

Empirical likelihood ratio test for median and mean residual lifetime

Mai Zhou and Jong-Hyeon Jeong

Department of Statistics, University of Kentucky, Lexington, KY 40506-0027, USA
Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, USA

SUMMARY

Prompted by several recent papers on the inference on median and mean residual life time, we note that the testing involving the mean or median residual life function in censored survival data can be obtained by an easy application of the general empirical likelihood ratio test. This approach has several advantages: (1) there is no need to estimate the variance/covariance at all, which may become prohibitively complicated for other procedures that require the estimation of such. (2) When inverting the tests to obtain confidence regions/intervals, this procedure inherits all the good properties of a likelihood ratio test. (3) Free software implementation of the test is readily available.

KEY WORDS: censored data; chi-square distribution; confidence interval; Wilks theorem.

1. INTRODUCTION

Jeong, Jung and Costantino [1] recently proposed a score-type test for the median residual life time. They argue that “the need for such estimates is becoming more critical in breast cancer research as long-term courses of secondary therapies are now being considered for patients who remain recurrence free after several years of initial treatment”. Using the concept of the median residual lifetimes in statistical inference would also help patients and physicians understand the efficacy of a new drug in a more intuitively straightforward way than using the traditional tools based on such as the hazard function or probability of survival.

Despite the practical usefulness of the median residual life function, however, it is well known that the inference on the median residual life function in survival data can be prohibitively challenging because it involves nonparametric estimation of the density function of the unknown failure time distribution under censoring. Therefore, Jeong *et al.* [1] considered a method directly dealing with an estimating equation to avoid estimation of the density function, like in the Berger *et al.* [2], proposing a score-type (possibly stratified) test to

compare median residual lifetimes between two or more groups. But their procedure still involves estimation of the variance of score function to evaluate the test statistic through the martingale theory. In this paper, we note that an even simpler procedure is available via the empirical likelihood ratio test for censored survival data, which does not require estimation of any variance at all. The empirical likelihood approach also inherits all the good properties of a likelihood ratio test and can handle more general types of censored data.

There are also a couple of recent papers dealing with the mean residual life function using the empirical likelihood approach [3,4]. Qin and Zhao [4], however, proposed a different approach using an estimating function that involves nuisance parameters. The version of empirical likelihood they proposed do not have a regular chi-squared limiting distribution under null hypothesis, unlike ours.

In Section 2, the median residual life function is defined and an empirical likelihood ratio test is proposed. In Section 3, the similar procedure applies to the mean residual life function. In Section 4, an existing statistical software that can facilitate the empirical likelihood ratio-based inference is briefly introduced. In Section 5, performance of the proposed empirical likelihood ratio test for the median residual life function is compared with the score-type test proposed by Jeong *et al.* [1] through a simulation study. In Section 6, the proposed method is illustrated with two real data examples. In Section 7, we conclude with a brief remark.

2. MEDIAN RESIDUAL LIFE

In survival data, the median residual lifetime at age x is defined as the median of the distribution of failure times among survivors beyond time x , i.e $P(T - x > \theta | T > x) = 0.5$, where θ is the median of the remaining lifetimes at time x . Therefore the median residual lifetime is quantitatively defined as the number θ that solves the equation

$$\frac{1 - F(x + \theta)}{1 - F(x)} = 0.5,$$

where $F(\cdot)$ is the cumulative distribution function of failure times. Other quantiles of the residual life distribution can be defined similarly. Even though we shall focus on developing the test for the median residual life function in the sequel, it can be easily modified to test any quantile residual life function.

Let us denote the median residual lifetime at age x as $Med(x)$. Clearly $\theta = Med(x)$ is also the solution to

$$1 - F(x + \theta) = 0.5\{1 - F(x)\} .$$

After rearranging the terms, we see that θ is the solution to

$$0.5 = F(x + \theta) - 0.5F(x) .$$

If we define a function $g_b(t)$ as

$$g_b(t) = I_{[t \leq (x+b)]} - 0.5I_{[t \leq x]} - 0.5, \quad (1)$$

then the hypothesis $H_0 : Med(x) = b$ can be tested by testing

$$H_0 : \int_0^\infty g_b(t) dF(t) = 0 .$$

This, in turn, can be accomplished by an empirical likelihood ratio test.

