

Estimation of Survival Distributions Under Right-Censoring When Sample Size Is Random

Agnieszka Rossa

Institute of Statistics and Demography, University of Łódź, Poland

e-mail: agrossa@uni.lodz.pl

Abstract: The paper deals with two classes of unbiased non-parametric estimators of survival and cumulative hazard functions in a population subject to right-censoring. Both classes of estimators are based on a sequential sampling scheme, and are similar to the well-known Kaplan-Meier and Nelson-Aalen estimators.

Keywords: Kaplan-Meier estimator; Nelson-Aalen estimator; Right-censoring; Sequential sampling; Survival analysis.

Subject Classifications: 62N02; 62L12.

1. INTRODUCTION

Let T be a continuous random variable denoting the time elapsed up to a well-defined event, hereafter referred to as the survival time. The survival function of T is then defined as

$$\bar{F}(x) = P(T > x), \quad (1.1)$$

and the cumulative hazard function for $x \in (0, \sup\{y : F(y) < 1\})$ takes the form

$$\Lambda(x) = \int_0^x \frac{dF(u)}{\bar{F}(u)}, \quad \text{where } F = 1 - \bar{F}. \quad (1.2)$$

Let Z be a non-negative random variable which censors on the right the survival time T . Denote by G the cumulative distribution function of Z . We will assume that T and Z are independent random variables in the same probability space $(\Omega, \mathcal{A}, P)$, and that E denotes the expectation with respect to P .

Due to right-censoring the variable T is possibly unobserved. The only available information is the smaller one of T and Z , that is

$$X = \min(T, Z), \quad (1.3)$$

and the censoring indicator

$$\Delta = \mathbf{1}(X = T), \quad (1.4)$$

where $\mathbf{1}(\cdot)$ denotes the indicator function.

Throughout the rest of the paper X will be called a censored survival time.

It follows from (1.3) that X is a random variable with a cumulative distribution function H , say, which is equal to

$$H = 1 - \bar{F}\bar{G}, \quad (1.5)$$

where $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$.

The Kaplan-Meier estimator (KM) (Kaplan and Meier 1958) and the Nelson-Aalen estimator (NA) (Nelson 1969; Aalen 1978) are usually used to estimate the survival and cumulative hazard functions (1.1) and (1.2), respectively, under random censorship model (1.3)–(1.4).

Statistical properties of KM and NA have been widely studied (see e.g., Efron 1967; Breslow and Crowley 1974; Peterson 1977; Winter et al. 1978; Chen et al. 1982; Chang 1991; Klein 1991; Stute 1994a,b; Pawlitschko 1999; Satten and Datta 2001). It has been shown, among other things, that both estimators are biased. The Kaplan-Meier and Nelson-Aalen estimators are closely related and have also been studied with the use of the counting processes and martingale theory (see Aalen 1978).

The Kaplan-Meier estimator $KM(x)$ is undefined if x is greater than the largest value of X observed in a sample and if the observation is right-censored. Such a disadvantage makes it sometimes impossible to estimate $\bar{F}(x)$ at a fixed point x , especially if x is large.

Gajek and Gather (1991) considered estimation of a survival distribution to be an element of a scale family $\{F_\theta, \theta \in \Theta\}$ of distributions. They showed that under Type I censoring the lower bound of the mean squared error of an estimator of θ^s is equal to 1, thus it is independent of the sample size n . It is well-known that the mean squared error can be expressed as the sum of the variance and the squared bias. As the variance approaches 0 for sufficiently large n , it follows from their result that under Type I censoring there does not exist an unbiased estimator of θ^s based on a fixed-size sample. The more general conclusion, adequate to the non-parametric right-censorship model, is that under Type I censoring there does not exist an unbiased estimator of a distribution function if the sample size is fixed.

In the paper sequential estimators of the survival function $\bar{F}$ and the cumulative hazard function Λ based on random-size samples are considered. A sequential approach in the estimation under the right-censorship model was used in the past (e.g., Gardiner and Susarla 1983, 1989). The approach taken here deals with estimating the survival and cumulative hazard functions by means of some unbiased and consistent estimators, which are similar to the well-known Kaplan-Meier and Nelson-Aalen estimators. The adopted approach allows to estimate $\bar{F}(x)$ or $\Lambda(x)$ at any fixed point x .

The paper is organized as follows. Section 2 introduces the standard Kaplan-Meier and Nelson-Aalen estimators. In Section 3, a special sequential sampling scheme is proposed. Section 4 presents random-sample-size estimators of $\bar{F}(x)$ and $\Lambda(x)$ and gives some their statistical properties. Some simulation results are given in Section 5 and Section 6 contains discussions.

2. KAPLAN-MEIER AND NELSON-AALEN ESTIMATORS

Consider a sample of the form

$$(X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_n, \Delta_n), \quad (2.1)$$

where (X_i, Δ_i) , $i = 1, 2, \dots, n$ are independent copies of (X, Δ) , and n is a fixed positive integer.

