Developments in Applied Statistics
Anuška Ferligoj and Andrej Mrvar (Editors)
Metodološki zvezki, 19, Ljubljana: FDV, 2003

A Non-parametric Mean Residual Life Estimator: An Example from Market Research

Kai Kopperschmidt[1] and Ulrich Pötter[2]

Abstract

The mean residual life (mrl) function dynamically describes the average time to an event, depending on the time since the previous event. It provides a forecast in parallel with the development of the underlying process. From a theoretical point of view, the mrl characterises the distribution of the process completely, but in contrast to other characterisations like the hazard rate, it has a direct interpretation in terms of average behaviour.

We use Kaplan–Meier integrals (weighted averages of residual times) to construct a nonparametric estimator of the mrl. We use results from Stute (1995) and Yang (1994) to describe the asymptotic behaviour of this estimator and derive an approximate variance formula.

We present a small simulation study and apply the estimator (and the variance formula) to data pertaining to purchase time behaviour from the Homescan Panel$^{\text{TM}}$, A.C. Nielsen, Germany.

1 Introduction

The mean residual life (mrl) function dynamically describes the average time to an event, depending on the time since the previous event. As an important example for functionals of the Kaplan–Meier estimator it has been studied by many authors, e.g., Gill (1983), Gijbels and Veraverbeke (1991), Yang (1994) and Stute (1995). From a theoretical point of view, the mrl characterises the distribution of the process completely, see e.g. Shaked and Shanthikumar (1991). The mrl function can therefore be used in model formulation just as densities or hazard functions are. The mrl function is defined as a conditional expectation of the time to an event given that that time is larger than a given value. Its computation thus involves integrals over unbounded intervals of the real line. While conditional expectations are easily interpreted and often the object of immediate interest in applications, the fact that integrals over unbounded domains are involved severely hampers the analysis

[1] A.C. Nielsen, Frankfurt, Germany; Kai.Kopperschmidt@Germany.ACNielsen.com
[2] University of Bochum, Bochum, Germany; Ulrich.Poetter@ruhr-uni-bochum.de

of estimators in the presence of censoring. This issue has been discussed extensively by Stute and Wang (1993), where a strong law of large numbers is given for such functionals. Instead of integrating with respect to the cumulative hazard rate – as was done in previous work – the authors use integrals with respect to the distribution function of the underlying random variable. Along the same lines, Stute (1995) provides a central limit theorem in this situation. A crucial role is played by the Kaplan–Meier weights, see Stute and Wang (1993, p. 1593) and Stute (1995, p. 423), respectively. These quantities generalise the well-known weights for non-parametric estimation in sampling theory, where the inverse inclusion probabilities are used as weighting factors. In the presence of independent censoring, however, the suitable weights are stochastic and vary between the observations.

2 An example from market research

A central goal of market research is to describe the market performance of "fast moving" consumer goods (fmcg). Those are products which are perishable or quickly used up, like food or detergents in contrast to cars, washing machines etc. Throughout this paper, we primarily have fmcg in mind when we talk about products. Both manufacturers and retailers have a strong interest in identifying the more or less successful items of a product class (pc). One way of doing this is to collect data on the purchase behaviour of households. Specifically, we may observe the purchase acts of the participating consumers for a pc of interest during a fixed period, say one year. We call the duration between two consecutive purchase acts by the same household an interpurchase time (ipt). As a pc consists of several items, different types of ipt occur: On the aggregate level, we have the time between two purchases in the pc, while for each product, there is also an item-specific ipt starting and ending with a purchase of the specific article.

In order to reduce problems of dependencies between observations we use only *one* ipt of each type per household in the sample. When interpreting the data, we also have to be aware of possible censorings. They occur at the end of the observation period, in case no repurchase has taken place by then.

Statistical inference concerning durations falls into the realm of survival analysis, where the most common quantity is the hazard rate. In this paper, however, we will focus on the mean residual life function (mrl) instead: Denote by T the random variable describing the length of the ipt and by F its cumulative distribution function. Then the mrl $m(t)$ at some time t equals the average remaining time to repurchase, given this event has not yet taken place:

$$m(t) := \mathbb{E}_T(T - t \mid T > t) = \frac{\int_t^\infty u - t \, dF(u)}{1 - F(t)} \tag{2.1}$$

