Estimation of the mean lifetime from vague data

Przemysław Grzegorzewski
Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
pgrzeg@ibspan.waw.pl

Abstract

In this paper we suggest how to estimate a mean lifetime in the presence of vague data, especially imprecise number of failures.

Keywords: Lifetime estimation, vague data.

1 Introduction

One of the most important problem of reliability analysis is to estimate the *mean lifetime* of the item under study. In technical applications this parameter is also called *mean time to failure* ($MTTF$) and is often included in the specification of a product. For example, producers are interested whether this time is sufficiently large, as large $MTTF$ allows them to extend a warranty time. Classical estimators require precise data obtained from strictly controlled reliability tests (for example, those performed by a producer at his laboratory). In such a case a failure should be precisely defined, and all tested items should be continuously monitored. However, in real situation these requirements might not be fulfilled. In the extreme case, the reliability data come from users whose reports are expressed in a vague way. The vagueness of the data has many different sources: it might be caused by subjective and imprecise perception of failures by a user, by imprecise records of reliability data, by imprecise records of the rate of usage, etc. Therefore we require both tools appropriate for modelling vague data and suitable statistical methodology to handle these data.

Grzegorzewski and Hryniewicz [1] considered the generalization of the exponential model which admit vagueness in lifetimes or censoring times but requires precise information about the number of observed failures (i.e. one knows whether given item has failed or it has survived). In their paper fuzzy sets were used for modelling vagueness of the lifetimes. However, sometimes we face situations when the number of observed failures is also vague. For example, it may be due to imprecise definition of the failure. We can also consider partial failures or information about the scale of the failure expressed by colloquial words. Hence in the present paper we suggest another generalization of the classical exponential model. We consider not only fuzzy lifetimes but situations in which the number of failures is fuzzy too.

2 Classical approach

The mean lifetime may be efficiently estimated by the sample average from the sample of the lifetimes $X_1, X_2, \ldots, X_n$ of n tested items. However, in the majority of practical cases the lifetimes of all tested items are not known, as the test is usually terminated before the failure of all items. It means that exact lifetimes are known for only a portion of the items under study, while remaining lifetimes are known only to exceed certain values. This feature of lifetime data is called *censoring*. More formally, a fixed censoring time $Z_i > 0$, $i = 1, \ldots, n$ is associated with each item. We observe X_i only if $X_i \leq Z_i$. Therefore our lifetime data consist of pairs $(T_1, Y_1), \ldots, (T_n, Y_n)$, where

$$T_i = \min\{X_i, Z_i\}, \qquad (1)$$

$$Y_i = \begin{cases} 1, & \text{if} \quad X_i = T_i \\ 0, & \text{if} \quad X_i = Z_i. \end{cases} \tag{2}$$

Numerous parametric models are used in the lifetime data analysis. Among them the most widely used are the exponential, Weibull, gamma and lognormal distribution models. Historically, the exponential model was the first lifetime model extensively developed and widely used in many areas of lifetime analysis: from studies on the lifetimes of various types of manufactured items to research involving survival or remission times in chronic diseases. In this model the lifetime is described by the probability density function

$$f(t) = \begin{cases} \frac{1}{\theta} e^{-\frac{t}{\theta}}, & \text{if} \quad t > 0 \\ 0, & \text{if} \quad t \leq 0. \end{cases} \tag{3}$$

where $\theta > 0$ is the mean lifetime. It is worth to notice that the hazard function in the considered model is constant. Although this assumption is very restrictive, the exponential model is still frequently used in practice because of two important features: its parameter θ is easily estimated and for lifetimes described by a probability distribution with increasing hazard it gives conservative approximation for the mean lifetime. Thus in this paper we assume that the exponential distribution model is the mathematical model which describes lifetimes of tested items. Note that

$$T = \sum_{i=1}^{n} T_i = \sum_{i \in O} X_i + \sum_{i \in C} Z_i \tag{4}$$

is the total survival time, where O and C denote the sets of items for whom exact lifetimes are observed and censored, respectively. Moreover, let

$$r = \sum_{i=1}^{n} Y_i \tag{5}$$

denote the number of observed failures. In the considered exponential model the statistic (r, T) is minimally sufficient statistic for θ and the maximum likelihood estimator of the mean lifetime θ is (assuming $r > 0$)

$$\widehat{\theta} = \frac{T}{r}. \tag{6}$$

It can be shown that statistic $\frac{2r\widehat{\theta}}{\theta}$ is approximately chi-square distributed with $2r + 1$ degrees of freedom. It appears that this approximation can be used for constructing satisfactory confidence intervals even for quite small sample sizes (e.g., see [2]). For example, the one-sided confidence interval with upper limit for θ on the confidence level $1 - \delta$ is

$$\left(0, \frac{2T}{\chi^2_{2r+1,\delta}}\right]. \tag{7}$$

where $\chi^2_{2r+1,\delta}$ is the quantile of order δ of the chi-square distribution with $2r+1$ degrees of freedom.

