Nonparametric Hypothesis Testing and Confidence Intervals with Doubly Censored Data

KUN CHEN AND MAI ZHOU

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027, U.S.A.

SUMMARY

The non-parametric maximum likelihood estimator (NPMLE) of the distribution function with doubly censored data can be computed using self-consistent algorithm (Turnbull, 1974). We extend the self-consistent algorithm to include a constraint on the NPMLE. We then show how this can be used to construct confidence intervals and test hypotheses based on the NPMLE via empirical likelihood ratio. Finally we present some numerical comparison of the performance of the above method with another method that make use of the influence functions.

AMS 1991 Subject Classification: Primary 62G10; secondary 62G05.

Key Words and Phrases: Modified Self-consistency, Empirical likelihood ratio, Influence function.

1. Introduction

When data are doubly censored, the nonparametric maximum likelihood estimator (NPMLE) of the distribution function have no explicit form and have to be computed via some iteration [Turnbull(1976), Chang and Yang(1987), Zhou(1995), Zhan and Wellner(1999)]. The estimation of the asymptotic variance of the NPMLE is even more involved [Chang(1990)].

Self-consistency algorithm is a general way of computing non-parametric distributional estimators when data are not completely observed [Efron (1967), Turnbull(1974), Tsai and Crowly(1985)]. In particular, a NPMLE of distribution function based on an i.i.d. sample must satisfy the self-consistent equation [Gill(1989)]. It is a special case of the EM algorithm when the parameter is the distribution function itself [Dempseter, Laird, and Rubin(1977)].

In this paper we extend the self-consistent algorithm to compute the NPMLE under a constraint $F(T) = \theta$ for doubly censored data. This constrained estimator is useful in the construction of empirical likelihood ratio. As an application we show how this in turn will allow us to find confidence intervals and test hypotheses for doubly censored data.

Another approach of constructing confidence intervals based on the NPMLE with doubly censored data is to estimate the influence function/variance of the NPMLE. But the influence function is not easy to estimate and this approach may not be very practical for larger samples. We will compare the two approaches for one example in this paper.

Doubly censored data can arise when observations are subject to both right and left censoring, i.e. the patients are only watched within a window of observational time, otherwise we only know it is below or above the window. Double censoring also arise from pairwise comparisons when there are right censoring in both group, as the following example explains.

Example: In paired comparison experiments, we have n pairs of observations based on two treatments. To estimate the treatment difference, it is customary to focus on the n pairwise differences. However, when right censoring occurs in either treatments, the pairwise difference can be right censored, or left censored, as the following table shows.

Trt 1	Trt 2	diff$(1-2)$
14+	6	8+
12	7	5
9	5+	4−

For a pair of observations, if the observation from treatment one is right-censored but treatment two is uncensored, their difference will be right-censored; if the observation from treatment one is uncensored but treatment two is right-censored, their difference will be left-censored; If both observations are uncensored, their difference is uncensored. If both observations are right-censored, their difference can take any value, we drop those data from further analysis.

Therefore, the differences of paired right-censored data can be doubly censored – having both left and right censored observations. Section 6 will treat this in more detail. $\diamondsuit$

The rest of this section is to formally introduce the relevant notation and basic assumptions.

Let $X_1, \cdots, X_n$ be positive random variables denoting the sample of lifetimes which is independent and identically distributed with a continuous distribution F_0. The censoring mechanism is such that X_i is observable if and only if it lies inside the interval $[Z_i, Y_i]$. The Z_i and Y_i are positive random variables with continuous distribution functions G_{L_0} and G_{R_0} respectively, and $Z_i \le Y_i$ with probability 1. If X_i is not inside $[Z_i, Y_i]$, the exact value of X_i cannot be determined. We only know whether X_i is less than Z_i or greater than Y_i and we observe Z_i or Y_i correspondingly.

The variable X_i is said to be left censored if $X_i < Z_i$ and right censored if $X_i > Y_i$. The available information may be expressed by a pair of random variables: $T_i,\ \delta_i$, where

$$T_i = \max(\min(X_i, Y_i), Z_i) \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } Z_i \le X_i \le Y_i \\ 0 & \text{if } X_i > Y_i \\ 2 & \text{if } X_i < Z_i \end{cases} \qquad i = 1, 2, \cdots, n. \tag{1.1}$$

The modified (constrained) self-consistent equation for doubly censored data is derived in Section 2. We generalize the self-consistent algorithm to deal with several constraints in Section 3. We then explain how the constrained self-consistent estimate may be used to obtain confidence intervals via empirical likelihood ratio in section 4. Influence function is discussed in Section 5. We carry out simulations and apply our algorithm to some doubly censored data in Section 6.

2. Modified Self-Consistent Equation Under a Constraint

Suppose for a given T and θ we are interested to test the hypothesis

$$H_0 : F(T) = \theta \qquad vs. \qquad H_1 : F(T) \neq \theta \ . \tag{2.1}$$

As we will see later, A key step to accomplish this is to be able to compute the NPMLE of $F(t)$ under H_0 from the data (1.1). We shall focus on computing NPMLE in this section.

First let us write down the log likelihood function ($\log L(F)$) for doubly censored data. The log likelihood function involves all n observations. However, We can decompose the likelihood function into two parts: for observations before and after time T.

$$\begin{aligned}
\log L(F) &= \log L_1 + \log L_2 \\
&= \sum_{\delta_i=1, i\le n_1} \log(w_i) + \sum_{\delta_i=0, i\le n_1} \log[1-F(t_i)] + \sum_{\delta_i=2, i\le n_1} \log[F(t_i-)] + \\
&\quad \sum_{\delta_i=1, i\le n_2} \log(v_i) + \sum_{\delta_i=0, i\le n_2} \log[1-F(s_i)] + \sum_{\delta_i=2, i\le n_2} \log[F(s_i-)] \\
&= \sum_{\delta_i=1, i\le n_1} \log(w_i) + \sum_{\delta_i=0, i\le n_1} \log[1-F(T)+F(T)-F(t_i)] \\
&\quad + \sum_{\delta_i=2, i\le n_1} \log[F(t_i-)] + \sum_{\delta_i=1, i\le n_2} \log(v_i) + \sum_{\delta_i=0, i\le n_2} \log[1-F(s_i)] \\
&\quad + \sum_{\delta_i=2, i\le n_2} \log[F(s_i-)-F(T)+F(T)]
\end{aligned} \tag{2.2}$$

where $\log L_1$ denotes the log likelihood function for n_1 observations before time T and $\log L_2$ denotes the log likelihood function for n_2 observations after time T ($n_1 + n_2 = n$). w_i is the jump at ith observation before time T. v_i is the jump at ith observation after time T. t_i is the observed time before time T. s_i is the observed time after time T.

Therefore, $w_i \ge 0$ and $\sum_{i=1}^{n_1} w_i = \theta$ corresponding to jumps before time T; similarly $v_i \ge 0$ and $\sum_{i=1}^{n_2} v_i = (1-\theta)$ for jumps after time T.

Notice that because $F(T) = \theta$, we have $F(T) - F(t_i) = w_{i+1} + \cdots + w_{n_1}$ and $F(s_i-) - F(T) = v_1 + v_2 + \cdots + v_{(i-1)}$. The first part of above likelihood, $\log L_1$, only involves w_i, the jumps of F before time T as in (2.3). Similarly the second part of the likelihood only involves the jumps of F after time T as in (2.4). Therefore we can maximize the whole likelihood in two *independent* steps: maximize $\log L_1$ and then maximize $\log L_2$.

If a distribution function, say $\hat{F}$, did maximize the log likelihood under H_0, then it must satisfy that the derivative of the log likelihood function at this $\hat{F}$ equal to 0. We define a directional change of w_i in the direction $h(.)$ and parametrized by λ to facilitate derivative (under H_0). Since w_i must sum up to θ, v_i must sum up to $(1-\theta)$, we define them as follows:

For jumps before time T, we define

$$w_i = w_i(\lambda) = \frac{\Delta\hat{F}(t_i)}{1+\lambda h(t_i)}\frac{\theta}{C(\lambda)} \qquad \text{with} \qquad C(\lambda) = \sum_{i=1}^{n_1} \frac{\Delta\hat{F}(t_i)}{1+\lambda h(t_i)} \ , \tag{2.3}$$

and $C(0) = \theta$.

