Restricted mean survival time and confidence intervals by empirical likelihood ratio

Mai Zhou

Published online: 31 Dec 2020.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Check for updates

Restricted mean survival time and confidence intervals by empirical likelihood ratio

Mai Zhou

Department of Statistics, University of Kentucky, Lexington, Kentucky, USA

ABSTRACT
Recently, an increasing number of researchers propose to use restricted mean survival time (RMST) to evaluate medical treatment outcomes when the proportional hazards assumption is in doubt. Using the empirical likelihood ratio for the Kaplan-Meier mean, we show how the statistical analysis for the RMST and difference/ratio of two RMSTs can be obtained. Examples are given using R software and package emplik. We argue that the confidence intervals produced with this new method has several intrinsic advantages over the current method implemented in R package survRM2. A simulation is also included.

1. Introduction

When the proportional hazards assumption is in doubt, many researchers propose to use the restricted mean survival time (RMST) as a measure to evaluate/compare treatments with right-censored data. Many papers appeared recently (Royston and Parmar 2011), (Zhao et al. 2012), (Uno et al. 2014), (Kim et al. 2017), (Abulizi et al. 2019) to name a few, developing the theory and illustrating the usefulness of the new statistic. The nonparametric RMST estimator, $\widehat{RMST}$, used by the above papers are all defined by integrated Kaplan-Meier estimator (Kaplan and Meier 1958). Tests and confidence intervals of RMST proposed are of the Wald type. To construct a 95% confidence interval, you first estimate the asymptotic variance of $\widehat{RMST}$ and form the confidence interval by $\widehat{RMST} \pm 1.96$ (*estimated standard deviation*). A more sophisticated analysis may apply a transformation before constructing the confidence interval. Software to do these calculations are available, for example, in the R package survRM2 (Uno et al. 2015).

Since the Kaplan-Meier estimator is a nonparametric maximum likelihood estimator (NPMLE), the $\widehat{RMST}$ proposed in the above papers is also a NPMLE. There are several well-known methods to construct confidence intervals based on MLEs or likelihood functions, see for example (Meeker and Escobar 1995). We propose and illustrate here the Wilks method of testing and constructing confidence intervals for the RMST (also the difference/ratio of two RMSTs). In the paper (Meeker and Escobar 1995), the Wilks method is recommended over the Wald method. What is new here is that the parametric likelihood function is replaced by a nonparametric (empirical) likelihood, and the observed data are right-censored. For discussion of empirical likelihood method see (Owen 2001), (Zhou 2016).

When sample size goes to infinite, the Wilks and the Wald confidence intervals are equivalent. However, for smaller samples Wilks intervals are often better. It is a generally accepted fact that the Wilks confidence intervals have several advantages over the Wald confidence intervals, see section 3 of (Meeker and Escobar 1995) and additional references there. The disadvantage they mentioned for Wilks is the computational difficulty. But with ever faster computers, innovative algorithms (Zhou

---

2005), (Zhou and Yang 2015), and publicly available software, this is much less of a problem nowadays.

We list here briefly the advantages of Wilks over Wald: 1). The Wilks confidence intervals are not necessary symmetric about the MLE, rather, it tries to reflect the skewness of the given data. 2). The Wilks confidence intervals are always within the parameter space, while a Wald confidence interval of a probability can include negative values, for example. 3). Once we obtained the Wilks confidence interval for a parameter θ, $[a, b]$ (say), the Wilks confidence interval for $g(\theta)$ is just $[g(a), g(b)]$ (assuming g is increasing). 4). There is no need to figuring out the variance of the MLE/NPMLE and estimate it.

The confidence intervals for survival probability based on the Kaplan-Meier estimator in SAS proc lifetest is Wald based. It offers several options of transformation: linear, log, logit, log-log, and arcsine-sqrt. This is partly due to an attempt of correcting 2) above. But there is no clear conscienses on what is the best transformation. R package survival offers similar options. If we use the Wilks method, these are non-issues. For the RMST problem at hand, the package survRM2 used a log transformation when working on the ratio of two RMSTs, but did not use any transformation when working with the difference of two RMSTs. In one of the above papers, they commented that it is not clear what transformation should be used on the difference of two RMSTs. We do not have to worry about transformation on the parameter when apply the Wilks method. It automatically applied the best transformation without us knowing it, so to speak.

2. Empirical likelihood ratio test

In survival analysis, many parameters of interest can be expressed as an integral with respect to the underlying survival function. Its natural estimator is then the same integral by replacing the unknown survival function with the Kaplan-Meier estimator, $\hat{F}_{km}(t)$. If this is the case, then we can perform a test and find its Wilks confidence interval by using the empirical likelihood ratio. The theoretical base of the empirical likelihood ratio tests and subsequent test-based confidence intervals are the following theorem.

Suppose $X_1, \cdots, X_n \sim F$ are i.i.d. lifetimes. However, they are subject to censoring and we only observe $T_i = \min(X_i, C_i)$, $\delta_i = I[X_i \leq C_i]$, where C_i are censoring times. In addition, we assume the censoring is noninformative.

