The transition from the sample data to the total aggregate of the final volume and the analysis of this transition laws

S V Teplyakova 1, A A Kotesova 1, S I Popov 1, A A Kotesov 2

1Don State Technical University, 1, Gagarin sq., Rostov-on-Don, 344010, Russia
2Azov Sea Port, 2, st. Peter, Azov, 346780, Russia

E-mail: a.kotesova@mail.ru

Abstract. One of the most important features in science and practice for ensuring the machines and their components’ reliability is the use one of the mathematics’ areas - theory of probability and mathematical statistics. Most textbooks on the theory of probability and mathematical statistics deal with infinitely large aggregates. Since even the sample’s representativeness does not allow fully describing the entire aggregate, this should be taken into account at the stage of calculation and design. These facts led to the need for the transition from the sample data to the general aggregate data of the final volume. For this transition, we used the graphical method, which allows going from the sample, for example, with a volume of n=10-100, to the general aggregate Nc=10^3-10^6 using the extrapolation method.

Introduction
The previous research work on the experimental determination of the various machines’ reliability for the years of 1970-1985 [2, 3, 4] established an unexpected pattern of failures occurrence that appeared in the early stages of the machine operation.

In the indicated period of time, a number of machine parts were calculated from the conditions for ensuring a given gamma-percent resource [5, 6]. However, the actual resource turned out to be several tens and even hundreds of times less than the estimated one, and accordingly, the economic damage increased.

There is no answer for such a significant discrepancy.

In the future, accumulation of various information related to the machine reliability problem, which made it possible to distinguish the part resource from the sample in the amount of n=50-100 and the set Nc of the identical parts in the amount operation of thousands of units.

After that, it was necessary to analyze the results of replacing the final aggregate to the final volume of the sample data by the general aggregate, the influence of relative ranges, the discrepancy between the minimum values of the aggregates and samples, and relative errors.

Then an assumption about a significant difference, for example, of the minimum sample resources as the part and the total aggregate in operation in the amount of Nas=10^3-10^6 units was formulated.

After that, the task to study the patterns of discrepancy in the minimum resources of the units found in the sample and calculated for the general aggregate of Nas=10^3-10^6 units was formulated.

In this case, the following laws were used to describe the statistical laws: the probabilistic law with the Weibull shift — for strength and resource, and the Fisher – Tippet law — for the acting stresses [7, 8, 9].
Main part
The reliability information collection showed that failures arise not only from the machines included in the sample, but from the totality of the same type of machines as a sample.

The probability paper (grid) for the Weibull law with a three parameters shift is presented in Figure 1. As an example, a selective distribution of the parts’ resource in the interval $F(x)=0.01-0.99$, i.e. for a single base of comparing the probability distributions with different relative ranges [11].

![Figure 1. The distribution of the sample, the aggregate and their parameters](image)

The principle of constructing a direct distribution and the transition from a selective distribution to an aggregate of the final volume is presented in the form of a block diagram in Figure 2.

After drawing the points of the selective distribution (cumulative), an approximating line is constructed in the interval $F(x)=0.01-0.99$. For the correct comparison of the various sample data and the aggregates with different ranges and the values $F(x)=0.01-0.99$ are adopted.

To construct a 3-parameter Weibull distribution on a probability grid, we used the formula

$$\ln[-\ln(1-F)] = b \cdot \ln\left(\frac{x-c}{a}\right) = 2.303 \cdot b \cdot [\ln(x - c) - \ln(a)].$$

(1)

To find the sample parameters of the Weibull distribution $a$, $b$, and $c$, a graphical method was used [7].

Figure 1 shows that the probabilistic grid and the graph consists of three parts: the upper part $F(x)=0.99-0.96$ is a continuation of the middle part of the sample line with $F(x)=0.01-0.99$ and is considered the top of the final volume general aggregate distribution.

The lower part of the direct sample distribution in the interval $F(x)=0.01-0.10-6$ also refers to the total aggregate of the final volume.

To determine the shape parameter of the distribution $b$ on the probability paper [3] there is an additional scale (shown at the top of the graph). Also, it can be determined by the angular coefficient of the line (the tangent of the angle).

The distribution scale parameter $a$ can be estimated, bearing in mind that $F (x = 0) = 0.632$. If we project a line point on the abscissa axis corresponding to the value of 63.2% on the ordinate axis, we will thereby obtain the numerical value of the parameter $a$.