We will discuss the mrl in more detail in Section 3. For the moment, we only note that the mrl is a conditional expectation and thus describes the mean behaviour of consumers. This feature makes it an interesting quantity for traditional marketing, where it is impossible to focus on specific, single consumers and one has to deal with the whole group of households instead. As an illustration, we present data pertaining to purchase time behaviour from the Homescan Panel™ of A. C. Nielsen, Germany. It is based on approximately 5,800 households surveyed in 2001. We focus on two items of an anonymised pc. Central facts of this pc and the two selected items, 'focus' and 'competitor', are given in the next table:

	penetration	market share	proportion of repurchasing HH
Class	0.58	-	0.78
Focus	0.03	0.014	0.50
Comp.	0.14	0.077	0.61

In words, 58% of the German households purchased the pc in 2001, 3% bought the focus item and 14% bought the competitor item at least once. With market shares of 1.4% and 7.7%, neither of the two products has an overwhelming influence on the development of the pc. On the aggregate level, censoring occurred in 22% of the observations. On the item level, censoring is rather heavy, 50% and 39%, respectively. In Section 4, we present a non-parametric estimator for the mrl in this situation, see (4.3). Applied to our data truncated at $\tau^* = 365$ days[2], the resulting curves are given in Figure 1.

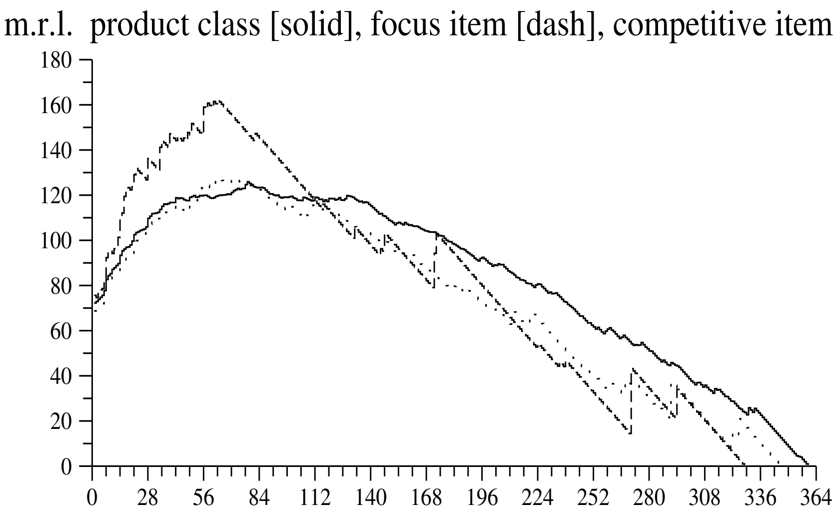

Figure 1: Estimated mrl functions according to (4.3) for the data presented in Section 2.

On a qualitative level, the estimated mrls suggest the following interpretation: The mrl of the pc serves as a reference. To some extent, it provides information

[2]Truncation to $[0, 365]$ means that we used only repurchase times that happened within a year. We are thus looking at the conditional mrl given that repurchase happened within a year. Purchase data from the panel households is available for the following months as well, if we neglect the slight but inevitable panel mortality. Consequently, one might aim at extending the truncation time τ^*, thus diminishing the loss of information. On the other hand, for a household to contribute to the actually used purchase data, it is required that its overall reporting quality exceeds a certain level over the whole period in question. As a result, extending the relevant observation period leads to a decreased sample size presumably offsetting any benefits.

on how the average consumer purchases the typical item of this pc. The mrl of the competitor stays rather close to the pc's mrl. This implies that its consumers perceive this specific item to 'represent' the pc: whenever the average consumer repurchases the pc, the customer of the competitor's item repurchases this specific product. Clearly, this makes the competitor a successful article within the pc. From the focus item's point of view, the situation is considerably less comfortable: Approximately two months (56 days) after the last purchase act, the average remaining time to repurchase of the focus item is 40 days longer than it is for the pc.

To determine whether the differences between the estimated mrl functions really support these arguments or merely occurred 'by chance', we have to introduce a stochastic model of our situation and estimate variances of the mrl estimates as well. We suggest several variance formulae in Section 5 and evaluate their performance in section 6. Concerning the real life data, it turns out that the corresponding estimations do not render any differences between the mrls significant.

3 Mean Residual Life function

The mrl function defined in (2.1) determines the distribution uniquely. If F is absolutely continuous one can compute the distribution function from the mrl function by

$$1 - F(t) = \frac{m(0)}{m(t)} \exp\left\{-\int_0^t \frac{1}{m(u)}\, du\right\}, \tag{3.1}$$

see Shaked and Shanthikumar (1991, p. 614). The mrl function can thus be used as a characterisation of the distribution in the same way as the density or the hazard function can.

