ISSN 8756-6990, Optoelectronics, Instrumentation and Data Processing, 2012, Vol. 48, No. 6, pp. 580–592. © Allerton Press, Inc., 2012.
Original Russian Text © N.S. Galanova, B.Yu. Lemeshko, E.V. Chimitova, 2012, published in Avtometriya, 2012, Vol. 48, No. 6, pp. 53–68.

Using Nonparametric Goodness-of-Fit Tests to Validate Accelerated Failure Time Models

N. S. Galanova, B. Yu. Lemeshko, and E. V. Chimitova

Novosibirsk State Technical University,
pr. Karla Marksa 20, Novosibirsk, 630092 Russia
E-mail: ekaterina.chimitova@gmail.com

Received March 27, 2012

Abstract—The construction of a reliability function from the results of accelerated failure time (AFT) models is considered. The constructed AFT models are verified by analyzing a sample of residuals. The fit of the residual sample to the baseline probability distribution is tested using modified nonparametric goodness-of-fit tests. In the absence of censoring in tests, it is proposed to use previously constructed models of distribution of statistics for testing composite hypotheses. In the case of censoring of type I or II, distributions of the goodness-of-fit test statistics are found by statistical modeling .

Keywords: accelerated failure time tests, AFT models, censored measurements, Kaplan–Meier estimates, nonparametric goodness-of-fit tests, modified tests.

DOI: 10.3103/S8756699012060064

INTRODUCTION

Reliability and survival problems involving lifetime type data are investigated in many areas of science and technology. In engineering calculations, such data are the failure times of some equipment or technical systems. In medicine, these are the time to a change of some biochemical indices, the time to remission after a certain type of treatment or the lifetime of patients, etc.

Typically, a common property of the data (measurement results, observations of test objects) analyzed in such problems is their incompleteness. For example, during tests, only some of the test objects may fail or some objects may be not tested for some reason. The test time may be limited, and a significant part of the objects may remain operational by the end of the experiment. Such data are called censored on the right. If experiments are limited in time, a censored sample of type I is obtained. Type II censoring occurs if tests are continued until the occurrence of a predetermined number of failures (measurement results).

Lifetime data have been analyzed in many publications, many methods of description and representation have been developed, and various models taking into account their characteristics for different applications have been proposed. Early papers [1–10] mostly considered simple models with different parametrizations of the reliability function. Among these models are, in particular, reliability functions based on the exponential probability distribution, Weibull distribution, gamma distribution, inverse Gaussian distribution, and others.

Accelerated failure time (AFT) models hold a special place in reliability theory; recently, they have attracted increased interest due to the growth in the production of highly reliable and highly technological products and systems. The design and production of such devices are geared to provide their trouble-free operation over a long period of time. During the time that can be allotted for reliability tests and studies under normal operating conditions, the probability of failure of a device is very low. In order to obtain sufficient data for analysis in such a situation, tests are carried out at loads exceeding the load calculated for normal conditions. Such tests are referred to as accelerated tests. The use of increased loads shortens the lifetime of the systems, and failures will occur within the time allotted for collection of experimental data.

AFT models are designed to evaluate the reliability functions of articles (systems) operating under normal operating conditions (under normal loading) from failure data obtained in accelerated tests. The construction of AFT models for various designs of experiments and analytical calculations for their parametrizations are discussed in detail in [5–8].

Although the construction of parametric AFT models has been the subject of numerous (mostly foreign) publications, many problems remain unsolved. First, the quality of AFT models (estimates of parameters) are significantly affected by the presence of censored observations. Second, there are a number of problems related to the validation of the constructed AFT models.

For example, maximum likelihood estimates (MLE) of distribution parameters from type I or II censored samples are asymptotically efficient, i. e., they asymptotically obey the normal distribution. However, in the case of limited sample sizes and in the presence of a significant portion of censored observations of type I or II, the distributions of the parameter estimates of the observed probability distributions are asymmetric and the estimates are biased [11, 12]. Naturally, this affects the quality of the distribution models.

In addition, censoring influences the procedures of validating the models. For example, in tests of composite hypotheses that complete samples fits some theoretical law and in estimation of the law parameters from the same sample, the distributions of nonparametric goodness-of-fit test statistics depend on the following factors: the law used to test the goodness-of-fit, the number and type of the estimated parameters, and the value of the shape parameter [13–24]. In the case of censored samples, these factors also include the dependence of the distributions of the statistics on the degree of censoring [25].

In the construction and analysis of AFT models, the above-mentioned problems related to the presence of censored data persist and even increase.

The most common method of testing the fit of a parametric AFT model to observation (measurement) results is to analyze the distributions of so-called residuals. If the hypothesis that the sample of residuals fits the baseline distributions is not rejected, this is evidence of the validity of the constructed AFT model. However, in publications devoted to accelerated failure test statistics, the validation of constructed AFT models is often not mentioned [1–3, 5–8] or the goodness-of-fit of residual samples to the baseline distributions is tested using graphical methods [9]. The reasons for this lie in the problems associated with the use of goodness-of-fit criteria for testing composite hypotheses [13–24], aggravated by the presence of censored observations in reliability studies, which affects the properties of estimates of model parameters [11, 12] and, in turn, is also reflected in distributions of the statistics of the goodness-of-fit tests used [25]. The need to validate AFT models exists, and this problem has been extensively studied, but at the moment we can project that it cannot be solved using a purely analytical approach.

The aim of this work is to investigate issues of validation of parametric AFT models through an analysis of residual samples.

AFT MODEL

Let T be a nonnegative random variable which determines the time of operation of the test object to failure. The reliability function is the probability of failure-free operation for some time t:

$$S(t) = P\{T > t\} = 1 - F(t), \tag{1}$$

where $F(t)$ is the corresponding probability distribution function.

The reliability of an object depends on some of its characteristics and operation conditions. The influence of these characteristics is taken into account by means of explanatory variables (covariates), which in reliability theory are commonly referred to as stresses or loads.

