Proceedings of the Korean Nuclear Society Autumn Meeting

Seoul, Korea, October 2001

Investigation of MLE in nonparametric estimation methods of reliability function

Kwang-Won Ahn, Yoonik Kim and Chang-Hyun Chung

Seoul National University

San 56-1 Shilim-dong Kwanak-gu

Seoul, 151-742, Korea

Kil Yoo Kim

Korea Atomic Energy Research Institute

Taejon, Korea, 306-600 Integrated Safety Assessment Team

Abstract

There have been lots of trials to estimate a reliability function. In the ESReDA 20th seminar, a new method in nonparametric way was proposed. The major point of that paper is how to use censored data efficiently.

Generally there are three kinds of approach to estimate a reliability function in nonparametric way, i.e., Reduced Sample Method, Actuarial Method and Product-Limit (PL) Method. The above three methods have some limits. So we suggest an advanced method that reflects censored information more efficiently.

In many instances there will be a unique maximum likelihood estimator (MLE) of an unknown parameter, and often it may be obtained by the process of differentiation. It is well known that the three methods generally used to estimate a reliability function in nonparametric way have maximum likelihood estimators that are uniquely exist. So, MLE of the new method is derived in this study. The procedure to calculate a MLE is similar just like that of PL-estimator. The difference of the two is that in the new method, the mass (or weight) of each point has an influence of the others but the mass in PL-estimator not.

I. Introduction

It is important to estimate correct reliability that reflects the information of the censored data, when we use the non-parametric method. Generally, there are several methods e. g, to estimate reliability in non-parametric ways, reduced sample method, actuarial method and Kaplan-Meier (Product-Limit) method and each of the estimator has MLE.

If the life time data of a unit come from several different tests or the life test truncated or terminated before all the units fail or the failed units are replaced good units, the MLE provides the most powerful and flexible technique for the estimation of parameters, although the computations are somewhat involved.

So, in this paper, the comparisons are conducted between the MLE of the above three estimators and the MLE of the new one. The procedure to calculate a MLE of new method is similar just like that of PL-estimator. The difference of the two is that in the new method, the mass of each point has an influence of the others but the mass in PL-estimator not

II. Generalized Maximum Likelihood Estimator

Assuming that the observation $\underset{\sim}{x}$ has a probability measure P_θ that satisfies

$$dP_\theta(\underset{\sim}{x}) = f_\theta(\underset{\sim}{x})d\mu(\underset{\sim}{x}) , \tag{1}$$

where $\mu(\underset{\sim}{x})$ is a dominating measure for the class $\{P_\theta\}$. Getting the maximum likelihood estimator of θ involves maximizing the likelihood

$$L(\theta) = f_\theta(\underset{\sim}{x}) . \tag{2}$$

Here assuming that the observation has a probability measure P_F that depends on the unknown distribution function F. The class $\{P_F\}$ has no dominating measure so we need a more general definition of maximum likelihood.

Kiefer and Wolfowitz suggest the following definition. Let $\mathbf{P} = \{P\}$ be a class of probability measures. For the elements P_1 and P_2 in $\mathbf{P}$, define

$$f(\underset{\sim}{x}; P_1, P_2) = \frac{dP_1(\underset{\sim}{x})}{d(P_1 + P_2)} , \tag{3}$$

the Radon-Nikodym derivative of P_1 with respect to $P_1 + P_2$. Define the probability measure $\hat{P}$ to be a generalized maximum likelihood estimator if

$$f(x; \hat{p}, p) \geq f(x; p, \hat{p}) \tag{4}$$

for any element $P \in \mathbf{P}$.

III. MLE under actuarial assumption

At the start of any interval x to x+1 time units after the initial event, there are two groups of patients:

(1) l_{1x} individuals who will be at risk for entire interval, (2) l_{2x} persons who, if they do not die beforehand, will be lost to observation or withdrawn alive during that interval.

Let q_x denote the probability of dying sometime during this interval. Then, under the actuarial assumption, the probability that one of the l_{2x} patients will die before his time of loss or withdrawal is, on the average, $q_x / 2$. Thus, the likelihood that the total number of deaths observed, d_x consists of d_{1x} from the first group and d_{2x} from the second is proportional to

$$L(q) = q^{d_1} (1-q)^{l_1-d_1} (\frac{q}{2})^{d_2} (1-\frac{q}{2})^{l_2-d_2} \tag{5}$$

l_1 and l_2 cannot be distinguished from the data, nor can d_1 and d_2. Only their totals $l = l_1 + l_2$ and $d = d_1 + d_2$ can be observed, as well as those actuarial lost or withdrawn alive, $w = l_2 - d_2$. Since, $l_1 - d_1 = l - l_2 - (d - d_2) = l - d - w$, we may rewrite

$$L(q) = \left(\frac{1}{2}\right)^{d_2} \left[q^d (1-q)^{l-d-w} (1-\frac{q}{2})^w \right] \tag{6}$$

Differentiating $\log L$ with respect to q and setting the result equal to zero, the MLE of q may be written in the form

$$\hat{q} = \frac{d}{l - w/(2-\hat{q})} \tag{7}$$

$$\hat{q} = \frac{2l - w + d - \sqrt{(2l - w + d)^2 - 8ld}}{2l} \tag{8}$$

IV. MLE of PL-estimator

The Kaplan-Meier PL estimator gives the GMLE of F. The proof proceeds as follows (assuming no ties).

