Journal of the Korean Statistical Society (1999), **28**: *2, pp 195–210*

Estimating Methods on Exponential Regression Models with Censored Data†

Il-Do Ha[1], Youngjo Lee[2] and Jae-Kee Song[3]

ABSTRACT

We consider a large class of exponential regression models with censored data and propose two modified Fisher scoring methods with corresponding algorithms. These proposed methods improve the Newton-Raphson method in estimating the model parameters. The simulated and real examples are illustrated in aspect of convergence.

Keywords: Censoring mechanism; Fisher scoring method; Genearized linear models; Link function; Quasi-likelihood.

1. INTRODUCTION

To the analysis of censored survival data on patients suffering from chronic disease (e.g., cancer), many authors previously have been treated the exponential regression models that a chosen function for the expected survival time of each patient is linearly related to covariates of a patient.

In fact, the function corresponds to a link function in generalized linear models (GLMs, Nelder and Wedderburn, 1972). For these exponential models, the most useful link is log, which has been discussed by Glasser (1967), Prentice (1973) and others. Other frequently used link is identity, e.g., Zippen and Armitage(1966) and Krall et al. (1975), and reciprocal link, e.g., Greenberg et al. (1974).

To analyze such data, we need to solve the likelihood equations of these exponential models. But, since the equations are nonlinear functions for the vector of the model parameters, these can be solved by iterative methods, such as the Newton-Raphson (N-R) method and the Fisher scoring (FS) method, etc. When

†This paper was partially supported by the Basic Science Research Institute Programs, Ministry of Education, 1998, Project No. BSRI-98-1403.

[1]Department of Statistics, Kyungsan University, Kyungpook, 712-240, Korea.
[2]Department of Statistics, Seoul National University, Seoul, 151-742, Korea.
[3]Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

the data may be censored, it does not be possible to figure out the Fisher information needed in the FS method, and so the likelihood equations are usually solved by the N-R method based on the observed information.

However, the observed information usually gives more complicated form. Moreover, the observed information often may not be positive definite at some points in the admissible regression parameter space for these exponential models. Therefore, the N-R method may not be converged if it has not the initial estimates near to optimal points of the regression parameters.

The above convergence problem in the implementation of the N-R method can be circumvented by using the modified FS method that the observed information is replaced by an alternative of the Fisher information.

The main purpose of this paper is to consider a large class of exponential regression models with censored data and then to propose two modified FS methods estimating the model parameters, and also to provide the corresponding algorithms.

This paper is organized as follows. In Section 2, the models considered are defined and the likelihood equations for the models are derived, and in Section 3 and Section 4, two modified FS methods are proposed and the corresponding algorithms including an useful initial estimate are also provided. In Section 5, for the models with three useful links (log, reciprocal and identity), two proposed FS methods and the N-R method are summarized and compared. In Section 6, our results are illustrated with the simulated and real examples. Finally, some concluding remarks are given in Section 7.

2. MODEL STRUCTURES AND LIKELIHOOD EQUATIONS

Let T_i be the survival time on study for the ith $(i = 1, 2, \cdots, n)$ individual (patient or subject) and C_i be the random censoring time associated with T_i. However T_i's may not all be observable due to the censoring mechanism, i.e., the observable quantities are

$$Y_i = \min(T_i, C_i), \quad \delta_i = I(T_i \leq C_i), \quad x_i^t = (x_{i1}, x_{i2}, \cdots, x_{ip}),$$

where $I(\cdot)$ is the indicator function and x_i^t is the $1 \times p$ vector of covariates associated with the ith individual.

Assume that T_i $(i = 1, 2, \cdots, n)$ are n independent survival times such that each T_i follows the exponential distribution with the mean $\mu_i = E(T_i)(> 0)$, which is dependent on the covariates x_i^t of the ith individual and that C_i $(i = 1, 2, \cdots, n)$

are n independent censoring times with a continuous distribution. Also assume that T_i and C_i $(i = 1, 2, \cdots, n)$ are independent.

Under the assumptions, we consider here the exponential regression model

$$\eta_i = g(\mu_i) = x_i^t\beta \quad (i = 1, 2, \cdots, n), \tag{2.1}$$

where η_i and $g(\cdot)$ are linear predictor and link function in the GLMs, respectively, and β is the $p \times 1$ vector of unknown model parameters.

In addition, the model (2.1) can be viewed as a location-scale model for $\log(T_i)$ $(i = 1, 2, \cdots, n)$, i.e.,

$$\log(T_i) = \log \mu_i + \epsilon_i \ (i = 1, 2, \cdots, n),$$

where $\mu_i = g^{-1}(x_i^t\beta)$ and the ϵ_i's follow independent standard extreme value distribution, see Lawless (1982, pp. 283).