As an example, consider a mrl function which is linear on some interval $I = [t_1, t_2]$. For $t \in I$, the mrl thus has the form

$$m(t) = a \cdot (t - t_1) + m(t_1),$$

where $m(t_1) > 0$ and $a \geq -1$. In case $a < 0$, we also have to ensure that $t_2 \leq t_1 - \frac{m(t_1)}{a}$.

For $a \neq 0$, the distribution function on I then equals a Pareto distribution scaled by $F(t_1)$:

$$F(t) = F(t_1) \cdot \left(\frac{a \cdot (t - t_1)}{m(t_1)} + 1\right)^{-\frac{a+1}{a}} \quad , \quad t \in I$$

Specifically, if $a = -0.5$, the distribution function is linear and the distribution on I is consequently uniform.

For $a = 0$, the distribution function on I is exponential:

$$F(t) = F(t_1) \cdot \exp\left(-\frac{t - t_1}{m(t_1)}\right)$$

Thus, if the mrl function is given on the interval I we get the behaviour of F restricted to I, and complete knowledge of F on I if $F(t_1)$ was known. Note that both formulae simplify considerably in terms of 'local coordinates' $F(t)/F(t_1)$ and $(t - t_1)/m(t_1)$.

In the presence of censoring there may be a point in time after which it is impossible to gather further information. Let τ be the least upper bound on the period of observation. Then the best one can hope for is to estimate integrals up to that point in time. The definition of the mrl function, however, requires an evaluation of the integrals to ∞, not only up to τ. Moreover, the truncation point τ is generally unknown (and not easily estimable).

We will therefore consider to estimate

$$m^*(t, \tau^*) := \mathbb{E}(T - t \,|\, T \in (t, \tau^*)) = \frac{\int_t^{\tau^*} u - t \, dF(u)}{\int_t^{\tau^*} dF(u)} \tag{3.2}$$

for some $\tau^* \leq \tau$. This truncated mrl function m^* need not be related to the mrl function m in any obvious way. However, m^* is still a conditional expectation with an easy interpretation and important applications. Moreover, as a generalisation of (3.1), the truncated mrl function uniquely identifies the distribution function up to τ^* through

$$F(\tau^*) - F(t) = \frac{m^*(0, \tau^*)}{m^*(t, \tau^*)} \exp\left(- \int_0^t \frac{du}{m^*(u, \tau^*)} \right) \tag{3.3}$$

To see this, note that for continuously differentiable (in t) m^*

$$\begin{aligned} \frac{\partial m^*(t, \tau^*)}{\partial t} &= \frac{\partial}{\partial t} \frac{\int_t^{\tau^*} u - t \, dF(u)}{\int_t^{\tau^*} dF(u)} \\ &= \left(-tf(t) - \int_t^{\tau^*} f(u) \, du + tf(t) \right) \frac{1}{F(\tau^*) - F(t)} \\ &\quad + \frac{f(t)}{(F(\tau^*) - F(t))^2} \int_t^{\tau^*} (u - t) f(u) \, du \\ &= -1 + m^*(t, \tau^*) \frac{f(t)}{F(\tau^*) - F(t)} \end{aligned}$$

Thus

$$-\frac{\partial}{\partial t} \ln \left(F(\tau^*) - F(t) \right) = \frac{\frac{\partial m^*(t, \tau^*)}{\partial t} + 1}{m^*(t, \tau^*)}$$

so that

$$F(\tau^*) - F(t) = \exp\left(- \int_0^t \frac{\partial m^*(v, \tau^*)/\partial v \,|_{v=u}}{m^*(u, \tau^*)} + \frac{1}{m^*(u, \tau^*)} \, du \right)$$

from which (3.3) follows by noting that the first fraction under the integral is the logarithmic derivative of $m^*(u, \tau^*)$.

4 Kaplan–Meier integrals

With complete data an estimator of the mrl function is

$$\widehat{m}(t) := \frac{\int_t^{\infty} u - t \, dF_n(u)}{\int_t^{\infty} dF_n(u)} = \frac{\sum_{i=1}^n (t_i - t) \mathbf{1}[t_i > t]}{\sum_{i=1}^n \mathbf{1}[t_i > t]} \tag{4.1}$$

