

Rank Estimators Versus Least Square Estimators for Estimating the Parameters of Semiparametric Accelerated Failure Time Model



Mostafa Karimi* and Noor Akma Ibrahim

Department of Mathematics, University Putra Malaysia, Malaysia

Submission: July 17, 2018; **Published:** February 27, 2019

***Corresponding author:** Mostafa Karimi, Department of Mathematics, Institute for Mathematical Research, University Putra Malaysia, Serdang, Malaysia

Abstract

Rank-based method and least square approach are the most common techniques for estimating the regression parameters of accelerated failure time model. In this paper, both inference procedures are considered, their advantages and disadvantages are explained, and their similarities and differences are discussed.

Keywords: Accelerated failure time model; Rank-based inference; Least square method; Semiparametric method; Censored data; Linear regression; Biostatistics

Introduction

Accelerated failure time model is an appealing regression model to biostatistics researchers due to its simple interpretation [1]. Estimating the regression parameters of the model through parametric methods is quite challenging in the presence of censored observations [2]. In such cases, semiparametric approaches are very common. Two main semiparametric methods for estimating the unknown parameters of the model are rank-based method [3], and least square method [4]. In this paper, both inference procedures are briefly explained and their main theoretical and computational aspects are considered. Both approaches are also compared and their advantages and disadvantages are discussed. The main focus of this study is on investigating the similarities and differences of two methods in theory as well as their performance in applications.

Inference procedures

Accelerated failure time model: For the i^{th} subject of a random sample of n subjects let T_i denote the failure time, C_i denote the censoring time, and Z_i denote the $p \times 1$ vector of corresponding covariates. Assume that conditional on covariates Z_i , failure times T_i and censoring times C_i are independent. The accelerated failure time model takes the form

$$\text{Log}(T_i) = \beta'Z_i + \epsilon_i$$

Where β is a p -vector of unknown model parameters, and ϵ_i are the error terms of the model for $i = 1, \dots, n$ with a common distribution function F which is unspecified [5]. The

data consists of $(\tilde{T}_i, \delta_i, Z_i)$ where \tilde{T}_i is the minimum of T_i and C_i , and δ_i is 1 if $T_i \leq C_i$ and 0 otherwise. The introduced model is a semiparametric linear regression model which relates the log-transformed failure times to the covariates.

Rank estimators

Define $\tilde{e}_i(b) = \text{Log}(\tilde{T}_i) - b'Z_i$, $\tilde{Y}_i(t; b) = I\{\tilde{e}_i(b) \geq t\}$ where $I\{\cdot\}$ is the indicator function, and $S^{(a)}(t; b) = n^{-1} \sum_{i=1}^n \tilde{Y}_i(t; b)(Z_i)^a$ for $a = 0, 1$. The weighted log-rank estimating function for the unknown parameter β is given by

$$U_\phi(b) = \sum_{i=1}^n \delta_i \phi(\tilde{e}_i(b), b) \{Z_i - \bar{Z}(\tilde{e}_i(b), b)\}$$

Where $\bar{Z}(e_i(b), b) = S^{(1)}(e_i(b); b) / S^{(0)}(e_i(b); b)$ and $\phi(e_i(b), b)$ is a weight function. The estimating function correspond to Gehan [6] if $\phi(e_i(b), b) = S^{(0)}(e_i(b); b)$ and log-rank if $S^{(0)}(e_i(b); b) = 1$ [7]. Let $\hat{\beta}_R$ denote the rank estimator for the unknown parameter of the model which is the solution of $\{U_\phi(b) = 0\}$. For estimating the unknown parameters of the model Jin et al. [3] proposed an iterative algorithm on the basis of the general weighted estimating function. The algorithm at its k th iteration is given by

$$\hat{\beta}_R^{(k)} = \arg \min_b L_R(b, \hat{\beta}_R^{(k-1)})$$

Where,

$$L_R(b, \hat{\beta}) = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \delta_i \phi(e_i(\hat{\beta}), \hat{\beta}) / S^{(0)}(e_i(\hat{\beta}); \hat{\beta}) \{e_i(b) - e_i(\hat{\beta})\} I\{e_i(b) \geq e_j(\hat{\beta})\}$$

According to Jin et al. [3] the rank estimator $\hat{\beta}_R^{(k)}$ is asymptotically normal for any k .

