Likelihood Ratio Based Confidence Intervals in Survival Analysis

S.A. MURPHY

Abstract

Confidence intervals for the survival function and the cumulative hazard function are considered. These confidence intervals are based on an inversion of the likelihood ratio statistic. To do this two extensions of the likelihood, each of which yields meaningful likelihood ratio hypothesis tests and subsequent confidence intervals, are considered. The choice of the best extension is difficult. In the failure time setting, the binomial extension is best in constructing confidence intervals concerning the survival function and the Poisson extension is best in constructing confidence intervals concerning the cumulative hazard. Simulations indicate that these two methods perform as well as or better than competitors based on asymptotic normality of the estimator.

Key words: Non-parametric Model; Infinite Dimensional Nuisance Parameter

S. A. Murphy is Assistant Professor, Department of Statistics, Pennsylvania State University, 326 Classroom Building, University Park, PA 16802. This work was partially supported by NSF grant DMS-9307255

1. Introduction

In this paper we consider likelihood based confidence intervals for functions of the cumulative hazard, $A(\cdot)$, and the survival function, $S(\cdot)$ for failure time data. In order to construct a likelihood based confidence interval for a parameter, say θ, one inverts the likelihood ratio test of $H_0 : \theta = \theta_0$. The confidence interval for θ consists of those θ_0 for which we are unable to reject H_0. Thomas and Grunkemeier (1975) first considered a likelihood based confidence set for $S(t_0)$, the survival function evaluated at the point t_0. They prove that the confidence set for $S(t_0)$ is a closed interval. This paper is a generalization of their early work; as in their case, we do not make parametric assumptions on the form of the survival distribution.

The possibility of likelihood ratio based confidence intervals in a semi-parametric setting is very attractive. As Hall and La Scala (1990) summarize, confidence intervals based on the likelihood ratio test possess nice properties. For example, the confidence interval does not have a predetermined shape whereas confidence intervals based on the asymptotic normality of an estimator have a symmetry which is implied by the asymptotic normality. Also likelihood ratio based confidence intervals respect the range of the parameter; if the parameter is a survival probability then the confidence interval will be contained in the closed interval zero to one. Another positive characteristic is that the likelihood ratio based confidence interval is transformation respecting, that is a likelihood ratio based confidence interval for $g(\theta)$ is given by g applied to each value in the confidence interval for θ. As Thomas and Grunkemeier (1975) illustrate, a likelihood ratio based confidence interval can be constructed in cases in which the maximum likelihood estimator does not exist. For these reasons, we consider likelihood ratio based confidence intervals in this semi-parametric setting.

We first give a discussion of two possible likelihood versions in the failure time setting. Both versions yield valid likelihood ratio tests and therefore valid confidence intervals. The appendix contains a rigorous proof for the asymptotic distribution properties of the confidence intervals for both the survival distribution, it's inverse and the cumulative hazard. The verification that the resulting confidence set is an interval is valid for any continuous function of the cumulative hazard (including S).

2. THE LIKELIHOOD

A crucial issue in discussing a likelihood ratio test is the identification of the likelihood. In parametric problems the likelihood is usually a well defined quantity. However in semiparametric problems several problems arise, (1) often there is a dominating measure only under absolute continuity of the distribution, (2) and if the distribution is assumed absolutely continuous then there is no maximum likelihood estimator.

To make the issues clear suppose a random sample, $X_1, \ldots, X_n$, from an unknown distribution, F, is observed. Stating the likelihood as $\prod_{i=1}^n f(X_i)$ assumes absolute continuity of the parameter, F, with respect to lebesgue measure. There is no maximum likelihood estimator of f since the likelihood can be made arbitrarily large by setting $f = 0$ except close to the X_i's, where we make f arbitrarily large while adjusting the support of f so that f always integrates to one. An alternate approach is to realize that the knowledge that F is absolutely continuous is not gleaned from our observations, indeed, the observations for finite sample size, n, may as well be from a discrete distribution with n support points. Therefore we expand the parameter space to include discrete distributions. Now problem (1) above must be confronted; one possibility is to follow Kiefer and Wolfowitz (1956) and compare any two possible F by using as the dominating measure the sum of the two F's. The maximum likelihood estimator is then defined as the F which does best in all two-way comparisons. This is equivalent to maximizing $d\boldsymbol{P}_F = \prod_{i=1}^n \Delta F(X_i)$ subject to the constraint that F is a distribution function; the data force the estimator to be a discrete distribution. As is well known, the maximum likelihood estimator is the empirical distribution function which places mass $1/n$ at each of the observations. Note that because of the lack of a dominating measure, $d\boldsymbol{P}_F$ is not a likelihood in the usual sense of a probability density which, evaluated at the X_i's can be maximized over the entire parameter space.

A similar situation exists for the likelihood of failure time data. Consider the scenario of simple random censoring. In this case the observations are, $(X_1, D_1), \ldots, (X_n, D_n)$. Each X_i is the minimum of the failure time, T_i and a censoring time, C_i. The variable, D_i is one if $X_i = T_i$ and zero otherwise. Assume that the T_i's form a random sample from a distribution with survival function, S, hazard rate function, α, and cumulative hazard, A. Likewise the C_i's also form a random sample from a distribution G and are independent of the T_i's. As in Andersen et. al. (1993) the likelihood for the observations, $(X_1, D_1), \ldots, (X_n, D_n)$ is proportional to,

$$\prod_{t>0} \big(Y(t)\alpha(t)dt\big)^{dN(t)} \big(1 - Y(t)\alpha(t)dt\big)^{1-dN(t)}, \tag{1}$$

where $N(t) = \sum_{i=1}^{n} \boldsymbol{I}\{X_i \leq t,\ D_i = 1\}$ and $Y(t) = \sum_{i=1}^{n} \boldsymbol{I}\{X_i > t\}$. This likelihood depends on absolute continuity of A and there does not exist a maximum likelihood estimator in the class of absolutely continuous cumulative hazards, i.e. one can make the above likelihood arbitrarily large by appropriate choice of α. If we allow A to belong to the class of all cumulative hazard functions the estimator will be discrete as in the previous paragraph. Following the Kiefer-Wolfowitz approach, Johansen (1978), showed that the we may equivalently maximize the function,

