Journal of Statistical Theory & Methods
1996, Vol. 7, pp 59~68

Weighted Estimation of Survival Curves for NBU Class Based on Censored Data[1]

Sangbock Lee[2]

Abstract In this paper, we consider how to estimate New Better Than Used (NBU) survival curves from randomly right censored data. We propose several possible NBU estimators and study their properties. Numerical studies indicate that the proposed estimators are appropriate in practical use. Some useful examples are presented.

Keywords : Bias, Cumulative hazard function, Nonparametric estimation, NBU Reliabilty function.

1. Introduction

In reliability theory and survival analysis of physical, biological, and other systems, the concept of aging has been found to be very useful. Selection of appropriate models of life distributions on the basis of specific aging criteria is an important step when performing reliability analysis. Barlow & Proschan(1975) and Bryson & Siddiqui(1969), among others, presented a detailed treatment of life distributions. They discussed aging concepts such as Increasing Failure Rate(IFR),Increasing Failure Rate Average(IFRA), Decreasing Mean Residual Life time(DMRL), New Better Than Used(NBU), and so on. We shall consider the problem of nonparametric estimating the NBU life distribution.

For a life distribution F of the lifetime X of a subject, let S=1-F and H=-log S denote its survival(or reliability) function and cumulative hazard function, respectively. A distribution function F with F(0)=0 is said to be NBU if

$$H(x+y) \geq H(x)+H(y) \qquad (1.1)$$

for all $x \geq 0$ and $y \geq 0$. Nonparametric inference on H is usually based on Nelson

---

[1] This research was supported in part by Korea Science & Engineering Foundation in 1994.
[2] Department of statistics, Catholic University of Taegu-Hyosung,712-702, Korea.

(1972) estimator H_n,

$$H_n(t) = \sum_{i=1}^{n} \frac{I(Z_i \le t, \delta_i = 1)}{n-i+1} \tag{1.2}$$

where $Z_i = \min(X_i, C_i)$, $\delta_i = I(X_i \le C_i)$, C_i is the censoring time variable with a distribution function G, i=1,..., n, and I(A) is the indicator function of a set A. Throughout in this paper assume without loss of generality that $Z_1 \le Z_2 \le ... \le Z_n$. We shall restrict the estimation problem to compact interval [0,T] ,where T is any point with F(T)<1 and G(T)<1.

Under model (1.1), we are interested in estimating cumulative hazard function

$$H(t) = \int_0^t h(u)du$$

and survival function S(t)= exp{-H(t)} for a subject. Note that IFR or IFRA distribution belongs to the NBU class due to Barlow & Proschan(1975) .

Practically, it is reasonable to require that if the true survival function satisfies the NBU property, then its estimator should also satisfy the same property. The product-limit estimator $\hat{S}_n$ (Kaplan & Meier,1958), being a step function and a biased estimator of S, will not be NBU with probability tending to one as Boyles & Samaniego(1984). Restricting the estimators to be a member of a specific subclass often results in more efficient than that of the unrestricted estimator by Reneau & Samaniego (1990). We shall consider that some weighted nonparametric estimators of the NBU survival curves satisfy NBU property from similar ideas of Chang & Rao(1993).

Since an NBU survival distribution has a superadditive cumulative hazard function, it is natural to consider

$$H_L(x) = Max\left[\inf_{0 \le x \le x+y \le T}\{H_n(x+y) - H_n(y)\}, H_n(x)\right] \tag{1.3}$$

as our estimator of H(x). The estimator H_L is the estimator of Wang(1987) and also that of Boyles & Samaniego(1984) for the uncensored case. H_L is also proved a strongly uniformly consistent NBU estimator of H by Wang(1987) under censored data model. Thus we can use $\hat{S}_u(x) = \exp\{-H_L(x)\}$ as an estimator of S(x) in the NBU class. A tipical defect of $\hat{S}_u$ is its bias and mean square error(MSE). Since $H_L(x) \le H_n(x)$, $\hat{S}_u$ is a positively bias in finite samples. In simulation studies, we showed that for small and moderate sample sizes the bias and MSE of $\hat{S}_u$ can be quite large when compared those of $\hat{S}_n$.

