Estimation of statistical characteristic of aeronautical materials and structures with the empirical distribution function under random censoring

L V Agamirov¹, V L Agamirov² and V A Vestyak³

¹ Moscow Power Engineering Institute (National Research University)
Krasnokazarmennaya street, 14, Moscow, 111250, Russia
² “Volga” Ltd, Sevastopolsky Prospect, 56/40, build. 1, Moscow, 117342, Russia
³ Moscow Aviation Institute (National Research University),
Volokolamskoe highway, 4, Moscow, 125993, Russia

E-mail: v.a.vestyak@mail.ru

Abstract. The technique of estimating the reliability of aviation structures in censored samples is considered, which are formed during targeted inspections of the critical sections of the power structure of the airframe of the aircraft due to the fact that the operating time at the aircraft fleet is different and in some of the examined zones there are no failures. The technique is based on the formation of an equivalent sampling frame based on the empirical distribution function. Due to the fact that the proposed method does not have a strictly mathematical justification, statistical simulations using the Monte Carlo method was performed on the basis of the normal and Weibull distributions. The results of the calculations showed good agreement of the results for estimates of the shift and scale parameters within 2-3% of the relative errors.

1. Introduction
The reduction in tests of primary structure members and cutting commissioning time of aeronautical equipment will mostly lead to the incompleteness of statistical data. The analysis of such data has become an important task in the engineering calculation practice.

Randomly censored observations take place in the process of analysis of a critical condition of vital structural members by the moment of their maintenance inspection (for instance, in case of cracks of certain size). A similar task turns up in case of limiting time of endurance and cyclic tests of structural members.

2. Algorithm
The purpose of this article is to develop algorithms to estimate parameters of some continuous distributions based on order statistics and least squares method, which, due to its unique properties, has a number of advantages over the maximum likelihood method usually applied in similar cases. The paper features a nonparametric method of construction of empirical distribution function with multiple censoring for resolving a problem [1]. Under this method, the distribution function \( F_k \) represented by values of full and partial operating time divided into \( m \) intervals each at any point \( k \) of a
random value $T$ under random (multiple, progressive) censoring will be estimated as per the following formula:

$$
F_k = 1 - \prod_{j=1}^{k} \left[ 1 - r_j / N - \sum_{j=1}^{k-1} (r_j + n_j) \right],
$$

(1)

where $r_j$ is the number of full operating times within the interval $[T_{j-1}, T_j]$, $j = 1 \ldots m$, $n_j$ is the number of partial operating times within the interval $[T_{j-1}, T_j]$, $t_i$ denote the values of full operating times within the interval $[T_{j-1}, T_j]$, $i = 1 \ldots m$, $T_0 = 0$, $n_j$ is the number of partial operating times within the interval $[T_{j-1}, T_j]$, $i = 1 \ldots n_j$, $k$ is the number of intervals over the segment $[0, t]$, $N = \sum_{j=1}^{m} (r_j + n_j)$ is a sample volume, $r = \sum_{j=1}^{m} r_j$ is the total number of full operating times, $n = \sum_{j=1}^{m} n_j$ is the total number of partial operating times, $N_{j} = N - \sum_{j=0}^{k-1} (r_j + n_j)$ is a hypothetical volume of censoring within the interval $[T_{j-1}, T_j]$.

As an estimation of conditional probabilities in each interval we will use a ratio of the number of objects that have reached a critical condition in a particular test (full operating times) to the number of objects that have not reached a critical condition (partial operating times) at the beginning of each interval provided that there is at least one object at the beginning of such interval, which has not reached a critical condition. Otherwise, the conditional probability of a failure over this interval will be zero. The idea of the method is that within each interval after termination of observation over $n_{j-1}$ failure-free objects at the end of the interval, the remaining $N_{j}$ failure-free objects are considered as a complete sample of objects which are still being tested and have equal probability of failure over the interval $[T_{j-1}, T_j]$. In the context of future development of an algorithm for resolving the problem, the following form of representation of an empirical distribution function at point $t$ seems to be more convenient, which lead to the same results as the equation (1) does:

$$
\hat{F}(t) = \begin{cases} 
\sum_{j=1}^{k} \left[ 1 - \hat{F}(t_{j-1}) \right] \frac{r_j}{N_{j}} & \text{if } N_j > 0, \\
0 & \text{if } N_j \leq 0
\end{cases}
$$

