On the small-sample performance of Efron's and of Gill's version of the product limit estimator under nonproportional hazards

Citation for published version (APA):

Geurts, J. H. J. (1987). On the small-sample performance of Efron's and of Gill's version of the product limit estimator under nonproportional hazards. *Biometrics*, 43(3), 683-692.

Document status and date:

Published: 01/01/1987

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

On the Small-Sample Performance of Efron's and of Gill's Version of the Product Limit Estimator Under Nonproportional Hazards

Johannes H. J. Geurts

Biometrics, Vol. 43, No. 3. (Sep., 1987), pp. 683-692.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28198709%2943%3A3%3C683%3AOTSPOE%3E2.0.CO%3B2-S>

Biometrics is currently published by International Biometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ibs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

On the Small-Sample Performance of Efron's and of Gill's Version of the Product Limit Estimator Under Nonproportional Hazards

Johannes H. J. Geurts

Department of Industrial Engineering and Management Science, Pav. F15,
Eindhoven University of Technology, P.B. 513,
5600 MB Eindhoven, The Netherlands

SUMMARY

Various proposals have been made to extend the product limit estimator to survival times beyond the largest observation in case that observation is censored. Two extreme extensions are examined with respect to bias and mean squared error (MSE). Their quality depends considerably on the "censoring constellation." The MSE of one extension appears to be robust against a wide variety of nonproportionalities of the hazard rates of the distributions of lifelength and time to censoring.

1. Introduction

In a multiply right-censored sample from a survival probability distribution, only the smaller of time to failure and censoring time is observable for each of the items included in the sample. Multiple right random (unplanned) censoring occurs in a variety of industrial, biological, and demographical situations.

Let $F(x)$, $x \geq 0$, be the cumulative distribution function (cdf) of time to failure X of an item and let $G(y)$, $y \geq 0$ be the cdf of random time to censoring Y of that item; X and Y are assumed to be mutually independent and F and G are assumed to be continuous and everywhere differentiable. The hazard rates of $F(x)$ and $G(y)$, respectively, are $z_x(x) = F'(x)/[1 - F(x)]$ and $z_y(y) = G'(y)/[1 - G(y)]$. Let $T = \min(X, Y)$ and let $H(t)$ be its cdf. In an ordered random sample $0 < t_1 < \dots < t_n$ of size n , the t_i are observable and, it is assumed, also the δ_i , where

$$\delta_i = \begin{cases} 1 & \text{if } t_i \text{ is a time to failure} \\ 0 & \text{if } t_i \text{ is a time to censoring} \end{cases}$$

The Kaplan–Meier (1958) estimator (KME), also called the product limit estimator, of $\bar{F}(x) = 1 - F(x)$ is defined as

$$\bar{F}_n(x) = 1, \quad 0 \leq x < t_1 \quad (1.1)$$

$$\bar{F}_n(x) = \bar{F}_n(t_i), \quad t_i \leq x < t_{i+1}, \quad i = 1, \dots, n-1 \quad (1.2)$$

$$\bar{F}_n(x) = \bar{F}_n(t_n), \quad x = t_n \quad (1.3)$$

$$\bar{F}_n(x) = 0, \quad x > t_n \mid \delta_n = 1 \quad (1.4)$$

Key words: Censored observations; Nonproportional hazards; Product limit estimator; Small-sample moments; Survival function.

where

$$\bar{F}_n(t_i) = \prod_{j=1}^i [1 - \delta_j/(n - j + 1)], \quad i = 1, \dots, n. \quad (1.5)$$

Originally $\bar{F}_n(x)$ for $x > t_n \mid \delta_n = 0$ was left undefined by Kaplan and Meier (1958). Later Efron (1967) slightly modified the KME by replacing (1.4) by

$$\bar{F}_n(x) = 0, \quad x > t_n. \quad (1.4a)$$

Gill (1980) proposed another modification, replacing (1.4) by

$$\bar{F}_n(x) = \bar{F}_n(t_n), \quad x > t_n. \quad (1.4b)$$

We shall refer to Efron's version, also called the self-consistent estimator, by KMEE and to Gill's by KMEG.

