BULETINUL Universităţii Petrol – Gaze din Ploieşti	Vol. LVIII No. 2/2006	31 - 38	Seria Matematică - Informatică - Fizică

The Statistical Estimation of the Indicators

Ionuț - Radu Bontaş, Alexandra Odăgescu

Academia de Studii Economice, Bucureşti, Catedra de Informatică Economică
e-mail: irbontas@yahoo.com

Abstract

Under the conditions where the machines and the modern installations are working, a great importance is given to the ensurance of a most efficient and complete usage of the work means on the entire service period. Therefore, if through the quality of the products we understand the rate or the level through which these correspond to the productive or individual consumption needs, then the reliability means beside these, the usage of these on the projected parameters, the safe and the continuous exploitation, in well determined conditions, over a given period of time. This paper presents estimation types, functions and statistical hypotesis regarding the random variables.

Key words: *reliability, estimation, statistical hypothesis*

Introduction

Reliability constitutes the sum of the following notions: probability, performance and mission to be accomplished, functioning and exploitation conditions, prescribed functioning time. The objects to which the reliability theory can be prescribed are *the product, the gadget, the system and the element.*

By *product* one can understand any material result of the production fated to the resolving of a certain practical problems.

The gadget is a product that represents the finished construction and it is made of objects like pieces, mechanisms, blocks, systems, devices, elements etc.

The system represents an object ansamble that function together in order to make up a function or all the functions independently.

The element represents a certain part of the system and it has the role of fulfilling a certain function inside the system. The elements inside a system can be tied-up in different ways: batch, parallel and bridge.

For the study of the reliability the element represents that part of the system that is characterised by its own reliability parameter. When the division of the elements is made of constructive criteria, the elements represent indivisible parts of the system.

It is said that a gadget is normally functioning if it fulfills its functions correctly. The properties of the gadget to correspond, in a certain moment, to all the requirements requested by the technical documentations of the working parameters for a normal functioning is called *functioning capacity*. A gadget capable to function normally, and which corresponds to the requested requirements asked by its parameters is in *good shape*.

In order for a product (gadget, system) to be "available" meaning that to be capable of fulfilling the function for which it was created in good conditions, it is necessary that, along a high reliability must be supervised, maintainance and repaired in order to keep its functioning state. This thing determines the reliability appreciation of a product to be considered and its *maintainability*, which represents the property of a product (system), expressed through the probability that this can be supervised, maintenance and repaired in a certain period of time.

The maintenance of a product (system) can become efficient only if the bench-marks of the product (system) are easy to access, spares are available and there is enough qualified work-force for the "service" operation.

The availability combines the concepts of reliability and maintainability and quantifies the capacity of an unit – which can be element, system etc. – to fulfill a service (function) requested at a certain moment.

Availability is best calculated only for the period which corresponds to the useful life of a product (system), it cannot be calculated for the initial period (youth) neither for the usage period (old age) of a product (system), periods in which the damages do not have an accidental character, and their number is huge comparing to the normal one.

In the study of the reliability of systems one can distinguish more types of reliability correspondent to the phases through which a system (product, gadget etc) is passing through, such:

Ante calculated reliability (preliminary, projected) - represents the reliability that is evaluated started from the conception of the gadget, from the dataset of its components and over the conditions of usage.

Experimental reliability (technical) – represents the reliability determined after the tests, in plant conditions, over the functioning of the gadget according to the standard regimes stated by the standards or the technical documentation.

Operational reliability (effective for the beneficiary) - represents the reliability determined in real exploitation conditions, considering internal and external factors.

We can talk even about a *nominal reliability* which represents the prescribed reliability from the specifications (internal standards, contracts etc).

In the reliability studies is necessary to determine with statistical methods the type of the repartition functions of the random variables that characterizes the elements' reliability, the estimation of the parameters of these repartitions as well as the verifying of the hypothesis regarding the random variables.

In the case in which the law of distribution is not known, the estimated values of the indicators are obtained through non parametrical methods.