In case the sample is available, for example, $n = 100$, it is possible to apply these values to the probability paper of the Weibull law for the probability $F (x) = 0.01-0.99$. Since the sample is taken from the aggregate of the final volume, it should be assumed that the direct aggregate on this paper will be a continuation of the direct sampling in one and the other direction. This assumption is confirmed by the modes’ equality (the practical equality of modes when they diverge for symmetric
distributions of 0.03%, for the asymmetric distributions, the difference is 1.3%). Hence the conclusion: the parameters of the scale $a$ and the shape $b$ coincide.

![Block diagram](image)

**Figure 2.** Block diagram for graph-analytical determination algorithm of aggregates’ parameters

The continuation of the direct sampling distribution in the lower part of the graph with the interval $F(x) = 0.01-10^{-4}$ is the beginning of the total aggregate for the final volume. The first value of the aggregate’s variation series is calculated by the formula

$$X_{1as} = X_{1as} + A^S\sqrt{-\ln(1 - F(x))}.$$  

(2)
The parameters $A_{\text{as}}$ and $B_{\text{as}}$ of the aggregate are found using the recommendations of GOST 11.008-75 [5], using the probabilistic grid of the Weibull law.

It was suggested that, when applying the experimental sample points to a probabilistic grid, followed by the straight-line approximation to obtain the aggregate’s direct distribution, it is enough to extend the sample line to the first value of the aggregate’s variation series. The probability of the first (minimum) value for the variation series will be $Q=1/N_{\text{as}}$, $N_{\text{as}}$ which is the volume of the final aggregate.

The assumption made for the determination of $A_{\text{as}}$ and $B_{\text{as}}$ parameters was verified by comparing the simulated values from $A_{\text{as}}$ and $B_{\text{as}}$ parameters with the experimental sample data. As a result of the comparison, the errors were $\delta=0,06\text{-}2,02\%$. To determine the distribution parameters of the statistical (experimental) data, it is possible to use the described graphical method for applying the general aggregate of the final volume instead of the sample information in the machine parts’ fatigue life calculations [1].

A rather low percentage of discrepancy $\delta=0,06\text{-}2,02\%$ made it possible to accept the numerical values of the sample parameters $a=A_{\text{as}}$ and $b=B_{\text{as}}$. The third parameter of the $C_{\text{as}}$ shift of the Weibull distribution aggregate is obtained by extrapolating the left branch of the distribution to the intersection with the abscissa axis. In this case, the obtained shift of the distribution turns out to be less than the first (minimum) value of the aggregate’s variation series, that is, this $C_{\text{as}}$ shift falls into the no-failure zone, therefore, instead of the shift of $C_{\text{as}}$, the first value of the aggregate’s variation series 1 can be used (Figure 3) [7].

![Figure 3](image)

**Figure 3.** The units’ resource allocation: 1 - according to the sample data; 2 - according to the final volume aggregate

It is known that the sample’s dispersion is smaller than the dispersion of the aggregate, therefore, in accordance with these conclusions and Figure 3, the left branch of the aggregate is longer than the sample and the dispersion (range) is larger than that of the sample one, i.e., the failure of the aggregate starts earlier than that of the sample one.

To ensure high reliability and even absolute reliability, it is necessary to have a guarantee margin for a resource exceeding a specified resource by 15-30%, and in some cases even more. For this, the simulation was performed, that is finding the minimum (first) value of the variation series.

The coefficient of the necessary resource increase is determined by the formula:

$$C_{\text{rel.r}} = \frac{T_{\text{pre-set}}}{T_{\text{f}}}$$

(3)

where $T_{\text{f}}$ is the factual minimum resource of the serial part, requiring an increase of $T_{\text{pre-set}} = 20$ thousand hours.

**Result**
To check whether the obtained sample lines belong to the original aggregate, the parameters of the sample series were calculated by computer simulation. The samples with a volume of $n = 100$ in the number of $m = 50$ pieces with different ranges were found: $R = 1.25; 20; 50$ (Figure 4). Further, for each sample, the corresponding aggregate parameters are determined. But, since it was previously said that the parameters’ values of the sample scale and the shape are equal to the parameters of the aggregate ($a = A$, $b = B$), and instead of the aggregate shift parameter, the first value of the variation series is considered, only $X_{1c}$ was compared.

Figure 4. Distributions of the resource samples and their aggregates with a range of $R = 50$

The obtained resource values are presented in Table 1.