Based on the observations (y_i, δ_i, x_i^t) for $i = 1, 2, , n$, the log-likelihood of β in the model (2.1) is

$$\ell(\beta) = \sum_{i=1}^{n} \ell_i,$$

where

$$\ell_i = \ell(\beta; y_i, \delta_i, x_i^t) = -\delta_i \ \log(\mu_i) - \frac{y_i}{\mu_i}.$$

To obtain the likelihood equations for β, we require an expression of $\partial\ell_i/\partial\beta_j$ for $j = 1, 2, \cdots, p$. Now, by the chain rule,

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}. \tag{2.2}$$

By combining the above equation for ℓ_i with (2.1) and (2.2), we have

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{(y_i - \delta_i\mu_i)x_{ij}}{\mu_i^2}(\frac{1}{g'(\mu_i)}), \quad j = 1, 2, \cdots, p, \tag{2.3}$$

where $g'(\mu_i) = \partial\eta_i/\partial\mu_i$. Then the likelihood equations for β are given by

$$\ell_j{}'(\beta) = \sum_{i=1}^{n} \frac{(y_i - \delta_i\mu_i)x_{ij}}{\mu_i^2}(\frac{1}{g'(\mu_i)}) = 0, \quad j = 1, 2, \cdots, p, \tag{2.4}$$

where $\ell_j{}'(\beta)$ is denoted by $\sum_{i=1}^{n} \partial\ell_i/\partial\beta_j$. Since the likelihood equations (2.4) are functions of the β_j's only through the mean function μ_i's in nonlinear fashion, the maximum likelihood (ML) estimator $\hat{\beta}$ of β is obtained by iterative methods, such as the N-R method and the FS method, etc. The p-dimensional score function can be written as

$$\ell'(\beta) = X^t W u, \tag{2.5}$$

where $\ell'(\beta)$ is the $p \times 1$ vector with the jth element $\ell_j{}'(\beta)$, X is the $n \times p$ model matrix whose ith row vector is x_i^t, W is the $n \times n$ diagonal weight matrix with the ith element $w_i = \{\mu_i^2[g'(\mu_i)]^2\}^{-1}$ and u is the $n \times 1$ vector with the ith element $u_i = (y_i - \delta_i \mu_i) g'(\mu_i)$.

3. FISHER SCORING-I METHOD

To solve the likelihood equations (2.4) via the N-R method, we need the negative second derivatives $-\ell''(\beta)$ of $\ell(\beta)$. Now, from (2.2) we have

$$\frac{-\partial^2 \ell_i}{\partial\beta_j \partial\beta_k} = \frac{-\partial^2 \ell_i}{\partial\mu_i^2}\left(\frac{\partial\mu_i}{\partial\beta_j}\frac{\partial\mu_i}{\partial\beta_k}\right) - \left(\frac{\partial\ell_i}{\partial\mu_i}\right)\left(\frac{\partial^2\mu_i}{\partial\beta_j\partial\beta_k}\right).$$

Thus,

$$-\ell''_{jk}(\beta) = \sum_{i=1}^{n} x_{ij} w_i^* x_{ik}, \quad j,k = 1,2,\cdots,p\ , \tag{3.1}$$

where $-\ell''_{jk}(\beta)$ is denoted by $\sum_{i=1}^{n} \frac{-\partial^2 \ell_i}{\partial\beta_j\partial\beta_k}$ and

$$w_i^* = -w_i\{(\delta_i - \frac{2y_i}{\mu_i}) - (y_i - \delta_i\mu_i)\frac{g''(\mu_i)}{g'(\mu_i)}\}. \tag{3.2}$$

In matrix notation, the equations (3.1) can be written as

$$-\ell''(\beta) = X^t W^* X, \tag{3.3}$$

where $-\ell''(\beta)$ is the $p \times p$ matrix with the (j,k)th element $-\ell_{jk}{}''(\beta)$ and W^* is the $n \times n$ diagonal weight matrix with the ith element w_i^*.

Here, (3.3) is positive definite if X has full column rank and W^* has positive elements on the main diagonal. Although X has full column rank the diagonal elements w_i^*'s in (3.2) are not easy to satisfy positive for all i, because they excessively depend on the data, the parameters and the link function. Further, since the observed information matrix in (3.3) may not be positive definite at some points in the admissible space for β and the expected value of Y_i in (3.3) can't be simply calculated due to censoring mechanism, both the N-R method and the FS method may not be applied directly.

