Journal of the Korean
Data & Information Science Society
1998, Vol. 9, No. 2, pp. 219 ~ 225

Quasi-Likelihood Approach for Linear Models with Censored Data

Il-Do Ha [1] · **Geon-Ho Cho** [2]

Abstract

The parameters in linear models with censored normal responses are usually estimated by the iterative maximum likelihood and least square methods. However, the iterative least square method is simple but hardly has theoretical justification, and the iterative maximum likelihood estimating equations are complicatedly derived. In this paper, we justify these methods via Wedderburn (1974)'s quasi-likelihood approach. This provides an explicit justification for the iterative least square method and also directly the iterative maximum likelihood method for estimating the regression coefficients.

Key Words and Phrases: censored data, generalized linear models, iterative maximum likelihood and least squares methods, Newton-Raphson method, quasi-likelihood.

1. Introduction

The estimation of linear models with censored normal responses was studied by several authors; see Wolynetz (1979), Schmee and Hahn (1979) and Aitkin (1981). Wolynetz (1979) presented an iterative maximum likelihood (ML) method and Aitkin (1981) obtained the EM equations, which are introduced by Dempster

[1] Full-time Instructor, Department of Statistics, Kyungsan University, Kyungpook, 712-240, Korea.
[2] Assistant Professor, Department of Statistics, Kyungsan University, Kyungpook, 712-240, Korea.

et al. (1977), equivalent to the ML estimating equations. Also, Schmee and Hahn (1979) described an iterative least squares (ILS) method. However, the ILS method is simple but hardly has theoretical justification, and the iterative ML estimating equations are complicatedly derived.

On the other hand, the Wedderburn (1974)'s quasi-likelihood (QL) approach, which greatly widened the scope of generalized linear models (GLMs) developed by Nelder and Wedderburn(1972), is widely used in data analysis where in paticular a likelihood is not available. Hence, the QL approach can be applied to the ILS method no using the likelihood of the normal linear model with censored data.

In this paper, we justify the above estimating methods via the QL approach. This provides an explicit justification for the ILS method and also directly the iterative ML method for estimating the regression coefficients.

The paper is organized as follows. In Section 2, we review the QL and its properties. In Section 3, the normal linear model with censored data and the estimating equations on the model are described. Also, the iterative ML and LS methods are discussed. Finally, a justfication of these methods via the QL approach is given in Section 4.

2. Quasi-Likelihood and Properties

Suppose that the ith observation's response variable Y_i $(i = 1, \cdots, n)$ have independent with only the first two moments

$$E(Y_i) = \mu_i, \quad \text{Var}(Y_i) = \phi V(\mu_i), \tag{1}$$

where μ_i is some known mean function of a set of unknown parameters $\beta_1, \cdots, \beta_p$, ϕ is unknown dispersion parameter, and $V(\cdot)$ is the variance function. In typical applications μ_i is determined by known covariates, $x_{i1}, \cdots, x_{ip}$ say, possibly through the following model equations

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

where η_i is the linear predictor, $g(\cdot)$ is some known GLM link function, $x_i^t =$

$(x_{i1}, \cdots, x_{ip})$ is the $1\times p$ vector of the covariates for observation i, and $\beta = (\beta_1, \cdots, \beta_p)^t$ is the $p \times 1$ vector of the model parameters.

For any $V(\mu_i)$ Wedderburn (1974) defined a quasi-likelihood (QL), more strictly a quasi-log-likelihood, q by the relation

$$\partial q/\partial\beta_j = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\phi V(\mu_i)} \cdot \frac{1}{g'(\mu_i)} \quad (j = 1, \cdots, p), \tag{3}$$

where $q = \sum_{i=1}^{n} q_i$ with the ith QL component q_i and $g'(\mu_i) = \partial g(\mu_i)/\partial\mu_i$. Here, $\partial q/\partial\beta_j = 0 \ \ (j = 1, \cdots, p)$ in (3) are called the QL estimating equations. Since the QL estimating equations are non-linear functions of β, solving them for maximum quasi-likelihood (MQL) estimate $\hat{\beta}$ of β usually requires Fisher scoring approach.