A stress $x(\cdot)$ is elevated with respect to a stress $y(\cdot)$ if the corresponding reliability functions satisfy the relation

$$S_{x(\cdot)}(t) \leq S_{y(\cdot)}(t), \quad t \geq 0. \tag{2}$$

The stress $x(\cdot)$ can be any operational parameter: voltage, pressure, humidity, temperature, etc.

The purpose of accelerated failure tests is to obtain data at elevated loads, which are then uses to evaluate the reliability function corresponding to the normal operation conditions.

For example, we consider the following possible test plans:

1. The observed objects are divided into l groups and are tested under stresses constant in time $\mathbf{x}_1, \ldots, \mathbf{x}_l$, i. e. n_i objects of the ith group are tested under a stress $\mathbf{x}_i$, $i = \overline{1, l}$.

2. All objects are tested under a step stress $x(t)$:

$$x(t) = \begin{cases} \mathbf{x}_1, & t_0 < t \le t_1, \\ \mathbf{x}_2, & t_1 < t \le t_2, \\ \dots & \\ \mathbf{x}_l, & t_{l-1} < t \le t_l. \end{cases} \tag{3}$$

The test plan can be a combination of these plans.

The sample of observations resulting from the tests is usually written in the form

$$\mathbf{X}_n = \{(X_1, \delta_1, \mathbf{x}^1), \dots, (X_n, \delta_n, \mathbf{x}^n)\}, \tag{4}$$

where $\delta_i = 1$ if X_i is the moment of failure (complete observation) and $\delta_i = 0$ if X_i is the moment of censoring; $\mathbf{x}^i$ is the covariate value at which the observation X_i was obtained.

A parametric AFT model of reliability can be defined as

$$S_{x(\cdot)}(t) = S_0\Big(\int_0^t r(x(s))ds\Big), \tag{5}$$

where $S_0(t) = 1 - F_0(t)$ is the baseline reliability function; $r(\cdot)$ is a non-negative function of stresses. In applications, the most frequently used model functions of stresses are as follows:

— the log-linear model $r(x) = \mathrm{e}^{\beta_0 + \beta_1 x}$ is used, for example, to analyze fatigue data in testing different electronic components;

— the power rule model $r(x) = \mathrm{e}^{\beta_0 + \beta_1 \ln x}$ is used in cases where the stress is voltage or mechanical loads;

— the Arrhenius model $r(x) = \mathrm{e}^{\beta_0 + \beta_1 / x}$ is used when the stress is, for example, temperature;

— the model $r(\mathbf{x}) = \mathrm{e}^{\beta_0 + \beta_1 x_1 + \dots + \beta_m x_m}$ is used in the case of multi-dimensional stresses (m is the dimension of the stress vector).

In the case of a step stress where the test plan is given by relation (3), the reliability function of the object observed under a stress $\mathbf{x}_i$ is given by [5]

$$S_{\mathbf{x}_i}(t) = S_0\Big(r(\mathbf{x}_i, \boldsymbol{\beta})(t - t_i) + \sum_{j=1}^{i} r(\mathbf{x}_j, \boldsymbol{\beta})(t_j - t_{j-1})\Big). \tag{6}$$

In parametric AFT models, it is assumed that the baseline reliability function $S_0(t) = 1 - F_0(t)$ is defined by some parametric family of distributions $F_0(t; \boldsymbol{\theta})$. As the baseline laws the following distributions are frequently accepted:

$$f_0(t, \boldsymbol{\theta}) = \frac{1}{\theta_0} \exp\Big(-\frac{t}{\theta_0}\Big)$$

— the exponential distribution;

$$f_0(t, \boldsymbol{\theta}) = \frac{1}{t\theta_1\sqrt{2\pi}} \exp\Big(-\frac{1}{2}\Big(\frac{1}{\theta_1}\ln\Big(\frac{t}{\theta_0}\Big)\Big)^2\Big)$$

— the lognormal distribution;

$$f_0(t, \boldsymbol{\theta}) = \frac{\theta_1}{t}\Big(\frac{t}{\theta_0}\Big)^{\theta_1} \exp\Big(-\Big(\frac{t}{\theta_0}\Big)^{\theta_1}\Big)$$

— the Weibull distribution;

$$f_0(t, \boldsymbol{\theta}) = \frac{\theta_1}{t\theta_0}\Big(\frac{t}{\theta_0}\Big)^{\theta_1}\Big(1 + \Big(\frac{t}{\theta_0}\Big)^{\theta_1}\Big)^{1/\theta_2 - 1} \exp\Big(1 - \Big(1 + \Big(\frac{t}{\theta_0}\Big)^{\theta_1}\Big)^{1/\theta_2}\Big)$$

— the generalized Weibull distribution;

$$f_0(t,\boldsymbol{\theta}) = \frac{1}{\theta_0 \Gamma(\theta_1)} \Big(\frac{t}{\theta_0}\Big)^{\theta_1 - 1} \exp\Big(-\frac{t}{\theta_0}\Big)$$

— the gamma distribution, and others.

Maximum likelihood estimates of the unknown parameters of an AFT model from a censored sample $\mathbf{X}_n = \{(X_1, \delta_1, \mathbf{x}^1), \ldots, (X_n, \delta_n, \mathbf{x}^n)\}$ can be obtained by maximizing the log-likelihood function

$$\ln L(\mathbf{X}_n; \boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^{n} (\delta_i \ln f_{x(\cdot)}(X_i) + (1-\delta_i) S_{x(\cdot)}(X_i)) \tag{7}$$

for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

In expression (7), the density $f(t)$ is obtained from the baseline density $f_0(t,\boldsymbol{\theta})$ in which the scale parameter θ_0 is replaced by a quantity inversely proportional to the function of the stress $1/r(x)$.

VALIDATION OF PARAMETRIC AFT MODELS

The validity of constructed parametric AFT models is usually tested by analyzing a sample of residuals which are calculated as follows:

$$Z_i = X_i r(\mathbf{x}^i, \hat{\boldsymbol{\beta}}), \quad i = \overline{1,n}, \tag{8}$$

for the case of time-independent stresses and

$$Z_i = r(\mathbf{x}_q^i, \hat{\boldsymbol{\beta}})(X_i - t_{i-1}) + \sum_{j=1}^{q} r(\mathbf{x}_j^i, \hat{\boldsymbol{\beta}})(t_j - t_{j-1}), \quad i = \overline{1,n}, \tag{9}$$

for a step stress of the form (3), where $\mathbf{x}_q^i$ is the stress value at which the observation X_i was recorded. If the data are well described by the constructed model, the residuals should fit the baseline distribution of failures $F_0(t;\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the obtained estimate of the parameter vector standardized by the scale parameter (at a scale parameter $\theta_0 = 1$).