$$\prod_{t>0} \big(Y(t)\Delta A(t)\big)^{\Delta N(t)} \big(1 - \Delta A(t)\big)^{Y(t)-\Delta N(t)}, \tag{2}$$

where the differentials have been replaced by differences due to the discreteness of the estimator. Note that (2) corresponds to the product of a sequence of conditional binomial probabilities, which arise from considering the distribution of $\Delta N(t)$ given the past up to time t, to be Binomial$\big(Y(t), \Delta A(t)\big)$. As explained by Andersen et. al. (1993, p. 180, 222), (2) is a likelihood if the failure time distribution is discrete. Kaplan and Meier (1958) use this likelihood in order to derive the estimator of the survival function, $S(t)$ which expressed as a function of A is $S(t) = \prod_{0<s\leq t} \big(1 - dA(s)\big)$. Thomas and Grunkemeier (1975) explain how, using the above likelihood, a likelihood ratio confidence interval for a survival probability can be constructed. In their notation, p_i is $1 - \Delta A(t)$ for t equal to the ith smallest observed failure time.

In (1) note that $\prod_{t>0} \big(1 - Y(t)\alpha(t)dt\big)^{1-dN(t)} = \exp\big(-\int_{t>0} Y(t)dA(t)\big)$. So another plausible extension which is discussed by Andersen et. al. (p. 223-226) is,

$$\prod_{t>0} \big(Y(t)\Delta A(t)\big)^{\Delta N(t)} \exp\Big(-\int_{t>0} Y(t)dA(t)\Big). \tag{3}$$

This extension is in the form of a likelihood of a sequence of conditional Poisson trials; $\Delta N(t)$ conditional on the past up to time t is Poisson$\big(Y(t)\Delta A(t)\big)$. The Poisson extension can be viewed as an approximation to the binomial version since for a continuous A, differences in A over small time intervals will be small relative to $Y(t)$. Unlike the binomial extension, this extension does not correspond to a discrete distribution for the failure time. However Johansen (1983) shows that the Poisson extension corresponds to the probability distribution of a dynamical Poisson process.

When A is absolutely continuous, both (2) and (3) are seen to be discrete approximations to the likelihood, (1) by interpreting $\Delta A(t)$ as $A(t+h) - A(t)$ for small h. However, neither extension is a likelihood in the sense of a density function for data from a continuous failure time distribution. It can be shown that both extensions lead to the same nonparametric maximum likelihood estimator, the Nelson-Aalen estimator given by, $\hat{A}(t) = \int_0^t Y(s)^{-1} dN(s)$. Therefore both extensions lead to the Kaplan-Meier estimator of S, $\hat{S}(t) = \prod_{0<s\le t} \left(1 - \Delta\hat{A}(s)\right)$ as the nonparametric maximum likelihood estimator of S. In general this can not be expected to occur; indeed the two extensions do not lead to algebraically equivalent estimators of A under $H_0 : A(t) = \theta_0$.

Both extensions possess the nice properties that, 1) there is an explicit estimator for the nonconstrained parameter, 2) the constraint can be written so that the score equation is of the form of a binomial or Poisson score and 3) the constrained estimator is explicit for a fixed Lagrange multiplier. As a result one is able to eliminate the constrained estimator of A or S from the likelihood ratio test, resulting is much simplification. This is exactly the situation in Owen's (1988) empirical likelihood; in that setting, constraints which are linear in F allow one to eliminate the constrained (under H_0) estimator of F from the likelihood ratio test.

In the failure time setting, one might question the use of the Poisson extension, since this version does not correspond to a discrete probability distribution for the failure time distribution. The Poisson extension to the likelihood does have advantages. First note that although there is no restriction on the height of the hazard function of a continuous failure time distribution, the binomial extension restricts $\Delta A(t)$ to be less than one. The Poisson extension makes no such restriction. The binomial extension makes this restriction for precisely the same reason it might be considered favorable; this extension corresponds to a discrete probability distribution for the failure time and the hazard function of a discrete failure time distribution is necessarily bounded above by one. This artificial restriction is not of any importance when one considers unconstrained estimation of A but when for example $A(t_0)$ is constrained to be equal to x_0, then if enough failures prior to t_0 do not occur in the data, the maximum likelihood estimator from the binomial extension may not even exist. The Poisson extension does not suffer from this problem. Because the Poisson extension does not enforce the restriction that $\Delta A(t) < 1$, it is available for more complicated models such as the frailty model (Andersen et. al., 1993) whereas the binomial is not. As a result we consider four possibilities in the next section: confidence intervals for both $A(t_0)$ and $S(t_0)$ produced by each of the binomial and Poisson extensions to the likelihood.

3. Likelihood ratio statistic

The analysis is simpler if we consider a general null hypothesis of the form, $H_0 : \boldsymbol{T}(A) = \theta_0$ where $\boldsymbol{T}$ is a function from the set of cumulative hazards to the reals. So $\boldsymbol{T}(A)$ is either $A(t_0)$ or $S(t_0) = \prod_{s \leq t_0}(1 - dA(s))$. Let $L_P(A)$, $L_B(A)$, be the logarithms of the Poisson and binomial extensions, respectively. If $\hat{A}_0$ denotes the estimator of A under H_0, then the likelihood ratio statistic is $2\big(L_P(\hat{A}) - L_P(\hat{A}_0)\big)$ for the Poisson extension and similarly for the binomial extension. From (3), we see that the likelihood ratio statistic for the Poisson extension is,

$$lrt_P = 2\int_0^{\tau} -\ln\left(\Delta\hat{A}_0(t)/\Delta\hat{A}(t)\right) - \left[1 - \Delta\hat{A}_0(t)/\Delta\hat{A}(t)\right] dN(t), \tag{4}$$

where $\Delta\hat{A}(t) = \Delta N(t)/Y(t)$. Likewise the likelihood ratio statistic for the binomial extension is

$$lrt_B = 2\int_0^{\tau} -\ln\left(\Delta\hat{A}_0(t)/\Delta\hat{A}(t)\right) + (\Delta\hat{A}(t)^{-1} - 1)\ln\left[\frac{1 - \Delta\hat{A}(t)}{1 - \Delta\hat{A}_0(t)}\right] dN(t). \tag{5}$$

To derive an equation for $\hat{A}_0$, we use the Lagrange multiplier method.

