Empirical likelihood ratio with arbitrarily censored/truncated data by EM algorithm

Mai Zhou [1]

University of Kentucky, Lexington, KY 40506 USA

Summary. Empirical likelihood ratio method (Thomas and Grunkmier 1975, Owen 1988, 1990, 2001) is a general nonparametric inference procedure that has many nice properties. Recently the procedure has been shown to work with some censored/truncated data with various parameters. But the computation of the empirical likelihood ratios with censored/truncated data and parameter of mean is non-trivial. We propose in this paper to use a modified self-consistency/EM algorithm (Turnbull 1976) to compute a class of arbitrarily censored/truncated empirical likelihood ratios where the constraint is of mean type.

Tests and confidence intervals based on the censored/truncated likelihood ratio performs well. Examples and simulations are given in the following cases: (1) right censored data with a mean parameter; (2) left truncated and right censored data with mean type parameter.

AMS 1991 Subject Classification: Primary 62G10; secondary 62G05.
Key Words and Phrases: Self consistency, maximization, constraints, Wilks theorem.

1. Introduction

Empirical likelihood ratio method was first used by Thomas and Grunkmier (1975) in connection with the Kaplan-Meier estimator. Owen (1988, 1990, 1991) and many others developed this into a general methodology. It has many desirable statistical properties, see the recent nice book of Owen (2001). A crucial step in applying the empirical likelihood ratio method is to find the maximum of the log empirical likelihood function (LELF) under some constraints. In all the papers mentioned above, that is achieved by using the Lagrange multiplier method. It reduces the maximization of n probabilities to a set of p monotone equations (for the multiplier λ), and p is fixed as n go to infinity. These equations can easily be solved, and thus empirical likelihood ratio can be easily computed.

Recently the empirical likelihood ratio method has been shown to work also with censored data and the parameter of mean. Pan and Zhou (1999) showed that for right censored data the empirical likelihood ratio with mean constraint also have a chi-square limit (Wilks theorem). Murphy and Van der Vaart (1997) demonstrated, among other things, that the Wilks theorem hold for doubly censored data too.

Theorem 1 (Pan and Zhou) *For the right censored data defined in (1) with a continuous distribution F, suppose the constraint equation is*

$$\int g(t)dF(t) = \theta_0$$

[1]Mai Zhou, Department of Statistics, University of Kentucky, Lexington, KY 40506 USA E-mail: mai@ms.uky.edu

where θ_0 is the true value (i.e. $\theta_0 = \int g dF_0$). If $g(t)$ satisfies certain regularity conditions, and $h(t)$ is another function that satisfy same regularity conditions, then as $n \to \infty$, the empirical likelihood ratio on a sub-family of distributions indexed by $h(\cdot)$ has the limit

$$-2 \log ELR(\theta_0, h) \xrightarrow{\mathcal{D}} r_h \times \chi^2_{(1)}$$

where the constant

$$r_h = \frac{Asy\ Var(\int g d\hat{F}_{KM}) \times \left(\int h^2 (1-G) dF + \int \frac{\int_t^\infty h^2(s) dF(s)}{1-F(t)} dG(t) - [\int h dF]^2 \right)}{\left(\int ghdF \right)^2},$$

and G is the CDF of the censoring times. Furthermore, the minimum value of the constant r_h over h is one.

Theorem 2 (Murphy and Van der Vaart) *For doubly censored observations (see definition in Example 2), suppose the distribution functions of the random variables involved are continuous and satisfy certain other regularity conditions. Let g be a left continuous function of bounded variation which, is not F_0-almost everywhere equal to a constant. If $\int g dF_0 = \theta_0$, then the likelihood ratio statistic for testing $H_0 : \int g dF = \theta_0$ satisfies $-2 \log ELR(\theta_0)$ converges to $\chi^2_{(1)}$ under F_0.*

For truncated data we refer readers to Li (1995), but the result is only for $g(t) = I_{[t \le C]}$.

One of the advantages of empirical likelihood method is that we can construct confidence intervals without estimating the variance of the statistic, which could be very difficult as in the situation of Theorem 2.

However, in the proofs of the Wilks theorem for the censored empirical likelihood ratio in the above two papers, the maximization of the log likelihood is more complicated then straight use of Lagrange multiplier. It is more of an existence proof rather than a constructive proof. In fact it involves least favorable sub-family of distributions and the existence of such, and thus it do not offer a viable computational method for the maximization of the empirical likelihood under constraint.

