Taylor & Francis
Taylor & Francis Group

Journal **…Journal of Statistical Computation and Simulation** Article ID **… GSCS189022**

TO: CORRESPONDING AUTHOR

AUTHOR QUERIES - TO BE ANSWERED BY THE AUTHOR

The following queries have arisen during the typesetting of your manuscript. Please answer the queries.

Q1	Please check the affiliation.	
Q2	Please check the edit of the sentence “Veteran's . . ”.	
Q3	Reference 10 is found in the list but not cited in the text. Please check.	

Production Editorial Department, Taylor & Francis
4 Park Square, Milton Park, Abingdon OX14 4RN

Telephone: +44 (0) 1235 828600
Facsimile: +44 (0) 1235 829000

Journal of Statistical Computation and Simulation
Vol. 00, No. 0, Month 2006, 1–10

Computation of the empirical likelihood ratio from censored data

KUN CHEN and MAI ZHOU*

Bayer Pharmaceuticals and University of Kentucky, Kentucky, USA Q1

(Received)

The empirical likelihood ratio method is a general nonparametric inference procedure that has many desirable properties. Recently, the procedure has been generalized to several settings including testing of weighted means with right-censored data. However, the computation of the empirical likelihood ratio with censored data and other complex settings is often nontrivial. We propose to use a sequential quadratic programming (SQP) method to solve the computational problem. We introduce several auxiliary variables so that the computation of SQP is greatly simplified. Examples of the computation with null hypothesis concerning the weighted mean are presented for right- and interval-censored data.

Keywords: Sequential quadratic programming; Maximization; Constraints; Wilks' theorem

AMS 1991 Subject Classification: Primary: 62G10; Secondary: 62G05

1. Introduction

The empirical likelihood ratio method was first proposed by Thomas and Grunkemeier [1]. Owen [2–4] and many others developed this into a general methodology. It has many desirable statistical properties, see Owen's recent book [5]. A crucial step in computing the empirical likelihood ratio, *i.e.* the Wilks statistic, is to find the maximum of the log empirical likelihood (LEL) function under some constraints. The Wilks statistic is just two times the difference of two such LEL functions maximized under different constraints. In all the articles mentioned earlier, this is achieved by using the Lagrange multiplier method. It reduces the maximization of empirical likelihood over $n-1$ variables to solving a set of r equations, $f(\lambda)=0$, for the r-dimensional multiplier λ. The number r is fixed as the sample size n increases. Furthermore, the functions f are monotone in each of the r coordinates. These equations can be easily solved numerically and thus the empirical likelihood ratio can be obtained.

Recently, the empirical likelihood ratio method has been generalized to several more complicated settings. For example, Pan and Zhou [6] showed that for right-censored data, the empirical likelihood ratio can also be used to test hypotheses about a weighted mean. Murphy

*Corresponding author. Email: mai@ms.uky.edu

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10629360600890998

Techset Composition Ltd, Salisbury GSCS189022.TeX Page#: 10 Printed: 17/7/2006

and van der Vaart [7] demonstrated, among other things, that Wilks' theorem for the empirical likelihood ratio also holds for doubly censored data.

However, the computation of the censored data empirical likelihood ratio in these settings remains difficult, as the Lagrange multiplier simplification is not available (see Example 1). Unlike the Owen paper [2], the proofs of Wilks' theorem for the censored data empirical likelihood ratio contained in Pan and Zhou [6] and Murphy and van der Vaart [7] do not offer a viable computational method. They provide existence proofs rather than constructive proofs. Therefore, a study of computational methods that can find the relevant empirical likelihood ratios numerically when analyzing censored data is needed.

Example 1: Suppose i.i.d. observations $X_1, \ldots, X_n$ with an unknown CDF $F_X(t)$ are subject to right censoring, so that we only observe

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

where $C_1, \ldots, C_n$ are censoring times, assumed independent of $X_1, \ldots, X_n$.