The KMEE and the KMEG can be considered as two extreme ways to "complete" the KME beyond the largest observation if that observation is censored. Brown, Hollander, and Korwar (1974) and Moeschberger and Klein (1985) investigate various properties of less extreme extensions. These all have in common that $0 \leq \bar{F}_n(x) < \bar{F}_n(t_n)$ for $x > t_n \mid \delta_n = 0$, where $\bar{F}_n(x)$ is monotonically decreasing in one way or another. Moeschberger and Klein (1985) address a particular form of planned progressive censoring while we address random censoring. Nevertheless, it can be concluded from their results that the performance of the various extensions is very dependent on what we shall call the "censoring constellation" which, in random censoring, may be described as the location and shape of the censoring distribution $G(t)$ relative to the failure-time distribution $F(t)$ and which we shall characterize below. In this paper we limit ourselves to comparing the two extreme extensions KMEE and KMEG under random censoring and various censoring constellations.

The small-sample behaviour of the KMEE was studied by Chen, Hollander, and Langberg (1982) for the case where the hazard rate of $F(t)$ is proportional to the hazard rate of $G(t)$ and by Geurts (1985) for some nonproportional hazards cases. Wellner (1985) provides an approximate variance formula for the KMEG and demonstrates that in the small-sample proportional hazards case the KMEG yields biases and variances that are almost everywhere smaller than the KMEE's, at least for values of t that are not large, i.e., values of $\bar{F}(t)$ that are large to medium.

In this paper we summarize in Section 2 some general results on the approximate variance formulae of the KMEE and the KMEG for $t \rightarrow \infty$. In Section 3 we present numerical bias and variance results for some particular small-sample nonproportional hazards cases of interest. In Section 4 we discuss these results and note that the mean squared error (MSE) of the KMEE exhibits considerable robustness against nonproportionalities of the hazard rates of $F(t)$ and $G(t)$ and that for moderate to small values of $\bar{F}(t)$ it is smaller than that of the KMEG. We also indicate the practical importance of these properties, which sometimes make the KMEE preferable over the KMEG.

2. Asymptotic Variances

The approximate variance of the KMEE, as given by "Greenwood's formula," mentioned by Kaplan and Meier (1958), is

$$VG_n(t) \equiv [\bar{F}(t)]^2 \cdot \int_0^t \frac{dF(z)}{[1 - F(z)]^2 n [1 - G(z)]}.$$

Table 1
Limiting values of $VG_n(t)$ and $VP_n(t)$ as $t \rightarrow \infty$ for various conditions on $r(t)$

(Case)	$r(t) = r$			$\lim_{t \rightarrow \infty} \sup r(t) < 1$ and $I(t) < \infty^a$ for $t < \infty$	$\lim_{t \rightarrow \infty} \inf r(t) > 1$
	<1 (1)	1 (2)	>1 (3)		
$\lim_{t \rightarrow \infty} VG_n(t)$	0	$(2n)^{-1}$	∞	0	∞
$\lim_{t \rightarrow \infty} VP_n(t)$	∞	∞	∞	∞	∞

$$^a I(t) = \int_0^t [1 - F(z)]^{-2-r(z)} dF(z).$$

Wellner (1985) gives the following approximation to the variance of the KMEG:

$$VP_n(t) \equiv [\bar{F}(t)]^2 \left(\exp \left\{ \int_0^t \frac{dF(z)}{[1 - F(z)]^2 n [1 - G(z)]} \right\} - 1 \right).$$

Kaplan and Meier (1958) show how the approximate variance formulae can be converted to approximate variance estimators by the insertion of sample estimates and the replacement of integrals by sums.

The approximations $VG_n(t)$ and $VP_n(t)$ agree as $n \rightarrow \infty$. However, for finite n they do not agree as $t \rightarrow \infty$. Here we shall give some theoretical results for $t \rightarrow \infty$; in the next section we shall numerically illustrate these results.