2.1 Empirical Likelihood Ratio Test

The empirical likelihood ratio tests, first proposed by Thomas and Grunkemeier [5] and Owen [6], attracted much attention since then. The empirical likelihood methods developed in the last 20 years has emerged as a very competitive nonparametric test procedure for quite general settings, including the test of a parameter defined by $\int g(t) dF(t)$ with censored survival data. It parallels the theory of the parametric likelihood ratio test, except the parametric likelihood is replaced by a nonparametric one. The book of Owen [7] summarized many of the results (Chapter 6 in particular). Other relevant papers include Murphy and van der Vaart [8], Pan and Zhou [9], and Zhou [10]. The following is an adaptation/summary of the relevant results from above sources suitable for our applications.

Suppose T_i , $i = 1, 2, \dots, n$, are independent and identically distributed (iid) event times of interest with a distribution $F(t)$. Due to censoring, we only observe a censored sample $Y_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$ is an indicator function. We assume the censoring time C_i is independent of the event time T_i .

Let p_i denote the probability mass over each observation Y_i , then the empirical likelihood (EL) for the above censored data is defined as

$$EL = \prod_{i=1}^n \{p_i\}^{\delta_i} \left\{ \sum_{Y_j > Y_i} p_j \right\}^{1-\delta_i} . \quad (2)$$

The maximization of the above EL with respect to p_i , subject to $p_i \geq 0$ and $\sum p_i = 1$, is well known to be achieved by (the jumps of) the Kaplan-Meier estimator computed from

(Y_i, δ_i) (Owen [7], pg 142). Let us denote the maximum empirical likelihood value achieved as $EL(KM)$.

In order to form the likelihood ratio, we also need to maximize the EL with respect to p_i under an extra constraint

$$\sum_{i=1}^n g(Y_i)p_i = \theta, \quad (3)$$

where $g(t)$ is a given function such that $0 < \text{Var } g(T) < \infty$ and θ is the value we wish to test.

The mean of the Kaplan-Meier estimator, i.e. $\sum g(Y_i)w_i$ with w_i being the jumps of the Kaplan-Meier estimator, may not always have a finite asymptotic variance. So we need the following extra condition to guarantee that:

$$\int \frac{g^2(t)}{1 - G(t-)} dF(t) < \infty, \quad (4)$$

where $G(\cdot)$ is the distribution function of the censoring variable C_i .

The Empirical Likelihood Theorem asserts that under the null hypothesis, $H_0 : \theta = \mathbb{E}g(T)$, $-2 \log$ empirical likelihood ratio has an asymptotic chi-squared distribution.

Theorem *Consider the right censored data and its empirical likelihood defined above. Suppose $\mathbb{E}g(T) = \theta$. Assume also that condition (4) holds. Then we have*

$$W(\theta) = -2 \log \frac{\max EL}{EL(KM)} \xrightarrow{\mathcal{D}} \chi_1^2 \quad \text{as } n \rightarrow \infty,$$

where the numerator max is carried out over all probabilities p_i that satisfy (3).

Testing the equality (or the ratio) of two median residual times from two samples (or from one sample at two different ages) can be carried out similarly as outlined in Jeong *et al.* [1].

Case I: If we are to test $H_0 : \text{Med}_1(x_1)/\text{Med}_2(x_2) = c$, where $\text{Med}_k(x_k)$ ($k = 1, 2$) denote the median residual time from sample k at age x_k , we shall first obtain two empirical likelihood ratio statistics for testing the two auxiliary hypothesis: $H_{01} : \text{Med}_1(x_1) = c\theta$ and $H_{02} : \text{Med}_2(x_2) = \theta$. Let us denote the two resulting test statistic by $W_1(\theta)$ and $W_2(\theta)$. Then the original hypothesis $H_0 : \text{Med}_1(x_1)/\text{Med}_2(x_2) = c$ can be tested by using the statistic $W = \inf_{\theta} \{W_1(\theta) + W_2(\theta)\}$, which follows a chi-square distribution with 1 degree of freedom under H_0 . Note that a special case of the null hypothesis $H_0 : \text{Med}_1(x)/\text{Med}_2(x) = c$ was considered in Jeong *et al.* [1].