Since (1.1) can be rewritten as $H(x) \ge H(t) + H(x-t)$ for all $x \ge 0$ and $0 \le t \le x$, an alternative NBU estimator $\hat{S}_l(x) = \exp\{-H_M(x)\}$ for S can be obtained using

$$H_M(x) = Min\left[\sup_{0 \le t \le x}\{H_n(x-t) + H_n(t)\}, H_n(x)\right] \tag{1.4}$$

The estimator $\hat{S}_l$ is the same as that of Reneau, Samaniego, & Boyles(1988) developed for the uncensored case. We shall study the consistency H_M, and derived from it the largest NBU survival estimator $\le \hat{S}_n$. The fact that $\hat{S}_l$ is bounded by $\hat{S}_n$ implies that it intends under biased estimator of S. As we shall point out in section 3, the bias and MSE of $\hat{S}_l$ can be quite large than those of $\hat{S}_n$ in small to moderate sample sizes.

In view of the fact that $\hat{S}_u$ and $\hat{S}_l$ are over and under biased estimators of $\hat{S}$, respectively, it is naturally to consider methods of adjusting these estimators to reduce their bias. In this paper, we shall present the results of of our researches of the feasibility of such adjustments to H_M and H_L. Correspondingly, S is estimated by the adjusting cumulative hazard function estimators. Chang & Rao(1992) and Reneau & Samaniego(1990) discussed methods of adjusting survival function estimator for the NBU class of a specified age. Chang & Rao(1993) discussed such methods for the NBU case.

It is easily shown that our methods of adjusting the NBU cumulative hazard function estimators are convenient for the censored case. These methods provide NBU survival curve estimators that are not only NBU but also perform very well, when compared to $\hat{S}_n$ and other previous NBU estimators.

In section 2, we shall show the stronly consistency and NBU property of (1.4). Using (1.3) and (1.4), we shall give two families of adjusted estimator for the estimation of an NBU survival function. We shall also show that the estimators based on either of the two methods satisfy the consistensy property. In section 3, we shall present the results of a simulation study to comapre the biases and MSE's of three different survival function estimators selected from Weibull classes.

Finally, Some examples and discussions are given.

2. Description of the NBU estimators

Through the remainder of this paper, we assume that the survival function S belongs to NBU class. The Following lemma asserts strong convergency of the estemator (1.4).

Lemma 2.1. If a distribution function F is NBU on $(0,\infty)$, then (1.4) converges strongly to the cumalative hazard function H(x) for all x in [0,T] with probability one.

Proof. Since the estimator (1.2) converges strong uniformly to the cumulative hazard function H on [0,T] by Shorack & Wellner(1986,p.304), (1.4) also converges strongly to H with probability one.

Next, we consider that the estimator (1.4) also belongs to NBU as the same idea of Chang & Rao(1993) for uncensored case.

Theorem 2.1. If a distribution function F is NBU on $(0,\infty)$, then (1.4) belongs to NBU class on [0,T].

Proof. By the Glivenko-Cantelli theorem, there exists a null set M such that for all ω in Ω/M,

$$\sup_x |H_n(x)-H(x)| \to 0$$

Our notation suppresses the fact that the sequences we will consider depend on ω. Since $H_n(0)=0$ and $H_n(T)$ is bounded for all n, we need establish only convergence of $H_n(x)$ for x in [0,T]. Let such an x be fixed. For any ω in $\Omega/_M$, we have

$$H_M(x)=\sup_{0\le t\le x}\{H_n(x-t)+H_n(t)\}\ge H_n(x)$$

which implies that, by Lemma 2.1,

$$\underline{\lim}_{n\to\infty} H_M(x)\ge \lim_{n\to\infty} H_n(x) = H(x),$$

We wish to show that for any $\varepsilon>0$,

$$\overline{\lim}_{n\to\infty} H_M(x)=\overline{\lim}_{n\to\infty}\sup_{0\le t\le x}\{H_n(x-t)+H_n(t)\} \le H(x)+\varepsilon \tag{2.1}$$