We consider the class of cases for which $1 - G(t) = [1 - F(t)]^{r(t)}$, where $r(t)$ is a positive function of t . The behaviour of $r(t)$ determines the limiting behaviour of $VG_n(t)$ and $VP_n(t)$ as indicated in Table 1. It should be noted that $r(t)$ is not identical to the "relative hazard rate" $z_y(t)/z_x(t)$. In the proportional hazards case only is $z_y(t)/z_x(t) = r(t) = r$. In general we have

$$z_y(t)/z_x(t) = r(t) - r'(t) \ln[1 - F(t)]/z_x(t).$$

For the proportional hazards case, i.e., $r(t) = r$, $0 < r < \infty$, we readily derive the results exhibited in Table 1 under $r(t) = r$. Straightforward limit arguments show that the results in the remaining two columns of Table 1 hold, where

$$I(t) = \int_0^t [1 - F(z)]^{-2-r(z)} dF(z).$$

The censoring constellation may be characterized by the behaviour of $r(t)$ together with the value of the expected fraction of observations censored,

$$E(FC) \equiv \Pr(X > Y) \equiv 1 - \int_0^\infty [1 - G(t)] dF(t).$$

3. Numerical Results

The asymptotic results of the previous section are not very informative. Therefore, we investigate the small-sample behaviour of the estimators for some specific distributions. We choose Weibull distributions for both $F(x)$ and $G(y)$:

$$F(x) \equiv 1 - \exp(-x^a) \quad \text{and} \quad G(y) \equiv 1 - \exp[-(\beta y)^b].$$

The choice of the Weibull family of distributions for $F(x)$ is justified by the fact that it is often used in survival-time analysis, especially in industrial situations. The choice of the

Table 2

Bias^a [$B = \bar{F}_n(t) - \bar{F}(t)$], variance,^a approximate variance (Var_{app}), and mean squared error^a of KMEE (upper entry) and KMEG (lower entry); $E(FC) = .3333$

$\bar{F}(t)$	b/a	$n = 10$				$n = 20$			
		$-B$	Var	Var_{app}	MSE	$-B$	Var	Var_{app}	MSE
.6065	$\frac{1}{2}$	.00	.032	.0304	.032	.00	.016	.0152	.016
		.00	.031	.0317	.032	.00	.016	.0155	.016
	1	.0001	.0281	.0274	.0281	.0000	.0138	.0137	.0138
		-.0000	.0280	.0284	.0280	.0000	.0138	.0140	.0138
	2	.00	.025	.0250	.025	.00	.012	.0125	.012
		.000	.025	.0258	.025	.00	.012	.0127	.012
.3679	$\frac{1}{2}$	.00	.037	.0333	.038	.00	.018	.0167	.018
		.00	.036	.0378	.036	.00	.018	.0177	.018
	1	.0054	.0361	.0314	.0362	.0003	.0164	.0157	.0164
		-.0009	.0334	.0354	.0334	-.0000	.0163	.0167	.0163
	2	.01	.034	.0286	.034	.00	.014	.0143	.014
		.00	.030	.0318	.030	.00	.014	.0151	.014
.1353	$\frac{1}{2}$	.01	.023	.0202	.023	.00	.011	.0101	.011
		-.01	.022	.0370	.023	.00	.011	.0135	.011
	1	.0285	.0238	.0233	.0246	.0129	.0140	.0117	.0142
		-.0193	.0254	.0471	.0258	-.0053	.0127	.0163	.0127
	2	.08	.017	.0347	.023	.06	.015	.0173	.018
		-.05	.029	.1033	.032	-.02	.016	.0289	.016
.0498	$\frac{1}{2}$	.009	.009	.0096	.009	.00	.005	.0048	.005
		-.02	.013	.1162	.013	-.01	.006	.0147	.007
	1	.0239	.0072	.0147	.0077	.0179	.0053	.0074	.0057
		-.0472	.0194	.9341	.0216	-.0224	.0088	.0457	.0093
	2	.0475	.0008	.0940	.0031	.0465	.0009	.0470	.0030
		-.11	.028	7×10^{13}	.042	-.08	.016	4×10^5	.023

^a The estimated standard deviation of the entries does not exceed one-half unit of the last decimal place shown; entries for $b/a = 1$ are exact.

Weibull for $G(y)$ is based on its flexibility, which allows it to approximate a wide variety of distributions of censoring times. The selection of values for the parameters of $F(x)$ and $G(y)$ is the result of the following considerations.

