Journal of Statistical Research *https://doi.org/10.47302/jsr.2021550101*
2021, Vol. 55, No. 1, pp. 7-17 *ISSN 0256 - 422 X*

AN ALTERNATIVE TO THE PRODUCT LIMIT METHOD: THE MULTISTATE APPROACH

M. ATAHARUL ISLAM
University of Dhaka, Bangladesh

EDITORIAL NOTE

The plan for this special volume of the Journal of Statistical Research started with Professor Dr. M. Ataharul Islam as one of the editors. Unfortunately, Professor Islam has passed away during the preparation of this volume on December 6, 2020. To honor Professor Islam's contribution to the statistical community we have published this significant methodological work of Islam from 27 years back, originally published in the Journal of Royal Statistical Society, Series D (The Statistician). Permission to re-publish the article in JSR has been granted by the Royal Statistical Society and the John Wiley & Sons, Inc. The original reference to the publication is here: M. Ataharul Islam: AN ALTERNATIVE TO THE PRODUCT LIMIT METHOD: THE MULTISTATE APPROACH, The Statistician 43, 237–245, 1994.

SUMMARY

A multistate alternative to the product limit method is developed. The approach proposed takes into account the censored and uncensored survivorship functions separately. The methods are useful in situations of informative censoring. This approach can be used for unequal probabilities of death for uncensored and censored states. The product limit estimates can be shown to be a special case of the proposed alternative if probabilities of death for uncensored and censored states are assumed to be equal.

Keywords and phrases: Censoring; Intercommunicating states; Multistate approach; Product limit method; Survival probability.

1 Introduction

The product limit method developed by Kaplan and Meier (1958) is one of the most commonly used techniques for estimating survivorship functions for samples of small and moderate sizes with censoring. The asymptotic distribution theory for the estimator has been discussed by Kaplan and Meier (1958), Efron (1967), Breslow and Crowley (1974), Meier (1975) and Peterson (1977). The product limit estimator is consistent and asymptotically unbiased (Kalbfleisch and Prentice, 1980). On the basis of Mantel (1987), Schemper (1991) generalized the product limit method for analysing continuous, multiple- event and non-monotonic processes. Aalen (1976), Matthews (1984) and Davis and Lawrance (1989) studied nonparametric estimators for the cumulative hazard rates in a multiple-decrement model. Aalen (1976) gave a heuristic justification for introducing an estimator that was analogous to the estimator developed by Kaplan and Meier (1958). This paper develops an alternative to the product limit method to take account of the censored and uncensored survivorship functions separately as well as to develop a method that can take into account unequal probabilities of death for uncensored and censored states. The methods are useful in situations of informative censoring. The product limit estimates can be obtained from the same approach under the assumption of equality of probabilities of death.

2 Product Limit Method

Let us consider that n individuals are observed and that the observed times of death or loss (censoring) are ordered as $t_1 < t_2 < \cdots < t_k$. Let us introduce an indicator c_i indicating whether the observed time t_i is for death or loss. Assume that no time is associated with both deaths and losses and let $c_i = 1$ if t_i is for death and $c_i = 0$ if t_i is for loss. Let us also define n_i as the number of individuals alive just before time t_i, d_i as the number of deaths or losses at time t_i, p_i as the probability of surviving the i th interval, given that the individual is alive at the beginning of the interval, and q_i as the probability of dying during the ith interval given that the individual is alive at the beginning of the interval.

If each observed time t_i is associated with a single death or loss then $n_i = n - i + 1$ and the estimate for p_i is

$$\hat{p}_i = \begin{cases} 1 - 1/n_i = (n-i)/(n-i+1) & \text{if} \quad c_i = 1, \\ 1 & \text{if} \quad c_i = 0. \end{cases}$$

Then for any time x after the beginning of the observation the product limit estimate for the survivorship function is

$$\begin{aligned} \hat{P}(x) &= \prod_{t_1 \leqslant x} \hat{p}_i \\ &= \prod_{t_i \leqslant x} \{(n-i)/(n-i+1)\}^{c_i}. \end{aligned} \tag{2.1}$$

For multiple deaths or losses at any time

$$\hat{P}(x) = \prod_{t_i \leqslant x} (1 - d_i/n_i)^{c_i} . \tag{2.2}$$

If deaths and losses are observed to be tied then we assume that the deaths occurred before the losses.

3 Multistate Approach

This section shows an alternative to the product limit method based on a multistate approach. Let us consider a cohort of n_0 individuals beginning the observation at time t_0. All the entries are assumed to be uncensored at time t_0. During each time interval any of the following four outcomes can be expected for any individual:

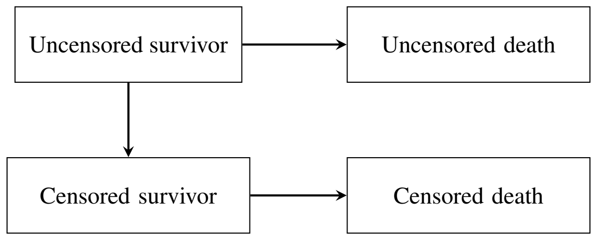
(a) the individual survives as an uncensored survivor,

(b) the individual progresses to death from the state of uncensored survivor,

(c) the individual survives after being lost to follow-up (censored survivor) or

(d) the individual progresses to death from the state of censored survivor.