Two equalities are key in both computation of the likelihood ratio statistic and in a asymptotic analysis of the statistic. The first equality is derived from the derivative of the log likelihood set equal to the Lagrange multiplier times the derivative of the constraint, $\boldsymbol{T}(A) - \theta_0$ and yields for our two likelihoods,

$$\Delta\hat{A}_0(t) = \left[1 + \lambda\overline{Y}(t)^{-1}\dot{\boldsymbol{T}}_{\hat{A}_0}(t)\right]^{-1}\Delta\hat{A}(t) \tag{6}$$

where $\overline{Y}(t) = Y(t)/n$ and $\dot{\boldsymbol{T}}$ is given Table 1, and λ is a Lagrange multiplier. Equation (6) and the entries in Table 1 are verified below. The second equality is the constraint and can be written as

$$\boldsymbol{T}(\hat{A}_0) = \boldsymbol{T}\Big(\int_0^{\cdot} \big[1 + \lambda \overline{Y}(t)^{-1} \dot{\boldsymbol{T}}_{\hat{A}_0}(t)\big]^{-1} d\hat{A}(t)\Big) = \theta_0. \tag{7}$$

Table 1: Values of $\dot{\boldsymbol{T}}_{\hat{A}_0}(t)$.

	Likelihood Version	
Constraint	Binomial	Poisson
$\prod_{s\le t_0}\big(1-\Delta\hat{A}_0(s)\big)=\theta_0$	$\mathbf{I}\{t \le t_0\}\theta_0$	$\mathbf{I}\{t \le t_0\}\theta_0\big(1-\Delta\hat{A}_0(t)\big)^{-1}$
$\hat{A}_0(t_0)=\theta_0$	$\mathbf{I}\{t \le t_0\}\big(1-\Delta\hat{A}_0(t)\big)$	$\mathbf{I}\{t \le t_0\}$

Equation (6) yields the interpretation that the constraint alters $\hat{A}$ by changing the risk set. At each time t, $Y(t)$ is the number of individuals at risk of failure and $\Delta\hat{A}(t)$ is the number of failures occurring at time t divided by the number of individuals at risk of failure. Therefore, $\Delta\hat{A}_0(t)$ is the number of failures occurring at time t divided by the size of the risk set, $Y(t)$ plus $n\lambda\dot{\boldsymbol{T}}_{\hat{A}_0}(t)$.

Note that for $\boldsymbol{T}(A)$ equal to the cumulative hazard or the survival function, equation (6) is either a linear or quadratic polynomial in $\hat{A}_0(t)$. Therefore in order to calculate the likelihood ratio statistic, we need only the value of λ. It is easy to estimate λ from equation (7) because both constraints are monotonic in λ.

To verify (6) and the contents of Table 1, we calculate both a derivative of the log likelihood and a derivative of the constraint by differentiating at each $\Delta A(t)$, t an observed failure time. The derivative of $L_P(A)$ evaluated at $A = \hat{A}_0$, is given by $\Delta N(t)/\Delta\hat{A}_0(t) - Y(t)$ for each t an observed failure time. Similarly the derivative of $L_B(A)$ with respect to each $\Delta A(t)$ and evaluated at $A = \hat{A}_0$ is, $\Delta N(t)/\Delta\hat{A}_0(t) - \big(Y(t) - \Delta N(t)\big)/\big(1-\Delta\hat{A}_0(t)\big)$. If $\boldsymbol{T}(A) = A(t_0)$ we get that the derivative of $\boldsymbol{T}(A)$ with respect to each $\Delta A(t)$ and evaluated at $A = \hat{A}_0$ is $\boldsymbol{I}\{t \le t_0\}$ and if $\boldsymbol{T}(A) = \pi_{t\le t_0}(1 - dA(t))$, the derivative is $-\boldsymbol{I}\{t \le t_0\}\theta_0\big(1-\Delta\hat{A}_0(t)\big)^{-1}$.

The Lagrange multiplier method says that the derivative of the log likelihood is proportional to the derivative of the constraint. Below we write this in a convenient form for both the Poisson likelihood and the binomial likelihood. First the Poisson extension:

$$\Delta N(t) - Y(t)\Delta\hat{A}_0(t) = n\lambda\dot{\boldsymbol{T}}_{\hat{A}_0}(t)\Delta\hat{A}_0(t) \tag{8}$$

and the binomial extension:

$$\Delta N(t) - \big(Y(t) - dN(t)\big)\big(1 - \Delta\hat{A}_0(t)\big)^{-1}\Delta\hat{A}_0(t) = n\lambda\dot{\boldsymbol{T}}_{\hat{A}_0}(t)\big(1 - \Delta\hat{A}_0(t)\big)^{-1}\Delta\hat{A}_0(t) \tag{9}$$

for some λ. The additional term, $\big(1 - \Delta\hat{A}_0(t)\big)^{-1}$ in the derivative of the constraint in (9) is to preserve the form of the binomial score. Note the n placed next to λ in both equations; this is for convenience. Other constants in t could have been placed next to λ instead. Note that the contents of Table 1 follow from (8) and (9). Equation (6) follows from equations (8) and (9) by solving for $\Delta\hat{A}_0(t)$.

The confidence set for $\boldsymbol{T}(A)$ is the set of all θ_0 for which the likelihood ratio statistic is less than or equal to the appropriate percentile of the chi-squared distribution. Because the two constraints are continuous in A, the likelihood ratio tests resulting from both the Poisson and binomial extensions yield confidence sets which are intervals. This fact follows from the concavity of the two likelihood functions. We prove this in the appendix.