For jumps after time T, we define

$$v_i = v_i(\lambda) = \frac{\Delta\hat{F}(s_i)}{1+\lambda h(s_i)}\frac{1-\theta}{D(\lambda)} \qquad \text{with} \qquad D(\lambda) = \sum_{i=1}^{n_2} \frac{\Delta\hat{F}(s_i)}{1+\lambda h(s_i)} \ . \tag{2.4}$$

Clearly $D(0) = 1-\theta$.

Remark 2.1: We may also use the following definition in the computation of derivatives.

$$w_i^*(\lambda) = \Delta\hat{F}(t_i)[1+\lambda h(t_i)]\frac{\theta}{C^*(\lambda)} \qquad \text{with} \qquad C^*(\lambda) = \sum_{i=1, t_i<T} \Delta\hat{F}(t_i)[1+\lambda h(t_i)] \ .$$

Since $W_i = \Delta\hat{F}(t_i)$ is the maximum, the derivative of the log likelihood with respect to λ must be zero for any direction $h(t)$. This leads to the equation (A.*)(see Appendix A) In particular, the choice $h(t) = I_{[t\le u]}$ for $-\infty < u < \infty$ will give us the modified self-consistent equation under the constraint (2.1).

We shall use the convention

$$1-F(t_i) = \sum_{j:t_j>t_i} \Delta F(t_j) \ ; \qquad F(t_i) = \sum_{j:t_j\le t_i} \Delta F(t_j) \ .$$

(a) The constrained self-consistent equation for $\hat{F}(u)$ when $u \le T$ is:

$$\begin{aligned} \hat{F}(u) \ &= \ \frac{\theta}{n_1}\Bigg\{ \sum_{\delta_i=1, i\le n_1} I[t_i\le u] + \sum_{\delta_i=0, i\le n_1} \frac{\frac{1-\theta}{\theta}\hat{F}(u)}{1-F(t_i)} \\ &\quad + \sum_{\delta_i=0, i\le n_1} \frac{\hat{F}(u)-\hat{F}(t_i)}{1-F(t_i)} I[t_i\le u] \ + \sum_{\delta_i=2, i\le n_1} \frac{\sum_{t_j<t_i}\Delta\hat{F}(t_j) I[t_j\le u]}{\hat{F}(t_i-)} \Bigg\} \ . \end{aligned} \tag{2.5}$$

where all $t_i \le T$.

The last term in (2.5) can also be simplified to

$$\sum_{\delta_i=2, i\le n_1} \frac{\hat{F}(\min(u, t_i-))}{\hat{F}(t_i-)} \ .$$

(b) Similarly, the constrained self-consistent equation for $u > T$ is:

$$\begin{aligned} \hat{F}(u) \ &= \ \theta + \frac{1-\theta}{n_2}\Bigg\{ \sum_{\delta_i=1, i\le n_2} I[s_i\le u] + \sum_{\delta_i=0, i\le n_2} \frac{\hat{F}(u)-\hat{F}(s_i)}{1-\hat{F}(s_i)} I[s_i\le u] \\ &\quad + \sum_{\delta_i=2, i\le n_2} \frac{\frac{\theta}{1-\theta}[\hat{F}(u)-\theta]}{\hat{F}(s_i-)} + \sum_{\delta_i=2, i\le n_2} \frac{\sum_{s_j<s_i}\Delta\hat{F}(s_j) I[s_j\le u]}{\hat{F}(s_i-)} \Bigg\} \ . \end{aligned} \tag{2.6}$$

where all $s_i > T$.

Again the last term in (2.6) can be simplified to

$$\sum_{\delta_i=2, i\leq n_2} \frac{\hat{F}(\min(u, s_i-)) - \theta}{\hat{F}(s_i-)} .$$

For the derivation of above self-consistent equations, please see Appendix A. We summerize the result in the following theorem.

Theorem 2.1: *The NPMLE of $F(\cdot)$ under hypothesis (2.1) based on data (1.1) must satisfy the modified self-consistent equations (2.5) and (2.6).* $\diamondsuit$

Using the above modified self-consistent equations, we can iteratively compute the estimator of distribution F under the constrain equation by combining the estimator before and after time T.

Theorem 2.2: *The estimator obtained by the above modified self-consistent equations is at least a local maximum of the likelihood function under* (2.1).

Proof: See Appendix B.

3. Modified Self-Consistent Equations Under Many Constraints

In this section, we extend the modified self-consistent algorithm to handle hypotheses (constraints) at several times:

$$H_0 : F(T_1) = \theta_1, F(T_2) = \theta_2, \ldots, F(T_k) = \theta_k$$

$$H_1 : F(T_j) \neq \theta_j \qquad \textit{for at least one } j, \quad 1 \leq j \leq k.$$

For simplicity, we only give the proof for k=2. Following the similar steps as in the simple hypothesis, we can decompose the log-likelihood function into three parts: before time T_1, between time T_1 and T_2 and after time T_2.

$$\begin{aligned}
\log L(F) &= \log L_1(F) + \log L_2(F) + \log L_3(F) \\
&= \sum_{\delta_i=1, i\leq n_1} \log(w_i) + \sum_{\delta_i=0, i\leq n_1} \log[1-F(t_i)] + \sum_{\delta_i=2, i\leq n_1} \log[F(t_i-)] \\
&\quad + \sum_{\delta_i=1, i\leq n_2} \log(z_i) + \sum_{\delta_i=0, i\leq n_2} \log[1-F(x_i)] + \sum_{\delta_i=2, i\leq n_2} \log[F(x_i-)] \\
&\quad + \sum_{\delta_i=1, i\leq n_3} \log(v_i) + \sum_{\delta_i=0, i\leq n_3} \log[1-F(s_i)] + \sum_{\delta_i=2, i\leq n_3} \log[F(s_i-)]
\end{aligned}$$

where w_i, z_i and v_i are defined as jumps on ith observation before time T_1, between time T_1 and T_2 and after time T_2 respectively; t_i, x_i and s_i are observations before time T_1, between time T_1 and T_2 and after time T_2 respectively. There are n_1 observations before time T_1, n_2 observations between time T_1 and T_2, and n_3 observations after time T_2. The total number of observations is $n_1+n_2+n_3 = n$. Note that the constraints are $\sum_{i=1}^{n_1} w_i = \theta_1$, $\sum_{i=1}^{n_2} z_i = \theta_2 - \theta_1$, and $\sum_{i=1}^{n_3} v_i = 1 - \theta_2$.

We define the jumps before time T_1 and after time T_2 similarly as in section 2. The jumps between time T_1 and T_2 can be defined as

$$z_i = z_i(\lambda) = \frac{\Delta \hat{F}(x_i)}{1+\lambda h(x_i)} \frac{\theta_2 - \theta_1}{B(\lambda)} \qquad \text{with} \qquad B(\lambda) = \sum_{i=1}^{n_2} \frac{\Delta \hat{F}(x_i)}{1+\lambda h(x_i)} \ ,$$

and $B(0) = \theta_2 - \theta_1$.

$\log L_1(F)$ and $\log L_3(F)$ here are the same as $\log L_1(F)$ and $\log L_2(F)$ in Section 2. We shall focus on $\log L_2(F)$, the log-likelihood function between time T_1 and T_2. The self-consistent equation for $T_1 \le u \le T_2$ is:

$$\begin{aligned} \hat{F}(u) \quad = \quad & \theta_1 + \frac{\theta_2 - \theta_1}{n_2} \left\{ \sum_{\delta_i = 1, i \le n_2} I(x_i \le u) \right. \\ & + \sum_{\delta_i=0, i\le n_2} \frac{\frac{1-\theta_2}{\theta_2-\theta_1}[\hat{F}(u) - \theta_1]}{1-\hat{F}(x_i)} + \sum_{\delta_i=0, i \le n_2} \frac{[\hat{F}(u) - \hat{F}(x_i)] I[x_i \le u]}{1-\hat{F}(x_i)} \\ & \left. + \sum_{\delta_i = 2, i \le n_2} \frac{[\hat{F}(\min(x_i, u)) - \theta_1]}{\hat{F}(x_i -)} + \sum_{\delta_i=2, i\le n_2} \frac{\frac{\theta_1}{\theta_2 - \theta_1}[\hat{F}(u) - \theta_1]}{\hat{F}(x_i-)} \right\} \end{aligned} \tag{3.1}$$

where $T_1 \le x_i \le T_2$ (For the details, see Appendix C).

