Operational process identification based on measurement results

L P Vershinina
Federal state Autonomous educational institution of higher education "Saint-Petersburg State University of Aerospace Instrumentation", Bolshaya Morskaya, 67, letter A, Saint-Petersburg, 190000, Russia

E-mail: zk-inf@yandex.ru

Abstract. Features of identification of technological process in real time are considered. It is shown that the use of standard statistical methods in conditions of insufficiency and inaccuracy of measuring information is limited. Special methods of processing the measurement results are given. An algorithm for operational evaluation and prediction of product parameter values based on measurement results, taking into account identification features, is proposed.

1. Introduction
Process identification is the most important stage creation and operation of automated technological process control system. In existing production, passive process identification methods are usually used. When organizing control in the automated control system, continuous metrological support of the technological process is performed. In this case, there is a problem of making decisions on process control in real time on the basis of measurement results in conditions of insufficient and inaccurate experimental data [1].

As is known, the purpose of measurement is to achieve the best estimate of the measured value. Among the methods for obtaining estimates, parametric methods are most widely used: the maximum likelihood method, the Bayes method, the method of conditional mathematical expectations, the method of moments, as well as special cases of these methods (the least squares method, the least modules method) [2]. It is assumed that the type of distribution function is known. Nonparametric methods do not require knowledge of the distribution function; however, they are not used enough in the tasks of identifying processes based on measurement results.

Violation of the conditions for applying a statistical method can lead to the fact that the prediction error will significantly exceed the measurement error. At the same time, there are methods of mathematical statistics that allow us to take into account the specifics of identifying the process in these conditions.

An algorithm based on such methods is proposed below.

2. Verification of statistical uniformity of measurement results
Continuous updating of the product range, change of quantitative and qualitative indicators, improvement of technological and constructive decisions are characteristic features of modern science absorbing industry. Among the most significant features of the process of manufacturing science
absorbing products we note the diversity of products, a wide variety of technological and control operations, frequent change and adjustment of process equipment [3].

Under these conditions, the researcher has at his disposal the test results of small batches of products that arrive sequentially as they are manufactured. Thus, there are two or more small samples.

Checking the uniformity of statistical data is relatively simple using nonparametric criteria. Hypotheses about the difference in means and variances are formulated as hypotheses about the position and scale parameters and are tested using special shift and scale test.

An example of such a criterion is the fast rank criterion [4]. Let \( n \) and \( m \) be the volumes of two samples. It is assumed that \( n, m \geq 4; n+m=20 \). After combining the samples, we rank the elements of the samples in ascending order.

We calculate for each sample the sums of ranks and the average ranks

\[
\bar{r}_1 = \frac{1}{n} \sum_{i=1}^{n} r_{i1}; \quad \bar{r}_2 = \frac{1}{m} \sum_{j=1}^{m} r_{j2}; \quad d = \bar{r}_1 - \bar{r}_2.
\]

When compared with a quantile of normal distribution with parameters \((0, s_d)\), where

\[
s_d = \left[ \frac{1}{2} \left( \frac{\sum r_{i1} + \sum r_{j2}}{6} \right) \left( \frac{1}{n} + \frac{1}{m} \right) \right]^{\frac{1}{2}},
\]

the shift hypothesis is rejected if

\[
\left| \frac{d}{s_d} \right| < u_{1+\alpha}.\frac{2}{n}
\]

For any type of initial distribution, the power of this criterion is not lower than 0.85.

Nonparametric rank criteria for comparing scale parameters are usually built on the basis of the corresponding shift criteria by changing the rules for assigning ranks or by changing the criteria statistics.

3. The measurement results are statistical uniform

We determine the type of distribution law and its parameters based on small samples. It is known that if \( x_1, x_2, \ldots, x_n \) is a sample from the population with the distribution function \( F(x) \), then the random variable \( y_i = F(x_i) \) is distributed uniformly on the interval \([0, 1]\). It follows that the criterion for the agreement of the sample with the distribution \( F(x) \) is to establish the uniformity of distribution \( y_i \).

For \( n > 5 \), a modified Moran criterion can be used to solve this problem [5].

The hypothesis about the uniformity of the distribution of \( y_i \) on the interval \([0, 1]\) is rejected if

\[
\frac{M_n + l/2 - b_1}{b_2} > \chi^2_\alpha(n),
\]

where

\[
M_n = \sum_{i=1}^{n+1} \ln D_i, \quad D_i = u_i - u_{i-1}, \quad u_0 = 0, \quad u_{n+1} = 1,
\]

\( u_i \) is order statistics of a random variable \( y_i \); \( l \) is the number of distribution parameters;

\( \chi^2_\alpha(n) \) is \( \alpha \)-quantile with \( n \) degrees of freedom;
If none of the standard distributions (normal, exponential, etc.) can be used as a distribution, then the contribution method can be used to estimate the density of the distribution over a small sample [6].

In the contribution method, a continuous function \( \phi_i(x_j) \), such as a beta function corresponds to each sample value of a random variable \( x_i \). In this case, the value of \( x_j \) must coincide with the mode of partial beta distribution \( \phi_i(x_j) \). The numerical experiment is used to obtain estimates of the parameters of partial beta contributions \( \phi_i(x_j) \). As a result, the desired distribution density has the form

\[
f(x_j) = \frac{1}{n} \sum_{i=1}^{n} \phi_i(x_j),
\]

where \( n \) is the sample size.

Further, using parametric methods of mathematical statistics, we construct an interval estimate of the parameter(s) based on the known distribution function [7].

