Noname manuscript No.
(will be inserted by the editor)

A Recursive Formula for the Kaplan-Meier Estimator with Mean Constraints

Mai Zhou · Yifan Yang

Received: date / Accepted: date

Abstract In this note we give a recursive computational algorithm for the 'constrained' Kaplan-Meier estimator. The constrain is assumed given in linear estimating equations or mean functions. We also illustrate how this leads to the empirical likelihood ratio test with right censored data. Speed comparison to the EM based algorithm favors the current procedure.

Keywords Wilks test · Empirical Likelihood Ratio · Right Censored Data · NPMLE

1 Introduction

Suppose that $X_1, X_2, \dots, X_n$ are i.i.d. nonnegative random variables denoting the lifetimes with a continuous distribution function F_0. Independent of the lifetimes there are censoring times $C_1, C_2, \dots, C_n$ which are i.i.d. with a distribution G_0. Only the censored observations, (T_i, δ_i), are available to us:

$$T_i = \min(X_i, C_i) \quad \text{and} \quad \delta_i = I[X_i \le C_i] \quad \text{for } \; i = 1, 2, \dots, n. \tag{1}$$

The empirical likelihood of the censored data in terms of distribution F is defined as

$$\begin{aligned} EL(F) &= \prod_{i=1}^{n} [\Delta F(T_i))]^{\delta_i} [1 - F(T_i)]^{1-\delta_i} \\ &= \prod_{i=1}^{n} [\Delta F(T_i))]^{\delta_i} \{ \sum_{j: T_j > T_i} \Delta F(T_j) \}^{1-\delta_i} \end{aligned}$$

where $\Delta F(t) = F(t+) - F(t-)$ is the jump of F at t. See for example [Kaplan and Meier(1958)] and [Owen(2010)]. The second line above *assumes* a discrete $F(\cdot)$. It is well known that the constrained or

Mai Zhou
Department of Statistics,University of Kentucky
Tel.: (859) 257-6912
Fax: (859) 323-1973
E-mail: mai@ms.uky.edu

Yifan Yang
Department of Statistics,University of Kentucky
Tel.: (859) 494-4072
E-mail: yifan.yang@uky.edu

unconstrained maximum of the empirical likelihood are both obtained by discrete F. Let $w_i = \Delta F(T_i)$ for $i = 1, 2, \ldots, n$. The likelihood at this F can be written in term of the jumps

$$EL = \prod_{i=1}^{n} [w_i]^{\delta_i} \{ \sum_{j=1}^{n} w_j I[T_j > T_i] \}^{1-\delta_i} \ ,$$

and the log likelihood is

$$\log EL = \sum_{i=1}^{n} \left\{ \delta_i \log w_i + (1-\delta_i) \log \sum_{j=1}^{n} w_j I[T_j > T_i] \right\} \ . \tag{2}$$

If we maximize the log EL above without constraint (I mean no extra constraints, the probability constraint $w_i \geq 0, \sum w_i = 1$ is always imposed), it is well known [Kaplan and Meier(1958)] that the Kaplan-Meier estimator $w_i = \Delta \hat{F}_{KM}(T_i)$ will achieve the maximum value of the log EL.

Empirical likelihood ratio method is first proposed by [Thomas and Grunkemeier(1975)] in the context of a Kaplan-Meier estimator. This method has been studied by [Owen(1988),Owen(2010),Li(1995), Murphy and van der Vaart(1997)] and [Pan and Zhou(1999)] among many others. When using the empirical likelihood with right censored data in testing a general hypothesis, [Zhou(2005)] gave an EM algorithm to compute the likelihood ratio. He compared the EM algorithm to the sequential quadratic programming method [Chen and Zhou(2007)], and concludes that the EM is better. Though quite stable, the EM can be slow in certain data settings. See more example in section 3. We shall give a new computation procedure for the constrained maximum of the log empirical likelihood above, and this lead to the computation of the empirical likelihood ratio for testing. This algorithm is recursive and much faster.