(2)

or

$$
\hat{F}_k = \begin{cases} 
\hat{F}_{k-1} + \frac{r_k}{N_{ek}} = \sum_{j=1}^{k} \frac{r_j}{N_{j}} & \text{if } N_j > 0, \\
0 & \text{if } N_j \leq 0
\end{cases}
$$

(3)

where the value

$$
N_{ek} = N_{ek-1} - \frac{n_{k-1}}{1 - \hat{F}_{k-1}}
$$

(4)

was dubbed as equivalent volume by the authors [1].
The flow-chart below represents an algorithm of calculation of empirical distribution function in accordance with the equations (3), (4) adapted to censoring conditions typical for random maintenance inspections of aeronautical structures.

![Flow-chart](image)

**Figure 1.** Flow-chart. Calculation of empirical distribution function under multiple censoring
Below is the corresponding Javascript-function:

```javascript
//******************Cumulative Distribution Function*******************
function cum(xcum,fcum,ycum) {
    //********Sorting*******************************************
    n=xcum.length;
    for (i = 0;i<n-1;i++) {
        for (j = i+1;j<n;j++) {
            if (Math.abs(xcum[i])>Math.abs(xcum[j])) {
                xr = xcum[i];xcum[i]=xcum[j];xcum[j]=xr;
            }
        }
    }
    //************************************************************
    //fcum[n] - cumulative distribution function; ycum[j] – failures
    f0=0.5/n;Neqv=n;j=0;ni=0;
    for (i=0;i<n;i++) {
        if(xcum[i]<0) {
            ni++;
        } else {
            Neqv=Neqv-ni/(1-f0);ds=0;
            if(Neqv>0) ds=1/(Neqv+1);
            fcum[j]=f0+ds;f0=fcum[j];ycum[j]=xcum[i];j++;ni=0;
        }
    }
}
```

In the algorithm the objects that have not reached a critical condition are taken to be negative. Moreover, in contrast to the model (3), (4), the equivalent volume adds one, as it is customary to do in the theory of order statistics [2] in order not to allow empirical probabilities to equal one in further parametrical calculations.

The parameters of continuous distributions with the parameter of shift and scale were estimated by the [3]. In contrast to an approach typical for the complete sample, in this paper we propose to employ the above estimation of empirical distribution function \( \tilde{F}_j \), \( j=1...k \) as an estimation of conditional probabilities \( P=i/(N+1) \), corresponding to order statistics \( i=1...N \) by which mathematical expectations and covariations of order statistics are calculated. Thus, this approach modifies a standard estimation technique taking account of random censoring of sample.

Functions of density and normal law distribution \( y=x \) (two-parameter logarithmic-normal \( y=\ln x \) ) will take the following form: \( f(z)=\frac{1}{\sqrt{2\cdot \pi}}\cdot \exp \left(-\frac{z^2}{2}\right) \), \( F(z)=\int_{-\infty}^{z} f(t) \cdot dt \), \( z=(y-a_w)/\sigma_w \).