Suppose (2.1) fails for some fixed ε_0. Then there exist a sequence $\{t_{ni}\}$such that

$$\{H_n(x-t_{n_i})+H_n(t_{n_i})\}>H(x)+\varepsilon_0 \text{ for all } i \tag{2.2}$$

The sequence $\{t_{ni}\}$ is bounded above by x in [0,T]. We will assume, with loss of generality, that $t_{n_i}\to t^*$, for otherwise, we could reconstruct a convergent subsequence satisfying (2.2). We further assume, again without loss of generality, that $\{t_{ni}\}$converges monotonically to t^*, and that $t_{n_i}\to t^*$ from below. We will have a contradiction to (2.2). If $t_{n_i}\to t^*$ from below, then

$$\{H_n(x-t_{n_i})+H_n(t_{n_i})\}\to\{H(x-t^{*-})+H(t^{*-})\}$$

where $H(t^-)=\lim_{z\to t^-}H(z)F$. But

$$\{H(x-t^{*-})+H(t^{*-})\}\le H(x)$$

for otherwise, there exist an $\varepsilon > 0$ such that

$$\{H(x-t^* - \varepsilon) + H(t^* - \varepsilon)\} > H(x)$$

contracting the fact that F is NBU. So, for sufficiently large n_i,

$$\{H_n(x - t_{n_i}) + H_n(t_{n_i})\} < H(x) + \varepsilon_0$$

contracting (2.2)

Corollary 2.1. $\hat{S}_l(x) = \exp\{-H_M(x)\}$ is the largest NBU estimator among smaller than or equal to $\hat{S}_n(x)$ on [0,T].
Proof. It is easily verified that $\hat{S}_l$ satisfies (1.1) from Theorem (2.1) and that $\hat{S}_l(x)$ is the largest estimator among $\leq \hat{S}_n(x)$ for all x in [0,T], by the fact that $H_M(x) \geq H_n(x)$ and due to Breslow & Crowly(1974) .

Since $\hat{S}_u$ and $\hat{S}_l$ can be considered the upper and lower bounds to the biased Kaplan-Meier estimator $\hat{S}_n$ it is reasonable to attempt to form an improved estimator either by averaging H_M and H_L. Accordingly, we propose the following two classes of estimators:

$$\hat{S}_{w\alpha}(x) = \exp\{-H_M(x)^{\alpha} H_L(x)^{(1-\alpha)}\} \quad 0 \leq \alpha \leq 1. \tag{2.3}$$

Equation (2.3) also belongs to the NBU class using the fact that both H_M and H_L are in the NBU class.

Theorem 2.2. $\hat{S}_{w\alpha}$, converges uniformly to S on [0,T].
Proof. By Slutsky Theorem and Skrokhod Theorem Theorem, those are easily derived.

3. A simulation study

As similar as the simulation design of Chang & Rao(1993), for each values, random samples of size 20 were simulated from each of several survival functions selected from a NBU family characterized by the following failure rate functions:

The Weibull(W) family with failure rate

$$h(x;\theta) = (1+\theta)x^{\theta} \quad x \geq 0, \ \theta \geq 0 \tag{3.1}$$

and censoring times from expontial distributions.

Note that $\theta = 0$ in the W families corresponds to an exponential survival function with failure rate equal to 1.