Outcomes (a) and (b) are observable but (c) and (d) are not observable. The following flow diagram shows the relationship between these states:



In practice the censored survivor or censored death states cannot be distinguished or estimated directly from observed data. However, an indirect estimation is possible from the observed data if we assume that the probabilities of dying for the uncensored and censored individuals are the same for each interval.

For simplicity we assume that $t_0 = 0$. Relationships between these states can be expressed by the following equations for the time interval $(0, x)$

$$n^{\mathrm{u}}(x) = n^{\mathrm{u}}(0) - {}^{\mathrm{u}}d^{\mathrm{c}}(x) - {}^{\mathrm{u}}d^{\mathrm{d}}(x), \tag{3.1}$$

$$n^{\mathrm{c}}(x) = n^{\mathrm{c}}(0) + {}^{\mathrm{u}}d^{\mathrm{c}}(x) - {}^{\mathrm{c}}d^{\mathrm{d}}(x), \tag{3.2}$$

where $n^{\mathrm{u}}(x)$ and $n^{\mathrm{c}}(x)$ are the respective numbers of survivors in uncensored and censored states at time x, ${}^{\mathrm{u}}d^{\mathrm{c}}(x)$ is the total number of individuals moved from an uncensored to a censored state during the time interval $(0, x)$, ${}^{\mathrm{u}}d^{\mathrm{d}}(x)$ is the total number of deaths from the uncensored state during $(0, x)$ and ${}^{\mathrm{c}}d^{\mathrm{d}}(x)$ is the total number of deaths from the censored state during $(0, x)$.

Dividing equations (3.1) and (3.2) by $n^{\mathrm{u}}(0)$ and letting $n^{\mathrm{c}}(0) = 0$, we obtain

$$P^{\mathrm{u}}(x) = 1 - {}^{\mathrm{u}}q^{\mathrm{c}}(x) - {}^{\mathrm{u}}q^{\mathrm{d}}(x) \tag{3.3}$$

$$P^{\mathrm{c}}(x) = {}^{\mathrm{u}}q^{\mathrm{c}}(x) - {}^{\mathrm{c}}\bar{q}^{\mathrm{d}}(x) \tag{3.4}$$

where ${}^{\mathrm{c}}\bar{q}^{\mathrm{d}}(x) = {}^{\mathrm{c}}d^{\mathrm{d}}(x)/n^{\mathrm{u}}(0), P^{\mathrm{u}}(x)$ is the probability of surviving the interval $(0, x)$ in an uncensored state, $P^{\mathrm{c}}(x)$ is the probability of surviving and being in the censored state at time x, ${}^{\mathrm{u}}q^{\mathrm{c}}(x)$ is the probability of moving from an uncensored to a censored state during the interval $(0, x)$ and ${}^{\mathrm{u}}q^{\mathrm{d}}(x)$ is the probability of dying during the interval $(0, x)$ in the uncensored state.

Adding equations (3.3) and (3.4) we obtain the probability of surviving to time x,

$$\begin{aligned} P(x) &= P^{\mathrm{u}}(x) + P^{\mathrm{c}}(x) \\ &= 1 - {}^{\mathrm{u}}q^{\mathrm{d}}(x) - {}^{\mathrm{c}}\bar{q}^{\mathrm{d}}(x). \end{aligned} \tag{3.5}$$

In equations (3.4) and (3.5), ${}^{\mathrm{c}}\bar{q}^{\mathrm{d}}(x)$ is not directly estimable from the observed data. By definition we have

$$\begin{aligned} {}^{\mathrm{c}}\bar{q}^{\mathrm{d}} &= \frac{{}^{\mathrm{c}}d^{\mathrm{d}}(x)}{n^{\mathrm{u}}(0)} \\ &= \frac{1}{n^{\mathrm{u}}(0)} \sum_{t_i \leqslant x} {}^{\mathrm{c}}d_i^{\mathrm{d}}. \end{aligned} \tag{3.6}$$

Let us consider the time intervals (t_i, t_{i+1}) of the survival time. Then we can denote the probability of moving from an uncensored to a censored survivor state during the time interval (t_i, t_{i+1}) by ${}^{\mathrm{u}}q_i^{\mathrm{c}}$, the probability of moving from an uncensored survivor to an uncensored death during the same interval by ${}^{\mathrm{u}}q_i^{\mathrm{d}}$ and the probability of moving from a censored survivor to a censored death by ${}^{\mathrm{c}}q_i^{\mathrm{d}}$. For simplicity let us assume that death does not occur to a censored individual in the interval when the individual enters the censored state.