The original Nelson-Aalen estimator can be expressed as follows

$$NA(x) = \sum_{i=1}^n \frac{\mathbf{1}(X_i \leq x, \Delta_i = 1)}{\sum_{j=1}^n \mathbf{1}(X_j \geq X_i)}. \quad (2.2)$$

Let $X_{n:n} = \max\{X_1, X_2, \dots, X_n\}$ and $\Delta_{[n]}$ be an indicator variable concomitant to $X_{n:n}$, i.e., $\Delta_{[n]} = \Delta_j$ if $X_{n:n} = X_j$. The Kaplan-Meier estimator can be then defined as

$$KM(x) = \begin{cases} \prod_{\{i: X_i \leq x\}} \left(1 - \frac{\Delta_i}{\sum_{j=1}^n \mathbf{1}(X_j \geq X_i)}\right) & \text{for } x \leq X_{n:n} \\ 0 & \text{for } x > X_{n:n}, \text{ if } \Delta_{[n]} = 1 \\ \text{undefined} & \text{for } x > X_{n:n}, \text{ if } \Delta_{[n]} = 0, \end{cases} \quad (2.3)$$

under an initial assumption that $KM(x) = 1$ if $\{i : X_i \leq x\} = \emptyset$.

Evaluation of $KM(x)$ is much easier if we use a formula based on the so-called ordered sample.

Let (x_i, δ_i) be a realization of (X_i, Δ_i) and $\leq_l$ be a relation defined as follows

$$(x_i, \delta_i) \leq_l (x_j, \delta_j) \iff (x_i < x_j) \vee (x_i = x_j \wedge \delta_i \geq \delta_j). \quad (2.4)$$

The ordered sample can be expressed as

$$(X_{1:n}, \Delta_{[1]}), (X_{2:n}, \Delta_{[2]}), \dots, (X_{n:n}, \Delta_{[n]}), \quad (2.5)$$

where $(X_{i:n}, \Delta_{[i]})$ represents an i -th observation in the sequence (2.1) ordered according to the relation $\leq_l$.

Now (2.3) is equivalent to

$$KM(x) = \begin{cases} \prod_{\{i: X_{i:n} \leq x\}} \left(1 - \frac{\Delta_{[i]}}{n-i+1}\right) & \text{for } x \leq X_{n:n} \\ 0 & \text{for } x > X_{n:n}, \text{ if } \Delta_{[n]} = 1 \\ \text{undefined} & \text{for } x > X_{n:n}, \text{ if } \Delta_{[n]} = 0. \end{cases} \quad (2.6)$$

3. SEQUENTIAL SAMPLING SCHEME

In this section a special type of the sequential sampling scheme is introduced. This scheme supplies random-size samples and allows to construct unbiased versions of the Kaplan-Meier and Nelson-Aalen estimators.

Assume that we observe a sequence $(X_1, \Delta_1), (X_2, \Delta_2), \dots$ until for a fixed number $k \geq 2$ of individuals we get $X_{i_j} \geq x_0$, $j = 1, 2, \dots, k$, where x_0 is a fixed, positive real value such that $x_0 < \sup\{x : H(x) < 1\}$ and H is a common cumulative distribution function of the X_i 's.

Let N_k be a total number of individuals observed. It follows that N_k is a random variable distributed according to the negative binomial distribution with parameters k and $p = 1 - H(x_0)$. Its probability distribution function takes the form

$$P(N_k = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n = k, k+1, \dots$$

The proposed sampling scheme provides us with the random-size sample

$$(X_1, \Delta_1), (X_2, \Delta_2), \dots, (X_{N_k}, \Delta_{N_k}). \quad (3.1)$$

The ordered sample can be derived from (3.1) by the use of the relation $\leq_l$

$$(X_{1:N_k}, \Delta_{[1]}), (X_{2:N_k}, \Delta_{[2]}), \dots, (X_{N_k:N_k}, \Delta_{[N_k]}). \quad (3.2)$$

4. THE PROPOSED ESTIMATORS

Definition 4.1. *Sequential Nelson-Aalen estimators of the cumulative hazard $\Lambda(x)$, $x \leq x_0$ are given by the formula*

$$NA_k(x) = \sum_{i=1}^{N_k} \frac{\mathbf{1}(X_i \leq x, \Delta_i = 1)}{\sum_{j=1}^{N_k} \mathbf{1}(X_j \geq X_i) - 1}, \quad x \leq x_0, \quad k \in \mathbf{N}, k \geq 2. \quad (4.1)$$

Definition 4.2. *Sequential Kaplan-Meier estimators of the survival probability $\bar{F}(x)$, $x \leq x_0$ are given by the formula*

$$KM_k(x) = \prod_{\{i: X_i \leq x\}} \left(1 - \frac{\Delta_i}{\sum_{j=1}^{N_k} \mathbf{1}(X_j \geq X_i) - 1}\right), \quad x \leq x_0, \quad k \in \mathbf{N}, k \geq 2, \quad (4.2)$$

under an initial assumption that $KM_k(x) = 1$ if $\{i : X_i \leq x\} = \emptyset$.

