# Modified Weighted Kaplan-Meier Estimator

Mohammad Shafiq
Department of Statistics
University of Peshawar
Peshawar, Pakistan

Shuhrat Shah
Department of Statistics
University of Peshawar
Peshawar, Pakistan

Alamgir
Department of Statistics
University of Peshawar
Peshawar, Pakistan

## Abstract

In many medical studies majority of the study subjects do not reach to the event of interest during the study period. In such situations survival probabilities can be estimated for censored observation by Kaplan Meier estimator. However in case of heavy censoring these estimates are biased and over estimate the survival probabilities. For heavy censoring a new method was proposed (Bahrawar Jan, 2005) to estimate the survival probabilities by weighting the censored observations by non-censoring rate. But the main defect in this weighted method is that it gives zero weight to the last censored observation. To over come this difficulty a new weight is proposed which also gives a non-zero weight to the last censored observation.

## 1. Introduction

### Survival Analysis

Survival analysis is a branch of statistics which deals with death in biological organisms and failure in mechanical systems. This is called reliability theory or reliability analysis in engineering. Death or failure is called an "event" in the survival analysis. So models of death or failure are generically termed “time-to-event models” (Mara 2005).

### Censoring

Censored data are those observations whose time to the occurrence of the event has not been observed completely. When some individuals are still alive at the end of the study or analysis, it means that the event of interest, usually death, has not occurred. In this case we only know that time taken for event is greater than time of study. This is called ‘right censoring’. For some individuals the time of entry into the control group is not known. For example if the ‘time to event’ is the time from contracting HIV until death. Since the time of contracting HIV is not exactly known, this is called ‘left censoring’. Where as neither the time of entry nor the event time is known for individual, is called ‘interval censoring’ (Svetlana 2002).

## Survival Function

The object of primary interest is the survival function, conventionally denoted by S, which is defined as:

$$S(t) = Pr(T > t)$$

where t is some time, T is the time of death, and "Pr" stands for probability. The survival function is the probability that the patient will survive till time t. Survival probability is usually assumed to approach zero as age increases. i.e., $S(t) \to 0$ as $t \to \infty$ (Johnson 1980/1999).

## Lifetime Distribution Function

The lifetime distribution function, conventionally denoted by *F(t)*, is defined as the complement of the survival function, i. e.

$$F(t) = Pr(T \leq t) = 1 - S(t)$$

and the derivative of F(t) (i.e., the density function of the lifetime distribution) is conventionally denoted by f(t), given by

$$f(t) = - S^{/}(t) \qquad \text{where}$$

f(t) is sometimes called the event density; it is the rate of death or failure events per unit time (Johnson 1980/1999).

## Hazard Function and Cumulative Hazard Function

The hazard function, conventionally denoted by λ, is defined as the event rate at time t conditional on survival until time t or later,

$$\lambda(t) = \lim_{dt \to 0} \frac{\Pr(t < T < t + dt / T > t)}{dt}$$

the numerator of this expression is the conditional probability that the event will occur in the interval (t, t+dt) given that it has not occurred before, and the denominator is the width of the interval.

## Kaplan-Meier Product- Limit Estimator

The Kaplan Meier estimator is the limit of the life table estimator when intervals are taken so small that only at most one observation occurs within an interval. This estimator gives a maximum likelihood estimate.

Let d(x) denote the number of deaths at time x. Generally it is either 0 or 1, but we allow the possibility of tied survival time in which case d(x) may be greater than 1. Let n(x) denotes the number of individuals at risk just prior to time x. Then the Kaplan Meier estimate can be expressed as

$$KM = S(t) = \Pi_{x \leq t} [1 - d(x)/n(x)]$$

Note that in the notation above the product changes only at times where we observe deaths, or in general events (Kaplan 1958). If the last observation is censored, the Kaplan-Meier estimator fails to estimate the tails of the survival function. Further, this method over estimates the survival distribution in case of heavy censoring (Breslow 1991).

This Kaplan Meier is further modified by Mr. Bahrawar Jan in his Ph.D thesis known as Weighted Kaplan Meier estimates, Bahrawar Jan (2004) for heavy censoring.

