


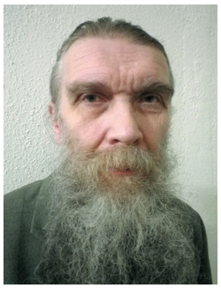

Varzhapetyan A.G., Gluschenko P.V., Gusman Y.A., Pichugin Y.A.

PROBABILISTIC ESTIMATES OF OPERATION OF DYNAMIC SYSTEMS SPECIFIED BY A LARGE AMOUNT OF PARAMETERS

This article offers an algorithm of probabilistic estimation of reliability of multi-parameter dynamic systems. Probabilistic estimates are based on multivariable distributions statistically received from the results of experimental testing. For this purpose we consider two separate cases – direct and indirect measurement of parameters. Reliability of dynamic systems is also considered in terms of validation of statistical hypotheses. The statistics used to validate hypotheses are based on the assessment of distribution parameter received experimentally.

***Keywords**: direct and indirect measurement of parameters, assessment of distribution parameters, statistical hypotheses.*

Introduction

Reliability is one of the determining properties specifying the quality of functioning of a technical system, inclusive of a dynamic system [1]. Let us note that depending on a type of input and output signals, there are continuous and discrete models of dynamic systems. This is described in more detail in paper [2]. One of the problems related to the assessment of reliability of a dynamic system is in turn the fact that the state of such system is normally specified by a large amount of parameters. On the one hand, this is stipulated by a large amount of interaction nodes within a dynamic system, and, on the other hand, by the fact that the nodes composing the system are the subsystems with a large amount of parameters specifying their state in working mode, i.e. during operation. That is why a conclusion about the reliability of a dynamic system can be made based on experimental testing of its nodes, or on experimental testing of a system as a whole. We shall assume that during experimental testing of the nodes, the parameters specifying their state can be measured directly. Let us also assume that when conducting experimental testing of a system, when indirect measurements are carried out, the task is reduced to a known model of linear regression when the specified loads act as factors, and a response is associated with a certain measured parameter specifying the system state as a whole.

Reliability assessment at the direct measurement of parameters

Let us assume that the technical parameters of a certain device or a node of a dynamic system are under testing. Technical parameters are well known to take a direct measurement more often. For these parameters it is the easiest to define the rules and permissible limits, deviations from which can mean definition of a failure or a defect of the object's functioning.

Anyway, we shall expect that with these limits we totally specify an acceptable region of the parameters ω. Making the task more specific, let us also assume that the state of a technical object (a dynamic system node) is specified by *m* parameters $\{\beta_j, j=1, 2, \ldots, m\}$, which shall be combined into a vector with respective dimensions $\{\beta = (\beta_1, \beta_2, ..., \beta_m)^{\mathrm{T}}$ ($^{\mathrm{T}}$ is a transposition sign). We shall also expect that as the result of the conducted series *n* of experimental testing we have obtained some data (the sample number *n*) of the direct measurement of parameters $\{\tilde{\beta}_i = (\tilde{\beta}_{1i}, \tilde{\beta}_{2i}, ..., \tilde{\beta}_{mi})^{\mathrm{T}}$, $i=1, 2, \ldots, n\}$.

Let us compute the following statistical estimates

$$\bar{\beta} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\beta}_i, \tag{1}$$

$$\hat{V}_{\tilde{\beta}} = \frac{1}{n-1}\sum_{i=1}^{n}(\tilde{\beta}_i - \bar{\beta})(\tilde{\beta}_i - \bar{\beta})^{\mathrm{T}}, \tag{2}$$

$$V_{\bar{\beta}} = \frac{1}{n}\hat{V}_{\tilde{\beta}}. \tag{3}$$

In formula (1) there is an estimate of mathematical expectation $E\beta = (E\beta_1, E\beta_2, ..., E\beta_m)^{\mathrm{T}}$, which coincides with mathematical expectation of the parameter measurement vector $E\tilde{\beta}$; (2) there is an estimate of the matrix $\hat{V}_{\tilde{\beta}}$ of dispersions and cross covariance of the parameter measurement vector $\tilde{\beta}$, that is

$$[\hat{V}_{\tilde{\beta}}]_{jj} = D(\tilde{\beta}_j),\ [\hat{V}_{\tilde{\beta}}]_{jk} = \mathrm{cov}(\tilde{\beta}_j, \tilde{\beta}_k)$$

(square brackets with indices denote the taking of a respective matrix element, D is a dispersion sign, cov is covariance); (3) there is a matrix of dispersions and covariance of estimate β. Therefore, we can consider two multivariable normal distributions $N(\bar{\beta}, \hat{V}_{\tilde{\beta}})$ and $N(\bar{\beta}, V_{\bar{\beta}})$, different in scope only, the scope of the second distribution is $\sqrt{n}$ times smaller.