Note that (4.1) as well as (4.2) define two classes of estimators of $\Lambda(x)$ and $\bar{F}(x)$, respectively, for any fixed integer $k \geq 2$ and for any fixed real value $x_0 \in (0, \sup\{x : H(x) < 1\})$.

It is worth also noting that on the right-hand sides of (4.1) and (4.2) there is $\sum_{j=1}^{N_k} \mathbf{1}(X_j \geq X_i) - 1$, as opposed to the sum $\sum_{j=1}^n \mathbf{1}(X_j \geq X_i)$ appearing in (2.2) and (2.3).

The formula equivalent to (4.2) is based on the ordered sample (3.2)

$$KM_k(x) = \prod_{\{i: X_{i:N_k} \leq x\}} \left(1 - \frac{\Delta_{[i]}}{N_k - i}\right), \quad x \leq x_0, \quad k \in \mathbf{N}, k \geq 2. \quad (4.3)$$

Proposition 4.1. *The sequential estimators $NA_k(x)$ are unbiased estimators of $\Lambda(x)$ for $x < x_0$. The variance of $NA_k(x)$ is expressed by the following equivalence*

$$V(NA_k(x)) = \int_0^x E\left(\frac{1}{M_k(u)}\right) d\Lambda(u) \quad \text{for } x \in (0, x_0), \quad (4.4)$$

where $M_k(u) = \sum_{j=1}^{N_k} \mathbf{1}(X_j > u)$, and the expectation $E(1/M_k(u))$, $u \in [0, x]$ is equal to

$$E\left(\frac{1}{M_k(u)}\right) = \left(\frac{p_u}{q_u}\right)^k \int_0^{q_u} v^{k-1} (1-v)^{-k} dv, \quad (4.5)$$

where $p_u = \bar{H}(x_0)/\bar{H}(u)$, $q_u = 1 - p_u$.

Proposition 4.2. *The sequential estimators $KM_k(x)$ are unbiased estimators of $\bar{F}(x)$ for $x \leq x_0$. If censoring times are fixed non-negative values then the variance of $KM_k(x)$ satisfies the inequality*

$$V(KM_k(x)) < \prod_{j=0}^J U_j - \bar{F}^2(x) \quad \text{for } x \in (0, x_0], \quad (4.6)$$

$$\text{where } U_j = p_{j+1|j}^2 + q_{j+1|j} \left[\frac{p_{j+1|j}}{p_j} \right]^k E\left(\frac{1}{M_{k,j}}\right) - q_{j+1|j}^2 \left[\frac{p_{j+1|j}}{p_j} \right]^k E\left(\frac{1}{M_{k,j}+1}\right), \quad (4.7)$$

$$M_{k,j} = \sum_{i=1}^{N_k} \mathbf{1}(X_i > y_j),$$

$$E\left(\frac{1}{M_{k,j}}\right) = \left(\frac{p_j}{q_j}\right)^k \int_0^{q_j} u^{k-1} (1-u)^{-k} du, \quad (4.8)$$

$$E\left(\frac{1}{M_{k,j} + 1}\right) = \left(\frac{p_j}{q_j}\right)^k \frac{1}{q_j} \int_0^{q_j} u^k (1-u)^{-k} du, \quad (4.9)$$

$$p_{j+1|j} = \bar{F}(y_{j+1}) / \bar{F}(y_j), \quad q_{j+1|j} = 1 - p_{j+1|j}, \quad p_j = \bar{H}(x_0) / \bar{H}(y_j), \quad q_j = 1 - p_j, \quad (4.10)$$

and $y_j, j = 1, 2, \dots, J$ are fixed censoring times, such that $0 < y_1 < \dots < y_J < x$.

Proofs of the Propositions 4.1 and 4.2 were given by Rossa (2005, pp. 40-56).

As it is only slight difference between modified and the standard estimators, the Kaplan-Meier variance estimator $\hat{V}(KM_k(x))$ can be defined as

$$\hat{V}(KM_k(x)) = KM_k^2(x) \sum_{i=1}^{N_k} \frac{\mathbf{1}(X_i \leq x, \Delta_i = 1)}{(M_{k,i} - 1)^2}, \quad x \in (0, x_0], \quad (4.11)$$

while the Nelson-Aalen variance estimator $\hat{V}(NA_k(x))$ can be given as an empirical counterpart of (4.4)

$$\hat{V}(NA_k(x)) = \sum_{i=1}^{N_k} \frac{\mathbf{1}(X_i \leq x, \Delta_i = 1)}{M_{k,i}^2}, \quad x \in (0, x_0), \quad (4.12)$$

where $M_{k,i} = \sum_{j=1}^{N_k} \mathbf{1}(X_j > X_i)$.

