A Wilks Theorem for the Censored Empirical Likelihood of Several Means

Mai Zhou
Department of Statistics
University of Kentucky

Abstract

In this paper we give a proof of the Wilks theorem for the empirical likelihood ratio test for the right censored data, when the hypothesis are formulated in terms of p estimating equations or mean functions. In particular we show that the empirical likelihood ratio test statistic is equal to a quadratic form similar to a Hotelling's T^2 statistic, plus a small error. The least favorable distribution is given explicitly in terms of advanced time transformation.

Keywords Likelihood ratio test; Chi square distribution; Right censored data; Multiple constraints.

1 Introduction

Empirical likelihood (EL) is a recently developed nonparametric statistical inference method similar to the parametric likelihood ratio test. Owen's 2001 book contains many important results. Specifically, Owen (1988) was the first paper to proof rigorously the asymptotic chi square distribution for the empirical likelihood ratio when there is a single parameter of mean type. Owen (1990) dealt with multiple parameters setting.

However, for right censored data, less results are available for empirical likelihoods. Thomas and Grunkemier (1975) proposed to use empirical likelihood for a single parameter of surviving probability (Kaplan-Meier estimator), and this is the first time where empirical likelihood was proposed as a better nonparametric inference method for construct confidence intervals. Li (1995), Murphy (1995) made the arguments in Thomas and Grunkemier rigorous.

When dealing with a single parameter of mean, and doubly censored data, Murphy and Van der Vaart (1997) contains a proof of asymptotic chi square distribution for the (censored) empirical likelihood ratio. But the conditions imposed, like boundedness, was often too restrictive in practice. Furthermore, we are not aware of a Wilks theorem for empirical likelihood dealing with multiple constraints of mean type for right censored data. These two points are precisely what the current paper try to provide.

In the analysis of (multiple) linear models, least squares type methods leads to the so called normal equations. If the linear model have p covariates, then we end up with p simultaneous estimating equations. To apply the EL techniques there, we need a Wilks theorem with multiple (p) parameters. Thus our result is needed in the EL analysis of the censored data regression models (AFT models), where typically multiple estimating equations of mean type, with terms that are not bounded, are concerned. (See for example Zhou and Li 2008, Zhou, Kim and Bathke 2012) Another example where Wilks theorem for EL is useful is in the testing concern the mean residual lifetimes, (Zhou and Jeong 2011), where the terms are again not bounded.

Some other existing work on censored data EL include: EL for the equality of k medians based on k-samples, see Naiknimbalkar and Rajarshi (1997); EL for the weighted hazards, see Pan and Zhou (2002). Li, Li and Zhou (2005) provides a review of EL results in survival analysis.

We end this section by introducing notation and the basic setup of this paper. The main theorem of this paper is in section 2. Some tedious calculations are put in the appendix.

Suppose that $X_1, X_2, \ldots, X_n$ 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 $C_1, C_2, \ldots, C_n$ which are i.i.d. with a distribution G_0. Only the censored observations, (T_i, δ_i), are available to us:

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

The empirical likelihood of the censored data in terms of distribution (see for example

Owen 2001 (6.9)) 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. The second line above assumes a discrete $F(\cdot)$. Let $w_i = \Delta F(T_i)$ for $i = 1, 2, \ldots, n$ then 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

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

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

2 A Wilks Theorem for Several Means

To form the ratio of two empirical likelihoods, we not only need to find the maximum of the log EL among all F, but we also need to find the maximum of log EL under a null hypothesis. We shall next specify the null hypothesis. A natural hypothesis to consider is the hypothesis involving linear or mean functionals. As Owen (1988) have shown, this includes the M- and Z- estimates.

Please note a similar argument as in Owen (1988) will show that we may restrict our attention in the EL analysis, i.e. search maximum under null hypothesis, to those discrete CDF F that are dominated by the Kaplan-Meier estimator: $F(t) \ll \hat{F}_{KM}(t)$. [that is, discrete distributions with jump location same as the Kaplan-Meier] Owen 1988 restricted his attention to those distribution functions F that are dominated by the empirical distribution.