The LEL function based on the censored observations (Z_i, δ_i) is

$$\mathrm{LEL}(\mathbf{w}) = \sum_{i=1}^{n} \left[\delta_i \log w_i + (1 - \delta_i) \log \left(\sum_{Z_j > Z_i} w_j \right) \right], \tag{2}$$

where $w_i = F_X(Z_i) - F_X(Z_i-)$.

The empirical likelihood ratio test is based on the Wilks statistic

$$\begin{aligned} -2\ \log\ R(H_0) &= -2 \log \frac{\max_{H_0}\ \mathrm{EL}(\mathbf{w})}{\max_{H_0+H_1} \mathrm{EL}(\mathbf{w})} \\ &= 2\big[\log\big(\max_{H_0+H_1}\ \mathrm{EL}(\mathbf{w})\big) - \log\big(\max_{H_0} \mathrm{EL}(\mathbf{w})\big)\big] \\ &= 2[\log(L(\tilde{\mathbf{w}})) - \log(L(\hat{\mathbf{w}}))] = 2[\mathrm{LEL}(\hat{\mathbf{w}}) - \mathrm{LEL}(\hat{\mathbf{w}})]. \end{aligned}$$

Here, $\tilde{\mathbf{w}}$ is the nonparametric maximum likelihood estimate (NPMLE) of probabilities without any constraint and $\hat{\mathbf{w}}$ is the NPMLE of probabilities under the H_0 constraint.

To compute Wilks' statistic for testing a hypothesis about a weighted mean of X, we need to find the maximum of the above LEL under the constraints

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

where μ is a given constant, specified by the null hypothesis. Although the asymptotic null distribution of the test statistic can be shown to be chi-squared with one degree of freedom, a straight application of the Lagrange multiplier method does not lead to a simple solution. The same difficulty arises also with the doubly censored data and other censoring cases. Thus, a viable computation algorithm for the maximization of the empirical likelihood ratio is needed.

We propose to use the sequential quadratic programming (SQP) method to find the constrained maximum. In particular, we show how one can introduce several auxiliary variables so that the computation of SQP for censored empirical likelihood is greatly simplified. In fact, this trick can be used to compute empirical likelihood ratios in many other cases (for example, doubly or interval-censored data), where a simple Lagrange multiplier computation is not available.

We briefly review the SQP method in section 2. We show how to use the SQP method to compute the maximum of the LEL function in section 3. Examples and simulations are given in section 4.

2. SQP method

There is a large amount of literature on nonlinear programming methods [see ref. 8 and references there in]. The general strictly convex (positive definite) quadratic programming problem is to minimize

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

subject to

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

where $\mathbf{x}$ and $\mathbf{a}$ are n-vectors, $\mathbf{G}$ the $n \times n$ symmetric positive definite matrix, $\mathbf{C}$ the $n \times m$ $(m < n)$ matrix, and $\mathbf{b}$ the m-vector and the superscript T denotes the transpose. In this article, the vector $\mathbf{x}$ is only subject to equality constraints $\mathbf{C}^{\mathrm{T}}\mathbf{x} - \mathbf{b} = \mathbf{0}$. This makes the QP problem easier. In the next section, we shall show how to introduce a few new variables in the maximization of the censored LEL (2) so that the matrix $\mathbf{G}$ is always diagonal, which further simplifies the computation. Therefore, instead of using a general QP algorithm, we have implemented our own version in R that takes advantage of the mentioned simplifications. The specific QP problem can be solved by performing one matrix QR decomposition, one backward solve, and one forward solve of equations.

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

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

where η is a column vector of length m. Taking the derivative with respect to $\mathbf{x}$ and setting 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

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

As the matrix $\mathbf{G}$ is diagonal, the inverse $\mathbf{G}^{-1}$ is easy to obtain. Finally, we need to solve for η. Substituting (5) into $\mathbf{C}^{\mathrm{T}}\mathbf{x} = \mathbf{b}$, we get $\mathbf{C}^{\mathrm{T}}(\mathbf{G}^{-1}[\mathbf{a} + \mathbf{C}\eta]) = \mathbf{b}$, which is, upon rewriting,

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

Once we get the solution to η from equation (6), we can substitute it back into equation (5) to calculate $\mathbf{x}$.