In fact, the QL estimating equations have the same estimating equations as those of GLMs and further the MQL estimator is exactly same as the ML estimator if there is a distribution of the GLM type having $\mathrm{Var}(Y_i) = \phi V(\mu_i)$. Also, since $E(\partial q/\partial\beta_j) = 0 \ \ (j = 1, \cdots, p)$ as long as $E(Y_i) = \mu_i$, the MQL estimator is consistent and robust against a misspecification of variance of Y_i. Further, McCullagh (1983) showed that under regular conditions, the MQL estimator $\hat{\beta}$ is consistent and have asymptotically normal distribution.

Clearly the MQL estimator is not affected by the value of ϕ, so that it can be calculated as if ϕ was known to be 1. However, in order to obtain its standard error some estimate of ϕ is required. Wedderburn suggested a moment estimate given by

$$\hat{\phi} = \frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{\mu}_i)/V(\hat{\mu}_i) = X^2/(n-p), \tag{4}$$

where X^2 is the Pearson statistic.

3. Estimating Equations on Model

Let T_i be the survival time on study for the ith $(i = 1, 2, \quad, n)$ 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

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

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 and C_i $(i = 1, 2, \cdots, n)$ are independent.

Under the above settings, we consider here the normal regression model

$$T_i = x_i^t \beta + e_i \ (i = 1, 2, \cdots, n), \tag{5}$$

where β is the $p \times 1$ vector of the unknown model parameters and error terms e_i's have i.i.d. $N(0, \sigma^2)$. Note that the response T_i in the model (5) may be a suitable transformation of the survival time on the study, e.g. $\log(T_i)$.

Based on the observations (y_i, δ_i, x_i^t) for $i = 1, 2, \quad, n$, the log-likelihood of (β, σ^2) in the model (5) is given by

$$\ell = \ell(\beta, \sigma^2) \propto -\frac{r}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i \in D} (y_i - \mu_i)^2 + \sum_{i \in C} \log \bar{\Phi}((y_i - \mu_i)/\sigma).$$

Here $D[C]$ are the indext set of individuals for uncensored [censored] observations, r is the number of uncensored observations, and Φ $(= 1 - \bar{\Phi})$ is the distribution function of $N(0, 1)$.

Then the ML estimating equations for (β, σ^2) are given by

$$\partial \ell / \partial \beta_j = \frac{1}{\sigma} \{ \sum_{i \in D} x_{ij} m_i + \sum_{i \in C} x_{ij} h(m_i) \} = 0 \quad (j = 1, \cdots, p) \tag{6}$$

and

$$\partial \ell / \partial \sigma^2 = -\frac{1}{2\sigma^2} \{ r - \sum_{i \in D} {m_i}^2 - \sum_{i \in C} m_i h(m_i) \} = 0, \tag{7}$$

where $m_i = (y_i - \mu_i)/\sigma$ and $h(\cdot)$ is the hazard function of $N(0, 1)$. The above two equations give the complicate non-linear functions for (β, σ^2). Thus, in order to their ML estimates the two equations are usually solved via the Newton-Raphson (N-R) method. That is, this requires the second derivatives of ℓ. However, if censoring is fairly heavy, the N-R method sometimes fail to converge unless initial parameter estimates are very close to the ML estimates. Also, negative values of σ^2 can arise during the iteration procedure; see Lawless (1982, pp. 315).

As an alternative to the N-R method, other iterative methods are suggested by several authors, see Wolynetz (1979), Schmee and Hahn (1979) and Aitkin (1981). For this, they considered the pseudo random variables

$$Y_i^* = Y_i \delta_i + E(T_i | T_i > Y_i)(1 - \delta_i) \ (i = 1, \cdots, n). \tag{8}$$

Here, since T_i has $N(\mu_i, \sigma^2)$,

$$E(T_i|T_i > Y_i) = \mu_i + \sigma h(m_i). \tag{9}$$

By plugging the equations (8) with (9) in the ML equations (6) and (7), the ML equations are reduced to

$$\partial\ell/\partial\beta_j = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i^* - \mu_i) = 0 \quad (j = 1, \cdots, p) \tag{10}$$

and

$$\partial\ell/\partial\sigma^2 = -\frac{1}{2\sigma^2}\{r - \sum_{i=1}^{n}(y_i^* - \mu_i)^2/\sigma^2 + \sum_{i=1}^{n}\lambda(m_i)\} = 0, \tag{11}$$

where $\lambda(x) = \partial h(x)/\partial x = h(x)[h(x) - x]$. Since the two ML equations give the simple linear forms for (β, σ^2), these equations can be easily solved iteratively for β and σ^2 to give the ML estimates; for more detail, see Lawless (1982, pp. 316).