Therefore the study of computational method that can find the relevant censored/truncated empirical likelihood ratios numerically is needed. A good computational method will make the above nice theoretical results practical. We propose in this paper to use a modified EM algorithm to achieve that. We have implemented the algorithm in R software (Gentleman and Ihaka 1998). In fact, the modified EM/self-consistent algorithm we propose can be used to compute empirical likelihood ratios in many other types of censored/truncated data cases as described in Turnbull (1976).

Of course, for problems where a simple Lagrange multiplier computation is available, it will usually be faster than the EM algorithm. Uncensored data, or right censored data with weighted *hazard constraint* are such cases, see (Pan and Zhou 2000) for details. But as we point out above, this is not the case for *mean type constraints* with censored/truncated data.

We end this section with two specific examples where we introduce the notation and setup of censored data with the mean constraint case.

Example 1 Suppose i.i.d. observations $X_1, \cdots, X_n \sim F(\cdot)$ are subject to right censoring so that we only observe

$$Z_i = \min(X_i, C_i) \ ; \qquad \delta_i = I_{[X_i \leq C_i]}, \qquad i = 1, 2, \ldots, n; \tag{1}$$

where $C_1, \cdots, C_n$ are censoring times.

The log empirical likelihood function (LELF) for the survival distribution based on the censored observations (Z_i, δ_i) is

$$L(p) = LELF = \sum_{i=1}^{n} \left[\delta_i \log p_i + (1 - \delta_i) \log \left(\sum_{Z_j > Z_i} p_j \right) \right] \ . \tag{2}$$

where $p_i = \Delta F(Z_i) = F(Z_i) - F(Z_i-)$.

To compute the empirical likelihood ratio (Wilks) statistic for testing the hypothesis: $mean(F) = \mu$, we need to find the maximum of the above LELF with respect to p_i under the constraint

$$\sum_{i=1}^{n} p_i Z_i = \mu \ , \qquad \sum_{i=1}^{n} p_i = 1 \ , \qquad p_i \geq 0 \ ; \tag{3}$$

where μ is given. Similar arguments to those of Li (1995) show that the maximization will force the $p_i = 0$ except when Z_i is an uncensored observation. We focus on finding those p_is. The straight application of Lagrange multiplier method leads to the equations

$$\frac{\delta_i}{p_i} + \sum_{k=1}^{n} (1 - \delta_k) \frac{I_{[Z_k < Z_i]}}{\sum_{Z_j > Z_k} p_j} - \lambda Z_i - \gamma = 0 \ ; \qquad \text{for} \ \ \delta_i = 1$$

which do not have a simple solution for p_i.

The calculations for the mean type constraint $\int g(t) dF(t) = \mu$ has similar difficulty and the solution with EM algorithm is also similar.

Example 2: Let $X_1, \cdots, X_n$ be positive random variables denoting the lifetimes which is i.i.d. 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 $Z_i \leq Y_i$ 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

$$T_i = \max(\min(X_i, Y_i), Z_i) \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } Z_i \le X_i \le Y_i \\ 0 & \text{if } X_i > Y_i \\ 2 & \text{if } X_i < Z_i \end{cases} \qquad i = 1, 2, \cdots, n. \tag{4}$$

The log empirical likelihood for the lifetime distribution F is

$$L(p) = \sum_{\delta_i=1} \log p_i + \sum_{\delta_i=0} \log(\sum_{Z_j > Z_i} p_j) + \sum_{\delta_i=2} \log(\sum_{Z_j < Z_i} p_j) \ . \tag{5}$$

We show how to use the EM algorithm to compute the maximum of the above empirical likelihood (under constraint), and also in many other truncated data cases. Examples and simulations are given in section 5.

2. Maximization of empirical likelihood with uncensored, weighted observations

The following is basically a weighted version of Owen (1990) Theorem 1. Suppose we have independent (uncensored, not truncated) observations $X_1, \cdots, X_n$ from distribution $F(\cdot)$. Associated with the observations are non-negative weights $w_1, \cdots, w_n$. The meaning of the weights are such that if $w_i = 2$, it means X_i is actually 2 observations tied together, etc. But we allow the weights to be fractions for the application later.