## Weighted Kaplan Meier

The Weighted Kaplan Meier is defined as

$$S^*(t) = \Pi_{x\leq t} W_j\,[1 - d(x)/n(x)]$$

Where

$$W_j = \left(\frac{n_j - c_j}{n_j}\right) \text{ is known as non-censoring rate.}$$

The greatest defect in the Weighted Kaplan Meier is that it gives zero weight to the last censored observation. So a new weight function is proposed to remove the deficiency. The proposed estimator is Modified Weighted Kaplan Meier Estimator.

## Proposed Modified Weighted Kaplan Meier.

The proposed Modified Weighted Kaplan Meier Estimator is

$$S^{**}(t) = \Pi_{x\leq t} W_j\,[1 - d(x)/n(x)]$$

Where the weight function is

$$W_j = 1\text{-Sin}\left(\frac{C_j * P_j}{n_j}\right) \text{ is known as non-censoring rate}$$

## Example

The proposed method is also supported by the analysis of real data set “Stanford Heart Transplant data” (Kalbflesch and Printice, 1980) which is classical survival data set.

| Time | Total No. of Deaths | Total No. of censored | No. at Risk | Prob of survival | Weighted Kaplan-Meier weights | Proposed .weight | Kaplan-Meier.S(t) | Weighted Kaplan-Meier.S(t) | Proposed .S(t) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 103 | 0.99029 | 1 | 1 | 0.99029 | 0.99029 | 0.99029 |
| 2 | 3 | 0 | 102 | 0.97059 | 1 | 1 | 0.961166 | 0.961166 | 0.961166 |
| 3 | 3 | 0 | 99 | 0.9697 | 1 | 1 | 0.932042 | 0.932042 | 0.932042 |
| 5 | 2 | 0 | 96 | 0.97917 | 1 | 1 | 0.912628 | 0.912628 | 0.912628 |
| 6 | 2 | 0 | 94 | 0.97872 | 1 | 1 | 0.893207 | 0.893207 | 0.893207 |
| 8 | 1 | 0 | 92 | 0.98913 | 1 | 1 | 0.883498 | 0.883498 | 0.883498 |
| 9 | 1 | 0 | 91 | 0.98901 | 1 | 1 | 0.873788 | 0.873788 | 0.873788 |
| 11 | 0 | 1 | 90 | 1 | 0.98889 | 0.98889 | 0.873788 | 0.864081 | 0.864081 |
| 12 | 1 | 0 | 89 | 0.98876 | 1 | 1 | 0.863967 | 0.854368 | 0.854368 |
| 16 | 3 | 0 | 88 | 0.96591 | 1 | 1 | 0.834514 | 0.825243 | 0.825243 |
| 17 | 1 | 0 | 85 | 0.98824 | 1 | 1 | 0.8247 | 0.815538 | 0.815538 |
| 18 | 1 | 0 | 84 | 0.9881 | 1 | 1 | 0.814886 | 0.805833 | 0.805833 |
| 21 | 2 | 0 | 83 | 0.9759 | 1 | 1 | 0.795248 | 0.786412 | 0.786412 |
| 28 | 1 | 0 | 81 | 0.98765 | 1 | 1 | 0.785426 | 0.7767 | 0.7767 |
| 30 | 1 | 0 | 80 | 0.9875 | 1 | 1 | 0.775609 | 0.766992 | 0.766992 |