The symbols used in table 1 have the following meanings:

$N=N(0)$	number of products with the same conditions on a duration of time t
$N(t)$	number of products in good shape at the moment t
q	number of attempts with duration t on which the product is placed
m	number of attempts with duration t in which the product is damaged
$d = N(0)-N(t)$	number of damaged products from duration $(0, t)$
t	chosen time duration
t_i	time of functioning of the product i (i=1,2,…,n)
N	number of attempts made over a product until it's damaged
t_j	attempt duration j (j=1,2,…,n)

For the statistical estimation of the reliability indicators experiments are organized in order to collect information regarding the characterization of the behavior of the elements and the system from the reliability point of view.

Table 1. Reliability indicators evaluation (Unknown functioning time distribution)

Indicator	Estimated	Value
	Punctual	What trust lapse ($1-\alpha$)
The repartition function of functioning time	(*) $F(t)=\left[N(0)-N(t)\right]/N(0)$ (**) $F(t)=\frac{m}{q}$	$\sum_{k=0}^{d} C_N^K F^K (1-F)^{N-K} = \alpha/2$ (*) $\sum_{k=d}^{N} C_N^K F^K (1-F)^{N-K} = \alpha/2$
The frequency of repartition of spoils	(*) $f(t)=\frac{\left[N(t)-N(t+\Delta t)\right]}{\Delta t \cdot N(0)}$	—
The reliability	(*) $R(t)=\frac{N(t)}{N(0)}$ (**) $R(t)=\frac{(q-m)}{q}$	$\sum_{k=0}^{d} C_N^K R^K (1-R)^{N-K} = \alpha/2$ $\sum_{k=0}^{N} C_N^K F^K (1-\underline{R})^{N-K} = \alpha/2$ (*) $\sum_{k=0}^{m} C_q^K \underline{R}^k (1-\underline{R})^{q-k} = \alpha/2$ $\sum_{k=m}^{q} C_q^K R^k (1-R)^{q-k} = \alpha/2$ (**)
The spoiling intensity	(*) $r(t)=\frac{[N(t)-N(t+\Delta t)]}{\Delta t * N(t)}$	—
The time of functioning average	(*) $m=\frac{\sum_{i=1}^{n} t_i}{N}$ (**) $m=\frac{\sum_{j=1}^{n} t_j}{n}$	—
The medium square deviation of the functioning time	(*) $D=\frac{\sum_{i=1}^{n}\left(t_i-m\right)^2}{N-1}$ (**) $D=\frac{\sum_{i=1}^{n}\left(t_i-m\right)^2}{n-1}$	—
The time dispersion of functioning	(*) $\sigma=\sqrt{\frac{\sum_{i=1}^{N}(t_i-m)^2}{N-1}}$	—

	(**) $\sigma = \sqrt{\dfrac{\sum_{i=1}^{N}(t_i - m)^2}{n-1}}$	
The time cuantila of functioning	t_F = the time until FN products are spoiled t_Q = the duration of tries in which the product spoils F_q times	—

The Necessity

The estimation represents the evaluation of one or more of the parameters of a population starting from a volume selection n under the condition of an initial hypothesis. If θ is a population parameter. The selection $\mathbf{X}_n$ ($x_1, x_2, \ldots, x_n$) represents observations over one random variable **X**, whose repartition function is specified.

Any function $\mathbf{t}_n$ ($x_1, x_2,\ldots, x_n$) that doesn't depend on θ but only on observations x_i, is also an random variable and it is called the statistics of the parameter θ.

The estimation of the parameter θ can be:

- *consistant estimation:* the estimation $\mathbf{t}_n$ ($x_1, x_2,\ldots, x_n$) is considered consistent if by growing the number of observations, the estimation function is becoming in probability towards the estimated parameter:
$$P_{n\to\infty}\{|\, t_n\text{-}\theta\, |> \varepsilon\,\}\to 0,\ \varepsilon > 0. \qquad (1.1)$$
- *centred estimation:* the *estimation* $\mathbf{t}_n$ ($x_1, x_2,\ldots, x_n$) is centered if an average value is the same with the estimated parameter θ:
$$[\mathbf{t}_n(x_1, x_2,\ldots, x_n)] = \theta. \qquad (1.2)$$
- *correct absolute estimation:* the estimation $\mathbf{t}_n(x_1, x_2,\ldots, x_n)$ is absolutelly correct if it is centered and its range goes to 0 by growing the number of observations:
$$M\,[\mathbf{t}_n(x_1, x_2,\ldots, x_n)] = \theta \qquad (1.3)$$
$$lim_{\ n\to\infty}\ D\,[\mathbf{t}_n(x_1, x_2,\ldots, x_n)] = 0. \qquad (1.3)'$$
- *efficient estimation:* the estimation $\mathbf{t}_n(x_1, x_2,\ldots, x_n)$ is efficient if it has minimum spread.

Methods of Estimation

The parameter estimation θ can be done through different methods, like:

The Maximum Truth-like Method

We call f (x, 0) the density function of the random variable **X** for which we have the volume selection n, $\mathbf{X}_n$ ($x_1, x_2,\ldots, x_n$). The estimation $\mathbf{t}_n$ ($x_1, x_2,\ldots, x_n$) is a maximum truth-like when $\mathbf{t}_n$ is a maximum point of the truth-like function:

$$L(x_1, x_2, ..., x_n ; \theta) = \sum_{i=1}^{n} f(x_i, \theta). \qquad (1.4)$$

The parameter θ results from the equation $\partial \ln L \big/ \partial\theta = 0$.

The Moments Method

Consists in equalising the empiric moments with the theoretical moments resulting unknown parameters:

$$\nu_k = \int_{-\infty}^{+\infty} x^k \cdot f(x,\theta)dx = \nu_k^* = \sum_{i=1}^{m} f_i \cdot x_i^k \tag{1.5}$$

$$\mu_k = \int_{-\infty}^{+\infty} (x-\nu_1)^k \cdot f(x,\theta)dx = \mu_k^* = \sum_{i=1}^{m} f_i \cdot (x_i - x)^k \ . \tag{1.6}$$

The Lining Method

Consists in the graphical representation of the empirical function for repartition through an axes coordinates system in which the repartition function represents a straight line. For this method there are used probabilistic papers.

The Smallest Squares Method

Consists in the parameter determination, under the circumstances of a minimum of the squares sum of the empirical functions deviations from the theoretical one:

$$\min [S] = \sum_{i=1}^{n} \left[F_n^*(x_i) - F(x_i,\theta)\right]^2 . \tag{1.7}$$

The parameter θ will be the solution of the equation $\partial S/\partial\theta = 0$.

This method is often used together with the lining method, in order to grow its accuracy.

The Estimation with the Help of Trust Distances

The method consists of the determination of two functions θ_i $(x_1, x_2, \ldots, x_n)$ and θ_S $(x_1, x_2, \ldots, x_n)$ such as parameter $\theta \in [\theta_i, \theta_s]$ with a given probability $1-\alpha = \gamma$.

$$P(\theta_i < \theta < \theta_s) = 1-\alpha = \gamma, \tag{1.8}$$

where:

$\gamma = 1-\alpha$ - trust level
α - meaning bridge
$[\theta_i, \theta_s]$ - trust distance

The trust distance can be previously defined bilateral one and the superior/inferior unilateral one:

$$P(\theta < \theta_s) = \alpha_2 \tag{1.8'}$$
$$P(\theta_i < \theta) = \alpha_1. \tag{1.8''}$$

If $F(x)$ and $f(x)$ are the repartition function and the density of repartition. According to the definition for the trust distance it results that θ_i and θ_s are the specifies of the repartition $F(x)$: $x_{\alpha/2}$ and $x_{1-\alpha/2}$ because only for that we have:

$$P\left(x_{\frac{\alpha}{2}} \le \theta < x_{\frac{1-\alpha}{2}}\right) = F\left(x_{\frac{1-\alpha}{2}}\right) - F\left(x_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha = \gamma \ . \tag{1.9}$$

In the table a parameter estimation of the main repartition functions is given, with the help of the bilateral trust distances.