The empirical likelihood based on the weighted observations is $\prod (p_i)^{w_i}$ and the log empirical likelihood is

$$\sum w_i \log p_i \ . \tag{6}$$

Theorem 3 *The maximization of the log empirical likelihood (6) with respect to p_i subject to the two constraint:*

$$\sum p_i = 1 \ , \qquad \sum g(X_i) p_i = \mu$$

is given by the formula

$$p_i = \frac{w_i}{\sum_j w_j + \lambda(g(X_i) - \mu)}$$

where λ is the solution of the equation

$$\sum_i \frac{w_i (g(X_i) - \mu)}{\sum_j w_j + \lambda(g(X_i) - \mu)} = 0 \ .$$

For μ in the range of the the $g(X_i)$'s there exist a unique solution of the λ and the p_i given above is also positive.

The proof of the above theorem is similar to (Owen 1990) and we omit the details here.

3. The constrained EM algorithm for censored data

There is a large amount of literature on the EM algorithm, see for example Dempster, Laird, and Rubin (1977). For the particular setting where the parameter is the CDF and observations are censored, see Efron (1967) and Turnbull (1974, 1976). In particular, Turbull (1976) covers a variety of censored/truncated data cases.

It is known that the EM algorithm will converge (eg. starting from the empirical distribution based on uncensored data only) for the nonparametric estimation of the survival function with right censored data. However, EM algorithm was not used in that situation because an explicit formula exists (the Kaplan-Meier estimator). With a constraint on the mean, there no longer exists any explicit formula for right censored data. For doubly censored data, it is even worse: there is no explicit formula for the NPMLE with or without mean constraints. EM algorithm may be used to compute both NPMLE. And thus this present opportunity for EM to play its roll and show its muscle here.

We describe below the EM algorithm for censored data.

E-Step: Given F, the weight, w_j, at location t_j can be computed as

$$\sum_{i=1}^{n} E_F \left[I_{[X_i=t_j]} | Z_i, \delta_i \right] = w_j \ .$$

We only need to compute the weight for those locations that either (1) t_j is a jump point for the given distribution F, or (2) t_j is an uncensored observation. In many cases (1) and (2) coincide (eg. the Kaplan-Meier estimator). The w_i for other locations is obviously zero. Also when Z_i is uncensored, the conditional expectation is trivial.

M-Step: with the (uncensored) pseudo observations $X = t_j$ and weights w_j from E-Step, we then find the probabilities p_j by using our Theorem 3 above. Those probabilities give rise to a new distribution F.

A good initial F to start the EM calculation is the NPMLE without the constraint. In the case of right censored data that is the Kaplan-Meier estimator. If that is not easily available, like in doubly censored (or other) cases, a distribution with equal probability on all the possible jump locations can also be used.

The EM iteration ends when the predefined convergence criterion is satisfied.

Example (continue)

Suppose the i^{th} observation is a right censored one: $(\delta_i = 0)$ the E Step above can be computed as follows

$$\text{for } \; t_j > Z_i \; ; \qquad E_F[I_{[X_i=t_j]}|Z_i, \delta_i] = \frac{\Delta F(t_j)}{1 - F(Z_i)}$$

and $E_F[\cdot] = 0$ for $t_j \leq Z_i$.

For uncensored observation Z_i, it is obvious that $E_F[\cdot] = 1$ when $t_j = Z_i$ and $E_F[\cdot] = 0$ for any other t_j.

For left censored observation Z_i, the E Step above can be computed as follows

$$\text{for } \; t_j < Z_i \; ; \qquad E_F[I_{[X_i=t_j]}|Z_i, \delta_i] = \frac{\Delta F(t_j)}{F(Z_i)}$$

and $E_F[\cdot] = 0$ for $t_j \geq Z_i$.

Remark: The E-step above is no different then Turnbull (1976). For interval censored or even set censored data the E-Step can also be computed similarly. Our modification is in the M-step.

4. Truncated and censored data

Similar idea of last section actually carry through for arbitrarily truncated and censored data as described in Turnbull (1976). We first describe in some details the left truncated and right censored observation case, since this seems to be the most commonly seen situation. We then briefly outline the algorithm for the general case and a theorem that basically says the constrained NPMLE is equivalent to the solution of the modified self-consistent equation.

4.1 Left truncated and right censored case

For left truncated observations, there is an explicit expression for the NPMLE of CDF, the Lynden-Bell estimator. Li (1995) discussed the empirical likelihood where the parameter is the probability $F(t)$. For left truncated and right censored observations, there is also an explicit NPMLE of the CDF. (Tsai, Jewell and Wang 1987). But to compute the NPMLE under the mean constraint, we need the EM algorithm described here.