In addition to the proportional hazards case where $r(t) = r$, we want to evaluate a case where $r(t)$ is decreasing and one where $r(t)$ is increasing. When $F(x)$ and $G(y)$ have been specified as above we have $r(t) = \beta^b t^{b-a}$ and therefore $r(t)$ is decreasing for $b/a < 1$, constant for $b/a = 1$, and increasing for $b/a > 1$. Practical experience suggests moderate values of b/a ; mathematical convenience led to the choice of $b/a = \frac{1}{2}$ and $b/a = 2$ in addition to $b/a = 1$. In the interest of comparability of our results with those of Chen et al. (1982), Wellner (1985), and Geurts (1985), we chose values of $E(FC) = \frac{1}{3}, \frac{1}{2}$, and $\frac{2}{3}$. Given the values of a and b , the value of $E(FC)$ then determines the value of the parameter β in $G(y)$ and thereby completely specifies $r(t)$.

If $F(x)$ and $G(y)$ are Weibull distributions as specified above and if a value of $E(FC)$ and a value of $F(x)$ for which the bias and the variance are to be evaluated are given, then the bias and variance results are independent of the particular value of a ; it is the ratio b/a that determines the result. For a proof we refer to Geurts (1985), who also provides the details of the calculation of the values of β that correspond to given values of $E(FC)$ for $b/a = \frac{1}{2}, 1$, and 2 . For reasons of convenience we choose $a = 1$ and therefore $b = \frac{1}{2}$, $b = 1$, and $b = 2$.

Table 3

Bias^a [$B = \bar{F}_n(t) - \bar{F}(t)$], variance,^a approximate variance (Var_{app}), and mean squared error^a of KMEE (upper entry) and KMEG (lower entry); $E(FC) = .5000$

$\bar{F}(t)$	b/a	$n = 10$				$n = 20$			
		$-B$	Var	Var_{app}	MSE	$-B$	Var	Var_{app}	MSE
.6065	$\frac{1}{2}$	.00	.042	.0370	.043	.00	.019	.0185	.019
		.00	.040	.0389	.040	.00	.019	.0190	.019
	1	.0017	.0347	.0316	.0348	.0000	.0162	.0158	.0162
		-.0001	.0333	.0330	.0333	.0000	.0162	.0161	.0162
	2	.00	.028	.0272	.028	.00	.014	.0136	.014
		.00	.028	.0282	.028	.00	.014	.0139	.014
.3679	$\frac{1}{2}$	.01	.057	.0445	.057	.00	.025	.0223	.025
		.00	.049	.0527	.050	.00	.024	.0242	.024
	1	.0339	.0589	.0432	.0600	.0061	.0265	.0216	.0265
		-.0073	.0474	.0509	.0475	-.0008	.0236	.0234	.0236
	2	.08	.066	.0443	.073	.03	.034	.0222	.035
		-.01	.045	.0524	.046	.00	.024	.0241	.024
.1353	$\frac{1}{2}$	.03	.030	.0314	.031	.02	.018	.0157	.018
		-.03	.037	.0836	.039	-.01	.017	.0249	.017
	1	.0742	.0232	.0491	.0287	.0532	.0200	.0245	.0228
		-.0723	.0470	.2487	.0522	-.0335	.0252	.0516	.0263
	2	.1325	.0014	.48	.0190	.1308	.0018	.24	.0189
		-.17	.051	4×10^9	.081	-.12	.034	9×10^3	.050
.0498	$\frac{1}{2}$	.019	.010	.0169	.011	.01	.008	.0085	.008
		-.055	.028	2.3	.032	-.02	.012	.0726	.012
	1	.0411	.0037	.0499	.0054	.0375	.0037	.0249	.0051
		-.1308	.0453	1×10^6	.0624	-.0818	.0236	6×10	.0303
	2	.05	.0000	9×10^2	.0025	.0498	.0000	5×10	.0025
		-.25	.051	$>10^{99}$	.116	-.21	.034	$>10^{99}$	.080

^a The estimated standard deviation of the entries does not exceed one-half unit of the last decimal place shown; entries for $b/a = 1$ are exact.

