COMPUTE CENSORED EMPIRICAL LIKELIHOOD RATIO BY SEQUENTIAL QUADRATIC PROGRAMMING

KUN CHEN AND MAI ZHOU

University of Kentucky

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 also work with censored data with various parameters and the nice properties also hold there. But the computation of the empirical likelihood ratios with censored data and/or complex setting is often non-trivial. We propose to use sequential quadratic programming (SQP) method to overcome the computational problem.

Examples are given in the following cases: (1) right censored data with a parameter of mean; (2) Interval censored data.

AMS 1991 Subject Classification: Primary 62G10; secondary 62G05.
Key Words and Phrases: 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 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 and much smaller then n. These equations can easily be solved.

Recently the empirical likelihood ratio method has been shown to also work with censored data and 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, 2000) demonstrated, among other things, that the Wilks theorem hold for doubly censored data too.

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

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

where θ_0 is the true value (i.e. $\theta_0 = \int g dF_0$) and $g(t)$ satisfies certain regularity conditions, then as $n \to \infty$, the empirical likelihood ratio

$$-2 \log ELR(\theta_0) \xrightarrow{\mathcal{D}} \chi^2_{(1)} \ .$$

Theorem 2 (Murphy and Van der Vaart) *For doubly censored observations, 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.*

However, in the proofs of the Wilks theorem for the censored empirical likelihood ratio in the above theorem/papers, the maximization of the log likelihood is more complicated then straight use of Lagrange multiplier. It is more of an existence proof rather then 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 computation method that can find the relevant empirical likelihood ratios numerically is needed and crucial for the above nice results to become reality in practice when analyzing censored data. We propose to use the sequential quadratic programming (SQP) method to achieve that. In fact it can compute empirical likelihood ratio in many other cases (for example, interval censored data) where a simple Lagrange multiplier computation is not available.

One drawback of the SQP method is that it becomes more memory/computation intensive for larger sample sizes, n. This should contrast to the Lagrange multiplier method mentioned above where p remains fixed as sample size n increases. However, we argue that this is not a major drawback because (1) the nice properties of the empirical likelihood ratio method are most visible for small to medium sample sizes. For large samples there often are alternative, equally good methods available. (2) By our implementation of the SQP method in R (Gentleman and Ihaka, 1996), we can easily handle (about 1 minute) sample sizes of up to 2000 on today's average PC (1 GHz, 256MB). No doubt our implementation may be improved (eg. implement it in C). And with memory getting cheaper, this drawback should diminish and not pose a major handicap for the usefulness of the SQP method in empirical likelihood.

To fix an idea we now introduce the notation and setup of the right-censored data with the mean constraint case.

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

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

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

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

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

To compute the Wilks statistic for test hypothesis, we need to find the maximum of the above LELF under the constraint

$$\sum_{i=1}^{n} w_i Z_i \delta_i = \mu \ , \qquad \sum_{i=1}^{n} w_i \delta_i = 1 \ , \qquad w_i \ge 0 \ ;$$

where μ is a given constant. The straight application of Lagrange multiplier method do not leads to a simple solution. Doubly censored data and other censoring leads to the same difficulty.

We show how to use the sequential quadratic programming method (SQP) to compute the above empirical likelihood and the likelihood ratio in section 3, and comment on how to similarly use it to compute empirical likelihood ratio in many other cases. Examples and simulations are given in section 5.