Suppose the observations are $(Y_1, Z_1, \delta_1), \cdots, (Y_n, Z_n, \delta_n)$ where the Y's are the left truncation times, Z's are the (possibly right censored) lifetimes. Denote by X the lifetimes before truncation/censoring. Censoring indicator δ assumes the usual meaning that $\delta = 1$ means Z is uncensored, $\delta = 0$ means Z is right censored. Truncation means for all i, $(Z_i > Y_i)$, and n is random. We assume Y is independent of X and both distributions are unknown.

The log likelihood pertaining the distribution of X is

$$L(p) = \sum_{i:\delta_i=1} \left(\log p_i - \log(\sum_{Z_j > Y_i} p_j) \right) + \sum_{i:\delta_i=0} \left(\log(\sum_{Z_j > Z_i} p_j) - \log(\sum_{Z_j > Y_i} p_j) \right) \ .$$

The NPMLE of the CDF puts positive probability only at the locations of observed, uncensored Z_i's. Denote those locations by t_j.

E-step Given a current estimate $F(\cdot)$ that have positive probability only at t_j's, we compute the weight

$$w_j = \sum_{i=1}^{n} E_F[I_{[X_i=t_j]}|X_i, \delta_i] + \sum_{i=1}^{n} I_{[t_j<Y_i]} \Delta F(t_j)/P_F(X > Y_i) \ ,$$

M-step with the pseudo observations t_j and associated weights w_j obtained in the E-step, we compute a new probability as described in Theorem 3, where the mean constraint weighs in.

The E-step above can be written more explicitly by noticing that (1) the E_F part can be computed same as in the example of the censored data case, and (2) the second term is (without summation)

$$I_{[t_j<Y_i]} \Delta F(t_j)/P_F(X > Y_i) = \frac{I_{[t_j<Y_i]} p_j}{\sum_k I_{[t_k>Y_i]} p_k}$$

where we used $p_j = \Delta F(t_j)$.

4.2 The general case

This subsection uses the same setup and notation of Turnbull (1976) and should be read along side that paper.

Our self-consistent equation with a mean constraint

$$\sum_{j=1}^{m} s_j g(t_j) = \mu \tag{7}$$

(in the context of Turnbull 1976) is just

$$\pi_j^*(s) = s_j \qquad j = 1, 2, \cdots m \tag{8}$$

where

$$\pi_j^*(s) = \frac{\sum_{i=1}^{N} \{\mu_{ij}(s) + \nu_{ij}(s)\}}{M(s) + \lambda(g(t_j) - \mu)} \ . \tag{9}$$

In the above t_j is any value picked (but fixed) inside the interval $[q_j, p_j]$, $M(s)$, μ_{ij}, ν_{ij} are as defined by Turnbull 1976, and λ is the solution of the following equation:

$$0 = \sum_{j=1}^{m} \frac{(g(t_j) - \mu) \times \sum_{i=1}^{N} \{\mu_{ij}(s) + \nu_{ij}(s)\}}{M(s) + \lambda(g(t_j) - \mu)} \ . \tag{10}$$

The function $g(\cdot)$ and constant μ are assumed given.

We now consider the equivalence of the modified self-consistency equation (8) with the constrained NPMLE. The log likelihood of the data is given by Turnbull (1976), equation (3.6). To maximize it under the mean constraint (7) and $\sum s_j = 1$, we proceed by Lagrange multiplier. Taking partial derivative of the target function

$$G = \sum_{i=1}^{N} \left\{ \log(\sum_{j=1}^{m} \alpha_{ij} s_j) - \log(\sum_{j=1}^{m} \beta_{ij} s_j) \right\} - \gamma(\sum_{j=1}^{m} s_j - 1) - \lambda(\sum_{j=1}^{m} s_j(g(t_j) - \mu))$$

with respect to s_j we get

$$d_j^*(s) = \sum_{i=1}^{N} \left\{ \frac{\alpha_{ij}}{\sum_{k=1}^{m} \alpha_{ik} s_k} - \frac{\beta_{ij}}{\sum_{k=1}^{m} \beta_{ik} s_k} \right\} - \gamma - \lambda(g(t_j) - \mu) \ . \tag{11}$$

For s to be the (constrained) NPMLE, those partial derivatives must be zero. Multiply each of the partial derivatives $d_j^*(s)$ by s_j and summation over j, we get $\gamma = 0$.