For a Weibull two-parameter distribution the following normalizing transformation was applied to the form with shift and scale parameters:

\[
F(x) = 1 - \exp \left(- \left(\frac{x}{c}\right)^b\right), \quad y = \ln x = a_w + z_w \cdot \sigma_w, \quad \sigma_w = 1/b, \quad a_w = \ln c, \quad z_w = \ln \left(\frac{1}{1-F(z_w)}\right), \quad F(z_w) = 1 - \exp[-\exp(z_w)], \quad f(z_w) = \exp[z_w - \exp(z_w)].
\]

Approximate interval estimations of the quantile of distribution for censored sample are defined by the following equations [5]:

\[
(5)
\]
\[ \hat{x}_{pl} \approx \hat{a} + t_{1-\beta} \left[ N-1, \Delta \right] \cdot \frac{\hat{\sigma}}{\sqrt{N}} , \]

\[ \hat{x}_{pu} \approx \hat{a} + t_{\beta} \left[ N-1, \Delta \right] \cdot \frac{\hat{\sigma}}{\sqrt{N}} , \]

where \( x_{pl}, x_{pu} \) are lower and upper confidence limits for quantile of distribution \( X \) of probability level \( P \); \( \beta \) is the level of confidence probability (normally \( \beta = 0.9 \) or \( 0.95 \)); \( t_{\gamma} \left[ f, \Delta \right] \) is quantile of level \( \gamma \) of Student non-central distribution with degree of freedom \( f = N-1 \) and non-centrality parameter \( \Delta \); \( \hat{a}, \hat{\sigma} \) are estimations of shift and scale of distribution correspondingly.

For a normal distribution the non-centrality parameter is found by formula \( \Delta = z_p \cdot \sqrt{N} \) (where \( z_p \) is quantile of level \( P \) of normalized normal distribution). For a Weibull distribution written in the form of a distribution with shift and scale parameters, as shown above, the non-centrality parameter will be found from the equation \( \Delta = \ln \ln \frac{1}{1-p} \cdot \sqrt{N} \), where \( p = F(z_w) = 1 - \exp \left[-\exp(z_w)\right] \).

As the proposed modified method of the replacement of order statistics with estimations of empirical distribution function does not have a strict mathematical justification, in order to verify the proposed model, we carried out a statistical modelling by the Monte Carlo method [4,6] with multiple (up to 2000 times) repetition of tests where every time full and censored observations were modelled, by calculating empirical distribution function (see Figure 1), estimation of parameter by least squares method and confidence probability of boundaries cover (6), (7) distribution quantiles. The preset two-sided probability \( \beta \) equaled 0.9. Level of distribution quantile \( P \) was set to be 0.05. The degree of the censoring of sample \( h \) was defined as ratio of the number of censored observations to the total number of tests \( h = (N-k) / N \). The calculations results demonstrated a good correspondence of the results for shift and scale parameters within 2-3% of relative errors. A slightly larger error of 7% was observed in estimating confidence probabilities for distribution quantiles, as a rule, downward the preset value of 0.9. Some results of modelling illustrating these data are presented in Tables 1,2 for scopes of tests of 10th, 15th and 20th degree of censoring from 0 to 0.5.

**Table 1.** Results of statistical modelling of randomly censored sample from a normal distribution with the following initial data: \( a = 3.0 \) of shift parameter (mathematical expectation); \( \sigma = 0.3 \) of scale parameter (mean square deviation); \( P = 0.05 \) of level of distribution quantile; \( \beta = 0.9 \) of confidence probability; volume of sample of \( N \); degree of sample censoring of \( h \).

