

On the Estimation of Three Parameters Lognormal Distribution Based on Fuzzy Life Time Data (Anggaran Taburan Lognormal Tiga Parameter Berdasarkan Data Masa Hayat yang Kabur)

MUHAMMAD SHAFIQ*, ALAMGIR & MUHAMMAD ATIF

ABSTRACT

Countless statistical tools are available to extract information from data. Life time modeling is considered as one of the most prominent fields of statistics, which is evident from the developments made in this field in the last few decades. Almost every statistic for life time analysis is based on precise life time observations, however, life time is not a precise measurement but more or less fuzzy. Therefore, in addition to classical statistical tools, fuzzy number approaches to describe life time data are more suitable. In order to incorporate fuzziness of the observations, fuzzy estimators for the three parameter lognormal distribution were suggested. The proposed estimators cover stochastic variation as well as fuzziness of the observations.

Keywords: Characterizing function; fuzzy number; life time; non-precise data

ABSTRAK

Terdapat banyak perkakasan statistik tersedia untuk mengekstrak maklumat daripada data. Pemodelan masa hayat dianggap sebagai salah satu bidang statistik yang paling menonjol. Ini jelas daripada pembangunan bidang ini sejak beberapa dekad yang lalu. Hampir setiap statistik untuk analisis masa hayat adalah berasaskan pemerhatian masa hayat yang tepat, walau bagaimanapun, masa hayat bukanlah suatu pengukuran yang tepat tetapi lebih atau kurang kabur. Oleh itu, sebagai tambahan kepada perkakas statistik klasik, pendekatan nombor kabur untuk menggambarkan data masa hayat adalah lebih sesuai. Dalam usaha untuk menggabungkan kekaburan daripada pemerhatian, penganggaran kabur untuk taburan tiga parameter lognormal telah dicadangkan. Penganggaran yang dicadangkan meliputi kelainan stokastik serta kekaburan daripada pemerhatian.

Kata kunci: Data tidak tepat; fungsi pencirian; masa hayat; nombor kabur

INTRODUCTION

Statistical decisions are based on data. Data are usually presented in the form of precise numbers or vectors or results of functions. In many situations, the observation cannot be recorded as precise number because of continuous nature or some irregular nature, e.g. wave nature of the water level (Wu 2004).

About the precise measurements (Barbato et al. 2013) mentioned that in the modern technology measurements the word ‘exact’ or ‘equal’ needs to be banned, because the characteristic exact is not possible to attain in reality. According to Viertl (2009), all the measurements obtained from continuous real variables cannot be precise numbers but are more or less imprecise. One should keep in mind that imprecision of single observation is different from stochastic variability and measurement errors, this imprecision of the measurement is called *fuzziness*.

In order to integrate fuzziness in decision making the idea of *fuzzy sets* was first presented by Zadeh in 1965. Most classical statistical tools are based on precise observations without considering fuzziness of the observations.

Life time is also of continuous nature, therefore, the measurement obtained on life time cannot be a precise number but rather fuzzy. Therefore, the analysis techniques related to life time data are required to be generalized in such a way that fuzziness of the observations is integrated in the estimation process.

ELEMENTS OF FUZZY SET THEORY

In classical set theory to represent whether an element t is in a subset A of a universal set M , a two valued characteristic function called *indicator function* is used as mentioned in (1):

$$I_A(t) = \begin{cases} 1 & \text{if } t \in A \\ 0 & \text{if } t \notin A \end{cases} \forall t \in M. \quad (1)$$

Fuzzy set theory is the generalization of classical set theory, therefore, the indicator function mentioned in (1) is generalized to the so-called *membership function* μ_{A^*} of a fuzzy subset A^* of M , i.e. (2):

$$\mu_{A^*}(t) \begin{cases} 1 & \text{if } t \in \text{core}(A^*) \\ \delta \in [0,1] & \text{if } t \text{ belongs to } A^* \\ & \text{to some degree } \delta \\ 0 & \text{if } t \notin A^* \end{cases} \forall t \in M, \quad (2)$$

where the core of a fuzzy subset A^* is the set of all points t in M such that $\mu_{A^*}(t) = 1$.