The hypothesis that the residual sample fits the distribution $F_0(t;\hat{\boldsymbol{\theta}})$ can be verified using different goodness-of-fit tests, whose use in this situation involves the following problems. First, measurements are usually censored, which affects the properties of estimates [11, 12] and the possibility of using specific goodness-of-fit tests [25]. Second, the tested hypothesis is composite since the parameters of the model were already evaluated from the test results and the same results are now to be used to test the hypothesis.

In tests of composite hypotheses, where unknown parameters are estimated from the same sample, nonparametric goodness-of-fit criteria are no longer independent of the distribution. In this case, the distributions $G(S \mid H_0)$ of the statistics S of nonparametric goodness-of-fit criteria for testing the hypothesis H_0 depend on several factors:

— the form of the baseline failure distribution $F_0(t;\boldsymbol{\theta})$; — the type and number of the parameters estimated from the sample; — the method of estimating the parameters; — and possibly, the value of the shape parameter (for example, the shape parameter of the gamma distribution [17] or the generalized Weibull distribution [20]).

Model distributions of the nonparametric Kolmogorov, Cramer–von Mises–Smirnov, and Anderson–Darling goodness-of-fit statistics for testing composite hypotheses using MLE for different combinations of parameters in the case of complete samples (non-censored observations) are presented in [17–24].

The hypothesis that a residual sample fits a distribution $F_0(t;\hat{\boldsymbol{\theta}})$ in the presence of censored measurements can be tested using modifications of Kolmogorov, Cramer–von Mises–Smirnov, and Anderson–Darling goodness-of-fit tests in which the statistics are expressed using the nonparametric Kaplan–Meier estimate, and not an empirical distribution function [26].

We denote by $a_1 < a_2 < \ldots < a_k = \tau$, $k \le n$ the values of the complete observations $(Z_i, \delta_i = 1)$ in the sample of residuals $(Z_1, \delta_1), (Z_2, \delta_2), \ldots, (Z_n, \delta_n)$. Then, the Kaplan–Meier estimate can be calculated by the formula

$$\hat{F}_n(t) = 1 - \prod_{a_i \le t} (1 - d_i/c_i), \tag{10}$$

where $d_i = \sum\limits_{Z_j = a_i} \delta_j$; c_i is the number observations for which $Z_j \ge a_i$, $j = 1, \ldots, n$.

In the modified Kolmogorov test for censored samples, the distance between the empirical and theoretical distributions is given by the quantity

$$D_n = \sup_{0 \le t \le \tau} \left| \hat{F}_n(t) - F_0(t; \boldsymbol{\theta}) \right|,$$

where $\hat{F}_n(t)$ is the Kaplan–Meier estimate of the distribution function and $F_0(t; \boldsymbol{\theta})$ is the theoretical distribution function corresponding to the tested hypothesis.

In order to reduce the dependence of the distribution of the statistics on the sample size for small n in the modified Kolmogorov test, it is appropriate to use the statistic with Bol'shev's correction [27]:

$$S_K^C = \frac{6nD_n + 1}{6\sqrt{n}}, \tag{11}$$

where $D_n = \max\{D_n^+, D_n^-\}$; $D_n^+ = \max_i\{\hat{F}_n(a_i) - F_0(a_i; \boldsymbol{\theta})\}$; $D_n^- = \max_i\{F_0(a_i; \boldsymbol{\theta}) - \hat{F}_n(a_{i-1})\}$.

In the modified Cramer–von Mises–Smirnov test, the distance between the distributions is given by

$$\omega^2 = \int_0^\tau (\hat{F}_n(t) - F_0(t; \boldsymbol{\theta}))^2 dF_0(t; \boldsymbol{\theta}).$$

The statistic of this test with the Kaplan–Meier estimate becomes

$$S_\omega^C = \frac{n}{3} F_0(a_1; \boldsymbol{\theta}) + n \sum_{j=1}^{k-1} \Big[\hat{F}_n^2(a_j)(F_0(a_{j+1}; \boldsymbol{\theta}) - F_0(a_j; \boldsymbol{\theta})) -$$

$$- \hat{F}_n(a_j)(F_0^2(a_{j+1}; \boldsymbol{\theta}) - F_0^2(a_j; \boldsymbol{\theta})) + \frac{1}{3}(F_0^3(a_{j+1}; \boldsymbol{\theta}) - F_0^3(a_j; \boldsymbol{\theta})) \Big]. \tag{12}$$

In the modified Anderson–Darling test, the measure of the distance is taken as

$$\Omega^2 = \int_0^\tau (\hat{F}_n(t) - F_0(t; \boldsymbol{\theta}))^2 \frac{dF_0(t; \boldsymbol{\theta})}{F_0(t; \boldsymbol{\theta})(1 - F_0(t; \boldsymbol{\theta}))},$$

and, accordingly, the test statistic has the form

$$S_\Omega^C = n \Big\{ - F_0(a_1; \boldsymbol{\theta}) + \sum_{j=1}^{k-1} \Big[F_0(a_j; \boldsymbol{\theta}) - F_0(a_{j+1}; \boldsymbol{\theta}) + \hat{F}_n^2(a_j)(\ln F_0(a_{j+1}; \boldsymbol{\theta}) - \ln F_0(a_j; \boldsymbol{\theta})) -$$

$$- (1 - \hat{F}_n(a_j))^2 (\ln(1 - F_0(a_{j+1}; \boldsymbol{\theta})) - \ln(1 - F_0(a_j; \boldsymbol{\theta}))) \Big] - \ln(1 - F_0(a_1; \boldsymbol{\theta})) \Big\}. \tag{13}$$