The results are summarized in Table 1 for $\theta = 0.5$ with 1000 replications. A weighed value of α is 0.6 which is calculated from PC techniques as following minimax criterion:

$$\pi(\alpha) = \sup_{x \geq 0} \left| \hat{S}_{w\alpha}(x) - S(x) \right| \quad (3.2)$$

In view of Tables, the magnitudes of the standard biases of the adjusted $\hat{S}_{w\alpha}$ were less than the magnitudes of the standard biases of $\hat{S}_{KA}$ and $\hat{S}_{NA}$ in most cases. And in most cases the MSE's of the adjusted estimates $\hat{S}_{w\alpha}$ are smaller than those of $\hat{S}_{KA}$ and $\hat{S}_{NA}$.

4. Examples and Discussions

The first data set in Table 2 reproduced Barlow & Campo(1975), contain 107 failure times of right rear brakes on D9G-66A caterpillar tractors. Doksum & Yandell(1984) showed that the survival function of these data belongs strongly to the IFR class(and therefor an NBU) by plotting methods. For the requirement of the model (1.1) for censored data, a plot toto check (1.1) using the product-limit

Table 1. MSE's and Bias for the estimators

(1) n=20, censoring ratio 0%

S \ X		0.2	0.5	0.8	1.7	2.0
$\hat{S}_{KA}(x)$	Bias	-.0017	-.0008	-.0081	-.0125	-.0125
	MSE	.0046	.0124	.0165	.0126	.0072
$\hat{S}_{NA}(x)$	Bias	.0006	.0082	.0101	.0542	.0687
	MSE	.0043	.0117	.0153	.0126	.0119
$\hat{S}_{W\alpha}(x)$	Bias	-.0004	.0046	.0027	-.0007	-.0057
	MSE	.0044	.0116	.0157	.0143	.0087

(2) n=20, censoring ratio 10%

S \ X		0.2	0.5	0.8	1.7	2.0
$\hat{S}_{KA}(x)$	Bias	.0024	.0024	-.0021	-.0024	-.0059
	MSE	.0036	.0127	.0194	.0122	.0077
$\hat{S}_{NA}(x)$	Bias	.0046	.0111	.0155	.0548	.0664
	MSE	.0034	.0120	.0179	.0124	.0114
$\hat{S}_{W\alpha}(x)$	Bias	.0037	.0076	.0081	.0102	.0013
	MSE	.0035	.0123	.0187	.0139	.0091

Table 1. (Continued.)
(3) n=20, censoring ratio 20%

S \ X		0.2	0.5	0.8	1.7	2.0
$\hat{S}_{KA}(x)$	Bias	-.0041	-.0062	-.0111	-.0346	-.0278
	MSE	.0047	.0146	.0224	.0159	.0079
$\hat{S}_{NA}(x)$	Bias	-.0016	.0041	.0125	.0837	.1104
	MSE	.0044	.0135	.0193	.0207	.0244
$\hat{S}_{W\alpha}(x)$	Bias	-.0026	.0000	.0019	-.0260	-.0240
	MSE	.0045	.0139	.0211	.0178	.0091

estimator, $\hat{S}_n$, can be develop by Chang & Rao(1993) and Hollande et al.(1985) as followings;

$$d_{ij} = \hat{S}_n(Z_i)\hat{S}_n(Z_j) - \hat{S}_n(Z_i + Z_j) \geq 0,\ 1 \leq i,\ j \leq n. \tag{4.1}$$

Therefore, a plot of d_{ij} versus $t_{ij} = Z_i + Z_j$ present a means for graphical check of the NBU assumption. Figure 1 shows such a plot for the tractor data. The plot indicates a strong NBU property as like as that of Chang & Rao(1993) for uncensored case. We present a plot of three survival curve estimators $\hat{S}_{KA}$, $\hat{S}_{NA}$ and $\hat{S}_{W\alpha}$ in Figure 2.