The first step in our analysis is to find a discrete CDF that maximizes the log $EL(F)$ under the (null) hypothesis (1), which are specified as follows:

$$\begin{aligned} \int_0^\infty g_1(t) dF(t) &= \mu_1 \\ \int_0^\infty g_2(t) dF(t) &= \mu_2 \\ \cdots \quad \cdots &\quad \cdots \\ \int_0^\infty g_p(t) dF(t) &= \mu_p \end{aligned} \tag{1}$$

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 (1) can be written as (for discrete CDF, and in terms of $w_i = \Delta F(T_i)$)

$$\begin{aligned} \sum_{i=1}^n g_1(T_i) w_i &= 0 \\ \sum_{i=1}^n g_2(T_i) w_i &= 0 \\ \cdots \quad \cdots &\quad \cdots \\ \sum_{i=1}^n g_p(T_i) w_i &= 0 \ . \end{aligned} \tag{2}$$

We must find the maximum of the log $EL(F)$ under these constraints. We shall accomplish that in two steps. First we construct a p parameter family of CDF that pass through (and dominated by) the Kaplan-Meier estimator, in the direction h. Then we find the CDF in this family that satisfy the constraints. Secondly, we will maximize the log EL over all possible h.

Notice that *any* discrete CDF dominated by the Kaplan-Meier estimator can *always* be written as (in terms of its jump)

$$\Delta F(T_i) = \Delta \hat{F}_{KM}(T_i)[1 - h(T_i)] \ , \quad i = 1, 2, \ldots, n;$$

for *some* h function. Of course this distribution usually do not satisfy the constraints above. This motivates us to define the following:

For any p given functions of t, $h = (h_1(t), \ldots, h_p(t))$, we define a family of distributions (indexed by $\lambda \in \mathbb{R}^p$ and dominated by the Kaplan-Meier estimator) by its jumps

$$\Delta F(T_i) = \Delta F_\lambda(T_i) = \Delta \hat{F}_{KM}(T_i)[1 - \lambda^\top h(T_i)] \tag{3}$$

where we require the total jumps sum to one (as any discrete CDF must) which lead to

$$\sum_i h(T_i)\Delta \hat{F}_{KM}(T_i) = 0. \tag{4}$$

The notation above $\lambda^\top h(T_i)$ is the inner product $\lambda_1 h_1(T_i) + \ldots + \lambda_p h_p(T_i)$.

Since the jumps of Kaplan-Meier estimator is between zero and one, at least for small values of λ (in a neighborhood of zero) the jumps of so defined F above are going to be between zero and one and thus, is a legitimate distribution for those small λ's. Obviously, when $\lambda = 0$, we have $F_{\lambda=0} = \hat{F}_{KM}$.

We shall also, WOLG, require that $\|h\|_2 = K$, for some fixed constant $K > 0$ (since $\lambda^\top h = (\lambda/a)^\top ah$). We shall take $\lambda = (\lambda_1, \ldots, \lambda_p)$ as the parameter who's value will be selected to make this distribution satisfy the hypothesis/constraints (1) above. The requirement that the distribution satisfy the constraints/hypothesis (1) will force the λ to take certain value, as in the following equations:

$$0 = \sum_{i=1}^n g_j(T_i)\Delta F_\lambda(T_i) \qquad j = 1, 2, \cdots, p. \tag{5}$$

Denote the solution of the above equation as λ^*. The fact that it has a unique solution can be guaranteed by the assumption that matrix A (defined in the Lemma 2 below) is invertible. (We, in fact, will assume a slightly stronger condition: that the condition number of A be *bounded* away from zero).

The next lemma will be useful later.

Lemma 1 *Define a vector gg of length p with elements $gg_j = \sum_i g_j(T_i)\Delta \hat{F}_{KM}(T_i)$. Under null hypothesis (1), we have, (since the true mean of g are assumed to be zero under null, no centering constants are needed)*

$$\sqrt{n}\; gg \xrightarrow{\mathcal{D}} N(0, \Sigma) \qquad \text{as } \; n \to \infty \; .$$