2 The computation of mean constrained maximum

In order to compute the empirical likelihood ratio, we need two empirical likelihoods: one with constraints, one without. The maximum of the empirical likelihood without constraint is achieved by F equals to the Kaplan-Meier estimator, as is well known. It remains to find the maximum of $\log EL$ under constraints.

A similar argument as in [Owen(1988)] will show that we may restrict our attention in the EL analysis, i.e. search max under constrains, to those discrete CDF F that are dominated by the Kaplan-Meier: $F(t) \ll \hat{F}_{KM}(t)$. [Owen 1988 restricted his attention to those distribution functions that $F(t) \ll$ the empirical distribution.]

The first step in our analysis is to find a discrete CDF that maximizes the log $EL(F)$ under the constraints (3), which are specified as follows:

$$\begin{aligned} \int_0^\infty g_1(t) dF(t) &= \mu_1 \\ \int_0^\infty g_2(t) dF(t) &= \mu_2 \\ \cdots \quad \cdots \quad & \cdots \\ \int_0^\infty g_p(t) dF(t) &= \mu_p \end{aligned} \tag{3}$$

where $g_i(t) (i = 1, 2, \ldots, p)$ are given functions satisfy some moment conditions (specified later), and μ_i $(i = 1, 2, \ldots, p)$ are given constants. Without loss of generality, we shall assume all $\mu_i = 0$. The constraints (3) can be written as (for discrete CDFs, and in terms of $w_i = \varDelta F(T_i)$)

$$\begin{aligned} \sum_{i=1}^n g_1(T_i) w_i &= 0 \\ \sum_{i=1}^n g_2(T_i) w_i &= 0 \\ \cdots \quad \cdots \quad & \cdots \\ \sum_{i=1}^n g_p(T_i) w_i &= 0 \ . \end{aligned} \tag{4}$$

We must find the maximum of the log$EL(F)$ under these constraints. We shall use the Lagrange multiplier to find this constrained maximum.

Without loss of generality assume w_i is only positive at the uncensored observations, except may be the last observation. Without loss of generality assume the observations are already ordered according to T and the smallest observation is an uncensored one ($\delta_1 = 1$). To see this, suppose the first observation is right censored and second one is uncensored. In this case, $\delta_1 = 0$, and $w_1 = 0$. Hence

$$w_1 = 0 \quad , \quad \sum_{i=2}^n w_i = 1. \tag{5}$$

The i-th term in log empirical likelihood is

$$\delta_i \log w_i + (1-\delta_i) \log \sum_{j=i+1}^n w_j .$$

This is true as observations are sorted according to T. Assuming $0 \times \log(0) = 0$, as shown in (5), the log empirical likelihood only depends on $w_2, \ldots, w_n$. Additionally, the first observation with $\delta = 0$ has no contribution to the constraints.

Assume there are n distinct values of T, which are already ordered. Let k be the number of censored observations among the n's. Thus we have only $n-k$ positive probability w_i's.

Introduce k new variables, one for each censored T observation, i.e. assume $\delta_j = 0$ and let:

$$S_j = \sum_{T_i > T_j} w_i = 1 - \sum_{T_i \le T_j} w_i \ . \tag{6}$$

This adds k new constraints to the optimization problem. We write the vector of those k constraints as $S_j - \sum_{T_i > T_j} w_i = 0$.

With these k new variables S_i, the log empirical likelihood in section 1 can be written simply as

$$\log EL = \sum_{\delta_i = 1} \log w_i + \sum_{\delta_i = 0} \log S_i \ . \tag{7}$$

The Lagrangian function for constrained maximum is

$$G = \log EL(w_i, S_j) + \lambda^\top \left(\sum_{\delta_i = 1} \delta_i w_i g(T_i) \right) - \eta \left(\sum_{i=1}^n w_i - 1 \right) - \gamma^\top \left(S - \sum_{T_i > \cdot} w_i \right) \ .$$

There are p constraints on the means, so length of λ is p; one constraints on the summation of w_i being one, so η is a scalar; and k constraints on the S_j, so the length of γ is k.