In order to apply the goodness-of-test tests, it is necessary, with the validity of the tested hypothesis, to know the distributions of the statistics of these tests that correspond to the plan and results of the test, the baseline distribution, and the type and degree of censoring. Distributions of the test statistics or percentage points required to test the hypothesis can be obtained using the approach of [17–24] based on computer technology and statistical modeling. It is clear that the presence of the above factors influencing the distribution of the statistics of the above-mentioned tests does not allow one to construct distributions of statistics that correspond to specific test conditions. The actual test results and the values of the factors affecting the distributions $G(S \mid H_0)$ of the goodness-of-fit test statistics S, with the validity of the tested hypothesis H_0, become known only after the construction of an AFT model. After that, computer simulation is used to construct an empirical distribution of the statistic (or even an approximate distribution model), from which it is possible to calculate the attained level of significance and decide on the validity of the model. This interactive algorithm for simulating distributions of goodness-of-fit test statistics provides a correct validation of the constructed model.

INTERACTIVE ALGORITHM FOR SIMULATING TEST STATISTIC DISTRIBUTIONS

The procedure for modeling (studying) the statistic distribution $G(S \mid H_0)$ of a test used is as follows.

1. Simulate a complete sample of failures $(X_1, \delta_1 = 1, \mathbf{x}^1), \ldots, (X_n, \delta_1 = 1, \mathbf{x}^n)$ in accordance with the tested model $F_x(t; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ is the MLE of the model parameters from the original sample. The failure time of the object under a constant stress $\mathbf{x}_i$ is modeled according to the equation

$$X = F_0^{-1}(y, \hat{\boldsymbol{\theta}})/r(\mathbf{x}_i, \hat{\boldsymbol{\beta}}), \tag{14}$$

where y is a pseudo-random variable uniformly distributed on the interval [0, 1].

The time of failure of an object under a step stress (3) is modeled as follows. The stress $\mathbf{x}_i$ is determined from the condition that the value of y falls into the interval $(b_{i-1}, b_i]$, $i = \overline{1, l}$, where $b_0 = 0$, $b_i = F_0\Big(\sum\limits_{j=0}^{i} r(\mathbf{x}_j, \hat{\boldsymbol{\beta}})(t_{j+1} - t_j)\Big)$, $b_l = 1$. If $i = 1$, the failure time is calculated in accordance with (14). Otherwise,

$$X = t_{i-1} + \Big[F_0^{-1}(y, \hat{\boldsymbol{\theta}}) - \sum_{j=1}^{i} r(\mathbf{x}_{j-1}, \hat{\boldsymbol{\beta}})(t_j - t_{j-1})\Big]\Big/ r(\mathbf{x}_i, \hat{\boldsymbol{\beta}}). \tag{15}$$

2. Convert (if necessary) the complete sample into a censored sample in accordance with the specified censoring scheme: $(X_1, \delta_1, \mathbf{x}^1), \ldots, (X_n, \delta_n, \mathbf{x}^n)$.
3. Determine the MLE of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ from the sample $(X_1, \delta_1, \mathbf{x}^1), \ldots, (X_n, \delta_n, \mathbf{x}^n)$.
4. Calculate the residuals for the tested model (using (8) or (9)).
5. Determine the test statistic value from the sample of residuals (using (11), (12) or (13)).
6. Repeat steps 1–5 N times to obtain the empirical distribution of the statistic $G_N(S_n \mid H_0)$.

The choice of N depends on the desired accuracy of modeling (for more details of the choice of N see [24]).

Based on the study of the problems associated with the use of the modified goodness-of-fit tests for the analysis of residual samples, an interactive algorithm for validating parametric AFT models was proposed, whose block diagram is shown in Fig. 1.

The proposed algorithms for modeling distributions of goodness-of fit test statistics for parametric AFT models based on samples of residuals were implemented in LiTiS software system of statistical survival analysis [28]. This system allows one to calculate MLEs of AFT model parameters for a wide range of baseline distributions and different stress functions, to correctly validate the constructed AFT models using modified goodness-of-fit tests, and to model the distributions of estimates and goodness-of-fit test statistics for the given AFT models and test plans.

DISTRIBUTIONS OF THE TEST STATISTICS IN THE CASE OF TIME-INDEPENDENT STRESSES

The present studies have shown, as expected, that in the case of complete data, where the relations for statistics (11)–(13) coincide with the classical ones, the distributions $G(S_n \mid H_0)$ of the tests considered are independent of the form of the function at stress $r(\cdot)$. Thus, to test the hypothesis that the sample of residuals fits a distribution $F_0(t; \hat{\boldsymbol{\theta}})$ (to validate parametric AFT models) using the modified goodness-of-fit tests, one can use the limiting distributions of statistics $G(S_n \mid H_0)$ presented in [17–23], and more fully in [24]. Results of statistical simulations have confirmed that even with small samples of residuals, the distributions of nonparametric test statistics coincide with approximations of distributions of these statistics obtained for models without covariates [17–24].

In the case of censored data, the distributions of statistics depend on both the degree and type of censoring. An important factor is the censoring scheme. For example, time limitation (type I censoring) or a limitation of the number of failures (type II censoring) can be specified identically for all n objects, or differently for each group of objects. In the latter case, the sample of residuals is a repeatedly censored sample.