The membership function maps the elements from the universal set M to the interval (Szeliga 2004).

FUZZY NUMBERS

According to Viertl (2011), a so-called *fuzzy number* t^* is a special fuzzy subset of \mathbb{R} it is determined by a real function of one real variable, the so-called *characterizing function* denoted by $\xi(\cdot)$, obeying the following three conditions:

$$\xi: \mathbb{R} \rightarrow [0, 1].$$

The characterizing function $\xi(\cdot)$ has bounded support, i.e. $\text{supp}[\xi(\cdot)] := \{t \in \mathbb{R} : \xi(t) > 0\} \subseteq [a, b]$; and

The set $C_\delta(t^*) := \{t \in \mathbb{R} : \xi(t) \geq \delta\}$ is called δ -cut. For a fuzzy number it is a finite union of non-empty compact intervals, i.e. $C_\delta(t^*) = \cup_{j=1}^{k_\delta} [a_{\delta,j}, b_{\delta,j}] \neq \emptyset \forall \delta \in [0,1]$ where k_δ represents the number of compact intervals in the δ -cut at level δ .

A so-called *fuzzy interval* is a special form of a fuzzy number, for this all δ -cuts are non-empty closed bounded intervals.

LEMMA

For a fuzzy number t^* having characterizing function $\xi(\cdot)$ the following lemma holds true:

$$\xi(t) = \max\{\delta \cdot I_{C_\delta(t^*)}(t) : \delta \in [0,1]\} \forall t \in \mathbb{R}.$$

For proof compare Viertl (2011).

Remark For a nested family of finite unions of compact intervals, i.e. $(A_\delta; \delta \in [0,1])$ obeying it is important to note that not all families of nested finite unions of compact intervals are the δ -cuts of a fuzzy number. But the following construction lemma holds:

CONSTRUCTION LEMMA

Let $A_g = \cup_{j=1}^{k_g} [a_{g,j}, b_{g,j}]$ be a nested family of non-empty subset of \mathbb{R} . Then the characterizing function of the generated fuzzy number is defined by

$$\xi(t) = \sup\{\delta \cdot I_{C_\delta(t^*)}(t) : \delta \in [0,1]\}$$

$$\forall t \in \mathbb{R}.$$

For details see Viertl and Hareter (2006).

EXTENSION PRINCIPLE

Let M and N be any spaces and G be an arbitrary function $G : M \rightarrow N$, then the extension principle is the generalization of the function G for a fuzzy argument value a^* in the set M .

Let a fuzzy element a^* having membership function $\mu : M \rightarrow [0,1]$, then the fuzzy value y^* which is denoted by $y^* = G(a^*)$ is the fuzzy element in space N whose membership function $\vartheta(\cdot)$ is defined by:

$$\vartheta(y) = \begin{cases} \sup\{\mu(a) : a \in M, G(a) = y\} & \text{if } \exists a : G(a) = y \\ 0 & \text{if } \nexists a : G(a) = y \end{cases}$$

$$\forall y \in N.$$

See Klir and Yuan (1995).

FUZZY VECTORS

A so-called *fuzzy vector* \underline{t}^* is a fuzzy subset of \mathbb{R}^n which is determined by a real function of n real variables t_1, t_2, \dots, t_n called *vector-characterizing function*, and is denoted by $\zeta(\cdot, \dots, \cdot)$, obeying the following three conditions:

$$\zeta: \mathbb{R}^n \rightarrow [0, 1];$$

The support of $\zeta(\cdot, \dots, \cdot)$ is a bounded set; and

For all $\delta \in [0, 1]$ the so-called δ -cut $C_\delta(\underline{t}^*) = \{\underline{t} \in \mathbb{R}^n : \zeta(\underline{t}) \geq \delta\}$ is non-empty, bounded, and a finite union of simply connected and closed sets.

If all δ -cuts of a n -dimensional sets, then the corresponding n -dimensional fuzzy vector is called *-dimensional fuzzy interval*.

THEOREM

For any continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, and fuzzy n -dimensional fuzzy interval \underline{t}^* the following holds true:

$$C_\delta(f(\underline{t}^*)) = \left[\min_{\underline{t} \in C_\delta(\underline{t}^*)} f(\underline{t}), \max_{\underline{t} \in C_\delta(\underline{t}^*)} f(\underline{t}) \right] \forall \delta \in [0,1]$$

For proof see Viertl (2011).