Next we shall take the partial derivatives and set them to zero. We shall show that $\eta = n$.

First we compute

$$\frac{\partial G}{\partial S_j} = \frac{1-\delta_j}{S_j} - \gamma_j$$

Setting the derivative to zero, we have

$$\gamma_j = (1 - \delta_j)/S_j \ . \tag{8}$$

Furthermore,

$$\frac{\partial G}{\partial w_k} = \frac{\delta_k}{w_k} + \lambda \delta_k g(T_k) - \eta + \gamma^\top (I[T_j < T_k, \delta_j = 0])$$

set the derivative to zero, (write k as i)

$$\eta = \frac{\delta_i}{w_i} + \lambda \delta_i g(T_i) + \gamma^\top (I[T_j < T_i, \delta_j = 0])$$

Multiply w_i on both sides and sum,

$$\sum_i w_i \eta = \sum_i \delta_i + \lambda \sum_i \delta_i w_i g(T_i) + (\sum_i w_i \gamma^\top I[T_j < T_i, \delta_j = 0]) \ .$$

Make use the other constraints, this simplifies to

$$\eta = (n - k) + 0 + \sum_i w_i \gamma^\top I[T_j < T_i, \delta_j = 0] \ . \tag{9}$$

We now focus on the last term above. Plug in the γ_j expression we obtained in (8) above, switch order of summation and it is not hard to see that

$$\sum_i w_i \gamma^\top I[T_j < T_i, \delta_j = 0] = \sum_i w_i \left(\sum_j \gamma_j I[T_j < T_i, \delta_j = 0] \right) \tag{10}$$

$$= \sum_j \sum_i \frac{w_i I[T_j < T_i, \delta_j = 0]}{S_j} = \sum_j 1 I[\delta_j = 0] = k \ . \tag{11}$$

Therefore equation (9) becomes $\eta = (n - k) + 0 + k$. Therefore $\eta = n$.

Thus, we have

$$w_i = \frac{\delta_i}{n - \lambda^\top \delta_i g(T_i) - \gamma^\top I[T_j < T_i, \delta_j = 0]} \tag{12}$$

where we further note (plug in the γ)

$$\gamma^\top I[T_j < T_i, \delta_j = 0] = \sum_j \frac{(1 - \delta_j)}{S_j} I[T_j < T_i, \delta_j = 0] = \sum_{j:\ \delta_j = 0} \frac{I[T_j < T_i]}{S_j} \ .$$

This finally gives rise to

$$w_i = w_i(\lambda) = \frac{\delta_i}{n - \lambda^\top \delta_i g(T_i) - \sum_{j:\delta_j = 0} \frac{I[T_j < T_i]}{S_j}} \tag{13}$$

which, together with (5), provides a recursive computation method for the probabilities w_i, provided λ is given:

1. Starting from the left most observation, and without loss of generality (as noted above) we can assume it is an uncensored data point: $\delta_1 = 1$. Thus

$$w_1 = \frac{1}{n - \lambda g(T_1)} \ .$$

2. Once we have w_i for all $i \le k$, we also have all S_j where $T_j < T_{k+1}$ and $\delta_j = 0$, by using ($S_j = 1 - \sum_i I[T_i \le T_j]w_i$), then we can compute

$$w_{k+1} = \frac{\delta_{k+1}}{n - \lambda^\top g(T_{k+1}) - \sum_{j:\ \delta_j=0} \frac{I[T_j < T_{k+1}]}{S_j}} \ .$$

So, this recursive calculation will give us w_i and S_j as a function of λ.