Furthermore the confidence interval for a survival probability is closed since (1) the parameter space for fixed sample size is a sequence of jump sizes each in the closed interval, zero to one, (2) the likelihood is continuous in the jump sizes and (3) the constraint is continuous. If the binomial extension is used to construct a confidence interval for the survival probability at time t, then as Thomas and Grunkemeier (1975) illustrate, the confidence interval will contain zero if and only if t is equal to the last observed failure and this failure occurs after the last observed censoring. This is, however, not the case for the Poisson extension. For example, if t is equal to the last observed failure which is followed by a censoring then zero may be in the confidence interval.

The confidence interval for the cumulative hazard at time t will be $[0, \infty)$ if there are no failures prior to time t. Otherwise the Poisson extension always leads to a finite closed confidence interval,since the likelihood ratio statistic can be made as large as desired by choice of a large θ_0. The binomial extension does not preclude an infinite confidence interval. For example, an infinite confidence interval can occur when there is only one observed failure and all censorings occur prior to this failure. In this case the binomial likelihood ratio test can not be made as large as desired by the choice of θ_0 large.

Theorem

Assume that $G(\tau-)$ is less than one and $S(\tau)$ greater than zero, then under H_0 : $\boldsymbol{T}(A_0) = \theta_0$, both lrt_P and lrt_B converge in distribution to a chi-squared distribution with one degree of freedom.

Remarks:

1) In this theorem, τ is finite. Suppose that $T(A) = S(t_0)$. As discussed earlier, zero will be in the confidence interval if and only if the last observed event is a failure and this failure occurs at time t_0 or earlier. The assumption, $S(\tau) > 0$, implies that for large sample sizes, zero will not be in the confidence interval for $S(t_0)$. Note that a likelihood ratio test of $H_0 : S(t_0) = 0$ is not feasible, since under H_0 all failures and observed censorings must occur prior to t_0 and we accept H_0 with probability one.

2) It may seem surprising that the Lagrange multiplier, λ, plays such a important role in the computations and in the statistical analysis. To see why this should occur, consider the finite dimensional problem with parameter γ and the null hypothesis, $H_0 : \boldsymbol{T}(\gamma) = \theta_0$. Let $\hat{\gamma}_0$ be the estimator of γ when $\boldsymbol{T}(\gamma) = \theta_0$ and let $\hat{\gamma}_1$ be the unconstrained estimator. Write the derivative of $\boldsymbol{T}$ evaluated at the true value of $\gamma = \gamma_0$ as $\dot{\boldsymbol{T}}_{\gamma_0}$ and denote the information matrix by $\mathbf{I}$. Then under the usual regularity conditions,

$$\hat{\gamma}_1 - \hat{\gamma}_0 \doteq \mathbf{I}^{-1}\dot{\boldsymbol{T}}_{\gamma_0}\lambda$$

and

$$\lambda \doteq \frac{\boldsymbol{T}(\hat{\gamma}_1) - \theta_0}{\dot{\boldsymbol{T}}_{\gamma_0}\mathbf{I}^{-1}\dot{\boldsymbol{T}}_{\gamma_0}}.$$

The second equation implies that λ converges to zero in probability. Using this fact and the first equation, the likelihood ratio test is approximately $2\big(L(\hat{\gamma}_0 + \mathbf{I}^{-1}\dot{\boldsymbol{T}}_{\gamma_0}\lambda) - L(\hat{\gamma}_0)\big)$ and the asymptotic distribution can be derived via a Taylor series about λ only instead of a Taylor series in γ. These observations become especially useful when γ is of high dimension as is case here.

3) The proof of this theorem is given in the appendix. An alternate proof given in Murphy (1994) is valid for a wider class of constraints than the two discussed here. In particular, the following proper-

ties of the constraints are used Murphy's (1994) proof. The constraint, $\boldsymbol{T}$, is compactly differentiable at the point A_0 with derivative, $d\boldsymbol{T}(A_0)(A) = \int_0^\tau \dot{\boldsymbol{T}}_{A_0}(t)dA(t)$. For each continuous A satisfying $\boldsymbol{T}(A) = \theta_0$, $\dot{\boldsymbol{T}}_A(t)$ is nonzero for t in an nonempty interval; $\dot{\boldsymbol{T}}_A(t)$ is either right continuous with a finite left hand limit or left continuous with a finite right hand limit, at each t in $(0, \tau)$; $\dot{\boldsymbol{T}}_A$ is of bounded variation on $[0, \tau]$ and if as m increases, A_m converges in supremum norm to A, then $\dot{\boldsymbol{T}}_{A_m}$ converges in supremum norm to $\dot{\boldsymbol{T}}_A$. Both constraints, $\pi_{t \leq t_0}\left(1 - dA(t)\right) = \theta_0$ and $A(t_0) = \theta_0$ satisfy the above conditions; see Andersen et. al. (1993, section 2.8) for compact differentiability of the former constraint. Other constraints which satisfy these conditions would be continuous functionals of either S or A such as the expected time lived in a restricted interval, $\int_0^\tau S(t)dt = \theta_0$.

4. Examples

In this section two likelihood ratio based confidence intervals are illustrated. In particular we consider a likelihood ratio based confidence interval for a quantile of $F = 1 - S$, and a pointwise confidence interval for A. In each case we construct three confidence intervals, one based on the binomial extension, the second based on the Poisson extension of the likelihood and the third based on a competitor. We will not evaluate the pointwise confidence intervals for survival probabilities as this was done by Thomas and Grunkemeier (1975). In order to illustrate these confidence sets, we generate the failure times from an exponential distribution with mean one and censoring times from a uniform distribution on the interval, zero to three. This will result in an average of one fourth of the failure times being censored at the upper quartile of the exponential distribution. Both simulations include 1000 samples of size 25 each.

