

# STATISTICAL ANALYSIS OF INTERVAL AND IMPRECISE DATA - APPLICATIONS IN THE ANALYSIS OF RELIABILITY FIELD DATA

Hryniewicz Olgierd

Systems Research Institute,  
Warszawa, Poland

## Keywords

reliability, statistical analysis, field data, interval data, fuzzy data

## Abstract

The analysis of field lifetime data is much more complicated than the analysis of the results of reliability laboratory tests. In the paper we present an overview of the most important problems of the statistical analysis of field lifetime data, and present their solutions known from literature. When the input information is partial or imprecise, we propose to use interval arithmetics for the calculation of bounds on reliability characteristics of interest. When this information can be described in a more informative fuzzy form, we can generalize our interval-valued results to their fuzzy equivalents.

## 1. Introduction

Statistical analysis of results of lifetime tests has its over 50 years lasting history. In contrast to methods usually applied for the analysis of ordinary statistical data, in case of lifetime data we have to take into account such specific features like censoring of observations or the existence of covariates. First applications of this methodology were designed for the analysis of reliability data. However, starting from the 1970's their main field of applications is the survival analysis applied not only to technical objects, but to human beings as well.

In classical textbooks on reliability it is always assumed that  $n$  independent objects (systems or components) are put on test, and in the ideal case of no censoring we observe the realizations of  $n$  independent and identically distributed (*iid*) random variables  $T_1, \dots, T_n$ . When the lifetime test is terminated after the  $r$ -th observed failure, e.g. when we observe a predetermined number of failures  $r$  at times  $t_{(1)} < \dots < t_{(r)}$ , and the remaining  $n-r$  objects survive a random censoring time  $t_{(r)}$ , we have the case of the type-II censoring. On the other hand, when the lifetime test is terminated at the predetermined time  $t_B$ , and the number of observed failures is a random variable, we have the case of the type-I censoring. In more general models, we may also assume the cases of individual random censoring (when we observe random variables  $X_i = \min(T_i, C_i)$   $i=1, \dots, n$ , where  $C_i, i=1, \dots, n$  are random, and independent from  $T_i, i=1, \dots, n$  censoring times), multiple censoring (when for each subgroup of tested objects there exists a predetermined censoring time), or progressive censoring (when a predetermined number of objects are withdrawn from the test after each observed failure). The detailed description of these censoring schemes may be found in classical textbooks on the analysis of lifetime data, such as the book of Lawless [16].

Unfortunately, the practical applicability of well known methods is often limited to the case of precisely designed laboratory tests, when all important assumptions made by statisticians are at least approximately fulfilled. These tests provide precise information about lifetimes and censoring times, but due to the restricted (usually low) number of observed failures and/or the restricted test times, the accuracy of reliability estimation is rather low. Moreover, some types of possible failures may not be observed in such tests (e.g. due to their limited duration), and the obtained estimates of reliability may be overestimated.

It is beyond any discussion that the most informative reliability data may come only from field experiments, i.e. from the exploitation of considered objects in real conditions. Unfortunately, we have very seldom statistical data that are obtained under field conditions and fit exactly to the well known theoretical

models. For example, test conditions are usually not exactly the same for all considered objects. Therefore, the random variables that describe their lifetimes are not identically distributed. Another serious practical problem is related to the lack of precision in reported lifetime data. Individual lifetimes are often imprecisely recorded. For example, they are presented in a grouped form, when only the number of failures which occurred during a certain time interval is recorded. Sometimes, times to failures are reported as calendar times, and this does not necessarily mean the same as if they were reported as times of actual operation. Finally, reliability data that come from warranty and other service programs are not appropriately balanced; there exists more information related to a relatively short warranty time, and significantly less information about the objects, which survived that time.

Statisticians who work with lifetime data have tried to build models that are useful for the analysis of field data. The majority of papers devoted to this problem are related to the methodology of dealing with data from warranty programs. These programs should be considered as the main source of reliability field data. Therefore, the presentation of the statistical problems of the analysis of warranty data shall be an important part of this paper. Some important mathematical models related to the analysis of warranty data are presented in the second section of this paper. In all these models it is assumed that all observations are described more or less precisely, and all necessary probability distributions are either known or evaluated using precisely defined statistical data. In many cases this approach is fully justified. However, close examination of real practical problems shows that in many cases available statistical data are reported imprecisely. We claim that making these data precise by force may introduce unnecessary errors. Therefore, we believe that in case of really imprecise data this fact should be taken into account in an appropriate way. In the third section of the paper we present the solution of some chosen practical problems when the input information is given in an interval form. These results are generalized in the fourth section to the case of fuzzy input data. Some conclusions and proposals for future investigations are presented in the last section of the paper.

## **2. Mathematical models of reliability field data coming from warranty programs**

Lifetime data collected during precisely controlled laboratory test provide important, albeit limited, information about reliability of tested equipment. This limitation has different reasons. First, the number of tested units and/or the duration of a lifetime test are usually very limited due to economic constraints. Second, controlled laboratory conditions do not reflect real usage conditions. For this reason, for example, some of important types of failures cannot be revealed during the test. Finally, only field data can provide useful information about dependencies between reliability characteristics of tested units and specific conditions of exploitation. In contrast to laboratory lifetime data, reliability field data may yield much more interesting information to a manufacturer. Unfortunately, the information that is characteristic for laboratory data is seldom available in case of field data. First of all, warranty programs that serve as the main source of reliability field data are not designed to collect precise data. For example, reliability data are collected only from those items that have failed during the warranty period. Moreover, this period may not be uniquely defined. It is a common practice to define the warranty period both in calendar (for example, one year) and operational (for example, in terms of mileage) time. Therefore, in many practical cases reliability data are intrinsically imprecise. Also exploitation conditions, important for the correct assessment of reliability, are not precisely reported. All those problems, and many others, make the statistical analysis of reliability field data a difficult problem. Therefore, despite its practical importance, the number of statistical papers devoted to the analysis of reliability field data is surprisingly low.

Statistical analysis of reliability field data coming from warranty programs can be roughly divided into two related, but distinct, parts: analysis of claims processes and the analysis of lifetime probability distributions. From the point of view of a manufacturer the most important information is contained in the description of the process of warranty claims. Comprehensive description of this type of analysis can be found in the papers by Lawless [17] and Kalbfleisch et al. [14]. In the analysis of claims processes statistical data are discrete, and are described by stochastic count processes like the Poisson process or its generalizations. By analysing count reliability data we can estimate such important characteristics as, e.g., the expected number of warranty claims during a specific period of time, the expected costs of such claims, etc. This type of information is extremely important for designing of warranty programs, planning of the supply of spare parts, and the evaluation of the efficiency of service activities, but does not yield precisely enough information about the intrinsic reliability characteristics of investigated units. Information of this

type is much more useful for improving the quality on the design stage of a product, especially for the comparison of different solutions, etc.

