Empirical Likelihood Test for Mean Residual Life Time and Median Residual Life Time

Mai Zhou

Since a confidence interval can be obtained by inverting the tests, we shall focus on the test problem and only mention the confidence interval procedure in passing until the example section.

1 Mean Residual Life Time

First note the mean residual time, at a given age x, is defined as

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

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

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

is equivalent to the following hypothesis

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

and also equivalent to

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

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

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

Now this test can be done easily by a one sample empirical likelihood test for censored data, with a hypothesis of mean type using `el.cen.EM2( )` from the R contributed package `emplik`. We first define the function $g(s) = [s - (x + \mu)] I_{[s > x]}$ in R code as

```
mygfun <- function(s, age, muage) {
as.numeric(s >= age) * ( s - (age+muage) )
}
```

and then to test $M(50) = 25$ we do

```
el.cen.EM2(x=, d=, fun=mygfun, mu=0, age=50, muage=25)
```

Similarly, we can (1) test mean residual lifetime with doubly censored data, since `el.cen.EM2( )` can handle doubly censored data. (2) test if the mean residual life time at two given time points are the same. (3) for two independent samples (or k independent samples), test if the two mean residual life times at a given age are the same. (4) similar to (3) but we can test the ratio of the two mean residual lifetimes equal to a given τ value.

To accomplish the above generalizations (2), (3) and (4), we need the following lemma.

Lemma Suppose we have a statistic $\hat{\theta}$ which is a 2x1 vector. We further suppose $\sqrt{n}(\hat{\theta}-\theta_0)$ is (approx.) normally distributed with mean zero (zero vector) and a 2x2 variance-covariance matrix Σ; where θ_0 is a vector of two identical entries.

Suppose the matrix A is such that (so $A = (\Sigma)^{-1}$)

$$Q(\theta_0) = (\hat{\theta} - \theta_0)^T A(\hat{\theta} - \theta_0) \approx \chi^2_2$$

Then we have

$$\min_{\theta_0} Q(\theta_0) \approx \chi^2_1 \ .$$

Some generalizations:

(1) replace θ_0 vector by $f_1(\theta_0)$, and $f_2(\theta_0)$, two continuous differentiable functions. Same conclusion hold.

(2) vectors more than 2 dimension. If we begin with vector of k dimension, then the minimum over one θ_0 will lead to a reduction of $df = 1$. i.e. from $df = k$ to $df = k - 1$.

(3) vectors of more than 2 dim. Suppose we begin with vectors of dim k, but the mean has several parameters, i.e. $f_1(\theta_{01}, \theta_{02})$... $f_k(\theta_{01}, \theta_{02})$. then the min over two θ will lead to a reduction of $df = 2$, i.e. from $df = k$ to $df = k - 2$.

2 Median Residual Life

By definition the median residual lifetime (at age x) is the number θ that solve the following

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

Other quantiles of the residual life distribution can be defined similarly. The test developed below can easily modified to test a quantile. But we shall focus on the median here.

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

$$1 - F(x + \theta) = 0.5[1 - F(x)] \ .$$

After easy calculation we see that θ is the solution to

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

Define a function $g_\theta(t)$ as

$$g_\theta(t) = I_{[t \leqslant (x+\theta)]} - 0.5 I_{[t \leqslant x]}$$

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

$$H_0 : \quad \int_0^\infty g_\theta(t) dF(t) = 0.5 \ .$$

This in turn, can be accomplished by an empirical likelihood test. A confidence interval can be obtained as those θ values that results a p-value larger then 0.1, etc.

Two generalizations:

1. In the same setup as above, but we test if the median residual life at two different ages are the same. Notice that for exponential distributions, the median residual times at all ages are the same [memoryless].

For example, we may want to test $H_0 : Med(x_1) = Med(x_2)$ based on a censored sample.

Here we can make use of the Lemma in its full strength. First we test $H_{00} : \ Med(x_1) = \theta, \ \ Med((x_2) = \theta$ for some given θ value.

We know the value must be inside $[\hat\theta_1, \hat\theta_2]$ where $\hat\theta_i$ are the MLE of median residual lifetime of age x_1 and x_2 respectively.

Then we minimize the log empirical likelihood ratio over θ. The resulting likelihood ratio we get is a chi square with df=1 under H_0.

2. Consider two independent samples. For testing if the two median residual times are equal from two samples. We can make use of the Lemma. [the version with diagonal matrix A]. Here the log empirical likelihood ratio for two sample should be the sum of the two individual log likelihood ratios for one sample.