Table 2. The estimation of the main repartition functions parameters

Repartition function		**Punctual estimation**	**Estimation with a bilateral trust distance**	
Normal	m	$\hat{m} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} x_i = x$ with the maximum truth-like method	σ^2 unknown	$x - t_{\frac{1+\gamma}{2}} \cdot (n-1) \cdot \frac{s}{\sqrt{n}}$
	σ	$\hat{\sigma}^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - x)^2 = s^2$ centred estimation	σ^2 known	$x - \Phi^{-1}_{\frac{1+\lambda}{2}} \cdot \frac{\sigma}{\sqrt{n}}$; $x + \Phi^{-1}_{\frac{1+\lambda}{2}} \cdot \frac{\sigma}{\sqrt{n}}$; $\frac{(n-1)s^2}{\lambda_{\frac{1+\gamma}{2}}(n-1)}$; $\frac{(n-1)s^2}{\lambda_{\frac{1+\gamma}{2}}(n-1)}$;
Log – normal	m	$\hat{m} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} \ln x_i = x$	As in the normal law	
	σ	$\hat{\sigma}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n}(x_i - x)^2 = s^2$		
Weibull ($\propto,\beta$)		$n\alpha - \sum_{i=1}^{n} x_i^{\beta}$; γ=0 $n/\beta + \sum_{i=1}^{n} \ln x_i - 1/\alpha \sum_{i=1}^{n} x_i^{\beta} \cdot \ln x_i = 0$	The estimation of the parameters α and β is made graphically on probabilistic paper	
Binomial		Maximum truth-like method: $\hat{q} = d/n$ d - number of damaged elements n - total number of elements	q_i and q_s are the solutions for the equations: $\sum_{x=d}^{n} C_n^x q_i^x (1-q_i)^{n-x} = \frac{\alpha}{2}$ $\sum_{x=d}^{n} C_n^x q_s^x (1-q_s)^{n-x} = \frac{\alpha}{2}$	
Poisson		Maximum truth-like method: $m = \frac{1}{n}\sum_{i=1}^{n} d_i$ n – number of experiments d_i – number of damaged elements in experimenting i	m_i and m_s are the solutions for the equations: $\sum_{x=d}^{\infty} \frac{e^{-m_i} m_i x}{x!} = \frac{\alpha}{2}; \sum_{x=0}^{d} \frac{e^{-m_s} m_s x}{x!} = \frac{\alpha}{2}$ $m_i = \frac{1}{2}\lambda^2_{\frac{1-\alpha}{2}}(2d); m_s = \frac{1}{2}\lambda^2_{\frac{1-\alpha}{2}}[2(d+1)]$	

Verifying the Statistical Hypothesis

The reliability theory uses mathematical statistic methods, mostly regarding the verifying of the proposals to be done over one or more repartitions that characterizes certain populations. Such a proposal is called statistical hypothesis, because it refers to a situation that can be true for one or more populations. The statistical hypothesis is a proposal over population and not over selection.

The statistical hypothesis can be:

- Parametrical, when refers to the parameters of a known repartition function;
- Unparametrical, when it refers to the unknown repartition function form.

In order to reject or accept a statistical hypothesis one can go to he observations, having, for example, the result (x_1, x_2, …, x_n) of the n observations. The multitude of all the possible observations is divided in two regions R_{n1} and R_{n2}. The made hypothesis is accepted if the observation's result (x_1, x_2, …, x_n) falls over R_{n1} and is rejected if it will fall aver R_{n2} named critical region. The studied hypothesis is written H_0 – null hypothesis, and the rest of the hypothesis H_1 – alternative hypothesis.

By taking the decision regarding the admission or the rejection of one hypothesis we can generate two types of errors:

- Rejecting hypothesis H_0 when it is true. This is a first grade error and has the probability:
$$\alpha = P\{(x_1, x_2, ..., x_n) R_{n_2} / H_0\}. \tag{1.10}$$
- Admitting hypothesis H_0 when it is false. This is a second grade error and has the probability:
$$\beta = P\{(x_1, x_2, ..., x_n) R_{n_1} / H_0\}. \tag{1.11}$$

Generally, the knowledge of the trust distance for a repartition's parameter allows us to construct the verification test of the hypothesis according to which it has given a certain value of this parameter.