Remark: In the special case of no constraint of mean, then there is no λ (or $\lambda = 0$) and we have

$$w_{k+1} = \frac{\delta_{k+1}}{n - \sum_{j:\ \delta_j=0} \frac{I[T_j < T_{k+1}]}{S_j}}$$

we can show that this is actually the jump of the Kaplan-Meier estimator, as it should be. (Proof: use the identity $(1-\hat{F})(1-\hat{G}) = 1 - \hat{H}$, we first get a formula for the jump of the Kaplan-Meier estimator: $w_i = 1/n \times 1/(1-\hat{G})$. This works for $\hat{F}$ as well as for $\hat{G}$. we next show that

$$1 - 1/n \sum_{j:\ \delta_j=0} \frac{I[T_j < T_{k+1}]}{S_j} = (1 - \hat{G}(T_{k+1}))$$

since the left hand side is just equal to the summation of jumps of $\hat{G}$ before T_{k+1}.)

Finally where is the λ going to come from? We may get it from the constraint equation

$$0 = \sum_i \delta_i w_i(\lambda) g(T_i) = \sum_i \frac{\delta_i g(T_i)}{n - \lambda^\top \delta_i g(T_i) - \sum_{j:\ \delta_j=0} \frac{I[T_j < T_i]}{S_j}} . \tag{14}$$

So, the iteration goes like this:

(1) pick a λ value, near zero. (remember zero λ gives the Kaplan-Meier)

(2) with this λ find all the w_i's and S_j's by the recursive formula.

(3) plug those w_i into the equation (14) above and you got a mean value, call it θ. [The w_i's you obtained in step (2) are actually the constrained Kaplan-Meier with the constraint being these θ instead of zero]. Check to see if θ is zero, if not; change the λ value and repeat, until you find a λ which gives rise to w_i and S_j that satisfy the mean zero requirement (14).

For one dimensional λ this is going to be easily handled by the `uniroot( )` function in R. For multi dimensional λ this calls for a Newton type iteration.

The empirical likelihood ratio is then obtained as

$$-2 \log ELR = -2\{\log EL(w_i, S_j) - \log EL(w_i = \Delta \hat{F}_{KM}(T_i))\} \ ;$$

for the first log EL inside the curly bracket above we use the expression (6) with the w_i, S_j computed from the recursive method in this section, and the second term is obtained by (2) with $w_i = \Delta \hat{F}_{KM}(T_i)$, the regular Kaplan-Meier estimator.

Tied observations poses no problem for the algorithms discussed, as we can incorporate a weight for each observation.

Under the assumption that the variance of $\int g(t) d\hat{F}_{KM}(t)$ are finite, we have a chi square limiting distribution for the above -2 log empirical likelihood ratio, under null hypothesis (The Wilks' Theorem). Therefore we reject the null hypothesis if the computed -2 empirical likelihood ratio exceeds the chi square 95% percentile with p degrees of freedom. See [Zhou(2010)] for a proof of this theorem.

3 Simulation

If the Jacobian matrix of mean zero requirement (14) is not singular, Newton method could be used to find the root of (14). As mentioned previously, the null parameter space which has no constraint corresponds to $\lambda = 0$.

We shall call the recursive computation for $w_i's$ plus the Newton iteration for λ as KMC method, which will also be the name of the R package.

In each Newton iteration when we try to find the root, it is helpful to know the bound of solution, or so called, feasible region. In the KMC computation, it is obvious that those $\lambda's$ such that $n-\lambda^\top \delta_i g(T_i) - \sum_{j:\sim\delta_j=0} \frac{I[T_j<T_i]}{S_j} = 0$ will lead (14) to ∞. Denote those roots as $\lambda_i^\star \in \mathbb{R}^p, i = 1,\ldots,n$. Any i-th entry of the desired λ root for (14) and 0 must satisfy

$$\exists j \text{ such that } \lambda^\star_{i_j i} < 0, \lambda_i < \lambda^\star_{i_{j+1} i} \quad \forall i = 1,\ldots,p$$

where $\lambda^\star_{ij}$ is the j-th entry of vector $\lambda^\star_i$ such that $\lambda^\star_{i_1 i} < \ldots < \lambda^\star_{i_n i}$, and $\lambda^\star_{0,i} \triangleq -\infty$, $\lambda^\star_{n+1,i} \triangleq +\infty$. So, our strategy is to start at 0 and try to stay within the feasible region at all times when carry out the Newton iterations.