For the numerical evaluation of the bias and the variance of the KMEE, Geurts (1985) employed a combination of analytical and simulation techniques. After small and obvious alterations this method was used also to evaluate the bias and variance of the KMEG. The results are given in Tables 2, 3, and 4. They cover each of the five cases in Table 1 as follows:

$$\begin{array}{lll}
 a = 1, \quad b = 1 & \beta = \frac{1}{2} \Rightarrow E(FC) = \frac{1}{3} \} & \text{Case (1);} \\
 \Rightarrow r(t) = \beta & \beta = 1 \Rightarrow E(FC) = \frac{1}{2} \} & \text{Case (2);} \\
 & \beta = 2 \Rightarrow E(FC) = \frac{2}{3} \} & \text{Case (3);} \\
 \\
 a = 1, \quad b = \frac{1}{2} & E(FC) = \frac{1}{3} \Rightarrow \beta = .2349 \} & \\
 \Rightarrow r(t) = (\beta/t)^{1/2} & E(FC) = \frac{1}{2} \Rightarrow \beta = .7491 \} & \text{Case (4);} \\
 & E(FC) = \frac{2}{3} \Rightarrow \beta = 2.1468 \} & \\
 \\
 a = 1, \quad b = 2 & E(FC) = \frac{1}{3} \Rightarrow \beta = .6824 \} & \\
 \Rightarrow r(t) = \beta^2 t & E(FC) = \frac{1}{2} \Rightarrow \beta = 1.1555 \} & \text{Case (5).} \\
 & E(FC) = \frac{2}{3} \Rightarrow \beta = 2.0636 \} &
 \end{array}$$

Tables 2, 3, and 4 also show values of the approximate variances of the KMEE and the KMEG, designated by Var_{app} . For the nonproportional hazards cases they were obtained

Table 4
Bias^a [$B = \bar{F}_n(t) - \bar{F}(t)$], variance,^a approximate variance (Var_{app}), and mean squared error^a of KME (upper entry) and KMEG (lower entry); $E(FC) = .6667$

$\bar{F}(t)$	b/a	$n = 10$				$n = 20$			
		$-B$	Var	Var_{app}	MSE	$-B$	Var	Var_{app}	MSE
.6065	$\frac{1}{2}$	.02	.072	.0506	.073	.00	.029	.0253	.029
		.00	.058	.0542	.058	.00	.028	.0262	.028
	1	.0250	.0639	.0427	.0646	.0017	.0242	.0213	.0242
		-.0022	.0478	.0453	.0478	-.0001	.0229	.0220	.0229
	2	.04	.064	.0376	.065	.00	.022	.0188	.022
		.00	.041	.0396	.041	.00	.020	.0193	.020
.3679	$\frac{1}{2}$	.07	.089	.0711	.094	.02	.050	.0355	.050
		-.03	.079	.0935	.080	-.01	.042	.0406	.042
	1	.1549	.0852	.0861	.1091	.0814	.0648	.0430	.0714
		-.0485	.0779	.1204	.0803	-.0160	.0462	.0507	.0464
	2	.338	.019	.311	.133	.32	.027	.156	.127
		-.12	.070	1.2	.087	-.08	.052	.292	.059
.1353	$\frac{1}{2}$	.066	.032	.0640	.037	.05	.026	.0320	.029
		-.10	.075	.5834	.088	-.05	.038	.0867	.040
	1	.1228	.0073	.2457	.0224	.1156	.0091	.1229	.0225
		-.2079	.0861	1×10^4	.1293	-.1419	.0560	1.5×10	.0761
	2	.1353	.0000	2×10^4	.0183	.1353	.0000	10^4	.0183
		-.35	.070	$>10^{99}$	.194	-.31	.053	$>10^{99}$	.148
.0498	$\frac{1}{2}$	.0334	.0083	.0420	.0094	.028	.007	.0210	.008
		-.15	.070	6×10^4	.096	-.08	.033	1.2×10	.040
	1	.0492	.0004	.6694	.0028	.0488	.0005	.3347	.0029
		-.2894	.0864	$>10^{99}$	.1702	-.2211	.0564	1×10^{56}	.1053
	2	.0498	.0000	$>10^{99}$	.0025	.0498	.0000	$>10^{99}$	.0025
		-.43	.071	$>10^{99}$	.26	-.39	.053	$>10^{99}$	.21

^a The estimated standard deviation of the entries does not exceed one-half unit of the last decimal place shown; entries for $b/a = 1$ are exact.

by numerically evaluating the integrals in the formulae for $VG_n(t)$ and $VP_n(t)$, respectively. For the proportional hazards cases $VG_n(t)$ and $VP_n(t)$ can be evaluated analytically.