Example 4.1. We shall illustrate the idea of the provided estimators using the data on durability of artificial heart valves.

An artificial heart valve is a device which is implanted in the heart of patients who suffer from valvular diseases in their heart. There are two main types of artificial valves, i.e. mechanical and biological ones. When one or two out of four natural valves of the heart (i.e. tricuspid, pulmonic, mitral or aortic) has a malfunction then a standard procedure is to replace the damaged valve by an artificial one. This requires an open-heart surgery.

Let us consider a population $\mathcal{G}$ of patients who have received a biological valve of a given type. Some of them have to be re-operated due to malfunctions of the implant. Let the subject of observation be the time T_i which elapsed up to the first re-operation of an i -th patient randomly drawn from

$\mathcal{G}$. It is clear that the true survival times T_i can be observed for those patients who were re-operated by the time of the analysis, for other patients survival times are right-censored. Moreover, some other random causes can yield independent right-censoring.

Due to censoring survival times T_i are unobserved random variables, but we can observe pairs (X_i, Δ_i) , where $X_i = \min(T_i, Z_i)$, is a censored survival time of an i -th patient, $\Delta_i = \mathbf{1}(X_i = T_i)$ is an indicator variable, and Z_i is a censoring time.

Let $\bar{F}$ and Λ denote a common survival function and a common cumulative hazard function of the T_i 's, respectively. Suppose that we need to estimate $\bar{F}(x)$ and $\Lambda(x)$ at any $x \leq 240$ (months).

As it was mentioned in Section 1, the original Kaplan-Meier and Nelson-Aalen estimators give biased estimates of $\bar{F}(x)$ and $\Lambda(x)$, respectively. Their bias is not negligible, especially if there is a large number of censored observations in a sample. What is more, the Kaplan-Meier estimator $KM(x)$ can be undefined if x exceeds the largest value of censored survival times observed in a sample. Thus, it can be impossible to estimate $\bar{F}(x)$ at a fixed x using the KM estimator.

A sequential approach described in Sections 3 and 4 gives unbiased estimators of $\bar{F}(x)$ and $\Lambda(x)$ at any $x \leq x_0$, where x_0 is fixed in advance. However, a specific sampling scheme has to be applied.

Let us assume that patients are successively drawn from $\mathcal{G}$ until censored survival times X_i (i.e. times to re-operation, possibly censored) of k of them exceed x_0 months. Let $k = 2$, $x_0 = 240$ (months) and assume that the following sample is observed

$$\begin{array}{ccccccc} (10, 1), & (244, 0), & (44, 1), & (210, 1), & (167, 0), & (151, 0), & (74, 1), \\ (238, 0), & (141, 0), & (96, 1), & (135, 0), & (54, 0), & (125, 0), & (119, 0), \\ (114, 0), & (109, 0), & (75, 0), & (133, 1), & (23, 1), & (10, 0), & (241, 0). \end{array}$$

Using the $\leq_l$ relation, the ordered sample takes the form

$$\begin{array}{cccccccc} (10, 1), & (10, 0), & (23, 1), & (44, 1), & (54, 0), & (74, 1), & (75, 0), \\ (96, 1), & (109, 0), & (114, 0), & (119, 0), & (125, 0), & (133, 1), & (135, 0), \\ (141, 0), & (151, 0), & (167, 0), & (210, 1), & (238, 0), & (241, 0), & (244, 0). \end{array}$$

The survival and cumulative hazard functions estimated by means of $KM_k(x)$ and $NA_k(x)$ for $k = 2$ and $x \in [0, 240]$ are given in (4.13) and

(4.14), respectively. Both estimators are plotted on Figures 1 and 2.

$$KM_2(x) = \begin{cases} 1 & \text{for } x < 10 \\ 0.950 & \text{for } 10 \leq x < 23 \\ 0.897 & \text{for } 23 \leq x < 44 \\ 0.844 & \text{for } 44 \leq x < 74 \\ 0.788 & \text{for } 74 \leq x < 96 \\ 0.727 & \text{for } 96 \leq x < 133 \\ 0.636 & \text{for } 133 \leq x < 210 \\ 0.424 & \text{for } 210 \leq x \leq 240, \end{cases} \quad (4.13)$$

$$NA_2(x) = \begin{cases} 0 & \text{for } x < 10 \\ 0.050 & \text{for } 10 \leq x < 23 \\ 0.106 & \text{for } 23 \leq x < 44 \\ 0.165 & \text{for } 44 \leq x < 74 \\ 0.232 & \text{for } 74 \leq x < 96 \\ 0.309 & \text{for } 96 \leq x < 133 \\ 0.434 & \text{for } 133 \leq x < 210 \\ 0.767 & \text{for } 210 \leq x \leq 240. \end{cases} \quad (4.14)$$