We shall assume that the acceptable region ω is a limited open set, which means that the closure of this set ϖ is compact. Openness here is interpreted by the fact that reaching of the vector of parameters of a limiting acceptable value is exceptionally undesirable. Moreover, even the fulfillment of such condition as, for instance, "estimate $\bar{\beta}$ is an inner point of the acceptable region ω", is not a guarantee for reliable operation. Only the respective probabilistic estimates can serve as such guarantee. To form these estimates, let us consider the distribution $N(\bar{\beta}, \hat{V}_{\tilde{\beta}})$. Density function of this distribution is expressed by the formula

$$f(x, \bar{\beta}, \hat{V}_{\tilde{\beta}}) = \frac{1}{\sqrt{(2\pi)^m \det \hat{V}_{\tilde{\beta}}}} \times e^{-\frac{1}{2}(x-\bar{\beta})^m \hat{V}_{\tilde{\beta}}^{-1}(x-\bar{\beta})}, \tag{4}$$

where a vector of variables x belongs to a *m*-dimensional real space R^m.

Let us assume that the limit ∂ω of the acceptable region of the parameters ω is defined by a final set of equations of the $\varphi_k(x)=0$ type, where $\{\varphi_k, k=1, 2, \ldots, k_\omega\}$ is a respective set of continuously differentiable functions. This set is consistent with a set of Lagrange functions of the following type

$$L_k(x, \lambda) = \ln f(x, \bar{\beta}, \hat{V}_{\tilde{\beta}}) - \lambda\phi_k(x), \tag{5}$$

and these functions are in conformity with a task system

$$\begin{cases} \dfrac{\partial L_k(x,\lambda)}{\partial x_j} = 0 \\ \phi_k(x) = 0 \end{cases} \quad (k = 1, 2, ..., k_\omega,\ j = 1, 2, ..., m). \tag{6}$$

The system (6) results in a set consisting of the k_ω solutions, from which we shall choose a value x_* delivering the largest value of density function (4). In this case the value of density function is usually called likelihood. That is why the value of vector of parameters $\beta_* = x_*$ shall be called a boundary point of maximum likelihood. By this point x_* let us define a reliability ellipsoid as follows [3]

$$W = \{x : f(x, \bar{\beta}, \hat{V}_{\tilde{\beta}}) = f(x_*, \bar{\beta}, \hat{V}_{\tilde{\beta}})\}.$$

We shall clarify the construction of this ellipsoid. It is easy to see that a probabilistic measure of the region ω

$$P(x \in \omega) = \int_{\omega} f(x, \bar{\beta}, \hat{V}_{\tilde{\beta}})dx,$$

Is in fact a probability that the value of parameters' vector belongs to ω, but this probability cannot be considered as a probability that a vector will not occasionally cross the limits of ∂ω. Such a probability is evidently a probabilistic measure of reliability ellipsoid

$$P(x \in W) = \int_{W} f(x, \bar{\beta}, \hat{V}_{\tilde{\beta}})dx. \tag{7}$$

In relation to formula (5) we should note that if the density function $f(x, \bar{\beta}, \hat{V}_{\tilde{\beta}})$ of distribution $N(\bar{\beta}, \hat{V}_{\tilde{\beta}})$ is substituted by the density function $f(x, \bar{\beta}, V_{\bar{\beta}})$ of distribution $N(\bar{\beta}, V_{\bar{\beta}})$, which is deduced by the substitution of the matrix $\hat{V}_{\tilde{\beta}}$ by the matrix $V_{\bar{\beta}}$ in formula (4), then we shall get the same value x_* and, respectively, the same reliability ellipsoid *W*. However, the probabilistic measure of this ellipsoid by the distribution $N(\bar{\beta}, V_{\bar{\beta}})$ is the probability that the mean value (or mathematical expectation, the unbiased estimate of which is the mean one) of parameter vector shall not cross the limits of ∂ω, in other words we have

$$P(\bar{x} \in W) = \int_{W} f(x, \bar{\beta}, V_{\bar{\beta}})dx. \tag{8}$$

For the purpose of technical diagnostics, the assessment of change of external conditions affecting an object's dynamics is not normally performed as the defect finding is usually made in special mode in which the parameters of ambient environment are fixed. In this situation the estimate (8) is rather substantiated. But in the case when due to technical, constructive or other features, the parameters of a tested device are not absolutely stable and can change depending on the operation conditions, a probabilistic measure (8) cannot be regarded as a reliability measure, and then the probabilistic measure (7) should be used.