As an alternative, we propose the modified FS method using an estimator of the Fisher information (hereafter we call it FS-I method) as follows:

Proposition 3.1. *Under the model (2.1) and regular conditions, the* $p \times p$ *Fisher information matrix* $I(\beta) = (I_{jk}(\beta))$ *is reduced to*

$$I_{jk}(\beta) = E(\sum_{i \in D} x_{ij} w_i x_{ik}), \quad j,k = 1,2,\cdots,p, \tag{3.4}$$

where D *is the index set of individuals for uncesored observations.*

Proof: Under the model (2.1) and regular conditions, we know that

$$E(\frac{\partial \ell_i}{\partial \beta_j}) = 0, \ j = 1,2,\cdots,p. \tag{3.5}$$

Therefore, from (2.3) and (3.5) we have

$$E(Y_i) = \mu_i E(\delta_i) \ \ (i = 1,2,\cdots,n). \tag{3.6}$$

Also, from (3.2) and (3.6) we obtain that

$$E(w_i^*) = w_i E(\delta_i) \ \ (i = 1,2,\cdots,n). \tag{3.7}$$

Combining the expectation of (3.1) with (3.7) completes the proof. □

Since $E(w_i^*) = w_i$ in the case of uncensoring, the Fisher information (3.4) can be easily calculated, i.e., $I_{jk}(\beta) = \sum_{i=1}^{n} x_{ij} w_i x_{ik}, \ \ j,k = 1,2,\cdots,p$. However when there is censoring, the Fisher information is not easily computed due to random sum in (3.4).

Therefore we propose a reasonable estimator of $I_{jk}(\beta)$ as follows:

$$\hat{I}_{jk}(\hat{\beta}) = \sum_{i \in D} x_{ij} \hat{w}_i x_{ik}, \quad j,k = 1,2,\cdots,p, \tag{3.8}$$

where $\hat{w}_i (= w_i(\hat{\beta}))$ for $i \in D$ is the value of the $w_i (= w_i(\beta))$ for $i \in D$ evaluated at $\beta = \hat{\beta}$. In matrix notation, (3.8) can be written as

$$\hat{I}(\hat{\beta}) = X_D^t \hat{W}_D X_D, \tag{3.9}$$

where $\hat{I}(\hat{\beta})$ is the $p \times p$ matrix with the (j,k)th element $\hat{I}_{jk}(\hat{\beta})$, X_D is the $r \times p$ matrix with ith row vector x_i^t for $i \in D$, r is the number of uncensoring and $\hat{W}_D$ is the $r \times r$ diagonal weight mairix with the ith element $\hat{w}_i$ for $i \in D$. Note that (3.9) is positive definite if X_D has full column rank and W_D has positive elements on the main diagonal. Here the diagonal elements of W_D are always positive, except trivial case (e.g., $\beta = 0$).

We here provide the following FS-I algorithm for obtaining the ML estimator $\hat{\beta}$ via the proposed FS-I method:

Step 0 : Obtain an initial estimate $\hat{\beta}^{(0)}$ of β.

Step 1 : Calculate the following with a chosen link function $g(\cdot)$

$$\hat{\mu}_i^{(0)} = g^{-1}(x_i^t \hat{\beta}^{(0)}) \quad (i = 1, 2, \cdots, n).$$

Step 2 : Calculate $\ell'(\hat{\beta}^{(0)})$ and $\hat{I}(\hat{\beta}^{(0)})$,
where $\ell'(\hat{\beta}^{(0)})$ is the score vector $\ell'(\beta)$ in (2.5) and $\hat{I}(\hat{\beta}^{(0)})$ is the estimate $\hat{I}(\beta)$ of the Fisher information in (3.9), evaluated at $\beta = \hat{\beta}^{(0)}$, respectively.

Step 3 : Calculate the next approximation $\hat{\beta}^{(1)}$ to β, as

$$\hat{\beta}^{(1)} = \hat{\beta}^{(0)} + [\hat{I}(\hat{\beta}^{(0)})]^{-1} \ell'(\hat{\beta}^{(0)}).$$

Step 4 : Repeat Step 1 to Step 3 by replacing $\hat{\beta}^{(0)}$ with $\hat{\beta}^{(1)}$ until convergence is hopefully achieved. One stop if $\hat{\beta}^{(0)}$ and $\hat{\beta}^{(1)}$ are close together and $\ell'(\hat{\beta}^{(1)})$ is close to zero.

For obtaining an initial estimate $\hat{\beta}^{(0)}$ of the FS-I algorithm, we use the initial estimate of β in the GLMs setting based on the data with regarding censored one as uncensored, see Nelder and Wedderburn (1972). That is,

$$\hat{\beta}^{(0)} = (X^t W^{(0)} X)^{-1} X^t W^{(0)} z^{(0)}, \tag{3.10}$$

where $W^{(0)}$ is the $n \times n$ diagonal matrix W with the ith element w_i in (2.5) evaluated at $\mu_i = y_i$ and $z^{(0)}$ is the $n \times 1$ adjusted dependent vector with the ith element $z_i^{(0)} (= g(\mu_i))$ evaluated at $\mu_i = y_i$. Note that the N-R method uses $-\ell''(\hat{\beta}^{(0)})$ instead of $\hat{I}(\hat{\beta}^{(0)})$ in the Step 3.