In this paper we limit the scope of our investigations to the statistical analysis of probability distributions of lifetimes. Throughout the paper we will denote by  $T$  the continuous random lifetime whose probability density function is denoted by  $f(t | \mathbf{x}; \boldsymbol{\theta})$ , where  $\mathbf{x}$  is a vector of parameters (covariates) that describe exploitation conditions, and  $\boldsymbol{\theta}$  is a vector of parameters that describe the lifetime distribution.

### 2.1. Estimation from truncated lifetime data

In case of the analysis of warranty data we often face situations when we observe both failure times  $t_i, i = 1, \dots$  and corresponding vectors of covariates  $\mathbf{x}_i, i = 1, \dots$  are observed only for failed units. Let  $T_c$  be a certain prespecified censoring time such that failures are observed only when  $T_i \leq T_c$ , where  $T_i, i = 1, \dots$  denote random variables describing lifetimes of failed units. If only lifetimes of failed units are available, and the form of the lifetime probability distribution is known, the statistical inference about parameters  $\boldsymbol{\theta}$  can be based on a truncated conditional likelihood function

$$L_T(\boldsymbol{\theta}) = \prod_{i: t_i \leq T_c} \frac{f(t_i | \mathbf{x}_i; \boldsymbol{\theta})}{F(T_c | \mathbf{x}_i; \boldsymbol{\theta})}, \tag{1}$$

where  $F(t | \mathbf{x}; \boldsymbol{\theta}) = P\{T \leq t | \mathbf{x}; \boldsymbol{\theta}\}$ . This likelihood function arises from the conditional (truncated at time  $T_c$ ) probability distribution of the random lifetime  $T$ . It is interesting to note, that the likelihood function (1) does not depend upon the number  $N$  of units in the considered population of tested items. Therefore, (1) is suitable for the estimation of  $\boldsymbol{\theta}$  when this number is unknown. Moreover, in case of a low proportion of failed items, this likelihood function can be quite uninformative, as it was noticed by Kalbfleisch and Lawless [13]. They showed using computer simulations that the variance of the estimators of unknown parameters is substantially larger than in the case when some information about non-failed units is available.

Estimation of  $\boldsymbol{\theta}$  using the likelihood function (1) is rather complicated, even in simple cases. A comprehensive presentation of this problem can be found in the book by Cohen [2]. A relatively simple solution was proposed by Cohen [1] for the lognormal probability distribution of lifetimes, i.e. when logarithms of observed lifetimes are distributed according to the normal distribution. In this case the maximum likelihood estimators based on (1), and the moment estimators based on the first two moments coincide, but computation requires either special tables or the usage of numerical procedures. Cohen [1] considered a single left truncation at  $X_0$ . In this case the  $k$ th sample moment is calculated from

$$v_k = \sum_{i=1}^n \frac{(x_i - X_0)^k}{n}. \tag{2}$$

For the estimation of the unknown parameters  $\sigma^2$  and  $\mu^*$  Cohen [1] proposed to use three first moments defined by (2). The obtained estimators can be calculated from the following simple formulae:

$$(\sigma^2)^* = \frac{v_2^2 - v_1 v_3}{v_2 - 2v_1^2}, \tag{3}$$

and

$$\mu^* = X_0 + \frac{v_3 - 2v_1 v_2}{v_2 - 2v_1^2}. \tag{4}$$

These formulae are derived for the left truncated sample. However, they can be applied in case of right truncated samples, in which case the odd moments are negative. The solution given by (3) and (4) is theoretically less efficient than the maximum likelihood estimators obtained from (1). However, Rai and Singh [21] have shown using extensive Monte Carlo simulations that there is no significant difference between these two methods. However, the efficiency of these estimators decreases, as expected, significantly when the percentage of truncated (i.e. not observed) lifetimes is larger than 30%.

## 2.2. Estimation from censored lifetime data with full information about censored lifetimes

In the previous subsection we considered the case when only the data from units failed prior to a certain time  $T_c$  are available. The situation when the information about non-failed units is available is well known as “censoring”. Following Hu and Lawless [9] let us present the general mathematical model of lifetime data. We consider population  $\mathbf{P}$  consisting of  $n$  units described by their lifetimes,  $t_i, i = 1, \dots, n$ , random censoring times,  $\tau_i, i = 1, \dots, n$ , and vectors of covariates  $\mathbf{z}_i, i = 1, \dots, n$ , respectively. Triplets  $(t_i, \tau_i, \mathbf{z}_i)$  are the realizations of a random sample from a distribution with joint probability function

$$f(t | \theta; \tau, \mathbf{z}) dG(\tau, \mathbf{z}), t > 0, \tau > 0, \mathbf{z} \in R^q, \tag{5}$$

where lifetimes and censoring times are usually considered independent given fixed  $\mathbf{z}$ , and  $G(\tau, \mathbf{z})$  is an arbitrary cumulative distribution function. Let  $O$  be the set of  $m$  units for whom the lifetimes are observed, i.e. for whom  $t_i \leq \tau_i, i = 1, \dots, n$ . The remaining  $n-m$  units belong to the set  $C$  of censored lifetimes for whom only their censoring times  $\tau_i$  and covariates  $\mathbf{z}_i$  are known. The function  $S(t | \theta; \tau, \mathbf{z}) = 1 - F(t | \theta; \tau, \mathbf{z})$ , where  $F(t | \theta; \tau, \mathbf{z})$  is the cumulative distribution function of the lifetime, is called in the literature the survivor function or the survival function. The likelihood function that describes the lifetime data is now given by [9]

$$L(\theta) = \prod_{i \in O} f(t_i | \theta; \tau_i, \mathbf{z}_i) dG(\tau_i, \mathbf{z}_i) \times \prod_{i \in C} S(t_i | \theta; \tau_i, \mathbf{z}_i) dG(\tau_i, \mathbf{z}_i) \tag{6}$$

The special cases of (6) are well known, and comprehensively described in many reliability textbooks, such as an excellent book by Lawless [16]. However, they are rather well suited for the description of laboratory life tests, where all censoring times are known, and the values of covariates that describe test conditions are under control. In this paper we recall only those results, which in our opinion are pertinent to the analysis of field lifetime data.