Theorem 1 (Empirical Likelihood Theorem for the Kaplan-Meier Integral) Let $(T_1, \delta_1), \cdots, (T_n, \delta_n)$ be n pairs of i.i.d. right-censored observations. Suppose $g(t)$ is a given function such that the asymptotic variance of the Kaplan-Meier mean $\int g(t) d\hat{F}_{km}(t)$ is positive and finite:

$$0 < AsyVar\left[\sqrt{n}\int g(t) d\hat{F}_{km}(t)\right] < \infty.$$

Then, under null hypothesis $H_0 : \int g(t) dF(t) = \mu$, we have,

$$-2\log R(\mu) = -2\log \frac{\sup\limits_{F:\ satisfy\ H_0} EL(F)}{EL(F = \hat{F}_{km})} \xrightarrow{D} \chi_1^2 \text{ as } n \to \infty,$$

where $EL(F)$ is the empirical likelihood for F based on the n right-censored data:

$$EL(F) = \prod_{i=1}^{n} [\Delta F(t_i)]^{\delta_i} [1 - F(t_i)]^{1-\delta_i}.$$

PROOF: See (Zhou 2016) Theorem 23 for a proof with a vector g. An easier to read version for univariate g and continuous F can be found in the Appendix.

A direct consequence of the Theorem 1 is that the p-value for testing $H_0 : \int g(t)dF(t) = \mu$ vs. $H_a : \int g(t)dF(t) \neq \mu$ can be computed as $P(\chi_1^2 > -2\log R(\mu))$ for large samples; where we denote by χ_1^2 a chi-square random variable with one degree of freedom. If we can test (obtain p-values) for various simple hypotheses by empirical likelihood ratio, then we will be able to construct a confidence interval:

$$95\% \text{ confidence interval} = \{\mu| -2\log R(\mu) \leq \chi_1^2(0.95)\}.$$

In words, the confidence interval/region consists of those values of the parameter μ that the simple hypothesis cannot be rejected. Those intervals are called Wilks confidence intervals.

The calculation of the log empirical likelihood ratio in the above theorem was studied in (Zhou 2005) and (Zhou and Yang 2015). The actual calculation in this paper is implemented in the R package emplik.

3. Restricted mean survival time

When the largest survival time in a sample is right-censored, the Kaplan-Meier estimator is non-uniquely defined beyond the last observed time. Thus the mean survival time estimator based on the Kaplan-Meier estimator is also non-uniquely defined. Some authors define the Kaplan-Meier to be zero beyond the largest censored observation in this case. We shall also take this convention.

The restricted mean survival time is used often in the comparison of cancer treatments when the proportional hazards assumption is in doubt. The idea is to compare the mean survival only on a limited time horizon, like 0 to 85 years, etc. By definition, the (true) RMST is

$$\mu(\tau) = \int_0^\tau 1 - F(s)ds$$

where τ is a pre-specified restriction time. We notice $\mu(\infty)$ is the ordinary mean value. After an integration by parts we can rewrite the true RMST as

$$\mu(\tau) = \int_0^\infty \min(s, \tau)\, dF(s).$$

An obvious estimate of the RMST is to replace $F(s)$ by the Kaplan-Meier $\hat{F}_{km}(s)$, so the estimate becomes (assume we always have Kaplan-Meier taking the value zero beyond the largest observation, which is the case for the Kaplan-Meier estimator in the package emplik)

$$\hat{\mu}(\tau) = \int_0^\infty \min(s, \tau)\, d\hat{F}_{km}(s).$$

The discussion above show that if we identify $g(t) = \min(t, \tau)$ then the empirical likelihood theorem in section 2 leads to a likelihood ratio test and a Wilks confidence interval for RMST.

For more discussions and references of RMST, in addition to the references mentioned in the introduction, please see the R package survRM2. Unlike the R package survRM2 which is specifically made to do statistical inferences for the RMST, the R package we are going to use, emplik, is made for empirical likelihood ratio type analysis in general. Therefore, some simple auxiliary functions need to be defined for the specific analysis of RMST. We illustrate those in the following examples.

Example 1. We use the data set kidney from the R package KMsurv to illustrate our method. The data set contains 119 cases of kidney dialysis patient and their catheter infection time is recorded. The maximum time recorded is 28.5 month. We pre-select the time restriction $\tau = 20$ (months). Suppose we are trying to test the hypothesis $H_0: \mu(20) = 15 vs. H_a : \mu(20) \neq 15$.

The actual R code and calculation is in Appendix. We see that the *p*-value of the test is 0.1125004. The existing function from the package emplik we used is el.cen.EM2. The newly defined (one-line) function for testing RMST is RMSTfun.

Next, we try to construct the 95% Wilks confidence interval for RMST. This is accomplished by the function findUL from the package emplik, which in turn, requires a function that returns log likelihood in testing a hypothesis about theta. Therefore, we need to define the (new) wrapper function myfun8. See Appendix for details.

We see the 95% confidence interval for $\mu(20)$ is [14.71036, 17.38996]. The step input of the function findUL is the initial search step size from the MLE. We set the option step = 1 in findUL. Ideally, the step should be similar to 1/2 of the standard deviation of the MLE. A wide range of step sizes should work. A step too small makes the calculation slower, while a step too large can whack the calculation.

The input MLE in the findUL is supposed to be the MLE of the parameter that we are trying to find its confidence interval. In our case, it is the MLE of RMST. But it does not have to be very accurate, a ballpark value will work. Here we used 15. May be 16 is the better choice. But the final confidence interval is the same.

