Journal of Natural Sciences Research www.iiste.org
ISSN 2224-3186 (Paper) ISSN 2225-0921 (Online)
Vol.6, No.5, 2016

Exponentiated Weibull Distribution on Censored Data in Clinical Trial

Eric Boahen*
Department of Statistics,Faculty of Mathematical Sciences,University for Development Studies –Tamale.

ANNAN RICHARD CLOTTEY
Faculty of Mathematical Sciences,University for Development Studies –Tamale.

Abstract
Recent studies in clinical trial, Weibull distribution play an important role because simulation studies indicated that, Weibull distribution is less biased than other distribution in maximum likelihood. It estimators in small samples are less biased as compared other forms of imputations. In this paper, Weibull distribution was applied to clinical censored data. The result indicated that, Weibull distribution is comparatively has less biased estimators, small variance and smaller confidence interval than mean imputation techniques.

1.1 INTRDUCTION

The Weibull distribution introduced by Professor Waloddi Weibull in 1939, is an important generalization of the exponential model with two positive parameters. The second parameter in the model allows greater flexibility of the model and different shape of the hazard function. The convenient of the Weibull model for empirical work stem on the one hand from this flexibility and on the other hand from the simplicity of the hazard and survival functions. The two parameter Weibull distribution has probability density function as

$f(t,\lambda,\gamma) = \lambda\gamma t^{\gamma-1}e^{(-\lambda t^{\gamma})}$ 1

Survival function $S(t) = e^{(-\lambda t^{\gamma})}$ 2

Hazard function $\lambda(t) = \lambda\gamma t^{\gamma-1}$ 3

And cumulative function as λt^{γ} 4

The mean $E(t) = \lambda^{-\gamma}\Gamma(1+\frac{1}{\gamma})$ 5

Variance $V(t) = \lambda^{-2/\gamma}\left[\Gamma\left(1+\frac{2}{\gamma}\right)\right] - \Gamma\left(1+\frac{1}{\gamma}\right)^2$ 6

Where Γ denote the gamma function with $\Gamma(k) = \int_0^{\infty} S^{k-1}e^{-s}ds, \quad k > 0$ 8

where $t \geq 0, \lambda > 0, \gamma > 0$ are all real parameters. At a point where $\gamma = 1$, Weibull density is equivalent to exponential distribution. At this constant hazard rate, Weibull distribution is seen as a generalization of exponential distribution. the hazard function decreases monotonously from ∞ at time zero to zero at time ∞ for $\gamma < 1$. the hazard function increase monotonously from zero at time zero to ∞ for $\gamma > 0$. Weibull distribution can also be generated as a limiting distribution of the minimum of a sample from a continuous distribution with a support on an open ended [0,x) for some value x ($0< x < \infty$). This extreme value attribute makes the Weibull distribution appropriate for the distribution of individual patients time to death in clinical trial, because there different causes of death which compete with each other and the first one to strike the patient will cause the death. A data follow Weibull distribution if the survival time of a patient is the sum of the survival times, where each one having a constant hazard rate. However, the Weibull distribution is inappropriate when the hazard rate is indicated as unimodal or bathtub- shape. A truncated death density of patients differ from censored death density. This is because, the total sampl size n of patients on study is unknown in truncated Weibull distribution, and if n deaths are recorded prior to some event time t, then these deaths represent the first n ordered observation from unknown number of ordered observations. In standard treatment of unlikely recovery disease, it is assumed that every individual on the study is at risk form the beginning and continues to be at risk until death or the patient is censored at a particular point ih time in the study. At this point patients who are not at risk are removed from the study and may or may not return to the study at later stage when they are attributed to the set of risk to the disease. Given the death density a patient to be $f(t,\lambda,\gamma)$, the corresponding truncated death density for the truncated patient assuming he/she was truncated prior to event time t is

$\varphi(t,\lambda,\gamma) = \frac{f(t,\lambda,\gamma)}{F(T,\lambda,\gamma)}$, where$F(T,\lambda,\gamma) = \int_0^T f(t,\lambda,\gamma)dt$, for $0 \leq t \leq T$ 9

1.2 WEIBULL CENSORED

Suppose $t_1, t_2, \ldots t_n$ are random samples of patients from a specific population. However, the right censored such as staggered entry of patients, loss to follow-up, competing risks, that is death from other cause or any combination of these may be an impediment to have the opportunity of observing these survival times. Let us denote d to be the censoring time for all the patients in the study. It is assumed that, the survival times of the first

