University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2004 *Paper* 153

A Note on Empirical Likelihood Inference of Residual Life Regression

Ying Qing Chen* Yichuan Zhao†

*Division of Biostatistics, School of Public Health, University of California, Berkeley, yqchen@stat.berkeley.edu

†Department of Mathematics and Statistics, Georgia State University, Atlanta, Georgia, dz2007@gmail.com

A Note on Empirical Likelihood Inference of Residual Life Regression

Ying Qing Chen and Yichuan Zhao

Abstract

Mean residual life function, or life expectancy, is an important function to characterize distribution of residual life. The proportional mean residual life model by Oakes and Dasu (1990) is a regression tool to study the association between life expectancy and its associated covariates. Although semiparametric inference procedures have been proposed in the literature, the accuracy of such procedures may be low when the censoring proportion is relatively large. In this paper, the semiparametric inference procedures are studied with an empirical likelihood ratio method. An empirical likelihood confidence region is constructed for the regression parameters. The proposed method is further compared with the normal approximation based method through a simulation study.

1 Introduction

A mean residual life function, $m(t)$, $t \geq 0$, is the expected remaining life given survival to t. Suppose T is a failure time, then $m(t) = E(T - t \mid T > t)$. It is an important function in economics, actuarial sciences, reliability and survival analysis to characterize life expectancy. Oakes and Dasu (1990) proposed a class of semiparametric models called the proportional mean residual life model, as an alternative to the widely used Cox proportional hazards model, to study the association between $m(t)$ and its associated covariates. The Oakes-Dasu model directly models the distribution of residual life and carries appealing interpretation in life expectancy. Specifically, an Oakes-Dasu proportional mean residual life model usually assumes that

$$m\{t \mid Z(t)\} = m_0(t) \exp\{\beta^{\mathrm{T}} Z(t)\}, \quad (1)$$

where $m(\cdot)$ are mean residual life functions, $Z(\cdot)$ are $p-$vector covariates and β are associated parameters. In a semiparametric version of this model, $m_0(\cdot)$ is usually unspecified. When a model satisfies both the proportional hazards and the proportional mean residual life assumptions, its underlying distributions belong to the Hall-Wellner class of distributions with linear mean residual life functions (Oakes & Dasu, 1990).

When there is no censoring, estimation procedures were developed in Maguluri and Zhang (1994). When censoring presents, Chen and Cheng (2004) recently developed quasi-partial score estimating equations for the regression parameters. Nevertheless, these large-sample normal approximation based estimation methods tend to have poor performance when the sample size is relatively small or the censoring proportion is relatively large. In this short note, we instead consider empirical likelihood method to estimate the parameters in model (1), as it is a powerful nonparametric method. In general, the empirical likelihood method has unique features, such as range respecting, transformation-preserving, asymmetric confidence interval, Bartlett correctability, and better coverage probability for small sample (Owen, 2001). In analysis of censored survival times, for example, empirical likelihood was used to derived pointwise confidence intervals for survival function with right censored data as early as in 1975 (Thomas & Grunkemeier, 1975).

In this short note we use the simple estimating equations in Chen and Cheng (1994) to construct an empirical likelihood-ratio based confidence region. The proposed confidence region and main asymptotic result are in Section 2. Results from a moderate simulation are in Section 3 to compare our method with normal approximation based method. Some alternative estimation method is discussed in Section 4.

2 Main results

In addition to the failure time T, let C be the potential censoring time. Conditional on the $p-$vector covariate Z, T and C are assumed to be independent. Suppose that the observed data set consists of n independent copies of (X_i, Δ_i, Z_i), $i = 1, \ldots, n$, where $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \le C_i)$. Here, $I(\cdot)$ is indicator function. Denote $Y_i(t) = I(X_i \ge t)$ and $N_i(t) = I(X_i \le t)\Delta_i$. Let $0 < \tau = \inf\{t : pr(X > t) = 0\} < \infty$.