A good indication that everything is OK, is in the output L value and U-value. They both should be equal to $\chi^2(1 - \alpha)$. Since the default confidence level is 95%; thus, in the above example they have value close to $3.84 = \chi_1^2(0.95)$.

If you want a 90% confidence interval (or other levels), you can add an input level = 2.705543 (that is chi square 90% percentile: qchisq(0.9, df = 1)) to the function find UL:

findUL (step = 1, fun = myfun8, MLE = 16, level = 2.705543, x = kidney$time, d = kidney$delta)

If we are dealing with only right censored data (as in the Kaplan-Meier estimator), a much faster version of the calculation (replacing el.cen.EM2) is available in the package kmc, function kmc.solve, see (Zhou and Yang 2015). But the package kmc is not as mature/stable as emplik, reflected in its version number, and short history.

Notice el.cen.EM2 can also deal with data that have both right censor and left censor; kmc.solve can only deal with right-censored data.

The smallest observation fed into kmc.solve must be an uncensored one. If you do have some censored observations that are smaller than any observed death, simply delete them. It is easy to show that those smallest censored observations do not carry any information in nonparametric modeling. Last, the kmc.solve only tests the mean equal to zero, so if you want to test other mean values, you need to first subtract it in the definition of the function.

By using the R package kmc to test the RMST, we get the same result but the speed is vastly improved. See Appendix. We caution that the code in kmc is less stable.

4. Difference/ratio of two restricted mean survival times

It is possible to formulate and prove a similar empirical likelihood theorem for the difference of two Kaplan-Meier means when dealing with two independent right-censored samples and another one for the ratio. However, we shall take a simpler and more general approach here. We first introduce an extra (nuisance) parameter so that under the new null hypothesis of two parameters, the log empirical likelihood ratio for the two-samples are completely separate. This makes the calculation of the log empirical likelihood for two-sample easier: it is the sum of the two one-sample log empirical likelihood ratio. In addition, Theorem 1 in section 2 applies to each of them, leading to an asymptotic distribution of $\chi_1^2 + \chi_1^2 = \chi_2^2$ by independence. The last step is to profile out the nuisance parameter we introduced in the first place.

To profile out the nuisance parameter is a common approach in parametric likelihood ratio analysis. The fact that it is also valid for empirical likelihood is of no surprise. See Owen (Owen 2001) Chapter 3. The novelty here is that we artificially introduced the nuisance parameter, only to be profiled out later.

Theorem 2 (Profile Empirical Likelihood) Suppose θ, v are two parameters with true values denoted by θ_0, v_0, and the empirical likelihood ratio $-2 \log R(\theta_0, v_0) \to \chi_2^2$ in distribution. Then, under mild regularity conditions, the profile log empirical likelihood ratio: $-2 \log \max_v R(\theta_0, v)$ has an asymptotic χ_1^2 distribution.

Proof: See (Qin and Lawless 1994) Corollary 5 for proof and exact conditions needed. Those smoothness conditions on the estimating functions are obviously satisfied here.

4.1. Difference of two RMSTs

Here we try to find the confidence interval for the difference of two RMSTs from two independent samples.

Let us first look at the testing of $H_0 : \mu_1(\tau) - \mu_2(\tau) = \theta$. After introducing a nuisance parameter $r = RMST \ of \ sample \ 1$, the null hypothesis about θ and r, $H_{01} : \mu_1(\tau) = r, \mu_1(\tau) - \mu_2(\tau) = \theta$, is equivalent to $H_{02} : \mu_1(\tau) = r, \mu_2(\tau) = r - \theta$. This H_{02} is two separate hypotheses about each of the sample; it can be solved one by one. The joint log likelihood is just the sum of the two (one sample) log empirical likelihoods.

Example 2 We use data from a study of treatments (D-penicillamine verses placebo) for primary biliary cirrhosis conducted at Mayo clinic to illustrate our method. In order to compare the results to those of R package survRM2, we use exactly the same data cases as theirs. There were 184 subjects in the treatment group, and 182 subjects in the placebo group. The censoring percentage is 41% for treatment group and 39% for placebo group, with maximum-observed time of 4556 and 4523 (days) respectively.

Here we take the restriction time $\tau = 10$ years = 3652.5 days.

For a simple test of hypothesis $H_0 : \mu_1(3652.5) - \mu_2(3652.5) = 0$ we can do, after defining the functions RMSTdiff and ThetafunD, the following to obtain the p-value.

ThetafunD(theta =0, x1 =pbcX1, d1 =pbcd1, x2 =pbcX2, d2 =pbcd2)

We omit the output to save space.

Next we find a 95% confidence interval for the difference $\mu_1(3652.5) - \mu_2(3652.5)$. See Appendix for the R code and output.

Clearly this coding consists of three blocks. The first block is the RMSTdiff function, which introduced a nuisance parameter (RMST of sample one, r) to make the calculation of the empirical likelihood easier. It then computes the log-likelihood ratio from each sample and returns the summation.