The asymptotic ($p \times p$) variance-covariance matrix $\Sigma = [\sigma_{jk}]$ (assumed to be non-singular) is given by

$$\sigma_{jk} = \int [g_j(x) - \bar{g}_j(x)][g_k(x) - \bar{g}_k(x)] \frac{dF_0(x)}{1 - G_0(x-)}$$

and it can be consistently estimated by $\hat{\Sigma} = [\hat{\sigma}_{jk}]$

$$\hat{\sigma}_{jk} = \sum_{i=1}^n [g_j(T_i) - \bar{g}_j(T_i)][g_k(T_i) - \bar{g}_k(T_i)] \frac{\Delta \hat{F}_{KM}(T_i)}{1 - \hat{G}_{KM}(T_i)} \;,$$

where $\bar{g}_j$ is the 'advanced-time transformation' of g_j defined by Efron and Johnstone (1991), either with respect to the F_0 (in σ_{jk} above) or with respect to the Kaplan-Meier estimator (in $\hat{\sigma}_{jk}$ above). See also Lemma A.1 in the appendix for advanced-time. Finally $\hat{G}_{KM}$ is the Kaplan-Meier estimator of the censoring distribution.

Proof: This theorem is just the asymptotic normality of the Kaplan-Meier integral, which has been treated by many others before. What is new is perhaps the variance-covariance formula. It can easily be proved by using the representation of Akritas (2000) which contains a univariate version of this lemma. Using the representation of Akritas, one then invoke the multivariate central limit theorem for the counting process martingales to finish proof. One such Central Limit Theorem for counting process martingales can be found in Kalbfleish and Prentice (2002) Chapter 5.

The assertion that the variance-covariance can be consistently estimated can be seen from the fact that it is an plug-in estimator, and the Kaplan-Meier estimator is uniformly strongly consistent. □

Solving equation (5) gives the following solution of λ for the constraint equations, which we call λ^*.

Lemma 2 *Assume the conditions of Lemma 1. Denote the solution of (5) as λ^*. Assume the condition number of the matrix A is bounded away from zero. Then the distribution F_λ that satisfy the constrain (5) must have*

$$\lambda^* A = gg$$

or

$$\lambda^* = A^{-1} gg$$

where the $p \times p$ matrix $A = [a_{jk}]$ is defined by the elements

$$a_{jk} = \sum_{i=1}^{n} g_j(T_i) h_k(T_i) \Delta \hat{F}_{KM}(T_i) \ . \tag{6}$$

Also, the elements may be written as

$$a_{jk} = \sum_i [g_j(T_i) - Eg_j][h_k(T_i) - Eh_k] \Delta \hat{F}_{KM}(T_i) \ .$$

Proof: Plug (3) in to (5) and solving equation (5) to arrive the stated result. The last expression of a_{jk} is valid since $\sum_i h_k(T_i)\Delta\hat{F}_{KM}(T_i) = 0$, so both of the a_{jk} expressions are valid expressions of the covariance.

We see that $\mathbb{E}h^2 = K^2 < \infty$ for all h. Also, we have that $\mathbb{E}g^2 < \infty$ (assumption of Lemma 1). To ensure the inverse matrix exist, we need to impose one more condition. Please see remark 2 below. □

Remark 1: We are to find the maximum of the $\log EL$ among all h. Yet we placed some restrictions on h. The assumption we need to place on h, for given g is that this matrix A should be invertible. For those h that A is not invertible, we may argue that this sub-family of distributions do not have a solution that satisfy the constraints (1), and thus can be ignored in the search of maximizing $\log EL$ *under constraints*. On the other hand, the least favorable h^* we calculated in (11) will lead to an A matrix that is identical to $\hat{\Sigma}$ defined in Lemma 1, which by assumption is a non singular variance-covariance matrix for sufficiently large n, and thus invertible.

Remark 2: Under the null hypothesis, gg is of order $O_p(1/\sqrt{n})$ (Lemma1). If the h function is such that the inverse matrix of A has a bounded condition number, then λ^* is also of order $O_p(1/\sqrt{n})$, uniformly for those h.

Remark 3: In order to work with the inverse matrix, we end up with a condition on h in terms of a matrix as below. Assume the following matrix is positive definite:

$$A^\infty = (a_{ij}^\infty), \quad \text{where} \quad a_{ij}^\infty = \int_0^\infty g_i(t)h_j(t)dF_0(t) \tag{7}$$

Notice once we assumed A^∞ is invertible, a finite sample version of this should also hold for large enough n, as needed in the Lemma 2. The meaning of this is that we need to avoid those h that are (almost) perpendicular to g. In (11) we see that the particular h we are interested is not perpendicular to g, in fact, far from it.