3 Median Residual Life Time in terms of hazard

We can easily write the median (or other quantiles) residual life hypothesis in terms of cumulative hazard function. Median residual life at age x is the value θ such that

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

Taking log on both sides, and notice cumulative hazard $H(s) = -\log[1 - F(s)]$, we have

$$H(x+\theta) - H(x) = -\log 0.5 = \log 2$$

The hypothesis that $H_0: \ Med(x) = \theta$ thus can be tested by testing

$$H_0: \int_x^{x+\theta} dH(t) = \log 2 \ .$$

The latter can be accomplished easily by empirical likelihood test in terms of hazard. The code/functions are in the R package emplik, and can handle right censored, left truncated data.

We need to define a function $g(t) = I_{[x \leqslant t < x+\theta]}$, and test the hypothesis $H_0: \ \int_0^\infty g(t) dH(t) = \log 2$. The theory of the empirical likelihood test for this hypothesis is spelled out in the paper (Pan and Zhou 1999).

```
gfun <- function(s, x, theta) {
as.numeric( (x <= s) & (s < (x+theta)) )
}
```

The implementation of the empirical likelihood for hazard has two versions: for continuous distribution/cumulative hazard and discrete distribution/hazards.

Continuous version: the R function is `emplikH2.test( )` . We may also use the two sample version, `emplikHs.test2( )`, with the second sample having zero weights. See the R help file of this function for more hints. The code may be more stable for two sample case.

Discrete version: the R function is `emplikH.disc( )` .

It is now clear, that we can test more complicated hypothesis:

(1) Other quantile of the residual life distribution.

(2) Test the equality of $Med(x_1) = Med(x_2)$, with one sample censored data. This will need to use the Lemma above.

(3) Test the equality of $Med(x)$ for two independent samples. This will need to use `emplikHs.test2( )` , the continuous version, or `emplikHs.disc2( )` the discrete version.

4 Examples and Simulations

We take the data set `cancer` from the R package `survival`. It contains 228 survival times of lung cancer patients with 63 right censored observations.

We shall find the 90% confidence interval for the mean (and median) residual life at one year (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 interval for mean residual life, it is often very helpful to know where is the 'center' of that confidence interval. We therefore also provide here the R code for estimating the Mean/Median Residual life time.

When testing for this value, you should get a P value of one, thus this is the 'center' of the confidence interval.

```
MMRtime <- function(x, d, age) {
#### estimate the Mean/Median Risidual lifetime over age. Depend on emplik.
temp <- WKM( x=x , d=d )
tivec <- temp$times
pivec <- temp$jump

if( age >= tivec[length(tivec)] ) stop("age too large")
if( age < tivec[1] ) warning("age smaller than first event time")

pivec[ tivec < age ] <- 0
Sage <- sum( pivec )

fenzi <- sum( (tivec - age)*pivec )
MRtime <- fenzi/Sage

Ptheta <- Sage/2
Cprob <- cumsum(pivec)
posi <- sum(Cprob < Ptheta)
theta <- tivec[posi+1]

list(MeanResidual = MRtime, MedianResidual = theta - age)
}
```

The following are R results when testing the mean/median residual times. (after loading the library emplik and survival)

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

$MedianResidual
[1] 258.75
```

For confidence interval of mean residual time we do

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

therefore the 90% confidence interval is [234.49389, 323.1998].

For the testing of median residual time, we first need to define a g function and then use `el.cen.EM2` to test.

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

Therefore the 90% confidence interval for the median residual time is [184.75, 321.7499]. Notice due to the discrete nature of the quantile function, we do not get exactly a p-value of 0.1. Some people have proposed smoothing technique to remedy this. One way of smoothing is to use a smoothed version of the g function we defined above. We leave this to reader.

The test of median residual time can also be done in terms of hazard, we also leave this to reader.

References

Guess, F. and Park, D.H. (1991) Nonparametric confidence bounds, using censored data, on the mean residual life. *IEEE Transactions on Reliability* **40**, 78-80.

Jeong, Jong-Hyeon, Jung, Sin-Ho and Costantino, Joseph P. (2008) Nonparametric Inference on Median Residual Life Function. *Biometrics* **64**, 157-163. March 2008

Zhao, Yichuan and Qin, Gengsheng (2006) Inference for the Mean Residual Life Function via Empirical Likelihood. *Communications in Statistics: Theory and Methods*, **35**, Number 6, 1025-1036.

Pan, X. R. and Zhou, M. (1999). Empirical Likelihood Ratio in terms of Cumulative Hazard Function for Censored Data. *J. Multivariate Analysis* **80**, 166-188.