4. Discussion

The absolute value of the bias of the KMEG is smaller than that of the KMEE at large to medium values of $\bar{F}(t)$ regardless of the values of b/a and $E(FC)$. A notable example of this phenomenon is provided by the values at $\bar{F}(t) = .3679$, $E(FC) = \frac{2}{3}$, $b/a = 2$ (see Table 4). However, at smaller values of $\bar{F}(t)$ the bias of the KMEG worsens while the bias of the KMEE improves; this becomes more notable as $E(FC)$ and b/a increase.

The exact small-sample variance of the KMEG is (slightly) smaller than that of the KMEE at large values of $\bar{F}(t)$, regardless of the values of $E(FC)$ and b/a . However, at smaller values of $\bar{F}(t)$ the variance of the KMEG remains rather stable while the variance of the KMEE decreases—the more so as $E(FC)$ and b/a increase—to such an extent that already at $\bar{F}(t) = .3679$, $E(FC) = \frac{2}{3}$, $b/a = 2$ (see Table 4) the variance of the KMEG is a multiple of that of the KMEE. Geurts (1985) gives an explanation of the phenomenon of the KMEE's decreasing bias and variance as $t \rightarrow \infty$.

While at large values of $\bar{F}(t)$ the asymptotic variance formulae $VG_n(t)$ and $VP_n(t)$ give fair approximations of the variances they are meant to approximate— $VP_n(t)$ generally performing better than $VG_n(t)$ —they both break down at smaller values of $\bar{F}(t)$, with $VG_n(t)$ holding up longer than $VP_n(t)$. The breakdown is accelerated by increasing $E(FC)$ and b/a . This phenomenon is consonant with Wellner's (1985) anticipation that $VP_n(t)$ should not be expected to be accurate when $n[1 - H(t)] < 1$. Beyond that, the numerical values of Tables 2, 3, and 4 provide a quantitative illustration of the qualitative results in Table 1.

Before we discuss the MSE we recall the fact that (the estimate of) the variance of an estimate is a misleading measure of the accuracy of the estimate when the estimator has a nonnegligible bias. When an estimator is biased the MSE of the estimate is a better measure of accuracy. The situation becomes even more cumbersome when the estimator of the variance of the estimate is biased too.

We have already seen the conditions of $\bar{F}(t)$, $E(FC)$, and b/a under which the bias of the KMEE and the bias of the KMEG become nonnegligible. The corresponding variance formulae, and with them the variance estimators, also suffer from increasing biases under these conditions. Since we have sample estimates of $E(FC)$ and $\bar{F}(t)$ —though the latter may be biased—we might make an attempt numerically to remove the bias of the estimate of $\bar{F}(t)$ and the estimate of its variance with the aid of Tables 2, 3, and 4, if only we knew (or had an estimate of) $r(t)$. Usually we do not know much about $r(t)$. However, Kardaun (1983) in an extensive analysis of a particular case manages to find some evidence for the proportionality of his hazard rates.

We now note an agreeable property of the KMEE already observed by Geurts (1985) and illustrated in Figures 1 and 2: for given values of $\bar{F}(t)$ and $E(FC)$ the MSE of the KMEE is remarkably insensitive to the shape of $r(t)$. Tables 2, 3, and 4 and Figures 1 and 2 show that the KMEG shares this property down to moderate values of $\bar{F}(t)$ but it is exactly where the possibility of bias begins to play a role and therefore the MSE becomes increasingly more important than the variance [i.e., at smaller values of $\bar{F}(t)$], that the MSE of the KMEG becomes increasingly dependent on $r(t)$ while the robustness of the MSE of

Figure 1. Mean squared error and squared bias of KMEE and of KMEG for selected values of $F(t)$; $n = 10$; $E(FC) = .3333$.
 -----: $b/a = \frac{1}{2}$; ———: $b/a = 1$; - · - · -: $b/a = 2$.