From the asymtotic normality of $\hat{\beta}$ under regular conditions and (3.9), we can obtain the estimated $p \times p$ asymtotic covariance matrix of $\hat{\beta}$, i.e.,

$$\widehat{\text{Cov}}(\hat{\beta}) = [\hat{I}(\hat{\beta})]^{-1} = [X_D^t \hat{W}_D X_D]^{-1}, \tag{3.11}$$

and so the standard errors (SEs) of $\hat{\beta}_j$ $(j = 1, 2, \cdots, p)$ are given by the square root of the (j, j)th diagonal element of (3.11).

4. FISHER SCORING-II METHOD

The FS-I method proposed in Section 3 uses uncensored observations only in estimating the Fisher information $I_{jk}(\beta)$. Thus the FS-I method may be improved by using full observations.

We consider the pseudo random variables

$$Y_i^* = Y_i\delta_i + E(T_i \mid T_i > Y_i)(1-\delta_i) \quad (i = 1, 2, \cdots, n), \tag{4.1}$$

where T_i and Y_i are survival time and observed time, respectively. Then

$$E(Y_i^*) = \mu_i \quad (i = 1, 2, \cdots, n), \tag{4.2}$$

where $\mu_i = E(T_i)$. Note that the expectation identity (4.2) that provides validity for estimation in the linear models with censored data was shown by Buckley and James (1979).

Since T_i has the exponential distribution with mean μ_i for $i = 1, 2, \cdots, n$, (4.1) becomes

$$Y_i^* = Y_i\delta_i + (Y_i + \mu_i)(1-\delta_i) \quad (i = 1, 2, \cdots, n). \tag{4.3}$$

We then propose the modified FS method using Y_i^*'s and Wedderburn's (1974) quasi-likelihood (QL) approach (hereafter we call it FS-II method) as follows:

Proposition 4.1. *Suppose that under the model (2.1), the variance structure of* Y_i^* *is specified by* $\mathrm{Var}(Y_i^*) = \phi\mu_i^2$, *where* $\mu_i = E(Y_i^*)$ *and* ϕ *is a dispersion parameter.*

Then

(a) the likelihood equations (2.4) is reduced to the form of the QL equations

$$q_j'(\beta) = \sum_{i=1}^{n} \frac{(y_i^* - \mu_i)x_{ij}}{\phi\mu_i^2}\left(\frac{1}{g'(\mu_i)}\right) = 0, \quad j = 1, 2, \cdots, p, \tag{4.4}$$

where $q_j'(\beta)$ *is denoted by* $\partial q(\beta)/\partial\beta_j$ *and* $q(\beta)$ *by the quasi-log-likelihood of* β *based on* $y_1^*, y_2^*, \cdots, y_n^*$,

(b) the $p\times p$ *quasi-Fisher information (or working Fisher information) matrix is given by*

$$F(\beta) = (X^tWX)/\phi, \tag{4.5}$$

where $F(\beta)$ *is denoted by* $\mathrm{Cov}(q'(\beta))$ *and* $q'(\beta)$ *is the* $p \times 1$ *vector with the* j*th element* $q_j'(\beta)$.

Proof: Let C be the index set of individuals for censored observations. Since the likelihood equations (2.4) can be rewritten as

$$\ell_j'(\beta) = \sum_{i \in D} \frac{(y_i - \mu_i)x_{ij}}{\mu_i^2}\left(\frac{1}{g'(\mu_i)}\right) + \sum_{i \in C} \frac{y_i x_{ij}}{\mu_i^2}\left(\frac{1}{g'(\mu_i)}\right) = 0, \quad j = 1, 2, \cdots, p, \tag{4.6}$$

from (4.3) we have

$$\ell_j'(\beta) = \sum_{i=1}^{n} \frac{(y_i^* - \mu_i)x_{ij}}{\mu_i^2}\left(\frac{1}{g'(\mu_i)}\right) = 0, \quad j = 1, 2, \cdots, p. \tag{4.7}$$

From the variance structure, the equations (4.7) is of the form of the QL equations based on the model (2.1) and the independent Y_i^*'s with the first two moments

$$E(Y_i^*) = \mu_i, \ \mathrm{Var}(Y_i^*) = \phi\mu_i^2.$$

This proves the part (a).