The assumption implied in the product limit method is that the probability of dying is the same for uncensored and censored individuals. Sometimes bias can be introduced due to this assumption in the estimates of survival probabilities because in many practical problems this is not the case. The approach proposed is flexible in incorporating an adjustment factor to estimate survival probabilities where

(a) censored probabilities of death may be equal to uncensored probabilities of death (case 1) or

(b) censored probabilities of death are not equal to uncensored probabilities of death (case 2).

Case 1 or the conventional product limit method assumption can be shown to be a special case of case 2. Some examples of case 2 are as follows:

(a) censored probabilities of death are less than uncensored probabilities of death for the whole time span under consideration;

(b) censored probabilities of death are greater than uncensored probabilities of death for the whole time span under consideration;

(c) censored probabilities of death are smaller at the early stage but uncensored probabilities of death are smaller at the later stage;

(d) censored probabilities of death are higher at the early stage but uncensored probabilities of death are higher at the later stage, etc.

Assuming that the probability of dying for uncensored and censored individuals are proportional at each time interval then we have ${}^{\mathrm{c}}q_i^{\mathrm{d}} = a_i\,{}^{\mathrm{u}}q_i^{\mathrm{d}}$ where a_i is the adjustment factor for the i th interval. If $a_i = 1$ then it reduces to the assumption of the product limit method (case 1; see Appendix A for examples of exact equivalence), and if $a_i \neq 1$ at least for one i then case 2 assumptions, some examples of which are given above, may be used. Then

$$
\begin{aligned}
{}^{\mathrm{c}}\bar{q}^{\mathrm{d}}(x) &= \frac{1}{n^{\mathrm{u}}(0)} \sum_{t_i \leqslant x} a_i\,{}^{\mathrm{u}}q_i^{\mathrm{d}} n^{\mathrm{c}}\left(t_i\right) \\
&= \sum_{t_i \leqslant x} a_i\,{}^{\mathrm{u}}q_i^{\mathrm{d}} p^{\mathrm{c}}\left(t_i\right). \qquad (3.7)
\end{aligned}
$$

Replacing equation (3.7) in equations (3.4) and (3.5), we obtain

$$
P^{\mathrm{c}}(x) = {}^{\mathrm{u}}q^{\mathrm{c}}(x) - \sum_{t_i < x} a_i\,{}^{\mathrm{u}}q_i^{\mathrm{d}} p^{\mathrm{c}}\left(t_i\right) \qquad (3.8)
$$

$$
\begin{aligned}
P(x) &= P^{\mathrm{u}}(x) + P^{\mathrm{c}}(x) \\
&= 1 - {}^{\mathrm{u}}q^{\mathrm{d}}(x) - \sum_{t_i < x} a_i\,{}^{\mathrm{u}}q_i^{\mathrm{d}} P^{\mathrm{c}}\left(t_i\right). \qquad (3.9)
\end{aligned}
$$

In the multistate approach, equation (3.9) replaces equations (2.1) and (2.2) of the product limit method.

3.1 Estimation

From the flow diagram it is evident that we can directly estimate only the probability of a transition from the state uncensored survivor (u) to states dead (d) and censored survivor (c) during the interval (t_i, t_{i+1}). No observations are available regarding transitions from the censored survivor state to dead. However, it can be shown that the probability of transitions from censored to dead can be obtained indirectly on the basis of estimates for the probability of transitions from uncensored to censored and dead under very restricted assumptions.

During each time interval (t_i, t_{i+1}), we consider that the following may happen to an individual who is surviving at the beginning of the interval:

(a) the individual may move to the absorbing state dead during the interval with probability ${}^{\mathrm{u}}q_i^{\mathrm{d}}$,

(b) the individual may move to the censored state with probability ${}^{\mathrm{u}}q_i^{\mathrm{c}}$ or

(c) the individual may survive the interval in the uncensored state with probability $1 - {}^{\mathrm{u}}q_i^{\mathrm{d}} - {}^{\mathrm{u}}q_i^{\mathrm{c}}$.

We assume that only one event takes place for an individual during an interval. The probability distribution for these transitions follows a multinomial distribution. Then the likelihood function can be shown to be

$$L \propto \left({}^{\mathrm{u}}q_i^{\mathrm{d}}\right)^{{}^{u}d_i^{\mathrm{d}}} \left({}^{\mathrm{u}}q_i^{\mathrm{c}}\right)^{{}^{\mathrm{u}}d_i^{\mathrm{c}}} \left(1 - {}^{\mathrm{u}}q_i^{\mathrm{d}} - {}^{\mathrm{u}}q_i^{\mathrm{c}}\right)^{n_i - {}^{u}d_i^{\mathrm{d}} - {}^{\mathrm{u}}d_i^{\mathrm{c}}}. \tag{3.10}$$