We could also calculate the analytical derivatives used in the Newton iteration. Define

$$f(\lambda) = \sum_i \delta_i w_i(\lambda) g(T_i)$$

To compute $\frac{\partial}{\partial\lambda} f(\lambda)$, we only need to calculate $\frac{\partial}{\partial\lambda} w_i$ and $\frac{\partial}{\partial\lambda} S_j = \frac{\partial}{\partial\lambda}\left(1 - \sum_{k=1}^{j} w_k\right) = -\frac{\partial}{\partial\lambda}\sum_{k=1}^{j} w_k$. There are no closed forms of such derivatives, but it could again be derived recursively.

(1) $w_1 = \frac{1}{n-\lambda g(T_1)}$, and $\frac{\partial}{\partial\lambda} w_1 = w_1^2 g(T_1)$
(2) $\frac{\partial}{\partial\lambda} w_{k+1}(\lambda) = \delta_{k+1}(w_{k+1})^2 \left(g(T_{k+1}) + \sum_{j:\delta_j=0}\left(I_{T_j<T_{k+1}}(S_j)^{-2}\frac{\partial}{\partial\lambda}\sum_{s=1}^{j}\frac{\partial}{\partial\lambda} w_s\right)\right)$

To evaluate the performance of this algorithm, a series of simulations had been done. We compared with standard EM algorithm (Zhou, 2005). Without further statement, all simulation works have been repeat 5,000 times and implemented in R language [Team et al(2005)]. R-2.15.3 is used on a Windows 7 (64-bits) computer with 2.4 GHz Intel(R) i7-3630QM CPU.

Experiment 1 : Consider a right censored data with only one constraint:

$$\begin{cases} X \sim \text{Exp}(1) \\ C \sim \text{Exp}(\beta) \end{cases} \tag{15}$$

Censoring percentage of the data are determined by different $\beta's$. Three models are included in the experiments

(1) $\beta = 1.5$, then 40% data are *uncensored*
(2) $\beta = 0.7$, then 58.9% data are *uncensored*
(3) $\beta = 0.2$, then 83.3% data are *uncensored*

The common hypothesis is (4), where $g(x) = (1-x)1_{(0\le x\le 1)} - e^{-1}$. We could verify that the true expectation is zero: $\int g(x)dF(x) = \int (1-x)1_{(0\le x\le 1)} e^{-x} dx - e^{-1}$=0.

To compare the performances of KMC and EM algorithm, we use four different sample sizes, i.e. 200, 1,000, 2,000, and 5,000 in the experiments. To make fair comparisons, $\|f^{(t)} - f^{(t+1)}\|_{\ell_2} \le 10^{-9}$ is used as the convergence criterion. Average spending time is reported to compare the computation efficiency in Table [1]. The no censored case is included for reference, this is equivalent to Newton solving λ without recursion. In all cases in our study, EM and KMC reported almost the same χ^2 test statistics and quantile to quantile plot is shown in Fig[1].

As shown in Table [1], we observed the following phenomenons:

Table 1 Average running time of EM/KMC (in second). "No Censor" column refers to time spend on solving empirical likelihood without censoring in R *el.test*(**emplik**). We use this as a comparison reference.