One of the features that distinguish reliability field tests from laboratory tests is the variety of test conditions. In the laboratory test these conditions are usually the same for all tested units. Only in case of accelerated lifetime these test conditions are different for different groups of tested units. In contrast to this situation, in reliability field tests usage conditions may be different for all tested units. Therefore, statistical methods that allow taking into account different test conditions are especially useful for the analysis of reliability field data.

There exist two general mathematical models that link lifetimes to test conditions and are frequently used in practice: proportional hazard models, and location-scale regression models. In the proportional hazard models the hazard function, defined as  $h(t; \theta, \mathbf{z}) = f(t; \theta, \mathbf{z}) / S(t; \theta, \mathbf{z})$ , is linked to the test conditions by the following equation

$$h(t | \mathbf{z}) = h_0(t)g(\mathbf{z}), \tag{7}$$

where functions  $h_0(\cdot)$  and  $g(\cdot)$  may have unknown parameters which have to be estimated from statistical data. Another representation of the proportional hazard model is the following:

$$S(t | \mathbf{z}) = S_0(t)^{g(\mathbf{z})}. \tag{8}$$

The most frequently used model is given by the following expression

$$h(t | \mathbf{z}) = h_0(t)e^{z\beta}, \tag{9}$$

where  $\mathbf{z}\beta = z_1\beta_1 + \dots + z_q\beta_q$ , and  $\beta_j$ 's are unknown regression coefficients. This model was investigated by many authors. To give an illustration of its application let us recall the results given in Lawless [16] for the case of the Weibull distribution of lifetimes.

The Weibull probability distribution is the most frequently used mathematical model of lifetime data. In the considered case of the proportional hazard model its survivor function is given by the following expression

$$S(t | \mathbf{z}) = \exp\left[-(te^{-z\beta})^\delta\right], \tag{10}$$

where  $\delta > 0$  is the shape parameter, responsible for the description of the type of failure processes. If we use the transformation  $Y = \log T$ , the logarithms of lifetimes are described by simple linear model

$$Y = \mathbf{z}\beta + \sigma W, \tag{11}$$

where  $W \sim \text{Weibull}$ , and the random variable  $W$  has a standard extreme value distribution with the probability density function  $\exp[w - \exp(w)]$ .

Suppose that  $n$  units are tested, and independent observations  $(x_i, \mathbf{z}_i), i = 1, \dots, n$  are available, where  $x_i$  is either a logarithm of lifetime or logarithm of censoring time of the  $i$ th tested unit. Additionally suppose that exactly  $r$  failures are observed. If we apply the maximum likelihood methodology to this model, we arrive at the following set of equations [16]:

$$-\frac{1}{\sigma} \sum_{i \in O} z_{il} + \frac{1}{\sigma} \sum_{i=1}^n z_{il} e^{x_i} = 0, l = 1, \dots, q \tag{12}$$

$$-\frac{r}{\sigma} - \frac{1}{\sigma} \sum_{i \in O} x_i + \frac{1}{\sigma} \sum_{i=1}^n x_i e^{x_i} = 0, \tag{13}$$

where  $x_i = (y_i - z_i\beta) / \sigma$ . The solution of  $q+1$  equations given by (12) and (13) yields the maximum likelihood estimators of  $\beta$  (and hence for the shape parameter  $\delta$ ), and regression coefficients  $\beta_1, \dots, \beta_q$ . The formulae for the calculation of the asymptotic covariance matrix of these estimators can be found in [16].

A second regression model commonly used for the analysis of lifetimes is the location-scale model for the logarithm of lifetime  $T$ . In this model the random variable  $Y = \log T$  has a distribution with the location parameter  $\mu(\mathbf{z})$ , and a scale parameter  $\sigma$ , which does not depend upon the covariates  $\mathbf{z}$ . This model can be written as follows:

$$Y = \mu(\mathbf{z}) + \sigma\xi, \tag{14}$$

where  $\sigma > 0$  and  $\xi$  is a random variable with a distribution that is independent on  $\mathbf{z}$ . Alternative representation of this model can be written as

$$S(t | \mathbf{z}) = S_0\left(\frac{t}{\alpha(\mathbf{z})}\right). \tag{15}$$

Both models, i.e. proportional hazard model and location-scale model, have been applied for different probability distributions of lifetimes. The detailed description of those results can be found, for example, in [16]. However, it is worth to note, that only in the case of the Weibull distribution (and the exponential distribution, which is a special case of the Weibull distribution) both models coincide.

When the type of the lifetime probability distribution is not known and the proportional hazards model seems to be appropriate we can apply distribution-free methods for the analysis of lifetimes. Let (8) be of the form

$$S(t | \mathbf{z}) = S_0(t)^{z\beta} \tag{16}$$

Cox [5] proposed a method for the separation of the estimation of the vector of regression coefficients  $\beta$  from the estimation of the survivor function  $S_0(t)$ . Suppose that the observed lifetimes are ordered as follows:  $t_{(1)} < \dots < t_{(m)}$ . Let  $R_i = R(t_{(i)})$  be the set of all units being at risk at time  $t_{(i)}$ , that is the set of all non-failed and uncensored units just prior to  $t_{(i)}$ . Note, that in this model censoring times of the remaining  $n - m$  units may take arbitrary values. For the estimation of  $\beta$  Cox [5] proposed to use a pseudo-likelihood function given by

$$L(\beta) = \prod_{i=1}^m \left( \frac{e^{z_{(i)}\beta}}{\sum_{l \in R_i} e^{z_{(i)}\beta}} \right) \tag{17}$$

Slight modification of (17) has been proposed in Lawless [16]. This modification allows for few multiple failures at times  $t_{(i)}, i = 1, \dots, m$ . Let  $D_i$  be the set of units that fail at  $t_{(i)}$ ,  $d_i$  be the number of those units, i.e.  $d_i = |D_i|$ , and  $\Xi_i = \sum_{l \in D_i} \mathbf{z}_l$ . The likelihood function is now given by [16]

$$L(\beta) = \prod_{i=1}^m \left[ \frac{e^{\Xi_i\beta}}{\left( \sum_{l \in R_i} e^{z_{(i)}\beta} \right)^{d_i}} \right] \tag{18}$$

The maximum likelihood estimators of the regression coefficients  $\beta$  are found from the following equations:

$$\sum_{i=1}^m \left( \Xi_{ir} - d_i \sum_{l \in D_i} z_{lr} e^{z_{(i)}\beta} / \sum_{l \in D_i} e^{z_{(i)}\beta} \right) = 0, r = 1, \dots, q, \tag{19}$$

where  $\Xi_{ir}$  is the  $r$ th component in  $\Xi_i = (\Xi_{i1}, \dots, \Xi_{iq})$ . Formulae for the calculation of the asymptotic covariance matrix of these estimators are given in [16]. When the vector of the regression coefficients  $\beta$  has been estimated, we can use a distribution-free method, such as the Kaplan-Meier estimator [15], for the estimation of  $S_0(t)$ .