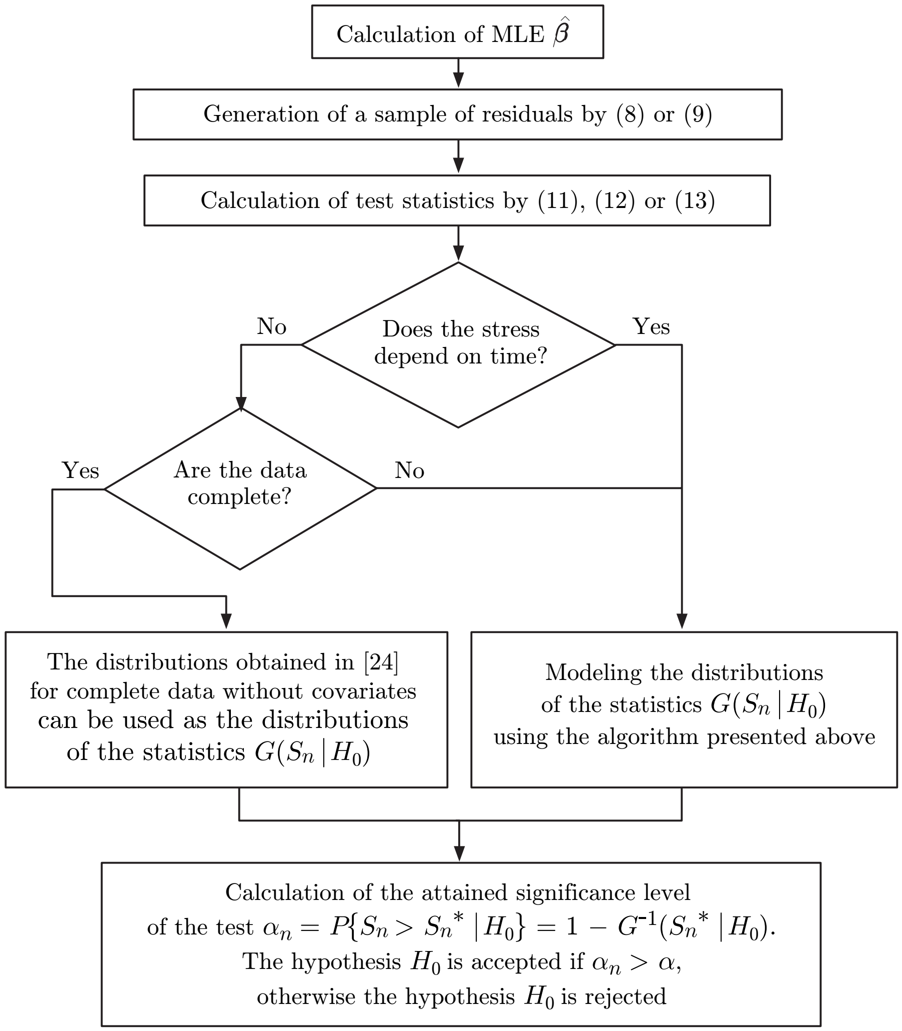

Fig. 1. Block diagram of the algorithm for testing the goodness-of-fit hypothesis for a parametric AFT model based on samples of residuals.

DISTRIBUTIONS OF TEST STATISTICS IN THE CASE OF STEP-STRESS LOADING

If objects were observed under step stress (3), distributions of statistics obtained on the basis of the samples of residuals (9) can differ significantly from distributions of statistics at stresses constant in time. As an example, Fig. 2 presents empirical distributions of Kolmogorov's statistic in testing the composite goodness-of-fit hypothesis for the Weibull parametric AFT model for the following two test plans.

1. Constant-stress test plan: all objects are divided into two groups with stresses $x_1 = 1$ and $x_2 = 2$ with $n_1 = 20$ and $n_2 = 20$ observations in each group.

2. Step-stress test plan: all $n = 40$ observations were modeled under a stress

$$x(t) = \begin{cases} x_1 = 1, & 0 < t \le 1, \\ x_2 = 2, & t > 1. \end{cases}$$

The log-linear model of the stress function $r(x) = e^{\beta_0 + \beta_1 x}$ was considered. The true values of the AFT model parameters are as follows: $\beta_0 = -1$, $\beta_1 = 1$, and the baseline distribution shape parameter $\theta_1 = 1$. The parameters β_0, β_1, and θ_1 were estimated using the maximum likelihood method. The number of simulated samples was $N = 5000$. For comparison, the figure shows the approximation obtained in [18] for the limiting distribution of Kolmogorov's statistic in testing the composite goodness-of-fit hypothesis for a

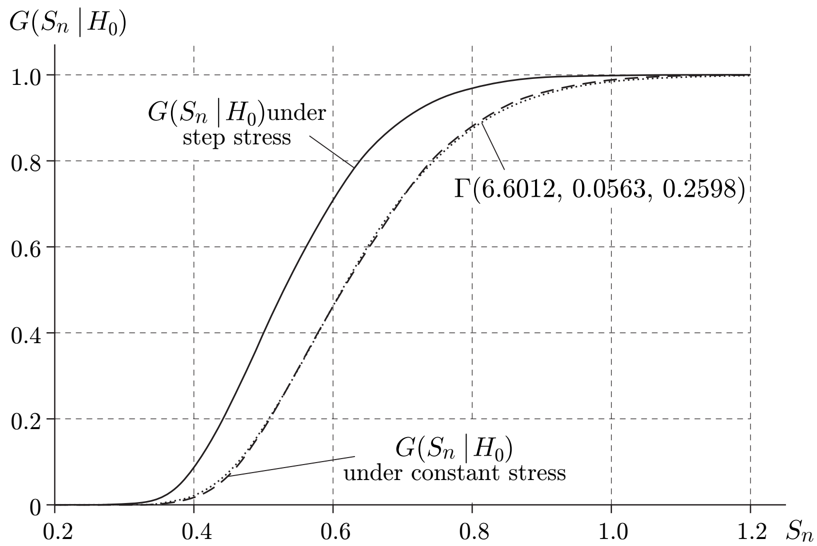

Fig. 2. Distribution of Kolmogorov's statistic for different test plans.

Weibull distribution for data without covariates (a gamma distribution with a bias parameter of 0.2598, a scale parameter of 0.0563, and a shape parameter of 6.6012).

As seen in Fig. 2, even in the case of complete data, the distribution of the goodness-of-fit test statistics for time-dependent covariates is significantly different from the approximation of the limiting distribution of the statistic. Thus, unlike in the case of time-constant stresses, in tests of parametric AFT models with time-dependent covariates, it is incorrect to use the test statistic distribution models constructed in [24] to calculate the attained significance level. Here distributions of statistics can be obtained by statistical modeling using the algorithm presented above.

CHOICE OF THE BASELINE RELIABILITY FUNCTION

In the construction of parametric models of reliability functions, including models with covariates and AFT models, one of the key problems is the choice of the baseline distribution or a family of distributions. In practice, as a rule, a priori information is not sufficient for an unambiguous choice of the baseline distribution or a family of distributions.