LIFE TIME DATA ANALYSIS

If a random variable T denotes life time, it will have an *observation space* $M_T \subseteq [0, \infty)$. In case of random sample t_1, t_2, \dots, t_n from T , each element of the sample is an element of the observation space, and the sample (t_1, t_2, \dots, t_n) is an element of the Cartesian product of n copies of M_T , i.e. $M_T \times M_T \times \dots \times M_T$ called the *sample space*, and is denoted by M_T^n .

But in case of fuzzy sample $(t_1^*, t_2^*, t_3^*, \dots, t_n^*)$ each element of the sample, i.e. $t_i^* i = 1(1)n$ is a fuzzy element of the

observation space M_T , but the sample is not a fuzzy element of the sample space M_T^n .

We need to attain a fuzzy element of the sample space, and this can be done by use of *triangular norms*, usually called *t-norms*.

For the vector-characterizing function $\zeta(\cdot, \dots, \cdot)$ of the *combined fuzzy sample* \underline{t}^* , the *minimum t-norm*, is applied, i.e.

$$C_\delta[\zeta(\cdot, \dots, \cdot)] = \min[\zeta_1(t_1), \zeta_2(t_2), \dots, \zeta_n(t_n)]$$

$$\forall (t_1, t_2, \dots, t_n) \in \mathbb{R}^n$$

and the δ -cuts of the combined fuzzy sample \underline{t}^* are obtained as the Cartesian product of the δ -cuts, i.e. $C_\delta[\zeta(\cdot, \dots, \cdot)] = \times_{i=1}^n C_\delta[\zeta_i(\cdot)] \forall \delta \in (0, 1]$ (Viertl 2011).

Numerical data handling requires statistical techniques to provide a framework to summarize and to draw inference from the raw data.

Statistical tools related to *life time* data analysis started in the 20th century and comprehensively developed in the last few decades. Life time can be simply defined as ‘the waiting time till a specified event occurs’. The event of interest may be death, failure, divorce and change of place. The main aim of these analyses was to obtain mean life time, to predict survival probabilities and compare survival curves (Deshpande & Purohit 2005).

The Lognormal distribution is considered in renown distributions to model life time data especially for highly positive skewed data.

Consider the event of interest is death/failure, then T is a non-negative random variable denoting the waiting time until the failure/death of a unit. For guaranty time γ , $\ln(T - \gamma)$ follows a normal distribution with mean μ and variance σ^2 , and has lognormal distribution with parameters (γ, μ, σ^2) represented by its density,

$$f(t|\gamma, \mu, \sigma) = \frac{1}{(t-\gamma)\sigma\sqrt{2\pi}} \exp\left\{-\frac{[\ln(t-\gamma-\mu)]^2}{2\sigma^2}\right\},$$

where $\gamma < t < \infty, \sigma^2 > 0, -\infty < \mu < \infty$.

The corresponding maximum likelihood estimators of the parameters are,

$$\hat{\mu} = \frac{\sum_{i=1}^n \ln(t_i - \hat{\gamma})}{n}, \tag{3}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n [\ln(t_i - \hat{\gamma}) - \hat{\mu}]^2, \tag{4}$$

$$\sum_{i=1}^n \frac{1}{t_i - \gamma} + \frac{1}{\sigma^2} \sum_{i=1}^n \frac{\ln(t_i - \gamma) - \mu}{t_i - \gamma} = 0, \tag{5}$$

by solving (3)-(5) iteratively to get an estimate of γ .

An estimator can also be obtained by putting (3) and (4) in (5), resulting in

$$\left[\sum_{i=1}^n \frac{1}{t_i - \gamma} \right] \left[\sum_{i=1}^n \ln(t_i - \gamma) - \sum_{i=1}^n (\ln(t_i - \gamma))^2 + \frac{1}{n} (\ln(t_i - \gamma))^2 \right] - n \sum_{i=1}^n \frac{\ln(t_i - \gamma)}{t_i - \gamma} = 0. \tag{6}$$

By solving (6) one can get $\hat{\gamma}$.