| \( h \) | \( N=10 \) | \( N=15 \) | \( N=20 \) | \( N=10 \) | \( N=15 \) | \( N=20 \) | \( N=10 \) | \( N=15 \) | \( N=20 \) |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0     | 2.9468  | 2.9570  | 2.9664  | 0.2956  | 0.2972  | 0.2969  | 0.8520  | 0.8635  | 0.8635  |
| 0.1   | 2.9690  | 2.9907  | 2.9905  | 0.3018  | 0.3070  | 0.3057  | 0.8545  | 0.8840  | 0.8760  |
| 0.2   | 2.9871  | 3.0047  | 3.0157  | 0.3081  | 0.3085  | 0.3094  | 0.8710  | 0.8800  | 0.8890  |
| 0.3   | 3.0124  | 3.0196  | 3.0340  | 0.3056  | 0.3056  | 0.3058  | 0.8925  | 0.8925  | 0.8920  |
| 0.4   | 3.0265  | 3.0457  | 3.0568  | 0.3004  | 0.3009  | 0.2977  | 0.8840  | 0.8640  | 0.8485  |
| 0.5   | 3.0422  | 3.0634  | 3.0686  | 0.2975  | 0.2922  | 0.2930  | 0.8785  | 0.8465  | 0.8360  |
Table 2. Results of statistical modelling of randomly censored sample from a Weibull distribution with the following initial date: \( a_w = 5.0 \) of shift parameter; \( \sigma_w = 0.2 \) of scale parameter.

| \( h \) | \( N=10 \) | \( N=15 \) | \( N=20 \) | \( N=10 \) | \( N=15 \) | \( N=20 \) | \( N=10 \) | \( N=15 \) | \( N=20 \) |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0     | 4.9855 | 4.9887 | 4.9898 | 0.2200 | 0.2147 | 0.2113 | 0.8575 | 0.8685 | 0.8705 |
| 0.1   | 5.0042 | 5.0131 | 5.0118 | 0.2207 | 0.2161 | 0.2119 | 0.8750 | 0.8925 | 0.9010 |
| 0.2   | 5.0189 | 5.0251 | 5.0280 | 0.2193 | 0.2135 | 0.2080 | 0.8855 | 0.8965 | 0.9165 |
| 0.3   | 5.0377 | 5.0356 | 5.0439 | 0.2144 | 0.2093 | 0.2017 | 0.9105 | 0.9215 | 0.9025 |
| 0.4   | 5.0488 | 5.0524 | 5.0568 | 0.2122 | 0.2025 | 0.1993 | 0.9145 | 0.9035 | 0.9000 |
| 0.5   | 5.0613 | 5.0676 | 5.0670 | 0.2085 | 0.1953 | 0.1916 | 0.9115 | 0.8940 | 0.8735 |

The following example considers procedures of statistical analysis of randomly censored samples by the method described above for logarithmic-normal distribution and Weibull distribution.

3. Example
The Table 3 presents two samples of results of fatigue bending test of 20 air-engine low-pressure compressor blades made of titanium alloy with symmetrical cycle of alternating voltage amplitude. Minuses denote the values of logarithms of useful life of the blades which have not reached a critical condition by the moment the test stopped. The task is to estimate the parameters for the first sample based on the normal distribution of durability logarithm and calculate the approximated 90% confidence limits for a quantile of level \( P=0.01 \), while for the second sample to estimate parameters of shift and scale based on Weibull distribution and confidence limits for a quantile of level \( P=0.025 \).

Table 3. Logarithms of durability of air-air-engine compressor blades

| \( \text{lg}N_i \) | \( \text{lg}N_k \) | \( F_k \) | \( z_{r+5} \) | \( \text{lg}N_i \) | \( \text{lg}N_k \) | \( F_k \) | \( z_{w+5} \) |
|-----|-----|-----|-----|-----|-----|-----|-----|
| -4.6730 | 4.7419 | 0.0739 | 3.5524 | -4.5038 | 4.5419 | 0.0739 | 2.4327 |
| 4.7419 | 4.8215 | 0.1268 | 3.8583 | 4.5419 | 4.7055 | 0.1268 | 3.0018 |
| -4.7889 | 4.8506 | 0.1797 | 4.0835 | -4.5988 | 4.7562 | 0.1797 | 3.3810 |
| 4.8215 | 4.8704 | 0.2326 | 4.2698 | 4.7055 | 4.8911 | 0.2363 | 3.6890 |
| 4.8506 | 4.9112 | 0.2856 | 4.4336 | 4.7562 | 4.9453 | 0.2929 | 3.9402 |
| 4.8704 | 4.9253 | 0.3385 | 4.5834 | -4.7704 | 4.9628 | 0.3494 | 4.1558 |
| 4.9112 | 4.9628 | 0.3914 | 4.7243 | 4.8911 | 4.9800 | 0.4060 | 4.3478 |
| 4.9253 | 4.9800 | 0.4443 | 4.8600 | 4.9453 | 5.0899 | 0.4685 | 4.5413 |
| 4.9628 | 5.0899 | 0.5028 | 5.0071 | 4.9628 | 5.1271 | 0.5311 | 4.7220 |
| 4.9800 | 5.1271 | 0.5613 | 5.1543 | 4.9800 | 5.1523 | 0.5936 | 4.8950 |
Index $k$ marks the values of objects which have reached a critical condition, $\hat{F}_k$ denotes estimations of empirical distribution function calculated by formulas (2)-(4), $z_p$ and $z_w$ are quantiles of normalized normal and Weibull distributions corresponding to those estimations.