Setting the first derivatives with respect to ${}^{\mathrm{u}}q_i^{\mathrm{d}}$ and ${}^{\mathrm{u}}q_i^{\mathrm{c}}$ equal to 0 we obtain the estimators

$${}^{\mathrm{u}}\hat{q}_i^{\mathrm{d}} = {}^{\mathrm{u}}d_i^{\mathrm{d}} / n_i^{\mathrm{u}} \tag{3.11}$$

and

$${}^{\mathrm{u}}\hat{q}_i^{\mathrm{c}} = {}^{\mathrm{u}}d_i^{\mathrm{c}} / n_i^{\mathrm{u}}. \tag{3.12}$$

As mentioned earlier, we cannot estimate ${}^{\mathrm{c}}q_i^{\mathrm{d}}$ directly from observed data; however, if we assume that ${}^{\mathrm{c}}q_i^{\mathrm{d}} = a_i {}^{\mathrm{u}}q_i^{\mathrm{d}}$, then

$${}^{\mathrm{c}}\hat{q}_i^{\mathrm{d}} = a_i {}^{\mathrm{u}}\hat{q}_i^{\mathrm{d}} = a_i {}^{\mathrm{u}}d_i^{\mathrm{d}} / n_i^{\mathrm{u}} \tag{3.13}$$

can be used as an indirect estimator.

3.2 Variance of estimators

Let us assume that the probability of dying in the censored state is proportional to the probability of dying in the uncensored state during the interval (t_i, t_{i+1}), i.e.

$$q_i^{\mathrm{d}} = {}^{\mathrm{c}}q_i^{\mathrm{d}} = a_i {}^{\mathrm{u}}q_i^{\mathrm{d}},$$

where

$$q_i^{\mathrm{d}} = \frac{d_i^{\mathrm{d}}}{n_i} = \left({}^{\mathrm{u}}d_i^{\mathrm{d}} + {}^{\mathrm{c}}d_i^{\mathrm{d}}\right) / \left(n_i^{\mathrm{u}} + n_i^{\mathrm{c}}\right).$$

For $a_i = 1$ this assumption reduces to the assumption of the product limit method. Let us also assume that the deaths from uncensored and censored states occur independently. Then we can show that the variance for the number of deaths from the uncensored survivor state is

$$\mathrm{var}\left({}^{\mathrm{u}}d_i^{\mathrm{d}}\right) = n_i^{\mathrm{u}} {}^{\mathrm{u}}q_i^{\mathrm{d}} \left(1 - {}^{\mathrm{u}}q_i^{\mathrm{d}}\right), \tag{3.14}$$

and similarly

$$\begin{aligned}\text{var}\left({}^{\text{c}}d_i^{\text{d}}\right) &= n_i^{\text{c}}{}^{\text{c}}q_i^{\text{d}}\left(1-{}^{\text{c}}q_i^{\text{d}}\right)\\ &= n_i^{\text{c}}a_i{}^{\text{u}}q_i^{\text{d}}\left(1-a_i{}^{\text{u}}q_i^{\text{d}}\right).\end{aligned} \tag{3.15}$$

Hence

$$\begin{aligned}\text{var}\left(d_i^{\text{d}}\right) &= \text{var}\left({}^{\text{u}}d_i^{\text{d}}+{}^{\text{c}}d_i^{\text{d}}\right)\\ &= n_iq_i^{\text{d}}\left(1-q_i^{\text{d}}\right),\end{aligned} \tag{3.16}$$

where $n_i = n_i^u + n_i^c$. Similarly we can show that,

$$\text{var}\left(\hat{q}_i^{\text{d}}\right) = q_i^{\text{d}}\left(1-q_i^{\text{d}}\right)/n_i. \tag{3.17}$$

Chiang (1968) discussed these relationships in detail for simple life-tables. With the multinomial distribution we obtain

$$\begin{aligned}\text{var}\left\{\hat{p}^{\text{u}}(x)\right\} &= \left\{1-{}^{\text{u}}q^{\text{d}}(x)-{}^{\text{u}}q^{\text{c}}(x)\right\}\frac{{}^{\text{u}}q^{\text{d}}(x)+{}^{\text{u}}q^{\text{c}}(x)}{n^{\text{u}}(0)}\\ &= P^{\text{u}}(x)^2\sum_{t_i\leqslant x}\frac{{}^{\text{u}}q_i^{\text{d}}+{}^{\text{u}}q_i^{\text{c}}}{1-{}^{\text{u}}q_i^{\text{d}}-{}^{\text{u}}q_i^{\text{c}}}.\end{aligned} \tag{3.18}$$