The left side of self-consistent equation (8) can then be written as

$$\pi_j^*(s) = \frac{s_j}{M(s) + \lambda(g(t_j) - \mu)} \left\{ d_j^*(s) + \lambda(g(t_j) - \mu) + \sum_{i=1}^{N} \left(\sum_{k=1}^{m} \beta_{ik} s_k \right)^{-1} \right\} \ .$$

Similar to Turnbull 1976, we finally have

$$\pi_j^*(s) = \left\{ 1 + \frac{d_j^*(s)}{M(s) + \lambda(g(t_j) - \mu)} \right\} s_j \ .$$

So the self-consistent equation becomes

$$\left\{ 1 + \frac{d_j^*(s)}{M(s) + \lambda(g(t_j) - \mu)} \right\} s_j = s_j \ .$$

Now a similar argument to Turnbull 1976 leads to the

Theorem 4 *The solution of the constrained log likelihood equation (11) is equivalent to the solution to the self-consistent equations (8).*

5. Empirical Likelihood Ratio Computation

Once the NPMLE of probabilities, p_i, are computed, we can plug them into the log likelihoods as in (2) or (5) or other cases to get the censored/truncated log empirical likelihood with mean constraint easily. This in turn allows us to compute the empirical log likelihood ratio statistic:

$$\begin{aligned} -2 \log R(H_0) &= -2 \log \frac{\max_{H_0} L(p)}{\max_{H_0+H_1} L(p)} & (12) \\ &= 2 \left[\log(\max_{H_0+H_1} L(p)) - \log(\max_{H_0} L(p)) \right] \\ &= 2 \left[\log(L(\tilde{p})) - \log(L(\hat{p})) \right] \ . & (13) \end{aligned}$$

Here $\tilde{p}$ is the NPMLE of probabilities without any constraint; $\hat{p}$ is the NPMLE of probabilities under H_0 constraint.

Both NPMLEs can be computed by EM algorithm, like in section 3. In some cases, there may be other (faster) methods available to compute $\tilde{p}$, the NPMLE without constraint. A case in point is the Kaplan-Meier estimator for the right censored data.

After we obtained the $\tilde{p}$ and $\hat{p}$, the likelihood ratio can be computed and Wilks theorem can then be used to find the P-value of the observed statistic. Thus we can use empirical likelihood ratio to test hypothesis and construct confidence intervals. To illustrate we will show some examples and simulation results in the next section.

6. Simulations and Examples

In this section example with real data and simulation results are reported for right-censored/left-truncated data to illustrate the usefulness of the proposed EM method, and also to illustrate the small to medium sample performance of the chi square approximation.

We have implemented this EM computation in R software (Gentleman and Ihaka 1996). It is available as a package `emplik` at one of the CRAN web site (http://cran.us.r-project.org). The R function `el.cen.EM` is for right, left or doubly censored observations with a mean type constraint. The R function `el.ltrc.EM` is for left truncated and right censored data with a mean type constraint.

6.1 Confidence Interval, real data, right censored

The first example concerns Veteran's Administration Lung cancer study data (for example available from the R package `survival`). We took the subset of survival data for treatment 1 and smallcell group. There are two right censored observations. The survival times are:

30, 384, 4, 54, 13, 123+, 97+, 153, 59, 117, 16, 151, 22, 56, 21, 18, 139, 20, 31, 52, 287, 18, 51, 122, 27, 54, 7, 63, 392, 10.

We use the EM algorithm to compute the log empirical likelihood with constraint $mean(F) = \mu$ for various values of μ. The log empirical likelihood has a maximum when $\mu = 94.7926$, which is the mean computed from the Kaplan-Meier estimator.

The 95% confidence interval for the mean survival time is seen to be $[61.70948, 144.912]$ since the log empirical likelihood was $3.841/2 = \chi^2(0.95)/2$ below the maximum value (= -93.14169) both when $\mu = 61.70948$ and $\mu = 144.912$. We see that the confidence interval is not symmetric

around the MLE, a nice feature of the confidence intervals derived from the empirical likelihood.