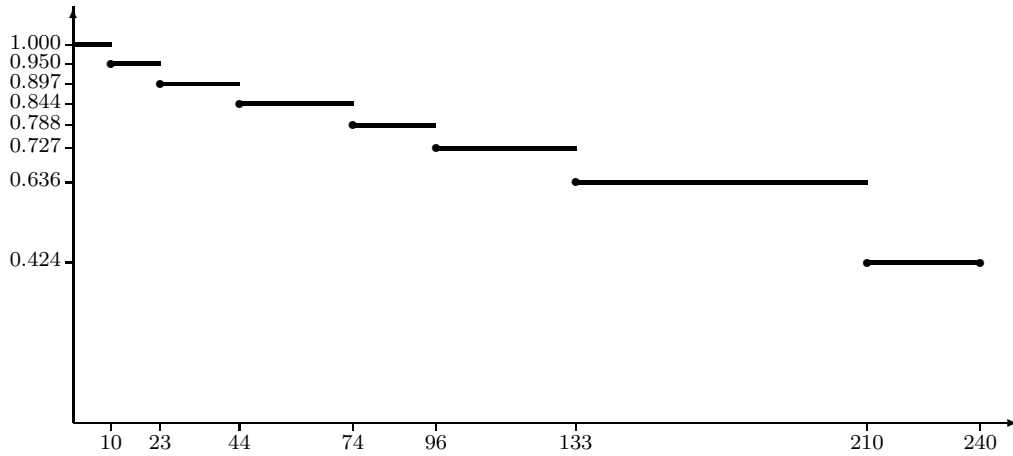


Figure 1. The Kaplan-Meier curve $KM_2(x)$, $x \in [0, 240]$.

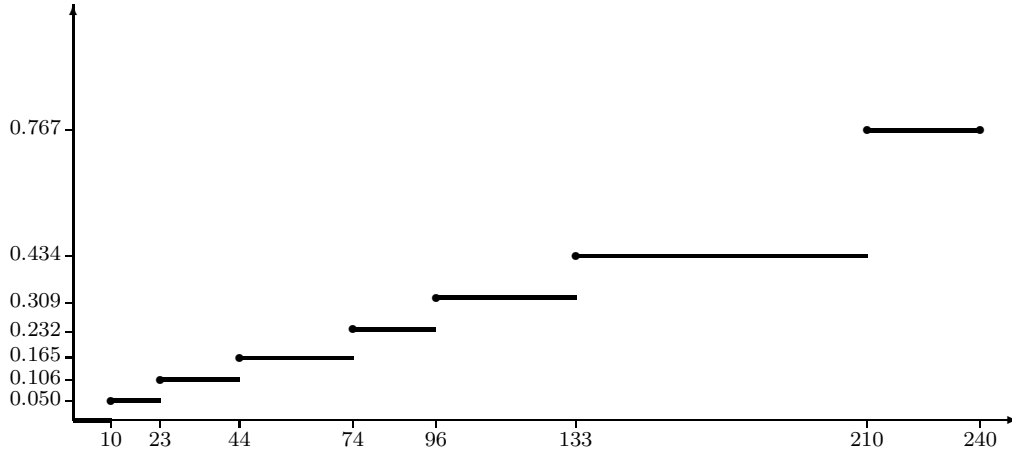


Figure 2. The Nelson-Aalen curve $NA_2(x)$, $x \in [0, 240]$.

5. A SIMULATION STUDY

To see the dependency of the variances of $KM_k(x)$ and $NA_k(x)$ on k and x_0 and to study some properties of the variance estimators $\hat{V}(KM_k(x_i))$ and $\hat{V}(NA_k(x_i))$ a simulation analysis was performed. In the study survival times T_i were simulated from:

- gamma distribution $\Gamma(\alpha, \beta)$ with the probability distribution function proportional to $x^\alpha \exp\{-\beta x\}$,
- Weibull distribution $\text{Wei}(\beta, \gamma)$ with the survival function (SDF) equal to $\exp\{-\beta x^\gamma\}$; a special case of this family of distributions is the exponential distribution $\text{Exp}(\beta)$ (for $\gamma = 1$),
- log-normal distribution $\log N(\mu, \sigma)$,
- Gompertz distribution $\text{Gom}(\beta, \gamma)$ with SDF equal to $\exp\{\gamma(1 - \exp\{-\beta x\})\}$,
- Pareto distribution $\text{Par}(\beta, \gamma)$ with SDF equal to $(1 + \beta x)^{-\gamma}$,
- Log-logistic distribution $\log L(\beta, \gamma)$ with SDF equal to $1/(1 + \beta x^\gamma)$.

Censoring times Z_i were simulated from an exponential distribution $\text{Exp}(\beta)$ with a fixed value of the mean time to censoring β yielding an assumed

censoring fraction $p = P(T_i > Z_i)$. Both survival and censoring times were then used to determine pairs (X_i, Δ_i) , where $X_i = \min(T_i, Z_i)$ and $\Delta_i = \mathbf{1}(X_i = T_i)$.