To prove the part (b), rewrite the quasi-score function as matrix notation, i.e.,

$$q'(\beta) = (X^t W u^*)/\phi, \tag{4.8}$$

where u^* is the $n \times 1$ vector with the ith element $u_i^* = (y_i^* - \mu_i)g'(\mu_i)$. By taking the covariance of $q'(\beta)$ with $E(q_j'(\beta)) = 0$, independency of Y_i^*'s and the variance structure, the proof of the part (b) is completed. □

For the true value of β in the model (2.1), the QL's form (4.4) is equivalent to the likelihood equations (2.4). Thus the solutions, say $\hat{\beta}^*$, of (4.4) become the ML estimates of β.

In fact, it is difficult to calculate the variance of Y_i^* in (4.3) due to censoring mechanism. However, $\mathrm{Var}(Y_i^*) = \phi\mu_i^2$ in Proposition 4.1 is used as a working variance, even if it may be different from the true variance of Y_i^*. Also, since $E(q_j'(\beta)) = 0$ ($j = 1, 2, \cdots, p$) as long as $E(Y_i^*) = \mu_i$ under the model (2.1), the ML estimator $\hat{\beta}^*$ will be consistent and robust against a misspecification of variance of Y_i^*, see Liang and Zeger (1986), McCullagh and Nelder (1989, Sec. 9).

On the other hand, we can see that (4.5) is positive definite, because the diagonal elements of W are always positive except trivial case (e.g., $\beta = 0$) and also it is always possible to select the columns of X to have full column rank.

We now provide the following FS-II algorithm for obtaining the ML estimator $\hat{\beta}^*$ via the proposed FS-II method:

Step 0 : Use (3.10) as an initial estimate $\hat{\beta}^{*(0)}$ of β.

Step 1 : Calculate the following with a chosen link function $g(\cdot)$

$$\hat{\mu}_i^{(0)} = g^{-1}(x_i^t\hat{\beta}^{*(0)}) \quad (i = 1, 2, \cdots, n),$$

and then calculate for $i \in C$

$$\hat{y}_i^{*(0)} = y_i + \hat{\mu}_i^{(0)}.$$

Step 2 : Calculate $q'(\hat{\beta}^{*(0)})$ and $F(\hat{\beta}^{*(0)})$,
where $q'(\hat{\beta}^{*(0)})$ is the quasi-score vector $q'(\beta)$ in (4.8) and $F(\hat{\beta}^{*(0)})$ is the quasi-Fisher information $F(\beta)$ in (4.5), evaluated at $\beta = \hat{\beta}^{*(0)}$, respectively.

Step 3 : Calculate the next approximation $\hat{\beta}^{*(1)}$ to β, as

$$\hat{\beta}^{*(1)} = \hat{\beta}^{*(0)} + [F(\hat{\beta}^{*(0)})]^{-1}q'(\hat{\beta}^{*(0)}),$$

i.e.,

$$\hat{\beta}^{*(1)} = \left(X^t\hat{W}^{(0)}X\right)^{-1}X^t\hat{W}^{(0)}\hat{z}^{*(0)},$$

where $\hat{W}^{(0)}$ and $\hat{z}^{*(0)}$ are the W in (2.5) and the $n \times 1$ adjusted dependent vector with the ith element $z_i^*(= g(\mu_i) + u_i^*)$, evaluated at $\beta = \hat{\beta}^{*(0)}$, respectively.

Step 4 : Repeat Step 1 to Step 3 by replacing $\hat{\beta}^{*(0)}$ with $\hat{\beta}^{*(1)}$ until convergence is hopefully achieved. One stop if $\hat{\beta}^{*(0)}$ and $\hat{\beta}^{*(1)}$ are close together and $\ell'(\hat{\beta}^{*(1)})$ is close to zero.

From the above Step 3, we can see that the ML estimator $\hat{\beta}^*$ is not affected by the value of the dispersion parameter ϕ and the FS-II algorithm is similar to the iterative weighted least squares (IWLS) to fit the GLMs in the case of uncensoring. Note that the SEs of $\hat{\beta}_j^*$'s can be obtained from the estimator (3.9) of Fisher information (or the observed information in (3.3)).

5. SPECIAL LINKS

In this section, we briefiy will compare with three methods (N-R, FS-I and FS-II) estimating the parameters in the model (2.1) with three useful links, such as log link ($\eta_i = \log(\mu_i)$), identity link ($\eta_i = \mu_i$) and reciprocal link ($\eta_i = 1/\mu_i$).