Define $f(\lambda) = \log EL(F_\lambda)$. It is easy to see that when $\lambda = 0$, F_λ becomes the Kaplan-Meier estimator. So the Wilks statistic is just $2[f(0) - \sup_h f(\lambda^*)] = -2\log ELR$.

Taking a Taylor expansion with $f(\lambda^*)$, we have

$$\text{Wilks statistics} = \inf_h \; 2\{f(0) - f(0) - \lambda^* f'(0) - 1/2[\lambda^*]^\top f''(0)\lambda^* + O_p(\lambda^*)\}$$

Notice that obviously $f'(0) = 0$ since the derivative of log likelihood at the maximum (i.e. the Kaplan-Meier) must be zero no matter what h. Recall the Kaplan-Meier estimator is the MLE that maximizes $\log EL$. This can also be readily checked, using the self-consistency identity (appendix) and the fact that $\sum_i h(T_i)\Delta\hat{F}_{KM}(T_i) = 0$.

We finally have

$$\text{Wilks statistics} = \inf_h \; [\sqrt{n}\lambda^*]^\top \left(-f''(0)/n\right) [\sqrt{n}\lambda^*] + o_p(1); \tag{8}$$

provided the $o_p(1)$ is uniform over h, which we pointed out is true under the assumption of condition number for the matrix A^∞ (Remark 2 and 3).

The rest of the analysis will focus on the first term on the right hand side of (8) above. First we calculate the second derivative f'', simplify it. Then we show the infimum over h of the right hand (ignore the $o_p(1)$ part) is achieved at an h satisfy the equation (11) below. And finally, for this particular h the above Wilks statistics becomes a Hotelling's T^2 and thus having asymptotically a chi square distribution with df$= p$.

Theorem 1 *Let $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ be n pairs of i.i.d. censored random variables as defined above. Suppose $g_i \quad i = 1, \ldots, p$ are given functions such that the $p \times p$ asymptotic variance-covariance matrix of the Kaplan-Meier mean estimator, $\int g_i(t) d\hat{F}_{KM}(t)$, is well defined and positive definite. That is, the $p \times p$ matrix Σ below, is well defined and positive definite:*

$$\Sigma = (\sigma_{jk}) = \int \frac{[g_j(x) - \bar{g}_j(x)][g_k(x) - \bar{g}_k(x)]}{(1 - G_0(x-))} dF_0(x) \; . \tag{9}$$

Assume also that the matrix A^∞ is invertible.

Then, under null hypothesis (1) as $n \to \infty$, we have

$$-2 \log ELR \xrightarrow{\mathcal{D}} \chi^2_{(p)} \quad as \quad n \to \infty$$

where $\log ELR = \sup_h \log EL(F_{\lambda^*}) - \log EL(\hat{F}_{KM})$.

In fact we have

$$-2 \log ELR = [\sqrt{n} gg]^\top \hat{\Sigma}^{-1} [\sqrt{n} gg] + o_p(1) \; ,$$

where gg and $\hat{\Sigma}$ are defined in Lemma 1.

A sufficient condition to guarantee the (single) integral in (9) to be well defined was pointed out by Akritas (2000)

$$0 < \int \frac{g_j^2(x)}{1 - G_0(x-)} dF_0(x) < \infty \ , \quad j = 1, \cdots, p.$$

To ensure the matrix Σ is nonsingular, we require that the g_i functions are so called linearly independent. In other words, the p constraints are all genuine, no redundancy.

Proof of Theorem 1. We proceed by proving two more lemmas.

Lemma 3 *The second derivative $f''(0)$ defined above is equal to $-nB$ with the elements of the matrix B defined below in (10).*

We now compute the second derivative $f''(0)$. Straight forward calculation show that this is a $p \times p$ matrix. The jk^{th} elements of the matrix $[-f''(0)/n] = B = [b_{jk}]$ is given by

$$\begin{aligned} b_{jk} &= \sum_{i=1}^{n} h_j(T_i) h_k(T_i) \frac{\delta_i}{n} \\ &+ \sum_{i=1}^{n} \frac{1-\delta_i}{n} \frac{[\sum_{m:T_m>T_i} h_j(T_m)\Delta\hat{F}_{KM}(T_m)][\sum_{m:T_m>T_i} h_k(T_m)\Delta\hat{F}_{KM}(T_m)]}{[1-\hat{F}_{KM}(T_i)]^2} . \end{aligned}$$