Figure 2. Mean squared error and squared bias of KME and of KMEG for selected values of $F(t)$; $n = 10$; $E(FC) = .6667$.
 -----: $b/a = \frac{1}{2}$; ———: $b/a = 1$; — · — · —: $b/a = 2$.

the KME persists. Additionally, the MSE of the KME there is smaller than that of the KMEG.

These phenomena begin to show up especially strongly at high values of $E(FC)$ and b/a for values of $F(t)$ as low as .75, say. Therefore, in the presence of a high percentage of randomly right-censored observations and in the absence of knowledge of the shape of $r(t)$, it would be prudent practice to estimate—if one has to—probabilities of survival beyond a censored largest observation by the KME rather than by the KMEG, i.e., to take them to be zero. And even if, unbeknownst to us, the censored largest observation does not fall beyond the 75th percentile but around the 65th of the lifelength distribution—the worst case for the KME, as can be seen from the tables and figures—then the damage would still be limited, i.e., the MSE of the KME would not be much larger than the MSE of the KMEG.

As mentioned before, the robustness and lower value of the MSE of the KME give the KME an advantage over the KMEG only for moderate to low values of $\bar{F}(t)$. At high values of $\bar{F}(t)$ the KMEG is to be preferred over the KME thanks to the KMEG's smaller bias and the better estimator of its (smaller) variance.

5. Example

In a case study Kardaun (1983) presents in his Table 1 survival data on 17 male patients with stage 2 larynx cancer. The data include the time interval (in years) between first treatment and either the death of the patient or the endpoint of the study. The observations that were terminated by the end of the study are censored observations. The 17 ordered observations are as follows (censored data are marked with an asterisk): .2, 1.8, 2.0, 2.2*, 2.6*, 3.3*, 3.6, 3.6*, 4.0, 4.3*, 4.3*, 5.0*, 6.2, 7.0, 7.5*, 7.6*, 9.3*. In order to avoid cumbersome two-way interpolations in the course of the example we have, for the sake of this example, augmented the data set by three fictitious observations assumed to have been terminated by death, viz. 7.8, 8.1, and 8.4 years. Note that for the augmented data set the fraction of censored observations is $FC = 10/20 = .5$.

Under the usual model of random censorship we can calculate the estimate $\bar{F}_n(t)$ of the survival function $\bar{F}(t)$ for any value of $t \geq 0$. For example, for both KMEE and KMEG we calculate from the augmented data set

$$\bar{F}_n(8.5) = \left(\frac{19}{20}\right)\left(\frac{18}{19}\right)\left(\frac{17}{18}\right)\left(\frac{13}{14}\right)\left(\frac{11}{12}\right)\left(\frac{7}{8}\right)\left(\frac{6}{7}\right)\left(\frac{3}{4}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = .1357.$$

The estimate $VG_{\text{est}}(8.5)$ of the variance of $\bar{F}_n(8.5)$ according to the KMEE is

$$\begin{aligned} VG_{\text{est}}(8.5) &= (.1357)^2[(20 \times 19)^{-1} + (19 \times 18)^{-1} \\ &\quad + (18 \times 17)^{-1} + (14 \times 13)^{-1} + (12 \times 11)^{-1} + (8 \times 7)^{-1} \\ &\quad + (7 \times 6)^{-1} + (4 \times 3)^{-1} + (3 \times 2)^{-1} + (2 \times 1)^{-1}] \\ &= .0150. \end{aligned}$$

The estimate $VP_{\text{est}}(8.5)$ corresponding to the KMEG is

$$VP_{\text{est}}(8.5) = (.1357)^2[\exp(*) - 1] = .0231 > VG_{\text{est}}(8.5) = .0150,$$

where $*$ represents the expression in square brackets in the right-hand side of the formula for $VG_{\text{est}}(8.5)$.