For the model (2.1), the requirement $\mu_i > 0$ for all i have to satisfy. That is, the admissible space B for β is given by

$$B = \{\beta | \ \mu_i = g^{-1}(x_i^t\beta) > 0, \ \ \forall \ i\}.$$

Thus, in the log link the admissible space for β has not restrictions on β, i.e.,

$$B1 = \{\beta | \ -\infty < x_i^t\beta < \infty, \ \ \forall \ i\},$$

whereas in the reciprocal and identity link the space has restrictions on β, i.e.,

$$B2 = \{\beta | \ x_i^t\beta > 0, \ \ \forall \ i\}.$$

Frequently there is no physical reason to choose one of the links, and this choice may be checked via the model checking based on residual analysis. However, in estimating the model (2.1) with these restrictions, the N-R method sometimes leads to numerical or statistical problems, e.g., Mantel and Myers (1971).

The initial estimate (3.10) can be used as common that of three methods. Moreover, it can be easily derived from not only the FS-I and FS-II iterations but also the N-R iterations, by regarding censored data as uncensored and using the data y as the initial estimate of μ. For three links, we can obtain the initial estimate, the observed information, the estimator of the Fisher information and the observed quasi-Fisher information, from (3.10),(3.3),(3.9) and (4.5), respectively. These will play an important role of finding the convergent solution of β in three methods, and also will contribute to the convergences and their rates. The results obtained are summarized by Table 5.1.

Table 5.1 illustrates the followings: (i) the implementation of two proposed methods (FS-I and FS-II) are easier and simpler than that of the N-R method, particularly in the identity link; (ii) as an alternative of the Fisher information, the FS-I method is based on only uncensored observations, but the FS-II method full observations; (iii) in the case of the reciprocal link, the N-R and the FS-I methods give the same results.

6. EXAMPLES

6.1. Simulated Examples

In order to compare the proposed FS-I and FS-II methods with the N-R method in aspect of convergence, we consider the model (2.1) with $p = 2$ and identity link, i.e.

$$\mu_i = \beta_0 + \beta_1 x_i \ (i = 1, 2, \cdots, n), \qquad (6.1)$$

Table 5.1: Comparison on convergences of N-R, FS-I and FS-II methods for each link.

link	$\hat{\beta}^{(0)} = (X^t W^{(0)} X)^{-1} \cdot X^t W^0 z^0$	$-\ell''(\hat{\beta}) = X^t \hat{W}^* X$ (N-R method)	$\hat{I}(\hat{\beta}) = X_D^t \hat{W}_D X_D$ (FS-I method)	$F(\hat{\beta}^*) = X^t \hat{W} X/\hat{\phi}$ (FS-II method)
log	$W^{(0)} = I_n^{\dagger}$, $z_i^{(0)} = log(y_i)$ for $i = 1, 2, \cdots, n$	$W^* = diag(\mu_i^{-1} y_i)$ for $i = 1, 2, \cdots, n$	$W_D = I_r^{\dagger}$	$W = I_n^{\dagger}$
reciprocal	$W^{(0)} = diag(y_i^2)$, $z_i^{(0)} = 1/y_i$ for $i = 1, 2, \cdots, n$	$W^* = diag(\mu_i^2 \delta_i)$ for $i = 1, 2, \cdots, n$	$W_D = diag(\mu_i^2)$ for $i \in D$	$W = diag(\mu_i^2)$ for $i = 1, 2, \cdots, n$
identity	$W^{(0)} = diag(y_i^{-2})$, $z_i^{(0)} = y_i$ for $i = 1, 2, \cdots, n$	$W^* = diag(-\mu_i^{-2} \delta_i + 2\mu_i^{-3} y_i)$ for $i = 1, 2, \cdots, n$	$W_D = diag(\mu_i^{-2})$ for $i \in D$	$W = diag(\mu_i^{-2})$ for $i = 1, 2, \cdots, n$

†: $I_n(I_r)$ denotes $n \times n(r \times r)$ identity matrix.

where the T_i's are independent exponential survival times with the mean μ_i. For simplicity the model parameters are set to $\beta_0 = \beta_1 = 1$, and the fixed covariates and the sample size n are set to the following two cases:

Case 1) $x_i = i/n, \quad n = 250$.

Case 2) $x_i = 0.1 \times i, \quad n = 200$.

Since the survival times are subjected to be censored to the right, we set the censoring times C_i's which are independent of T_i's to be distributed as an uniform distribution $U(0, \lambda_i)$ with parameter $\lambda_i = \mu_i / CR$. Here CR is the nominal censoring rate, and set to 25% [20%] for Case 1 [Case 2], respectively.

For generations of random variates and estimation of the model parameters, IMSL package and SAS/IML are used, respectively. For three methods the same estimate is used as an initial value $\hat{\beta}^{(0)}$. That is, in identity link we use the initial estimate (3.10); also see Table 5.1. As the convergence criterion, we use the maximum relative changes (upper bound 10^{-3}) of the previous and current estimates for β.