As derived in (2004), the following estimating equations can be used to estimate the parameter β in model (1):

$$U(\beta) = \sum_{i=1}^{n} \int_0^\tau \{Z_i - \overline{Z}(t)\}\{\widehat{m}_0(t;\beta)dN_i(t) - Y_i(t)\exp(-\beta^T Z_i)dt\} = 0, \tag{2}$$

where $\overline{Z}(t) = \sum_{i=1}^n Y_i(t)Z_i / \sum_{i=1}^n Y_i(t)$ and

$$\widehat{m}_0(t;\beta) = \left[\exp\left\{-\int_0^t \frac{\sum_i dN_i(t)}{\sum_i Y_i(t)})\right\}\right]^{-1} \int_t^\tau \exp\left\{-\int_0^u \frac{\sum_i dN_i(s)}{\sum_i Y_i(s)}\right\} \frac{\sum_i Y_i(u)\exp(-\beta^T Z_i)}{\sum_i Y_i(u)} du.$$

Denote $\widehat{\beta}$ and β_* the estimated and true parameter of β, respectively. Then as shown in Chen and Cheng (2004), under the regularity conditions, the random vector

$$n^{1/2}(\widehat{\beta} - \beta_*) \xrightarrow{\mathcal{D}} N(0, A^{-1}VA^{-1}),$$

where A and V can be consistently estimated by their empirical estimators,

$$\widehat{A} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \overline{Z}(t)\}^{\otimes 2} Y_i(t)\exp(-\widehat{\beta}^T Z_i)dt, \text{ and}$$

$$\widehat{V} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \{Z_i(t) - \overline{Z}(t)\}^{\otimes 2} Y_i(t)\widehat{m}_0(t;\widehat{\beta})\{\exp(-\widehat{\beta}^T Z_i)dt + d\widehat{m}_0(t;\widehat{\beta})\},$$

respectively. Thus an asymptotic $100(1-\alpha)\%$ confidence region for β based on the above normal approximation is given by

$$\mathcal{R}_1 = \{\beta : n(\widehat{\beta}-\beta_*)^{\mathrm{T}}\widehat{A}\widehat{V}^{-1}\widehat{A}(\widehat{\beta}-\beta_*) \le \chi_p^2(\alpha)\}, \tag{3}$$

where $\chi_p^2(\alpha)$ is the upper α-quantile of the chi-square distribution with degrees of freedom p. Apparently the accuracy of $\mathcal{R}_1$ mainly depends on the large-sample normal approximation and also the proportion of censoring. For relatively small sample size or large censoring proportions, its accuracy may be compromised.

Now consider the empirical likelihood approach, instead. For $i = 1, 2, \ldots, n$, we define

$$W_i = \int_0^\tau \{Z_i - \overline{Z}(t)\}\{\widehat{m}_0(t;\beta_*)dN_i(t) - Y_i(t)\exp(-\beta_*^T Z_i)dt\},$$

and summarize the following results in Chen and Cheng (2004) as a lemma.

Lemma 1. *Under regularity conditions in Chen and Cheng (2004), (i)* $n^{-1/2}\sum_{i=1}^n W_i \xrightarrow{\mathcal{D}} N(0, V)$*, and (ii)* $n^{-1}\sum_{i=1}^n W_i W_i^T \to V$ *in probability.*

Thus the associated empirical likelihood is

$$L(\beta_*) = \sup\left\{\prod_{i=1}^n p_i : \sum p_i = 1, \sum_{i=1}^n p_i W_i = 0, p_i \ge 0, i = 1, \ldots, n\right\}.$$