The test λ^2. It is an unparameter test that needs a greater volume of statistical data. The test λ^2 is applied due to the following procedure:

- The line of observations (x_1, x_2, …, x_n) is breeded ordonated. The axis $(0,\infty)$ is divided in distances $[0,X_1)$, $[X_1,X_2)$, …, $[X_{N-1},\infty)$.
- We note the theoretical probability p_i in order to a value x_i to fall in the distance $[x_{i-1},x_i)$:
$$p_i = \int_{x_i}^{x_{i-1}} dF(x) = \int_{x_i}^{x_{i-1}} f(x)dx\,, \tag{1.12}$$

where $F(x)$ is the hypothetical repartition function.

- We consider m_i the number of values x_i that fall in $[x_{i-1},x_i)$ interval. When the volume of the selection n is big, the size:
$$X^2 = \sum_{i=1}^{N} \frac{(m_i - np_i)^2}{np_i}$$ has a repartition with (N-1) liberty grades.

If the parameters of $F(x)$ are statistically determined, then p_i is easily calculated and X^2 has a repartition λ^2 with (N-s-1) liberty grades (s is the number of estimated parameters of the theoretical repartition function).

The hypothesis H_0: $F(x)$ is the repartition function that reflects the analyzed selection, is accepted if:

$$X^2 \le \lambda_\alpha^2(N-1) \quad or \quad X^2 \le \lambda_\alpha^2(N-s-1). \tag{1.13}$$

The Kolmogorov test is unparametrical and it is used in the case where the empirical repartition function is given by points.

If $F(x)$ is the hypothetical repartition function of the variable x and $(x_1, x_2, \ldots, x_n)$ the right selection. Than we will statistical calculate:

$$D_n = \max\left(D_n^+, D_n^-\right), \tag{1.14}$$

where:

$$D_n^+ = \max_{1\le k\le n}\left|\frac{H_k}{n} - F(x_k)\right| \tag{1.15}$$

$$D_n^- = \max_{1\le k\le n}\left|F(x_k) - \frac{H_{k-1}}{n}\right|. \tag{1.16}$$

If $F(x)$ is continuous, Kolmogorov theory settles:

$$\lim P\left(\sqrt{n}\ D_n < y\right) = \begin{cases} 0 & for\ y \le o \\ k(y) & for\ y > 0 \end{cases}, \tag{1.17}$$

where $k(y) = \sum_{-\infty}^{+\infty} (-1)^k e^{-2k^2y^2}$.

The hypothesis H_0: $F(x)$ is a repartition function that reflects the analyzed selection, it is accepted if $D_n \le D_n(\alpha)$ where $D_n(\alpha)$ is the critical deviation for the meaningness level.

References

1. Mihoc, Gh., Muja, A., Diatcu, F. – *Bazele matematice ale teoriei fiabilității*, Ed. Dacia, 1976
2. Toma, M., Odăgescu, I. - *Metode numerice și subrutine*, Ed. Tehnică 1980
3. *** -*STAS 2876 – 74*, Erori de măsurare. Terminologie
4. *** -*STAS 10307 – 75*, Standard de fiabilitate. Fiabilitatea produselor industriale

Estimarea statistică a indicatorilor

Rezumat

În condițiile în care se lucrează cu mașini și instalații moderne, complexe, o importanță primordială o prezintă asigurarea unei utilizări cât mai eficiente și mai complete a acestor mijloace de muncă pe întreaga lor durată de serviciu. Deci, dacă prin calitatea produselor înțelegem gradul sau nivelul prin care acestea corespund necesităților de consum productiv sau individual, atunci fiabilitatea înseamnă pe lângă acestea, folosirea lor la parametrii proiectați, exploatarea lor sigură și continuă, în condiții bine determinate, pe parcursul unei durate de timp date. Acest articol prezintă tipuri de estimări, funcții și ipoteze statistice privind variabilele aleatoare.