### 2.3. Estimation from censored lifetime data with incomplete information about censored lifetimes

In case of real field lifetime data the full information about the non-failed units is often unavailable, even in the case when there exists full and precise information about all failed units. Consider, for example, the data from warranty programs. Suppose that we do our analysis at a certain moment of time using the data (lifetimes and values of covariates) on all units that have failed by that moment. As we usually do not have information about the units which have not failed, we neither know their censoring times nor the values of their corresponding covariates. Moreover, we may also not know even the total number of units  $n$ . However, if even partial information about these units is available, it can be used for the improvement of the efficiency

of estimation. This information may come, for example, from the follow-ups of certain units during the warranty period or monitoring of some units after their warranty has been expired.

Suzuki [22], [23] was one of the first researchers who considered the case of incomplete information coming from field reliability data. In [23] he considered the case when a certain fraction  $p^*$  of units is additionally monitored during their warranty period. Thus, we have lifetimes of all units that have failed during the warranty period and all censoring times that do not failed during the warranty period but have been monitored. Under the assumption of random censoring times independent from random times to failure Suzuki [23] derived the maximum likelihood estimator of the survivor function  $S(t)$  that generalizes the estimator proposed by Kaplan – Meier [15]. In [22] Suzuki applied his methodology to find estimators of the parameters of such lifetime distributions like the exponential distribution or the Weibull distribution. Consider, for example, the exponential distribution with the survivor function  $S(t) = \exp(-\lambda t), \lambda > 0, t > 0$ .

Let  $t_1, \dots, t_m$  be the observed lifetimes of  $m$  failed units, and  $t_1^*, \dots, t_k^*$  be the known censoring times of those  $k$  monitored units that have not failed during the warranty period. The censoring times of the remaining  $n_l = n - m - k$  units that have not failed during the warranty period are unknown. The maximum likelihood estimator of the hazard rate  $\lambda$  is given as [22]

$$\lambda^* = \frac{m}{\sum_{i=1}^m t_i + \left(1 + \frac{n_l}{k}\right) \sum_{i=1}^k t_i^*} \tag{20}$$

In the similar case of the Weibull distribution Suzuki [22] derived modified maximum likelihood equations.

In [22] Suzuki considered also another problem related to the analysis of warranty data. In modern warranty systems the warranty “time” is often bi-dimensional. For example, for newly sold cars warranties are defined both in terms of calendar time and mileage. Thus, failures that occurred during the calendar-time warranty period but after the moment when the maximum mileage had been exceeded are not reported. Formulae used for the calculation of respective estimators are more complicated in this case. A more general model, when the additional information about covariates is available, was considered by Kalbfleisch and Lawless [13].

The results of Suzuki [22], [23] originated the paper by Oh and Bai [20] who considered the case when monitoring of certain units taken randomly from the whole population of considered objects is monitored not only during a warranty period, but also during some after-warranty period. They assumed that: (i) each failure that occurs during a warranty period  $(0, T_1]$  is reported with probability 1; (ii) each failure that occurs during an after-warranty period  $(T_1, T_2]$  is reported with probability  $p$ , and (iii) each unreported unit either fails during the after-warranty period but is not reported with probability  $1-p$  or survives time  $T_2$ . Let  $f(t; \theta)$  be the probability density function of the lifetime, and  $S(t; \theta)$  be the corresponding survivor function of the considered objects. We assume that the vector of parameters  $\theta$  is unknown, but we know probability  $p$ . In this case the log-likelihood function is given by [20]

$$\begin{aligned} \log L(\theta) = & \sum_{i \in D_1} \log\{f(t_i; \theta)\} + \sum_{i \in D_2} [\log p + \log\{f(t_i; \theta)\}] \\ & + n_3 \log[(1 - p)S(T_1; \theta) + pS(T_2; \theta)] \end{aligned} \tag{21}$$

where  $D_1$  is the set of units which failed during the warranty period  $(0, T_1]$ ,  $D_2$  is the set of units failed and reported during the after-warranty period  $(T_1, T_2]$ , and  $n_3$  is the number of units (both failed and not failed) not reported during  $(0, T_2]$ . Maximum likelihood estimators of  $\theta$  can be found, as usual, by the maximization of (21). Oh and Bai [20] considered also a more difficult problem when the probability of revealing failures during the after-warranty period is unknown. To solve this problem they applied the EM maximum likelihood algorithm and proposed an iterative

procedure for finding the estimators of  $\square$ . For both cases of known and unknown  $p$  Oh and Bai [20] calculated the asymptotic covariance matrix of the obtained estimators. Another approach was used by Hu et al. [11] who have found non-parametric estimators of the probability distribution of the time to failure  $f(t)$  when the additional information about the probability distribution of censoring times is available. Hu et al. [11] assumed that times to failure and censoring times are described by mutually independent discrete random variables and found moment and maximum likelihood estimators of  $f(t)$ .

The problem of two time scales mentioned in the paper by Suzuki [22] has attracted many researchers. The general discussion of the alternative time scales in modelling lifetimes is considered in the paper by Duchesne and Lawless [7]. In the considered in this paper context of the analysis of field reliability data this problem was considered by several authors. For example, Lawless et al. [18] considered the following linear transformation of the original calendar time  $t$  to an operational (usage) time  $u$

$$u_i(t) = \alpha_i t, \quad t \geq 0, \quad (22)$$

where  $\alpha_i$  is a random usage rate described by the cumulative probability function  $G(\alpha)$ . Jung and Bai [12] have used this approach for the analysis of lifetime data coming from warranty programs when warranty periods were defined in two time scales (e.g. calendar time and mileage). The results of their computations are rather difficult for real applications, and cannot be applied without a specialized software. Moreover, this model requires the knowledge of  $G(\alpha)$ , and this probability distribution is rarely known for practitioners.

Another approach for solving this problem was proposed by Jung and Bai [12] who described lifetime data by a bivariate Weibull distribution. They calculated a very complicated log-likelihood function that can be used for the estimation of the parameters of this distribution when the data are reported both in calendar time and operational time. They showed an example where this approach may be more appropriate than the linear transformation model proposed by Lawless et al. [18].