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

$$\begin{aligned} (\mathbf{R}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}})\eta &= \mathbf{b} - \mathbf{R}\mathbf{Q}\mathbf{G}^{-1/2}\mathbf{a}, \\ (\mathbf{R}\mathbf{R}^{\mathrm{T}})\eta &= \mathbf{b} - \mathbf{R}\mathbf{Q}\mathbf{G}^{-1/2}\mathbf{a}, \\ \mathbf{R}^{\mathrm{T}}\eta &= \mathbf{R}^{-1}\mathbf{b} - \mathbf{Q}\mathbf{G}^{-1/2}\mathbf{a}. \end{aligned} \tag{7}$$

Equation (7) can be solved by using back-substitution (twice) and matrix–vector multiplication (once), which are low cost operations.

We are interested in maximizing the LEL or minimizing the negative LEL over all possible probabilities. This is a nonlinear programming problem. As it is hard to find a minimum of the negative LEL directly in many cases, and the negative LEL is often convex at least near the minimum, we use a quadratic function to approximate it. Starting from an initial probability

$\mathbf{w}^0$, we replace the nonlinear target function (negative LEL) with a quadratic function that has the same first and second derivatives at $\mathbf{w}^0$. The QP method is used to find the minimum of the quadratic function subject to the same constraints. Denote the location of the minimum by $\mathbf{w}^1$. Then, we update the quadratic approximation which now has the same first and second derivatives as the negative LEL at $\mathbf{w}^1$. The QP method is used again to find the minimum of the new quadratic function under the same constraints. Iteration ends when a predefined convergence criterion is satisfied. The convergence criterion can be based on the values of the negative LEL, which should decrease at each iteration. When the value of the negative LEL no longer decreases, we stop the iteration.

One way to improve convergence and guarantee that the negative LEL decreases at each iteration is the technique of damping: write the updated value of the solution as $\mathbf{x}^{(s)} = \mathbf{x}^{(s-1)} + \mathbf{x}$, we shall only accept $\mathbf{x}^{(s)}$ if it decreases the negative LEL, otherwise we shall search along the line $\mathbf{x}_\xi^{(s)} = \mathbf{x}^{(s-1)} + \xi\mathbf{x}$ for $0 \le \xi < 1$ until it decreases the negative LEL value.

When the information matrix (of the LEL) is positive, the quadratic approximation is good at least in a neighbourhood of the true MLE. Thus, in the case of convergence, the solution gives the correct MLE under the given constraints.

3. Empirical likelihood maximization with right-censored data

We now describe the SQP method that solves the problem in Example 1. The implementations for doubly censored data and interval-censored data are similar. We only give the details of the right-censored data here.

For the right-censored data as in equation (1), the LEL is given in equation (2). It is well known that the maximizer of the LEL has the following property: $w_i > 0$ only when the corresponding $\delta_i = 1$. We shall restrict the search of a maximizer for the LEL under the mean constraint to those w_i's. See Owen [2, p. 238] for a discussion on this type of restriction.

We describe below two ways to implement the SQP method for finding the constrained MLE.

The first implementation of QP is to take $\mathbf{w}$ in equation (2) as $\mathbf{x}$. The knowledge of $w_i = 0$ when $\delta_i = 0$ helps to reduce the number of variables to k (the number of uncensored data). The length of the vector $\mathbf{a}$ is k and the matrix $\mathbf{G}$ 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 very expensive.

The second and better method of using the SQP with censored data is to introduce some auxiliary variables $R_l = P(Z \ge Z_l)$, one for each censored observation; this enlarges the dimension of the vectors ($\mathbf{a}$, $\mathbf{x}$, $\mathbf{b}$) and the matrices ($\mathbf{G}$, $\mathbf{C}$) in equations (3) and (4), but simplifies the matrix $\mathbf{G}$. In fact, $\mathbf{G}$ will be diagonal, so that we can directly plug in the inverse of the decomposition matrix of $\mathbf{G}$. This speeds up the computation tremendously.