After several rounds of tedious simplifications (see appendix for details) we have

$$b_{jk} = \sum_{i=1}^{n} [h_j(T_i) - \bar{h}_j(T_i)][h_k(T_i) - \bar{h}_k(T_i)][1 - \hat{G}_{KM}(T_i)]\Delta\hat{F}_{KM}(T_i) \ . \tag{10}$$

Lemma 4 (Matrix Cauchy-Schwarz inequality) *For any h we have*

$$[A^{-1}]^\top B A^{-1} \geq \hat{\Sigma}^{-1}$$

where the $\geq$ means the matrix inequality for positive-definite matrices. And the equality is achieved by h^* that satisfy (11) below. For $j = 1, 2, \ldots, p$

$$[h_j^*(x) - \bar{h}_j^*(x)] = \frac{g_j(x) - \bar{g}_j(x)}{1 - \hat{G}_{KM}(x-)} \qquad a.s. \quad \hat{F}_{KM} \ . \tag{11}$$

Obviously any h that is a constant multiple of (11) also achieve the equality. (i.e. h has a constant free play).

Proof: First we rewrite the entries of matrix A as

$$a_{jk} = \sum_{i=1}^{n} \frac{g_j(T_i) - \bar{g}_j(T_i)}{\sqrt{1 - \hat{G}_{KM}(T_i)}} \ [h_k(T_i) - \bar{h}_k(T_i)]\sqrt{1 - \hat{G}_{KM}(T_i)} \ \Delta\hat{F}_{KM}(T_i) \ .$$

This is valid see the advanced transformation identity in appendix later. So, in terms of the expectation with respect to the Kaplan-Meier, we have

$$a_{jk} = E \ \frac{(g-\bar{g})}{\sqrt{1-G}} \ (h-\bar{h})\sqrt{1-G} = E(\alpha\beta), \qquad \text{say}$$

and furthermore,

$$b_{jk} = E \ (h-\bar{h})\sqrt{1-G} \ (h-\bar{h})\sqrt{1-G} = E(\beta^2) \ ,$$

$$\hat{\sigma}_{jk} = E \ \frac{g-\bar{g}}{\sqrt{1-G}} \ \frac{g-\bar{g}}{\sqrt{1-G}} = E(\alpha^2) \ .$$

The inequality can then follow easily from the well known matrix Cauchy-Schwarz inequality (see Tripathi (1999) and references therein). □

Using Lemma 4, we see that for any y, the quadratic form, $y^\top A^{-1\top} B A^{-1} y \geq y^\top \hat{\Sigma}^{-1} y$. And the equality is achieved for h satisfy (11).

Using Lemma 1, 2, 3 and this Cauchy-Schwarz inequality, we see that

$$\text{Wilks Statistics} = \inf_h [\sqrt{n} gg]^\top [A^{-1}]^\top B A^{-1} [\sqrt{n} gg] + o_p(1) = [\sqrt{n} gg]^\top \hat{\Sigma}^{-1} [\sqrt{n} gg] + o_p(1) \ . \quad (12)$$

As $n \to \infty$ the right hand side is clearly converging to a chi square distribution under null hypothesis with $df = p$. The proof of Theorem 1 is now complete. □

Remark 4: In other words, the least favorable direction is given by (11). Or, if we define a parametric model by the equation (3), this is what Stein (1956) called the most difficult parametric problem.

Remark 5: If we want to profile out part of the parameters in the above empirical likelihood ratio, we shall get a chi square distribution with reduced degrees of freedom. This can be easily seen from the asymptotic representation of the empirical likelihood ratio as a Hotelling's T^2, for which the similar profiling result is well known.

3 Discussion

We make some critical use of the "advanced time change" of Efron and Johnstone (1990). The least favorable distribution we found is also specified in terms of the advanced time change. See the equation (3) and (11). We also used a multivariate version of the Akritas' (2000) Central Limit Theorem for the Kaplan-Meier integral, in particular the variance/covariance expression.

The computation of the empirical likelihood ratio discussed in this paper has been available using the R package `emplik`. For example the empirical likelihood ratio test as we stated in Theorem 1 in section 3 can be computed by the function `el.cen.EM2` there.