The aim of assessment of reliable operation can be considered in terms of validation of statistic hypotheses. We can consider the hypothesis that the mathematical expectation of the parameter vector does not belong to the acceptable region $H_{\varpi} : E\beta \notin \omega$, which can be equivalently expressed by the way that the mathematical expectation of this vector belongs to the supplement of the acceptable region in R^m, $H_{\varpi} : E\beta \in \varpi$. As it is shown in papers [4] and [5], in order to validate the hypothesis H_{ϖ}, it is sufficient to check the hypothesis $H_* : E\beta = \beta_*$. Indeed, if on a certain significance level α the hypothesis H_* is discarded, the hypothesis H_{ϖ} shall be discarded on the same significance level as well. The statistics appropriate for this validation are formed as follows. Let Q be an orthogonal matrix carrying the matrix $V_{\hat\beta}$ to diagonal form, that is

$$Q^m V_{\hat\beta} Q = \Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_m),$$

then in case of a true hypothesis $H_* : E\beta = \beta_*$ the components of the next vector

$$s = \Lambda^{-1/2} Q^m (\bar\beta - \beta_*),$$

$$\text{where } \Lambda^{-1/2} = diag\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, ..., \frac{1}{\sqrt{\lambda_m}}\right) \tag{9}$$

shall follow Student's distribution with an amount of degrees of freedom (n–1). Indeed, let us consider the vector $u = Q^m\beta$, whose every j-th component being a linear combination of normally distributed variables, i.e.

$$u_j = q_{j1}\beta_1 + q_{j2}\beta_2 + ... + q_{jm}\beta_m,$$

is also distributed by a normal probability law, and therefore the same (normal distribution but with other parameters) is also attributed to every j-th component of the vector $v = Q^m (\bar\beta - \beta_*)$, and this component is respectively equal to

$$v_j = \frac{1}{n}\sum_{i=1}^{n} u_{ji} - u_{j*} = \bar u_j - u_{j*},$$

where $u_{j*} = q_{j1}\beta_{1*} + q_{j2}\beta_{2*} + ... + q_{jm}\beta_{m*}$,

And the variables β_{k*}, k=1, 2, …, m are the components of the vector β_*, i.e. the coordinates of a boundary point of maximum likelihood. Therefore, the elements of the matrix $\Lambda = Q^m V_{\hat\beta} Q$ are respectively equal to

$$\lambda_j = \frac{1}{n(n-1)}\sum_{i=1}^{n}(u_{ji} - \bar u_j)^2 = \frac{\hat\sigma_j^2}{n},$$

where $\hat\sigma_j^2$ is an estimate of dispersion of the element u_j, and

$$s_j = \frac{\sqrt{n}(\bar u_j - u_{j*})}{\hat\sigma_u}.$$

The last equation leaves no doubts as to the above mentioned statement concerning the distribution of components of the vector s. Besides, we shall remind that as the sample number increases, Student's distribution asymptotically tends to a standard normal distribution $N(0,1)$. Consequently, the distribution of the variable

$$q = (\bar\beta - \beta_*)^m V_{\hat\beta}^{-1} (\bar\beta - \beta_*) = s_1^2 + s_2^2 + ... + s_m^2 \tag{10}$$

asymptotically tends to the chi square distribution with m of degrees of freedom. Having chosen the significance level α, we can validate the hypothesis $H_* : E\beta = \beta_*$, by checking the following inequations

$$|s_j| < t_{n-1}^{\alpha/2}, \quad j = 1, 2, ..., m, \tag{11}$$

$$q < \chi_m^2(\alpha), \tag{12}$$

where $\chi_m^2(\alpha)$ and $t_{n-1}^{\alpha/2}$ are quantiles of the chi square distribution and Student's distribution, respectively. And Student's criterion is considered to be two-sided. If even one of the inequations (11) and (12) is not satisfied, the hypothesis H_* is discarded. It gives confidence that the mathematical expectation of the vector of parameters shall not exceed the limits of the acceptable region.