In the simulation study a prescribed number $M = 10000$ of repetitions was considered. In each repetition pairs $(X_1, \Delta_1), (X_2, \Delta_2), \dots$ were simulated until for k of them the inequality $X_{i_j} \geq x_0$, $j = 1, 2, \dots, k$ was satisfied, where a positive real value x_0 and an integer k were fixed in advance. Next, the estimators (4.1), (4.2), (4.11), (4.12) at some $x \in [0, x_0]$ were evaluated. Finally, the variances $V(KM_k(x))$, $V(NA_k(x))$ of the estimators (4.1), (4.2) as well as the expectations $E(\hat{V}(KM_k(x)))$, $E(\hat{V}(NA_k(x)))$ of the estimators (4.11), (4.12) were approximated by means of the following formulae

$$V(KM_k(x)) \approx \frac{1}{M} \sum_{j=1}^M \left(KM_k^{\{j\}}(x) - \bar{F}(x) \right)^2, \quad (5.1)$$

$$V(NA_k(x)) \approx \frac{1}{M} \sum_{j=1}^M \left(NA_k^{\{j\}}(x) - \Lambda(x) \right)^2, \quad (5.2)$$

$$E(\hat{V}(KM_k(x))) \approx \frac{1}{M} \sum_{j=1}^M \hat{V}^{\{j\}}(KM_k(x)), \quad (5.3)$$

$$E(\hat{V}(NA_k(x))) \approx \frac{1}{M} \sum_{j=1}^M \hat{V}^{\{j\}}(NA_k(x)). \quad (5.4)$$

where $KM_k^{\{j\}}(x)$, $NA_k^{\{j\}}(x)$, $\hat{V}^{\{j\}}(KM_k(x))$ and $\hat{V}^{\{j\}}(NA_k(x))$ denote estimates of the respective estimators obtained in an j -th repetition.

Figures 3 and 4 exhibit typical behavior of $V(KM_k(x))$, $V(NA_k(x))$ and $E(\hat{V}(KM_k(x)))$, $E(\hat{V}(NA_k(x)))$ for various k , x_0 for the whole range of $x \in [0, x_0]$.