With censored observations it is natural to replace F_n by the Kaplan-Meier estimator

$$1 - \hat{F}_n(t) = \prod_{i=1}^{n} \left(1 - \frac{\delta_{(i)}}{n - i + 1}\right)^{\mathbf{1}[z_{(i)} \leq t]} \tag{4.2}$$

where $z_i = \min\{t_i, u_i\}$ are the censored observations, $\delta_i = \mathbf{1}[t_i \leq u_i]$ are the censoring indicators and $z_{(1)} \leq \ldots \leq z_{(n)}$ are the ordered values of the observed times while the $\delta_{(i)}$ are the corresponding censoring indicators. For a given upper limit τ^*, the equation (4.1) becomes

$$\widehat{m^*}(t, \tau^*) := \frac{\int_t^{\tau^*} u - t \, d\hat{F}_n(u)}{\int_t^{\tau^*} d\hat{F}_n(u)} = \frac{\sum_{i=1}^{n} (t_i - t)\mathbf{1}[t_i > t] w_i}{\sum_{i=1}^{n} \mathbf{1}[t_i > t] w_i} \tag{4.3}$$

where the w_i are the jump sizes $\hat{F}_n(t_i) - \hat{F}_n(t_i-)$ of the Kaplan–Meier estimator. Note that if the largest observation is censored we will not force $1 - \hat{F}_n$ to be zero after that observation.

In order to analyse the estimator (4.3) we propose a simple stochastic model that will allow the calculation of approximate variances (within that model, of course) and that might be used to gauge the performance of the estimator on real data sets. We will assume that the observations arise from independent and identically distributed copies of the random variables (T, U), where T and U are independent and T has distribution function F while U has distribution function G. Then $1 - H(t) = (1 - F(t))(1 - G(t))$ is the survivor function of the random variable $Z := \min\{T, U\}$. We set $\delta := \mathbf{1}(T \leq U)$ and let $\tau := \inf\{t \mid H(t) = 1\}$ be the least upper bound of the support of Z. In the following we will assume that F is absolutely continuous while we allow for an arbitrary distribution of G. Note, however, that Stute's (1995) results are valid for general F and general G, while Yang's results (1994) allow G to vary with the observations but requires absolute continuous F. Our restriction to absolutely continuous F and identically distributed censoring times is however often applicable and simplifies the formulations considerably. Extending the integrals with respect to $\hat{F}_n$ up to τ requires delicate considerations on the behaviour of $\hat{F}_n$ near τ. The point is well discussed by Stute and Wang (1993) and by Gill (1994). We will here avoid an explicit discussion by either truncating to τ^* or by assuming $\tau = \infty$. In the simulations we will use a distribution function F with a compact support strictly included in the support of G.

5 Variance formulae

The weights in the formula for $\widehat{m^*}$ will in the presence of censoring depend on all the censored observations preceding a given event time. The weights w_i are therefore not independent random variables in our model. There is, however, a general representation of Kaplan–Meier integrals in terms of sums of independent random variables given by Stute (1995, p. 425). Using that representation, standard results on sums of independent random variables can be used to derive variances of the estimator $\widehat{m^*}$. In the case of absolutely continuous F, the representation simplifies considerably and can be written as

$$\int_0^\infty \phi_k(u)\, d\hat{F}_n(u) - \int_0^\tau \phi_k(u)\, dF(u) =$$
$$\frac{1}{n}\sum_{i=1}^n \int_0^\infty \frac{\phi_k(u) - \mathbb{E}(\phi_k(T)\,|\,T > u)}{1 - G(u_-)}\, dM_i(u) + o_P(n^{-1/2}) \quad k = 1, 2$$

where

$$\phi_1(u) := \mathbf{1}[u > t](u - t) \quad , \quad \phi_2(u) := \mathbf{1}[u > t]$$

and where

$$M_i(t) := \mathbf{1}[t_i \le t, \delta_i = 1] - \int_0^t \frac{\mathbf{1}[z_i \ge u]}{1 - F(u)}\, dF(u)$$

is the martingale for the counting process $\mathbf{1}[t_i \le t, \delta_i = 1]$ with respect to the standard filtration (see the Appendix for a derivation).

The representation gives among other results a variance formula via a standard martingale argument (details are deferred to the Appendix):

$$\mathrm{Var}\left(\int_0^\infty \frac{\phi_k(u) - \mathbb{E}(\phi_k(T)\,|\,T > u)}{1 - G(u_-)}\, dM_i(u)\right) = \tag{5.1}$$
$$\int_0^\infty \frac{(\phi_k(u) - \mathbb{E}(\phi_k(T)\,|\,T > u))^2}{1 - G(u_-)}\, dF(u) =: \sigma_k^2\ ,\ k = 1, 2$$

This formula was also derived via a direct argument by Yang (1994).