Case II: If we are to test the ratio of two median residual lifetimes from the same sample but at two different ages x_1 and x_2 , i.e. $H_0 : \text{Med}(x_1)/\text{Med}(x_2) = c$, we may proceed

similarly as above, except we need to replace the two auxiliary hypothesis by $H_{00} : Med(x_2) = \theta, Med(x_1) = c\theta$. Denote the Wilks likelihood ratio statistic for H_{00} by $W_3(\theta)$. Then the test of $H_0 : Med(x_1)/Med(x_2) = c$ can be obtained by using $W = \inf_{\theta} W_3(\theta)$. This statistic again will have a chi-square distribution with 1 degree of freedom under H_0 .

For the score-type test, this latter case would be much more involved, since the covariance between $Med(x_1)$ and $Med(x_2)$ needs to be estimated. This is even more so when we are dealing with the mean residual time. Due to the discrete nature of the quantile function, some have advocated for a smoothed quantile empirical likelihood ratio test [11]. One way of smoothing the residual median is to replace the indicator functions in (1) by a smoothed version.

In addition to the advantage of no need for nonparametric estimation of the density function for the variance calculation, the empirical likelihood ratio inference inherits the nice properties of a likelihood ratio-based confidence region, i.e. range respecting and transform invariant.

3. MEAN RESIDUAL LIFE

The mean residual lifetime of a random variable T , at a given age x , is defined as

$$M(x) = E(T|T \geq x) - x = \frac{\int_x^{\infty} s dF(s)}{1 - F(x)} - x = \frac{\int_x^{\infty} 1 - F(s) ds}{1 - F(x)}.$$

For a given x value, we first notice that the hypothesis

$$H_0 : M(x) = \mu$$

is equivalent to the following hypothesis

$$H_0 : \frac{\int_x^{\infty} s dF(s)}{1 - F(x)} = (x + \mu) ,$$

which is also equivalent to

$$H_0 : \int_x^{\infty} s dF(s) = [1 - F(x)](x + \mu) .$$

This in turn can be written as (since $\int_x^{\infty} dF(s) = 1 - F(x)$)

$$H_0 : \int_x^{\infty} [s - (x + \mu)] dF(s) = 0 . \quad (5)$$

This test can be performed by a one-sample empirical likelihood ratio test for censored survival data, similar to the median case, but with a different definition of the g -function, i.e. $g(s) = [s - (x + \mu)]I_{(s > x)}$.

Testing the ratio of two mean residual times from two independent samples (or from the same sample but at two different time points) can be done following the same procedure outlined in Section 2.1.

4. AVAILABLE STATISTICAL PACKAGE

A publicly downloadable software implementation of the empirical likelihood ratio tests with censored survival data is **emplik**, which is an extension package to be used with the R software [12]. In particular, the function `el.cen.EM2` inside the package **emplik** carries out the above test. A real data example of calculating confidence intervals using the function is described in the APPENDIX, together with Section 6.

Since the procedure `el.cen.EM2()` can handle doubly censored data as well, the same test procedure outlined above can test median residual lifetime with doubly censored data. Left truncated and right censored data can be treated similarly, but another function `emplikH2.test()` inside the **emplik** package needs to be used after reformulating the hypothesis in terms of cumulative hazard.

5. A SIMULATION STUDY

A simulation study was performed to compare the two-sample testing procedure from Jeong *et al.* [1] and one based on the empirical likelihood approach. For both groups simulated, failure times were generated identically from a Weibull distribution with censoring proportions of 0%, 10%, 20% and 30% similarly as in Jeong *et al.* [1]. For a fair comparison, the non-smoothed version of the empirical likelihood ratio test was considered. The proportion of rejecting the null hypothesis of the equality of the two medians were compared for different sample sizes at various time points. Table I summarizes the results. In general, the results from Jeong *et al.*'s method tend to be slightly conservative than ones from the EL approach, especially when the sample size becomes large.

Table I. Empirical coverage probabilities for the Jeong *et al.* (2008) and Empirical Likelihood method.

n	t_0	Jeong et al. (2008)				Empirical Likelihood			
		0%	10%	20%	30%	0%	10%	20%	30%
50	0	.978	.978	.981	.976	.976	.975	.981	.979
50	1	.980	.979	.981	.976	.978	.979	.979	.976
50	2	.974	.973	.977	.976	.975	.976	.974	.976
50	3	.984	.986	.977	.979	.985	.989	.982	.985
100	0	.971	.970	.971	.977	.969	.969	.967	.972
100	1	.971	.973	.976	.979	.968	.969	.972	.977
100	2	.974	.976	.976	.978	.971	.973	.975	.975
100	3	.979	.981	.981	.982	.975	.976	.981	.981
500	0	.965	.966	.966	.968	.953	.952	.947	.957
500	1	.964	.966	.968	.968	.954	.957	.955	.958
500	2	.969	.969	.967	.970	.962	.961	.956	.960
500	3	.974	.972	.973	.969	.960	.958	.962	.959