The empirical likelihood for the integrated hazards was studied by Pan and Zhou (2002).

Appendix:

SIMPLIFICATION OF THE SECOND DERIVATIVE $-f''(0)/n$.

The first 'simplification' uses the self consistency identity (Lemma A.2 below, with $g = h_j h_k$) to the first term of $-f''(0)/n$.

$$
\begin{aligned}
b_{jk} &= \sum_{i=1}^{n} h_j(T_i)h_k(T_i)\Delta\hat{F}_{KM}(T_i) - \sum_{i=1}^{n} \frac{1-\delta_i}{n} \frac{\sum_{T_j > T_i} h_j(T_j)h_k(T_j)\Delta\hat{F}_{KM}(T_j)}{1-\hat{F}_{KM}(T_i)} \\
&+ \sum_{i=1}^{n} \frac{1-\delta_i}{n} \frac{[\sum_{m:T_m>T_i} h_j(T_m)\Delta\hat{F}_{KM}(T_m)][\sum_{m:T_m>T_i} h_k(T_m)\Delta\hat{F}_{KM}(T_m)]}{[1-\hat{F}_{KM}(T_i)]^2} .
\end{aligned}
$$

The last two summations above (those with $1-\delta_i$) can be combined by using the identity $\mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y) = \mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y)$. In fact we use it n times, each with a different conditional distribution, to get the following.

$$
\begin{aligned}
b_{jk} &= \sum_{i=1}^{n} h_j(T_i)h_k(T_i)\Delta\hat{F}_{KM}(T_i) \\
&- \sum_{i=1}^{n} \frac{1-\delta_i}{n} \frac{\sum_{m:T_m>T_i} [h_j(T_m) - E_{\hat{F}}(h_j|t>T_i)][h_k(T_m) - E_{\hat{F}}(h_k|t>T_i)]\Delta\hat{F}_{KM}(T_m)}{1-\hat{F}_{KM}(T_i)} .
\end{aligned}
$$

Notice, we have many expressions of covariances in the above. The first summation can also be written as $\sum [h_j - \mathbb{E}h_j][h_k - \mathbb{E}h_k]\Delta\hat{F}_{KM}$ since $\sum h_j \Delta\hat{F}_{KM} = 0$. Our second simplification uses the advanced time identity (see below) to re-write the covariance expressions. We use the identity $n+1$ times for either the Kaplan-Meier CDF or the conditional Kaplan-Meier CDF to get

$$
\begin{aligned}
b_{jk} &= \sum_{i=1}^{n} [h_j(T_i) - \bar{h}_j(T_i)][h_k(T_i) - \bar{h}_k(T_i)]\Delta\hat{F}_{KM}(T_i) \\
&- \sum_{i=1}^{n} \frac{1-\delta_i}{n} \frac{\sum_{m:T_m>T_i} [h_j(T_m) - \bar{h}_j(T_m)][h_k(T_m) - \bar{h}_k(T_m)]\Delta\hat{F}_{KM}(T_m)}{1-\hat{F}_{KM}(T_i)} .
\end{aligned}
$$

Third simplification uses the self-consistency identity again (with $g = [h_j - \bar{h}_j][h_k - \bar{h}_k]$). This allows us to combine the two terms. We get,

$$b_{jk} = \sum_{i=1}^{n} [h_j(T_i) - \bar{h}_j(T_i)][h_k(T_i) - \bar{h}_k(T_i)] \frac{\delta_i}{n} \ .$$

Fourth and final simplification uses the following identity to replace δ_i/n,

$$\frac{\delta_i}{n(1 - \hat{G}_{KM}(T_i))} = \Delta \hat{F}_{KM}(T_i) \ .$$

This identity can be proved from the well known fact for the Kaplan-Meier estimator $[1 - \hat{F}_{KM}(t)][1 - \hat{G}_{KM}(t)] = 1 - H(t) = 1/n \sum_{i=1}^{n} I[T_i > t]$.