Note that we can invert the test of $H_0 : S(t_0) = p_0$ in two ways, either we consider all p_0 for which H_0 is not rejected or we consider all t_0 for which H_0 is not rejected. The former method yields a confidence interval for the survival probability, $S(t_0)$ and the latter method yields a confidence interval for $S^{-1}(p_0)$, where $S^{-1}(p_0)$ is defined by $\inf\{t \geq 0 : S(t) \leq p_0\}$. For example, the upper quartile of the failure time distribution is given by $S^{-1}(.25)$. The confidence interval for $S^{-1}(p_0)$ is the set $\{t_0 : H_0 : S(t_0) = p_0 \text{ is accepted}\}$ or denoting the likelihood ratio test by $lrt(t_0)$, then the $(1 - \alpha)$th percent confidence interval is given by $\{t_0 : lrt(t_0) \leq c_\alpha^2\}$ where c_α is the $(1-\alpha/2)$th percentile of a standard normal distribution.

To calculate $lrt(t_0)$ for an arbitrary value of t_0 and constraint $\boldsymbol{T}(A) = \theta_0$, substitute for $\Delta\hat{A}_0$ in (7) by using the values from Tables 1 and 4. The equation, (7), will now have only one unknown, λ. Next solve (7) for λ; this is easy since the constraint is monotone in λ. Now substitute the solution for λ and the value of $\Delta\hat{A}_0$ from Table 4 into the equation for the likelihood ratio (either equation, (4) or (5)) to get $lrt(t_0)$). Note that the integral in the likelihood ratio has upper endpoint equal to t_0 since $\Delta\hat{A}_0(t) = \Delta\hat{A}(t)$ for t larger than t_0. Since lrt is constant between the observed failure times, we have that if failure time t_k is in the confidence interval then $[t_k, t_{k+1})$ is in the confidence interval. So the confidence interval is the union of half open intervals with left endpoints corresponding to those failure times t_k for which the null hypothesis is accepted.

Any test of $H_0 : S(t_0) = p_0$ can be inverted to yield a confidence interval for $S^{-1}(p_0)$; Brookmeyer and Crowley, (1982), invert a test of $H_0 : S(t_0) = p_0$ based on asymptotic normality of the Kaplan-Meier estimator. A competitor to the likelihood ratio statistic is the inversion of the test of $H_0 : S(t_0) = p_0$ based on the asymptotic normality of $\log\big(-\log(\hat{S}(t_0))\big)$. Kalbfleisch and Prentice (1980) suggest that this transform will be better approximated by a normal distribution than an untransformed version. The small sample properties of confidence intervals for survival probabilities based on this transformation and other transformations is considered by Borgan and Liestøl (1992).

The confidence interval will be one sided whenever the test statistic evaluated at the last observed failure time is not significant. In Table 2, we see that for 95% confidence intervals this occurs in over 90% of the 1000 samples. The binomial extension of the likelihood yields a more easily computable confidence interval ((6) is linear in $\Delta\hat{A}_0(t)$), and our simulations indicate that this extension combined with the use of the chi-squared reference distribution produces as good as or better, confidence interval than the log-log transform.

Table 2: 95% Confidence Intervals for the Upper Quartile of the Failure Distribution

		Form of Confidence Interval		
Method	Confidence Level	$[\theta_L, \infty)$	$[\theta_L, \theta_U)$	Width[a]
Binomial LRT	.95	913	87	.94
Poisson LRT	.98	930	70	1.18
Log-Log Transform	.96	913	87	.95

[a]NOTE: The average width of the finite intervals

Andersen et. al. (p. 277, 1993) illustrate two methodologies for the confidence interval. The first method is as in the simulation above; the confidence interval is derived from an inversion of the test statistic based on asymptotic normality of $\log\big(-\log(\hat{S}(t))\big)$. The second method estimates the percentile by inverting $\hat{S}$ and then forms the confidence interval from this estimator plus and minus a z percentile times the estimated standard error. The two methods are used to give a confidence interval for the lower quartile of the survival function for 80 male patients whose melanomas were removed in a radical operation. Of the 80 men, 29 were observed to die of the cancer and the death times of the remaining 59 were censored. The data are given in Andersen et. al. along with a detailed description (pp. 11, 709). The 95% confidence interval based on the log-log transform is 2.13 years to 5.76 years and the 95% confidence interval based on the estimated lower quartile is 1.91 years to 4.81 years. The 95% confidence intervals resulting from the binomial and Poisson extensions are very similar to the confidence interval based on the log-log transform and are 2.19 years to 5.81 years and from 2.15 years to 5.81 years, respectively.

Next consider confidence intervals for the cumulative hazard function. Bie et. al. (1987) review several different methods and give Monte Carlo simulation results for pointwise confidence intervals for the cumulative hazard. One of the confidence intervals which performs well is based on the asymptotic normality of variance stabilizing transformation for the no censoring situation. This is $\arcsin\big(\exp(-\hat{A}(t_0)/2)\big)$. In the simulation below t_0 is set to the upper quartile of the exponential distribution.

The calculation of the likelihood ratio test statistic for a specified value of θ_0 is similar to the previous example in that one uses Tables 1 and 4 to substitute into equation (7) and solve for λ. This value of λ is then used to calculate the likelihood ratio test statistic. As before the upper limit of the integral in the likelihood ratio test statistic is t_0 since $\Delta\hat{A}_0(t) = \Delta\hat{A}(t)$ for t greater than t_0. One straightforward way of calculating the lower confidence bound from the likelihood ratio test is to begin at $\theta_0 = \hat{A}(t_0)$. Of course this value of θ_0 will be in the confidence interval; next step down to $\theta_0 = \hat{A}(t_0)/2$; if the likelihood ratio test does not reject this value of θ_0, again reduce θ_0 by a half. Once the test rejects the value of θ_0, move up one half of the way to the previous θ_0. Continue until convergence. Calculating the upper confidence bound is similar except the algorithm starts off by stepping one half the way up to an arbitrarily large value (in the simulation, three times the upper confidence limit of the competitor is used).