Parameters of normal law of distribution of durability logarithm $\hat{N}_{a}$, $\hat{N}_{\sigma}$ calculated by the least squares method equaled 5.0638 and 0.2190 correspondingly. Quantile of 0.01 degree was estimated as $\hat{x}_p = \hat{N}_a - 2.326 \cdot \hat{N}_{\sigma} = 4.5544$ with $\hat{x}_p = 4.3014$ of lower confidence limit, and $\hat{x}_u = 4.6865$ of the upper one. Similar estimations of shift and scale for Weibull distribution were: $\hat{w}_a = 5.1713$, $\hat{w}_\sigma = 0.2144$. Quantile of 0.025 degree was estimated as $\hat{x}_p = \hat{w}_a - 3.6762 \cdot \hat{w}_\sigma = 4.3830$. With $\hat{x}_p = 3.9296$ of lower confidence limit, and $\hat{x}_u = 4.5872$ of upper one. Empirical distribution functions of durability logarithm are represented in Figures 2 and 3 together with empirical values.
4. Conclusion
1) The paper has studied a technique of point and interval estimation of parameters of distributions with statistical analysis of fatigue tests of aeronautical structures based on empirical distribution function under censoring, and by the least squares method.
2) To verify the technique we carried out a statistical modelling by the Monte Carlo method based on normal law of distribution and Weibull distribution under random censoring, which offered 2-3% of relative error in estimating shift and scale parameters, and about 7% of error in estimating confidence probabilities for distribution quantiles, which allows for recommending the outcome solutions for further investigation and practical application.
3) We studied an example of statistical analysis of randomly censored samples by the proposed method for a logarithmic-normal and Weibull distributions.

References
[1] Skripnik V M, Nazin A E and Prikhodko Y G 1988 Analiz nadyozhnosti tekhicheskikh system po tsenzurirovannym vyborkam [Analysis of reliability of engineering systems by censored samples] (USSR, Moscow, Radio i svyaz)
[2] Shulenin V P 2012 Matematicheskaya statistika [Mathematical Statistics, Part II, Non-parametrical Statistics] (Publ. NTL)
[3] Kendall M G and Stuart A 1958 The Advanced Theory of Statistics: Distribution Theory 1 433
[4] Kendall M G and Stuart A 1961 The Advanced Theory of Statistics: Inference and Relationship 2 676
[5] Agamirov L V, Agamirov V L and Vestyak V A 2017 Statistical analysis of results of tests of aeronautical equipment under random censoring Programnye product i systemy 1 (30) 124
[6] Agamirov L V, Agamirov V L and Vestyak V A 2011 A calculation method of obtaining quantiles of fatigue characteristics distribution of constructive elements Vestnik MAI 18 (4) 71
[7] Balakrishnan N and Sandhu R A 1995 A simple simulation algorithm for generating progressive type-II censored samples Amer. Statist. 49 229