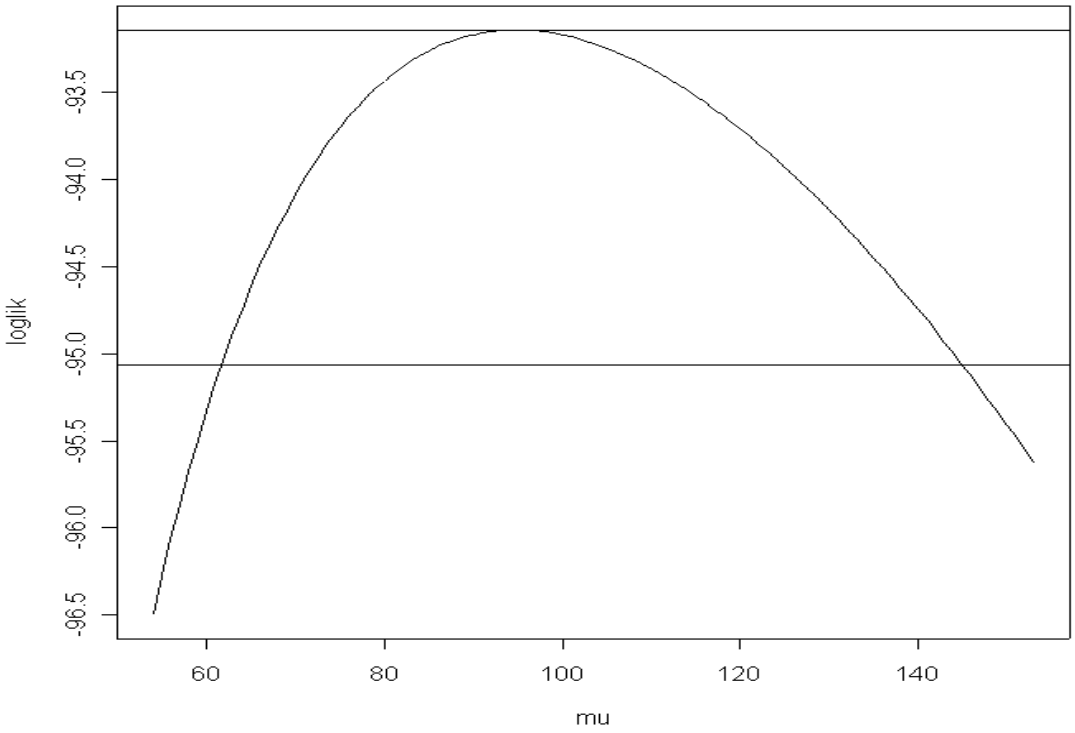

Figure 1: Log likelihood for μ near maximum

6.2 Simulation: right censored data

It is generally believed that for the smaller sample sizes the likelihood ratio/chi square based inference is more accurate then those obtained by Wald method. The Q-Q plot from the following simulation shows that the chi square distribution is a pretty good approximation of the -2 log empirical likelihood ratio statistic for right censored data and mean parameter.

We randomly generated 5000 right-censored samples, each of size $n = 50$ as in equation (1), where $X \sim exp(1)$ and $C \sim exp(0.2)$ and $g(t) = I_{[t\leq 1]}$, or $g(t) = (1-t)I_{[t\leq 1]}$. i.e. the constraint is $\int_0^1 g(t)d(1-\exp(-t)) = \mu$. Both plots look similar, we only show here the one with $g(t) = (1-t)I_{[t\leq 1]}$.

We computed 5000 empirical likelihood ratios, using the Kaplan Meier estimator's jumps as $(\tilde{p})$ which maximizes the denominator in (13) and the modified EM method of section 3 gave $(\hat{p})$ that maximizes the numerator under H_0 constraint. The Q-Q plot is based on 5000 empirical likelihood ratios and χ^2_1 percentiles, and is shown in Figure 1 Two vertical lines were drawn at the point 3.84