Table 3: 95% Confidence Intervals for Cumulative Hazard Function

Method	Confidence Level	Width[a]
Binomial LRT	.92	1.43
Poisson LRT	.94	1.72
Arcsin-Transform	.94	1.76

[a]NOTE: The average width of the finite intervals

In the above simulations, the use of the arcsin-transform resulted in an infinite confidence interval for 18 of the 1000 simulated data sets. Infinite confidence intervals occur in data sets in which the estimated standard deviation of $\arcsin\big(\exp(-\hat{A}(t_0)/2)\big)$ is very large. All of the other confidence intervals were finite. The poorer performance of the binomial based confidence interval is related to the constraint that jumps of the cumulative hazard be smaller than one. This constraint causes values of θ_0 in the null hypothesis, $H_0 : A(t_0) = \theta_0$ which are close to $\hat{A}(t_0)$ to be rejected, thereby resulting in an overly short confidence interval. The Poisson extension of the likelihood yields a more easily computable confidence interval ((6) is linear in $\Delta\hat{A}_0(t)$), and our simulations indicate that this extension combined with the use of the chi-squared reference distribution produces as good as or better, confidence interval than the arcsin-transform.

5. Summary

As illustrated here, there may be multiple extensions of the likelihood in the semiparametric setting, each of which may yield meaningful likelihood ratio hypothesis tests and subsequent confidence intervals. The choice of the best extension is difficult. It appears that in the failure time setting, the binomial extension is computationally easiest to use in constructing confidence intervals concerning the survival function and the Poisson extension is computationally easier to use in constructing confidence intervals concerning the cumulative hazard. In our simulations these two methods performed as well as or better than competitor based on asymptotic normality of the estimator. More through simulations are needed. However our limited simulations indicate that the choice of extension to the likelihood should depend on the parameter of interest. Work is needed on how one should choose the extension of the likelihood.

Note that Owen's (1988) empirical likelihood confidence interval is a likelihood ratio confidence interval in a nonparametric setting. In parametric settings and in the empirical likelihood setting, the coverage error of a likelihood ratio based confidence interval is of order n^{-1} (Hall and La Scala, 1990). DiCiccio, et. al. (1991) prove that the empirical likelihood ratio test is Bartlett correctable; that is, a simple correction can be made to the test so that the coverage error is of order n^{-2}. It is of interest to prove both of these results in this setting. Perhaps the work of Mykland (1994) will be useful in this regard.

Laska and Meisner, (1992) propose the use of a likelihood ratio based confidence interval for equality of cure rates. The methodology presented here should follow over to that situation. Another related area is the formation of likelihood ratio based confidence bounds for both the survival function and the cumulative hazard function. A theoretical analysis similar to the analysis given in the appendix indicates the possibility of such confidence bounds.

6. Appendix

Let $y(t) = S(t)\big(1 - G(t-)\big)$ and denote the true value of A by A_0. The proofs to follow repeatedly use the fact that $\sup_{0\le t\le\tau} |\hat{A}(t) - A_0(t)|$ and $\sup_{0\le t\le\tau} |\overline{Y}(t) - y(t)|$ converge in probability to zero.

Andersen et. al. (p. 193-4, 1993) illustrate that the assumptions of the theorem are sufficient for the consistency of $\hat{A}$ in supremum norm on $[0, \tau]$ and give the asymptotic distribution theory for $\sqrt{n}(\hat{A} - A_0)$.

Proof that the confidence sets are intervals

This proof is valid for continuous $\boldsymbol{T}$. Let m be the number of observed failure times. Put u_i equal to $\Delta \hat{A}_0(t)/\Delta A(t)$ and Y_i equal to $Y(t)$ for t the ith smallest failure time. The likelihood ratio test derived from the binomial extension can be written as, $lrt_B(\mathbf{u}) = 2\sum_{i=1}^{m} -\ln(u_i) - (Y_i - 1)\ln\left(\frac{Y_i - u_i}{Y_i - 1}\right)$ which is a sum of convex functions each in u_i. Let $\boldsymbol{T}^{-1}(\theta_0)$ be the set of m dimensional vectors, $\mathbf{u}$ for which $\boldsymbol{T}$ evaluated at an A with ith jump size u_i/Y_i is equal to θ_0; in the following we write $\boldsymbol{T}(\mathbf{u}) = \theta_0$.

The confidence set is given by the set of θ_0 for which $\min_{u \in \boldsymbol{T}^{-1}(\theta_0)} lrt_B(\mathbf{u}) \le c_\alpha^2$ where c_α is the $(1 - \alpha/2)$th percentile of a standard normal distribution. Since $lrt_B(\mathbf{u})$ is convex in $\mathbf{u}$, the set of $\mathbf{u}$ for which $lrt_B(\mathbf{u}) \le c_\alpha^2$ is a closed convex set. This implies that the confidence set is convex as follows. Suppose θ_1 and θ_2 are in the confidence set. Let θ^* be a convex combination of the two θ's and let $\mathbf{u}_j$ satisfy $\boldsymbol{T}(\mathbf{u}_j) = \theta_j$ and $lrt_B(\mathbf{u}_j) \le c_\alpha$ for $j = 1, 2$. Define the function, $g(\epsilon) = \boldsymbol{T}(\epsilon \mathbf{u}_1 + (1 - \epsilon)\mathbf{u}_2)$. This function is continuous and $g(1) = \theta_1$ and $g(0) = \theta_2$. Therefore there exists an ϵ^* in $(0, 1)$ for which $g(\epsilon^*) = \theta^*$ and since $lrt_B(\mathbf{u}) \le c_\alpha$ is a closed convex set, $\epsilon^* \mathbf{u}_1 + (1 - \epsilon^*)\mathbf{u}_2$ must be a member of this set. We have found a $\mathbf{u}$ for which $lrt_B(\mathbf{u}) \le c_\alpha$ and $\boldsymbol{T}(\mathbf{u}) = \theta^*$; this means θ^* must be a member of the confidence set. A similar proof holds for the Poisson extension. Note that this proof does not require that $\dot{\boldsymbol{T}}$ be independent of A.