We finally get what we want.

$$b_{jk} = \sum_{i=1}^{n} [h_j(T_i) - \bar{h}_j(T_i)][h_k(T_i) - \bar{h}_k(T_i)][1 - \hat{G}_{KM}(T_i)] \Delta \hat{F}_{KM}(T_i) \ .$$

Lemma A.1: (Advanced-time identity) [Efron and Johnstone 1990] Define the 'advanced time' transformation for a function $g(t)$ with respect to a CDF $F(\cdot)$ as

$$\bar{g}(s) = \bar{g}_F(s) = \frac{\int_{(s,\infty)} g(x) dF(x)}{1 - F(s)} = E_F[g(X) | X > s] \ .$$

Then we have

$$\mathrm{Var}_F(g) = \int [g(x) - E_F g]^2 dF(x) = \int [g(x) - \bar{g}(x)]^2 dF(x)$$

and

$$\mathrm{Cov}_F(g, h) = \int [g(t) - E_F g] h(t) dF(t) = \int [g(t) - \bar{g}(t)][h(t) - \bar{h}(t)] dF(t)$$

where $E_F g = \int g(x) dF(x)$.

Proof: The result for the variance is directly from Efron and Johnstone (1990). The result for the covariance can be proved similarly. □

Lemma A.2 (Self-consistency identity) For the Kaplan-Meier estimator, $\hat{F}_{KM}$, we have that for any function $g(\cdot)$

$$\sum_i g(T_i) \Delta \hat{F}_{KM}(T_i) \ = \ \sum_i \frac{\delta_i}{n} g(T_i) + \sum_i \frac{(1 - \delta_i)}{n} \frac{\sum_{T_j > T_i} g(T_j) \Delta \hat{F}_{KM}(T_j)}{1 - \hat{F}_{KM}(T_i)} \ .$$

Proof: The probability corresponding to $g(T_k)$ on the left hand side is $\Delta \hat{F}_{KM}(T_k)$. The probabilities associated with $g(T_k)$ on the right hand side is precisely those given by Turnbull (1976) self-consistent equation. It is well known that Kaplan-Meier estimator is self consistent. □

Acknowledgement

The research was supported in part by National Science Foundation grant DMS-0604920.

References

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

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

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data* 2nd Ed. Wiley, New York

Kaplan, E. and Meier, P. (1958). Non-parametric estimator from incomplete observations. *Journal of American Statistical Association*, **53**, 457-481.

Li, G. (1995). On nonparametric likelihood ratio estimation of survival probabilities for censored data. *Statist. & Prob. Letters*, **25** , 95-104.

Murphy, S.A. (1995). Likelihood ratio-based confidence intervals in survival analysis. *Journal of American Statistical Association*, **90** 1399-1405.

Murphy, S. and van der Vaart, A. (1997). Semiparametric likelihood ratio inference. *Annals of Statistics* **25**, 1471-509.

Li, G., Li, R. and Zhou, M. (2005). Empirical likelihood in survival analysis. In *Contemporary Multivariate Analysis and Design of Experiments* Edited by J. Fan and G. Li. pp. 337-350. The World Scientific Publisher.

Tripathi, G. (1999). A matrix extension of the Cauchy-Schwarz inequality. *Economics Letters* **63**, 1-3.

Naiknimbalkar, U.V. and Rajarshi, M.B. (1997). Empirical likelihood ratio test for equality of k medians in censored data. *Statistics & Probability Letters*, **34**, 267-273.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237-49.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, **18** 90-120.

Owen, A. (2001). *Empirical Likelihood.* Chapman & Hall/CRC, London.

Pan, X.R. and Zhou, M. (2002). Empirical likelihood in terms of cumulative hazard function for censored data. *Journal of Multivariate Analysis* **80**, 166-188.

Thomas, D.R. and Grunkemeier, G.L. (1975). Confidence Interval Estimation of Survival Probabilities for Censored Data. *Journal of the American Statistical Association*, **70**, 865-871.

Turnbull, B. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of Royal Statistical Society, Series, B* **38**, 290-295.

Zhou, M. and Li, G. (2008). Empirical likelihood analysis of the Buckley-James estimator. *Journal of Multivariate Analysis*, **99**, 649 - 664.

Zhou, M. and Jeong, J. (2011) Empirical Likelihood Test for Mean and Median Residual Life Times. *Statistics in Medicine* **30**, 152-159.

Zhou, M., Kim, M. and Bathke, A. (2012) Empirical Likelihood Analysis for the Heteroscedastic Accelerated Failure Time Model. *Statistica Sinica*, **22**, 295-316.