The second block, ThetafunD, is to profile out the nuisance parameter r. It is a wrapper for the R function optimize. Two points of caution: in some cases, the log-likelihood ratio from the first bloc may not be smooth [however, for the RMST function here, it is smooth]. An example of a non-smooth log likelihood is the testing of quantile/medians. If that is the case, we have to use an exhaustive search or optimize will have to be applied after some smoothing. Another point is the lower and upper bound of the search. This must take into account the value and approximate confidence interval for the nuisance parameter. Here the nuisance parameter is RMST of sample 1. It has a value of approximate 2612 and (95%) confidence interval approximately 2400 to 2815. This makes us use the lower = 1500 days and upper = 3100 days. When in doubt, a larger range can be used. The drawback is the search will be slower.

The third block is the findUL function to construct confidence interval. Here we know roughly the width of the final confidence interval should be at least 400 days. So we use step =100. The final 95% confidence interval of the difference, RMST1 $-$ RMST2 is: $[-340.4272, 243.4025]$. Please note this is in terms of days. To compare to the results from survRM2 we need to convert them to the years, which is $[-340.4272/365.25, 243.4025/365.25] = [-0.9320389, 0.6663997]$ (years). As a comparison, the

Wald confidence interval from the R package survRM2 example is $[-0.939, 0.665]$. Our confidence interval is shorter and is not symmetric about the MLE ($= -0.137$). As discussed in section 1, the advantage of Wilks confidence intervals should be more profound over Wald if the distribution of $\hat{\mu}_1(\tau) - \hat{\mu}_2(\tau)$ is more skewed.

4.2. Ratio of two RMSTs

We now try to find the confidence interval for the ratio of two RMSTs from two independent samples.

Example 3 Using the same data from section 4.1 example 2, we first consider testing $H_0 : \mu_1(\tau)/\mu_2(\tau) = \theta$. Here we introduced an extra nuisance parameter $r =$ the RMST of sample 2. The new null hypothesis before profiling is $H_{02} : \mu_2(\tau) = r, \mu_1(\tau) = r\theta$.

The R code and sample output are in Appendix.

Again, there are three steps: (1) find the log empirical likelihood ratio for testing H_{02} by RMSTratio. (2) profile the nuisance parameter r out (producing a p-value for testing H_0). (3) find the confidence interval using findUL.

The nuisance parameter in the difference of two RMSTs case is RMST1, and here the nuisance parameter is RMST2. However, the two RMSTs are not too far apart and the range $[1500, 3100]$ is wide enough so that it works for both case in the optimize function in step 2.

The true MLE of θ here is not 1, but it is close enough. We therefore take the input MLE $=1$ in findUL function. A reasonable choice of the step input in findUL function for ratio is anywhere between 0.1 and 0.2. Please note this is the 95% confidence interval for the ratio of RMST1/RMST2. In this case it is $[0.8782903, 1.09761]$. This is almost the same as the interval given by the R package survRM2: $[0.878, 1.096]$.

One nice property of the confidence interval obtained here is that for the confidence interval of the reciprocal of the ratio above, RMST2/RMST1, is just $[1/1.09761, 1/0.8782903]$ due to invariance.

5. Simulation

We have conducted a small simulation to show the empirical likelihood test, although an asymptotic test, performs pretty well for moderate sample sizes. In Table 1 for both samples the lifetimes X_i are generated from a Weibull distribution with shape parameter 0.9 and scale parameter 12.7. The censoring times C_i are generated from a uniform $(0, 12.5)$ distribution.

Simulated type I errors are based on 5000 runs. We have used the t-distribution calibration for empirical likelihood as suggested by Owen (Owen 2001) Equation (3.3): i.e., the $-2 \log R$ is deemed 5% significant if it is larger than $[qt(0.975, df = n1 + n2 - 2)]^2$, where $qt()$ denotes the quantile function of a student t-distribution.

The type I errors for empirical likelihood are slightly inflated compared to the nominal, but they are less profound than the results of asymptotic Wald test. Even without the t-distribution calibration, the EL test's type I error in sample sizes 30 +30, $\tau = 8$ case would be 5.92% vs. Wald test 6.52%, and for 30 +30, $\tau = 10$, 5.73% vs. Wald test 6.44%, etc.

Table 1. Nominal versus actual type I error for empirical likelihood (EL) test and Wald test.

sample sizes	restriction time	EL test (5%)	Wald test (5%)
25, 25	$\tau = 8$	5.74%	7.1%
25, 25	$\tau = 10$	6.12%	7.02%
30, 30	$\tau = 8$	5.30%	6.52%
30, 30	$\tau = 10$	5.73%	6.44%
40, 40	$\tau = 8$	5.30%	6.02%
40, 40	$\tau = 10$	5.36%	6.08%

Table 2. Nominal versus actual error.

sample sizes	restriction time	distributions	EL test (5%)	Wald test (5%)
35, 25	$\tau = 10$	H_0	6.04%	6.51%
35, 25	$\tau = 10$	H_A	6.12%	6.83%

For the same sample size, we see that when the restriction time are higher ($\tau = 10$), the type I errors (of both test) are less accurate than a lower restriction time ($\tau = 8$). This may due to the larger variance of Kaplan-Meier in the tail.

In Table 1, the probability of an observation end up being uncensored and occurs before restriction time $\tau = 8$ ($\tau = 10$ respectively) is about 35.2% (37.2%). The probability of an observation being larger than $\tau = 8$ ($\tau = 10$) is about 18.6% (8.8%).