Table 1. The results of calculating the minimum units’ resource.

| $R$  | $TP_{an, min}$ | $TP_{an, max}$ | $R_{as}$ | $K_{rel.r.}$ |
|------|---------------|----------------|----------|--------------|
| 1.25* | 41            | 49             | 1.2      |              |
| 20    | 3100          | 6800           | 2.2      | 6.45         |
| 50    | 620           | 3500           | 5.64     | 32.26        |

* - range of steel hardness

To verify this assumption, a computer experiment by calculating the sample values from the set represented by the sample line continuation was performed [4, 5]:

$$x = c + a \cdot \sqrt[3]{-\ln(1 - F(x))}. \quad (4)$$

The obtained modeling errors for the ranges $R = 50$ are $\delta = 0.05$-$1.75\%$; for $R = 100$ $\delta = 0.06$-$2.02\%$; for $R = 200$ $\delta = 0.07$-$2.6\%$. It should be noted that the shift parameter for the aggregate of the final volume does not have physical meaning, since it turns out to be less than the smallest (first) value of the aggregate’s variation series. Therefore, the possibility of using the shift parameter of such an aggregate has a limited possibility.

Thus, it is correct to use not the shift parameter $C_{as}$, but the first value of the variation series of the aggregate $X_{1a}$. To verify such a replacement, calculations of $a$, $b$, $c$ parameters of the sample ($n = 100$, $R = 50$) were performed. Parameters $A$, $B$, $C_{as}$, and $X_{1c}$ were found graphically. After approximating the sample of resources, the parts received $a = 3.7 \times 10^3$; $b = 1.59$; $c = 2.7 \times 10^3$ hours and $A = 3.7 \times 10^5$; $B = 1.59$; $C_{as} = 59.9$ hours and $X_{1c} = 1100$ hours. A similar calculation of $x$ values was performed for the ranges $R = 100$ and $R = 200$ (for fatigue processes in detail). The results in the form of a direct aggregate in the interval $F(x) = 0.01$-$0.99$ and the simulated sample values are presented in Figure 5.
Figure 5. The modeling sample values with the ranges R=50, 100, 200 and combining them with the aggregate’s distribution line

This figure presents 1-7 determination points of the relative modeling error.

Simulation accuracy, i.e. the determination of the error between the sample points and the corresponding values on the direct distribution for the ranges R = 50 is \( \delta = 0.05\text{-}1.75\% \); for R = 100 is \( \delta = 0.06\text{-}2.02\% \); for R = 200 is \( \delta = 0.07\text{-}2.6\% \). The values of these errors indicate a fairly good agreement between the computer experiment and the original direct Weibull distribution.

To determine the failure probability, it is necessary to construct the distribution densities of the three-parameter Weibull law for the sample and the total aggregate of the final volume. The failure probability is the area under the distribution curve - for a sample of \( Q_e \) and a combination of details \( Q_{as} \) is presented in Figure 6.

Figure 6. Failure probability for the sample and the final aggregate of the three-parameter Weibull law

The probability of failure to the pre-set resource, namely, the probability of a pre-set value of \( x \) falling into a segment from \( \alpha \) to \( \beta \), is represented through the distribution density. Geometrically, the probability of getting the quantity \( x \) onto the \([\alpha, \beta]\) segment is equal to the area of the distribution curve based on this segment [6, 7, 8, 9]. Obviously, it is equal to the sum of the probability elements in its segment, i.e. the integral \( Q(\alpha < x < \beta) = \int_{\alpha}^{\beta} f(x)\,dx \).

The formula for calculating \( rQ \) discrepancies relative to the probability of the aggregate’s failure:

\[
r_Q = \frac{Q_{as}}{Q_e}
\]  

(5)
where $Q_{as}$ is the probability of the aggregate failure, $Q_s$ is the sample failure probability. The data obtained by the calculation are shown in Table 2.

Table 2. The probability of failure occurrence for aggregates’ volume $N_c=10^3-10^6$.

| $N_{as}$ | $10^3$ | $10^4$ | $10^5$ | $10^6$ |
|----------|--------|--------|--------|--------|
| $Q_{as}$ | 0,118  | 0,180  | 0,251  | 0,293  |
| $r_Q$    | 5,62   | 8,57   | 11,95  | 13,95  |

$Q_s=0,021$

It can be seen from the graph that the discrepancy $r_Q$ is 5 to 14 times for the volume aggregates $N_c=10^3-10^6$, i.e. this discrepancy in the failure probability will correspond to the number of failures and, accordingly, the costs of their elimination [10, 11, 12, 13].

The dependence of the $r_Q$ discrepancy for the minimum values of the $x_{1c}$ resource aggregates (Fig. 7) can be calculated by the formula:

$$r_{x1} = \frac{x_{1s}}{x_{1as}}$$

(6)

Figure 7. The dependence of the $r_{x1}$ discrepancy for $x_{1c}$ minimum values $x_{-1c}$ of the aggregates’ volume $N_c=10^3-10^6$ and the sample $x_{1s}$

To determine the dependence of the $r$ discrepancy on the aggregate volume $N_{as}$ by the parameters A, B, C for the aggregates $N_{as}=10^3-10^6$, differential curves of three-parameter Weibull distributions for the sample and four general aggregates of finite volume were calculated [14,15, 16, 17, 18].