For proof see Cohen and Whitten (1980).

Dealing with fuzzy life time data some work has been done like, fuzzy Bayesian estimation on lifetime data (Wu 2004), Bayesian reliability analysis for fuzzy lifetime data (Huang et al. 2006), reliability estimation based on fuzzy lifetime data (Viertl 2009) and reliability estimation in Rayleigh distribution based on fuzzy lifetime data (Pak et al. 2013), but still in most of the publications fuzziness is ignored.

In this study generalized parameter estimators for the three parameter lognormal distributions are proposed which are based on fuzzy life time observations.

GENERALIZED ESTIMATORS FOR THE THREE PARAMETER LOGNORMAL DISTRIBUTION

A so-called trapezoidal fuzzy number is denoted by $t^* = [m, s, l, r]$ and their characterizing function is defined by the following expression,

$$\xi(t) = \begin{cases} \frac{t-m+s+l}{l} & f \text{ or } m-s-l < t < m-s \\ 1 & f \text{ or } m-s < t < m+s \\ \frac{m+s+r-t}{r} & f \text{ or } m+s < t < m+s+r \\ 0 & otherwise \end{cases}$$

A characterizing function of a trapezoidal fuzzy number is depicted in Figure 1.

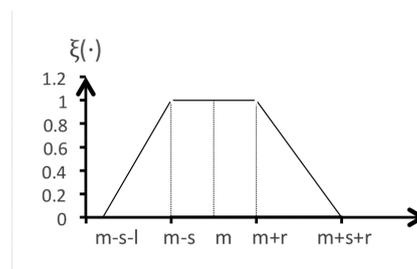


FIGURE 1. Characterizing function of trapezoidal fuzzy number

Let $(t_1^*, t_2^*, t_3^*, \dots, t_n^*)$ be fuzzy life time observations with characterizing functions $\xi_1(t_1), \xi_2(t_2), \dots, \xi_n(t_n)$, respectively.

The corresponding δ -cuts of the fuzzy life times are denoted by $C_\delta(t_i^*) = [t_{i,\delta}, \bar{t}_{i,\delta}] \forall \delta \in [0,1]$ where $t_{i,\delta}$ and $\bar{t}_{i,\delta}$ represent lower and upper ends of the δ -cuts of t_i^* , $i = 1(1)n$.

A sample of fuzzy life time observations with characterizing functions is depicted in Figure 2.

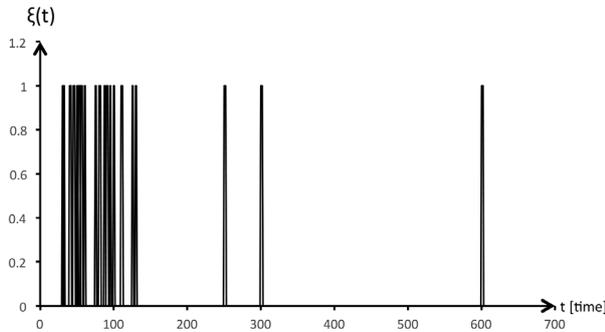


FIGURE 2. Characterizing functions of fuzzy life time observations

The generalized (fuzzy) estimators based on fuzzy life time observations are denoted by $\hat{\gamma}^*$, $\widehat{\sigma^2}^*$, and $\hat{\mu}^*$.

The corresponding δ -cuts of the fuzzy estimators are denoted by:

$$C_\delta(\hat{\gamma}^*) = [\underline{\gamma}_\delta, \bar{\gamma}_\delta] \quad \forall \delta \in [0,1]$$

$$C_\delta(\hat{\mu}^*) = [\underline{\mu}_\delta, \bar{\mu}_\delta] \quad \forall \delta \in [0,1],$$

and

$$C_\delta(\widehat{\sigma^2}^*) = [\underline{\sigma^2}_\delta, \bar{\sigma^2}_\delta] \quad \forall \delta \in [0,1]$$