2. The Sequential Quadratic Programming Method

There is a large amount of literature on the nonlinear programming methods, see for example Nocedal and Wright (1999) and reference there. The general strictly convex (positive definite) quadratic programming problem is to minimize

$$f(\mathbf{x}) = -\mathbf{a}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} \ , \tag{3}$$

subject to

$$s(\mathbf{x}) = \mathbf{C}^T \mathbf{x} - \mathbf{b} \ge \mathbf{0} \ , \tag{4}$$

where $\mathbf{x}$ and $\mathbf{a}$ are n-vectors, $\mathbf{G}$ is an $n \times n$ symmetric positive definite matrix, $\mathbf{C}$ is $n \times m$ $(m < n)$ matrix, $\mathbf{b}$ is an m-vector, and superscript T denotes the transpose. The algorithm of Goldfarb and Idnani (1983) to solve the QP problem is very stable numerically and available in many packages like Matlab or R. In this paper, the vector $\mathbf{x}$ is only subject to equality constraints $\mathbf{C}^T\mathbf{x} - \mathbf{b} = \mathbf{0}$. This makes the QP problem easier to solve. In next section we shall show that we can introduce a few new variables in the maximizing of the censored LELF (2) so that the matrix $\mathbf{G}$ is always diagonal, which further simplifies the computation. Therefore instead of use general QP algorithm, we implemented our own version in R which takes advantage of the said simplifications. The specific QP problem can be solved by performing one matrix QR decomposition, one backward solve, and one forward solve of equations.

Iteration is needed in the QP since the quadratic approximation to LELF is only good locally. The matrix $\mathbf{G}$ is the second derivative of LELF and is diagonal (see next section), the column vector $\mathbf{a}$ (of length n) is the first derivative of LELF. Both derivatives should be evaluated at the best current solution $\mathbf{x}^{(i-1)}$.

Since all our constraints are *equality* constraints, one way of solve the minimization problem (3) is to use (yet again) the Lagrange multiplier:

$$\min_{x,\eta} \quad -\mathbf{a}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} - \eta^T[\mathbf{C}^T\mathbf{x} - \mathbf{b}]$$

where η is a column vector of length m. Taking derivative with respect to $\mathbf{x}$ and set it equal to zero, we get

$$\mathbf{G}\mathbf{x} - \mathbf{a} - \mathbf{C}\eta = \mathbf{0} \ .$$

We can solve $\mathbf{x}$ in terms of η to get

$$\mathbf{x} = \mathbf{G}^{-1}[\mathbf{a} + \mathbf{C}\eta] \ . \tag{5}$$

Since the matrix $\mathbf{G}$ is diagonal, the inverse $\mathbf{G}^{-1}$ is easy to get. Finally we need to solve for η. Substitute (5) into $\mathbf{C}^T\mathbf{x} = \mathbf{b}$ we get

$$\mathbf{C}^T(\mathbf{G}^{-1}[\mathbf{a} + \mathbf{C}\eta]) = \mathbf{b}$$

which is, upon rewriting,

$$\mathbf{C}^T\mathbf{G}^{-1}\mathbf{C}\eta = \mathbf{b} - \mathbf{C}^T\mathbf{G}^{-1}\mathbf{a} \ . \tag{6}$$

Once we get the solution η from this equation we can substitute it back into (5) above to get $\mathbf{x}$.

In the iteration, the matrix $\mathbf{C}$ and vector $\mathbf{b}$ do not change, only the diagonal matrix $\mathbf{G}$ and the vector $\mathbf{a}$ needs to be updated, because they are the derivatives of the log likelihood function at current solution.

One way to solve (6) is to use QR decomposition. If $\mathbf{C}^T\mathbf{G}^{-1/2} = \mathbf{RQ}$ then the above equation can be rewritten as

$$(\mathbf{RQQ}^T\mathbf{R}^T)\eta = \mathbf{b} - \mathbf{RQG}^{-1/2}\mathbf{a}$$
$$(\mathbf{RR}^T)\eta = \mathbf{b} - \mathbf{RQG}^{-1/2}\mathbf{a}$$
$$\mathbf{R}^T\eta = \mathbf{R}^{-1}\mathbf{b} - \mathbf{QG}^{-1/2}\mathbf{a} \tag{7}$$

The solution of (7) can be obtained by using back-substitute (twice) and one matrix-vector multiplication, which are low cost operation numerically.