Reliability field data may be collected and stored also in other forms that are far from those known in classical textbooks. Coit and Dey [3], and Coit and Jin [4] consider the case, typical for the collection of real reliability data, when data from different test programs are available in a form  $(r, T_r)$ , where  $r$  is the number of observed failures, and  $T_r$  is the cumulative time on test for the data record with  $r$  failures. Coit and Dey [3] considered the case when lifetimes are distributed according to the exponential distribution. They proposed the test for the verification of this assumption.

Coit and Jin [4] considered a case when lifetimes are distributed according to the gamma distribution

$$f(t) = \lambda^k t^{k-1} e^{-\lambda t} / \Gamma(k), \quad t > 0, k > 0, \lambda > 0 \quad (23)$$

They considered the case typical for the analysis of field data for repairable objects, where a single data record consists of the number of observed failures and total time between those failures. Let  $T_{rj}$  be the  $j$ th cumulative operating time for the data record with exactly  $r$  failures (Note, that no censoring is considered in this case);  $n_r$  be the number of data records with exactly  $r$  failures;  $m$  be the maximum number of failures for any considered data record;  $M$  be the total number of observed failures, i.e.  $M = \sum_{r=1}^m r n_r$ ; and  $\bar{t} = \sum_{r=1}^m \sum_{j=1}^{n_r} T_{rj} / M$  be the average time to failure. The maximum likelihood estimator of the shape parameter  $k$  can be found from the equation [4]

$$\sum_{r=1}^m r n_r \psi(rk) - M \ln k = K', \quad (24)$$

where

$$K' = \sum_{r=1}^m \sum_{j=1}^{n_r} r \ln T_{rj} - M \ln \bar{t}, \quad (25)$$



and

$$\psi(rk) = \ln(rk) - \frac{1}{2rk} - \frac{1}{12(rk)^2} + \frac{1}{120(rk)^4} - \dots \quad (26)$$

is the digamma function. The estimator of the parameter  $\square$  is simply given by  $\hat{\kappa} = \bar{\kappa} / \bar{t}$ .

Another type of reliability field data was considered in papers by Usher [24] and Lin et al. [19]. These authors considered the case of so-called masked data. This type of data is observed when lifetimes of whole systems are observed, but the exact cause of failure (i.e. a component that failed) can be isolated only to some subset of components. Unfortunately, the problem of estimation of lifetime characteristics has been solved only either in the case of two-component systems [24] or in the case when there exists additional prior information about reliability of considered components.

### 2.4. Estimation of the failure rate from field data

Hu and Lawless [10] considered the case when reliability data sets contain information not only on times to first failures, but also on times to consecutive failures if the observed units failed several times during a warranty period. In such cases, which are typical for the reliability analysis of repairable objects, the most frequently used model that describes the process of failures is a Poisson process characterized by a failure rate  $\square$ . When the failure rate varies in time the process of failures is called the non-homogeneous Poisson process, and the reliability characteristic of interest is the time-dependent failure rate  $\lambda(t)$ . Hu and Lawless [10] considered parametric and non-parametric estimation of  $\lambda(t)$  in two cases: when only data on failed units are reported (i.e. in case of data truncation, when the number of non-failed units is unknown), and when the population's size and the distribution of individual censoring times are known. In the first case the estimator of the failure rate  $\lambda(t)$  can be found iteratively. In the second case the complexity of computations depends on the amount of knowledge about the population size and the distribution of censoring times.

### 3. Statistical analysis of reliability field data with incomplete interval-type information

In the previous section we have presented different mathematical models that can be used for the analysis of reliability field data. This type of lifetime data is in general more difficult to analyse using classical statistical methods. What is typical to field data is the existence of missing, unobserved or imprecisely reported data. If we want to analyse such data using a thorough statistical approach we immediately are in troubles. First of all, additional statistical information is needed which is necessary for the description of missing or imprecisely reported data in terms of the theory of probability. For example, if lifetime data are imprecisely reported due to the unknown delay time, see [17] for the description of the problem, the probability distribution of the delay time has to be identified using independent investigation. The same problem arises when we need to know the usage rate. The probability distribution of  $\alpha_i$  in (22) has to be estimated even in the case when there exist doubts whether such unique distribution ever exists. Another group of problems arises even in those cases when the additional information is available. Mathematical models used for the estimation of reliability characteristics become very complicated, and in many cases are rather intractable for an average user. Specialized software is needed, and this software is rarely commercially available. In all these and similar cases there exists, however, additional imprecise information about *possible* values of the quantities of interest. This information may be expressed in a form of *intervals* of possible values of model parameters or values of imprecisely reported observations. It has to be noted that this type of the representation of imprecision is *not* equivalent to the assumption that quantities with unknown or imprecisely reported values are *uniformly* distributed on those intervals. The interpretation of these intervals should be rather made in the spirit of the classical theory of measurement. If such unknown or imprecisely reported quantity is represented by the interval of its possible values it may be understood as if that value could be represented by *any* probability distribution defined over such interval. Thus, the application of interval data yields both pessimistic and optimistic *bounds* for the reliability characteristics of

interest. In this section we present some examples of the usage of this approach in dealing with reliability field data.

Let us begin with the simplest model of life data. Suppose that  $M$  units are tested during a fixed time period  $T_0$ . Let  $t_1, \dots, t_m$  be the observed times of  $m$  reported failures. The failures of the remaining  $M-m$  units have not been reported by the time  $T_0$ , and we assume that for these units  $T_0$  is their censoring time. As usual, we assume that  $f(t; \theta)$  is the density function of the time to failure, and  $\theta$  is a vector of its unknown parameters. The maximum likelihood estimators of  $\theta$  can be found by the maximization of the log-likelihood function

$$L(\theta) = \sum_{i=1}^m \log f(t_i; \theta) + (M - m)S(T_0; \theta), \tag{27}$$