Let $p = (p_1, \ldots, p_n)^{\mathrm{T}}$ be a vector of probabilities such that $\sum_{i=1}^n p_i = 1$, where $p_i \ge 0$, $i = 1, 2, \ldots, n$. Since $\prod_{i=1}^n p_i$ attains its maximum at $p_i = 1/n$, the empirical likelihood ratio at the true value β_* is then

$$R(\beta_*) = \sup\left\{\prod_{i=1}^n np_i : \sum p_i = 1, \sum_{i=1}^n p_i W_i = 0, p_i \ge 0, i = 1, \ldots, n\right\}.$$

By using Lagrange multipliers, we have

$$-2\log R(\beta_*) = 2\sum_{i=1}^n \log\{1 + \lambda^{\mathrm{T}} W_i\}, \tag{4}$$

where λ satisfies the equation

$$\frac{1}{n}\sum_{i=1}^n \frac{W_i}{1+\lambda^T W_i} = 0. \tag{5}$$

Suppose that $\{Z_i(t)\}$ are uniformly bounded by a constant. Now define

$$V_i = \int_0^\tau \{Z_i - \mu_Z(t)\} m_*(t) dM_i(t),$$

where $\mu_Z(t)$ is the limit of $\overline{Z}(t)$ as $n \to \infty$. Then $E|V_i|^2 < \infty$. According to the proof of Lemma 3 in Owen (1990), we have $\max_{1\le i\le n} |V_i| = o_p(n^{1/2})$. Note

$$W_i = \int_0^\tau \left\{Z_i - \overline{Z}(t)\right\} m_*(t) dM_i(t) + o_p(1).$$

By the martingale representations of V_i and W_i, we can prove that $|V_i - W_i| = o_p(1)$. Then we have

$$\max_{1\le i\le n} |W_i| = o_p(n^{1/2}), \text{ and} \tag{6}$$

$$\frac{1}{n}\sum_{i=1}^n |W_i|^3 = o_p(n^{1/2}). \tag{7}$$

Let $\lambda = \rho\theta$, where $\rho \ge 0$ and $|\theta| = 1$. Recall $\Gamma_n = 1/n \sum_{i=1}^n W_i W_i^T = V + o_p(1)$, where V is the limit of $1/n \sum_{i=1}^n W_i W_i^T$. Let $\sigma_p > 0$ be the smallest eigenvalue of V. Then, $\theta\Gamma_n\theta \ge \sigma_p + o_p(1)$. According to Lemma 1, $1/n\,|\sum_{i=1}^n W_i| = O_p(n^{-1/2})$. By (6), the equations in (5) and the argument used in Owen (1990), we know that

$$|\lambda| \;=\; O_p(n^{-1/2}). \tag{8}$$

Consider a Taylor expansion to the right-hand side of (4),

$$-2\log R(\beta_*) = 2\sum_{i=1}^n \left\{\lambda^T W_i - \frac{1}{2}(\lambda^T W_i)^2\right\} + r_n, \tag{9}$$

where $|r_n| = O_p(1)\sum_{i=1}^n |\lambda^{\mathrm{T}} W_i|^3$. Hence, by (7), $|r_n| = O_p(1)|\lambda|^3 \sum_{i=1}^n |W_i|^3 = o_p(1)$. Furthermore, since

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \frac{W_i}{1+\lambda^{\mathrm{T}} W_i} &= \frac{1}{n}\sum_{i=1}^n W_i \left(1 - \lambda^{\mathrm{T}} W_i + \frac{(\lambda^{\mathrm{T}} W_i)^2}{1+\lambda^{\mathrm{T}} W_i}\right)\\ &= \frac{1}{n}\sum_{i=1}^n W_i - \left(\frac{1}{n}\sum_{i=1}^n W_i W_i^{\mathrm{T}}\right)\lambda + \frac{1}{n}\sum_{i=1}^n \frac{W_i(\lambda^{\mathrm{T}} W_i)^2}{1+\lambda^{\mathrm{T}} W_i} = 0,\end{aligned}$$

it follows that

$$\lambda = \left(\sum_{i=1}^n W_i W_i^{\mathrm{T}}\right)^{-1} \sum_{i=1}^n W_i + o_p(1). \tag{10}$$