3 Vague data

3.1 Fuzzy lifetimes

Now suppose that the lifetimes are no longer crisp but vague. A generalization of the exponential model which admit vagueness in lifetimes was considered by Grzegorzewski and Hryniewicz [1]. They model imprecise lifetimes by fuzzy numbers. Thus applying the extension principle to (4) we get a fuzzy total survival lifetime $\widetilde{T}$ (which is also a fuzzy number) and a fuzzy estimator of $MTTF$ in the presence of vague data is given by

$$\widetilde{\theta} = \frac{\widetilde{T}}{r}. \tag{8}$$

For more details and the discussion on fuzzy confidence intervals we refer the reader to [1].

3.2 Up and down states

Very often, in practice, we deal with not only critical failures but also with non-critical failures, usually described in a common language. For example, one is anxious because of the sudden strange noise in the car. However, he can still drive this car. Such situation corresponds to a failure which is not critical (at least in this moment).

In order to take into account such non-critical failures let us describe the state of each observed item at the time L_i. Let G denote a set of all items which are capable at their censoring times L_i. Therefore we can assign to each item $i = 1, \ldots, n$ its degree of belongingness $g_i = \mu_G(i)$ to G, where $g_i \in [0, 1]$. If $g_i = 1$ then given item haven't had any failure before L_i and it works perfectly at the time moment L_i, while for $g_i = 0$ it have had a critical failure before or at time moment L_i. If $g_i \in (0, 1)$ then item under study neither works perfectly nor is completely failed. So G is a fuzzy set with a finite support. Alternatively, one can consider a set D of faulty items and, of course, the degree of belongingness to D is equal to $d_i = \mu_D(i) = 1 - g_i$. Further on we'll call g_i and d_i as a degree of the up state and down state, respectively. Now we have to find a fuzzy counterpart of the number of observed failures r. Below we propose several methods for failure counting. Depending on the output we can divide them into two groups: crisp or fuzzy methods.

3.3 Crisp failure counting

Let $g_1, \ldots, g_n$ $(d_1, \ldots, d_n)$ denote degrees of up states (down states) of all items under test. The most natural way for counting failures is to consider critical failures only or to treat all kinds of failures similarly. This two approaches correspond to optimistic (or liberal) and pessimistic (conservative) viewpoints, respectively. Thus the number of failures observed in accordance with the optimistic viewpoint is

$$R_{opt} = \sum_{i=1}^{n} I(d_i = 1) = n - \sum_{i=1}^{n} I(g_i > 0), \quad (9)$$

where I is an indicator function, while the number of failures obtained in accordance with the pessimistic viewpoint is

$$R_{pes} = n - \sum_{i=1}^{n} I(g_i = 1) = \sum_{i=1}^{n} I(d_i > 0). \quad (10)$$

More generally, one can take into account only failures with some degree of down state (up state). Then we get

$$R_\xi = \sum_{i=1}^{n} I(d_i > \xi) = n - \sum_{i=1}^{n} I(g_i \geq 1 - \xi), \quad (11)$$

where $\xi \in (0,1)$. Measures (9) – (11) are crisp, since $R_{opt}, R_{pes}, R_\xi \in N \cup \{0\}$. It is clear that $R_{opt} \leq R_\xi \leq R_{pes}$ for each $\xi \in (0,1)$. The methods of failure counting described above are in some sense reductive. Actually, they abandon the whole information on particular degrees of up states or down states and utilize only part of that information – whether these degrees exceed a given level. However, sometimes it would be useful to take into account all accessible information. Then a following method for failure counting might be used

$$R^c = \sum_{i=1}^{n} d_i = |D| = n - \sum_{i=1}^{n} g_i = n - |G|, \quad (12)$$

where $|D|$ and $|G|$ denotes the cardinality of fuzzy set D and G, respectively.