Example 6.1. For Case 1, the actural CR was 24%. The results are summarized in Table 6.1; the FS-I, the FS-II, and the N-R methods converge after the 3th, the 7th, and the 14th iteration, respectively. Thus two FS-I and FS-II methods achieve more rapid convergence than the N-R method. Note that the final estimates of the model parameters are obtained as $(1.01, 1.13)$ and these seems to have good agreement with the imposed population values $\beta_0 = \beta_1 = 1$.

Table 6.1: Comparision of N-R method and FS-I, II methods in Case 1.

Iteration	N-R		FS-I		FS-II	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
initial	0.022[a)]	0.012				
1	0.034	0.019	1.026	1.087	0.833	0.720
2	0.050	0.028	1.008	1.126	0.981	0.994
3	0.075	0.042	1.007	1.129	1.004	1.087
4	0.110	0.063	unchanged		1.007	1.116
5	0.162	0.094			1.007	1.126
6	0.237	0.139			1.007	1.128
7	0.400	0.206			1.007	1.129
8	0.476	0.303			unchanged	
9	0.642	0.441				
10	0.817	0.626				
11	0.951	0.848				
12	1.005	1.043				
13	1.008	1.123				
14	1.007	1.130				
15	unchanged					
SE	0.181	0.394	0.175	0.379	0.175	0.379

a): represents that the same initial estimate is used in N-R , FS-I and II.

Example 6.2. For Case 2, the actural CR was 15%. Some interesting patterns emerge when the simulated censored data are performed by three methods. According to Table 6.2, the FS-I and the FS-II methods converge, but the N-R method does not converge with the initial estimate (3.10); if we use the one-step iteration estimate of FS-I as a new N-R starting point, the N-R method converges, but it takes two iterations longer than the FS-I.

Additionally, from Example 6.1 and Example 6.2 we can see that all three methods provide nearly the same SEs; the SEs in the FS-II were obtained from (3.9).

6.2. Real Examples

We here illustrate the performances of the purposed methods with real examples.

Example 6.3. The survival data on 17 AG positive lukemia patients with the corresponding log white blood counts (x) were given by Zippin and Armitage

Table 6.2: Comparision of N-R method and FS-I, II methods in Case 2.

Iteration	N-R		FS-I		N-R		FS-II	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
initial	0.042[a)]	-0.003						
1	0.063	-0.004	21.755	-0.999	—	—	18.358	-0.760
2	0.094	-0.007	9.539	0.179	15.614[b)]	-0.606	7.595	0.303
3	0.140	-0.010	1.763	0.801	4.307	0.152	2.830	0.718
4	0.209	-0.015	1.371	0.843	2.768	0.392	1.723	0.812
5	0.311	-0.022	1.387	0.840	1.655	0.640	1.471	0.833
6	0.462	-0.033	1.386	0.840	1.467	0.779	1.409	0.838
7	0.679	-0.048	unchanged		1.393	0.834	1.392	0.840
8	0.983	-0.070			1.386	0.840	1.388	0.840
9	1.366	-0.097			unchanged		1.387	0.840
10	4.580	-0.325					unchanged	
11	6.509	-0.461						
12	8.955	-0.635						
13	11.631	-0.824						
⋮		⋮						
39	185188459	-13133933						
40	370200566	-26255359						
41	740705743	-52532322						
42	divergent							
SE			0.545	0.104	0.535	0.102	0.545	0.104

a): See Footnote a) in Table 6.1.
b): respresents that the first iteration estimate of FS-I is used for convergence of N-R.

(1966). The 17 patients are consisted of 12 deaths and 5 survivors. Then the model (6.1) is fitted via three methods. The convergence results are appeared in Table 6.3 and give the similar results to Example 6.1, although the sample size is too small. The fitted model is as follows: (i) for the N-R, $\hat{\mu} = 257.34 - 45.84x$ with corresponding SEs 121.93 and 25.99; (ii) for the FS-I, $\hat{\mu} = 257.34 - 45.84x$ with corresponding SEs 128.87 and 27.72; (iii) for the FS-II, $\hat{\mu} = 257.34 - 45.84x$ with corresponding SEs 128.87 and 27.72.

Example 6.4. The survival data for 137 advanced lung cancer patients with the months (x) from diagnosis to entry into the study, are taken from the data set discussed by Prentice (1973). Of the 137 patients 9 survivors exist. Then the model (6.1) is also fitted via three methods. The convergence results are appeared in Table 6.4 and give the similar results to Example 6.2; in particular the N-R method does not converge at all but converges slowly when the first iteration value of the FS-I is used as a new N-R starting point.

Table 6.3: Comparision of N-R method and FS-I, II methods in Zippin and Armitage's(1966) data.