Similarly, we have

$$\sum_{i=1}^{n} \frac{\lambda^{\mathrm{T}} W_i}{1+\lambda^{\mathrm{T}} W_i} = \sum_{i=1}^{n} (\lambda^{\mathrm{T}} W_i) - \sum_{i=1}^{n} (\lambda^{\mathrm{T}} W_i)^2 + \sum_{i=1}^{n} \frac{(\lambda^{\mathrm{T}} W_i)^3}{1+\lambda^{\mathrm{T}} W_i} = 0. \tag{11}$$

Since

$$\sum_{i=1}^{n} \frac{(\lambda^{\mathrm{T}} W_i)^3}{1+\lambda^{\mathrm{T}} W_i} = o_p(1), \tag{12}$$

we know that $\sum_{i=1}^{n} (\lambda^{\mathrm{T}} W_i)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda^{\mathrm{T}} W_i + o_p(1)$. As a result, the following is true

$$\begin{aligned} -2 \log R(\beta_*) &= \sum_{i=1}^{n} \lambda^{\mathrm{T}} W_i + o_p(1) \\ &= \left(n^{-1/2} \sum_{i=1}^{n} W_i \right)^{\mathrm{T}} \left(n^{-1} \sum_{i=1}^{n} W_i W_i^{\mathrm{T}} \right)^{-1} \left(n^{-1/2} \sum_{i=1}^{n} W_i \right) + o_p(1) \\ &= \chi_p^2. \end{aligned}$$

Hence we establish a theorem as:

THEOREM 1. *Assume $\{Z_i(t)\}$ are uniformly bounded by a constant. Then $-2 \log R(\beta_*)$ converges in distribution to χ_p^2, where χ_p^2 is a chi-square distribution with degrees of freedom p.*

According to this theorem, an asymptotic $100(1-\alpha)\%$ empirical likelihood confidence region for β is thus constructed as

$$\mathcal{R}_2 = \{\beta : -2 \log R \leq \chi_p^2(\alpha)\}, \tag{13}$$

where $\chi_p^2(\alpha)$ is defined before.

3 SIMULATIONS

A small-scale simulation is conducted to compare the performance of the empirical likelihood procedure with the normal approximation procedures. In Chen and Cheng (2004), their simulations were conducted for relatively large sample size with relatively small proportion of censoring. In order to compare the results, we adopt a similar simulation setup as theirs. That is, we consider two covariates for each subject, Z_1 and Z_2, respectively, with Z_1 being

Table 1: Summary of simulation studies: 95% nominal coverage probabilities of normal approximation and empirical likelihood method

			$\beta_* = (0,0)^{\mathrm{T}}$		$\beta_* = (1,1)^{\mathrm{T}}$	
n	Censoring %	Method	Z_1	Z_2	Z_1	Z_2
50	25%	Normal	0.873	0.882	0.865	0.852
50	25%	EL	0.932	0.937	0.940	0.935
50	50%	Normal	0.797	0.804	0.811	0.817
50	50%	EL	0.917	0.922	0.934	0.925
200	25%	Normal	0.952	0.944	0.942	0.958
200	25%	EL	0.957	0.955	0.956	0.947
200	50%	Normal	0.957	0.942	0.938	0.947
200	50%	EL	0.944	0.952	0.949	0.933

Normal, normal approximation method; EL, empirical likelihood method.

binary of 0 and 1 and Z_2 being uniform on [0,1]. The baseline mean residual life function is $t+1$, corresponding to a Pareto distribution with survival function of $1/(t+1)^2$. Failure times are generated according to model (1), with true parameters of β to be $(0,0)^{\mathrm{T}}$ and $(1,1)^{\mathrm{T}}$, respectively. Independent censoring times are generated from uniform on $[o,c]$, with different c selected to result in 25% and 50% of censoring, respectively. The sample size for each simulation is 50, and 200, representing relatively small and large samples, respectively. The simulation results are tabulated in Table (1). Each entry of the table is based on 1,000 simulated data sets. As shown in the table, both of the methods work reasonably well with right coverage probabilities of 95% when sample size is relatively large. But for sample size, the normal approximation method apparently has relatively larger under-coverage, while the empirical likelihood method has better coverage.