We are interested in maximizing LELF, or minimize the negative LELF. This is a nonlinear programming problem. Since it is hard to find minimum of negative LELF directly in many cases, and the negative LELF is often convex, we use a quadratic function to approximate it. Starting from an initial probability w, we replace the nonlinear target function (-LELF) by a quadratic function which has the same first and second derivatives as the negative LELF at the location of initial probability. QP method is used to find the minimum of the quadratic function subject to the same given constraints. Then we update the quadratic approximate which now has the same first and second derivatives as the negative LELF at the location of the updated probability (the minimum of previous quadratic function). QP method is used again to find the minimum of the new quadratic approximate function under the same constraints. Iteration ends until the predefined convergence criterion is satisfied. When the information matrix is positive, the quadratic approximation is good at least in a neighborhood of true MLE. Thus, when converged, the solution gives the correct MLE under given constraint.

3. Likelihood Maximization with Right Censored Data

We now describe the SQP method that solves the problem in Example 1. The implementation for doubly censored data, interval censored data are similar. We only give the detail for right censored data here.

When data are right-censored as in (1), the LELF becomes more complicated, we can rewrite (2) as

$$LELF = L(w) = \sum_{i=1,\delta_i=1}^{k} \log w_i + \sum_{l=1,\delta_l=0}^{n-k} \log \left(\sum_{Z_j > Z_l, \delta_j=1, 1\le j \le k} w_j \right) , \qquad (8)$$

where k is the number of uncensored observation, Z is the ordered observation, i.e. $Z_1 \le Z_2 \le \ldots \le Z_n$. We describe below two ways to implement SQP method for finding the constrained MLE.

The first, simple minded implementation of QP, would just take the $\mathbf{w}$ in (8) as $\mathbf{x}$, which has length k, the length of the vectors $\mathbf{a}$ is k and matrixes $\mathbf{G}$ is $k \times k$. The second derivative matrix $\mathbf{G}$ in the quadratic approximation is dense and the computation of the inverse/QR decomposition is expensive numerically.

A second and better way of use the SQP with censored data will first introduce some auxiliary variables $R_l = P(Z \ge Z_l)$, one for each censored observations, this enlarges the dimension of vectors ($\mathbf{a}$, $\mathbf{x}$, $\mathbf{b}$) and matrix ($\mathbf{G}$, $\mathbf{C}$) in (3), but simplifies the matrix $\mathbf{G}$. In fact $\mathbf{G}$ will be diagonal, so that we can directly plug-in the inverse of decomposition matrix of $\mathbf{G}$. It speeds up the computation tremendously.

We illustrate the two methods for the problem described in Example 1. In method one, since we know $w_i > 0$ only when the corresponding $\delta_i = 1$, we would separate the observations into two groups: $Z_1 < \cdots < Z_k$ for those with $\delta = 1$ and $Z_1^* < \cdots < Z_{n-k}^*$ for those with $\delta = 0$. The first derivative of log likelihood function is:

$$\frac{\partial L(w)}{\partial w_i} = \frac{1}{w_i} + \sum_{l=1}^{n-k} \frac{I_{[Z_i > Z_l^*]}}{\sum_{Z_j > Z_l^*, \delta_j=1, 1\le j\le k} w_j} ,$$

Let us define $M_l = \displaystyle\sum_{Z_j > Z_l, \delta_j=1, 1\le j \le k} w_j$, Then

$$\mathbf{a} = \begin{pmatrix} \dfrac{1}{w_1} + \displaystyle\sum_{l=1}^{n-k} \dfrac{I_{[Z_1 > Z_l^*]}}{M_l} \\ \dfrac{1}{w_2} + \displaystyle\sum_{l=1}^{n-k} \dfrac{I_{[Z_2 > Z_l^*]}}{M_l} \\ \vdots \\ \dfrac{1}{w_k} + \displaystyle\sum_{l=1}^{n-k} \dfrac{I_{[Z_k > Z_l^*]}}{M_l} \end{pmatrix} .$$