We also find that $var\{\hat{P}(x)\}$ can be expressed as

$$\begin{aligned}\text{var}\{\hat{P}(x)\} &= \text{var}\left\{1-{}^{\text{u}}\hat{q}^{\text{d}}(x)-{}^{\text{c}}\hat{\hat{q}}^{\text{d}}(x)\right\}\\ &= \text{var}\left\{{}^{\text{u}}\hat{q}^{\text{d}}(x)\right\}+\text{var}\left\{{}^{\text{c}}\hat{\hat{q}}^{\text{d}}(x)\right\}\\ &= \frac{{}^{\text{u}}q^{\text{d}}(x)\left\{1-{}^{\text{u}}q^{\text{d}}(x)\right\}}{n_{\text{u}}(0)}+\frac{\text{var}\left\{\sum_{t_i<x}{}^{\text{c}}d^{\text{d}_1}\right\}}{n_{\text{u}}(0)^2}\\ &= \frac{1}{n_{\text{u}}(0)}\left[{}^{\text{u}}q^{\text{d}}(x)\left\{1-{}^{\text{u}}q^{\text{d}}(x)\right\}+\frac{1}{n_{\text{u}}(0)}\sum_{t_i<x}n_i^{\text{c}}a_i{}^{\text{u}}q_i^{\text{d}}\left(1-a_i{}^{\text{u}}q_i^{\text{d}}\right)\right].\end{aligned} \tag{3.19}$$

If $a_i = 1$ then the variance can be obtained under the assumption of equality of probabilities of death for uncensored and censored states.

4 Examples

To show examples of the alternative approach, we have considered a data set adapted from Miller (1981). On the basis of this data set on lengths of complete remission (in weeks) from acute myelogenous leukaemia, it has been shown that the multistate approach gives exactly the same result as the product limit method under the given conditions. Using the estimates (3.11) and (3.12) in equations (3.3) and (3.8), we obtained the survival probabilities in the uncensored and censored states. For the total survival probability we added the estimates for the uncensored and censored states which can be estimated independently by equation (3.9). The results are displayed in Table 1. We assumed that the proportion of survivors in the censored state at the beginning

Table 1: Probability of surviving in uncensored and censored states by the multistate approach and the product limit method

Time x (weeks)	Multistate method			Product limit method
	$P^{\mathrm{u}}(x)$	$P^{\mathrm{c}}(x)$	$P(x)$	$P(x)$
9	0.909	–	0.909	0.909
13	0.818	–	0.818	0.818
13+	0.727	0.091	0.818	
18	0.636	0.080	0.716	0.716
23	0.545	0.068	0.613	0.614
28+	0.454	0.159	0.613	
31	0.363	0.127	0.490	0.491
34	0.273	0.095	0.368	0.368
45+	0.182	0.186	0.368	
48	0.091	0.093	0.184	0.184
161+	0	0.184	0.184	

of the experiment is 0, i.e. $P^c(0) = 0$. Two different assumptions on censored death probabilities have been used to show the nature of bias incurred by the product limit assumption of equal uncensored and censored deaths in the case of deviations from such an assumption. For this purpose the data set presented in Table 1 was used. The first assumption considers that the probability of death from the censored state is half the probability of death from the uncensored state ($a_i = 0.5$) since the time of censoring (Assumption 1). The second assumption considers that the probability of death from the censored state is 1.5 times the probability of death from the uncensored state ($a_i = 1.5$) since the time of censoring (Assumption 2). These results are summarized in Table 2. Figure 1 compares the survival probabilities at different times on the basis of these assumptions with the survival probabilities estimated in Table 1 on the basis of the assumption of the product limit method, i.e. $a_i = 1$. These differences are quite substantial. The deviations will be even greater for an increase in the number of censored cases.

5 Conclusion

A multistate approach has been proposed in this paper as an alternative to the product limit method for estimating the probability of surviving a time interval. The approach proposed introduces an adjustment factor to take into account unequal probabilities of death from uncensored and censored survivor states. The approach allows a choice of the assumption about the deaths among the censored individuals. The simplest of these assumptions is that the deaths among uncensored and censored individuals are the same, which is the assumption implied in the estimates from the product limit method. This paper considered deviations from such an assumption to

Table 2: Estimates of survival probabilities based on various assumptions on the censored probabilities of death by the multistate approach

Time x	Values of $P(x)$ from the following assumptions:		
(weeks)	Assumption 1	Assumption 2	Product limit
9	0.909	0.909	0.909
13	0.818	0.818	0.818
13+	0.818	0.818	
18	0.722	0.710	0.716
23	0.625	0.604	0.614
28+	0.625	0.604	
31	0.517	0.468	0.491
34	0.407	0.338	0.368
45+	0.407	0.338	
48	0.260	0.130	0.184
161+	0.260	0.130	

show the effect on the estimates for the probability of surviving. This flexibility in the assumptions of censored deaths will allow researchers to incorporate assumptions about such deaths that may vary in many medical problems from the standard product limit assumption. One of the major advantages of the proposed technique is that it can be extended to several intercommunicating states and the probability of surviving for each of those components can be obtained without making the theory complicated.

Acknowledgements

The author is greatly indebted to a referee for his valuable suggestions on the basis of which the main focus of the paper has been shifted. The author is also grateful to the Population Council, University of Dhaka, and the University of Pennsylvania for supporting this work.

A Appendix

The exact equivalence between the product limit method and the alternative proposed for the assumption of equality of probabilities of death for uncensored and censored individuals can be shown by some examples.