$d \leq n$ patients are recorded and the survival times of the remaining $n-d$ patients are not recorded. This assumption stipulated that the data is singly censored. In singly censored Weibull distribution, location parameter is distorted and both scale and shape parameters are easily captured. To estimate these parameters, the likelihood function

$$L(\lambda,\gamma) = \frac{n!}{(n-d)!}(\lambda\gamma)^d \prod_{i=1}^{d} t_i^{\gamma-1} e^{-[(\lambda\sum_{i=1}^{d-1} t_i^{\gamma}+(n-d)+1)\lambda t_d^{\gamma}]} \quad 10$$

From equation one, it follows that $\frac{\partial^2 logL}{\partial\lambda^2} = -\frac{d}{\lambda^2}$ 11

$$\frac{\partial^2 logL}{\partial\lambda\partial\gamma} = -\sum t_i^{\gamma}(\log t_i) \quad 12$$

and

$$\frac{\partial^2 logL}{\partial\gamma^2} = -\frac{d}{\gamma^2} - \lambda\sum t_i^{\gamma}(\log t_i)^2 \quad 13$$

For large samples, the variance –covariance matrix of the estimators $\hat{\lambda}$ and $\hat{\gamma}$ is given by inverting the matrix of the second partial derivatives. This means that the approximate values $Var\hat{\lambda}$, $Var\hat{\gamma}$, and $Cov(\hat{\lambda},\hat{\gamma})$ is the elements of the matrix

$$\begin{bmatrix} Var\hat{\lambda} & Cov(\hat{\lambda},\hat{\gamma}) \\ Cov(\hat{\lambda},\hat{\gamma}) & Var\hat{\gamma} \end{bmatrix} = \begin{bmatrix} \frac{d}{\lambda^2} & \sum t_i^{\gamma}(\log t_i) \\ \sum t_i^{\gamma}(\log t_i) & \frac{d}{\gamma^2} + \lambda\sum t_i^{\gamma}(\log t_i)^2 \end{bmatrix}^{-1} \quad 14$$

The vector estimator $\begin{pmatrix}\hat{\lambda}\\ \hat{\gamma}\end{pmatrix}$ has an approximate bivariate normal distribution in large samples with mean vector $\begin{pmatrix}\lambda\\ \gamma\end{pmatrix}$.

An approximate $(1-\alpha)100$ percent confidence region for the mean $\begin{pmatrix}\lambda\\ \gamma\end{pmatrix}$ is given by

$$\frac{d}{\lambda^2}(\hat{\lambda}-\lambda)^2 + \left[\frac{d}{\gamma^2} + \lambda\sum t_i^{\gamma}(\log t_i)^2\right](\hat{\gamma}-\gamma)^2 + 2\sum t_i^{\gamma}(\log t_i)(\hat{\lambda}-\lambda)(\hat{\gamma}-\gamma) = x^2(1-\alpha:2) \quad 15$$

Since λ and γ are two parameters whose corresponding death density function is $f(t:\lambda,\gamma)$, it follows that the estimators $\hat{\lambda}$ and $\hat{\gamma}$ are the maximum likelihood estimators of λ and γ. The maximum likelihood function of the data $(t_1, t_2, t_3 \ . \ . \ . \ t_n)$ of λ and γ is $L(\lambda,\gamma) = \prod_{i=1}^{n} f(t:\lambda,\gamma)$. In this case an approximate large sample 100(1-α) percent elliptical confidence region for λ and r. this region is given by

$$\sigma^{11}(\hat{\lambda},\hat{\gamma})(\lambda-\hat{\lambda},)^2 + \sigma^{22}(\hat{\lambda},\hat{\gamma})(\gamma-\hat{\gamma}) + 2\sigma^{12}(\hat{\lambda},\hat{\gamma})(\lambda-\hat{\lambda})(\gamma-\hat{\gamma}) = x^2(1-\alpha:2),$$

where

$$\sigma^{ij}(\hat{\lambda},\hat{\gamma}) = -E\left[\frac{\partial^2 Log_e L(\lambda,\gamma)}{\partial\lambda\partial\gamma}\right] \quad 16$$

And $x^2(1-\alpha:2) is\ the$ 100(1-α) percent of the chi- square distribution with 2 degree of freedom. If $\sigma^{12}(\hat{\lambda},\hat{\gamma}) = 0$, the elliptical region defined by (1) has its major and minor axes parallel to the λ and r axes. To add to this, if $\sigma^{11}(\hat{\lambda},\hat{\gamma}) = \sigma^{22}(\hat{\lambda},\hat{\gamma})$, the region is circular. In the general case. When $\sigma^{12}(\hat{\lambda},\hat{\gamma}) \neq 0$ and $\sigma^{11}(\hat{\lambda},\hat{\gamma}) \neq \sigma^{22}(\hat{\lambda},\hat{\gamma})$, then the elliptical confidence region for (λ,γ) does not have its major and minor axes parallel to the λ and r axes. In this case the major and minor axes are determined by rotation. Supposing $\emptyset$ is the angle that the major and minor makes with the λ and γ axes, $\emptyset$ is determined by

$$\tan 2\emptyset = \frac{2\sigma^{12}(\hat{\lambda},\hat{\gamma})}{\sigma^{11}(\hat{\lambda},\hat{\gamma}) - \sigma^{22}(\hat{\lambda},\hat{\gamma})} \quad 17$$