Taking the second derivative with respect to $w_i, i = 1, 2, \ldots, k$:

$$\frac{\partial^2 L(w)}{\partial^2 w_i} = -\frac{1}{w_i} - \sum_{l=1}^{n-k} \frac{I_{[Z_i > Z_l^*]}}{M_l^2} \ ,$$

and with respect w_q and $i \neq q$:

$$\frac{\partial^2 L(w)}{\partial w_i \partial w_q} = -\sum_{l=1}^{n-k} \frac{I_{[Z_i > Z_l^*]} I_{[Z_q > Z_l^*]}}{M_l^2} = \frac{\partial^2 L(w)}{\partial w_q \partial w_i} \ .$$

Therefore the second derivatives $\mathbf{G}$ is

$$\mathbf{G} = \begin{pmatrix} \frac{1}{w_1^2} + \sum_{l=1}^{n-k} \frac{I_{[Z_1 > Z_l^*]}}{M_l^2} & \sum_{l=1}^{n-k} \frac{I_{[Z_1 > Z_l^*]} I_{[Z_2 > Z_l^*]}}{M_l^2} & \ldots & \sum_{l=1}^{n-k} \frac{I_{[Z_1 > Z_l^*]} I_{[Z_k > Z_l^*]}}{M_l^2} \\ \sum_{l=1}^{n-k} \frac{I_{[Z_2 > Z_l^*]} I_{[Z_1 > Z_l^*]}}{M_l^2} & \frac{1}{w_2^2} + \sum_{l=1}^{n-k} \frac{I_{[Z_2 > Z_l^*]}}{M_l^2} & \ldots & \sum_{l=1}^{n-k} \frac{I_{[Z_2 > Z_l^*]} I_{[Z_k > Z_l^*]}}{M_l^2} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{l=1}^{n-k} \frac{I_{[Z_k > Z_l^*]} I_{[Z_1 > Z_l^*]}}{M_l^2} & \sum_{l=1}^{n-k} \frac{I_{[Z_k > Z_l^*]} I_{[Z_2 > Z_l^*]}}{M_l^2} & \ldots & \frac{1}{w_k^2} + \sum_{l=1}^{n-k} \frac{I_{[Z_k > Z_l^*]}}{M_l^2} \end{pmatrix} ,$$

$$\mathbf{x} = \begin{pmatrix} w_1 - w_1^\star \\ w_2 - w_2^\star \\ \vdots \\ w_k - w_k^\star \end{pmatrix} \quad , \quad \mathbf{C} = \begin{pmatrix} 1 & Z_1 \\ 1 & Z_2 \\ \vdots & \vdots \\ 1 & Z_k \end{pmatrix} \ .$$

We always use an initial value $\mathbf{w_0}$ that is a probability but may not satisfy the mean constraint. Therefore $\mathbf{b_0} = (0, \mu - \bar{Z})$. For subsequent iterations we have $\mathbf{b} = (0, 0)$ since the current value of $\mathbf{w}$ already satisfies both constraints.

For this QP problem, the matrix $\mathbf{G}$ must be positive definite. We shall show that indeed it is so here.

Theorem 3: The $\mathbf{G}$ matrix defined above is positive definite.

PROOF: Separate G into two parts: $\mathbf{G} = \mathbf{A} + \sum_{l=1}^{n-k} \mathbf{B}_l$, where

$$\mathbf{A} = \begin{pmatrix} \frac{1}{w_1^2} & 0 & \ldots & 0 \\ 0 & \frac{1}{w_2^2} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \frac{1}{w_n^2} \end{pmatrix} ;$$

$\mathbf{B}_l = e\ \mathbf{v}\ \mathbf{v}^T$, where

$$e = \frac{1}{(M_l)^2} \ , \qquad \mathbf{v}^T = \Big(\ I_{[Z_1 > Z_l^*]}\ ,\quad I_{[Z_2 > Z_l^*]}\ ,\quad \ldots\quad I_{[Z_k > Z_l^*]}\ \Big)\ .$$