For the characterizing functions of the fuzzy estimator $\hat{\gamma}^*$ a generating family of intervals is obtained through the following proposed equations:

$$\hat{\gamma}_{1,\delta} := \left[\sum_{i=1}^n \frac{1}{t_{i,\delta} - \gamma} \right]$$

$$\left[\sum_{i=1}^n \ln(t_{i,\delta} - \gamma) - \sum_{i=1}^n (\ln(t_{i,\delta} - \gamma))^2 + \frac{1}{n} (\ln(t_{i,\delta} - \gamma))^2 \right]$$

$$-n \sum_{i=1}^n \frac{\ln(t_{i,\delta} - \gamma)}{t_{i,\delta} - \gamma} = 0 \quad \forall \delta \in [0,1], \tag{7}$$

$$\hat{\gamma}_{2,\delta} := \left[\sum_{i=1}^n \frac{1}{\bar{t}_{i,\delta} - \gamma} \right] \left[\sum_{i=1}^n \ln(\bar{t}_{i,\delta} - \gamma) - \sum_{i=1}^n (\ln(\bar{t}_{i,\delta} - \gamma))^2 + \frac{1}{n} (\ln(\bar{t}_{i,\delta} - \gamma))^2 \right]$$

$$-n \sum_{i=1}^n \frac{\ln(\bar{t}_{i,\delta} - \gamma)}{\bar{t}_{i,\delta} - \gamma} = 0 \quad \forall \delta \in [0,1]. \tag{8}$$

Solving the two equations for $\hat{\gamma}_{1,\delta}$ and $\hat{\gamma}_{2,\delta}$, then lower and upper ends of the family of intervals for the fuzzy estimator $\hat{\gamma}^*$ are defined by $\underline{\gamma}_\delta = \min[\hat{\gamma}_{1,\delta}, \hat{\gamma}_{2,\delta}]$ and $\bar{\gamma}_\delta = \max[\hat{\gamma}_{1,\delta}, \hat{\gamma}_{2,\delta}]$.

$$(A_\delta(\hat{\gamma}^*)) = [\underline{\gamma}_\delta, \bar{\gamma}_\delta] \quad \forall \delta \in [0,1].$$

From this generating family of intervals, the characterizing function is obtained by the mentioned Construction lemma. For the fuzzy life time observations data from Figure 2 the characterizing function is depicted in Figure 3.

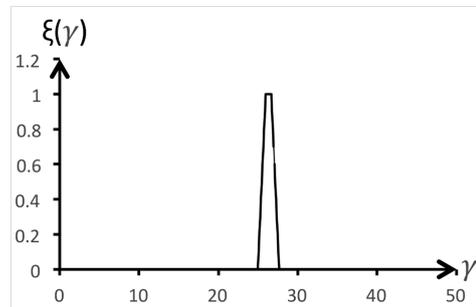


FIGURE 3. Characterizing function of the fuzzy estimator

For the fuzzy estimator $\hat{\mu}^*$ using the theorem given earlier, the generating family of intervals is obtained through the following equations:

$$(A_\delta(\hat{\mu}^*)) = \left[\frac{\sum_{i=1}^n \ln(t_{i,\delta} - \bar{\gamma})}{n}, \frac{\sum_{i=1}^n \ln(\bar{t}_{i,\delta} - \underline{\gamma})}{n} \right]; \quad \forall \delta \in [0,1]. \tag{9}$$

From this generating family of intervals the characterizing function is obtained by the mentioned Construction lemma. For the example fuzzy life time observations data from Figure 2 the characterizing function is depicted in Figure 4.

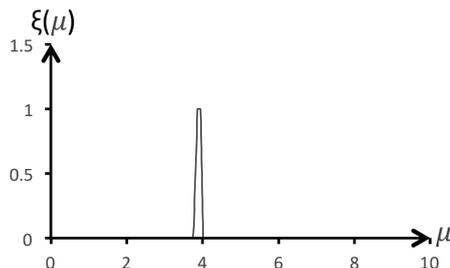


FIGURE 4. Characterizing function of the fuzzy estimator

For the characterizing functions of the fuzzy estimator the generating family of intervals is obtained through the following equations:

$$\overline{\sigma^2}_\delta := \frac{1}{n} \begin{cases} \sum_{i=1}^n [\ln(\underline{t}_{i,\delta} - \bar{\gamma}_\delta) - \bar{\mu}_\delta]^2 & \text{if } \underline{t}_{i,\delta} - \bar{\gamma}_\delta > 0 \\ \sum_{i=1}^n [\ln(\underline{t}_{i,\delta} - \underline{\gamma}_\delta) - \bar{\mu}_\delta]^2 & \text{if } \underline{t}_{i,\delta} - \bar{\gamma}_\delta < 0 \end{cases} \quad (10)$$

$$\sigma^2_\delta := \frac{1}{n} \left\{ \sum_{i=1}^n [\ln(\underline{t}_{i,\delta} - \underline{\gamma}_\delta) - \bar{\mu}_\delta]^2 \right\} \\ \left(A_\delta(\sigma^{2*}) \right) = \left[\sigma^2_\delta, \overline{\sigma^2}_\delta \right] \quad \forall \delta \in [0,1]. \quad (11)$$

From this generating family of intervals the characterizing function is obtained by the mentioned Construction lemma. For the example fuzzy life time observations data from Figure 2 the characterizing function is depicted in Figure 5.

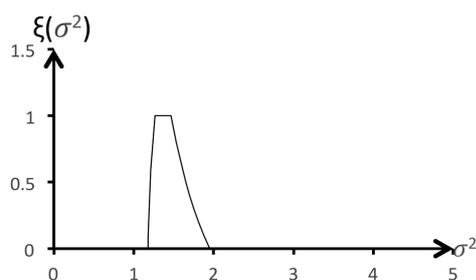


FIGURE 5. Characterizing function of the fuzzy estimator

CONCLUSION

In many situations, life time data distributions are highly positive skewed. In such situations the lognormal distribution is considered to be the best fit. The estimation techniques to estimate parameters of the lognormal distribution are usually based on precise life time observations, while in reality life times are not a precise measurement but fuzzy.

Therefore, for appropriate estimations in addition to classical tools fuzzy number methods are more suitable. In this study fuzzy estimators are proposed to assimilate fuzziness of life time observations in estimation. The results obtained are more suitable to model life time data in real situations.

REFERENCES

- Barbato, G., Germak, A., Genta, G. & Barbato, A. 2013. *Measurements for Decision Making. Measurements and Basic Statistics*. Bologna: Esculapio.
- Cohen, A.C. & Whitten, B.J. 1980. Estimation in the three-parameter lognormal distribution. *Journal of the American Statistical Association* 75(370): 399-404.
- Deshpande, J.V. & Purohit, S.G. 2005. *Life Time Data: Statistical Models and Methods*. Singapore: World Scientific Publishing.
- Huang, H.Z., Zuo, M.J. & Sun, Z.Q. 2006. Bayesian reliability analysis for fuzzy lifetime data. *Fuzzy Sets and Systems* 157(12): 1674-1686.
- Klir, G.J. & Yuan, B. 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. New Jersey: Prentice-Hall.
- Pak, A., Parham, G. & Saraj, M. 2013. Reliability estimation in Rayleigh distribution based on fuzzy lifetime data. *International Journal of System Assurance Engineering and Management* 5(4): 487-494.
- Szeliga, E. 2004. Structural reliability - fuzzy sets theory approach. *Journal of Theoretical and Applied Mechanics* 42(3): 651-666.
- Viertl, R. 2011. *Statistical Methods for Fuzzy Data*. Chichester: Wiley.
- Viertl, R. 2009. On reliability estimation based on fuzzy lifetime data. *Journal of Statistical Planning & Inference* 139(5): 1750-1755.
- Viertl, R. & Hareter, D. 2006. *Beschreibung und Analyse unscharfer Information: Statistische Methoden für unscharfe Daten*. Wien: Springer.
- Wu, H.C. 2004. Fuzzy Bayesian estimation on lifetime data. *Computational Statistics* 19(4): 613-633.
- Zadeh, L. 1965. Fuzzy sets. *Information and Control* 8(3): 338-353.

Muhammad Shafiq*
Department of Economics
Kohat University of Science and Technology
Kohat

Alamgir & Muhammad Atif
Department of Statistics
University of Peshawar
Pakistan

*Corresponding author; email: mshafiq_stat@yahoo.com

Received: 10 June 2015

Accepted: 14 June 2016