VOLUME-8, ISSUE-7, JULY-2019 • PRINT ISSN No. 2277 - 8160

Original Research Paper

Statistics

MODELLING CENSORED SURVIVAL DATA WITH Q-EXPONENTIAL DISTRIBUTIONS

Dr. N. Sundaram

Department of Statistics, Presidency College (Autonomous), Chennai-600005

ABSTRACT This research paper includes q-Exponential (q-E), Topp-Leone q-Exponential (TLq-E), and Topp-Leone Nadarajah-Haghighi (TLNH) models which can be used for the first time in modeling survival analysis with censoring. The maximum likelihood estimation of model parameters was also obtained using incomplete observations. We assessed the performance of the models using maximum likelihood estimators. The usefulness of the model was illustrated by using a diabetic data. The proposed model provides better fit compared to the other three models. TLNH model is better than the other two models based on the -2LL, AIC and quantile estimate of survival time.

KEYWORDS : q-Exponential, Topp-Leone q-Exponential, Topp-Leone Nadarajah-Haghighi, AIC.

INTRODUCTION

The exponential distribution (Gupta and Kundu, 1999, 2001) has wide range of applications including life testing experiments, reliability analysis, applied statistics and clinical studies. This distribution is a special case of the two parameter q-E, three parameter TLq-E and TLNH models. This research article concentrates on q-E, TLq-E and TLNH models.

q-Exponential (q-E) Model

An important characteristic of q-E distribution is that it has two parameters q and α providing more flexibility with regard to its decay, differently from exponential model. Picoli. and Malacarne (2003). If the survival time T follows the q-E. Its probability density function, distribution function and survival function are respectively

$$f(t)_{q-E} = (2-q)\alpha[1-(1-q)\alpha t]^{\left[\frac{1}{1-q}\right]},$$

$$t > 0, \alpha > 0, q < 2, q \neq 0, \qquad (1)$$

$$F(t)_{q-E} = 1 - [1-(1-q)\alpha t]^{\left[\frac{2-q}{1-q}\right]}, \text{ and} \qquad (2)$$

$$S(t)_{q-E} = [1-(1-q)\alpha t]^{\left[\frac{2-q}{1-q}\right]}. \qquad (3)$$

Topp-Leone q-Exponential (TLq-E) Model

Nicy Sebastian et al. have been defined the TLq-E model in 2019. If the survival time T follows the TLq-E probability density function, distribution function and survival function are as follows.

$$f(t)_{TLq-E} = 2\alpha\beta(2-q)[1-(1-q)\alpha t]^{\left[\frac{3-q}{1-q}\right]}$$

$$\left\{1-[1-(1-q)\alpha t]^{2\left[\frac{2-q}{1-q}\right]}\right\}^{\beta-1},$$

$$t > 0, \alpha, \beta > 0, q < 2, q \neq 0, \qquad (4)$$

$$F(t)_{TLq-E} = \left\{1-[1-(1-q)\alpha t]^{2\left[\frac{2-q}{1-q}\right]}\right\}^{\beta}, \text{ and } (5)$$

$$S(t)_{TLq-E} = 1 - \left\{1-[1-(1-q)\alpha t]^{2\left[\frac{2-q}{1-q}\right]}\right\}^{\beta}. \ (6)$$

Topp-Leone Nadarajah-Haghighi (TLNH) Model

TLNH model is defined by Yousof and Korkmaz (2017). If the survival time T follows the TLNH its probability density function, distribution function and survival function are respectively

$$f(t)_{TLNH} = 2\alpha\beta q[1+qt]^{\alpha-1}e^{2[1-(1+qt)^{\alpha}]}$$

$$\left[1-e^{2[1-(1+qt)^{\alpha}]}\right]^{\beta-1}, t > 0, \alpha, \beta, q > 0. \qquad (7)$$

$$F(t)_{TLNH} = \left[1-e^{2[1-(1+qt)^{\alpha}]}\right]^{\beta}, \text{ and} \qquad (8)$$

$$S(t)_{TLNH} = 1 - \left[1-e^{2[1-(1+qt)^{\alpha}]}\right]^{\beta}. \qquad (9)$$

PARAMETRIC MAXIMUM LIKELIHOOD FOR SURVIVAL MODEL

The n pairs of observations for ith individual is $(t_i, \delta_i; i=1,2,\ldots,n)$, where δ_i is an indicator variable

$$\delta_i = \begin{cases} 1 \ if \ t_i \ is \ uncensored \\ 0 \ if \ t_i \ is \ censored. \end{cases}$$