Table 2. Failure times of right rear brakes(* censored data)

56 83 104 116* 244 305 429 452* 453 503 552 614 661* 673 683 685 753 763 806 834 838 862 897* 904 981* 1007 1008 1049* 1069 1107 1125 1141* 1153 1154 1193* 1201 1253* 1313 1329 1347 1454 1464* 1490 1491 1532 1549 1568 1574 1586 1599 1608 1723* 1769 1795 1927 1957 2005 2010 2016* 2022 2037 2065 2096 2139 2150 2156 2160* 2190 2210 2220 2248 2285* 2325 2337* 2351 2437 2454 2546* 2565 2584 2624 2675 2701 2755 2887 2879* 2922 2986 3092 3160 3185 3191 3439 3617 3638 3756* 3826 3995 4007 4159 4300* 4487 5074 5579 5623 6869* 7739

As a second example, we consider 43 patients suffering from chronic granulosytic leukemia in Bryson & Siddiqui(1969). NBU survival function assumptions are also reasonable for these data in view of Figure 3. Figure 4 is shown survival curvers for leukemia data as the same as those of Example 1.

Table 3. Data of patients from chronic granulosytic leukemia(* censored data)

7 47 58* 74 177 232 273 285 317 429 440* 445 455 468 495* 497 532 571 579 581 650 702 715* 779 881 900 930 960 968 1077* 1109 1314 1367* 1534 1712 1784 1877 1886 2045 2056 2260* 2429 2509