3.4 Fuzzy failure counting

The basic advantage of the methods for counting failures given above is that they are easy to handle, since their output are crisp. Unfortunately, such approach doesn't reflect reality very well, especially that the test results are often non-precise but vague. Moreover, the requirements are also sometimes vague. It seems that the best way to summarize fuzzy descriptions of the test results is to use fuzzy failure counting measures. We consider observed degrees of down states and count the number of failures we would get if the rejection limit would be fixed on each degree of down state (naturally, the lower is the rejection limit, more failures we observe). Thus we get a following fuzzy measure of the number of failures:

$$R_{opt}^f = |D|_f, \quad (13)$$

where $|D|_f$ denotes fuzzy cardinality of fuzzy set D. We may also start from up states. Therefore

$$R_{pes}^f = n - |G|_f, \quad (14)$$

where $|G|_f$ denotes fuzzy cardinality of fuzzy set G. However, contrary to the crisp counting $|D|_f \neq n - |G|_f$. It is seen that such fuzzy measure of the number of failures is a finite fuzzy set. It is also a normal fuzzy set (since we assume, as in the classical approach, that there exist at least one critical failure).

4 MTTF estimation

By the extension principle, we may define a fuzzy estimator of the mean lifetime $\tilde{\tilde{\theta}}$ in the presence of vague data as

$$\tilde{\tilde{\theta}} = \frac{\tilde{T}}{\tilde{R}}, \quad (15)$$

where $\widetilde{T}$ is the fuzzy total survival time and R denotes the number of vaguely defined failures (crisp or fuzzy). Actually (15) provides a family of estimators that depend on the choice of R. Namely, one can choose his preferred measure of failure counting $R \in \{R_{opt}, R_{pes}, R_\xi, R^c, R^f_{opt}, R^f_{pes}\}$ and get, as a result, an estimator $\widetilde{\widetilde{\theta}}_{opt}$, $\widetilde{\widetilde{\theta}}_{pes}$, $\widetilde{\widetilde{\theta}}_\xi$ $(0 < \xi < 1)$, $\widetilde{\widetilde{\theta}}^c$, $\widetilde{\widetilde{\theta}}^f_{opt}$, $\widetilde{\widetilde{\theta}}^f_{pes}$, respectively. It is not difficult to prove that $\widetilde{\widetilde{\theta}}$ is a fuzzy number.

Besides finding the fuzzy estimator of the mean lifetime we can also construct fuzzy confidence intervals for $MTTF$.

Theorem Assume $\widetilde{T}$ denotes the fuzzy total survival time with a membership function μ_T and R is the number of observed failures. Moreover, let $\widetilde{T}^U_\alpha = \sup\{t \in \mathcal{R} : \mu_T(t) \geq \alpha\}$ and let $\chi^2_{df,\delta}$ denote the quantile of order $\delta \in (0,1)$ from the chi-square distribution with df degrees of freedom, where either $df = 2R + 1$ for $R \in \{R_{opt}, R_{pes}, R^c, R_\xi\}$, $\xi \in (0,1)$ or $df = 2R^U_\alpha + 1$ for $R \in \{R^f_{opt}, R^f_{pes}\}$, $R^U_\alpha = \max\{x \in \mathcal{N} : \mu_R(x) \geq \alpha\}$. Then a fuzzy set $\Pi = \Pi(\widetilde{T})$ given by a family of its $\alpha-$cuts

$$\Pi_\alpha(\widetilde{T}) = \left(0, \frac{2\widetilde{T}^U_\alpha}{\chi^2_{df,\delta}}\right], \quad \alpha \in (0,1] \qquad (16)$$

is the one-sided confidence interval with upper limit for the mean lifetime θ on the confidence level $1 - \delta$.

5 Conclusions

We have shown how to estimate the mean lifetime in the presence of vague lifetimes and vague number of observed failures. The suggested methodology can also be used for models other than exponential.

References

[1] Grzegorzewski P., Hryniewicz O. (1999). Lifetime Tests for Vague Data. In: *Computing with Words in Information/Intelligent Systems*, L.A. Zadeh, J. Kacprzyk (eds.), Part 2. Physica-Verlag, Heidelberg, pp. 176–193.

[2] Lawless J.F. (1982). *Statistical Models and Methods for Lifetime Data.* Wiley, New York.