Least square estimators

When there is no censored observations the least square estimator of the unknown model parameters is obtained by solving the following estimating equation:

$$\sum_{j=1}^n (Z_i - \bar{Z})(\text{Log}(T_i) - b'Z_i) = 0$$

This estimating equation cannot be used when data contains censored observations since the actual value of T_i is unknown for subject i when $\delta_i = 0$. For obtaining the least square estimators in the presence of censored data Jin et al. [4] proposed an iterative algorithm which at its k^{th} iteration is given by

$$\hat{\beta}_S^{(k)} = L_S \left(\hat{\beta}_S^{(k-1)} \right)$$

Where

$$L_S(b) = \left\{ \sum_{j=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})' \right\}^{-1} \left\{ \sum_{j=1}^n (Z_i - \bar{Z})(\hat{Y}_i(b) - \bar{Y}(b))' \right\}$$

In this equation, $\bar{Y}(b) = n^{-1} \sum_{i=1}^n \hat{Y}_i(b)$ and $\hat{Y}_i(b) = E(\text{Log}(T_i) | \tilde{t}_i, \delta_i, Z_i)$ which is proposed by Buckley & James [8] and can be approximated by

$$\hat{Y}_i(b) = \delta_i \text{Log}(\tilde{t}_i) + (1 - \delta_i) \left\{ \frac{\int_{\tilde{t}_i(b)}^{\infty} u d\hat{F}(u)}{1 - F(\tilde{t}_i(b))} + b'Z_i \right\}$$

Where \hat{F} is the Kaplan–Meier estimator of F . According to Jin et al. [4] the least square estimator $\hat{\beta}_S^{(k)}$ is asymptotically normal if the initial value $\hat{\beta}_S^{(0)}$ is asymptotically normal.

Discussion

Both the rank-based method and the least square approach are semiparametric inference procedures since the probability distribution of error terms of the model is completely unknown. One advantage of rank-based inference over the least square method is that it does not involve estimating the distribution of the error terms, while obtaining least square estimators requires the Kaplan–Meier estimator of the distribution of the error terms. This makes the least square method and its corresponding algorithm more complicated than the rank-based method, both theoretically and computationally. Note that, both algorithms need a consistent estimator of the model parameter such as Gehan estimator for their initial values. Thus, the least square approach requires to obtain a rank estimator prior to the computational stage of its associated algorithm. In addition, it has been established that rank estimators are always asymptotically normal [9,10] while the asymptotic normality of least square estimators strongly depend on the asymptotic normality of the initial value of their

corresponding algorithm. However, the results of the simulation studies by Jin et al. [4] illustrated that there was no significant difference between the efficiency of rank estimators and least square estimators. More precisely, the rank estimators were slightly more efficient under extreme-value error, and the least square estimators were slightly more efficient under logistic and normal errors.

Conclusion

For estimating the regression parameters of semiparametric accelerated failure time model both rank estimators and least square estimators are common. From a theoretical point of view, rank-based inference procedure involves less technical difficulties since it does not require estimating the probability distribution of the error terms while least square approach involves Kaplan–Meier estimator of the distribution of the error terms. Moreover, the asymptotic normality of rank estimators does not depend on the distribution of the initial value of its associated algorithm. In application, the results of simulation studies show that there is no significant difference between the efficiency of rank estimators and least square estimators. Therefore, in studies that researcher is free to choose between these two methods rank estimators are definitely more recommended than least square estimators.

References

1. Karimi M, Shariat A (2017) Semiparametric Accelerated Failure Time Model as a New Approach for Health Science Studies. Iran J Public Health 46(11): 1594-1595.
2. Karimi M, Ibrahim NA, Bakar MR, Arasan J (2017) Rank-based inference for the accelerated failure time model in the presence of interval censored data. Numerical Algebra, Control & Optimization 7(1): 107-112.
3. Jin Z, Lin DY, Wei LJ, Ying Z (2003) Rank-based inference for the accelerated failure time model. Biometrika 90(2): 341-353.
4. Jin Z, Lin DY, Ying Z (2006) On least-squares regression with censored data. Biometrika 93(1): 147-161.
5. Kalbfleisch JD, Prentice RL (2011) The statistical analysis of failure time data. John Wiley & Sons, New York, USA.
6. Gehan EA (1965) A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52(1-2): 203-224.
7. Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 50(3): 163-170.
8. Buckley J, James I (1979) Linear regression with censored data. Biometrika 66(3): 429-436.
9. Tsiatis AA (1990) Estimating regression parameters using linear rank tests for censored data. The Annals of Statistics 1: 354-372.
10. Ying Z (1993) A large sample study of rank estimation for censored regression data. The Annals of Statistics 1: 76-99.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/BBOAJ.2019.09.555757](https://doi.org/10.19080/BBOAJ.2019.09.555757)

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

<https://juniperpublishers.com/online-submission.php>