If a probability measure $\hat{P}$ gives positive probability to $\underset{\sim}{x}$ and P does not gives positive probability to $\underset{\sim}{x}$, then $f(x; p, \hat{p}) = 0$. Thus to check for $P \in \mathbf{P}$ it is sufficient to check it for those P with $P\{\underset{\sim}{x}\} > 0$ and in this case reduces to

$$\hat{P}\{\underset{\sim}{x}\} > P\{\underset{\sim}{x}\} . \tag{9}$$

Since $\hat{S}$ puts positive mass on the point $\underset{\sim}{x} = ((y_1, \delta_1), \Lambda\ , (y_n, \delta_n))$, we need only consider probability measures P which put positive mass on this point and show that $\hat{S}$ maximizes

$P\{((y_1,\delta_1),\Lambda\ ,(y_n,\delta_n))\}$. For any such P

$$L = P\{((y_1,\delta_1),\Lambda\ ,(y_n,\delta_n))\}$$
$$= \prod_{i=1}^{n} P\{T = y_{(i)}\}^{\delta_{(i)}} P\{T > y_{(i)}\}^{1-\delta_{(i)}} . \tag{10}$$

Let P assign probability p_i to the half-open interval $[y_{(i)}, y_{(i+1)})$, where $y_{(n+1)} = +\infty$. For fixed $p_1, p_2, \Lambda\ , p_n$ the likelihood L is maximized by setting $P\{T = y_{(i)}\} = p_i$ if $\delta_{(i)} = 1$. If $\delta_{(i)} = 0$, then L is maximized by setting $P\{y_{(i)} < T < y_{(i+1)}\} = p_i$. Thus for fixed $p_1, p_2, \Lambda\ , p_n$ the maximum value for L is

$$\prod_{i=1}^{n} p_i^{\delta_{(i)}} \left(\sum_{j=i}^{n} p_j \right)^{1-\delta_{(i)}} . \tag{11}$$

Let $\lambda_i = \frac{p_i}{\sum_{j=i}^{n} p_j}$, $i = 1,\Lambda\ , n$.

Then, since $1-\lambda_i = \frac{\sum_{j=i+1}^{n} p_j}{\sum_{j=i}^{n} p_j}$ and $\sum_{j=1}^{n} p_j = 1$, we have $\sum_{j=i}^{n} p_j = \prod_{j=1}^{i-1}(1-\lambda_j)$, and since $\lambda_n = 1$, we get

$$\prod_{i=1}^{n} p_i^{\delta_{(i)}} \left(\sum_{j=i}^{n} p_j \right)^{1-\delta_{(i)}} = \prod_{i=1}^{n} \lambda_i^{\delta_{(i)}} \prod_{j=1}^{i-1}(1-\lambda_j) = \prod_{i=1}^{n-1} \lambda_i^{\delta_{(i)}} (1-\lambda_i)^{n-i} . \tag{12}$$

It is well known from binomial sampling theory that each product is maximized by

$$\hat{\lambda}_i = \frac{\delta_{(i)}}{n-i+\delta_{(i)}} = \frac{\delta_{(i)}}{n-i+1} . \tag{13}$$

$$\hat{p}_i = \hat{\lambda}_i \left(\sum_{j=i}^{n} \hat{p}_j \right) = \hat{\lambda}_i \prod_{j=1}^{i-1}(1-\hat{\lambda}_j) = \frac{\delta_{(i)}}{n-i+1} \prod_{j=1}^{i-1}\left(1 - \frac{\delta_{(j)}}{n-j+1}\right). \tag{14}$$

This is corresponds to $\hat{S}$. The argument for ties is identical.

V. The advanced algorithm of computing the reliability

We introduced another method of computing the survival function. At first, we should formulize the PL estimator as the basic mass of our estimator. Assume no ties.