where  $S(T_0; \theta) = P(T > T_0)$  is the survivor function. The problem stated above is a classical statistical problem extensively investigated for numerous probability distributions of lifetimes. Consider now its more realistic version. First of all let us assume that the reported failure times  $T_i$  do not represent real failure times due to some random delay. For example, a transmission leakage in a car may be reported after a visit to a service centre, and not after observing its first signs on a garage floor [21]. Let  $D_i$  be a random delay time. Hence, the real time to failure is described by an unobserved random variable. Note however, that even if observed lifetimes  $T_i$  are distributed according to a well known probability distribution, e.g. the Weibull distribution, then the distribution of  $X_i$  may be completely different, even when the distribution of delays  $D_i$  is known. In real situation the distribution of  $D_i$  is usually very difficult to estimate, so the derivation of a more or less precise probabilistic model for the description of  $X_i$  is usually hardly possible. The existence of delays in the reporting of failures may cause additional complication. As a matter of fact we may not be sure if all failures have been reported by the censoring time  $T_0$ . We do not consider this possibility in our model, as its thorough description seems to be very complicated. Now, let us consider another serious problem with the analysis of reliability field data. In the majority of practical cases reliability engineers are rather not interested in the description of reliability in terms of calendar time, but in terms of operational or usage time. In the previous section we discussed some basic problems that arise when we want to model this situation. Even in the simplest case of a linear transformation of a calendar time to a usage time we have to know the daily usage rate  $U_i$  that is a random variable whose distribution is very difficult to estimate. In practice this can be done only for products like cars when the usage time is continuously monitored in an automatic way. In all other practical situations the usage rate may be only estimated from imprecise statements of users. Let  $Z_i$  be the lifetime in terms of usage time. Then we have  $Z_i = (T_i - D_i)U_i$ . In face of all difficulties mentioned above the derivation of the probability distribution of  $Z_i$  seems to be hardly possible. Finally, let us notice that different usage rates influence the values of censoring times of non-failed units. These censoring times are now the realizations of a random variable  $Z_0 = T_0 U$ , where  $U$  represents a random usage rate for non-failed units. It is quite obvious that this distribution can be estimated either using expert opinions or from a specially designed statistical experiment.

The discussion presented above shows quite clearly that even in the simplest case of the analysis of real field data the precise mathematical description of the problem becomes very difficult or even mathematically intractable. However, we still have to analyse the data in the form they are available to us. Therefore, there is a need to propose approximate methods that should be simple enough in order to be applied in practice. In this section of the paper we propose to represent our lack of knowledge in terms of intervals representing the values of considered characteristics or quantities.

In order to simplify further notation let us denote by  $\square x$  a compact interval  $[x_{min}, x_{max}]$ . The lack of knowledge about the precise value of the time to a real failure let us describe by assuming that the real time to failure takes place in the interval  $\square t_i$ , where  $t_{i,max}$  is equal to the reported failure time  $t_i$ . Similarly, we assume that the usage rate for each observed failed unit is described by the interval  $\square u_i$ , and the usage rate for all censored unit belongs to the interval  $\square u$ . Hence, we can calculate the interval the usage time to a failure belongs to. This can be done using the rules of simple interval arithmetics; the lower bound of the interval  $\square z_i$  is equal to  $z_{i,min} = t_{i,min} u_{i,min}$ , and the upper bound is given by  $z_{i,max} = t_{i,max} u_{i,max}$ . Similarly, the lower bound for the usage censoring time is given by  $Z_{0,min} = T_0 u_{min}$ , and the upper bound by  $Z_{0,max} = T_0 u_{max}$ . We should also make the assumption that the probability distribution of lifetimes belongs to

a certain class of probability distributions. This assumption is a crucial one, as strictly speaking this distribution is different from that which describes observed times to failures measured in the calendar time. However, when the intervals of interest are not very wide this assumption seems to be practically acceptable.

In the next step of our analysis we calculate a multivariate interval  $\Delta\theta = [\theta_{min}, \theta_{max}]$  that describes the estimated values of  $\theta$ . Lower and upper bounds of  $\theta$  can be found by solving two optimisation **problems**.

$$\theta_{min} = \inf_{z_i \in \Delta z_i, Z_0 \in \Delta Z_0} \arg \max_{\theta} L(\theta) \tag{28}$$

$$\theta_{max} = \sup_{z_i \in \Delta z_i, Z_0 \in \Delta Z_0} \arg \max_{\theta} L(\theta), \tag{29}$$

where  $L(\theta)$  is the log-likelihood function given by

$$L(\theta) = \sum_{i=1}^m \log f(z_i; \theta) + (M - m)S(Z_0; \theta). \tag{30}$$

The optimisation problem defined by (28) – (29) may be, in a general case, difficult, as the interval computations for non-linear functions are usually time consuming. However, in some practical cases the optimisation problem may be significantly simplified. In the case of the exponential distribution the lower and upper bound for the hazard rate  $\lambda$  are given by simple formulae

$$\lambda_{min} = \frac{m}{\sum_{i=1}^m z_{i,max} + (M - m)Z_{0,max}} \tag{31}$$

$$\lambda_{max} = \frac{m}{\sum_{i=1}^m z_{i,min} + (M - m)Z_{0,min}}. \tag{32}$$

Unfortunately, in the case of the Weibull distribution the interval for the possible estimated values of the shape parameter  $\theta$  cannot be calculated using separately lower and upper bounds for observed lifetimes and censoring times. Only the bounds for the scale parameter (or its reciprocal) can be calculated in such a way. In general, simple computations are possible only then if a lifetime distribution is of a location-scale type. In such a case, the bounds for a location parameter can be calculated using lower and upper bounds for lifetimes and censoring times separately.

Let us consider now another relatively simple example of a practical application of the interval approach in the analysis of reliability field data. In section 2.2 of this paper we presented a mathematical model of lifetimes when the probability distribution of these random variables depends also on certain covariates, which describe usage conditions. These conditions may be described by a vector of covariates  $z$ , and the dependence of lifetimes on these covariates may be described by different mathematical models. Assume now, that this dependence is described by the proportional hazard model defined by equations (7) – (9). In this model probability distribution of lifetimes depends on the values of covariates via  $z\beta = z_1\beta_1 + \dots + z_q\beta_q$ , where  $\beta_l$ 's are unknown regression coefficients. Estimation of these coefficients in the proportional hazard model was proposed by Cox [5], and is briefly presented in section 2.2.

In case of reliability field experiments each investigated unit may be used in different conditions. Theoretically, these conditions may be defined quite precisely, and described by real numbers. However, in practice it is much more convenient to describe usage conditions by categorical variables. In such a case each covariate  $z_j, j=1, \dots, p$  may adopt only a finite number of possible values  $z_{j,l}, j=1, \dots, p; l=1, \dots, n_j$ . If the set of these values can be identified for each of  $k$  failed units we can find the estimators of  $\theta$  by solving equations (19). However, in certain circumstances the users may face difficulties with a precise identification of the values of covariates  $z$ . Let us suppose, for example, that exploitation conditions vary in time, and it is

not obvious whether these conditions should be labelled as moderate or severe. In such situation the necessity to choose only one value of the covariate that describes the severity of exploitation conditions may distort a final reliability analysis. Introduction of another probabilistic model for the description of this situation may be too difficult from a practical point of view. Therefore, it seems to be much more convenient to use a set-valued description of the considered covariates. In case of covariates described by real numbers we can directly use the notation introduced previously, i.e.  $\Delta z_j = [z_{j,min}, z_{j,max}]$ ,  $j = 1, \dots, p$ . However, we also can use this notation in case of ordered categorical data. Let  $\square$  be a multivariate interval that describes the estimated values of the regression coefficients  $\beta$  in the presence of interval data  $\Delta z_i$ ,  $i = 1, \dots, m$ , where  $m$  is the number of observed failures. The lower and upper bounds for  $\square$  can be found by solving the optimisation problems

$$\beta_{min} = \inf_{z_i \in \Delta z_i} \arg \max_{\beta} L(\beta) \tag{33}$$

$$\beta_{max} = \sup_{z_i \in \Delta z_i} \arg \max_{\beta} L(\beta) , \tag{34}$$

where  $L(\beta)$  is the log-likelihood function given by (18).