| Time | Total No. of Deaths | Total No. of censored | No. at Risk | Prob of survival | Weighted Kaplan-Meier weights | Proposed .weight | Kaplan-Meier.S(t) | Weighted Kaplan-Meier.S(t) | Proposed .S(t) |
|---|---|---|---|---|---|---|---|---|---|
| 31 | 0 | 1 | 79 | 1 | 0.98734 | 0.98734 | 0.775609 | 0.757281 | 0.757281 |
| 32 | 1 | 0 | 78 | 0.98718 | 1 | 1 | 0.765665 | 0.747573 | 0.747573 |
| 35 | 1 | 0 | 77 | 0.98701 | 1 | 1 | 0.755719 | 0.737862 | 0.737862 |
| 36 | 1 | 0 | 76 | 0.98684 | 1 | 1 | 0.745774 | 0.728152 | 0.728152 |
| 37 | 1 | 0 | 75 | 0.98667 | 1 | 1 | 0.735833 | 0.718446 | 0.718446 |
| 39 | 1 | 1 | 74 | 0.98649 | 0.98649 | 0.98667 | 0.725892 | 0.699164 | 0.699292 |
| 40 | 2 | 0 | 72 | 0.97222 | 1 | 1 | 0.705726 | 0.679742 | 0.679866 |
| 43 | 1 | 0 | 70 | 0.98571 | 1 | 1 | 0.695642 | 0.670028 | 0.67015 |
| 45 | 1 | 0 | 69 | 0.98551 | 1 | 1 | 0.685562 | 0.660319 | 0.66044 |
| 50 | 1 | 0 | 68 | 0.98529 | 1 | 1 | 0.675477 | 0.650606 | 0.650725 |
| 51 | 1 | 0 | 67 | 0.98507 | 1 | 1 | 0.665392 | 0.640892 | 0.641009 |
| 53 | 1 | 0 | 66 | 0.98485 | 1 | 1 | 0.655312 | 0.631183 | 0.631298 |
| 58 | 1 | 0 | 65 | 0.98462 | 1 | 1 | 0.645233 | 0.621475 | 0.621589 |
| 61 | 1 | 0 | 64 | 0.98438 | 1 | 1 | 0.635154 | 0.611768 | 0.61188 |
| 66 | 1 | 0 | 63 | 0.98413 | 1 | 1 | 0.625074 | 0.602059 | 0.602169 |
| 68 | 2 | 0 | 62 | 0.96774 | 1 | 1 | 0.60491 | 0.582637 | 0.582743 |
| 69 | 1 | 0 | 60 | 0.98333 | 1 | 1 | 0.594826 | 0.572924 | 0.573029 |
| 72 | 2 | 0 | 59 | 0.9661 | 1 | 1 | 0.574661 | 0.553502 | 0.553603 |
| 77 | 1 | 0 | 57 | 0.98246 | 1 | 1 | 0.564582 | 0.543794 | 0.543893 |
| 78 | 1 | 0 | 56 | 0.98214 | 1 | 1 | 0.554498 | 0.534081 | 0.534179 |
| 80 | 1 | 0 | 55 | 0.98182 | 1 | 1 | 0.544417 | 0.524372 | 0.524468 |
| 81 | 1 | 0 | 54 | 0.98148 | 1 | 1 | 0.534335 | 0.51466 | 0.514754 |
| 85 | 1 | 0 | 53 | 0.98113 | 1 | 1 | 0.524252 | 0.504949 | 0.505041 |
| 90 | 1 | 0 | 52 | 0.98077 | 1 | 1 | 0.51417 | 0.495239 | 0.495329 |
| 96 | 1 | 0 | 51 | 0.98039 | 1 | 1 | 0.504088 | 0.485527 | 0.485616 |
| 100 | 1 | 0 | 50 | 0.98 | 1 | 1 | 0.494006 | 0.475817 | 0.475903 |
| 102 | 1 | 0 | 49 | 0.97959 | 1 | 1 | 0.483923 | 0.466105 | 0.46619 |
| 109 | 0 | 1 | 48 | 1 | 0.97917 | 0.97917 | 0.483923 | 0.456396 | 0.456479 |
| 110 | 1 | 0 | 47 | 0.97872 | 1 | 1 | 0.473625 | 0.446684 | 0.446766 |
| 131 | 0 | 1 | 46 | 1 | 0.97826 | 0.97826 | 0.473625 | 0.436973 | 0.437053 |
| 149 | 1 | 0 | 45 | 0.97778 | 1 | 1 | 0.463101 | 0.427264 | 0.427342 |
| 153 | 1 | 0 | 44 | 0.97727 | 1 | 1 | 0.452575 | 0.417552 | 0.417628 |
| 165 | 1 | 0 | 43 | 0.97674 | 1 | 1 | 0.442048 | 0.40784 | 0.407914 |
| 180 | 0 | 1 | 42 | 1 | 0.97619 | 0.97619 | 0.442048 | 0.398129 | 0.398202 |
| 186 | 1 | 0 | 41 | 0.97561 | 1 | 1 | 0.431267 | 0.388419 | 0.388489 |
| 188 | 1 | 0 | 40 | 0.975 | 1 | 1 | 0.420485 | 0.378708 | 0.378777 |
| 207 | 1 | 0 | 39 | 0.97436 | 1 | 1 | 0.409704 | 0.368998 | 0.369065 |
| 219 | 1 | 0 | 38 | 0.97368 | 1 | 1 | 0.39892 | 0.359286 | 0.359352 |
| 263 | 1 | 0 | 37 | 0.97297 | 1 | 1 | 0.388138 | 0.349575 | 0.349638 |
| 265 | 0 | 1 | 36 | 1 | 0.97222 | 0.97223 | 0.388138 | 0.339863 | 0.339929 |
| 285 | 2 | 0 | 35 | 0.94286 | 1 | 1 | 0.365959 | 0.320444 | 0.320505 |
| 308 | 1 | 0 | 33 | 0.9697 | 1 | 1 | 0.354871 | 0.310734 | 0.310794 |
| 334 | 1 | 0 | 32 | 0.96875 | 1 | 1 | 0.343781 | 0.301024 | 0.301082 |
| 340 | 1 | 1 | 31 | 0.96774 | 0.96774 | 0.96879 | 0.332691 | 0.281915 | 0.282275 |
| 342 | 1 | 0 | 29 | 0.96552 | 1 | 1 | 0.321219 | 0.272194 | 0.272542 |
| 370 | 0 | 1 | 28 | 1 | 0.96429 | 0.96429 | 0.321219 | 0.262474 | 0.26281 |
| 397 | 0 | 1 | 27 | 1 | 0.96296 | 0.96297 | 0.321219 | 0.252752 | 0.253078 |
| 427 | 0 | 1 | 26 | 1 | 0.96154 | 0.96155 | 0.321219 | 0.243031 | 0.243347 |
| 445 | 0 | 1 | 25 | 1 | 0.96 | 0.96001 | 0.321219 | 0.23331 | 0.233616 |
| 482 | 0 | 1 | 24 | 1 | 0.95833 | 0.95835 | 0.321219 | 0.223588 | 0.223886 |
| 515 | 0 | 1 | 23 | 1 | 0.95652 | 0.95654 | 0.321219 | 0.213867 | 0.214156 |