We used a Dell OptiPlex 9020 desktop computer with an i7-4790 CPU for simulation. The computational time of a p-value for testing the difference of two RMSTs is about 1.1 seconds, for sample size (200, 200) using the empirical likelihood method; and about 0.9 seconds for sample size (50, 50). The Wald test is obtained using the R package survRM2, and is almost instant.

In Table 2, we also carried out another simulation in the un-equal sample size setup and included an alternative setting, where for the first sample the lifetimes are generated from a Weibull (1.9, 12.7) distribution and the rest remain the same as in Table 1. In the alternative case, the error is "percent of nominal 95% confidence intervals NOT catching the true parameter value".

We see the same improvements on the finite sample accuracy of the empirical likelihood (EL) tests over the Wald tests.

6. Discussion

The empirical likelihood method avoids the need for calculating/estimating the variance of the RMST. Also, no transformation of the parameter needs to be considered. On the other hand, we are left with some optimization problems, which (at least in this case) can easily be solved by a computer. In this respect, it is similar to the bootstrap method: calculation/estimating the variance is replaced by computer algorithms. But the empirical likelihood method always gives you the same p-value testing a given null hypothesis while the bootstrap result varies depending on the random number generator and the number of bootstrap iterations.

The empirical likelihood method is highly recommended over Wald method when (1) the sample sizes are small to moderate; (2) the sampling distribution of the estimator, $\hat{\mu}_1(\tau) - \hat{\mu}_2(\tau)$, is skewed; (3) either a variance estimator is not available or there are several competing ones and no clear winner.

We would like to point out that the variance estimator used by R package survRM2 (Horiguchi and Ono 2020) and by SAS proc lifetest (Guo and Liang 2019) differs by a factor $m/(m-1)$ where m is the total number of observed events before restriction time τ.

The restricted mean time lost (RMTL) is related to RMST: $RMTL = \tau - RMST$. The statistical analysis of RMTL and the difference of two RMTLs can be easily obtained from those for the RMST by simple arithmetic. The analysis for the ratio of two RMTLs can be similarly done by the same method illustrated in this paper. Notice that we have $RMTL(\tau) = \int_0^\infty \max(\tau - s, 0) dF(s)$. We need to replace the R function RMSTfun by RMTLfun defined by: RMTLfun <- function(x) {pmax(600-x, 0) – 100} for testing RMTL = 100, with $\tau = 600$, for example.

The empirical likelihood method can also be used to test the hypothesis of several RMSTs with different restriction times. This amounts to use several g's to impose several constraints.

References

Abulizi, X., H. J. Ribaudo, and P. Flandre. 2019. The use of the restricted mean survival time as a treatment measure in HIV/AIDS clinical trial: Reanalysis of the ACTG A5257 trial. *Journal of Acquired Immune Deficiency Syndromes* 81 (1):44–51. doi:10.1097/QAI.0000000000001978.

Guo, C., and Y. Liang. 2019. Analyzing restricted mean survival time using SAS/STAT. SAS global forum proceedings Paper SAS3013-2019, SAS Institute Inc.

Horiguchi, M., and H. Ono. 2020. On permutation tests for comparing restricted mean survival time with small sample from randomized trials. *Statistics in Medicine* 39 (20):2655–2670. doi:10.1002/sim.8565.

Kaplan, E., and P. Meier. 1958. Nonparametric estimation from incomplete observations. *Journal of Tha American Statistical Association* 53:457–481. doi:10.1080/01621459.1958.10501452.

Kim, D. H., H. Uno, and L. J. Wei. 2017. Restricted mean survival time as a measure to interpret clinical trial results. *JAMA Cardiology* 2 (11):1179–1180. doi:10.1001/jamacardio.2017.2922.

Meeker, W. Q., and L. Escobar. 1995. Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician* 49 (1):48–53.

Owen, A. 2001. *Empirical likelihood.* New York: Chapman & Hall/CRC Press.

Qin, J., and J. Lawless. 1994. Empirical likelihood and general estimating equations. *Annals of Statistics* 22:300–325. doi:10.1214/aos/1176325370.

Royston, P., and M. K. B. Parmar. 2011. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine* 30:2409–2421. doi:10.1002/sim.4274.

Uno, H., B. Claggett, L. Tian, E. Inoue, P. Gallo, T. Miyata, D. Schrag, M. Takeuchi, Y. Uyama, L. Zhao, et al. 2014. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of Clinical Oncology* 32:2380–2385. doi:10.1200/JCO.2014.55.2208.

Uno, H., L. Tian, A. Cronin, and C. Battioui. 2015. survRM2: Comparing restricted mean survival time. R package version 1.0-1. https://CRAN.R-project.org/package=survRM2.

Zhao, L., L. Tian, H. Uno, S. D. Solomon, M. A. Pfeffer, J. S. Schindler, and L. J. Wei. 2012. Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study. *Clinical Trials* 9:570–577. doi:10.1177/1740774512455464.

Zhou, M. 2005. Empirical likelihood ratio with arbitrarily censored/truncated data by EM algorithm. *Journal of Computational and Graphical Statistics* 14 (3):643–656. doi:10.1198/106186005X59270.