4. Empirical Likelihood Ratio Tests And Confidence Intervals

Let $\tilde{F}$ be the NPMLE of F which maximizes log likelihood, log $L(F)$ defined in (2.2), over all distributions and $\hat{F}$ denote the NPMLE of F under H_0, which maximizes the log likelihood only among distributions that satisfy H_0. We define the empirical likelihood ratio function as:

$$R(H_0) = \frac{L(\hat{F})}{L(\tilde{F})} \ .$$

Our likelihood ratio test statistic is:

$$\begin{aligned} -2\log R(H_0) \quad &= \quad -2 \log \frac{\max_{H_0} L(F)}{\max_{H_0+H_1} L(F)} \\ &= \quad 2 \left[\log(\max_{H_0+H_1} L(F)) - \log(\max_{H_0} L(F)) \right] \\ &= \quad 2 \left[\log(L(\tilde{F})) - \log(L(\hat{F})) \right] \ . \end{aligned}$$

The method described in section 2 will enable us to compute the constrained NPMLE $\hat{F}$ which maximizes the log likelihood under $H_0 : F(T) = \theta$. Using the usual self-consistent algorithm, we can compute NPMLE $\tilde{F}$ without any constraint. So once we have these estimates, we can easily compute the empirical likelihood ratio. We use chi-square theory to carry out hypothesis testing and construct confidence intervals. For theory about empirical likelihood ratio, see Owen (1987) and Murphy and Van der Varrt (1995).

If the observed $-2\log R(H_0)$ is greater than $\chi^2_{1,\alpha}$ (the $100(1-\alpha)$th percentile of χ^2 with 1 degree freedom), we reject H_0 at α significance level. To construct the confidence interval for $F(T)$, we can test different hypotheses with fixed T and various θ's and form the $1-\alpha$ confidence interval for $F(T)$ as

$$\left\{\theta : -2\log R(H_0 : F(T) = \theta) \leq \chi^2_{1,\alpha}\right\} \ . \tag{4.1}$$

We can also the construct confidence interval for percentile $F^{-1}(\theta)$ as follows. Test many hypotheses with fixed θ value and various T values, and form the confidence interval as:

$$\left\{T : -2\log R(H_0 : F(T) = \theta) \leq \chi^2_{1,\alpha}\right\} \ . \tag{4.2}$$

In particular, a confidence interval for the median can be obtained with $\theta = 1/2$.

5. Influence Function of NPMLE and Its Estimation

Influence function (or influence curve) is a general technique to obtain the variance of a random process (and more). In the analysis of $\tilde{F}(\cdot)$ with doubly censored samples, there are three influence functions corresponding to right, left and non-censored observations. Chang (1990) computed those asymptotic influence functions for the process $\sqrt{n}(\hat{F}_n(t) - F(t))$, he obtained the following:

$$\begin{aligned}\sqrt{n}(\hat{F}_n(t) - F(t)) \quad &= \quad \int_0^T IC_1(t,s)dq_1^{(n)}(s) + \int_0^T IC_0(t,s)dq_0^{(n)}(s) \\ &\quad + \int_0^T IC_2(t,s)dq_2^{(n)}(s) + o_p^{(n)}(1) \ ,\end{aligned}$$

where

$$q_j^{(n)}(t) = \sqrt{n}\left[\left(\frac{1}{n}\sum_i I_{[z_i \leq t, \delta_i = j]}\right) - E\left(\frac{1}{n}\sum_i I_{[z_i \leq t, \delta_i = j]}\right)\right] \qquad\qquad j = 0, 1, 2 \ .$$

¿From the above, we could try to estimate the three asymptotic influence functions, $I\hat{C}_j(t,s)$ and then estimate the variance of $\sqrt{n}(\hat{F}(t) - F(t))$ by

$$\begin{aligned}&\int_0^T I\hat{C}_1^2(t,s)dq_1^{(n)}(s) + \int_0^T I\hat{C}_0^2(t,s)dq_0^{(n)}(s) + \int_0^T I\hat{C}_2^2(t,s)dq_2^{(n)}(s) \\ &- \left(\int_0^T I\hat{C}_1(t,s)dq_1^{(n)}(s) + \int_0^T I\hat{C}_0(t,s)dq_0^{(n)}(s) + \int_0^T I\hat{C}_2(t,s)dq_2^{(n)}(s)\right)^2 \ .\end{aligned}$$

However, those influence functions are only defined via Fredholm integral equations that involve unknown distribution.

We plug-in the (self consistent) estimate of those distribution functions, discretize the Fredholm integral equations into matrix equations and solve for $\hat{IC}$. For details see Chang (1990) and *Numerical Recipes in C*.

In this approach we need to solve a matrix equation that resulted from discretizing corresponding integral equations for influence functions. If we discretizing at a few points, then the estimate would not be very good. If we use a lot of discretizing points, computation is slow. Therefore for large sample sizes, this approach is very computationally expensive (to find the inverse of a large matrix). In our experience, when censored observations (both right and left) total exceed 500, it becomes slow in our implementation, since we discretizing at observed censoring times.

Nevertheless, this approach is worth exploring and it is interesting to compare to the approach of empirical likelihood described in Section 4.

6. Applications, Simulations and Examples

6.1 Applications: Paired Comparison

In section 1, we gave an example indicating that doubly censored data may result from paired comparison experiments. We shall specify a model for this case and illustrate the use of the testing procedures to test the hypothesis for drug effect when we do not want to make parametric assumptions.

A reasonable model for the paired experiment is as follows: for ith subject (or pair) ($i = 1, 2, \ldots, n$) we observe Y_{1i} and Y_{2i} where

$$Y_{1i} = \tau_d + S_i + \epsilon_{1i} \ ,$$

$$Y_{2i} = \tau_p + S_i + \epsilon_{2i} \ ,$$

where $\tau_d(\tau_p)$ is the main effect for drug (placebo), S_i is the subject effect, ϵ_{ki} is the random error. The difference of Y_{1i} and Y_{2i} is:

$$D_i = (\tau_d - \tau_p) + (\epsilon_{1i} - \epsilon_{2i}),$$

which is free from S_i, a fact that lead many test procedures to be based on the D_i's. If we assume ϵ_{1i} and ϵ_{1i} are exchangable, then the median of D_i is $\tau_d - \tau_p$. Thus a test of $H_0 : \tau_d - \tau_p = 0$ can be carried out by testing if the median of D_i is zero. Double censoring on the D_i requires our test as described in section 4 and 5.

In the case where ϵ_{ki} are i.i.d. with a distribution of $\exp(\lambda) - 1/\lambda$ (mean zero exponential), D_i has double exponential distribution with location parameter $\tau_d - \tau_p$. Since the sample median is

MLE of location parameter for double exponential distribution, we can expect the test to perform well in this case.

We carried out some simulations for this procedure summerized in Table 6.1.

Table 6.1: The Percentage Of Rejecting $H_0 : \tau_d = \tau_p$ *At* $\alpha = 0.05$

Difference	Size	No Censored	Light Censored	Medium Censored	Heavy Censored
0.0	$n = 100$	0.053	0.054	0.058	0.044
	$n = 25$	0.077	0.048	0.053	0.044
0.3	$n = 100$	0.968	0.907	0.819	0.633
	$n = 25$	0.487	0.374	0.284	0.188
0.2	$n = 100$	0.767	0.635	0.487	0.373
	$n = 25$	0.276	0.205	0.146	0.120

When the null hypothesis is true, the percentages of rejecting H_0 are very close to the nominal level 0.05 for small and large samples. When there is difference between drug and placebo, the rejecting percentages decrease with the increases of censoring observations.

6.2 Simulation: Hypothesis Testing

In our first simulation, we took normally distributed samples of size 100 and size 25 respectively for each run and each entry in Table 6.3 was based on 5,000 runs.

Table 6.2: Generating Normally Distributed Samples $i = 1, 2, \cdots, n.$

	X_i	Z_i	Y_i
light-censored ($10\% - 20\%$ censored)	$N(\mu = 10, \sigma = 2)$	$N(\mu = 6, \sigma = 2)$	$\exp(1) + Z_i + 8$
medium-censored ($20\% - 40\%$ censored)	$N(\mu = 10, \sigma = 2)$	$N(\mu = 7, \sigma = 2)$	$\exp(1) + Z_i + 5$
heavy-censored ($40\% - 60\%$ censored)	$N(\mu = 10, \sigma = 2)$	$N(\mu = 9, \sigma = 2)$	$\exp(1) + Z_i + 2$

Table 6.3 illustrates the probabilities of rejecting $H_0 : F(T) = \theta$ at nominal significance level $\alpha = 0.05$ (or 0.10). The percentages were computed as the number of -2log-likelihood ratios greater than critical value $\chi^2_{1,0.05} = 3.84$ (or $\chi^2_{1,0.1} = 2.71$) divided by 5000. From Table 6.3 we can see the probabilities of rejecting H_0 are pretty close to the nominal level $\alpha = 0.05$ (or 0.10).