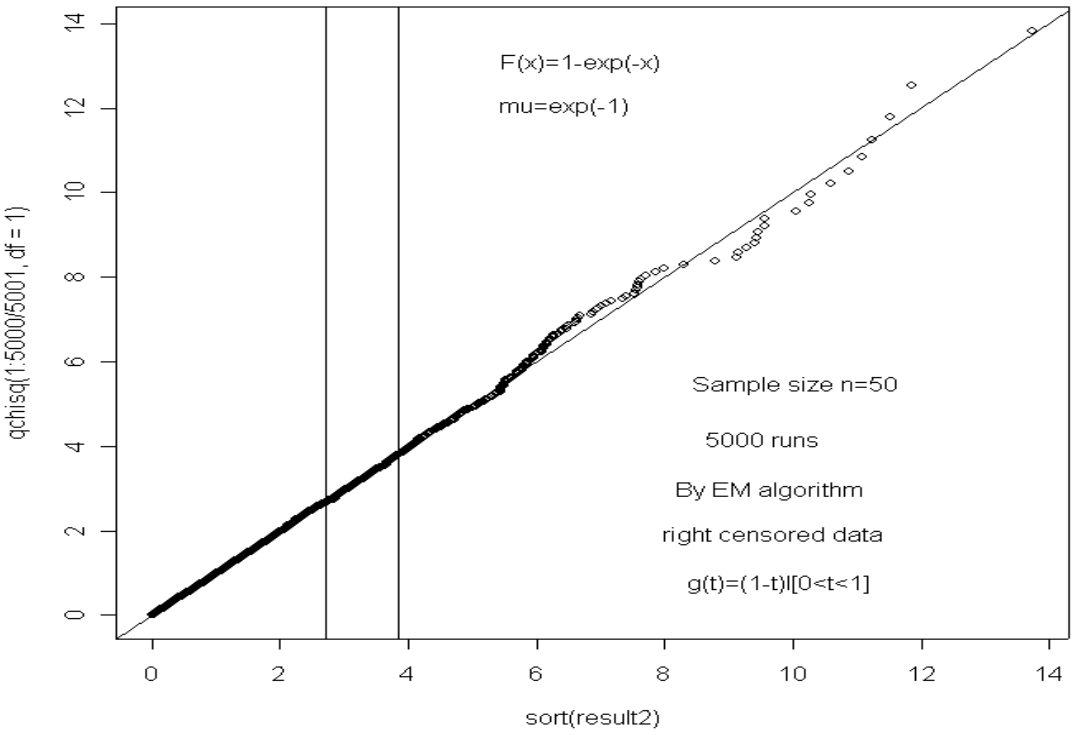

Figure 2: A Q-Q plot for right censored likelihood ratio

and 2.71 which are the critical values of χ^2_1 with nominal level 5% or 10%. From the Q-Q plot, we can see that the χ^2_1 approximation is pretty good since the -2log-likelihood ratios are very close to χ^2_1 percentiles. Only at the tail of the plot, the differences between -2log-likelihood ratios and χ^2_1 are getting bigger.

6.3 Simulation and example – Left truncated, right censored Case

We generate (left) truncation times, Y, as shifted exponential distributed random variables, $exp(4)-0.1$. We generate lifetimes X distributed as $exp(1)$ and censoring times C distributed as $exp(0.15)$. The truncation probability $P(Y > X)$ is around 13.4%. The censoring probability $P(X > C)$ is around 13%.

The triplets, (Y, X, C), are rejected unless we have $Y < X$ and $Y < C$. In that case we return the triplets Y, $\min(X, C)$ and $d = I_{[X \leq C]}$. In the simulation, 50 triplets $\{Y,\ \min(X, C),\ d\}$ are generated each time a simulation is run. The function $g(t)$ we used is $t(1-t)I_{[0<t<1]}$. The mean of this function is $\mu = (3e^{-1} - 1)$.

Lastly, let us look at a small data taken from the book of Klein and Moeschberger (1997). The