As a measure of the preference of a particular baseline probability distribution $F_0(x;\boldsymbol{\theta})$, it is possible to use the attained significance levels obtained in test of the hypotheses that the sample of residuals fits the baseline distribution $F_0(x;\hat{\boldsymbol{\theta}})$ using goodness-of-fit tests.

Example 1. We illustrate this with the example of choosing the baseline distribution function for the problem of constructing the reliability function of electrical insulating liquids for which the results of accelerated failure tests are presented in [2]. In the tests, all objects were divided into seven groups. Within each group, the objects were observed under a constant high voltage of 26 kV to 38. The purpose of the accelerated tests was to evaluate the reliability function of the insulating liquids under a normal voltage of 20 kV. Note that all these tests ended in failure of the test object; therefore, the failure data are complete (no censored observations) (Table 1). It should be noted that the time from the beginning of a test to failure significantly reduces with increasing voltage; n_i indicates the number of tests carried out under the corresponding load.

The stress function is selected in the same form as in [2]: $r(x) = e^{\beta_0 + \beta_1 \ln x}$. The following possible baseline distributions were considered : the exponential, Weibull, generalized Weibull, and gamma distributions.

Since the data are complete (censored measurements are absent), the models of [17–24] can be taken as the limiting distributions of the statistics of the modified goodness-of-fit tests .

Table 2 presents the obtained estimates of the parameters of the models, a 95% confidence interval for the estimates of the model parameters, and values of the log-likelihood function.

Table 3 shows the calculated values of the goodness-of-fit test statistics and the attained significance levels.

Table 1. Test plans and failure times

Voltage, kV	n_i	Failure time, min
26	3	5.79, 1579.52, 2323.7
28	5	68.85, 426.07, 110.29, 108.29, 1067.6
30	11	17.05, 22.66, 21.02, 175.88, 139.07, 144.12, 20.46, 43.40, 194.90, 47.30, 7.74
32	15	0.40, 82.85, 9.88, 89.29, 215.10, 2.75, 0.79, 15.93, 3.91, 0.27, 0.69, 100.58, 27.80, 13.95, 53.24
34	19	0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 3.16, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71, 72.89
36	15	1.97, 0.59, 2.58, 1.69, 2.71, 25.50, 0.35, 0.99, 3.99, 3.67, 2.07, 0.96, 5.35, 2.90, 13.77
38	8	0.47, 0.73, 1.40, 0.74, 0.39, 1.13, 0.09, 2.38

Table 2. Parameter estimates and confidence intervals

Baseline Law	Estimates	Confidence interval		$\ln L$
Exponential	$\hat{\beta}_0 = -64.1303$	−64.17	−64.10	−305.55
	$\hat{\beta}_1 = 17.481$	17.47	17.49	
Weibull	$\hat{\beta}_0 = -63.8973$	−63.98	−63.81	−300.83
	$\hat{\beta}_1 = 17.457$	17.43	17.48	
	$\hat{\theta}_1 = 0.7762$	0.74	0.82	
Gamma	$\hat{\beta}_0 = -64.3374$	−64.42	−64.26	−301.61
	$\hat{\beta}_1 = 17.434$	17.41	17.46	
	$\hat{\theta}_1 = 0.6923$	0.65	0.73	
Generalized Weibull	$\hat{\beta}_0 = -62.8442$	−62.98	−62.72	−300.47
	$\hat{\beta}_1 = 17.402$	17.36	17.44	
	$\hat{\theta}_1 = 0.9238$	0.87	0.98	
	$\hat{\theta}_2 = 1.6220$	1.53	1.72	

Table 3. Results of testing the hypothesis that the sample of residuals fits the baseline distributions of failures

Baseline Law	Criterion					
	Kolmogorov		Cramer–von Mises–Smirnov		Anderson–Darling	
	S_n^*	$P\{S_n > S_n^*\}$	S_n^*	$P\{S_n > S_n^*\}$	S_n^*	$P\{S_n > S_n^*\}$
Exponential	1.46	0.00	0.51	0.00	2.95	0.00
Weibull	0.67	0.33	0.06	0.29	0.39	0.38
Gamma	0.86	0.10	0.11	0.09	0.62	0.12
Generalized Weibull	0.55	0.76	0.04	0.71	0.28	0.70

Table 4. Test plans and failure times

Temperature, °C	Failure times, h
150	10 censored at the time 8064
170	1764, 2772, 3444, 3542, 3780, 4860, 5196 3 censored at the time 5448
190	408, 408, 1344, 1344, 1440 5 censored at the time 1680
220	408, 408, 504, 504, 504 5 censored at the time 528

Table 5. Baseline distribution models

Baseline distribution	Estimates	Confidence interval		$\ln L$
Exponential	$\hat{\beta}_0 = 15.2813$	15.20	15.36	-155.3647
	$\hat{\beta}_1 = -10.8408$	-10.88	-10.81	
Weibull	$\hat{\beta}_0 = 12.9681$	12.92	13.02	-146.2894
	$\hat{\beta}_1 = -9.5471$	-9.57	-9.53	
	$\hat{\theta}_1 = 3.0867$	2.76	3.44	
Gamma	$\hat{\beta}_0 = 14.5602$	14.50	1462	-147.3728
	$\hat{\beta}_1 = -9.5845$	-9.61	-9.56	
	$\hat{\theta}_1 = 4.4626$	4.26	4.68	
Generalized Weibull	$\hat{\beta}_0 = 10.5386$	10.43	10.57	-145.8709
	$\hat{\beta}_1 = -9.6778$	-9.73	-9.66	
	$\hat{\theta}_1 = 2.6329$	2.60	2.73	
	$\hat{\theta}_2 = 0.0010$	0.00	0.0013	
Lognormal	$\hat{\beta}_0 = 13.2782$	13.21	13.34	-148.5752
	$\hat{\beta}_1 = -9.6591$	9.63	9.69	
	$\hat{\theta}_1 = 0.5927$	0.54	0.65	

Because in this example, there were no censored measurements, in the tests and selection of the most appropriate AFT model, the results of [17–24] were used as the distributions of the goodness-of-fit test statistics. As is evident in Table 3, at a significance level $\alpha \leq 0.09$, the goodness-of-fit hypothesis is not rejected for all models, except for the model constructed on the basis of the exponential law. However, for the AFT model constructed on the basis of the generalized Weibull distribution, the attained significance levels for all tests are much higher. Therefore, this model is the preferred over all the baseline distributions considered.