We took exponentially distributed samples of size 100 and size 25 respectively for our second simulation. This simulation was also based on 5000 samples.

Table 6.3: The Percentage Of Rejecting $H_0 : F(T) = \theta$ At $\alpha = 0.05$ And 0.10

T	θ	Sample Size	Light Censored	Medium Censored	Heavy Censored
10	0.5	$n = 100$	5.38% / 10.36%	5.02% / 10.04%	4.66% / 12.26%
		$n = 25$	4.62% / 10.76%	4.82% / 10.12%	4.92% / 10.14%
8	0.1586553	$n = 100$	5.04% / 10.2%	5.02% / 10.4%	5.70% / 10.44%
		$n = 25$	5.12% / 12.8%	6.08% / 11.52%	5.58% / 10.72%

Table 6.4: Generating Exponentially Distributed Samples $i = 1, 2, \cdots, n$.

	X_i	Z_i	Y_i
light-censored (10% – 20% censored)	exp(2)	exp(15)	$\exp(1) + Z_i + 1$
medium-censored (20% – 40% censored)	exp(2)	exp(8)	$\exp(1) + Z_i + 0.3$
heavy-censored (40% – 60% censored)	exp(2)	exp(5)	$\exp(1) + Z_i$

Again Table 6.5 shows that the probabilities of rejecting H_0 are around the nominal level $\alpha = 0.05$ (or 0.10).

Table 6.5: The Percentage Of Rejecting $H_0 : F(T) = \theta$ At $\alpha = 0.05$ And 0.10

T	θ	Sample Size	Light Censored	Medium Censored	Heavy Censored
0.5	0.6321206	$n = 100$	4.18% / 9.06%	4.50% / 9.58%	4.94% / 9.98%
		$n = 25$	4.58% / 9.68%	4.70% / 9.74%	5.40% / 10.2%
0.3	0.4511884	$n = 100$	4.72% / 8.64%	4.56% / 9.34%	4.48% / 9.38%
		$n = 25$	4.52% / 10.18%	4.82% / 10.18%	5.46% / 10.68%

Figure 6.1 and 6.2 are Q-Q plots of -2log-likelihood ratios for the above two simulations verse the $\chi^2_{(1)}$ percentiles. At the point 3.84 (or 2.71), if the -2log-likelihood ratio line is above the dashed line (45° line), the rejecting probability is greater than 5% (or 10%). Otherwise, the rejecting probability is less than 5% (or 10%).

Remark 6.1: The Q-Q plots for exponentially distributed light-censored simulations (Figure 6.1(d) and 6.2 (d)) are somehow more discrete than the others. We did 5000 uncensored simulations with exponential distribution, the Q-Q plot is similar to Figure 6.1(d) and 6.2 (d). Since the constraint is $F(T) = \theta$, $\hat{F}$ and $\tilde{F}$ only depend on the number of observations(n_1) before and (n_2) after time T. If there are the same n_1 and n_2 in two uncensored samples, -2log-likelihood ratio will be the same. That is why uncensored and light-censored plots look more discrete. The censoring somehow smoothed the plot, and made the chi-square approximation better.

6.3 Example: Confidence intervals – A case study

We compare the confidence intervals obtained via empirical likelihood ratio as in section 4 to confidence intervals obtained by directly estimating the asymptotic variance of $\hat{F}(T)$ and then form the Wald $(1-\alpha)$ confidence interval:

$$\hat{F}(T) \pm Z_{(1-\alpha/2)}\sqrt{\hat{Var}[\hat{F}(T)]}\ . \tag{6.1}$$

where $Z_{(1-\alpha/2)}$ is the $100(1-\alpha/2)$th percentile of a standard normal distribution. Better confidence intervals may be obtained on a transformed scale. We may use the log-log transformation to obtain: $(\hat{S}(\cdot) = 1 - \hat{F}(\cdot))$,

$$[\hat{S}(T)^b, \hat{S}(T)^{1/b}]\ , \qquad \text{where} \qquad b = \exp\left[\frac{Z_{(1-\alpha/2)}\sqrt{\hat{Var}[\hat{S}(T)]}}{\hat{S}(T)|\log \hat{S}(T)|}\right]\ . \tag{6.2}$$

or the logit transformation to obtain:

$$[\frac{e^a}{1+e^a}, \frac{e^b}{1+e^b}]\ , \tag{6.3}$$

where

$$a = \log\frac{\hat{F}(T)}{1-\hat{F}(T)} - \frac{Z_{(1-\alpha/2)}}{\hat{F}(T)(1-\hat{F}(T))}\sqrt{\hat{Var}[\hat{F}(T)]}$$

and

$$b = \log\frac{\hat{F}(T)}{1-\hat{F}(T)} + \frac{Z_{(1-\alpha/2)}}{\hat{F}(T)(1-\hat{F}(T))}\sqrt{\hat{Var}[\hat{F}(T)]}\ .$$

It is not easy to obtain the Wald type confidence interval for the median (or other quantiles). But we can invert the test of $F(T) = 0.5$ with estimated variance. Therefore the confidence set for median may be obtained as the set of points T that satisfy the following condition:

$$\left\{T : -Z_{(1-\alpha/2)} \le \frac{\hat{F}(T) - 0.5}{\sqrt{\hat{Var}[\hat{F}(T)]}} \le Z_{(1-\alpha/2)}\right\}\ . \tag{6.4}$$

To construct confidence intervals, we will use the following doubly censored data as our example.

Turnbull and Weiss (1978) reported part of a study conducted at Stanford-Palo Alto Peer Counseling Program (see Hamburg et al. (1975) for details of the study). In this study, 191 California high school boys were asked "When did you first use marijuana?" The answers are either the exact age (uncensored observations), or "I never used it" which are right-censored observations at the boys' current ages, or "I have used it but can not recall just when the first time was" which are left-censored observations. Table 6.6 shows the results of this study.

The estimated median age of high school boys who use marijuana is 14 years old.

Suppose we are interested to obtain 95% confidence interval for the median age of first time marijuana use. Table 6.7 shows the 95% confidence intervals for the median age with (4.2) and (6.4). Table 6.8 shows four different kinds of 95% confidence intervals for θ at the median age with (4.1), (6.1), (6.2) and (6.3).

Table 6.7: 95% Confidence Intervals For Median Age(14)

method		Confidence Interval
$T : -2 \log R(H_0 : F(T) = 0.5) \leq 3.84$	(4.2)	$(11.00000, 16.99991)$
$T : -1.96 \leq \frac{\hat{F}(t)-0.5}{\sqrt{V\hat{a}r[\hat{F}(t)]}} \leq 1.96$	(6.4)	$(10, 19)$

Table 6.8: 95% Confidence Intervals For $F(14)$

method		Confidence Interval
$\theta : -2 \log R(H_0 : F(14) = \theta) \leq 3.84$	(4.1)	$(0.2798736, 0.7195857)$
$[1 - \hat{S}(14)^b, 1 - \hat{S}(14)^{1/b}]$	(6.2)	$(0.0378062, 0.9999983)$
$[\frac{e^a}{1+e^a}, \frac{e^b}{1+e^b}]$	(6.3)	$(0.01719094, 0.9842504)$
$\hat{F}(14) \pm 1.96\sqrt{V\hat{a}r[\hat{F}(14)]}$	(6.1)	$(-0.533258, 1.511003)$

From Table 6.8 we can see the 95% confidence interval obtained by empirical likelihood ratio (4.1) is narrower than any other three intervals with (6.1), (6.2) and (6.3). The two transformed intervals obtain by (6.2) and (6.3) are wider than (4.1) but close to each other. We do not know which transformation is better. We believe the empirical likelihood ratio approach is better because we do not need to know what is the best transformation, and simulation in the previous section show the chi square approximation is pretty accurate. The Wald confidence interval (6.1) includes the numbers less than 0 and greater than 1. These are not reasonable confidence limits for a distribution since a distribution must be between 0 and 1. Therefore, in this example we conclude that the empirical likelihood ratio approach works better than Wald's approach.