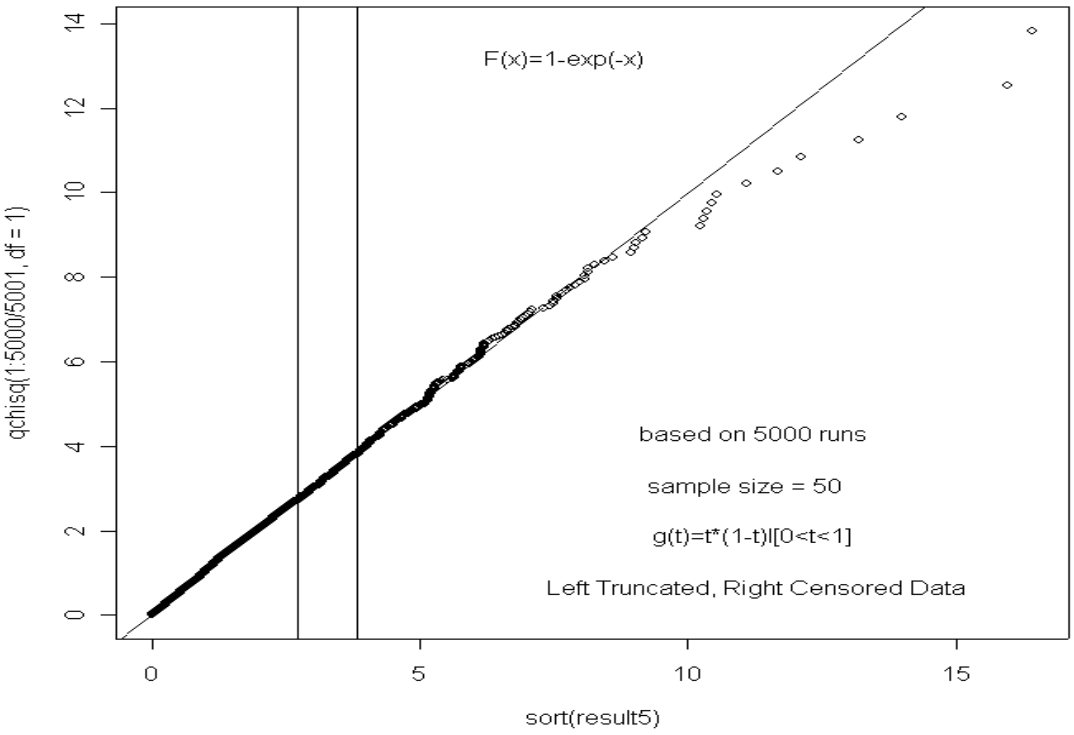

Figure 3: Q-Q plot of -2log-likelihood ratios vs. $\chi^2_{(1)}$ percentiles for sample size 50

survival times of female psychiatric inpatients as reported in Table 1.7 on page 16 of the above book. $Y = (51, 58, 55, 28, 25, 48, 47, 25, 31, 30, 33, 43, 45, 35, 36)$;
$Z = (52, 59, 57, 50, 57, 59, 61, 61, 62, 67, 68, 69, 69, 65, 76)$ and $d = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1)$. The mean computed from Tsai-Jewell-Wang estimate is 63.18557. The plot of -2 log likelihood ratio against changing μ value is similar to Figure 1 and is omitted. The -2 log likelihood ratio have a minimum of zero at $\mu = 63.18557$ as it should be. A 95% confidence interval for μ are those values of μ that the -2log likelihood ratio is less than 3.84. In this case it is $[58.78936, 67.81304]$.

7. Discussion

The computational algorithm proposed in this paper covers a wide variety of censored/truncated data cases as in Turnbull (1976). It enables us to compute the NPMLE of CDF under a mean type constraint. It also enables us to compute the $-2\log$ empirical likelihood ratio in those cases. Coupled with empirical likelihood theory (Wilks theorem), the latter can be used to do inference on the NPMLE of CDF.

The asymptotic theory for the constrained NPMLE and for the empirical likelihood ratio lags behind the computation. There is yet a result that covers all the cases described in Turnbull (1976), but see Owen (2001), Murphy and van der Vaart (1997) for some known cases. We conjecture that for left truncated and right censored observations, the asymptotic χ^2 distribution remains valid for the empirical likelihood ratio with a mean constraint (Wilks theorem).

One of the advantages of EM algorithm is that it requires minimal computer memory. In the iteration, we only need to store the current copy of $F(\cdot)$ and vector w. In contrast, the Sequential Quadratic Programming method, (Chen and Zhou 2001, Owen 2001), which try to minimize the censored empirical likelihood (2) by quadratic approximation, needs to store matrices of size $n \times n$. This advantage is most visible for samples of size above 500 in our experience.

References

Chen, K. and Zhou, M. (2001). Computing the censored empirical likelihood via sequential quadratic programming. Dept. Statist. Univ. of Kentucky Tech Report. Accepted for publication.

Gentleman, R. and Ihaka, R. (1996). R: A Language for data analysis and graphics. *J. of Computational and Graphical Statistics*, **5**, 299-314.

Klein and Moeschberger (1997). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer, New York

Li, G. (1995). Nonparametric likelihood ratio estimation of probabilities for truncated data. *JASA* **90**, 997-1003.

Murphy, S. and van der Vaart, A. (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.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**, 90-120.

Owen, A. (1991). Empirical Likelihood for Linear Models. *The Annals of Statistics*, **19** 1725-1747.

Owen, A. (2001). *Empirical likelihood.* Chapman & Hall, London.

Pan, X.R. and Zhou, M. (1999). Using one parameter sub-family of distributions in empirical likelihood with censored data. *J. Statist. Planning and Infer.*

Pan, X.R. and Zhou, M. (2001). Using one parameter sub-family of distributions in empirical likelihood with censored data. *J. Statist. Planning and Infer.*

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Amer. Statist. Assoc.* **70**, 865-871.

Tsai, W-Y, Jewell, N.P. and Wang, M-C. (1987). The product limit estimate of a survival curve under right censoring and left truncation. *Biometrika* **74**, 883-886.

Turnbull, B. (1976), *The empirical distribution function with arbitrarily grouped, censored and truncated data. JRSS* **B**, 290-295.

Turnbull, B. (1974), *Nonparametric estimation of a survivorship function with doubly censored data. JASA* 169-173.