We illustrate the two methods for the problem described in Example 1. In the first method, as $w_i > 0$ 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 the LEL function is:

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

Let us denote $M_l = \sum_{Z_j > Z_l^*, \delta_j = 1, 1 \le j \le k} w_j$, then the **a** vector in the QP problem (3) will be

$$\mathbf{a} = \left(\frac{1}{w_1} + \sum_{l=1}^{n-k} \frac{I_{[Z_1 > Z_l^*]}}{M_l}, \quad \frac{1}{w_2} + \sum_{l=1}^{n-k} \frac{I_{[Z_2 > Z_l^*]}}{M_l}, \ldots, \frac{1}{w_k} + \sum_{l=1}^{n-k} \frac{I_{[Z_k > Z_l^*]}}{M_l} \right)^{\mathrm{T}}.$$

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

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

and for $i \neq q$:

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

and, therefore, the matrix **G** is given by the negative of those second derivatives.

Finally,

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

We always use an initial value $\mathbf{w}_0$ that is a probability, but it may not satisfy the mean constraint. Therefore, $\mathbf{b}_0 = (0, \mu - \bar{Z})$, where $\bar{Z} = \sum w_{0i} Z_i$. For subsequent iterations, we have $\mathbf{b} = (0, 0)$, as the current value of **w** already satisfies both constraints.

In the second and better SQP implementation, we introduce new variables

$$R_l = R(Z_l) = \sum_{Z_j > Z_l, \delta_j = 1, 1 \le j \le k} w_j,$$

one for each right-censored observation Z_l. If we identify **x** in equation (3) as the vector $(\mathbf{w}, \mathbf{R})$, then the LEL function (2) becomes

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

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

$$\frac{\partial \mathrm{LEL}(w, R)}{\partial w_i} = \frac{1}{w_i}, \quad i = 1, 2, \ldots, k,$$
$$\frac{\partial \mathrm{LEL}(w, R)}{\partial R_l} = \frac{1}{R_l} \quad l = 1, 2, \ldots, n-k.$$

Therefore, the vector $\mathbf{a}$ ($n \times 1$) in the quadratic programming problem (3) becomes much simpler with entries equal to either $1/w_i$ or $1/R_l$, depending on the censoring status of the observation. The second derivatives of L with respect to $(\mathbf{w}, \mathbf{R})$ are

$$\frac{\partial \mathrm{LEL}(w,R)}{(\partial w_i)^2} = -\frac{1}{w_i^2}, \quad \frac{\partial^2 \mathrm{LEL}(w,R)}{(\partial R_l)^2} = -\frac{1}{R_l^2}, \quad \frac{\partial^2 \mathrm{LEL}(w,R)}{\partial w_i \partial R_l} = 0,$$
$$i = 1, 2, \ldots, k, \quad l = 1, 2, \ldots, n-k.$$

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

$$(1): \quad R_1 = \sum_{Z_j > Z_1^*, \delta_j = 1, 1 \le j \le k} w_j,$$
$$\vdots$$
$$(n-k): \quad R_{n-k} = \sum_{Z_j > Z_{n-k}^*, \delta_j = 1, 1 \le j \le k} w_j.$$

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

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

would make the constraint matrix $\mathbf{C}$ to be of size $n \times (n-k+2)$. The first two columns of $\mathbf{C}$ 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 the columns depends on the positions of the 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 and 0 if censored.

Example 2: 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 vector 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 that we want to test a null hypothesis $\sum w_i Z_i \delta_i = w_1 + 3w_2 + 5w_3 = \mu$. We have the LEL function

$$\mathrm{LEL}(w, R) = \log\ w_1 + \log\ w_2 + \log\ w_3 + \log\ R_1 + \log\ R_2,$$

where $R_1 = w_2 + w_3$ and $R_2 = w_3$. In this case, the relevant vectors and matrices 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 $R^\star$ are the current values and w and R will be the updated values after one QP.