| Time | Total No. of Deaths | Total No. of censored | No. at Risk | Prob of survival | Weighted Kaplan-Meier weights | Proposed .weight | Kaplan-Meier.S(t) | Weighted Kaplan-Meier.S(t) | Proposed .S(t) |
|---|---|---|---|---|---|---|---|---|---|
| 545 | 0 | 1 | 22 | 1 | 0.95455 | 0.95456 | 0.321219 | 0.204146 | 0.204424 |
| 583 | 1 | 0 | 21 | 0.95238 | 1 | 1 | 0.305923 | 0.194425 | 0.19469 |
| 596 | 0 | 1 | 20 | 1 | 0.95 | 0.95002 | 0.305923 | 0.184704 | 0.184959 |
| 630 | 0 | 1 | 19 | 1 | 0.94737 | 0.94739 | 0.305923 | 0.174983 | 0.175228 |
| 670 | 0 | 1 | 18 | 1 | 0.94444 | 0.94447 | 0.305923 | 0.165261 | 0.165498 |
| 675 | 1 | 0 | 17 | 0.94118 | 1 | 1 | 0.287929 | 0.15554 | 0.155763 |
| 733 | 1 | 0 | 16 | 0.9375 | 1 | 1 | 0.269933 | 0.145819 | 0.146028 |
| 841 | 0 | 1 | 15 | 1 | 0.93333 | 0.93338 | 0.269933 | 0.136097 | 0.1363 |
| 852 | 1 | 0 | 14 | 0.92857 | 1 | 1 | 0.250652 | 0.126376 | 0.126564 |
| 915 | 0 | 1 | 13 | 1 | 0.92308 | 0.92315 | 0.250652 | 0.116655 | 0.116837 |
| 941 | 0 | 1 | 12 | 1 | 0.91667 | 0.91676 | 0.250652 | 0.106934 | 0.107112 |
| 979 | 1 | 0 | 11 | 0.90909 | 1 | 1 | 0.227865 | 0.097213 | 0.097374 |
| 995 | 1 | 0 | 10 | 0.9 | 1 | 1 | 0.205079 | 0.087491 | 0.087637 |
| 1032 | 1 | 0 | 9 | 0.88889 | 1 | 1 | 0.182292 | 0.07777 | 0.0779 |
| 1141 | 0 | 1 | 8 | 1 | 0.875 | 0.87533 | 0.182292 | 0.068049 | 0.068188 |
| 1321 | 0 | 1 | 7 | 1 | 0.85714 | 0.85763 | 0.182292 | 0.058327 | 0.05848 |
| 1386 | 1 | 0 | 6 | 0.83333 | 1 | 1 | 0.15191 | 0.048606 | 0.048733 |
| 1400 | 0 | 1 | 5 | 1 | 0.8 | 0.80133 | 0.15191 | 0.038885 | 0.039051 |
| 1407 | 0 | 1 | 4 | 1 | 0.75 | 0.7526 | 0.15191 | 0.029164 | 0.02939 |
| 1571 | 0 | 1 | 3 | 1 | 0.66667 | 0.67281 | 0.15191 | 0.019442 | 0.019774 |
| 1586 | 0 | 1 | 2 | 1 | 0.5 | 0.52057 | 0.15191 | 0.009721 | 0.010294 |
| 1799 | 0 | 1 | 1 | 1 | 0 | 0.15853 | 0.15191 | 0 | 0.001632 |