A.1 Example 1

Let us consider n ordered survival times, out of which the first individual died at time t_1, the second is lost to follow-up (censored) after time t_2 and the third died at time t_3. We want to estimate $P(x)$ where $t_i \leqslant x$. The

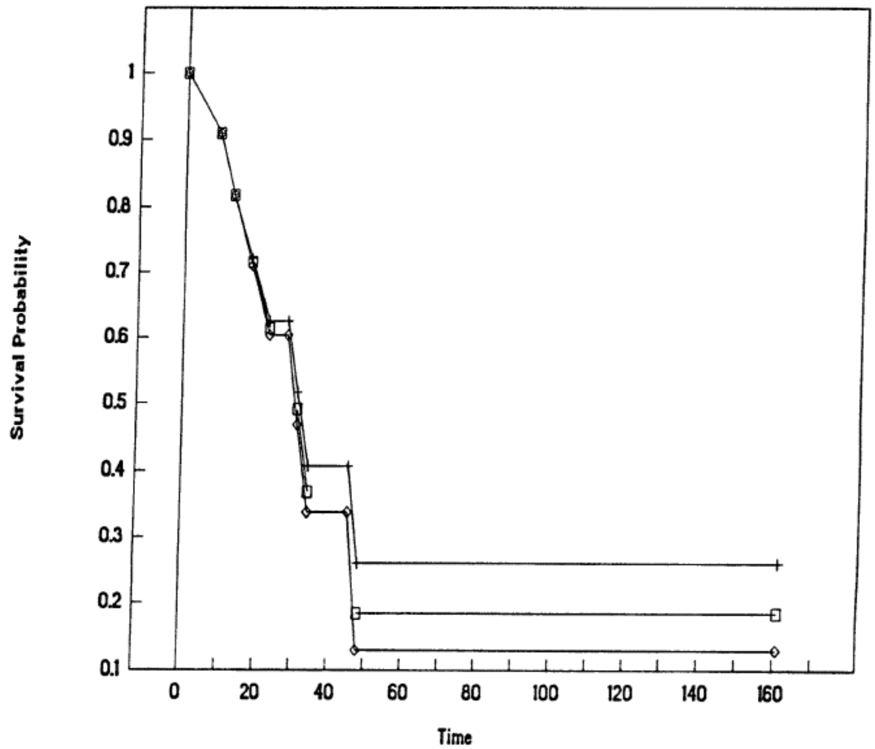


Figure 1: Comparison of survival probability estimates based on various assumptions about the censored probabilities of death: □, product limit assumption ($a_i = 1$); +, assumption 1 ($a_i = 0.5$); ◇, assumption 2 ($a_i = 1.5$).

product limit estimates can be obtained from equation (2.2) as follows where $d_1 = 1, c_1 = 1, n_1 = n, d_2 = 0, c_2 = 0, n_2 = n-1, d_3 = 1, c_3 = 1$ and $n_3 = n-2$:

$$\begin{aligned}\hat{P}(x) &= \left(1 - \frac{1}{n_1}\right) \times 1 \times \left(1 - \frac{1}{n_3}\right) \\ &= \frac{n-1}{n} \times 1 \times \frac{n-3}{n-2} \\ &= \frac{n^2 - 4n + 3}{n^2 - 2n}. \end{aligned} \tag{A.1}$$

Equation (A.1) can be expressed in terms of equation (3.9) for ${}^{u}\hat{q}^{\mathrm{d}}(x) = 2/n, {}^{\mathrm{u}}\hat{q}_1^{\mathrm{d}} = 1/n, \hat{P}^{\mathrm{c}}(t_1) = 0, {}^{\mathrm{u}}\hat{q}_2^{\mathrm{d}} = 0, \hat{P}^{\mathrm{c}}(t_2) = 1/n, {}^{u}\hat{q}_3^{\mathrm{d}} = 1/(n-2)$ and $\hat{P}^{\mathrm{c}}(t_3) = 1/n$ as shown below:

$$\begin{aligned}\hat{P}(x) &= 1 - \frac{2}{n} - \frac{1}{n}\frac{1}{n-2} \\ &= \frac{n^2 - 4n + 3}{n^2 - 2n}. \end{aligned} \tag{A.2}$$

A.2 Example 2

Let us consider n ordered survival times, out of which the first observation is a death at time t_1, the second observation is censored at time t_2, the third observation is death at time t_3, the fourth observation is censored at time t_4 and the fifth observation is death at time t_5. We want to estimate $P(x)$ where $t_i \leqslant x$. Using equation (2.2) we can show for $d_1 = 1, c_1 = 1, n_1 = n, d_2 = 0, c_2 = 0, n_2 = n-1, d_3 = 1, c_3 = 1, n_3 = n-2, d_4 = 0, c_4 = 0, n_4 = n-3, d_5 = 1, c_5 = 1$ and $n_5 = n-4$ that

$$\begin{aligned}\hat{P}(x) &= \frac{n-1}{n} \times 1 \times \frac{n-3}{n-2} \times 1 \times \frac{n-5}{n-4} \\ &= 1 - \frac{3}{n} - \frac{3n-9}{n\left(n^2-6n+8\right)}. \end{aligned} \tag{A.3}$$