The function $\boldsymbol{T}(A) = S^{-1}(p_0)$ is not continuous in A. However, three facts, (1) S is a continuous function of A, (2) $S(t)$ is non-increasing in t and (3) $lrt(\mathbf{u}) \le c_\alpha$ excludes jumps of A which are of size zero, yield the existence of the ϵ^* in the above proof.

Proof of theorem

Since A_0 is assumed to be continuous, we assume that $\Delta N(t)$ is a 0, 1 random variable in the following. In order to derive the asymptotic distribution, we use equation (6) to eliminate $\Delta A_0(t)$

from the likelihood ratio test statistic. That is, we solve for $\Delta A_0(t)$ as a function of λ in (6) and then substitute $\Delta A_0(t)$ into the likelihood ratio to get a function in λ only. For large n, the formula for $\Delta A_0(t)$ is given in Table 4.

Table 4 : Values for $\Delta \hat{A}_0(t),\ t \leq t_0$

	Likelihood Version	
Constraint	Binomial	Poisson
$S(t_0) = \theta_0$	$\big(Y(t) + n\lambda\big)^{-1} \Delta N(t)$	$\frac{1+n\lambda+Y(t)-\sqrt{(1+n\lambda+Y(t))^2-4Y(t)}}{2Y(t)} \Delta N(t)$
$A(t_0) = \theta_0$	$2\Big(n\lambda + Y(t) + \sqrt{(n\lambda + Y(t))^2 - 4n\lambda}\Big)^{-1} \Delta N(t)$	$\big(Y(t) + n\lambda\big)^{-1} \Delta N(t)$

In particular, the formula for $\Delta \hat{A}_0(t)$ for the Poisson version and constraint, $S(t_0) = \theta_0$ is only correct for large n. For small n the second root of the quadratic equation defined by (6) may be the maximum likelihood estimator of $\Delta A_0(t)$. Note that setting $\lambda = 0$ in all four equations will yield the Nelson-Aalen estimator of A.

Recall that λ is determined by the constraint. For both likelihood extensions, $\hat{A}_0(t_0)$ is decreasing function of λ, whereas $\prod_{s \leq t_0} \big(1 - \Delta \hat{A}_0(s)\big)$ is an increasing function of λ. We use this monotonicity to prove that λ will converge to zero in probability. To illustrate the method, consider the constraint, $\prod_{s \leq t_0} \big(1 - \Delta \hat{A}_0(s)\big) - \theta_0 = 0$ and the Poisson likelihood extension. This is the hardest of the four proofs.

The first step is to prove that λ converges to zero in probability. Let ϵ be positive and smaller than $y(t_0)$. Set

$$A_{(-\epsilon)}(\cdot) = \int_0^{\cdot} \frac{1 + n(-\epsilon) + Y(t) - \sqrt{(1 + n(-\epsilon) + Y(t))^2 - 4Y(t)}}{2Y(t)} dN(t)$$

and likewise set

$$A_{\epsilon}(\cdot) = \int_0^{\cdot} \frac{1 + n\epsilon + Y(t) - \sqrt{(1 + n\epsilon + Y(t))^2 - 4Y(t)}}{2Y(t)} dN(t).$$

In the next paragraph we see that $A_{(-\epsilon)}(\cdot)$ converges almost surely in supremum norm to $\int_0^{\cdot} y(t)\big((-\epsilon)+ y(t)\big)^{-1} dA_0(t)$ and that $A_{\epsilon}(\cdot)$ converges almost surely in supremum norm to $\int_0^{\cdot} y(t)\big(\epsilon + y(t)\big)^{-1} dA_0(t)$. The continuity of the product integral (see Andersen et.al., p. 114) implies then that, $\prod_{s \leq t_0} (1 -$

$\Delta A_{(-\epsilon)}(s)\big) - \theta_0$ converges to the negative value, $\pi_{s\leq t_0}\left(1 - y(t)\big((-\epsilon)+y(t)\big)^{-1}dA_0(t) - \theta_0\right.$. Likewise $\prod_{s\leq t_0}\big(1-\Delta A_\epsilon(s)\big) - \theta_0$ converges to the positive value, $\pi_{s\leq t_0}\left(1 - y(t)\big(\epsilon + y(t)\big)^{-1}dA_0(t) - \theta_0\right.$. By the monotonicity of $\hat{A}_0(t_0)$ in λ, $P\big[\lambda\in[-\epsilon,\epsilon]\big]$ is equal to the probability of the intersection of $\prod_{s\leq t_0}\big(1-\Delta A_\epsilon(s)\big) - \theta_0$ greater than zero and $\prod_{s\leq t_0}\big(1-\Delta A_{(-\epsilon)}(s)\big) - \theta_0$ less than zero. Therefore $P\big[\lambda\in[-\epsilon,\epsilon]\big]$ converges to one as the sample size increases. Since the above is true for any small positive ϵ, we have that λ converges to zero in probability.

To prove that for small positive ϵ, $A_{(-\epsilon)}(\cdot)$ converges almost surely, in supremum norm, to $\int_0^{\cdot} y(t)\big((-\epsilon)+y(t)\big)^{-1}dA_0(t)$, first rewrite the integrand in the definition of $A_{(-\epsilon)}$ by dividing and multiplying by n. We get $A_{(-\epsilon)}(\cdot) = \int_0^{\cdot}\frac{n^{-1}+(-\epsilon)+\bar{Y}(t)-\sqrt{(n^{-1}+(-\epsilon)+\bar{Y}(t))^2-4n^{-1}\bar{Y}(t)}}{2n^{-1}}d\hat{A}(t)$. Use a Taylor series in n^{-1} about zero to write the numerator in the integrand as, $\left(1-\frac{2(\lambda'(t)+(-\epsilon)+\bar{Y}(t))-4\bar{Y}(t)}{2\sqrt{(\lambda'(t)+(-\epsilon)+\bar{Y}(t))^2-4\lambda'(t)\bar{Y}(t)}}\right)n^{-1}$ for $\lambda'(t)$ between zero and n^{-1}. Now use the strong law of large numbers to get that the integrand converges in supremum norm to $y(t)\big((-\epsilon)+y(t)\big)^{-1}$; also $\hat{A}$ converges in supremum norm to $A_0(t)$. Since $y(t)\big((-\epsilon)+y(t)\big)^{-1}$ is of bounded variation, we combine the above to get the desired convergence. The proof that $A_\epsilon(\cdot)$ converges almost surely in supremum norm to $\int_0^{\cdot} y(t)\big(\epsilon+y(t)\big)^{-1}dA_0(t)$ is similar.