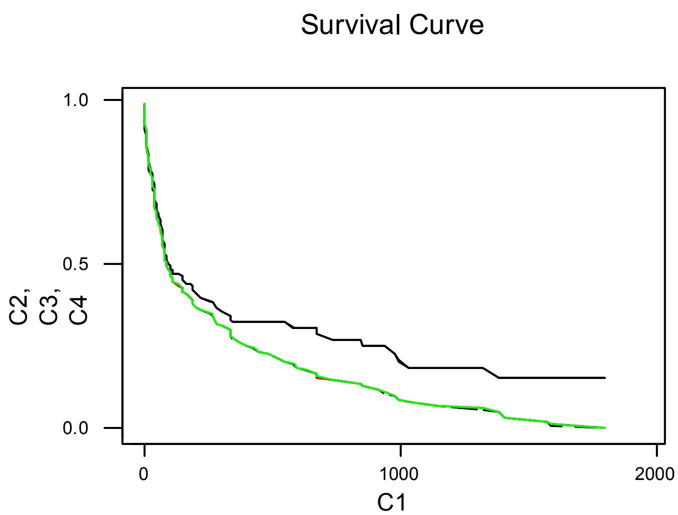


**Figure 1**

In the above example, Time shows the number of months. S(t) shows that the subject will survive up to time t. The Kaplan-Meier estimator gives highest probabilities of survival while the weighted Kaplan-Meier estimator gives small probabilities than Kaplan-Meier but it gives "0" probability of survival to the last censored observation.

Both the proposed Estimator and Weighted Kaplan Meier give same weight to all censored observations. They also give same probability of survival but the important point is that the Weighted Kaplan Meier Estimator gives zero weight to the last observation, which is censored while the proposed Estimator gives it some nonzero weight and has a small probability of survival.

## 2. Conclusion

The proposed weighted function was tested on "Stanford Heart Transplant Data" and was compared with traditional Kaplan-Meier estimator and weighted Kaplan-Meier estimator. It was found that the Kaplan-Meier estimator gave very high probability of survival. This over estimation was controlled by Weighted Kaplan-Meier estimator but the survival probability estimated by this method was zero at last censored observation. The survival probabilities given by proposed estimator were same as given by weighted Kaplan-Meier estimator but the important point is that proposed method gives a non-zero survival probability to the last censored observation.

## References

1. Bahrawar Jan (2004). Improved Inferences in the context of Survival/Failure Time. Ph.D Thesis University of Peshawar.
2. Breslow, N.E (1991). Introduction to Kaplan and Meier (1958) Nonparametric estimation from incomplete observations. In Breakthroughs in Statistics II, S. Kotz and N.L. Johnson (eds), app: 311-318. New York: Springer.
3. Johnson R.E and Johnson M. (1980/1999). Survival Models and Data Analysis. New York: John Wiley & Sons.
4. Kalbflesch, J. D. and Prentice, R. L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.
5. Kaplan, E.L and Meier, P (1958). Non-parametric estimation from incomplete observations. Journal of the American Statistical Association 53:457-481
6. Mara Tableman (2005). Methods of Analyzing survival Data.
7. Svetlana Borovkova (2002). Analysis of Survival Data.