Zhou, M. 2016. *Empirical likelihood methods in survival analysis.* Boca Raton: CRC Press, Taylor & Francis Group.

Zhou, M., and Y. F. Yang. 2015. A recursive formula for the Kaplan-Meier estimator with mean constraints and its application to empirical likelihood. *Computational Statistics* 30 (4):1097–1109. doi:10.1007/s00180-015-0567-9.

Appendix A. Example R code for calculation

R calculation for Example 1. Obtain *p*-value.

```
library(KMsurv)
library(emplik)
data(kidney)
RMSTfun <- function(x) {pmin(x, 20) – 15}
el.cen.EM2(x = kidney$time, d = kidney$delta, fun = RMSTfun, mu = 0, maxit = 50)
## $loglik
## [1] −131.105
##
## $times
## [1] 0.5 0.5 0.5 0.5 0.5 0.5 1.5 2.5 2.5 3.5 3.5 4.5 4.5 5.5 6.5 8.5 8.5
## [18] 9.5 10.5 11.5 15.5 15.5 16.5 18.5 23.5 26.5 28.5
##
## $prob
## [1] 0.01122308 0.01122308 0.01122308 0.01122308 0.01122308 0.01122308 0.01243132 0.01279473
## [9] 0.01279473 0.01381316 0.01381316 0.01533046 0.01533046 0.01599773 0.01746515 0.02043914
## [17] 0.02043914 0.02186346 0.02285866 0.02485554 0.03538362 0.03538362 0.03297763 0.03220609
## [25] 0.06200016 0.10114908 0.39333447
##
## $lam
## [1] 2.061926
##
## $iters
```

```
## [1] 50
##
## $'-2LLR'
## [1] 2.51874
##
## $Pval
## [1] 0.1125004
```
Construct confidence interval for Example 1.
```
myfun8 <- function(theta, x, d){
el.cen.EM2(x,d,fun = function(t){pmin(t,20)-theta},mu = 0,maxit = 50)
}
findUL(step = 1, fun = myfun8, MLE = 15, x = kidney$time, d = kidney$delta)
## $Low
## [1] 14.71036
##
## $Up
## [1] 17.38996
##
## $FstepL
## [1] 6.103516e-05
##
## $FstepU
## [1] 6.103516e-05
##
## $Lvalue
## [1] 3.839861
##
## $Uvalue
## [1] 3.839985
```
Example 1 Calculation using R package kmc version 0.2–3. First, find the p-value.
```
library(KMsurv)
library(kmc)
data(kidney)
RMSTfun <- function(x) {pmin(x, 20) – 15}
kmc.solve(x = kidney$time, d = kidney$delta, g = list(RMSTfun))
```
Example 1 Using package kmc version 0.2–3 to Find the confidence interval:
```
myfun9 <- function(theta, x, d){
ff <- function(t){pmin(t, 20) – theta}
kmc.solve(x, d, g = list(ff))
}
findUL(step = 1, fun = myfun9, MLE = 16, x = kidney$time, d = kidney$delta)
## $Low
## [1] 406.4569
##
## $Up
## [1] 535.4033
##
## $FstepL
## [1] 6.103516e-05
##
## $FstepU
## [1] 6.103516e-05
##
## $Lvalue
## [1] 3.84
##
## $Uvalue
## [1] 3.84
```
R code for example 2
```
mypbc <- pbc[1:312,]
```

```
pbcX1 <- mypbc$time[mypbc$trt = = 1]
pbcX2 <- mypbc$time[mypbc$trt = = 2]
pbcD1 <- mypbc$status[mypbc$trt = = 1]
pbcD2 <- mypbc$status[mypbc$trt = = 2]
pbcd1 <- as.numeric(pbcD1 = = 2)
pbcd2 <- as.numeric(pbcD2 = = 2)
RMSTdiff <- function(r, x1, d1, x2, d2, theta){
temp1 <- el.cen.EM2(x = x1, d = d1, fun = function(x){pmin(x, 3652.5)}, mu = r)
temp2 <- el.cen.EM2(x = x2, d = d2, fun = function(x){pmin(x, 3652.5)}, mu = r-theta)
return(temp1$"-2LLR" + temp2$"-2LLR")
}
ThetafunD <- function(theta, x1, d1, x2, d2) {
temp <- optimize(f = RMSTdiff, lower = 1500, upper = 3100, x1 = x1, d1 = d1, x2 = x2, d2 = d2, theta = theta)
cstar <- temp$minimum
val <- temp$objective
list("-2LLR" = val, cstar = cstar, Pval = 1-pchisq(val, df = 1))
}
findUL(step = 100, fun = ThetafunD, MLE = 0, x1 = pbcX1, d1 = pbcd1, x2 = pbcX2, d2 = pbcd2)
## $Low
## [1] −340.4272
##
## $Up
## [1] 243.4025
##
## $FstepL
## [1] 1e-06
##
## $FstepU
## [1] 1e-06
##
## $Lvalue
## [1] 3.84
##
## $Uvalue
## [1] 3.84
R code and sample out put for Example 3.
RMSTratio <- function(r, x1, d1, x2, d2, theta){
temp1 <- el.cen.EM2(x = x1,d = d1,fun = function(x){pmin(x, 3652.5)},mu = r*theta)
temp2 <- el.cen.EM2(x = x2,d = d2,fun = function(x){pmin(x, 3652.5)},mu = r)
return(temp1$"-2LLR" + temp2$"-2LLR")
}
ThetafunR <- function(theta, x1, d1, x2, d2) {
temp <- optimize(f = RMSTratio, lower = 1500, upper = 3100, x1 = x1, d1 = d1, x2 = x2, d2 = d2, theta = theta)
cstar <- temp$minimum
val <- temp$objective
list("-2LLR" = val, cstar = cstar, Pval = 1-pchisq(val, df = 1))
}
findUL(step = .2, fun = ThetafunR, MLE = 1, x1 = pbcX1, d1 = pbcd1, x2 = pbcX2, d2 = pbcd2)
## $Low
## [1] 0.87829
##
## $Up
## [1] 1.097608
##
## $FstepL
## [1] 6.103516e-05
##
## $FstepU
## [1] 6.103516e-05
##
```

```
## $Lvalue
## [1] 3.840018
##
## $Uvalue
## [1] 3.839883
```