Example 2. As an example of selection of the model from censored measurements, we consider data on engine failure. The test plan and failure data are given in [29], and their analysis is given in [30, 3].

In this present study, the objects were divided into four groups, each of which was observed under increased temperature. The test time was limited for each group. Failure was observed for 17 of the observed 40 objects. The purpose of the study was to estimate the reliability of the objects under normal operating conditions — at a temperature of 130 °C. Table 4 shows the increased stress-load values for each of the groups and the failure times.

As in [3], the stress function was the log-linear model $r(z(x)) = e^{\beta_0 + \beta_1 z(x)}$, where the load value (in Celsius degrees) is recalculated according to the relation $z(x) = 1000/(273.2 + x)$.

Table 5 shows the obtained estimates, the 95% confidence interval, and the values of the likelihood function for different parametric AFT models.

The values of $\ln L$ for the considered AFT models suggest that the models based on the baseline laws of the Weibull distribution and the generalized Weibull are preferred. To determine the most suitable AFT

Table 6. Results of testing of the hypothesis that the sample of residuals fits the baseline distribution

Baseline distribution	Criterion					
	Kolmogorov		Cramer–von Mises–Smirnov		Anderson–Darling	
	S_n^*	$P\{S_n > S_n^*\}$	S_n^*	$P\{S_n > S_n^*\}$	S_n^*	$P\{S_n > S_n^*\}$
Weibull	1.64	0.48	0.36	0.48	1.93	0.46
Generalized Weibull	1.63	0.42	0.35	0.41	1.90	0.40

model, using nonparametric goodness-of-fit tests, we tested the hypothesis that samples of residuals fit the corresponding baseline law. The distributions of the statistics of the tests used to test the composite hypotheses relative to the laws considered (with the degree of censoring available in this case) are unknown, but they are needed to draw the final conclusion. We emphasize that in testing composite hypotheses relative to the generalized Weibull distribution, the distributions of the statistics of the corresponding goodness-of-fit tests $G(S_n \mid H_0)$ depend on the value of the parameter θ_2 of this distribution.

Since the experiment described here involved type I censoring, to find the empirical distributions $G(S_n \mid H_0)$ of the statistics (11)–(13) of the modified nonparametric Kolmogorov, Cramer–von Mises–Smirnov and Anderson–Darling goodness-of-fit tests, we used a computer simulation technique [31] and the above algorithm for modeling the distributions of statistics of the goodness-of-fit tests applied to the analysis of AFT models. Empirical distributions of the statistics were modeled in accordance with the test conditions specified in Table 4 (loads, censoring times, and the number of groups and the number of objects in the group with $n = 40$). Then, the attained significance levels $G(S_n \mid H_0)$ were determined from the constructed distributions $P\{S_n > S_n^*\}$ of the statistics and the values of the goodness-of-fit statistics S_n^* calculated from samples of residuals.

The resulting values of the goodness-of-fit test statistics S_n^* and the attained significance levels $P\{S_n > S_n^*\}$ are given in Table 6. We see that in both cases, the tested hypothesis is not rejected, but for the model constructed using the Weibull distribution, the attained significance levels are somewhat higher. Thus, based on the tests of the hypotheses, preference is given to the simpler AFT model of the reliability function constructed using the Weibull distribution as the baseline failure law. At the same time, the values of the test statistic show that, in this case, the AFT model obtained with the choice of the generalized Weibull distribution as the baseline law, which has the shape parameter, is closer to the Kaplan–Meier estimate calculated from the results of tests and which is an analog of an empirical distribution function.

Survival and reliability problems often involve type III censoring, in which the censoring times are independent random variables. Problems encountered in the construction and validation of AFT models with random censoring are more complicated and less well studied. A fundamental difficulty is that, as a rule, there is no information on the distribution of the censoring times. In the present work, we confined ourselves to the results for complete samples and samples with types I and II censoring; the problems associated with the validation of AFT models with random censoring of test results will be considered in the future.

CONCLUSIONS

In the present paper, we considered the construction of accelerated failure test models for the most commonly used distribution functions of different loads used in reliability and survival analysis.

The usefulness of a constructed AFT model is determined by how exactly the predicted reliability function obtained from the results of accelerated failure tests (at elevated loads) of test objects corresponds to the unknown reliability function that characterizes the survival of objects under normal conditions.

The validity of constructed parametric AFT models can be judged from an analysis of the sample of residuals. Decision on the validity of the model can be made from the results of testing the hypothesis that the sample of residuals fits the baseline distribution law. This is done using the modified nonparametric goodness-of-fit tests.

In the case of complete data (in the absence of censored observations) and at loads constant in time, the models constructed in [17–24] for the classical tests for composite hypotheses are used as the as limiting distributions of the statistics of the modified Kolmogorov, Cramer–von Mises–Smirnov, and Anderson–Darling tests,

In the presence of I or II type censored observations, the required distributions of the statistics of the modified goodness-of-fit tests can be obtained without major difficulties by statistical modeling.

It is shown that the best baseline law and, hence, the best AFT model of the models considered can be selected by consistently analyzing the sample of residuals corresponding to some set of baseline probability distributions and determining the maximum attained significance level for the goodness-of-fit tests.

This work was supported by the Ministry of Education and Science of the Russian Federation (Grant No. 8.1274.2011) and the Federal target program "Scientific and scientific-pedagogical personnel of innovative Russia."