Let $\mathbf{m}$ be any k-vector,

$$\mathbf{m}^T \mathbf{B}_l \mathbf{m} = \mathbf{m}^T (e \mathbf{v}\mathbf{v}^T)\mathbf{m} = e(\mathbf{v}^T\mathbf{m})^T(\mathbf{v}^T\mathbf{m}) = e\mathbf{z}^T\mathbf{z} = e\sum_{i=1}^{k} z_i^2 \geq 0\ .$$

This shows $\mathbf{B}_l$ is semi-positive definite for $l = 1, 2, \ldots, n-k$, $\sum_{l=1}^{n-k} \mathbf{B}_l$ is also semi-positive definite. Therefore, since $\mathbf{A}$ is clearly positive definite, $\mathbf{G}$ is the summation of a positive and a semi-positive definite matrix. It is positive definite. $\diamondsuit$

In the second and better SQP implementation, we introduce new variables $R_l = \sum_{Z_j > Z_l^*, \delta_j = 1, 1 \leq j \leq k} w_j$; one for each censored observation Z_l. If we identify $\mathbf{x}$ as the vector $(\mathbf{w}, \mathbf{R})$, then the log likelihood function becomes

$$L(\mathbf{x}) = L(w, R) = \sum_{i=1, \delta_i = 1}^{k} \log w_i + \sum_{l=1, \delta_l = 0}^{n-k} \log R_l\ .$$

To find the quadratic approximation of $L(\mathbf{x})$, we need to compute two derivatives. The first derivative with respect to w and R is

$$\frac{\partial L(w, R)}{\partial w_i} = \frac{1}{w_i}\ , \qquad i = 1, 2, \ldots, k,$$

$$\frac{\partial L(w, R)}{\partial R_l} = \frac{1}{R_l}\ , \qquad l = 1, 2, \ldots, n-k.$$

So the vector $\mathbf{a}$ $(n \times 1)$ in the quadratic programming (3) becomes much simpler with entries either equal to $\dfrac{1}{w_i}$ or $\dfrac{1}{R_l}$ depending on the censoring status of the observation. The second derivative of L with respect to w and R is

$$\frac{\partial^2 L(w, R)}{\partial^2 w_i} = -\frac{1}{w_i^2}\ , \qquad \frac{\partial^2 L(w, R)}{\partial^2 R_l} = -\frac{1}{R_l^2}\ , \qquad \frac{\partial^2 L(w, R)}{\partial w_i \partial R_l} = 0\ ,$$

$$i = 1, 2, \ldots, k, \qquad\qquad l = 1, 2, \ldots, n-k.$$

Therefore the matrix $\mathbf{G}$ $(n \times n)$ in the quadratic approximation (3) is diagonal. The i^{th} diagonal element of $\mathbf{G}$ is either $\dfrac{1}{w_i^2}$ or $\dfrac{1}{R_l^2}$ depending on whether this observation is censored or not. Since $\mathbf{G}$ is a diagonal matrix, it is trivial to find the inverse of the decomposition matrix of $\mathbf{G}$, say $\mathbf{H}^{-1}$, such that $\mathbf{H}^T\mathbf{H} = \mathbf{G}$. $\mathbf{H}^{-1}$ is also a diagonal matrix with entries w_i or R_l corresponding to censored status. Many QP solvers including the one in R package `quadprog` can directly use $\mathbf{H}^{-1}$ to calculate the solution much faster than the previous method. Now, because we introduced new variables R_l, they bring $(n-k)$ extra constraints, i.e.

