vladivostok, Russia. No. 8602749 DOI: 10.1109/FarEastCon.2018.8602749

[2] Kukharova T.V., Utkin V.A., Boev I.V. Observation and Prediction Systems Modeling for Human Mental State. 2018 International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon 2018, 2018, October 2-4, Vladivostok, Russia. No. 8602831 DOI: 10.1109/FarEastCon.2018.8602831

[3] Martirosyan A.V., Martirosyan, K.V., Chernyshev A.B. Application of Fourier Series in Distributed Control Systems Simulation. 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EConRus) January 28-31, St.Petersburg, Russia. 2019. Pp. 609–613. DOI: 10.1109/EConRus.2019.8656865

[4] Martirosyan A.V., Martirosyan K.V., Chernyshev A.B. Methods of distributed systems' structured modeling. 2016 IEEE NW Russia Young Researchers in Electrical and Electronic Engineering Conference February 2-3. St.Petersburg, Russia. 2016. P. 283–289. DOI: 10.1109/EConRusNW.2016.7448175

[5] Martirosyan A.V., Martirosyan K.V. Uality improvement information technology for mineral water field's control. 2016 IEEE Conference on Quality Management, Transport and Information Security, Information Technologies October 4-11. Nalchik, Russia. 2016. P. 147-151. DOI: 10.1109/ITMQIS.2016.7751925

[6] Ilyushin Y., Mokeev A. Distribution of temperature in a spatially onedimensional object as a result of the active point source. ARPN Journal of Engineering and Applied Sciences 14(6).2019, pp. 1238-1243

[7] Afanaseva O., Ilyushin Y. Analysis and synthesis of distributed icedrill heating control system of mountain reconnaissance drilling rig. International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management: 18th International Multidisciplinary Scientific Geoconference, SGEM 2018, Bulgaria, Albena 2–8 July 2018. Stef92 Technology Ltd., 2018, Volume 18, Issue 2.2, pp. 41-48. DOI: 10.5593/sgem2018/2.2/S08.006

[8] Afanaseva O., Ilyushin Y. Analysis and processing of the hydrolitospheric plast information remote sensing through the theory of systems with distributed. International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management. 18th International Multidisciplinary Scientific Geoconference, SGEM 2018, Bulgaria, Albena, 2–8 July 2018. Stef92 Technology Ltd., 2018, Volume 18, Issue 2.2, pp. 35-40. DOI: 10.5593/sgem2018/2.2/S08.005

[9] Ilyushin Y.V., Novozhilov I.M. Software implementation of a pulse regulator of a distributed distributed control object. 2017. Proceedings of 2017 IEEE 2nd International Conference on Control in Technical Systems, CTS 2017, No. 8109555, pp. 315-317. DOI: 10.1109/CTS2017.8109555

[10] Ilyushin Y.V., Novozhilov I.M. Development of a technique for the synthesis of a pulsed regulator of a distributed control system. 2017. Proceedings of 2017 IEEE 2nd International Conference on Control in Technical Systems, CTS 2017, No. 8109517, pp. 168-171. DOI: 10.1109/CTS2017.8109517