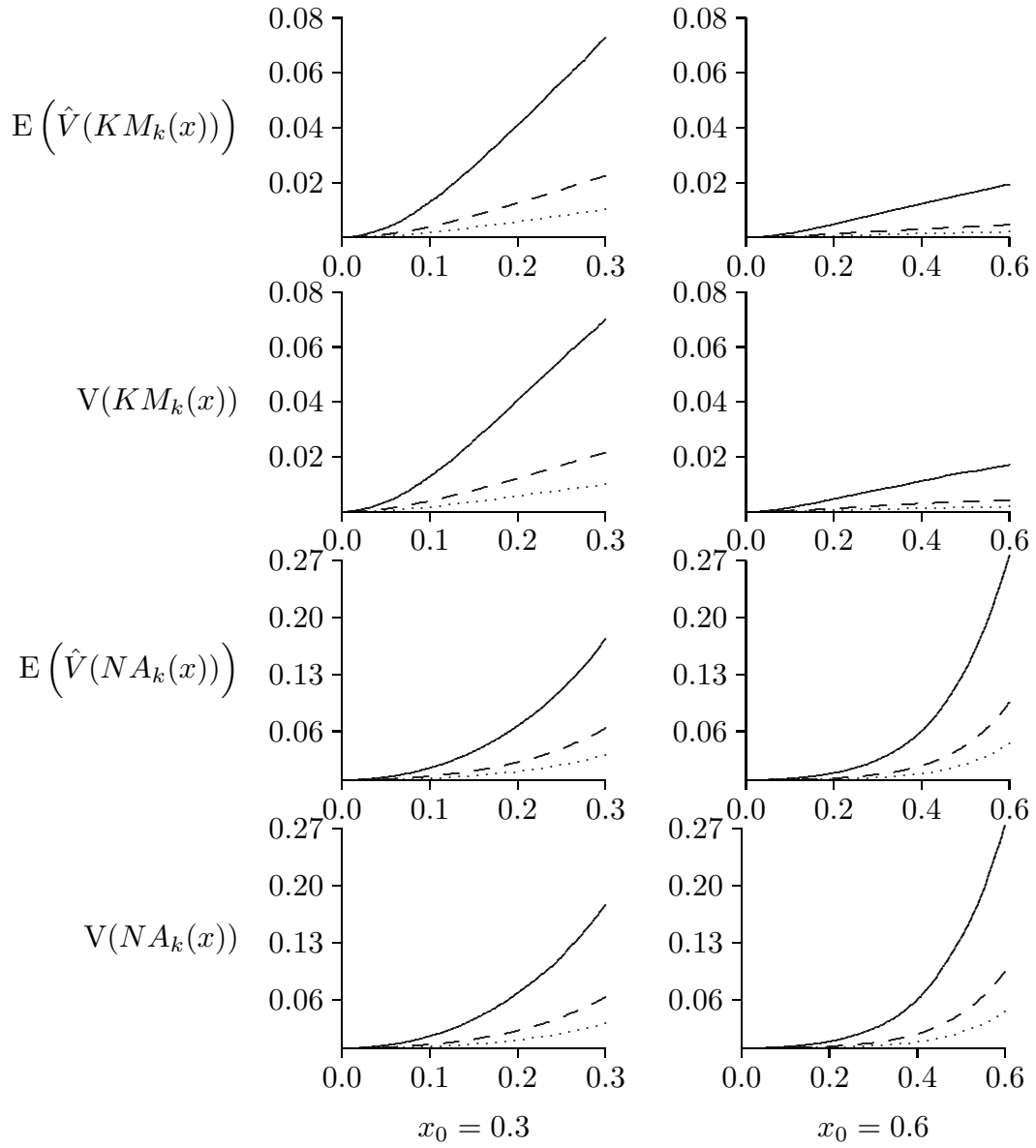


Figure 3. Expectations $E(\hat{V}(KM_k(x)))$, $E(\hat{V}(NA_k(x)))$ and variances $V(KM_k(x))$, $V(NA_k(x))$ for $x \in [0, x_0]$, $k = 2, 5, 10$ (solid, dashed and dotted lines, respectively), $F \sim \Gamma(2; 5)$, censoring fraction $p \approx 0.7$.