6.4 Software Used

The simulation was carried out using Splus 3.4 for Unix on the HP workstations. The Splus functions that computes the constrained NPMLE with doubly censored data will be uploaded to Statlib in the near future, as an update to the function `d009newr` that was there since June, 1995.

References

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

Chang, M. N. and Yang, G. L. (1987), Strong Consistency of a Nonparametric Estimator of the Survival Function with Doubly Censored Data. *Ann. Statist.* 15, 1536-1547.

Chang, M. N. (1990). Weak Convergence in Doubly Censored Data. *Ann. Statist.* 18, 390-405

Dempseter, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum Likelihood From Incomplete Data via The EM Algorithm with Discussion). *J. Roy. Statist. Soc.* Ser. B 39 1-38.

Efron, B. (1967). The Two Sample Problem with Censored Data. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* 4, 831-883.

Gill, R. (1989), Non-and Semi-parametric Maximum Likelihood Estimator and the Von Mises Method (I) *Scand. J. Statist.* **16**, 97-128.

Hamberg, B.A., Kraemer, H.C., and Jahnke, W. (1975). A Hierarchy of Drug Use in Adolescence Behavioral and Attitudinal Correlates of Substantial Drug Use. *American Journal of Psychiatry* 132 (1975): 1155-1163.

Kaplan, E. and Meier, P. (1958), Non-parametric Estimator From Incomplete Observations *JASA* 53, 457–481.

Murphy, S. and Van der Varrt, (1997). Semiparametric Likelihood Ratio Inference. *Ann. Statist.* 25, 1471-1509.

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

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1993) *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge Univ. Press.

Turnbull, B. (1974), Nonparametric Estimation of a Survivorship Function with Doubly Censored Data. *JASA* 169-173.

Turnbull, B. (1976), The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *JRSS* B 290-295.

Turnbull, B. and Weiss (1978). A Likelihood Ratio Statistic for Testing Goodness of Fit with Randomly Censored Data. *Biometrics* 34 (1978): 367-375.

Wei-Yann Tsai and John Crowley (1985), A Large Sample Study of Generalized Maximum Likelihood Estimator From Incomplete Data Via Self-Consistency. *Ann. Statist.* 13, No. 4, 1317-1334.

Zhan, Y. and Wellner, J. (1999). A hybrid algorithm for computation of the NPMLE from censored data. *JASA*, **92**, 945-959.

Zhou, M.(1995) http://lib.stat.cmu.edu/s/d009newr.

Appendix A

Derivation of the Self-consistent Equation Before Time T

(a) The log likelihood function before time T:

$$\begin{aligned}\log L_1 &= \sum_{\delta_i=1, i\le n_1} \log(w_i) + \sum_{\delta_i=0, i\le n_1} \log[1-F(t_i)] + \sum_{\delta_i=2, i\le n_1} \log[F(t_i-)] \\ &= \sum_{\delta_i=1, i\le n_1} \log(w_i) + \sum_{\delta_i=0, i\le n_1} \log\left[(1-\theta) + \sum_{t_j>t_i, j\le n_1} w_j\right] + \sum_{\delta_i=2, i\le n_1} \log\left(\sum_{t_j<t_i, j\le n_1} w_j\right) .\end{aligned}$$

To facilitate derivative we substitute $w_i(\lambda)$ as in (2.3)

$$\begin{aligned}\log L_1 &= \sum_{\delta_i=1, i\le n_1} \log \frac{\Delta\hat F(t_i)}{1+\lambda h(t_i)}\frac{\theta}{C(\lambda)} + \sum_{\delta_i=0, i\le n_1} \log\left[(1-\theta) + \sum_{t_j>t_i, j\le n_1} \frac{\Delta\hat F(t_j)}{1+\lambda h(t_j)}\frac{\theta}{C(\lambda)}\right] \\ &\quad + \sum_{\delta_i=2, i\le n_1} \log \sum_{t_j<t_i, j\le n_1} \frac{\Delta\hat F(t_j)}{1+\lambda h(t_j)}\frac{\theta}{C(\lambda)}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\delta_i=1, i\le n_1} \log \frac{\theta}{C(\lambda)} + \sum_{\delta_i=1, i\le n_1} \log \frac{\Delta\hat{F}(t_i)}{1+\lambda h(t_i)} + \sum_{\delta_i=0, i\le n_1} \log \frac{\theta}{C(\lambda)} \\
&\quad + \sum_{\delta_i=0, i\le n_1} \log \left[\frac{1-\theta}{\theta} C(\lambda) + \sum_{t_j>t_i, j\le n_1} \frac{\Delta\hat{F}(t_j)}{1+\lambda h(t_j)} \right] \\
&\quad + \sum_{\delta_i=2, i\le n_1} \log \frac{\theta}{C(\lambda)} + \sum_{\delta_i=2, i\le n_1} \log \sum_{t_j<t_i, j\le n_1} \frac{\Delta\hat{F}(t_j)}{1+\lambda h(t_j)} \\
&= n_1 \log \frac{\theta}{C(\lambda)} + \sum_{\delta_i=1, i\le n_1} \log \frac{\Delta\hat{F}(t_i)}{1+\lambda h(t_i)} + \sum_{\delta_i=0, i\le n_1} \log \left[\frac{(1-\theta)C(\lambda)}{\theta} + \sum_{t_j>t_i, j\le n_1} \frac{\Delta\hat{F}(t_j)}{1+\lambda h(t_j)} \right] \\
&\quad + \sum_{\delta_i=2, i\le n_1} \log \sum_{t_j<t_i, j\le n_1} \frac{\Delta\hat{F}(t_j)}{1+\lambda h(t_j)} \,.
\end{aligned}
$$

Now we are ready to take derivative with respect to λ,

$$
\begin{aligned}
\frac{\partial \log L_1}{\partial \lambda} &= -n_1 \frac{C'(\lambda)}{C(\lambda)} - \sum_{\delta_i=1, i\le n_1} \frac{h(t_i)}{1+\lambda h(t_i)} \\
&\quad + \sum_{\delta_i=0, i\le n_1} \frac{\frac{1-\theta}{\theta} C'(\lambda) - \sum_{t_j>t_i} \frac{\Delta\hat{F}(t_j)h(t_j)}{[1+\lambda h(t_j)]^2}}{\frac{1-\theta}{\theta} C(\lambda) + \sum_{t_j>t_i} \frac{\Delta\hat{F}(t_j)}{1+\lambda h(t_j)}} \\
&\quad - \sum_{\delta_i=2, i\le n_1} \frac{\sum_{t_j<t_i} \frac{\Delta\hat{F}(t_j)h(t_i)}{[1+\lambda h(t_j)]^2}}{\sum_{t_j<t_i} \frac{\Delta\hat{F}(t_j)}{1+\lambda h(t_j)}} \,.
\end{aligned}
$$

If we set $\lambda = 0$, the above can be simplified to

$$
\begin{aligned}
\frac{\partial \log L_1}{\partial \lambda}\Big|_{\lambda=0} &= \frac{n_1}{\theta} \sum_{i=1}^{n_1} \Delta\hat{F}(t_i)h(t_i) - \sum_{\delta_i=1, i\le n_1} h(t_i) - \sum_{\delta_i=0, i\le n_1} \frac{\frac{1-\theta}{\theta} \sum_{i=1}^{n_1} \Delta\hat{F}(t_i)h(t_i)}{(1-\theta) + \sum_{t_j>t_i} \Delta\hat{F}(t_j)} \\
&\quad - \sum_{\delta_i=0, i\le n_1} \frac{\sum_{t_j>t_i} \Delta\hat{F}(t_j)h(t_j)}{(1-\theta) + \sum_{t_j>t_i} \Delta\hat{F}(t_j)} - \sum_{\delta_i=2, i\le n_1} \frac{\sum_{t_j<t_i} \Delta\hat{F}(t_j)h(t_j)}{\sum_{t_j<t_i} \Delta\hat{F}(t_j)} \,.
\end{aligned}
$$

This derivative must be zero since $\hat{F}$ is the NPMLE.