The vector $\mathbf{b}_0$ will depend on the starting value $\mathbf{w}_0$. We always use a starting value $\mathbf{w}_0$ that is a probability, but it may not satisfy the weighted mean constraint. 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 be zero. Suppose we take $\mathbf{w}_0$ to be the discrete uniform probability, then

$$\mathbf{b}_0 = \left(0, \quad \mu - \bar{Z}, \quad 0, \quad 0\right), \quad \mathbf{b} = \left(0, \quad 0, \quad 0, \quad 0\right).$$

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 1 To compare the two methods, we generated a random sample of size $n = 100$, where X is taken from $N(1, 1)$ and C from $N(1.5, 2)$. On the same computer, the first method took about 25–30 min to find the maximum of the likelihood, whereas, the second method only took 1–2 s. The difference is remarkable. The computation took about five iterations of QP. Of course, this comparison is very much hardware-dependent, but it at least is an indication of what could happen.

Remark 2 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 interval-censored data where for an interval-censored observation, the log-likelihood term is $\log\ x_i$, and x_i now equals the sum of the probabilities located inside the interval of observation i.

4. Empirical likelihood ratio computation

The SQP method is a very powerful method to find the maximizer of an LEL function under constraints, which, in turn, allows us to compute the empirical likelihood ratio statistic. After we obtain $\tilde{w}$ and $\hat{w}$, Wilks' theorem can then be used to compute the P-value of the observed statistic. Thus, we can use the empirical likelihood ratio to test hypotheses and construct confidence intervals.

We have implemented this SQP in the R software [9]. The R function, `el.cen.test`, that computes the empirical likelihood ratio for right-censored observations with one mean constraint has been packaged as part of the emplik package and posted on CRAN (http://cran.us.r-project.org). Our implementation of QP in R uses the R functions `backsolve()`, `qr()`, which, in turn, call the corresponding LINPACK routines.

To illustrate the application, we will show the simulation results for right-censored data and give one example for interval-censored data.

4.1 *Confidence interval, real data, right-censored*

Q2 Veteran's Administration Lung cancer study data are available from the R package `survival`. We took the subset of survival data for treatment 1 and the small cell 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 used the empirical likelihood ratio to test the null hypothesis that the 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], as the empirical likelihood ratio test statistic $-2\,\mathrm{LogLikRatio} = 3.841$ 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, and this is typical for confidence intervals based on the likelihood ratio tests.

4.2 *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 $N(1, 1)$ and C from $N(1.5, 1)$. Censoring percentage is around 10–20%. The software R is used in the implementation. We tested 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 equation (9), and used the SQP method to find $(\hat{w})$, which maximizes the numerator under the H_0 constraint. The Q–Q plot based on 5000 empirical likelihood ratios and χ_1^2 percentiles is shown in figure 1. At point 3.84 (or 2.71), which is the critical value of χ_1^2 with nominal level 5% (or 10%), if the -2 log-likelihood ratio line is above the dashed line (45° line), the probability of rejecting H_0 is >5% (or 10%). Otherwise, the rejection probability is <5% (or 10%). From the Q–Q plot, we can see that the χ_1^2 approximation is pretty good. Only at the tail of the plot, the differences between the percentiles of -2 log-likelihood ratios and χ_1^2 are getting bigger.

4.3 *Example – interval-censored case*

As mentioned earlier, the SQP method can also be used to compute the (constrained) NPMLE with interval-censored data. We used the breast cosmetic deterioration data from ref. [9]