Appendix B. Empirical Likelihood Theorem for Right Censored Data: Proof

We shall assume the underlying survival function $1 - F$ is continuous for easier reading. The feature of the distribution (or survival) function we are testing (mean value of $g(X)$) is specified by $\int g(t)dF(t)$, i.e. $H_0 : \int g(t)dF(t) = \mu$. So at a minimum we need to assume the asymptotic variance of this statistics, $AsyVar\sqrt{n}[\int g(t)d\hat{F}_{km}(t) - \mu]$, is finite. WLOG, we assume the true $\mu = 0$.

1st Step. The first major step is to re-write the asymptotic variance of Kaplan-Meier mean $\int g(t)d\hat{F}_{km}$. There are several equivalent expressions for the variance, but we shall focus on one that is given by Akritas (2000).

Assume F is continuous, we have

$$AsyVar[\sqrt{n}\int g(t)d\hat{F}_{km}(t)] = \int [g(t) - \bar{g}(t)]^2 \frac{dF(t)}{1 - G(t-)} \tag{1}$$

where G is the censoring distribution.

We call the attention of readers to the use of $\bar{g}$, the advanced time:

$$\bar{g}(s) = \frac{\int_{(s,\infty)} g(x)dF(x)}{1 - F(s)}. \tag{2}$$

Actually, we should probably write it as $\bar{g}_F(s)$ instead of $\bar{g}(s)$ since the definition uses F. The reference of the advanced time is Akritas (2000), Efron and Johnstone (1990). The nice thing about this variance formula (1) is that it allows us to use the Cauchy-Schwarz inequality to show some maximum result. After all, the empirical likelihood ratio is a " **maximum** likelihood ratio".

2nd step Let us recall the Cauchy-schwarz inequality. The integral version of it is

$$\left[\int_A \xi(x)\eta(x)dv\right]^2 \leq \int_A \xi^2(x)w^2(x)dv \int_A \frac{\eta^2(x)}{w^2(x)}dv$$

here dv is a measure on the set A, w is positive. The max of left hand side is achieved if and only if $\xi w = \eta/w$ (a.s. dv), and easy to check, when this happens we have the equality holds.

We will be using the inequality with $\eta(x) = [g - \bar{g}]$; $\xi(x) = [f - \bar{f}]$; $w = \sqrt{1 - G_-}$ and $dv = dF$.

So, the actual inequality we shall use to find the maximum of empirical likelihood ratio is

$$\int \frac{[g - \bar{g}]^2}{1 - G_-}dF \geq \frac{\left[\int (f - \bar{f})(g - \bar{g})dF\right]^2}{\int [f - \bar{f}]^2[1 - G_-]dF}.$$

We notice the left hand side of this inequality is the Asymptotic Variance of the Kaplan-Meier mean.

Another way to express the Cauchy-Schwarz inequality above is

$$\sup_f \frac{\left[\int (f - \bar{f})(g - \bar{g})dF\right]^2}{\int [f - \bar{f}]^2[1 - G_-]dF} = \int \frac{[g - \bar{g}]^2}{1 - G_-}dF.$$

Or

$$\inf_f \frac{\int [f - \bar{f}]^2 [1 - G_-] dF}{\left[\int (f - \bar{f})(g - \bar{g}) dF\right]^2} = \frac{1}{\int \frac{[g - \bar{g}]^2}{1 - G_-} dF}. \tag{3}$$

Third Step. Since the empirical likelihood ratio is defined via a constrained maximum, we first take care the constrain. We "arrange" (for lack of a better word) the constrained maximum into two components, a scalar λ and a function f, where the scalar takes care of the constraint and the arbitrary f will be maximized by Cauchy-Schwarz inequality.

Recall the log empirical likelihood ratio statistic is

$$-2\log \frac{\max\limits_{\int g dF = 0} \prod\limits_{\delta_i = 1} \Delta F(t_i) \prod\limits_{\delta_i = 0} 1 - F(t_i)}{\prod\limits_{\delta_i = 1} \Delta \hat{F}_{km}(t_i) \prod\limits_{\delta_i = 0} 1 - \hat{F}_{km}(t_i)}. \tag{4}$$

Let us focus on the numerator of the above. The maximum is taken over all F that satisfy the constraint. We write the jumps of F to be maximized as

$$\Delta F(t_i) = \Delta \hat{F}_{km}(t_i)[1 - \lambda f(t_i)]. \tag{5}$$

This is to say that we only look at those F that are dominated by the Kaplan-Meier. This is just a tilt of the Kaplan-Meier in the **direction** f with **magnitude** λ. This will simplify the above expression, and before we go on we examine some of the requirements for f.