The total likelihood function for each model can be written in the following way

For q-Exponential model:

$$L_{q-E} = \prod_{i=1}^{n}\{f_{T(U)}(t_i)_{q-E}\}^{\delta_i} \ \{S_{T(C)}(t_i)_{q-E}\}^{1-\delta_i}, \qquad (10)$$

For *TLq*-Exponential model:

$$L_{TLq-E} = \prod_{i=1}^{n}\{f_{T(U)}(t_i)_{TLq-E}\}^{\delta_i} \ \{S_{T(C)}(t_i)_{TLq-E}\}^{1-\delta_i}, \qquad (11)$$

For *TLNH* model:

$$L_{TLNH} = \prod_{i=1}^{n}\{f_{T(U)}(t_i)_{TLNH}\}^{\delta_i} \ \{S_{T(C)}(t_i)_{TLNH}\}^{1-\delta_i}, \qquad (12)$$

where T(U) and T(C) are uncensored and censored survival times respectively. Under the assumption that the censoring times and survival times are independent, the above equations (10, 11 and 12) do not involve any unknown lifetime parameter. So these products can be treated like constants when maximizing L for each model.

DATABASE (Lee and Wang, 2003)

In this data set, one hundred and forty-nine diabetic patients were followed for 17 years. Out of this 17% were censored observations. Survival time and status are considered in this paper.

DATA ANALYSIS

In this section we used the diabetic data for the comparison of parametric models using graphic and algebraic methods. Time *T* cannot assume zero values since log *T* is not defined for $T = 0$. The q-E model is specified completely by the two parameters q and α. Figure 1 gives the survival smooth curves that is obtained from the q-E model for q = 0.5147 and α=0.06, from which an idea of the flexibility of the model may be obtained. Figure1 shows that the median survival time is approximately 7 years and other quartile estimates are $\hat{Q}_1 = 3$ years, and $\hat{Q}_3 = 12$ years (Table 1).

The TLq-E model is specified completely by the three parameters q, α and β. Figure 1 gives the survival smooth curves is obtained from the TLq-E model for q = 0.6497, α=0.06, and β=4.2608 from which an idea of the flexibility of the model may be obtained. Figure1 shows that the median survival time is approximately 11 years and other quartile estimates are $\hat{Q}_1 = 7$ years, and $\hat{Q}_3 = 14$ years (Table 1).

The TLNH model is specified completely by the three parameters q, α and β. Figure 1 gives the survival smooth curves obtained from the TLNH model for q = 0.0013, α= 43.5875 , and β=3.2948, from which an idea of the flexibility of the model may be obtained. Figure 1 shows that the median survival time is approximately 11 years and other quartile estimates are $\hat{Q}_1 = 7$ years, and $\hat{Q}_3 = 15$ years (Table 1).

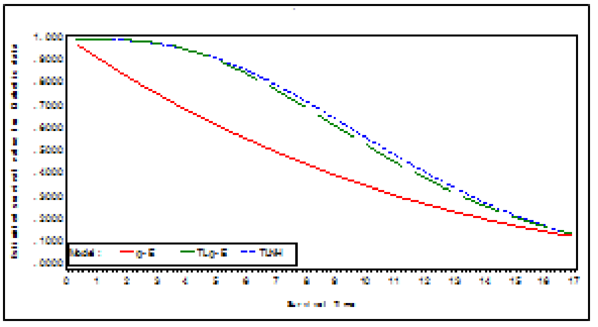

Figure 1 Survival curve for diabetic data

Figure 2 displays the comparison between the probability density function of the parametric q-E, TLq-E and TLNH (Table 2). It can be observed for the diabetic data, TLNH model fits better than q-El and TLq-E models, also TLNH model gives better survival times.

The q-E model is the simplest with two parameters to be estimated and is a special case of the TLq-E and TLNH models. The choice of parametric family to be used depends on the shape of the probability density function. We infer from Figure 2, three parametric models have a better fit compared to the other two parametric models.