Figure 1. Data of Tractor

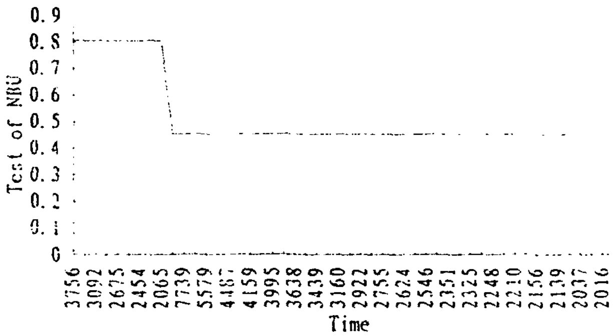


Figure 2. Survival Curves

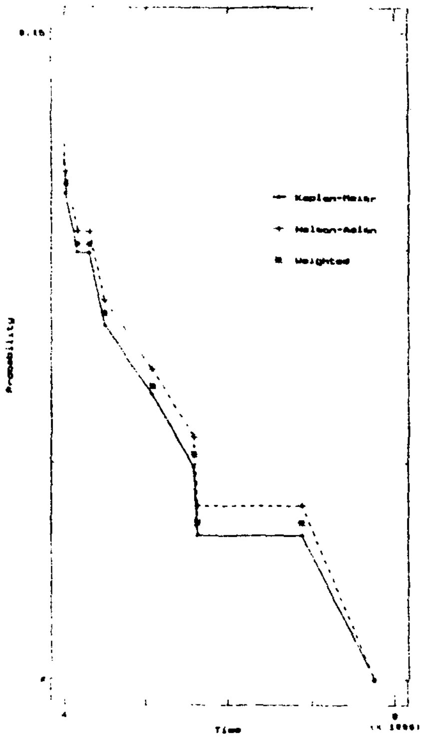


Figure 3. Chronic granulosytic leukemia

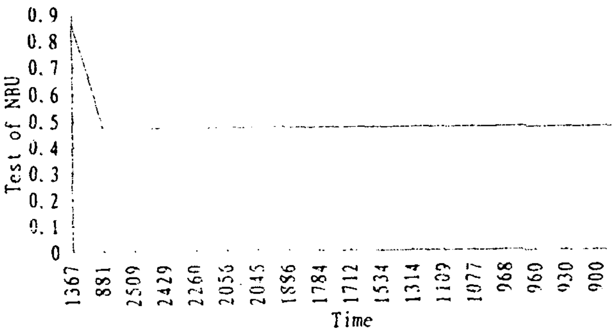
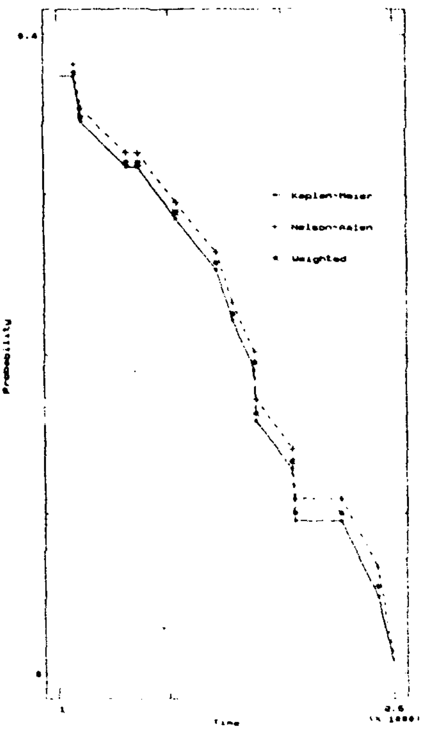


Figure 4. Survival Curves



Acknowledgement

Thanks to referees for useful comments.

References

Barlow, R. E., and Campo, R. (1975), *Total Time on Test Processes and Applications to Failure Data Analysis, Reliability and Fault Tree Analysis*, Philadelphia :Society for Industrial and Applied Mathematics, pp.451-481.

Barlow, R. E., and Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing*, New York:Holt, Rinehart & Winston.

Boyles, R. A., and Samaniego, F. J. (1984), *Estimating a Survival Curve When New is Better Than Used*, Operations Research, 32, 732-740.

Breslow, N., and Crowley, J. (1974), *A Large Sample Study of the Life Table and Product Limit Estimates under Random Censorship*, Annals of Statistics, 2, 437-453.

Bryson, M. C.,and Siddiqui, M. M. (1969), *Some Criteria for Aging*, Journal of erican Statistical Association, 64, 1472-1483.

Chang, M. N., and Rao, P. V.(1992), *On the Estimation of a Survival Function When New is Better Than Used of a Specified Age*, Journal of Noparametric Statistics, 2, 45-58.

Chang, M. N., and Rao, P. V. (1993), *Improved Estimation of Survival Functions in the New-Better-Than-Used Class*, Technometrics, 35, 192-203.

Doksum, K. A., and Yandell, Y. S. (1984), *Tests for Exponentiality*, in Handbook of Statistics(Vol. 4), eds. P.R.Krishnaiah and P.K. Sen, NewYork: Elsevier, pp.591-611.

Efron, B. (1982), *The Jackknife, the Bootsrap and Other Resamplings*, Philadelphia: Society fot Industrial and Applied Mathematics.

Hollande, M.,Park, D. H.,and Proschan, F. (1985), *Testing whether New is Better Than Used of a Specified Age, with Randomly Censored Data*, The Canadian Journal of Statistics, 13, 45-52.

Kaplan, E. L., and Meier, P. (1958), *Nonparametric estimation from incomplete observations*, Journal of the American Statistical Association, 53, 457-481.

Nelson, W. (1972), *Theory and Applications of Hazard Plotting for Censored Failure Data*, Technometrics, 14, 945-966.

Reneau, D. M., and Samaniego, F. J. (1990), *Estimating the Survival Curve When New is Better Than Used of a Specified Age*, Journal of the American Statistical Association, 85, 123-131.

Reneau, D. M., Samaniego, F. J., and Boyles, R. A. (1988), *Estimating a Survival*

Curve When New is Better Than Used with respect to a Distinguished Set, Technical Report 146, University of California, Davis, Div. of Statistics.

Shorack, G. R., and Wellner, J. A.(1986), *Empirical Processes with Applications to Statistics*, John Wiley & Sons:New York, pp.304-306.

Wang, J. (1987), *Estimating IFRA and NBU Survival Curves Based on Censored Data*, Scandinavian Journal of Statistics, 14, 199-210.