Thus we get

$$
\begin{aligned}
\sum_{i=1}^{n_1} \Delta\hat{F}(t_i)h(t_i) &= \frac{\theta}{n_1} \left\{ \sum_{\delta_i=1, i\le n_1} h(t_i) + \sum_{\delta_i=0, i\le n_1} \frac{\frac{1-\theta}{\theta} \sum_{i=1}^{n_1} \Delta\hat{F}(t_i)h(t_i)}{1-F(t_i)} \right. \\
&\quad \left. + \sum_{\delta_i=0, i\le n_1} \frac{\sum_{t_j>t_i} \Delta\hat{F}(t_j)h(t_j)}{1-F(t_i)} + \sum_{\delta_i=2, i\le n_1} \frac{\sum_{t_j<t_i} \Delta\hat{F}(t_j)h(t_j)}{\hat{F}(t_i-)} \right\} \quad (A.*)
\end{aligned}
$$

If we take $h(t_i) = I[t_i \le u]$, for $u \le T (A.*)$ becomes equation (2.5).

Derivation of the Self-consistent Equation After Time T

Similarly we can obtain the constrained self-consistent equation for $\hat{F}$ when $u > T$.

(b) The log likelihood function after time T is

$$\begin{aligned}
\log L_2 &= \sum_{\delta_i=1,i\le n_2} \log(v_i) + \sum_{\delta_i=0,i\le n_2} \log[1-F(s_i)] + \sum_{\delta_i=2,i\le n_2} \log[F(s_i-)] \\
&= \sum_{\delta_i=1,i\le n_2} \log(v_i) + \sum_{\delta_i=0,i\le n_2} \log(\sum_{s_j>s_i;j\le n_2} v_j) + \sum_{\delta_i=2,i\le n_2} \log(\theta + \sum_{s_j<s_i;j\le n_2} v_j)
\end{aligned}$$

Again substituting $v_i(\lambda)$ as in (2.4),

$$\begin{aligned}
\log L_2 &= \sum_{\delta_i=1,i\le n_2} \log \frac{\Delta\hat{F}(s_i)}{1+\lambda h(s_i)} \frac{1-\theta}{D(\lambda)} + \sum_{\delta_i=0,i\le n_2} \log \sum_{s_j>s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)} \frac{1-\theta}{D(\lambda)} \\
&\quad + \sum_{\delta_i=2,i\le n_2} \log \left[\theta + \sum_{s_j<s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)} \frac{1-\theta}{D(\lambda)} \right] \\
&= \sum_{\delta_i=1,i\le n_2} \log \frac{1-\theta}{D(\lambda)} + \sum_{\delta_i=1,i\le n_2} \log \frac{\Delta\hat{F}(s_i)}{1+\lambda h(s_i)} + \sum_{\delta_i=0,i\le n_2} \log \frac{1-\theta}{D(\lambda)} \\
&\quad + \sum_{\delta_i=0,i\le n_2} \log \sum_{s_j>s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)} + \sum_{\delta_i=2,i\le n_2} \log \frac{1-\theta}{D(\lambda)} \\
&\quad + \sum_{\delta_i=2,i\le n_2} \log \left[\frac{\theta}{1-\theta} D(\lambda) + \sum_{s_j<s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)} \right] \\
&= n_2 \log \frac{1-\theta}{D(\lambda)} + \sum_{\delta_i=1,i\le n_2} \log \frac{\Delta\hat{F}(s_i)}{1+\lambda h(s_i)} + \sum_{\delta_i=0,i\le n_2} \log \sum_{s_j>s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)} \\
&\quad + \sum_{\delta_i=2,i\le n_2} \log \left[\frac{\theta}{1-\theta} D(\lambda) + \sum_{s_j<s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)} \right] .
\end{aligned}$$

Taking derivative with respect to λ, it follows:

$$\begin{aligned}
\frac{\partial \log L_2}{\partial\lambda} &= -n_2 \frac{D'(\lambda)}{D(\lambda)} - \sum_{\delta_i=1,i\le n_2} \frac{h(s_i)}{[1+\lambda h(s_i)]^2} \\
&\quad - \sum_{\delta_i=0,i\le n_2} \frac{\sum_{s_j>s_i} \frac{\Delta\hat{F}(s_j)h(s_j)}{[1+\lambda h(s_j)]^2}}{\sum_{s_j>s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)}} \\
&\quad + \sum_{\delta_i=2,i\le n_2} \frac{\frac{\theta}{1-\theta} D'(\lambda) - \sum_{s_j<s_i} \frac{\Delta\hat{F}(s_j)h(s_i)}{[1+\lambda h(s_j)]^2}}{\frac{\theta}{1-\theta} D(\lambda) + \sum_{s_j<s_i} \frac{\Delta\hat{F}(s_j)}{1+\lambda h(s_j)}} .
\end{aligned}$$

If we set $\lambda = 0$ and the derivative must be zero, thus

$$\begin{aligned}
\sum_{i=1}^{n_2} \Delta\hat{F}(s_i)h(s_i) &= \frac{1-\theta}{n_2} \left\{ \sum_{\delta_i=1,i\le n_2} h(s_i) + \sum_{\delta_i=0,i\le n_2} \frac{\sum_{s_j>s_i} \Delta\hat{F}(s_j)h(s_j)}{\sum_{s_j>s_i} \Delta\hat{F}(s_j)} \right. \\
&\quad \left. + \sum_{\delta_i=2,i\le n_2} \frac{\frac{\theta}{1-\theta} \sum_{i=1}^{n_2} \Delta\hat{F}(s_j)h(s_j)}{\theta + \sum_{s_j<s_i} \Delta\hat{F}(s_j)} + \sum_{\delta_i=2,i\le n_2} \frac{\sum_{s_j<s_i} \Delta\hat{F}(s_j)h(s_j)}{\theta + \sum_{s_j<s_i} \Delta\hat{F}(s_j)} \right\} .
\end{aligned}$$

If we take $h(s_i) = I[s_i \le u]$ for $u > T$, the above equation becomes (2.6).

Appendix B

Proof of Theorem 2.2

Taking the second derivative of log likelihood function $\log L_1$ with respect to λ and setting $\lambda = 0$ and $h(t_i) = I_{[t_i \le u]}$, we can simplify the derivative as follows:

$$\begin{aligned}
\frac{\partial^2 \log L_1}{\partial \lambda^2}|_{\lambda=0} &= -\frac{2n_1}{\theta}\hat{F}(u) + \frac{n_1}{\theta^2}[\hat{F}(u)]^2 + \sum_{\delta_i=1, i\le n_1} I_{[t_i\le u]} \\
&+2\sum_{\delta_i=0,i\le n_1} \frac{\frac{1-\theta}{\theta}\hat{F}(u)}{1-\hat{F}(t_i)} + 2\sum_{\delta_i=0,i\le n_1}\frac{\hat{F}(u)-\hat{F}(t_i)}{1-\hat{F}(t_i)} I_{[t_i\le u]} \\
&-\sum_{\delta_i=0,i\le n_1} \frac{\left\{\frac{1-\theta}{\theta}\hat{F}(u) + [\hat{F}(u)-\hat{F}(t_i)]I_{[t_i\le u]}\right\}^2}{[1-\hat{F}(t_i)]^2} \\
&+2\sum_{\delta_i=2,i\le n_1}\frac{\hat{F}(\min(u,t_i-)}{\hat{F}(t_i-)} - \sum_{\delta_i=2,i\le n_1}\left\{\frac{\hat{F}(\min(u,t_i-)}{\hat{F}(t_i-)}\right\}^2 .
\end{aligned}$$