1.3 APPLICATION OF WEIBULL DISTRIBUTION TO CENSORED DATA IN CLINICAL TRIAL.

The length in months of patients survival time who were suffering from acute renal calculi were recorded as 0.001, 0.030, 0.071, 0.185, 0.345, 0.435, 0.469, 0.470. 0.505, 0.664, 0.806, 0.970, 1.033, i.550, 1.550, 2.046+, 3.532, 7.057, 9.098+, 57.628 where 2.046+ and 9.098+ months denoted by plus sign were followed to completion and were in remission when study ended. It is very useful to obtain the maximum likelihood estimates $\hat{\lambda}$ and $\hat{\gamma}$ of λ and γ

$$\frac{\sum_{i=1}^{18} t^{\hat{\gamma}} logt_i + \sum_{i=1}^{2} T^{\hat{\gamma}}\ logT_i}{\sum_{i=1}^{18} t^{\hat{\gamma}} + \sum_{i=1}^{2} T^{\hat{\gamma}}} - \frac{1}{\hat{\gamma}} + 0.624 = 0 \quad 18$$

Where $0.624 = {}^{1}/_{18}\sum_{i=1}^{18} t^{\hat{\gamma}} logt_i$

Setting

$$G(\gamma) = \frac{\sum_{i=1}^{18} t^{\hat{\gamma}} logt_i + \sum_{i=1}^{2} T^{\hat{\gamma}}\ logT_i}{\sum_{i=1}^{18} t^{\hat{\gamma}} + \sum_{i=1}^{2} T^{\hat{\gamma}}} - \frac{1}{\hat{\gamma}} + 0.624 \quad 19$$

From interpolation $G(0.4) = -0.703$ and $G(0.5) = 0.166$. from this, $\hat{\gamma}$ lies between 0.4 and 0.5.

$\hat{\gamma}$	$G(\hat{\gamma})$
0.4	-0.703

$\hat{\gamma} = 0.481$ 0.000
0.5 0.166
From the above interpolation $\hat{\gamma} = 0.481$
The maximum likelihood of $\hat{\lambda}$ is

$$\hat{\lambda} = \frac{18}{\sum_{i=1}^{18} t^{\hat{\gamma}} + \sum_{i=1}^{2} T^{\hat{\gamma}}} = \frac{18}{26.245} = 0.686 \quad (20)$$

$\hat{\lambda} = 0.686$

The asymptotic variance – covariance for the estimator $\begin{pmatrix}\hat{\lambda}\\ \hat{\gamma}\end{pmatrix}$ is

$$\begin{bmatrix} \widehat{Var\lambda} & Cov(\hat{\lambda},\hat{\gamma}) \\ Cov(\hat{\lambda},\hat{\gamma}) & Var\hat{\gamma} \end{bmatrix} = \begin{bmatrix} \frac{d}{\lambda^2} & \sum t_i^{\gamma}(\log t_i) \\ \sum t_i^{\gamma}(\log t_i) & \frac{d}{\gamma^2} + \lambda \sum t_i^{\gamma}(\log t_i)^2 \end{bmatrix}^{-1} \quad (21)$$

$$= \begin{bmatrix} 38.249 & 38.715 \\ 38.715 & 182.538 \end{bmatrix}^{-1} = \begin{bmatrix} 0.033 & -0.007 \\ -0.007 & 0.007 \end{bmatrix} \quad (22)$$

It should be noted that, the approximate estimate of this vector distribution $\begin{pmatrix}\hat{\lambda}\\ \hat{\gamma}\end{pmatrix}$ has a bivariate normal distribution with mean $\begin{pmatrix}\lambda\\ \gamma\end{pmatrix}$ and variance- covariance matrix $\begin{bmatrix} 0.033 & -0.007 \\ -0.007 & 0.007 \end{bmatrix}$. The 95% joint elliptical confidence is
$38.249(\lambda - 0.686)^2 + 182.538(\gamma - 0.481)^2 + 79.430(\lambda - 0.686)(\gamma - 0.481)$
$= 5.98$