The second step is to derive the asymptotic distribution of $\sqrt{n}\lambda$ from the constraint, $\prod_{s\leq t_0}\big(1-\Delta\hat{A}_0(s)\big)-\theta_0 = 0$. Recall that the only unknown in this constraint is λ. Use a Taylor series in λ about zero, to get that for some λ' between λ and zero, $\prod_{s\leq t_0}\big(1-\Delta\hat{A}(s)\big)-\theta_0 =$

$$\prod_{s\leq t_0}\big(1-\Delta A_{\lambda'}(s)\big)\int_0^{t_0}\Big(n^{-1}\sqrt{(1+n\lambda'+Y(t))^2-4Y(t)}\big(1-\Delta A_{\lambda'}(t)\big)\Big)^{-1}dA_{\lambda'}(t)\lambda$$

where $A_{\lambda'}(\cdot) = \int_0^{\cdot}\frac{1+n\lambda'+Y(t)-\sqrt{(1+n\lambda'+Y(t))^2-4Y(t)}}{2Y(t)}dN(t)$. Using the same Taylor series argument as used in showing consistency of λ, one can show that $A_{\lambda'}$ and the above integrand converge almost surely in supremum norm to A_0 and to y^{-1}, respectively. The continuity in $A_{\lambda'}$ of the terms on the right side of the above equality combined with the assumed continuity of $A_0(t)$ in t implies that

$$\sqrt{n}\Big(\prod_{s\leq t_0}\big(1-\Delta\hat{A}(s)\big)-\theta_0\Big) = \Big(\theta_0^2\int_0^{t_0}y(t)^{-1}dA_0(t)+o(1)\Big)\sqrt{n}\lambda$$

where $o(1)$ is a term converging almost surely to zero. Therefore the asymptotic distribution of $\sqrt{n}\lambda$ is the distribution of $\Big(\theta_0^2\int_0^{t_0}y(t)^{-1}dA_0(t)\Big)^{-1}U(t_0)$. The distribution of $U(t_0)$ follows from the results on p. 263 of Andersen et. al. (1993). That is, $U(t_0)$is a mean zero normal random variable with variance, $\theta_0^2 \int_0^{t_0} y(t)^{-1} dA_0(t)$.

The last step is to use the consistency of λ to zero and its asymptotic distribution to show that the likelihood ratio test based on the Poisson extension has an asymptotic chi-squared distribution. This derivation is based on a Taylor series in λ about zero. From equation (4), the Taylor series yields,

$$n\lambda^2\theta_0 \int_0^{t_0} \frac{-n^{-1} - \lambda' + \bar{Y}(t)}{\left(\left(n^{-1} + \lambda' + \bar{Y}(t)\right)^2 - 4n^{-1}\bar{Y}(t)\right)^{3/2}} d\bar{N}(t).$$

The integral converges almost surely to $\int_0^{t_0} y(t)^{-1} dA_0(t)$. This combined with the asymptotic distribution of λ yields an asymptotic chi-squared distribution for the likelihood ratio test.

5. References

Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer Verlag.

Bie, O., Borgan, Ø. and Liestøl, K. (1987), "Confidence Intervals and Confidence Bands for the Cumulative Hazard Rate Function and their Small Sample Properties," *Scandinavian Journal of Statistics*, **14**, 221-233.

Borgan, Ø. and Liestøl, K. (1990), "A Note on Confidence Intervals and Bands for the Survival Function Based on Transformations," *Scandinavian Journal of Statistics*, **17**, 35-42.

Brookmeyer, R. and Crowley, J.J. (1982), "A Confidence Interval for the Median Survival Time," *Biometrics*, **38**, 29-41.

DiCiccio, T.J., Hall, P. and Romano, J.P. (1991), "Empirical Likelihood is Bartlett Correctable," *Annals of Statistics*, **19**, 1053-1061.

Hall, P. and La Scala, B. (1990), "Methodology and Algorithms of Empirical Likelihood," *International Statistical Review*, **58**, 109-127.

Johansen, S. (1978), "The Product Limit Estimator as Maximum Likelihood Estimator," *Scandinavian*

Journal of Statistics, **5**, 195-199.

Johansen, S. (1983), "An Extension of Cox's Regression Model," *International Statistical Review*, **51**, 258-262.

Kalbfleisch, J.D. and Prentice, R. L. (1980), *The Statistical Analysis Failure Time Data*, New York: Wiley.

Kaplan, E.L. and Meier, P. (1958), "Non-parametric Estimation from Incomplete Observations," Journal of the American Statistical Association, **53**, 457-481,562-563.

Kiefer, J. and Wolfowitz, J. (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Nuisance Parameters," *Annals of Mathematical Statistics*, **27**, 887-906.

Laska, E.M. and Meisner, M.J. (1992), "Nonparametric Estimation and Testing in a Cure Model," *Biometrics*, **48**, 1223-1234.

Murphy, S. (1994), "Likelihood Ratio Based Confidence Intervals in Survival Analysis," *Technical Report # 94-02*, Dept. of Statistics, The Pennsylvania State Univ.

Mykland, P. (1994), "Dual Likelihood," Unpublished manuscript.

Owen, A.B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, **75**, 237-249.

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.