REFERENCES

1. J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data* (Springer, New York, 2003).
2. J. F. Lawless, *Statistical Models and Methods for Lifetime Data* (John Wiley and Sons, Hoboken, 2002).
3. J. D. Kalbeisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data* (John Wiley and Sons, New York, 2002).
4. C. Huber, N. Limnios, and M. Nikulin, *Mathematical Methods in Survival Analysis, Reliability and Quality of Life* (Wiley-ISTE, New Jersey, 2008).
5. V. Bagdonavicius and M. Nikulin, *Accelerated Life Models. Modeling and Statistical Analysis* (Chapman and Hall/CRC, New York, 2002).
6. M. Nikulin, *Accelerated life models, Encyclopedia of Statistics in Quality and Reliability* (John Wiley and Sons, New York, 2007), pp. 503–506.
7. M. Nikulin, "Step Stress Accelerated Life Testing: Classical Methods," Ibid, pp. 57–62.
8. A. V. Antonov and M. C. Nikulin, *Statistical Models in Reliability Theory: Textbook* (Abris, Moscow, 2012).
9. E. T. Lee and J. W. Wang, *Statistical Methods for Survival Data Analysis* (John Wiley and Sons, Hoboken, 2003).
10. D. R. Cox and D. Oakes, *Analysis of Survival Data* (Finansy i Statistika, Moscow, 1988; Chapman and Hall/CRC, 1984).
11. B. Yu. Lemeshko, S. Y. Gil'debrant, and S. N. Postovalov, "On the Estimation of Reliability Parameters from Censored Samples," Zavodskaya Laboratoriya. Diagnostika Materialov **67** (1), 52–64 (2001).
12. B. Yu. Lemeshko, "On the Estimation of the Distribution Parameters and Testing of Hypotheses from Censored Samples," Metody Menedzmenta Kachestva, No. 4, 32–38 (2001).
13. B. Yu. Lemeshko and S. N. Postovalov, "On the Distribution of Nonparametric Goodness-of-Fit Tests as Estimated by the Sample Parameters of Experimentally Observed Laws," Zavodskaya Laboratoriya. Diagnostika Materialov **64** (3), 61–72 (1998).
14. B. Yu. Lemeshko and S. N. Postovalov, "Application of the Nonparametric Goodness-of-Fit Tests for Testing Composite Hypotheses," Avtometriya, **37** (2), 88–102 (2001).
15. *R 50.1.037 - 2002. Recommendations for standardization. Applied Statistics. Rules of Testing the Goodness of Fit of Experimental and Theoretical Distributions. Part II. Nonparametric Tests* (Izd. Standartov, Moscow, 2002) [in Russian].
16. B. Yu. Lemeshko and A. A. Maklakov, "Nonparametric Tests for Testing Composite Hypotheses on Goodness-of-Fit to Exponential Family Distributions," Avtometriya **40** (3), 3–20 (2004).
17. B. Yu. Lemeshko, S. B. Lemeshko, and S. N. Postovalov, "Statistic Distribution Models for Some Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses," Commun. Statist. Theory and Methods **39** (3), 460–471 (2010).
18. B. Yu. Lemeshko and C. B. Lemeshko, "Models of Distributions of Nonparametric Goodness-of-Fit Test Statistics in Verifying Composite Hypotheses Using Maximum-Likelihood Estimators, Part. I," Izmeritel'naya Tekhnka, No. 6, 3–11 (2009).
19. B. Yu. Lemeshko and S. B. Lemeshko, "Models of Distributions of Nonparametric Goodness-of-Fit Test Statistics in Verifying Composite Hypotheses Using Maximum-Likelihood Estimators. Part II", Izmeritel'naya Tekhnka, No. 8, 17–26 (2009).
20. K. A. Akushkina, S. B. Lemeshko, and B. Yu. Lemeshko, "Models for Statistical Distributions in Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses of the Generalized Weibull Distribution," in *Proc. of the 3rd Intern. Conf. on Accelerated Life Testing, Reliability-Based Analysis and Design*, 19–21 May, 2010, Clermont-Ferrand, France, pp. 125–132.
21. B. Yu. Lemeshko, C. B. Lemeshko, M. C. Nikulin, and H. Saaidia, "Modeling Distributions of Nonparametric Goodness-of Fit Test Statistics in Testing Composite Hypotheses with Respect to the Inverse Gaussian Law," AIT, No. 7, 83–102 (2010).
22. B. Yu. Lemeshko and S. B. Lemeshko, "Models of Statistic Distributions of Nonparametric Goodness-of-Fit Tests in Composite Hypotheses Testing for Double Exponential Law Cases," Commun. Statist. Theory and Methods **40** (16), 2879–2892 (2011).

23. B. Yu. Lemeshko and S. B. Lemeshko, "Construction of Statistic Distribution Models for Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses: The Computer Approach," Quality Technology and Quantitative Management, **8** (4), 359–373 (2011).
24. B. Yu. Lemeshko, C. B. Lemeshko, C. N. Postovalov, and E. V. Chimitova, *Statistical Analysis, Modeling, and Investigation of Probabilistic Regularities: Computer Approach* (Izd. NGTU, Novosibirsk, 2011).
25. B. Yu. Lemeshko, E. V. Chimitova, and T. A. Pleshkova, "Testing Simple and Composite Goodness-of-Fit Hypotheses from Censored Samples," Nauch. Vestn. NGTU, No. 4, 13–28 (2010).
26. E. L Kaplan and P. Meier, "Nonparametric Estimation from Incomplete Observations," J. Amer. Statist. Association **5** (282), 457–481 (1958).
27. A. H. Bol'shev and H. V. Smirnov, *Tables of Mathematical Statistics* (Nauka, Moscow, 1983) [in Russian].
28. *LiTiS — The Software System for Statistical Analysis of Lifetime Data*. http://amsa.conf.nstu.ru/amsa2011/Litis.msi.
29. D. E. Crawford, "Analysis of Incomplete Life Test Data on Motorettes," Insulation/Circuits **16** (10), 43–48 (1970).
30. R. L. Prentice, "Exponential Survivals with Censoring and Explanatory Variables," Biometrika **60** (2), 279–288 (1973).
31. B. Yu. Lemeshko and C. N. Postovalov, *Computer Technologies of Data Analysis and Investigation of Statistical Regularities: Textbook* (Izd. NGTU, Novosibirsk, 2004).