Censoring Rate	N	EM	KMC(nuDev)	KMC(AnalyDev)	No Censor
60% $\beta = 1.5$	200	0.175	0.011	0.028	0.005
	1000	3.503	0.106	0.211	0.007
	2000	13.935	0.349	0.692	0.033
	5000	73.562	1.801	3.663	0.036
41% $\beta = 0.7$	200	0.064	0.010	0.029	0.000
	1000	1.058	0.115	0.268	0.010
	2000	4.104	0.385	0.836	0.020
	5000	22.878	2.367	4.693	0.037
17% $\beta = 0.2$	200	0.014	0.008	0.029	0.002
	1000	0.117	0.071	0.240	0.009
	2000	0.425	0.240	0.694	0.018
	5000	2.702	1.220	3.282	0.026

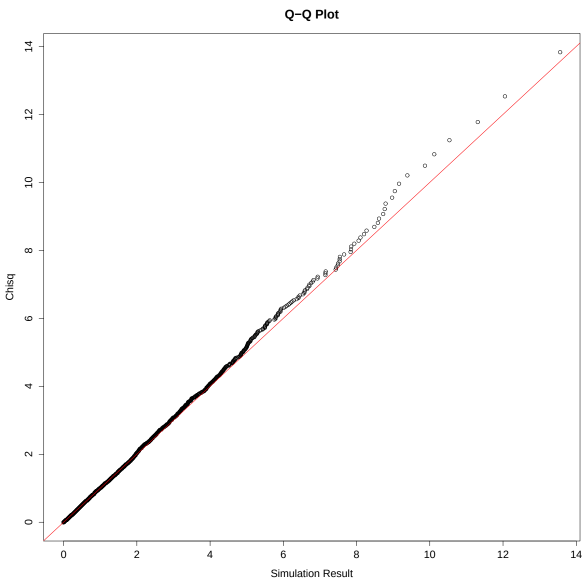

Fig. 1 Q-Q Plot for KMC with $\beta = .2$ N=1,000

(1) KMC always outperformed EM algorithm in speed at different simulation settings.
(2) Computation complexity of EM increased sharply with the percentage of censored data increasing. This is reasonable, since more censored data needs more E-step computation. But censored rate did not affect KMC much.
(3) Sample size is related to the computation complexity. We could see the running time of both EM and KMC increased along with the sample size.
(4) Another phenomenon is that, the computation of numeric derivative and analytic derivative of KMC is similar. This fact is straightforward, as KMC depends on solving (14) which is defined iteratively.

To summarize, when sample size is small and censored rate is low, the performance of EM and KMC is similar. But either in the large sample case or heavily censored case, KMC far outperformed EM algorithm with the same stopping criterion.

Experiment 2 : Consider a right censored data setting with two constraints. The i.i.d. right censored data are generated from (1) using

$$\begin{cases} X \sim \text{Exp}(1) \\ C \sim \text{Exp}(.7) \end{cases} \tag{16}$$

with the following hypothesis:

$$H_0 : \sum_i g_j(T_i)w_i = 0; \quad j = 1,2 \quad \text{where} \begin{cases} g_1(x) = (1-x)1_{(0\le x\le 1)} - e^{-1} \\ g_2(x) = 1_{(0\le x\le 1)} - 1 + e^{-1} \end{cases} \tag{17}$$

It is straightforward to verify that both g functions have expectation zero. In this simulation study, we

Table 2 Average running time of EM/KMC (in second)

Censoring Rate	N	EM	KMC(nuDev)
41%	200	3.055	0.033
41%	500	55.601	0.083

observed that EM spent great amount of time (3s$\sim$55s per case) to meet the converge criterion, while the average running time of KMC was considerable shorter (0.03s$\sim$0.08s). This dramatic result shows in multi-dimensional case, KMC runs much faster than EM algorithm. Only numerical derivatives were used in our simulations, one could implement the analytic ones using iteration shown previously. But in multi-dimensional case, the iterative type of derivatives do not have advantage over numeric ones. We recommend to use KMC with numeric derivative if one has more than one hypothesis even the sample size is small.
Experiment 3 : Other than exponential setting, considering a right censored data with one constraints:

$$\begin{cases} X \sim \text{Exp}(1) \\ C \sim U(0,\eta) \end{cases} \tag{18}$$

with hypothesis

$$H_0 : \sum_i g(T_i)w_i = 0; \quad \text{with} \quad g(x) = (1-x)1_{(0\le x\le 1)} - e^{-1}$$

we carried out some experiments on censoring time C_i from uniform distribution. There were two models:

(1) $\eta = 2$, then 56.77% data are uncensored
(2) $\eta = 4$, then 75.46% data are uncensored

Table 3 Average running time of EM/KMC (in second) of one constraint and Uniform distributed censored time

Censoring Rate	N	EM	KMC(nuDev)	KMC(AnalyDev)	No Censor
43%	200	0.104	0.014	0.038	0.004
$\eta = 2$	2000	9.700	0.654	1.288	0.107
25%	200	0.029	0.008	0.031	0.003
$\eta = 4$	2000	3.040	0.309	0.858	0.136

We found that the result shown in Tab[3] is very similar to Tab[1], which infers that different distribution of censored time will not affect the simulation result too much.

4 Discussion and Conclusion

In this manuscript, we proposed a new recursive algorithm, KMC, to calculate log empirical likelihood ratio statistics of right censored data with linear types of hypothesis. Our method used Lagrange multiplier method directly, and recursively computes the test statistics.

Numerical simulations show this new method has an advantage against traditional EM algorithm in the sense of computational complexity. Our simulation work also shows that the performance of KMC does not depend on the censoring rate, and outperformed EM algorithm at every simulation setting. We recommend to use KMC in all cases but particular large gain are expected in the following cases:

(1) Sample size is large (e.g. > 1000 observations);
(2) Data are heavily censored (e.g. censored rate $> 40\%$);
(3) There are more than one constraint.

On the other hand, and somewhat surprisingly, the analytic derivative did not help speed up computation in our simulation study. Besides, since KMC with numeric method could be extended to more than one constraints case, we highly recommend using numeric derivative in KMC rather than analytical one.

One of the issues of KMC is the initial value choosing, as is the case for most Newton algorithms. The performance of root solving relies on the precision of numerical derivative and Newton method. Our current strategy uses the M-step output of EM algorithm with only two iterations. Other better initial values are certainly possible. In addition, current KMC only works on right censored data, while EM algorithm works for right-, left- or doubly censored data; or even interval censored data. We were unable to find a proper way to derive such recursive computation algorithm in other censoring cases. (Software is available to download at http://github.com/yfyang86/kmc)

Acknowledgements Mai Zhou's research was supported in part by NSF grant DMS 1007666.

References

[Chen and Zhou(2007)] Chen K, Zhou M (2007) Computation of the empirical likelihood ratio from censored data. Journal of Statistical Computation and Simulation 77(12):1033–1042

[Kaplan and Meier(1958)] Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. Journal of the American statistical association 53(282):457–481

[Li(1995)] Li G (1995) On nonparametric likelihood ratio estimation of survival probabilities for censored data. Statistics & Probability Letters 25(2):95–104

[Murphy and van der Vaart(1997)] Murphy SA, van der Vaart AW (1997) Semiparametric likelihood ratio inference. The Annals of Statistics pp 1471–1509

[Owen(1988)] Owen AB (1988) Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75(2):237–249

[Owen(2010)] Owen AB (2010) Empirical likelihood. CRC press

[Pan and Zhou(1999)] Pan XR, Zhou M (1999) Using one-parameter sub-family of distributions in empirical likelihood ratio with censored data. Journal of statistical planning and inference 75(2):379–392

[Team et al(2005)] Team RDC, et al (2005) R: A language and environment for statistical computing

[Thomas and Grunkemeier(1975)] Thomas DR, Grunkemeier GL (1975) Confidence interval estimation of survival probabilities for censored data. Journal of the American Statistical Association 70(352):865–871

[Zhou(2005)] Zhou M (2005) Empirical likelihood ratio with arbitrarily censored/truncated data by em algorithm. Journal of Computational and Graphical Statistics 14(3):643–656

[Zhou(2010)] Zhou M (2010) A wilks theorem for the censored empirical likelihood of means. URL www.ms.uky.edu/∼mai/research/Note3.pdf