$n = total\ \#\ of\ \ observations.$

$y_{(1)} < y_{(2)} \cdots < y_{(n)}$: *ordered statistics of observations*

$y_1 < y_2, \cdots, < y_j \ (i = 1, 2, \cdots, j): \textit{observations which censoring censoring occur}$

$\textit{mass of } y_{(k)} = w_{(k)}$

$$w_{(k)} = \frac{1}{n} \ (\textit{mass at start}) \tag{15}$$

$$w_{(k)} = \frac{1}{n} \ (k < i_1) \tag{16}$$

$$w_{(k)} = 0 \ \ (k = i_1) \tag{17}$$

After first redistribution,

$$w_{(k)} = \frac{1}{n} + \frac{1}{n - i_1} \cdot \frac{1}{n} \ \ (k > i_1) \tag{18}$$

$$w_{(k)} = 0 \ \ (k = i_2)\,. \tag{19}$$

After second redistribution,

$$w_{(k)} = \frac{1}{n} + \frac{1}{n - i_1} \cdot (\frac{1}{n}) + \frac{1}{n - i_2}\{\frac{1}{n} + \frac{1}{n - i_1} \cdot (\frac{1}{n})\} \ \ (k > i_2) \tag{20}$$

. . . .

After k-th redistribution, the mass of $y_{(k)} (= w_{(k)})$ is defined as following equations.

$$w_{(k)} = w_{(k-1)} + \frac{1}{n - i_h} \cdot w_{(k-1)} \ \ (k > i_h) \tag{21}$$

Therefore, the mass after last redistribution is defined as following equations.

$$w_{(k)} = \frac{1}{n} \ (k < i_1) \tag{22}$$

$$w_{(k)} = 0 \ \ (k = i) \tag{23}$$

$$w_{(k)} = w_{(k-1)} + \frac{1}{n - i_h} w_{(k-1)} \ \ (i_h < k < i_{h+1}) \tag{24}$$

$$w_{(k)} = w_{(k-1)} + \frac{1}{n - i_j} w_{(k-1)} \ (k > i_j) \tag{25}$$

If we consider the ties, the results are the same as following equations.

$$w_{(k)} = \frac{d_1}{n} \ (k < i_1) \tag{26}$$

$$w_{(k)} = 0 \ \ (k = i) \tag{27}$$

$$w_{(k)} = w_{(k-1)} + \frac{d_{i_k}}{n - i_h} w_{(k-1)} \ \ (i_h < k < i_{h+1}) \tag{28}$$

$$w_{(k)} = w_{(k-1)} + \frac{d_{i_j}}{n - i_j} w_{(k-1)} \ (k > i_j) \tag{29}$$

$d_i = \textit{the \# of } i - th \textit{ observations}$

Until now, we derive a mass function, which will be used in the following theorem. We use this mass function when we redistribute to the right of censoring point. This theorem will support the PL-estimator. Because Kaplan-Meier (PL) estimator is rational if the survival tendency is fixed. But in the real world, the events do not occur with same tendency. In this theorem, we distribute a same mass at first. This assumption will be sufficient in the statistical point of view.

F(last observation)=1

S(last observation)=0

But it is rational considering the weighting of mass, which is located in the right of censoring point when we redistribute the mass of censoring point to the right. From now, let's derive a formula of estimating a survival function, reflecting a tendency of data. The purpose of this paper is to estimate a correct survival function. Therefore we develop a model that is more realistic.

$w_{(i,k)} = mass\ of\ \ y_{(k)}\ \ according\ \ to\ the\ i-th\ redistribution$

$d_k = \#\ of\ \ ties\ at\ \ y_{(k)}$

$$w_{(i_0,k)} = \frac{d_k}{n} \tag{30}$$

$$\begin{aligned} w_{(i_j,k)} &= w_{(i_{j-1},k)} && (k < i_j) \\ &= 0 && (k = i_j) \\ &= w_{(i_{j-1},k)} + w_{i_j} \cdot \frac{w_{(k)}}{\sum_{k>i_j}^{n} w_{(k)}} && (k > i_j) \end{aligned} \tag{31}$$

VI. MLE of the advanced algorithm

Just like a MLE of PL-estimator, the followings are the MLE of MS estimator.

$$\prod_{i=1}^{n} w_{(i_j,k)}{}^{\delta_{(i)}} \left(\sum_{j'=i}^{n} w_{(j'_j,k)} \right)^{1-\delta_{(i)}} . \tag{32}$$

Let $\lambda_i = \dfrac{w_{(i_j,k)}}{\sum_{j'=i}^{n} w_{(j'_j,k)}}$, $\quad i = 1, \Lambda\ , n$.