4. The measurement results are not statistical uniform

If the statistics are not uniform, we identify the nature of the trend (linear, nonlinear). Further estimation and prediction of parameter values, in general, is made on the basis of trend’s extrapolation.

If additional information is available, special statistical methods are used. Here are some examples.

Example 1. Information about the properties of an empirical distribution can be very general. For example, if the distribution is symmetric with respect to the grouping center of a random variable, then you can refine the estimate of the grouping center.

For an unknown probability distribution law, the estimate for the distribution center, in accordance with the Chebyshev inequality, has the form:

\[
\bar{x} - \frac{\sigma}{\sqrt{n(1-\alpha)}} \leq M(x) \leq \bar{x} + \frac{\sigma}{\sqrt{n(1-\alpha)}},
\]

where \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \), \( \alpha \)-significance level.

For a symmetric distribution, we obtain a narrower confidence interval:

\[
\bar{x} - \frac{\sigma}{\sqrt{n(1-\alpha)}} \leq M(x) \leq \bar{x} + \frac{\sigma}{\sqrt{n(1-\alpha)}}.
\]

In a small sample, the Antill-Kersting-Zucchini test [8] and the Boos test [9] can be used to test the hypothesis of symmetric distribution.

Example 2. Let it be established that the initial distribution at \( t=0 \) is normal \( N(a_0, \sigma_0) \), and the drift of the mean in time is described by the formula

\[
a(t)=a_0+2\lambda \sigma_0 t, \quad \lambda= (\max a(t)-\min a(t))/(2\sigma_0).
\]

This situation occurs, for example, when the process parameters drift. Then, at time \( t \), the distribution will be a composition of normal and uniform distribution with density \( f(z, \lambda) \) and distribution function \( F(z, \lambda) \):

\[
f(z, \lambda)=\sigma\Phi(z\lambda)+\Phi(z\lambda)/(2\lambda),
\]

\[
F(z, \lambda)=\Phi(z\lambda)+\Phi(z\lambda)/(2\lambda),
\]
\[ F(z, \lambda) = \left\{ 1 + [(z \eta + \lambda) \Phi(z \eta + \lambda) - (z \eta - \lambda) \Phi(z \eta - \lambda) + \varphi(z \eta + \lambda) - \varphi(z \eta - \lambda)] / \lambda \right\} / 2, \]

where \( \eta = (1 + \lambda^2 / 3)^{1/2} \); \( z = (x - a_0 - \lambda \sigma_0) / (\sigma_0 \eta) \); \( \Phi(y) \) is the Laplace function; \( \varphi(y) \) is the distribution density \( N(0, 1) \).

Consider a situation that is often encountered in the automatic manufacture of parts. When the properties of the workpieces change, the scattering of the initial normal distribution changes. There is a systematic shift of the accuracy of the measurement process. Then the error distribution can be described as follows.

Let the drift of the standard deviation be described by the formula

\[ \sigma(t) = \sigma_0 (1 + 2 \gamma t) \]

where

\[ \gamma = (\text{max} \sigma(t) - \text{min} \sigma(t)) / (2 \sigma_0) \]

Then the density \( \psi(y, \gamma) \) and the distribution function \( \Psi(y, \gamma) \) will have the form:

\[ \psi(y, \gamma) = \frac{\theta}{4 \sqrt{6 \pi \gamma}} \left[ G \left( -\frac{y^2 \theta^2}{6} \right) - G \left( -\frac{y^2 \theta^2}{6(1 + 2 \gamma)} \right) \right] \]

\[
\Psi(y, \gamma) = \frac{1}{2} + \frac{\theta}{4 \sqrt{6 \pi \gamma}} \left[ G \left( -\frac{y^2 \theta^2}{6} \right) - G \left( -\frac{y^2 \theta^2}{6(1 + 2 \gamma)} \right) \right] dy,
\]

where

\[ G(y) = \int_{-\phi}^{y} e^u du, \quad \phi = \left( 3 + 6 \gamma + 4 \gamma^2 \right)^{1/2}, \quad y = \frac{x - a_0}{\sigma_0 \theta} \]

The values of the functions \( f(z, \lambda) \), \( F(z, \lambda) \), \( \psi(y, \gamma) \) and \( \Psi(y, \gamma) \) are tabulated and given in [10].

5. Conclusions

In conclusion, we note that rigid conditions for process identification make it difficult to use widely used parametric methods. At the same time, the identification of individual properties of the available experimental data provides information for the selection of special techniques and processing methods. The usefulness of using unconventional parametric and nonparametric methods of mathematical statistics for evaluating and predicting the parameters of manufactured products, taking into account the features of identification is shown.

References

[1] Vershinina L P and Vershinin M I 2020 Journal of Physics: Conf. Ser. 1515 052023
[2] Ordinarceva N P and Furman O V 2012 News of higher educational institutions Volga region Technical science 3(23) 55-61
[3] Vershinina L P and Vershinin M I 2019 IOP Conf. Series: Materials Science and Engineering 537 042014
[4] Kenuj M G 1979 Fast statistical calculations. Simplified assessment and verification methods (Moscow: Statistika) p 69
[5] Cheng R C H and Stephens M A 1989 Biometrika 76(2) 385-92
[6] Popov S A and Shapovalov V I 1982 Reliability and quality control 11 29-32
[7] Himmelblau D 1973 Analysis of processes by statistical methods (Moscow: Mir) p 958
[8] Antille A, Kersting G and Zucchini W 1982 JASA 77(379) 639-46
[9] Boos D D 1982 JASA 77(379) 647-51
[10] Borodachev N A, Abdashitov R M and Veselova I M 1973 Precision of production in mechanical engineering and instrument making (Moscow: Mashinostroenie) p 567