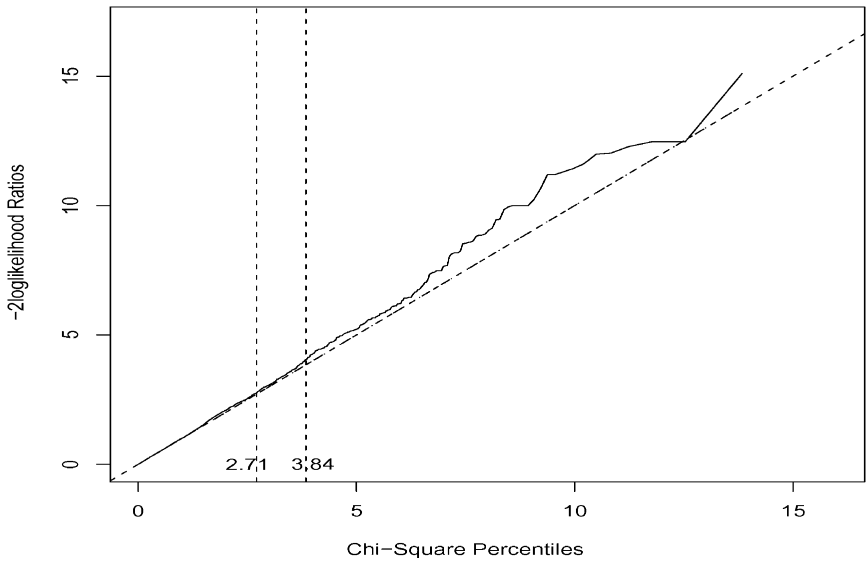

Figure 1. Q-Q plot of -2 log-likelihood ratios versus $\chi^2_{(1)}$ percentiles for sample size = 300.

Table 1. Restricted set of intervals and the associated probabilities.

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
$-2\text{LLR}(H_0)$		0	7.782341

as an example. The data consist of 46 early breast cancer patients, who were treated with radiotherapy, but there are only eight intervals with positive probabilities. We used SQP to compute the probabilities for these eight intervals under the constraint $\sum_{i=1}^{8} X_i p_i = \mu$, where μ is the population mean which we want to test, X_i is the midpoint of each interval, and p_i is the probability of the corresponding interval. Table 1 lists the probabilities for two different constraints. The mean of the unconstrained NPMLE is 33.5809; therefore, the hypothesis H_0: $\mu = 33.5809$ is equivalent to imposing no constraint and the P-value is 1.

5. Discussion

One drawback of the SQP method is that it becomes more memory/computationally intensive for larger sample sizes. The cost increases at the rate of n^2. This is in contrast to the Lagrange multiplier method mentioned earlier, where (when available) r remains fixed as the sample size n increases. However, we argue that this is not a major drawback for SQP because (1) the advantages of the empirical likelihood ratio method are most pronounced for small-to-medium sample sizes. Often for large samples, there are alternative, equally effective, and easily computable statistical methods available, as, for example, the Wald method. (2) By our implementation of the SQP method in R, we can easily handle sample sizes of up to 2000 on today's average PC (20 s on a 3 GHz, 512 MB PC). With computer hardware getting cheaper, this drawback should diminish and not pose a major handicap for the SQP method for most applications.

Of course, not all constrained maximization problems have a solution. If the H_0 constraint is too faraway from the sample mean, this may well happen. See ref. [2, p. 238] for further discussion. When this happens, we should define the likelihood ratio to be zero, implying that this is an impossible H_0.

There may be simpler methods available to compute $\tilde{w}$, the NPMLE without constraint. In the case of Example 1, this is the well-known Kaplan–Meier estimator.

Acknowledgement

I would like to thank Arne Bathke and an anonymous referee for careful reading of this article and many suggestions that led to a clearer presentation.

References

[1] Thomas, D.R. and Grunkemeier, G.L., 1975, Confidence interval estimation of survival probabilities for censored data. *American Statistical Association*, **70**, 865–871.
[2] Owen, A., 1988, Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.
[3] Owen, A., 1990, Empirical likelihood ratio confidence regions. The *Annals of Statistics*, **18**, 90–120.
[4] Owen, A., 1991, Empirical likelihood for linear models. *The Annals of Statistics*, **19**, 1725–1747.
[5] Owen, A., 2001, *Empirical Likelihood* (London: Chapman & Hall).
[6] Pan, X.R. and Zhou, M., 1999, Using one parameter sub-family of distributions in empirical likelihood with censored data. *Journal of Statistical Planning and Inference*, **75**, 379–392.
[7] Murphy, S. and van der Vaart, A., 1997, Semiparametric likelihood ratio inference. The *Annals of Statistics*, **25**, 1471–1509.
[8] Nocedal, J. and Wright, S., 1999, *Numerical Optimization* (New York: Springer).
[9] Gentleman, R. and Ihaka, R., 1996, R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.
Q3 [10] Goldfarb, D. and Idnani, A., 1983, A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, **27**, 1–33.