Since $\sum_{all} \Delta F(t_i)$ must be one, we must have $f(t_i)$ satisfy $\sum_{all} f(t_i) \Delta \hat{F}_{km}(t_i) = 0$.

Also the maximum require the constraint $\int g(t) dF(t) = 0$ (recall WLOG we assumed $\mu = 0$), this leads to an equation for the λ that satisfy

$$\lambda \sum_{all} g(t_i) f(t_i) d\hat{F}_{km}(t_i) = \sum_{all} g(t_i) d\hat{F}_{km}(t_i)$$

or we shall write the above as

$$\lambda A = gg,$$

where the definition of A and gg are defined by the above equation. Thus we have $\lambda = gg A^{-1}$.

Finally, (4) becomes

$$-2\log \max_{\int g dF = 0} \prod_{\delta_i = 1} [1 - \lambda f(t_i)] \frac{\prod_{\delta_i = 0} \sum_{s > t_i} \Delta \hat{F}_{km}(s_j)[1 - \lambda f(s_j)]}{\prod_{\delta_i = 0} 1 - \hat{F}_{km}(t_i)}$$

$$= -2 \max_f \log \prod_{\delta_i = 1} [1 - \lambda f(t_i)] \prod_{\delta_i = 0} \frac{\sum_{s > t_i} \Delta \hat{F}_{km}(s_j)[1 - \lambda f(s_j)]}{1 - \hat{F}_{km}(t_i)}.$$

Forth Step Here we calculate the derivatives and write out a Taylor expansion for

$$u(\lambda) = \log \prod_{\delta_i = 1} [1 - \lambda f(t_i)] \prod_{\delta_i = 0} \frac{\sum_{s > t_i} \Delta \hat{F}_{km}(s_j)[1 - \lambda f(s_j)]}{1 - \hat{F}_{km}(t_i)}.$$

$$u(\lambda) = u(0) + \lambda u'(0) + 1/2\lambda u''(0)\lambda + \cdots$$

Easy to check, $u(0) = \log(1) = 0$ and $u'(0) = 0$. (Without calculation, we can explain why $u'(0) = 0$: this log likelihood ratio $u(\lambda)$ achieves its maximum value 0 at $\lambda = 0$, therefore the derivative at $\lambda = 0$ must be 0.)

All then left to complete the above 3 term Taylor expansion is the calculation of $u''(0)$ which involves f. After this we can use Cauchy-Schwarz inequality of step 2 to find the maximum of u over f.

Long and tedious calculation show (see (Zhou 2016) for details)

$$u''(0)/n = -\sum [f - \bar{f}]^2 [1 - \hat{G}_{km}(t_i-)] \Delta \hat{F}_{km}(t_i).$$

The Taylor expansion for the $-2\log$ empirical likelihood ratio (before maximization) becomes

$$(\sqrt{n}\lambda)\frac{u''(0)}{n}(\sqrt{n}\lambda) + \cdots = (\sqrt{n}gg)A^{-1}\frac{u''(0)}{n}A^{-1}(\sqrt{n}gg) + \cdots.$$

Final Step is to use Cauchy-Schwarz to find the maximum of the middle part of above (which is the only place involving f)

$$A^{-1}\frac{u''(0)}{n}A^{-1} = \frac{\sum [f-\bar{f}]^2[1-\hat{G}_{km}(t_i-)]\Delta\hat{F}_{km}(t_i)}{[\sum gf\Delta\hat{F}_{km}(t_i)]^2} \tag{6}$$

and once we find the maximum achieved for this expression (6), the chi square limit is easy.

We recall the mean of f with respect to $\hat{F}_{km}$ is zero, so we may rewrite the denominator of above as the square of covariance between f and g (under Kaplan-Meier measure). Therefore,

$$\sum gf\Delta\hat{F}_{km}(t_i) = \sum [g-\bar{g}][f-\bar{f}]\Delta\hat{F}_{km}(t_i).$$

Here we used a formula of covariance with advanced time, similar to the variance formula we discussed earlier. This equality is proved in (Zhou 2016), Lemma 26.

Now use (3) to see that the minimum of $-2 \times \log$ likelihood ratio over f is

$$(\sqrt{n}gg)\frac{1}{\int \frac{(g-\bar{g})^2}{1-\hat{G}_{km}(t-)}d\hat{F}_{km}(t)}(\sqrt{n}gg) + \cdots.$$

This last expression, under null hypothesis, clearly has an asymptotic chi square distribution, with degree of freedom 1.

ADDITIONAL REFERENCE FOR THE APPENDIX

Akritas, M. (2000). The central limit theorem under censoring. *Bernoulli* 6: 1109–1120.

Efron, B. and Johnstone, I. (1990). Fisher's information in terms of the hazard rate. *Ann. Statist.*, 18, 38–62.