Note that Wolynetz (1979) presented the ML equations (10) and (11) and Aitkin (1981) obtained the EM equations equivalent to (10) and (11). This ML method gives simpler forms than the N-R method but is still complicatedly derived. On the other hand, Schmee and Hahn (1979) described an ILS method, with Y_i^* instead to the response T_i in the model (5). The ILS method provides the same equations as (10) for β, but different from (11) for σ^2. In order to estimate σ^2, Schmee and Hahn (1979) used the residual mean squares estimate of σ^2, based on the response Y_i^*. Notice that the ML and ILS methods are essentially the same except for the estimate of the error variance σ^2.

Here, we can see that the ILS method provides simpler forms than the ML and the N-R methods but almost has not theoretical justification. In next section, we justify the ILS method and also investigate the ML method, via the QL approach.

4. A Justification via QL Approach

The model (5) can be rewritten as the normal linear models in which T_i ($i = 1, 2, \cdots, n$) are independent with

$$\mu_i = x_i^t\beta \text{ and } \mathrm{Var}(T_i) = \phi V(\mu_i), \tag{12}$$

where $\mu_i = E(T_i)$, $\phi = \sigma^2$, and $V(\mu_i) = 1$. Under the model (12), T_i's may be subjected to be censored, so only Y_i's are observed, but

$$E(Y_i) \neq \mu_i.$$

We here consider the pseudo random variables Y_i^*'s in (8). Then they give

$$E(Y_i^*) = \mu_i \quad (i = 1, 2, \cdots, n), \tag{13}$$

where $\mu_i = E(T_i) = x_i^t\beta$. Note that the expectation identity (13) that provides validity for estimation in the linear models with censored data was shown by Buckley and James(1979).

From the ML equations (10) for β and the expectation identity (13), we can see that the ML equations result from the QL equations with

$$E(Y_i^*) = \mu_i = x_i^t\beta \quad \text{and} \quad \text{Var}(Y_i^*) = \phi = \sigma^2$$

based on the pseudo random variable Y_i^* with (11); also see the QL equations for (3). So, the MQL estimator of β becomes the ML estimator of Wolynetz (1979) and Aitkin (1981), and also the LS estimator of Schmee and Hahn (1979). Here, it is difficult to calculate the variance of Y_i^* due to censoring mechanism. However, by the consistent and robust properties of the MQL estimator mentioned in Section 3, there needs no this calculation. That is, since $E(\partial\ell/\partial\beta_j) = 0 \;\; (j = 1, \cdots, p)$ as long as $E(Y_i^*) = \mu_i$ under the model (5) or (13), the corresponding MQL estimator is consistent and robust against a misspecification of variance of Y_i^*.

Further, for the estimation of σ^2, we can use the estimates (4); this gives the same estimate as the ILS estimate for σ^2. Now, the parameters β and σ^2 are asymtotically parameter orthogonality, in sense of Cox and Reid (1987). Thus, the MQL (i.e., ML and LS) estimates of β are not almost affected by an estimate of σ^2.

After all, the QL approach provides an explicit justification for the ILS method of Schmee and Hahn (1979) and also justifies the ML method for β of Wolynetz (1979) and Aitkin (1981).

References

1. Aitkin, M. A. (1981). A note on the regression analysis of censored data. *Technometrics*, 23, 161-163.

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

3. Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with disussion). *Journal of the Royal Statistical Society B*, 49, 1-39.

4. Dempster, A. P., Laird, N. M. and Rubin. D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39, 1-38.

5. Lawless, J.F. (1982). *Statistical Models and Methods for Life Time Data.* John Wiley and Sons, New York.

6. McCullagh, P. (1983), Quasi-likelihood functions. *Annals of Statistics*, 11, 59-67.

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

8. Schmee, J. and Hahn, G. J. (1979). A simple method for regression analysis with censored data. *Technometrics*, 21, 417-423.

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

10. Wolynetz, M.S.(1979). Maximum likelihood estimation in a linear model from confined and censored normal data. Algorithm AS139, *Applied Statistics*, 28, 195-206.