Substituting $\hat{F}(u)$ as in (2.5),

$$\begin{aligned}
\frac{\partial^2 \log L_1}{\partial \lambda^2}|_{\lambda=0} &= -\sum_{\delta_i=1,i\le n_1} I_{[t_i\le u]} + \frac{n_1}{\theta^2}[\hat{F}(u)]^2 \\
&-\sum_{\delta_i=0,i\le n_1} \frac{\left\{\frac{1-\theta}{\theta}\hat{F}(u) + [\hat{F}(u)-\hat{F}(t_i)]I_{[t_i\le u]}\right\}^2}{[1-\hat{F}(t_i)]^2} \\
&-\sum_{\delta_i=2,i\le n_1}\left[\frac{\hat{F}(\min(u,t_i-))}{\hat{F}(t_i-)}\right]^2 \\
&= -n_1\left\{\frac{1}{n_1}\left[\sum_{\delta_i=1,i\le n_1} I^2_{[t_i\le u]} + \sum_{\delta_i=0,i\le n_1}\frac{\left\{\frac{1-\theta}{\theta}\hat{F}(u) + [\hat{F}(u)-\hat{F}(t_i)]I_{[t_i\le u]}\right\}^2}{[1-\hat{F}(t_i)]^2}\right.\right. \\
&\left.\left. + \sum_{\delta_i=2,i\le n_1}\left(\frac{\hat{F}(\min(u,t_i-))}{\hat{F}(t_i-)}\right)^2\right] - \left[\frac{\hat{F}(u)}{\theta}\right]^2\right\} \\
&= -n_1\left\{\frac{1}{n_1}\sum_{i=1}^{n_1} a_i^2 - \left[\frac{1}{n_1}\sum_{i=1}^{n_1} a_i\right]^2\right\}
\end{aligned}$$

where

$$\begin{aligned}
\sum_{i=1}^{n_1} a_i^2 &= \sum_{\delta_i=1,i\le n_1} I^2_{[t_i\le u]} + \sum_{\delta_i=0,i\le n_1}\frac{\left\{\frac{1-\theta}{\theta}\hat{F}(u) + [\hat{F}(u)-\hat{F}(t_i)]I_{[t_i\le u]}\right\}^2}{[1-\hat{F}(t_i)]^2} \\
&+ \sum_{\delta_i=2,i\le n_1}\left[\frac{\hat{F}(\min(u,t_i-))}{\hat{F}(t_i-)}\right]^2
\end{aligned}$$

and

$$\sum_{i=1}^{n_1} a_i = \sum_{\delta_i=1,i\le n_1} I_{[t_i\le u]} + \sum_{\delta_i=0,i\le n_1}\frac{\frac{1-\theta}{\theta}\hat{F}(u) + [\hat{F}(u)-\hat{F}(t_i)]I_{[t_i\le u]}}{[1-\hat{F}(t_i)]}$$

$$+\sum_{\delta_i=2,i\le n_1}\frac{\hat{F}(\min(u,t_i-)}{\hat{F}(t_i-)}$$

By (2.5) and Cauchy-Schwartz Inequality, it is clear that the second derivative of $\log L_1$ is non-positive. The second derivative of $\log L_2$ is also non-positive by following the similar steps. $\diamondsuit$

Appendix C

Derivation of the self-consistent Equations Between Time T_1 and T_2

$$\begin{aligned}
\log L_2 &= \sum_{\delta_i=1,i\le n_2}\log(z_i)+\sum_{\delta_i=0,i\le n_2}\log[1-F(x_i)]+\sum_{\delta_i=2,i\le n_2}\log[F(x_i-)] \\
&= \sum_{\delta_i=1,i\le n_2}\log(z_i)+\sum_{\delta_i=0,i\le n_2}\log\left[1-F(T_2)+F(T_2)-F(x_i)\right] \\
&\quad+\sum_{\delta_i=2,i\le n_2}\log\left[F(x_i-)-F(T_1)+F(T_1)\right] \\
&= \sum_{\delta_i=1,i\le n_2}\log(z_i)+\sum_{\delta_i=0,i\le n_2}\log\left[(1-\theta_2)+\sum_{x_j>x_i,i\le n_2}z_j\right]+\sum_{\delta_i=2,i\le n_2}\log\left[\sum_{x_j<x_i,i\le n_2}z_j+\theta_1\right]
\end{aligned}$$

Plug $z_i(\lambda)$ into $\log L_2$ into above equation, we get

$$\begin{aligned}
\log L_2 &= \sum_{\delta_i=1,i\le n_2}\log\left[\frac{\Delta\hat{F}(x_i)}{1+\lambda h(x_i)}\frac{\theta_2-\theta_1}{B(\lambda)}\right]+\sum_{\delta_i=0,i\le n_2}\log\left[(1-\theta_2)+\sum_{x_j>x_i,j\le n_2}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)}\frac{\theta_2-\theta_1}{B(\lambda)}\right] \\
&\quad+\sum_{\delta_i=2,i\le n_2}\log\left[\sum_{x_j<x_i,j\le n_2}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)}\frac{\theta_2-\theta_1}{B(\lambda)}+\theta_1\right] \\
&= \sum_{\delta_i=1,i\le n_2}\log\frac{\theta_2-\theta_1}{B(\lambda)}+\sum_{\delta_i=1,i\le n_2}\log\frac{\Delta\hat{F}(x_i)}{1+\lambda h(x_i)} \\
&\quad+\sum_{\delta_i=0,i\le n_2}\log\left[\frac{\theta_2-\theta_1}{B(\lambda)}\frac{B(\lambda)}{\theta_2-\theta_1}(1-\theta_2)+\sum_{x_j>x_i,j\le n_2}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)}\frac{\theta_2-\theta_1}{B(\lambda)}\right] \\
&\quad+\sum_{\delta_i=2,i\le n_2}\log\left[\frac{\theta_2-\theta_1}{B(\lambda)}\sum_{x_j>x_i,j\le n_1}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)}+\frac{\theta_2-\theta_1}{B(\lambda)}\frac{\theta_1}{\theta_2-\theta_1}\right] \\
&= n_2\log\frac{\theta_2-\theta_1}{B(\lambda)}+\sum_{\delta_i=1,i\le n_2}\log\frac{\Delta\hat{F}(x_i)}{1+\lambda h(x_i)} \\
&\quad+\sum_{\delta_i=0,i\le n_2}\log\left[\frac{1-\theta_2}{\theta_2-\theta_1}B(\lambda)+\sum_{x_j>x_i,j\le n_2}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)}\right] \\
&\quad+\sum_{\delta_i=2,i\le n_2}\log\left[\sum_{x_j<x_i,j\le n_2}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)}+\frac{\theta_1}{\theta 2-\theta_1}B(\lambda)\right] .
\end{aligned}$$

Taking derivative with respective to λ,

$$\frac{\partial\log L_2}{\partial\lambda} = -n_2\frac{B^{'}(\lambda)}{B(\lambda)}-\sum_{\delta_i=1,i\le n_1}\frac{h(x_i)}{1+\lambda h(x_i)}$$

$$+\sum_{\delta_i=0, i\le n_1} \frac{\frac{1-\theta_2}{\theta_2-\theta_1}B^{'}(\lambda) - \sum_{x_j>x_i}\frac{\Delta\hat{F}(x_j)h(x_j)}{[1+\lambda h(x_j)]^2}}{\frac{1-\theta_2}{\theta_2-\theta_1}B(\lambda) + \sum_{x_j>x_i}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)}}$$

$$+\sum_{\delta_i=2, i\le n_2} \frac{\sum_{x_j<x_i, j\le n_2}\frac{-\Delta\hat{F}(x_j)h(x_i)}{[1+\lambda h(x_j)]^2} + \frac{\theta_1}{\theta_2-\theta_1}B^{'}(\lambda)}{\sum_{x_j<x_i, j\le n_2}\frac{\Delta\hat{F}(x_j)}{1+\lambda h(x_j)} + \frac{\theta_1}{\theta_2-\theta_1}B(\lambda)}\ .$$

If we set $\lambda = 0$, the above equation can be simplified to

$$\begin{aligned}\frac{\partial \log L_2}{\partial \lambda}|_{\lambda=0} &= \frac{n_2}{\theta_2-\theta_1}\sum_{i=1}^{n_2}\Delta\hat{F}(x_i)h(x_i) - \sum_{\delta_i=1, i\le n_2} h(x_i) \\ &- \sum_{\delta_i=0, i\le n_2}\frac{\frac{1-\theta_2}{\theta_2-\theta_1}\sum_{i=1}^{n_2}\Delta\hat{F}(x_i)h(x_i) + \sum_{x_j>x_i}\Delta\hat{F}(x_j)h(x_j)}{(1-\theta_2)+\sum_{x_j>x_i}\Delta\hat{F}(x_j)} \\ &- \sum_{\delta_i=2, i\le n_2}\frac{\sum_{x_j<x_i}\Delta\hat{F}(x_j)h(x_j) + \frac{\theta_1}{\theta_2-\theta_1}\sum_{i=1}^{n_2}\Delta\hat{F}(x_i)h(x_i)}{\theta_1+\sum_{x_j<x_i}\Delta\hat{F}(x_j)}\ .\end{aligned}$$

Since the derivative is equal to zero, we rearrange the terms and obtain

$$\begin{aligned}\sum_{i=1}^{n_2}\Delta\hat{F}(x_i)h(x_i) &= \frac{\theta_2-\theta_1}{n_2}\left\{\sum_{\delta_i=1, i\le n_2} h(x_i)\right. \\ &+ \sum_{\delta_i=0, i\le n_2}\frac{\frac{1-\theta_2}{\theta_2-\theta_1}\sum_{i=1}^{n_2}\Delta\hat{F}(x_i)h(x_i) + \sum_{x_j>x_i}\Delta\hat{F}(x_j)h(x_j)}{(1-\theta_2)+\sum_{x_j>x_i}\Delta\hat{F}(x_j)} \\ &+ \left.\sum_{\delta_i=2, i\le n_2}\frac{\sum_{x_j<x_i}\Delta\hat{F}(x_j)h(x_j) + \frac{\theta_1}{\theta_2-\theta_1}\sum_{i=1}^{n_2}\Delta\hat{F}(x_i)h(x_i)}{\theta_1+\sum_{x_j<x_i}\Delta\hat{F}(x_j)}\right\}\ .\end{aligned}$$

Now we take $h(x_i) = I_{[x_i\le u]}$ for $T_1 \le u \le T_2$, the above equation can be rewritten as (3.1).