6. REAL EXAMPLES

First, we take a data set **cancer** from the R package **survival**. It contains 228 survival times from lung cancer patients with 63 right-censored observations. We shall find the 90% confidence interval for the mean and median residual lifetimes at year one (365.25 days) i.e. confidence interval for $M(365.25)$ and $Med(365.25)$.

When inverting the empirical likelihood ratio tests to get the confidence intervals, it is often very helpful to know where is the ‘center’ of that confidence interval, i.e. when testing for this ‘center’ value, one should get a p-value of one. For the empirical likelihood ratio tests described in the previous sections, the ‘center’ is given by the nonparametric maximum likelihood estimator based on the Kaplan-Meier estimator.

Detailed steps to evaluate a 90% confidence interval by using the function `el.cen.EM2` are illustrated in the APPENDIX. Following the steps, we find the 90% confidence interval for the median residual time as $[184.75, 321.75]$. Notice that, due to the discrete nature of the quantile function, we do not get an exact p-value of 0.1. Smoothing the indicator function in (1), however, always enable us to get the exact p-value. Another benefit of smoothing is (potentially) a more accurate p-value, as indicated by Chen and Hall [11]. If we use the linear smooth function or the cubic smoother function with a bandwidth of $1/20$, we get a 90% confidence interval of $[184.74, 321.71]$ and $[184.77, 321.73]$, respectively. These intervals are practically very close to one from the non-smoothed approach. They are also very similar to the confidence interval estimate obtained by the score-type test as $[184.75, 321.74]$.

The second example comes from a breast cancer study (NSABP Protocol B-04) as de-

scribed in Jeong *et al.* [1]. The data include 586 node positive patients and 1079 node-negative patients. In this example, we first estimate the median residual lifetimes among node-positive and node-negative patients separately by using the empirical likelihood approach and then statistically compare them by using the 95% confidence intervals of the ratio estimated from both Jeong *et al.*'s (J) and the empirical likelihood (EL) ratio method. From Table II we see that the two approaches provided almost identical results for the 95% confidence intervals for the ratio of the two medians.

Table II. Estimated median residual lifetimes in node-negative and node-positive groups, the ratios, and 95% confidence intervals for the ratios (NSABP B-04 data).

t_0	<u>Median Residual Lifetime</u>		<u>Ratio</u>	<u>95% CI</u>	
	Node-Negative	Node-Positive		J	EL
0	12.46 (11.2,13.5)	6.87 (6.4,7.4)	0.55	(0.49, 0.63)	(0.49, 0.63)
2	12.44 (11.2,13.6)	6.93 (5.9,8.1)	0.56	(0.47, 0.70)	(0.47, 0.70)
4	13.05 (11.8,14.8)	8.24 (6.8,10.2)	0.63	(0.49, 0.81)	(0.49, 0.81)
6	13.40 (12.5,14.3)	8.75 (7.7,10.6)	0.65	(0.54, 0.81)	(0.56, 0.82)
8	12.91 (11.9,13.8)	10.19 (8.8,11.6)	0.79	(0.66, 0.93)	(0.67, 0.93)
10	12.48 (11.2,13.7)	9.66 (8.2,11.8)	0.77	(0.62, 1.00)	(0.62, 1.00)
12	11.85 (10.6,13.0)	9.66 (7.5,12.6)	0.82	(0.63, 1.08)	(0.63, 1.08)

7. A REMARK

In this note, we proposed a method to infer the median or mean residual lifetimes by using the empirical likelihood ratio approach for censored survival data. A major advantage of the proposed method is no need for nonparametric estimation of any kind of variance for statistical inference, especially for the median case. The results from the proposed method were similar to ones from a score-type test statistic recently proposed, implying that the empirical likelihood ratio method may be an important alternative, but simpler, method for inferring median or mean residual lifetimes in censored survival data.

Furthermore, this empirical likelihood approach can be generalized to provide easy statistical inference for the (median or mean) regression model analysis on the residual lifetimes. We shall study this in a forthcoming paper.