4 Discussion

In this short note, we use an empirical likelihood method to construct confidence regions for the parameters in the proportional mean residual life model. This method is shown to be relatively more accurate in coverage probabilities in small sample size, compared with the normal approximation method in Chen and Cheng (2004). As seen in the development, the empirical likelihood method was applied to the estimating equations proposed by Chen and Cheng (2004). In fact, an empirical likelihood estimator for β can be obtained as $\widetilde{\beta} = \text{argmax}_\beta\{R(\beta)\}$, which can be further shown

$$n^{1/2}(\widetilde{\beta} - \beta_*) \xrightarrow{\mathcal{D}} N(0, A^{-1}VA^{-1}).$$

In addition, empirical likelihood method can be extended to the weighted version of estimating equations straightforwardly as well. The efficient estimator of β would be obtained by choosing the optimal weight function.

In fact, the empirical likelihood method can be used to construct confidence regions with an alternative approach, although it involves estimation of censoring distribution. Consider a synthetic variable $\widetilde{T}(G,t) = S_c(t)X\Delta/S_c(X)$, for $t > 0$, where $S_c(\cdot)$ is the survival function of censoring distribution. Then

$$\begin{aligned} E\{\widetilde{T}(t;S_c) \mid X > t; Z\} &= E_T\left[E_C\left\{\left.\frac{S_c(t)TI(C \ge T)}{S_c(T)}\right| C > t; Z\right\}\middle| T > t; Z\right] \\ &= E_T\left[\left\{\frac{S_c(t)TS_c(T \mid C > t)}{S_c(T)}\right\}\middle| T > t; Z\right] = E(T \mid T > t; Z). \end{aligned}$$

Thus, the following estimating equations can be used to estimate $m_0(t)$ and β jointly:

$$\sum_{i=1}^{n} Y_i(t)\left\{\widetilde{T}_i(t;\widehat{S}_c) - m_0(t)\exp(\beta^{\mathrm{T}}Z_i)\right\} = 0, \tag{14}$$

$$\sum_{i=1}^{n} \int_0^\tau Y_i(t)Z_i\left\{\widetilde{T}_i(t;\widehat{S}_c) - m_0(t)\exp(\beta^{\mathrm{T}}Z_i)\right\} dt = 0, \tag{15}$$

where $\widehat{S}_c$ is some consistent estimator of S_c, such as the Kaplan-Meier estimator when the censoring is considered as homogeneous. By plugging in (15) with (14), thus the following estimating equations can be used to estimate β:

$$\sum_{i=1}^{n} \int_0^\tau Y_i(t)\widetilde{T}_i(t;\widehat{S}_c)\left\{Z_i - \overline{Z}(t)\right\} dt = 0.$$

Thus similar empirical likelihood method proposed previously should apply to construct alternative confidence regions. When censoring is heterogenous across individual subjects, more model assumptions are then needed to use this approach.

References

Chen, Y. Q. & Cheng, S. (2004), Semiparametric regression analysis of mean residual life with censored survival data, *Biometrika*, in press.

Cox, D. R. (1972), Regression models and life-tables (with discussion), *J. R. Statist. Soc.* B, **34**, 187–220.

Maguluri, G. & Zhang, C. H. (1994), Estimation in the mean residual life regression model, *J. Roy. Statist. Soc. Ser. B*, **56**, 477–489.

Oakes, D. & Dasu, T. (1990), A note on residual life, *Biometrika* **77**, 409–410.

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

Owen, A. (2001), *Empirical Likelihood*, New York: Chapman & Hall/CRC.

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