Table 6.6: Marijuana Use In High School Boys

Age	Number of Exact Observations	Number Who Have Yet to Smoke Marijuana	Number Who Have Started Smoking at an Earlier Age
10	4	0	0
11	12	0	0
12	19	2	0
13	24	15	1
14	20	24	2
15	13	18	3
16	3	14	2
17	1	6	3
18	0	0	1
> 18	4	0	0

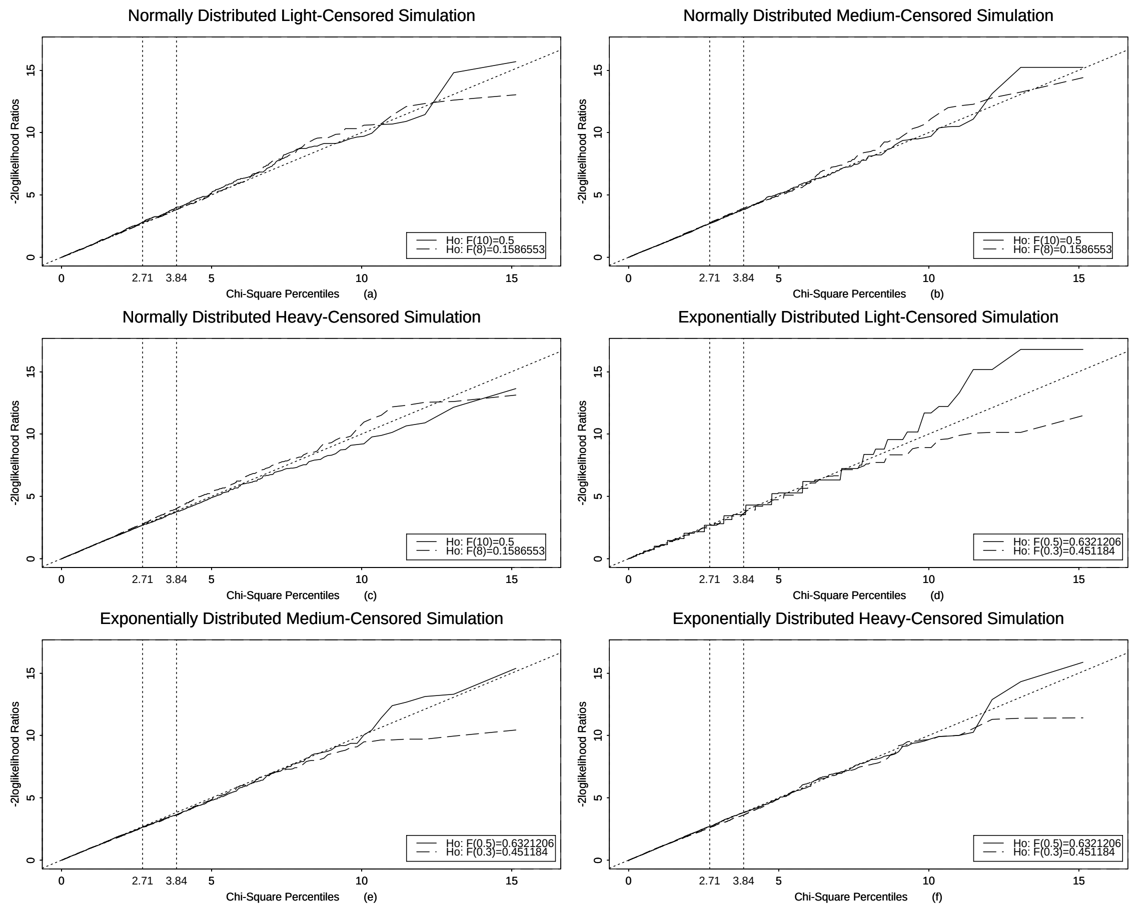

Figure 6.1: Q-Q Plot of -2log-likelihood Ratios vs. $\chi^2_{(1)}$ Percentiles For Sample Size $= 100$

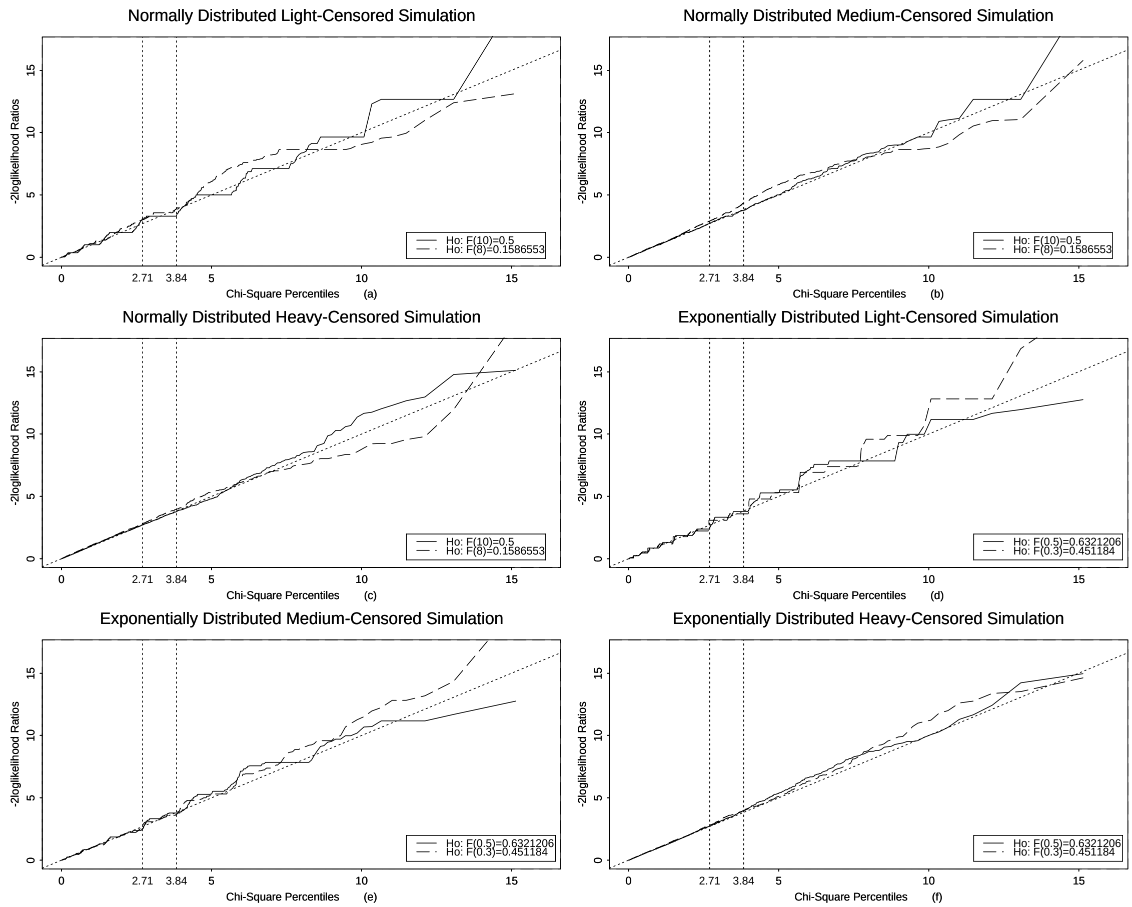

Figure 6.2: *Q-Q Plot of* -2*loglikelihood Ratios vs.* $\chi^2_{(1)}$ *Percentiles For Sample Size* $= 25$