The solution of (33) – (34) is, in general, difficult. However, in many cases the dependence of reliability upon covariates has a monotonic nature. In this case the lower and upper bounds of  $\square$  defined by (33) – (34) may be found using appropriately chosen (depending on the direction of the dependence) boundary values of  $\Delta z_i$ ,  $i = 1, \dots, m$ .

The limited volume of this paper allows us to present only a general description of relatively simple models for the analysis of reliability field data. These models are more complicated than the simplest lifetime models, but are applicable in such cases when a proper probabilistic analysis of reliability field data is either very difficult or even impossible. In order to overcome these problems we have to deal with some information of subjective nature. This is the price we have to pay if we want to solve more realistic problems.

#### 4. Statistical analysis of reliability field data with imprecise fuzzy information

In the previous section we considered the case when the information which is necessary for a proper evaluation of reliability in terms of the theory of probability and mathematical statistics may be incomplete and imprecise. Our lack of full information we represented in terms of intervals describing the quantities of interest. Representation of uncertainty by intervals has its origins in the theory of measurement. If no additional information is present, this methodology allows the calculation of the bounds for reliability characteristics of interest. These bounds may be interpreted as “the worse” and “the best” possible values which take into account any type of variability of imprecisely or partially known values of field lifetime data. However, one can argue that this type of representation of uncertainty may not reflect the complexity of available information. For example, let us suppose that the daily usage rate of certain equipment is reported by its user as “about five hours a day”. From further inquiry one may get information that it means “between four and six hours a day”. Note, that this information does not tell anything about the way the usage rate varies in time. Therefore, the representation of uncertainty in a form of an interval seems to be quite appropriate. On the other hand, it is easy to note that the original information, “about five hours a day”, carries additional information. One may believe that the real usage rate is more often closer to five hours than to any other number of hours. This still vague information, which does not allow building any probability distribution, may be described formally using the theory of fuzzy sets introduced by Lotfi A. Zadeh[25].

Fuzzy sets are the generalization of ordinary sets. In order to define a fuzzy set we have to specify a so called *universe of discourse*  $X$ , i.e. an ordinary set that contains all elements that are relevant for the description of a vaguely defined (or described) object. In the considered in this paper reliability context it might be a set (or a subset) of positive real numbers, when we describe the usage rate, a set of integers, when we describe a partially known number of units on test, or a set of labels, when we describe the severity of working conditions. A *membership function*  $\mu_A : X \rightarrow [0,1]$  such that  $\mu_A(x)$  tells us to which degree an element  $x \in X$  belongs to the fuzzy set  $A$ , is the second part of the definition of a fuzzy set. Thus, a fuzzy set  $A$  in a universe of discourse  $X$  is a set of pairs

$$A = \{\mu_A(x), x\} \tag{35}$$

This formalism is very useful for the description of vague and imprecise concepts, as the value of the membership function  $\mu_A(x) \in [0,1]$  describes our degree of belief that the value  $x$  describes the considered concept.

Each fuzzy set has a unique representation in terms of so called  $\alpha$ -cuts, or  $\alpha$ -level sets. The ordinary (non-fuzzy) set  $A_\alpha = \{x \in X : \mu_A(x) \geq \alpha\}$ , for each  $\alpha \in (0,1]$ , is called the  $\alpha$ -cut of the fuzzy set  $A$ , and the set of all  $\alpha$ -cuts uniquely defines this fuzzy sets.

When the universe of discourse is represented by the set of real numbers we can generalize the concept of a real number and define a *fuzzy number*. The fuzzy subset  $A$  of the real line  $R$ , with the membership function  $\mu_A : R \rightarrow [0,1]$  is a fuzzy number iff

- a)  $A$  is normal, i.e. there exists an element  $x_0$  such that  $\mu(x_0) = 1$ ;
- b)  $A$  is fuzzy convex, i.e.  $\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geq \mu_A(x_1) \wedge \mu_A(x_2)$ ,  $\forall x_1, x_2 \in R, \forall \lambda \in [0,1]$ ;
- c)  $\mu_A$  is upper semi-continuous;
- d)  $\text{supp } A$  is bounded.

From the definition given above one can easily find that for any fuzzy number  $A$  there exist four real numbers  $a_1, a_2, a_3, a_4$  and two functions: non-decreasing function  $\eta_A : R \rightarrow [0,1]$ , and non-increasing function  $\xi_A : R \rightarrow [0,1]$ , such that the membership function  $\mu_A$  is given by

$$\mu_A(x) = \begin{cases} 0 & \text{if } x < a_1 \\ \eta_A(x) & \text{if } a_1 \leq x < a_2 \\ 1 & \text{if } a_2 \leq x < a_3 \\ \xi_A(x) & \text{if } a_3 \leq x < a_4 \\ 0 & \text{if } a_4 < x \end{cases} \tag{36}$$

Functions  $\eta_A$  and  $\xi_A$  are called the left side and the right side of a fuzzy number  $A$ , respectively. A special, and very useful in practice, case of a general fuzzy number is a trapezoidal fuzzy number defined by the following membership function

$$\mu_x(x) = \begin{cases} 0 & \text{if } x < x_1 \\ (x - x_1)/(x_2 - x_1) & \text{if } x_1 \leq x < x_2 \\ 1 & \text{if } x_2 \leq x < x_3 \\ (x_4 - x)/(x_4 - x_3) & \text{if } x_3 \leq x < x_4 \\ 0 & \text{if } x_4 \leq x \end{cases} \tag{37}$$

Note, that real-valued intervals considered in the previous section of this paper can be looked upon as trapezoidal fuzzy numbers for which  $x_1 = x_2 = x_{min}$ , and  $x_3 = x_4 = x_{max}$ .