As it has been indicated above, such approach to reliability assessment is valid only if the parameters of a technical device under testing are constant values, and variation of testing data occurs only due to measurement uncertainty. In case of variable parameters, when there is a dependence on operational conditions – such as rate and type of load, etc, the conclusion about the reliability of a tested technical device can be made only by checking of the hypothesis $H'_* : \beta = \beta_*$, which is interpreted in the way that the situation is possible when the parameters will reach the boundaries of the acceptable region ω. The check of such hypothesis is based not on the distribution of estimate of the mean value $N(\bar\beta, V_{\hat\beta})$, but on the distribution of a measurement data vector $N(\bar\beta, \hat V_{\hat\beta})$, the scope of which is more by $\sqrt{n}$ times.

It means that the statistics (9) or the components of the vector s decrease by $\sqrt{n}$ times, and the statistics (10), i.e. q decreases by n times. Having discarded the hypothesis H'_*, we shall discard the hypothesis $H'_{\varpi} : \beta \in \varpi$ as well, which is in this context interpreted in the way that the vector values of the parameters β may occasionally turn out to be beyond the boundaries of the acceptable region ω, i.e. at some time there may occur an emergency situation.

Reliability of a dynamic system at an indirect assessment of parameters

To assess the reliability of a dynamic system, a preliminary choice of type of test load as well as of possible restrictions for its parameters is made at the initial phase of the development (or choice) of a diagnostics method. Impossibility of direct assessment of the parameters leads to a necessity to use a model. In many cases for such a model we can use a linear model of a multiple regression type. If the parameter estimates cannot be found with a satisfactory accuracy, either the model itself, or a set of certain data should be changed. And when we speak about the change of data set we mean the increase of a number of values of the regressors which here should be interpreted as the specified modes or loads.

Let us consider the following model

$$y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_m z_{im} + \varepsilon_i \quad (i=1,\ 2,\ \ldots,\ n). \qquad (13)$$

Here y_i is a total characteristics of the system with a specified acceptable level $y_{\max}$; $\{z_{i1}, z_{i2}, \ldots, z_{im}\}$ is a set of the known loads (modes) specified at the time moments i during a testing period n; a set of parameters $\{\beta_1, \beta_2, \ldots, \beta_m\}$ specifies a reaction to the prescribed loads; ε_i is an error of the model (13) that can be interpreted as an error occurring due to measurement inaccuracy, or as the result of unaccounted loads. As a vector-matrix form, the model (13) shall be written as follows

$$Y = Z\beta + \varepsilon,$$

where Y is a vector with n dimensions of the values y_i; ε is a vector with n dimensions of model (13) inaccuracy; Z is a matrix with dimensions $n \times m$, the i-th line of which is equal to $Z_i = (1, z_{i1}, z_{i2}, \ldots, z_{im})$; β is still a vector of the parameters but with dimensions (m+1). If we assume that inaccuracies do not correlate, that they have the same dispersion and are distributed by multidimensional normal law, i.e.. $\varepsilon \sim N(0, \sigma^2 I)$ (0 is a zero column, I is an identity matrix), then the estimate by a least squares method (LS) of the parameter vector is

$$\hat{\beta} = (Z^{\mathrm{T}} Z)^{-1} Z^{\mathrm{T}} Y,$$

and the matrix of cross covariance of this estimate is as follows

$$V_{\hat{\beta}} = \hat{\sigma}^2 (Z^{\mathrm{T}} Z)^{-1},$$

where $\hat{\sigma}^2 = \dfrac{1}{n-m-1} \sum_{i=1}^{n} (y_i - Z_i \hat{\beta})^2$ is an unbiased estimate σ^2.

For a specified set of loads $\{\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_m\}$, a condition of the reliable operation of the system can be written by an inequation

$$\tilde{Z} \hat{\beta} < y_{\max},$$

where $\tilde{Z} = (1,\ \tilde{z}_1,\ \tilde{z}_2, \ldots, \tilde{z}_m)$.