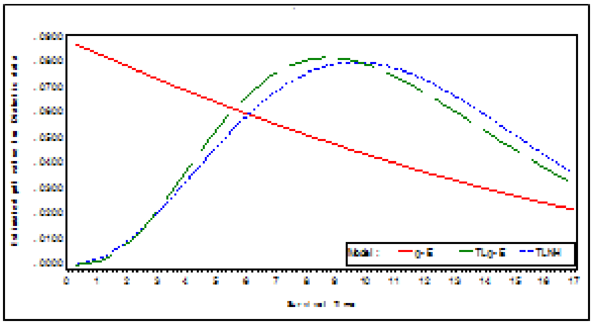

Figure 2 Probability density curve for diabetic data

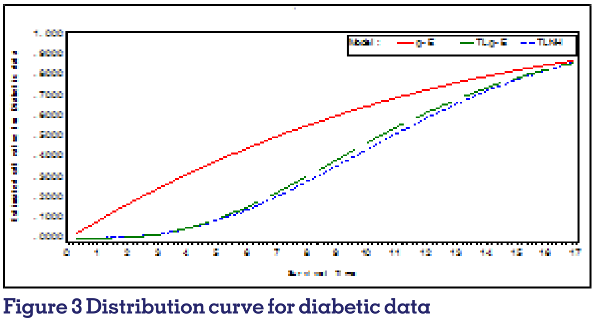

Figure 3 Distribution curve for diabetic data

Figure 3 represents the closeness of the models and q-E model far away from the other two models. TLq-E and TLNH models were very close each other at the origin and at the end of the survival time. Also infer that TLNH model gives better survival time based on the quantile estimate (Table 1).

Table 1 Quantile estimates for diabetic data

Quantile Estimate	Model		
	q-E	TLq-E	TLNH
$\hat{Q}_1$	3	7	7
Median	7	11	11
$\hat{Q}_3$	12	14	15

MODEL COMPARISONS

From Table 2, TLNH model has the smallest AIC and -2LL values. Because of this, TLNH model fits better than the other two models q-E and TLq-E.

Table 2 Optimization Results for diabetic data

Model	Estimate			-2LL	AIC
	$\hat{q}$	$\hat{\alpha}$	$\hat{\beta}$		
q-E	0.5147	0.0600	-	843	847
TLq-E	0.6497	0.0600	4.2608	770	776
TLNH	0.0013	43.5875	3.2948	751	757

CONCLUSION

The q-E, TLq-E and TLNH models plays an important role in describing many probability distributions. Among these TLq-E and TLNH models are in good agreement. Thus, to decide which distribution gives a better description, it is necessary to consider a sufficiently large range. In this work, in order to investigate the referred controversy, we focused systems that takes a sufficiently large range of data into account. Moreover, as to obtain a further illumination on this controversy, for the first time we applied censored data in the above said models. We applied q-E, TLq-E and TLNH models to analyze survival distributions of diabetic patients. We verified that the diabetic patient data is well described by TLq-E and TLNH models based on the quantile estimate (Table 1). We hope that the TLq-E and TLNH models can be useful in other situations. Finally we conclude that the TLNH model performs better (reduced AIC and -2LL) with optimizes results described in Table 2. These results show that the TLNH model has the lowest AIC, and -2LL values. Hence, it could be chosen as the best model under these circumstances. We also plot the estimated survival curve, probability curve and cumulative distribution curve of the q-E, TLq-E and TLNH models for the diabetic patient data set (Figure 1, Figure 2 and Figure 3). Clearly, the TLNH model provides a closer fit to the empirical survival function of diabetic data.

REFERENCES

1. Elisa T. Lee and John Wenyu Wang (2003). Statistical methods for survival data analysis. Third edition, John Wiley & Sons, Inc., publication.
2. Gupta, R.D. and Kundu, D. (1999). Generalized exponential distribution. Australian and New Zealand Journal of Statistics, Vol. 41, no. 2, pp. 173-88.
3. Gupta, R. D. and Kundu, D. (2001). Exponentiated exponential family; an alternative to gamma and Weibull. Biometrical Journal, Vol. 43, pp. 117–130.
4. Haitham M.Yousof and Mustafa Ç. Korkmaz (2017), Topp-Leone Nadarajah-Haghighi distribution. Journal of Statisticians: Statistics and Actuarial Sciences. IDIA 10(2), 119-128.
5. Nicy Sebastian, Rasin R. S. and Silviya P. O (2019). Topp-Leone generated q-exponential distribution and its applications. Math.ST - Statistics Theory, arXiv:1903.07028v1.
6. Picoli S Jr., R.S. Mendes and L.C (2003). Malacarne, q-exponential, Weibull, and q-Weibull distributions: an empirical analysis, Physica A, 324, 678-688.