Membership functions of fuzzy numbers that are defined as functions of other fuzzy numbers may be calculated using the following *extension principle* introduced by Zadeh, and described in Dubois and Prade [6] as follows:

Let  $X$  be a Cartesian product of universe  $X = X_1 \times X_2 \times \dots \times X_r$ , and  $A_1, \dots, A_r$  be  $r$  fuzzy sets in  $X_1, \dots, X_r$ , respectively. Let  $f$  be a mapping from  $X = X_1 \times X_2 \times \dots \times X_r$  to a universe  $Y$  such that  $y = f(x_1, \dots, x_r)$ . The extension principle allows us to induce from  $r$  fuzzy sets  $A_i$  a fuzzy set  $B$  on  $Y$  through  $f$  such that

$$\mu_B(y) = \sup_{x_1, \dots, x_r: y=f(x_1, \dots, x_r)} \min[\mu_{A_1}(x_1), \dots, \mu_{A_r}(x_r)] \quad (38)$$

$$\mu_B(y) = 0 \text{ if } f^{-1}(y) = \emptyset \quad (39)$$

One can prove, see e.g. books by Dubois and Prade [6] or by Zimmermann [27], that the application of the extension principle is equivalent to the application of the interval arithmetics on  $\square$ -cuts.

Fuzzy sets, and their special instances – fuzzy numbers, have been applied in solving different reliability problems. An extensive overview of these applications can be found in Hryniewicz [8]. If we want to apply this approach to the analysis of field lifetime tests we can directly apply the results presented in the previous section. In order to do so let us notice that the calculations presented in that section are exactly the same as the calculations that should be done for given  $\square$ -cuts representing fuzzy data.

## 5. Conclusion

Probabilistic models that have been proposed for the description of field lifetime data, and are relatively easy to be applied in practice, usually do not describe all the aspects of this type of data. If we want to build models, which better describe reality, then immediately these models become very complicated. Moreover, additional assumptions have to be made in order to describe complex phenomena characteristic for this problem. In this paper we have proposed an alternative but only approximate approach where unknown values of model parameters are represented in terms of intervals. By applying the interval arithmetics we can calculate the bounds on the values of respective reliability characteristics. If additional but still imprecise information is available we propose to generalize the interval-valued calculations to fuzzy-valued ones. The results of these calculations can be interpreted as *possibility distributions* in the sense of Zadeh [26], defined on sets of possible values of vague quantities. It has to be stressed, however, that if appropriate probabilistic information is available it should not be replaced with the fuzzy one. Fuzziness in our models does not replace randomness, but supplements it if we have to use imprecisely perceived notions or vague statistical data.

## References

- [1] Cohen, C. A. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics*, 1, 217 – 237.
- [2] Cohen, C. A. (1991). *Truncated and censored samples: theory and applications*. Marcel Dekker, New York.
- [3] Coit, D. W. & Dey, K. A. (1999). Analysis of grouped data from field-failure reporting systems. *Reliability Engineering and System Safety*, 65, 95 – 101.
- [4] Coit, D. W. & Jin, T. (2000). Gamma distribution parameter estimation for field reliability data with missing failure times. *IEEE Transactions*, 32, 1161 – 1166.
- [5] Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, ser.B*, 34, 187 – 202.
- [6] Dubois, D. & Prade, H. (1980). *Fuzzy Sets and Systems. Theory and Applications*. Academic Press, New York.
- [7] Duchesne, T. & Lawless, J. F. (2000). Alternative time scales and failure time models. *Lifetime Data Analysis*, 6, 157-179.

- 
- [8] Hryniewicz, O. (2007). Fuzzy sets in the evaluation of reliability. In: *Computational Intelligence in Reliability Engineering. New Metaheuristics, Neural and Fuzzy Techniques in Reliability*, Levitin, G. (Ed.), Springer, Berlin., 363 – 386.
- [9] Hu, J. X. & Lawless, J. F. (1996a). Estimation from truncated lifetime data with supplementary information on covariates and censoring times. *Biometrika*, 83(4), 747-761.
- [10] Hu, J. X. & Lawless, J. F. (1996b). Estimation of rate and mean functions from truncated recurrent event data. *Journal of the American Statistical Association*, 91, 300-310.
- [11] Hu, J. X., Lawless, J. F. & Suzuki, K. (1998). Nonparametric estimation of a lifetime distribution when censoring times are missing. *Technometrics*, 40, 3-13.
- [12] Jung, M. & Bai, D. S. (2007). Analysis of field data under two-dimensional warranty. *Reliability Engineering and System Safety*, 92, 135-143.
- [13] Kalbfleisch, J. D. & Lawless, J. F. (1988). Estimation of reliability in field-performance studies. *Technometrics*, 30, 365-388.
- [14] Kalbfleisch, J. D., Lawless, J. F. & Robinson, J.A. (1991). Methods for the analysis and prediction of warranty claims. *Technometrics*, 33, 273-285.
- [15] Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53, 457 – 481.
- [16] Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York.
- [17] Lawless, J. F. (1998). Statistical analysis of product warranty data. *International Statistical Review*, 66, 41-60.
- [18] Lawless, J. F., Hu, J. & Cao, J. (1995). Methods for the estimation of a lifetime distributions and rates from automotive warranty data. *Lifetime Data Analysis*, 1, 227-240.
- [19] Lin, D. K. J., Usher, J. S. & Guess, F.M (1996). Bayes estimation of component-reliability from masked system-life data. *IEEE Transactions on Reliability*, 45, 233 – 237.
- [20] Oh, Y. S. & Bai, D. S. (2001). Field data analyses with after-warranty failure data. *Reliability Engineering and System Safety*, 72, 1-8.
- [21] Rai, B. & Singh, N. (2003). Hazard rate estimation from incomplete and unclean warranty data. *Reliability Engineering and System Safety*, 81, 79 – 82.
- [22] Suzuki, K. (1985). Estimation of lifetime parameters from incomplete field data. *Technometrics*, 27, 263-272.
- [23] Suzuki, K. (1985). Nonparametric estimation of lifetime distributions from a record of failures and follow-ups. *Journal of the American Statistical Association*, 80, 68-72.
- [24] Usher, J. S. (1996). Weibull component reliability-prediction in the presence of masked data. *IEEE Transactions on Reliability*, 45, 229-232.
- [25] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338 – 353.
- [26] Zadeh, L.A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3 – 28.
- [27] Zimmermann, H. J. (1996). *Fuzzy Set Theory and its Applications* (Third Edition), Kluwer, Boston.
-