Let $f(x, \hat{\beta}, V_{\hat{\beta}})$ be the density function of normal distribution $N(\hat{\beta}, V_{\hat{\beta}})$. For Lagrange function

$$L_k(x, \lambda) = f(x, \hat{\beta}, V_{\hat{\beta}}) - \lambda(\tilde{Z} \hat{\beta} - y_{\max})$$

the solution of the respective Lagrange task has quite a simple form

$$x_* = \hat{\beta} + \lambda V_{\hat{\beta}} \tilde{Z}^{\mathrm{T}},\ \lambda = (y_{\max} - \tilde{Z}\hat{\beta})(\tilde{Z} V_{\hat{\beta}} \tilde{Z}^{\mathrm{T}})^{-1}.$$

Further, by assigning $\beta_* = x_*$, we can, as in case of direct measurements, define the reliability ellipsoid

$$W = \{x : f(x, \hat{\beta}, V_{\hat{\beta}}) = f(x_*, \hat{\beta}, V_{\hat{\beta}})\},$$

as x_* is a point of tangency of the reliability ellipsoid with a hyperplane $\tilde{Z} x = y_{\max}$, which specifies a separation between unsafe and safe modes of system operation.

A probabilistic measure of this ellipsoid by the distribution $N(\hat{\beta}, V_{\hat{\beta}})$, is actually the probability of reliable (safe) operation of the system. And it is very important to perform correct specifications of the largest possible loads $\{\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_m\}$. To check the hypothesis $H'_* : \beta = \beta_*$ in case of indirect measurements we also have the statistics

$$s = \Lambda^{-1/2} Q^m (\hat{\beta} - \beta_*) \text{ and}$$

$$q = (\hat{\beta} - \beta_*)^m V_{\hat{\beta}}^{-1} (\hat{\beta} - \beta_*),$$

Where an orthogonal matrix Q, still leads to diagonal form of the matrix $V_{\hat{\beta}}$. And the components of the vector s in case of a true hypothesis $H'_* : \beta = \beta_*$ shall have Student's distribution with a number of degrees of freedom equal to $(n-m-1)$, and the number of degrees of freedom of the statistics q shall be equal to (m+1). In addition to these statistics, the variable

$$\gamma = (m+1)^{-1} (\hat{\beta} - \beta_*)^{\mathrm{T}} V_{\hat{\beta}}^{-1} (\hat{\hat{\beta}} - \beta_*) = \frac{(m+1)^{-1} (\hat{\beta} - \beta_*)^{\mathrm{T}} Z^{\mathrm{T}} Z (\hat{\beta} - \beta_*)}{(n-m-1)^{-1} \sum_{i=1}^{n} (y_i - Z_i \hat{\beta})^2}$$

in this assumptions will have exactly the F-distribution with a number of degrees of freedom $(m+1,\ n-m-1)$.

Note: If to pass to centered values in (13), it will lead to the exclusion of the parameter β_0, which does not reflect connection with the load fluctuations. In this case the number of degrees of freedom of the statistics q shall be equal to m, and the number of degrees of freedom of the statistics γ shall be equal to $(m, n-m-1)$.

Conclusion

Mathematical apparatus offered by this paper can be successfully used when performing the assessments of technical systems reliability. The obtained estimates define a lower limit of a confidence interval of reliability parameters. Accumulation of statistics shall enable to let the estimate move to the right towards real values. To reduce time, we can use the existing and widely applied software tools such as Mapple, QStat.

References

1. **Gluschenko P.V.** Technical diagnostics: Modeling in Diagnosis and Prognosis in the Status of Technical Objects./P.V. Gluschenko – 3rd issue – M.:Vuzovskaya kniga, 2013. – 248 p.

2. **Antokhina Y.A., Varzhapetyan A.G., Gluschenko P.V.**// Improvement of quality of dynamic objects functioning / SPb, Economy and management No.4, 2014, p. 49-56.

3. **Pichugin Y.A.** Probabilistic assessment of risk and reliability in economy // Scientific session SUAI: Book of reports: in 3 parts. P1. Technical sciences / SPb.: SUAI, SPb., 2013. –221 p. (P. 195-196).

4. **Pichugin Y.A.** Check of complex statistical hypothesis in linear regression analysis // Scientific session SUAI: Book of reports: in 3 parts. P1. Technical sciences/ SPb.: SUAI, SPb., 2014. – 310 p. (P. 275-277).

5. **Pichugin Y.A.** Geometric aspects of check of complex statistical hypotheses in mathematical modeling // NTV SPbSPU. Physical and mathematical sciences. No.2 (218)2015. P. 123-137.