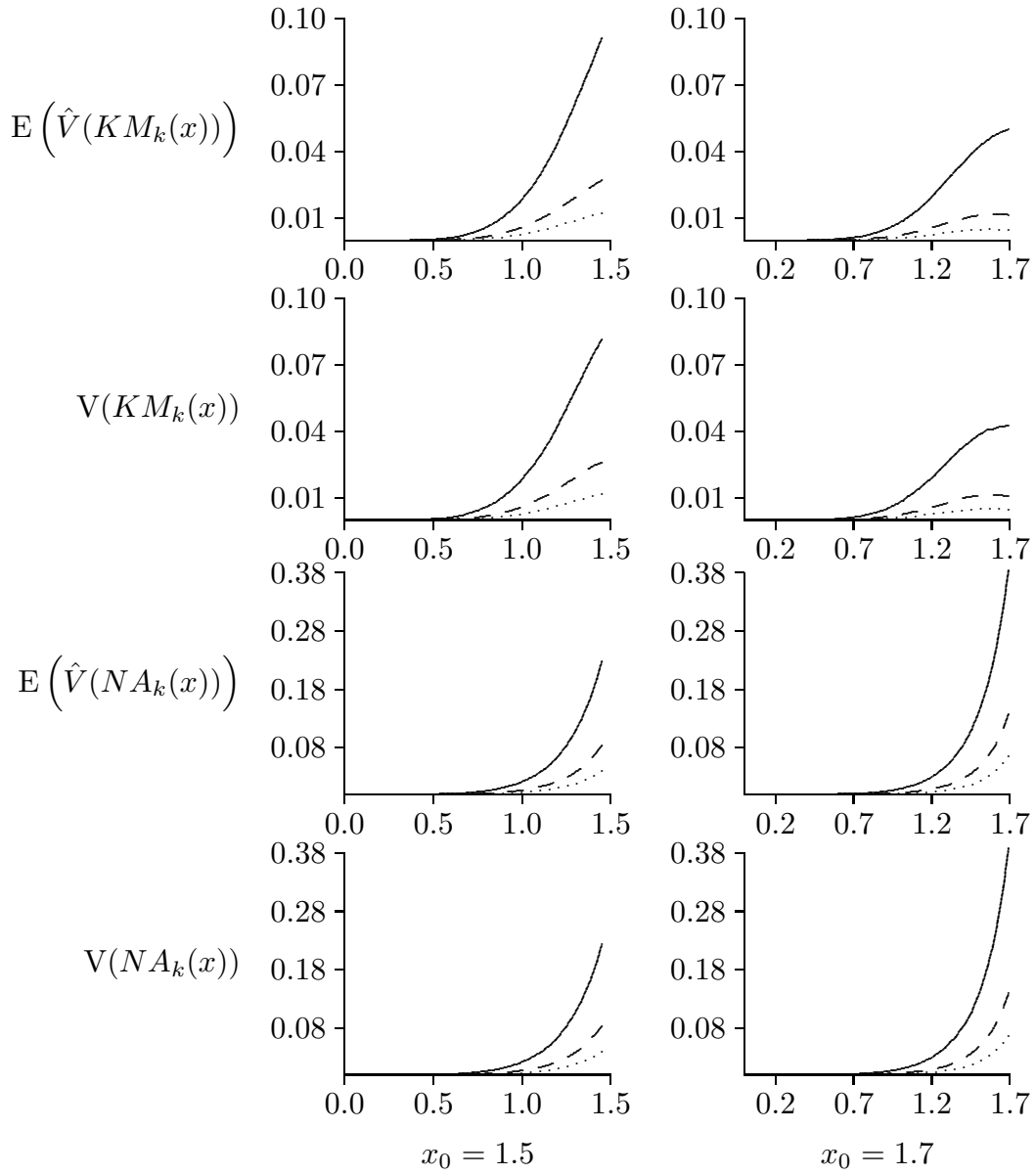


Figure 4. Expectations $E(\hat{V}(KM_k(x)))$, $E(\hat{V}(NA_k(x)))$ and variances $V(KM_k(x))$, $V(NA_k(x))$ for $x \in [0, x_0]$, $k = 2, 5, 10$ (solid, dashed and dotted lines, respectively), $F \sim \text{Wei}(0.1; 5)$, censoring fraction $p \approx 0.7$.

6. DISCUSSIONS

In the paper two classes of sequential estimators $NA_k(x)$ and $KM_k(x)$ were proposed. Both classes are based on a specific sequential sampling scheme. In the scheme two parameters have to be fixed in advance, i.e. an integer $k \geq 2$ and a positive value x_0 such that $x_0 < \sup\{x : H(x) < 1\}$. In order to choose a proper value of x_0 , even if H remains unknown, it is sufficient to know the maximal possible values t and z , say, of the survival and censoring times, respectively. Then for any $x_0 \in (0, \min(t, z))$ there is $H(x_0) < 1$.

It is worth also noting that $NA_k(x)$ and $KM_k(x)$ are consistent estimators because they are unbiased with variances converging to 0 as $E(N_k) \rightarrow \infty$.

ACKNOWLEDGMENTS

I thank a referee and the Associate Editor for their helpful remarks and comments. I am grateful to the Editor-in-Chief for spending his own valuable time on improving the layout of the manuscript.

REFERENCES

- Aalen, O. (1978). Nonparametric Inference for a Family of Counting Processes, *Annals of Statistics* 6: 701-726.
- Breslow, N. E. and Crowley, J. J. (1974). Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship, *Annals of Statistics* 2: 437-454.
- Chang, M. N. (1991). Moments of the Kaplan-Meier Estimator, *Sankhya, Series A* 53: 27-50.
- Chen, Y. Y., Hollander, M., and Langberg, N. A. (1982). Small-Sample Results for the Kaplan-Meier Estimator, *Journal of American Statistical Association* 77: 141-144.
- Efron, B. (1967). The Two-Sample Problem with Censored Data, in *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability* 4, pp. 831-853, Berkeley: University of California Press.

- Gajek, L. and Gather, U. (1991). Estimating a Scale Parameter Under Censorship, *Statistics* 22: 529-549.
- Gardiner, J. C. and Susarla, V. (1983). Sequential Estimation of the Mean Survival Time Under Random Censorship, *Sequential Analysis* 2: 201-223.
- Gardiner, J. C. and Susarla, V. (1989). The Performance of a Sequential Procedure for the Estimation of the Median Survival Time with Specified Accuracy, *Sequential Analysis* 8: 191-204.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations, *Journal of American Statistical Association* 53: 457-481.
- Klein, P. J. (1991). Small Sample Moments of Some Estimators of the Variance of the Kaplan-Meier and Nelson-Aalen Estimators, *Scandinavian Journal of Statistics* 18: 333-340.
- Nelson, W. B. (1969). Hazard Plotting for Incomplete Failure Data, *Journal of Quality Technology* 1: 27-52.
- Pawlitschko, J. (1999). A Comparison of Survival Function Estimators in the Koziol-Green Model, *Statistics* 32: 277-291.
- Peterson, A. V. (1977). Expressing the Kaplan-Meier Estimator as a Function of Empirical Subsurvival Functions, *Journal of American Statistical Association* 72: 854-858.
- Rossa, A. (2005). *Estimation of Survival Distributions Under Right-Censoring and Applications*, monograph in Polish, Łódź: University of Łódź Press.
- Satten, G. A. and Datta, S. (2001). The Kaplan-Meier Estimator as an Inverse-Probability-of-Censoring Weighted Average, *American Statistician* 55: 207-209.
- Stute, W. (1994a). Improved Estimation Under Random Censorship, *Communications in Statistics – Theory & Methods* 23: 2671-2682.

Stute, W. (1994b). The Bias of Kaplan-Meier Integrals, *Scandinavian Journal of Statistics* 21: 475-484.

Winter, B. B., Földes, A., and Rejtő, L. (1978). Gliwienko-Cantelli Theorems for the PL Estimate, *Problems of Control and Information Theory* 7: 213-225.