Summary

The advantage of a graphical solution to the problem of replacing a sample with an aggregate and the more complete account of the number of failures is a graphical coincidence on the probability paper of the Weibull law for direct aggregate and sampling. In this case, the parameters of the shape and scale for the aggregate and the sample are numerically equal, and the first values of the sample series lie on one straight line.

Thus, the described method for determining the parameters of the sample and the aggregate of the Weibull distributions final volume makes it possible to guarantee a high or even absolute units’ reliability.

References

[1] Kotesova A A 2013 Refined definition of the aggregate resource by sampling data for a single-bucket excavator Engineering Bulletin of the Don 2 (25) 120.
[2] Kasyanov V E, Kotesova A A, Teplyakova S V 2013 Simplified determination of discrepancies between the minimum resources of samples and aggregates for critical machine parts Don Engineering Bulletin 2 (25) 119.

[3] Zaitseva M M, Kotesova A A, Kotesov A A 2011 Analysis of options for the manufacture of a single-bucket excavator (Deposited manuscript) 164.

[4] Kotesova A A 2012 Optimization of the failure-free operation probability of the single-bucket excavator boom Engineering Bulletin of the Don 20 (2) 404-406.

[5] Kasyanov V, Deryushev V, Shulkin L, Kosenko E, Kotesova A 2018 Endurance tests of single machines production MATEC Web of Conferences 02107.

[6] Kasyanov V E, Rogovenko T N, Topilin I V 2003 Determination of the minimum resource of the basic elements of machines Bulletin of mechanical engineering 3 9.

[7] Kotesova A A, Teplyakova S V, Popov S I, Kopylov F C 2019 Ensuring assigned fatigue gamma percentage of the components International Scientific Conference «Construction and Architecture: Theory and Practice of Innovative Development»: IOP Conference Series: Materials Science and Engineering 698. doi:10.1088/1757-899X/698/6/066029

[8] Ivanov V V, Popov S I, Selemeneva E M, Babazhanov N T 2019 Study of technological characteristics of the process of formation of vibration mechanochemical oxide coating XV International scientific-technical conference «Dynamics of technical systems» (DTS-2019): AIP Conference Proceedings 2188. doi.org/10.1063/1.5138389

[9] Ivanov V V, Popov S I, Marchenko Ju V, Marchenko E V, Dontsov N S, Timofeev S A 2019 Thickness of vibrational mechanic-chemical solid-lubricant coating in friction pairs of transport engineering products XII International Scientific Conference on Agricultural Machinery Industry (INTERAGROMASH 2019): IOP Conference Series: Earth and Environmental Science 403. doi:10.1088/1755-1315/403/1/012115

[10] Korotky A A, Marchenko E V, Ivanov V V, Popov S I, Marchenko Ju V, Dontsov N S 2019 Model of forming vibration mechanic-chemical solid lubrication coating on surface of steel rope XII International Scientific Conference on Agricultural Machinery Industry (INTERAGROMASH 2019): IOP Conference Series: Earth and Environmental Science 403. doi:10.1088/1755-1315/403/1/012116

[11] Ivanov V V, Popov S I, Kirichek A V 2018 Investigation of optimal chemical composition of cast aluminum alloys for vibrational mechanical-chemical polishing and deposition of protective and decorative coatings XI International Conference on Mechanical Engineering, Automation and Control Systems (MEACS 2017): IOP Conference Series: Materials Science and Engineering 327. doi:10.1088/1757-899X/327/3/032026

[12] Ivanov V V, Popov S I, Kirichek A V 2017 Qualitative Characteristics of MoS2 Solid-Lubricant Coating Formed by Vibro-Wave Impact of Free-Moving Indenters Key Engineering Materials 736, 18-22. DOI:10.4028/ www.scientific.net/ KEM.736.18

[13] Teplyakova S V 2014 Graphic-analytical method of definition parameters of Weibull distribution is offered Scientific journal «Scientific review» 11(2)

[14] Teplyakova S V, Kotesova A A 2014 To ensure the practical availability of items of construction machinery Scientific Internet – journal «Science World» 2

[15] Teplyakova S V 2016 Ensuring the reliability of machine parts during the assigned fatigue life (dissertation of the candidate of technical Sciences) 145