Iteration	N-R		FS-I		FS-II	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
initial	42.440[a)]	-8.287				
1	62.054	-12.110	220.575	-37.646	170.955	-29.851
2	89.521	-17.457	262.536	-47.044	242.708	-43.568
3	126.471	-24.633	256.327	-45.600	253.695	-45.333
4	172.878	-33.601	257.530	-45.879	256.350	-45.712
5	*224.314*	*-43.444*	*257.307*	*-45.827*	*257.053*	*-45.802*
6	268.695	-51.705	257.349	-45.837	257.254	-45.826
7	289.315	-55.023	257.341	-45.835	257.315	-45.832
8	283.772	-52.954	257.343	-45.835	257.334	-45.834
9	270.167	-49.283	unchanged		257.340	-45.835
10	260.653	-46.728			257.342	-45.835
11	257.588	-45.902			unchanged	
12	257.349	-45.836				
13	257.343	-45.835				
14	unchanged					
SE	121.933	25.987	128.869	27.720	128.868	27.720

a): See Footnote a) in Table 6.1.
Note: The upper bound of maximum relative changes was used as 10^{-5}.

Table 6.4: Comparision of N-R method and FS-I, II methods in Prentice's(1973) data.

Iteration	N-R		FS-I		N-R		FS-II	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
initial	3.477[a)]	-0.063						
1	5.197	-0.095	152.739	-2.446	—	—	142.757	-2.272
2	7.753	-0.142	149.739	-2.320	143.450[b)]	-1.946	148.905	-2.284
3	11.535	-0.211	147.763	-2.148	149.439	-2.180	147.108	-2.084
4	*17.091*	*-0.313*	*143.463*	*-1.713*	*152.141*	*-2.444*	142.040	-1.544
5	25.163	-0.463	138.442	-1.013	145.846	-2.102	138.949	-1.021
6	36.695	-0.678	140.217	-1.113	151.492	-2.390	140.166	-1.112
7	52.733	-0.983	140.201	-1.111	170.470	-3.716	140.199	-1.112
8	74.101	-1.401	140.202	-1.112	214.427	-6.916	unchanged	
9	100.674	-1.959	unchanged		122.419	-3.287		
⋮	⋮				⋮			
21	-401671.9	14348.892			140.215	-1.113		
22	-803152.8	28688.483			140.202	-1.112		
⋮	⋮				unchanged			
34	-3.2892×10^9	117470043						
35	-6.5769×10^9	234888188						
36	divergent							
SE			*13.541*	*0.512*	*13.488*	*0.500*	*13.540*	*0.512*

a): See Footnote a) in Table 6.1.
b): See Footnote b) in Table 6.2.

7. CONCLUDING REMARKS

We proposed two modified FS methods (FS-I and FS-II), in order to improve the N-R method with convergence problem in estimating the model (2.1) with censored data. For this the several examples were illustrated.

As the results we have observed the following facts: (i) two proposed FS methods than the N-R method are more easily implemented and appeared in giving better convergence; (ii) the selected initial estimate provides good convergence for not only two FS methods but also the N-R method. In addition, from Table 5.1 we can see that the proposed FS methods also work well under the various model (2.1) with three links (identity, log, and reciprocal) and $p \geq 2$ (e.g., $p = 8$).

Finally, the proposed methods may be used in estimating other regression models with censored data.

REFERENCES

Buckley, J. and James, I. (1979). "Linear regression with censored data," *Biometrika,* **66**, 429-436.

Glasser, M. (1967). "Exponential survival with covariance," *Journal of the American Statistical Association,* **62**, 561-568.

Greenberg, R., Bayard, S., and Byar, D. (1974). "Selecting concomitant variables using a likelihood ratio step-down procedure and a method of testing goodness of fit of an exponential survival model," *Biometrics,* **30**, 601-608.

Krall, J., Uthoff, V., and Harley, J. (1975). "A step-up procedure for selecting variables associated with survival," *Biometrics,* **31**, 49-57.

Lawless, J. F. (1982). *Statistical models and methods for life time data,* John Wiley and Sons, New York.

Liang, K. Y. and Zeger, S. L. (1986). "Longitudinal data analysis using generalized linear models," *Biometrika,* **73**, 13-22.

Mantel, N. and Myers, M. (1971). "Problems of convergence of maximum likelihood iterative procedures in multiparameter situations," *Journal of the American Statistical Association,* **66**, 484-491.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models,* 2nd ed., Chapman and Hall, London.

Nelder, J. A. and Wedderburn, R. W. M. (1972). "Generalized linear models," *Journal of the Royal Statistical Society,* **A 135**, 370-384.

Prentice, R. L. (1973). "Exponential survival with censoring and explanatory variables," *Biometrika,* **60**, 279-288.

Wedderburn, R. W. M. (1974). "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," *Biometrika* **61**, 439-447.

Zippen, C. and Armitage, P. (1966). "Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter," *Biometrics,* **22**, 665-672.