Equation (A.3) can be expressed in terms of equation (11) for ${}^{\mathrm{u}}\hat{q}^{\mathrm{d}}(x) = 3/n, {}^{\mathrm{u}}\hat{q}_1^{\mathrm{d}} = 1/n, \hat{P}^{\mathrm{c}}\left(t_1\right) = 0, {}^{\mathrm{u}}\hat{q}_2^{\mathrm{d}} = 0, \hat{P}^{\mathrm{c}}\left(t_2\right) = 1/n, {}^{\mathrm{u}}\hat{q}_3^{\mathrm{d}} = 1/(n-2), \hat{P}^{\mathrm{c}}\left(t_3\right) = 1/n, {}^{\mathrm{u}}\hat{q}_4^{\mathrm{d}} = 0, \hat{P}^{\mathrm{c}}\left(t_4\right) = 2/n - (1/n)\{1/(n-2)\}, {}^{\mathrm{u}}\hat{q}_5^{\mathrm{d}} = 1/(n-4)$ and $\hat{P}^{\mathrm{c}}\left(t_5\right) = 2/n - (1/n)\{1/(n-2)\}$ as follows:

$$\begin{aligned}\hat{P}(x) &= 1 - {}^{\mathrm{u}}\hat{q}^{\mathrm{d}}(x) - \left\{{}^{\mathrm{u}}\hat{q}_3^{\mathrm{d}}\hat{P}^{\mathrm{c}}\left(t_3\right) + {}^{\mathrm{u}}\hat{q}_5^{\mathrm{d}}\hat{P}^{\mathrm{c}}\left(t_5\right)\right\} \\ &= 1 - \frac{3}{n} - \left\{\frac{1}{n-2}\frac{1}{n} + \frac{1}{n-4}\left(\frac{2}{n} - \frac{1}{n}\frac{1}{n-2}\right)\right\} \\ &= 1 - \frac{3}{n} - \frac{3n-9}{n\left(n^2-6n+8\right)}. \end{aligned} \tag{A.4}$$

These examples show the exact equivalence between the estimates obtained by the product limit method and the proposed method for the assumption of equality of probabilities of death for uncensored and censored states.

References

Aalen, O. (1976), “Nonparametric inference in connection with multiple decrement models,” *Scandinavian Journal of Statistics*, 3, 15–27.

Breslow, N. E. and Crowley, J. (1974), “A large sample study of the life table and product limit estimates under random censorship,” *The Annals of statistics*, 2, 437–453.

Chiang, C. L. (1968), “Introduction to stochastic processes in biostatistics”, 213–219. New York: Wiley.

Davis, T. P. and Lawrance, A. J. (1989), “The likelihood for competing risk survival analysis,” *Scandinavian journal of statistics*, 5, 23–28.

Efron, B. (1967), “The two sample problem with censored data,” in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. IV, pp. 831–853. New York: Wiley.

Kalbfleisch, J. D. and Prentice, R. L. (1980), *The statistical analysis of failure time data*, New York: Wiley.

Kaplan, E. L. and Meier, P. (1958), “Nonparametric estimation from incomplete observations,” *Journal of the American statistical association*, 53, 457–481.

Mantel, N. (1987), "Truncation as a unifying principle," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 36, 21–24.

Matthews, D. E. (1984), "Some observations on semi-markov models for partially censored data," *Canadian Journal of Statistics*, 12, 201–205.

Meier, P. (1977), "Estimation of a distribution function from incomplete observations," In *Prespectives in Probability and Statistics* (ed. J. Gani), 67–87. New York: Academic Press.

Miller Jr, R. G. (1981), *Survival analysis*, New York: Wiley.

Peterson, A. V. (1977), "Exploring the Kaplan-Meier estimator as a function of empirical subsurvival functions," *Journal of the American Statistical Association*, 72, 854–858.

Schemper, M. (1991), "Generalisation of the Kaplan–Meier Estimator for Standardised Processes," *Journal of the Royal Statistical Society: Series D (The Statistician)*, 40, 17–25.

Journal of Statistical Research *https://doi.org/10.47302/jsr.2021550102*
2021, Vol. 55, No. 1, pp. 17-19 *ISSN 0256 - 422 X*

Invited commentary on the paper by M. Ataharul Islam:

AN ALTERNATIVE TO THE PRODUCT LIMIT METHOD: THE MULTISTATE APPROACH, The Statistician 43, 237-245, 1994

NIELS KEIDING
Section of Biostatistics, University of Copenhagen, Denmark
Email: hnj501@ku.dk

This paper is one of the first journal publications from Dr. Islam, published at the end of his three-year long post-doc experience in the U.S., at the prestigious institutions Population Studies Center (Jane Menken) at the University of Pennsylvania, and at the East-West Center in Hawaii.