APPENDIX

Here we describe in detail some of the computations presented in Section 6. First the packages `emplik` and `survival` need to be loaded into R [package `survival` is only needed here to supply the data set `cancer`].

```

> data(cancer)
> time <- cancer$time
> status <- cancer$status-1
> MMRtime(x=time, d=status, age=365.25)
$MeanResidual
[1] 275.9997

$MedianResidual
[1] 258.75

```

The following is the result from testing the mean residual times through the confidence interval approach. First we need to define the g function for the residual mean life.

```

> mygfun <- function(s, age, muage) {as.numeric(s >= age)*(s-(age+muage))}
> el.cen.EM2(x=time, d=status, fun=mygfun, mu=0, age=365.25, muage=234.49389)$Pval
[1] 0.1000000
> el.cen.EM2(x=time, d=status, fun=mygfun, mu=0, age=365.25, muage=323.1998)$Pval
[1] 0.1

```

Therefore the 90% confidence interval for mean residual time at 365.25 days is [234.49389, 323.1998].

For testing of the median residual time, we first need to code the g_θ function defined in (1) and then use `el.cen.EM2` to test.

```

> mygfun2 <- function(s, age, Mdage) {as.numeric(s <= (age+Mdage))-0.5*as.numeric(s <= age) -0.5}
> el.cen.EM2(x=time, d=status, fun=mygfun2, mu=0, age=365.25, Mdage=184.75)$Pval
[1] 0.1135797
> el.cen.EM2(x=time, d=status, fun=mygfun2, mu=0, age=365.25, Mdage=321.7499)$Pval
[1] 0.1192006

```

This imply a 90% confidence interval for the median residual time is [184.75, 321.7499]. Note we do not get an exact p-value of 0.1 here. For the smoothed quantile, first define a (linearly) smoothed g function, then find the confidence limits.

```

> mygfun22 <- function(s, age, Mdage) {
myfun7(s, theta=(age+Mdage), epi=1/20)-0.5*myfun7(s, theta=age, epi=1/20) -0.5 }
> myfun7 <- function(x, theta=0, epi) {
if(epi <= 0) stop("epi must > 0")
u <- (x-theta)/epi
return( pmax(0, pmin(1-u, 1)) ) }
> el.cen.EM2(x=time, d=status, fun=mygfun22, mu=0, age=365.25, Mdage=184.7416765)$Pval
[1] 0.1000000
> el.cen.EM2(x=time, d=status, fun=mygfun22, mu=0, age=365.25, Mdage=321.71153607)$Pval
[1] 0.1000000

```

ACKNOWLEDGEMENTS

Dr. Mai Zhou's research was supported in part by NSF grant DMS-0604920. Dr. Jong-Hyeon Jeong's work was supported in part by the National Institute of Health (NIH) Grants 5-U10-CA69974-09 and 5-U10-CA69651-11.

REFERENCES

1. Jeong JH, Jung SH, Costantino JP. Nonparametric inference on median residual life function. *Biometrics* 2008; **64**:157-163.
2. Berger RL, Boos DD, Guess FM. Tests and confidence sets for comparing two mean residual life functions. *Biometrics* 1988; **44**:103-115.
3. Zhao Y, Qin G. Inference for the mean residual life function via empirical likelihood. *Communications in Statistics: Theory and Methods* 2006; **35**:1025-1036.
4. Qin G, Zhao Y. Empirical likelihood inference for the mean residual life under random censorship. *Statistics and Probability Letters* 2007; **77**:549-557.
5. Thomas DR, Grunkemeier GL. Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association* 1975; **70**:865-871.
6. Owen A. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 1988; **75**:237-249.
7. Owen A. *Empirical Likelihood*. Chapman & Hall, London, 2001.
8. Murphy S, van der Vaart A. Semi-parametric likelihood ratio inference. *Annals of Statistics* 1997; **25**:1471-1509.
9. Pan XR, Zhou M. Empirical likelihood ratio in terms of cumulative hazard function for censored data. *J. Multivariate Analysis* 2002; **80**:166-188.
10. Zhou M. Empirical likelihood ratio with arbitrarily censored/truncated data by EM algorithm. *Journal of Computational and Graphical Statistics* 2005; **14**:643-656.
11. Chen SX, Hall P. Smoothed empirical likelihood confidence-intervals for quantiles. *Annals of Statistics* 1991; **21**:1166-1181.
12. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>, 2008.