$$(1): \qquad R_1 = \sum_{Z_j > Z_1^*, \delta_j = 1, 1 \le j \le k} w_j \ ,$$

$$\vdots$$

$$(n-k): \qquad R_{n-k} = \sum_{Z_j > Z_{n-k}^*, \delta_j = 1, 1 \le j \le k} w_j \ .$$

These plus the two original constraints (using the original $Z_1 < \cdots < Z_n$)

$$\sum_{i=1}^{n} w_i \delta_i = 1 \ , \qquad \sum_{i=1}^{n} w_i Z_i \delta_i = \mu \ ,$$

would make the constrained matrix $\mathbf{C}$ to be $n \times (n-k+2)$. The first two columns for the above two original constraints will be

$$\begin{pmatrix} \delta_1 & \delta_1 Z_1 \\ \delta_2 & \delta_2 Z_2 \\ \vdots & \vdots \\ \delta_n & \delta_n Z_n \end{pmatrix} .$$

The rest of columns depend of the positions of censored observations. If the observation is censored, the entry is 1. All entries before this observation are 0. The entries after this observation are -1 if uncensored, 0 if censored.

Example: For a concrete example of second QP implementation, suppose there are five ordered observations $\mathbf{Z} = (1, 2, 3, 4, 5)$ and censoring indicators $\delta = (1, 0, 1, 0, 1)$. The weight will be $w = (w_1, 0, w_2, 0, w_3)$ and the probability constraint is that $\sum w_i \delta_i = w_1 + w_2 + w_3 = 1$. Suppose we want to test a null hypothesis $\sum w_i Z_i \delta_i = w_1 + 3w_2 + 5w_3 = \mu$. We have the log likelihood function

$$L(w, R) = \log w_1 + \log w_2 + \log w_3 + \log R_1 + \log R_2 \ ,$$

where $R_1 = w_2 + w_3$ and $R_2 = w_3$. In this case, the relevant vectors and matrixes are:

$$\mathbf{a} = \begin{pmatrix} 1/w_1^\star \\ 1/R_1^\star \\ 1/w_2^\star \\ 1/R_2^\star \\ 1/w_3^\star \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \frac{1}{(w_1^\star)^2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{(R_1^\star)^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{(w_2^\star)^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{(R_2^\star)^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{(w_3^\star)^2} \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 3 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 5 & -1 & -1 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} w_1 - w_1^\star \\ R_1 - R_1^\star \\ w_2 - w_2^\star \\ R_2 - R_2^\star \\ w_3 - w_3^\star \end{pmatrix},$$

where $w^\star$ and $R^\star$ is the current value, and w and R will be the updated values after one QP.

The vector $\mathbf{b}$ will depend the starting value of $\mathbf{w}$. We always use a starting value $\mathbf{w}$ that is a probability. But if the starting value of $\mathbf{w}$ is such that $\sum w_i Z_i \delta_i = \bar{Z} \neq \mu$, then the vector $\mathbf{b_0}$ below is appropriate. However, after one QP iteration the new $\mathbf{w}$ will satisfy $\sum w_i Z_i \delta_i = \mu$ and thus for subsequent QP the vector $\mathbf{b}$ should all be zero.

$$\mathbf{b_0} = \begin{pmatrix} 0, & \mu - \bar{Z}, & 0, & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 0, & 0, & 0, & 0 \end{pmatrix}.$$

The decomposition of the matrix $\mathbf{G}$ is $\mathbf{H}$ and we have:

$$\mathbf{H}^{-1} = \begin{pmatrix} w_1^\star & 0 & 0 & 0 & 0 \\ 0 & R_1^\star & 0 & 0 & 0 \\ 0 & 0 & w_2^\star & 0 & 0 \\ 0 & 0 & 0 & R_2^\star & 0 \\ 0 & 0 & 0 & 0 & w_3^\star \end{pmatrix}.$$

Remark 2. To compare the two methods, we generated random sample of size $n = 100$, where X from $N(1,1)$, C from $N(1.5,2)$. On the same computer, the first method took about $25-30$ minutes to compute the likelihood, however, the second method only took $1-2$ seconds. The difference is remarkable.