The title of the paper focused on the methodology: embedding the problem of modelling censoring in survival analysis into multistate modelling. Fundamental work generalizing central survival analysis methods to multistate models had been around for some time, notably by Aalen and Johansen (1978) and also Fleming (1978a,b).

The key aspect of the paper is however different: Islam challenged the standard assumption in survival analysis that censoring and survival should be independent. In this paper Islam took a different view: what if the mortality for those censored is different from the mortality for those uncensored? He gave no practical or hypothetical examples where that would be a realistic assumption, but there are scattered remarks such as

... *in many practical problems* [the assumption ... that the probability of dying is the same for uncensored and censored individuals] *this is not the case* (Section 3)

This flexibility in the assumptions of censored deaths will allow researchers to incorporate assumptions about such deaths that *may vary in many medical problems from the standard product limit assumption (Section 3)* with no reference or other indication of being based on reality.

Since the assumption of independence between death and censoring penetrates the total survival analysis literature, a brief introduction may be justified. In the simplest situation one considers n independent, identically distributed life times X_i with distribution function $F(x)$. Not all life times are observed; for some it is only known that the individual was still alive at some censoring time C_i . Under the assumption that censoring is independent of the life time Kaplan and Meier (1958) derived the product limit estimator of $1 - F(x)$ and indicated that this could be interpreted as the nonparametric maximum likelihood estimator of $1 - F(x)$.

The usual interpretation of these basic matters in survival analysis is that individuals who are censored are subject to the same mortality risk as those whose life length is observed, and this issue was and is a central part of elementary as well as advanced survival analysis as taught and used everywhere. Andersen et al. (2021) (p. 189) recently gave an excellent, very careful non-mathematical description of the current understanding of censoring.

To return to Islam's paper:

As indicated in the paper's title, Islam proposed the obvious four-state model (uncensored survivor, uncensored death, censored survivor, censored death) and derived some basic formulae regarding estimation. These estimators depend on unobservable quantities unless the standard assumption of independence between survival and censoring is fulfilled.

The results were illustrated not through Islam's own practical experience but by reinterpreting a small set of data from the well-known textbook by Miller (1981) on length of remission for leukemia patients. Using the formulae developed from the multistate model, he derived modified product-limit estimates from various assumptions of increased or decreased mortality of the censored patients.

References

Aalen, O. O. and Johansen, S. (1978), "An empirical transition matrix for non-homogeneous Markov chains based on censored observations," *Scandinavian Journal of Statistics*, 5, 141–150.

Andersen, P. K., Perme, M. P., van Houwelingen, H. C., Cook, R. J., Joly, P., Martinussen, T., Taylor, J. M., Abrahamowicz, M., and Therneau, T. M. (2021), "Analysis of time-to-event for observational studies: Guidance to the use of intensity models," *Statistics in Medicine*, 40, 185–211.

Fleming, T. R. (1978a), "Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks," *The Annals of Statistics*, 6, 1057–1070.

— (1978b), "Asymptotic distribution results in competing risks estimation," *The Annals of Statistics*, 6, 1071–1079.

Received: April 20, 2021

Accepted: April 30, 2021

Journal of Statistical Research *https://doi.org/10.47302/jsr.2021550103*
2021, Vol. 55, No. 1, pp. 19-20 *ISSN 0256 - 422 X*

Invited commentary on the paper by M. Ataharul Islam:

AN ALTERNATIVE TO THE PRODUCT LIMIT METHOD: THE MULTISTATE APPROACH, The Statistician 43, 237-245, 1994

ODD O. AALEN
University of Oslo, Norway
Email: o.o.aalen@medisin.uio.no

Dr. M. Ataharul Islam has written an interesting paper on the Kaplan-Meier product-limit estimator. This estimator is one of the fundamental tools of medical statistics, used to construct survival curves for censored data. The Kaplan-Meier paper from 1958 has 60,000 references by 2021, and is thus one of the most referenced papers in all of statistics.

The aim of the Kaplan-Meier approach is to estimate the probability that a certain event will not occur before a given time. Plotting these probabilities against time gives a survival curve. The essential feature of the product limit estimator is its ability to correct for censoring (incompletely observed times) in a simple manner. The price of simplicity, however, is an assumption about non-informative censoring, meaning that censoring is not related to the likelihood of the event occurring.

However, often this assumption is not met. One may instead have informative censoring, and the Kaplan-Meier approach does not work. The aim of Dr. Islam's paper is to develop a method for this case. His multistate approach is transparent and various choices of the effect of informative censoring can be made.

Dr. Islam's method can be seen as sensitivity analysis; what is the effect when the standard assumption of the product-limit estimator does not hold. In view of the importance of the

Kaplan-Meier approach, there should be more attention to this issue. One will rarely have good knowledge about the effect of censoring. Sensitivity analysis along the lines of Dr. Islam's approach seems a very useful tool which should be applied more.

Received: March 19, 2021

Accepted: April 30, 2021