Our (augmented) sample size is $n = 20$, our estimate of $E(FC)$ is $.5$, and our estimate of $\bar{F}(8.5)$ is $.14$. We now refer to Table 3. (In doing so we implicitly assume that both the survival-time distribution and the distribution of censoring times that gave rise to our data set can be represented by Weibull distributions.) We do not know the value of b/a but we see that if $b/a = 2$ then in the neighbourhood of $\{E(FC) = .5000, \bar{F}(8.5) = .1353\}$ the KMEE and the KMEG both are seriously biased, the KMEE on average underestimating the true value of $\bar{F}(8.5)$ by $.13$ and the KMEG on average overestimating $\bar{F}(8.5)$ by $.12$. Therefore, in the absence of knowledge about the value of b/a we should express the accuracy of our estimate of $\bar{F}(8.5)$ by (an estimate of) the MSE rather than by (an estimate of) the variance. Our estimates $VG_{\text{est}}(8.5)$ and $VP_{\text{est}}(8.5)$ also cannot be trusted in the neighbourhood of $\{E(FC) = .5000, \bar{F}(8.5) = .1353\}$, as the differences of the entries under "Var" and "Var_{app}" indicate.

The KMEE helps us out of this quandary of how to find an indication of the accuracy of the estimate of $\bar{F}(8.5)$. As can be seen in Table 3, its MSE is quite insensitive to the actual value of b/a and an MSE value of $.02$ gives a fair indication of the accuracy of the estimate $\bar{F}_n(8.5) = .14$, regardless of the unknown value of b/a . That the KMEE's MSE value of $.02$ is not far from the values of $VG_{\text{est}}(8.5) = .015$ and $VP_{\text{est}}(8.5) = .023$ is a lucky coincidence.

In extending the KME to values of t beyond the censored largest observation $t = 9.3$, the KMEE would yield the value zero with estimated variance $VG_{n,\text{est}} = 0$ and the KMEG would yield $.1357$ with estimated variance $VP_{n,\text{est}} = .0231$. Again, the estimates may be biased and their variance estimates cannot be trusted. However, for true values of $\bar{F}(t) \leq .1353$, Table 3 gives an MSE value—depending on the unknown value of $\frac{1}{2} \leq b/a \leq 2$ —between $.0228$ and 0 for the KMEE and an MSE between $.012$ and values greater than $.08$ for the KMEG. So, in the absence of knowledge of the value of b/a we would prefer the KMEE's estimate over the KMEG's.

ACKNOWLEDGEMENTS

I wish to thank the referees—one of whom supplied two additional references—for their helpful comments, and the editor for his very careful conduct of the reviewing process.

RÉSUMÉ

De nombreuses suggestions ont été faites pour étendre l'estimateur de Kaplan et Meier (1958, *Journal of the American Statistical Association* **53**, 457–481) à des temps de survie au-delà de la plus grande observation, dans le cas où celle-ci est censurée. On examine deux extensions extrêmes relativement au biais et à l'erreur quadratique moyenne (MSE). Leurs qualités dépendent considérablement de la "constellation des censures." L'erreur quadratique moyenne d'une des extensions semble robuste relativement à une large classe de non proportionnalités des taux de hasard des distributions de la durée de vie et de l'instant de censure.

REFERENCES

- Brown, B. W., Jr., Hollander, M., and Korwar, R. M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In *Reliability and Biometry: Statistical Analysis of Lifelength*, F. Proschan and R. J. Serfling (eds), 327–358. Philadelphia: Society for Industrial and Applied Mathematics.
- Chen, Y. Y., Hollander, M., and Langberg, N. A. (1982). Small-sample results for the Kaplan–Meier estimator. *Journal of the American Statistical Association* **77**, 141–144.
- Efron, B. (1967). The two-sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **4**, 831–852.
- Geurts, J. H. J. (1985). Some small-sample nonproportional hazards results for the Kaplan–Meier estimator. *Statistica Neerlandica* **39**, 1–13.
- Gill, R. D. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tract No. 124. Amsterdam: Mathematisch Centrum.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.
- Kardaun, O. (1983). Statistical survival analysis of male larynx cancer patients—a case study. *Statistica Neerlandica* **37**, 103–125.
- Moeschberger, M. L. and Klein, J. P. (1985). A comparison of several methods of estimating the survival function when there is extreme right censoring. *Biometrics* **41**, 253–259.
- Wellner, J. A. (1985). A heavy censoring limit theorem for the product limit estimator. *The Annals of Statistics* **13**, 150–162.

Received April 1985; revised February 1987.