Remark 3. The same trick also works for other types of censoring. The key is to introduce some new variables so that the log likelihood function is just $\sum \log x_i$. This, for example, works for the interval censored data, and x_i is the sum of the probabilities located inside the interval of observation i.

4 Likelihood Ratio Computation

The SQP method can be applied in many more complicated cases such as doubly censored, interval censored, and k-sample problems. It is a very convenient and powerful way to find the maximum of log likelihood function under constraints which in turn allows us to compute the empirical likelihood ratio statistic:

$$
\begin{aligned}
-2\log R(H_0) &= -2\log \frac{\max_{H_0} L(w)}{\max_{H_0+H_1} L(w)} && (9)\\
&= 2\left[\log(\max_{H_0+H_1} L(w)) - \log(\max_{H_0} L(w))\right] && (10)\\
&= 2\left[\log(L(\tilde{w})) - \log(L(\hat{w}))\right] \ . && (11)
\end{aligned}
$$

Here $\tilde{w}$ is the NPMLE of probabilities without any constraint, $\hat{w}$ is the NPMLE of probabilities under H_0 constraint. Both NPMLEs can be computed by SQP. In addition, in many cases, there are other methods available to compute $\tilde{w}$, the NPMLE without constraint, but not for the $\hat{w}$. After we obtained the $\tilde{w}$ and $\hat{w}$, Wilks theorem can then be used to compute the P-value of the observed statistic. Thus we can use empirical likelihood ratio to test hypothesis and construct confidence intervals. To illustrate this application, we will show the simulation result for right censored data and give one example for interval censored data.

5. Simulations and Examples

In this section, right-censored data simulation and interval censored example are given to illustrate the application of SQP method.

We have implemented this SQP in R software (Gentleman and Ihaka 1996). The R function `el.cen.test` to do right censored observations with one mean constraint has been packaged as part of `emplik` package and posted on CRAN. It is available inside the package `emplik` at one of the CRAN web site (http://cran.us.r-project.org). Others are available from the author.

5.1 Confidence Interval, real data, right censored

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 data. 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 empirical likelihood ratio to test many null hypothesis that mean is equal to μ (for various values of μ). The 95% confidence interval for the mean survival time is seen to be $[61.708, 144.915]$ since the empirical likelihood ratio test statistic -2LogLikRatio= 3.841 both when $\mu = 61.708$ and $\mu = 144.915$.

The MLE of the mean is 94.7926 which is the integrated Kaplan-Meier estimator. We see that the confidence interval is not symmetric around the MLE, a nice feature of the confidence interval based on the likelihood ratio tests.

5.2 Simulation: right censored data

We randomly generated 5000 right-censored samples, each of size $n = 300$ as in equation (1), where X is taken from $N(1, 1)$, C from $N(1.5, 1)$. Censoring percentage is around $10\% - 20\%$. Software R is used in the implementation. We test the null hypothesis $H_0 : \sum_{i=1}^{n} w_i Z_i \delta_i = \mu = 1$, which is true for our generated data.

We computed 5000 empirical likelihood ratios, using the Kaplan Meier estimator's jumps as $(\tilde{w})$ which maximizes the denominator in (9) and we use SQP method to find $(\hat{w})$ that maximizes the numerator under H_0 constraint. The Q-Q plot based on 5000 empirical likelihood ratios and χ^2_1 percentiles are shown in Figure 1 At the point 3.84 (or 2.71) which is the critical value of χ^2_1 with nominal level 5% (or 10%), if the -2log-likelihood ratio line is above the dashed line ($45°$ line) the probability of rejecting H_0 is greater than 5% (or 10%). Otherwise, the rejecting probability is less than 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.