Then, since $1-\lambda_i = \dfrac{\sum_{j'=i+1}^{n} w_{(j'_j,k)}}{\sum_{j'=i}^{n} w_{(j'_j,k)}}$ and $\sum_{j'=1}^{n} w_{(j'_j,k)} = 1$, we have $\sum_{j'=i}^{n} w_{(j'_j,k)} = \prod_{j'=1}^{i-1} (1-\lambda_{j'})$, and since

$\lambda_n = 1$, we get $\prod_{i=1}^{n} w_{(i_j,k)}{}^{\delta_{(i)}} \left(\sum_{j'=i}^{n} w_{(j'_j,k)} \right)^{1-\delta_{(i)}} = \prod_{i=1}^{n} \lambda_i{}^{\delta_{(i)}} \prod_{j'=1}^{i-1} (1-\lambda_{j'}) = \prod_{i=1}^{n-1} \lambda_i{}^{\delta_{(i)}} (1-\lambda_i)^{n-i}$. (33)

It is well known from binomial sampling theory that each product is maximized by

$$\hat{\lambda}_i = \frac{\delta_{(i)}}{n-i+\delta_{(i)}} = \frac{\delta_{(i)}}{n-i+1}. \tag{34}$$

$$\hat{w}_{(i_j,k)} = \hat{\lambda}_i \left(\sum_{j'=i}^{n} \hat{w}_{(j'_j,k)} \right) = \hat{\lambda}_i \prod_{j'=1}^{i-1} (1-\hat{\lambda}_{j'}) = \frac{\delta_{(i)}}{n-i+1} \prod_{j=1}^{i-1} \left(1 - \frac{\delta_{(j)}}{n-j+1} \right). \tag{35}$$

This is corresponds to $\hat{S}$. The argument for ties is identical.

VII. Summary and conclusion

There are three kinds of approach to estimate a reliability function in nonparametric way, i.e., Reduced Sample Method, Actuarial Method and Product-Limit (PL) Method. The above three methods have some limits as follow. In the Reduced Sample Method, first, the censored data are ignored and calculation is performed using only the observed data. Therefore much information about the censored data is lost. Secondly, in the Actuarial Method, it does not seem to be a correct assumption that the half of the censored data drops out at the censored point. This method not only loses but also misuses much information. In any aspect, these two methods are not more rational than the Product-Limit Method. The Product-Limit Method has been widely used because the logic is more reasonable than those of the others, in which if censoring occurs, we are to redistribute the weight of the censored point equally to the remaining points. But there also exists a limit of redistribution of equal values in the last method. So we suggest an advanced method that reflects censored information more efficiently.

The paper publish contains a new algorithm of calculating a reliability function. There are two steps in the algorithm: first, to calculate reliability functions using three nonparametric methods such as reduced sample method, actuarial method and PL-estimator; second, using the result of the above methods to recalculate reliability functions. So, the idea of this method is similar to that of the PL-estimator. The difference between PL-estimator and the method proposed occurs when the weight of censored point is redistributed. The one redistribute the weight of censored point equally but the other redistributes the weight of censored point considering the weight of each point that are located at the right of censored point.

Using the above methods, we have shown that the MLEs of actuarial method, PL-estimator and advanced algorithm. For the case of actuarial method, the procedure is very simple and the results are expressed as a compact form. The only difference is the sample size so called “effective sample size”. In the actuarial method, the censoring information is reflected on the effective sample size. So, the analysis of the result is natural. For the case of PL-estimator and new one, it is necessary to derive a GMLE (Generalized Maximum Likelihood Estimator). After we derive a GMLE, the MLEs of PL-estimator and new one are obtained just following the procedure of it. But the difference of the two exists. The difference of the two is that in the new method, the mass (or weight) of each point has an influence of the others but the mass in PL-estimator not.

References

1. Kwang-Won Ahn, Yoonik Kim, Chang-Hyun Chung and Kil Yoo Kim, "The non-parametric method estimating the survival or the reliability function considering the information of censored data." *ESREDA* (2001).

2. Kwang-Won Ahn, Yoonik Kim, Chang-Hyun Chung and Kil Yoo Kim, "The Development of A New Algorithm to Calculate A Survival Function in Non-parametric Ways." *Proceedings of the Korean Nuclear Society Spring Meeting* (2001).

3. Berkson and Gage, *Proc. Staff Meet*. Mayo Clin. (1950).

4. Cutler and Ederer, J. chronic Dis. (1958).

5. Breslow and Crowley, Ann. Stat. (1974).

6. Thomas and Grunkemeier, JASA. (1975).

7. Kaplan and Meier, JASA. (1958).

8. Embury et al. , West. J. Med.(1977).

9. Peterson, JASA. (1977).

10. Efron, *Proc. Fifth Berkeley Symp. IV*. (1967).

11. Rupert G. Miller, Jr. Gail Gong. Alvaro Munoz. *Survival Analysis*. Stanford University. (1981).

12. David G, Kleinbaum. *Survival analysis: a self-learning text*. Springer-verlag New York, Inc. (1995).

13. Eugene K. Harris, Adelin Albert. *Survivorship analysis for clinical studies*. Marcel Dekker, Inc. (1991).

14. Chiang, *Stochastic Processes in Biostatistics* (1968), chapter 9.

15. Leiderman et al., Nature. (1973).