$\tan 2\emptyset = \frac{2\sigma^{12}(\hat{\lambda},\hat{\gamma})}{\sigma^{11}(\hat{\lambda},\hat{\gamma}) - \sigma^{22}(\hat{\lambda},\hat{\gamma})}$

$\sigma^{11}(\hat{\lambda},\hat{\gamma}) = \frac{d}{\lambda^2}$, $\sigma^{12}(\hat{\lambda},\hat{\gamma}) = \sum t_i^{\gamma}(\log t_i)$, $\sigma^{22}(\hat{\lambda},\hat{\gamma}) = \frac{d}{\gamma^2} + \lambda \sum t_i^{\gamma}(\log t_i)^2$

$\sigma^{11}(\hat{\lambda},\hat{\gamma}) = \frac{d}{\lambda^2} = 0.033$

$\sigma^{12}(\hat{\lambda},\hat{\gamma}) = \sum t_i^{\gamma}(\log t_i) = -0.007$

$\sigma^{22}(\hat{\lambda},\hat{\gamma}) = \frac{d}{\gamma^2} + \lambda \sum t_i^{\gamma}(\log t_i)^2 = 0.007$

$tan2\emptyset = \frac{2(-0.007)}{0.033+0.007} = \frac{-0.014}{0.026} = -0.538$

$\emptyset = tan^{-1} - 0.27 = 15.06^{\circ}$

The separate approximation 95 percent confidence interval for λ and γ is $\hat{\lambda} \pm 1.96\sqrt{var\hat{\lambda}}$ and $\hat{\gamma} \pm 1.96\sqrt{var\hat{\gamma}}$ respectively. It should be noted that $var\hat{\lambda}$ and $var\hat{\gamma}$ are the diagonal elements of the variance covariance matrix.

$\hat{\lambda} \pm 1.96\sqrt{var\hat{\lambda}} = 0.686 \pm 1.96\sqrt{0.033} = (0.33 < \lambda < 1.04)$

$\hat{\gamma} \pm 1.96\sqrt{var\hat{\gamma}} = 0.481 \pm 1.96\sqrt{0.007} = (0.317 < \gamma < 0.645)$

Table 1.1

Method	Mean	Variance	Confidence interval
Weibull	0.686	0.033	$0.33 < \lambda < 1.04$
Mean substitution	4.295	12.355	-2.594< λ < 11.184

1.4 Conclusion

From table above, the mean of Weibull distribution months has a minimum variance as compared to variance of the mean for the mean imputation seen as the maximization of the mean survival time of patients. In this case, the survival time of the patients from Weibull distribution estimate is very close to the true mean survival time, even though part of the data is censored the available known information has been accounted for and maximized through maximum likelihood. From the table, mean imputation has large confidence interval of -2.594< λ < 11.184 as compared to that of Weibull $0.33 < \lambda < 1.04$. Through population mean is more closer to sample mean in Weibull distribution the mean imputation. It is seen from the paper that substitutions in all forms including imputation distort censored information and consequently lead to bias. It is not surprising that Helsel (1990) and other publications consistently shown that using substitution and any other form of imputation in censored observations unnecessary introduce bias in summary statistics.

REFERNCES

P. Armitage,. "The comparison of survival curves" *Journal of the Royal Statistical Society, vol.A122, 279-300,* 1959

J. Berkson. and R. P Gage "Survival curve for cancer patients following treatment" *Journal of the American*

Statistical Association vol. 47, 501-515, 1952
Adam W., Naoke W., (1999). Socioeconomic Inequalities in Child Malnutrition iDeveloping World .*World Bank Policy Research Working Paper No. 2434*. (Accessed:March 3, 2015).
Afifi A. A. and S.P. Azen (1972), statistical analysis: A computer oriented Approach, Academic Press, New York
Aitken, A. C. (1935), On least-squares and Linear Combinations. Proceedings of the Royal Society of Edinburgh
Kaplan, E. L and O. Meier, 1958 nonparametric estimation from incomplete observations *Journal of the American statistical association vol. 53, 457-481*
Armitage, P. 1959 The comparison of survival curves *Journal of the Royal Statistical Society, vol.A122, 279-300*
Berkson J. and R. P Gage 1952 Survival curve for cancer patients following treatment *Journal of the American Statistical Association vol. 47, 501-515*
Anderson, T. W. (1964). Sequential Analysis with Delayed Observations, Journal of the American Statistical Association, Vol.59, pp. 1006-1015
Armitage, P. (1960) Sequential Medical trials, Charles C. Thomas, Springfield, III
Bain, L. J. and C. E. Antle (1967), Estimation of Parameter in the Weibull Distribution, Technometrics, Vol. 9, pp.621-628
Barlow, R. E. and F. proschan (1965). Mathematical Theory of Reliability. John Willey and Sons. New York