5.3 Example – Interval Censored Case

As we mentioned before, SQP method can also be used to compute the (constrained) non-parametric MLE with interval censored data. We use the breast cosmetic deterioration data from Gentleman and Geyer (1994) as an example. The data consist of 46 early breast cancer patients who were treated with radiotherapy, but there are only 8 intervals with positive probabilities. We use SQP to compute the probabilities for these 8 intervals under constraint $\sum_{i=1}^{8} X_i p_i = \mu$,

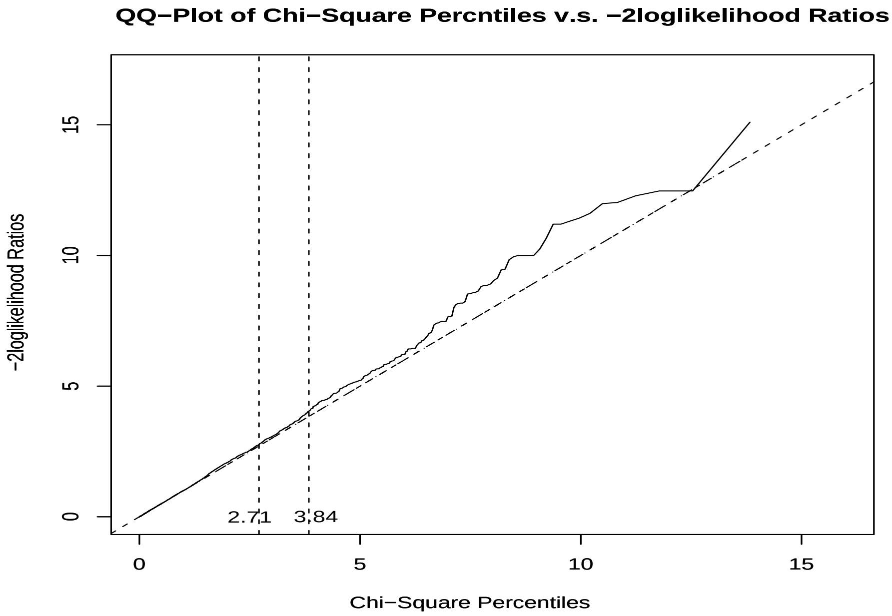

Figure 1: Q-Q Plot of -2*log-likelihood Ratios vs.* $\chi^2_{(1)}$ *Percentiles for Sample Size* $= 300$

where μ is population mean which we want to test, X_i is the midpoint of each interval, p_i is the probability of corresponding interval. If we let $\mu = \sum_{i=1}^{8} X_i p_{0i} = 33.5809$ (i.e. no constraint), where p_{0i} is the probability from the above paper, we get the same probability for each interval. Table 3.1 lists the probabilities for two different constraints.

References

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

Gentleman, R. and Geyer, C. J. (1994), Maximum likelihood for interval censored data: Consistency and computation *Biometrika*, **81, 3**, 618-623.

Goldfarb, D. and Idnani, A (1983), A numerically stable dual method for solving strictly convex quadratic programs *Mathematical Programming*, **27**, 1-33.

left	right	$H_0 : \mu = 33.5809$	$H_0 : \mu = 40$
4	5	0.04634407	0.01954125
6	7	0.03336178	0.01543886
7	8	0.08866270	0.03917190
11	12	0.07075012	0.03524150
24	25	0.09264346	0.05263571
33	34	0.08178547	0.06119782
38	40	0.12087966	0.09192321
46	48	0.46557274	0.68484974
$-2LLR(H_0)$		0	7.782341

Table 3.1: Restricted set of intervals and the associated probabilities

Murphy, S. and van der Vaart, A. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471-1509.

Murphy, S. and van der Vaart, A. (2000). On profile likelihood. *JASA*, **90**, 1399-1405.

Nocedal, J. and Wright, S. (1999). *Numerical Optimization.* Springer, New York.

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.*

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

Department of Statistics
University of Kentucky
Lexington, KY 40506-0027
mai@ms.uky.edu