While the above formulae are useful in theoretical work, practical computations will also have to rely on a direct empirical counterpart of Stute's representation. We start from the ordered values $z_{(i)}, i = 1, \ldots, n$ and the corresponding values $\delta_{(i)}$ and $\phi_{ki} := \phi_k(z_{(i)})$. We then define (assuming no ties):

$$\begin{aligned}
\gamma_i &:= \exp\left(\sum_{j=1}^{i-1} \frac{1 - \delta_{(j)}}{n - j}\right) \quad i = 2, \ldots, n \quad \gamma_1 := 1 \\
a_{ki} &:= \delta_{(i)}\phi_{ki}\gamma_i \quad i = 1, \ldots, n \\
b_{ki} &:= \frac{1 - \delta_{(i)}}{n - i} \sum_{j=i+1}^n a_{kj} \quad i = 1, \ldots, n-1 \quad b_{kn} := 0 \\
c_{ki} &:= \sum_{j=1}^{i-1} \frac{b_{kj}}{n - j} \quad i = 2, \ldots, n \quad c_{k1} := 0
\end{aligned}$$

Finally, we set

$$A_{ki} := a_{ki} + b_{ki} - c_{ki} \quad i = 1, \ldots, n,\ k = 1, 2 \tag{5.2}$$

Writing

$$\widehat{m^*}(t, \tau^*) = \frac{\int_t^{\tau^*} u - t\, d\hat{F}_n(u)}{\int_t^{\tau^*}\, d\hat{F}_n(u)} =: \frac{\hat{A}_1}{\hat{A}_2} \tag{5.3}$$

we get an expression for the variance of the fraction using the delta method:

$$\operatorname{Var}\left(\widehat{m}^*(t,\tau^*)\right) \approx \frac{\hat{A}_{1n}^2\hat{\sigma}_2^2}{n\hat{A}_{2n}^4} + \frac{\hat{\sigma}_1^2}{n\hat{A}_{2n}^2} - 2\frac{\hat{A}_{1n}}{n\hat{A}_{2n}^3}\operatorname{Cov}(\hat{A}_{1n},\hat{A}_{2n}) \tag{5.4}$$

where $\hat{\sigma}_1^2$ is the estimated variance of $\hat{A}_1$ and $\hat{\sigma}_2^2$ is the estimated variance of $\hat{A}_2$.

6 A small simulation study

The variance approximations of the previous section can be implemented in various ways. In this section we will use simulations to evaluate some of the possible choices. To keep things simple, we use only the exponential distribution with expectation 1 truncated to $[0,2]$ for F, i.e. $F(t) = (1-\exp(-t))/(1-\exp(-2))$ on $t \in [0,2]$. The truncation assures finiteness of all moments with respect to F and thus all regularity requirements for the validity of the variance formulae are fulfilled. We evaluate the mrl at 0.2, 0.5, 1.0 and 1.5. The mrl at these points is 0.6435, 0.5691, 0.4180 and 0.2293. The survival probabilities $1-F(t)$ at the evaluation points are 0.79, 0.55, 0.27, and 0.10. As censoring distributions we use the exponential distribution with expectations 5 and 1 corresponding to censoring probabilities of 0.12 and 0.43, respectively. The corresponding expected proportions at risk are 0.65, 0.33, 0.10, and 0.02 using a censoring distribution with expectation 1, and 0.76, 0.49, 0.22, and 0.08 using a censoring distribution with expectation 5. Sample sizes are 200 and 1000. We use 1000 simulations for each of the combinations.

Table 1 gives the mrl at the evaluation points together with summary statistics for the estimated mrl from the simulations. The mrl estimator is slightly downward biased for smaller numbers of observations and for later times. This was to be expected from the supermartingal structure of Kaplan-Meier integrals as discussed by Stute and Wang (1993) and the bias of the Kaplan-Meier estimator itself as given in Fleming/Harrington (1991, p. 99). Possible consequences and remedies are described by Miller (1983) and Gill (1994, section 8).

In our situation, the asymptotic variances from (5.1) and (5.4) can be computed explicitly. They are also given in Table 1. For $n = 1000$, the agreement with the variances estimated in the simulation experiment is excellent. For $n = 200$, however, the variances of the mrl estimates tend to be larger than the asymptotic variances suggest. Note that the variances are multiplied by 100 in the case of $n = 200$ and by 1000 in the case of $n = 1000$.

Turning to the empirical variance estimators, a simple approach would be to use the empirical variance of the terms A_{1i}/A_{2i}, thus avoiding the use of (5.4). Even when using the empirical variances of the A_{ki} separately, one might try simpler versions of (5.4). Yang (1994, p. 342) proposed to use just the variance of the numerator. Since the denominator is consistent, an appeal to Slutsky's lemma shows this to be a consistent variance estimator. But this would work well only if the expectation of the numerator would be 0. A further possibility is to use both empirical variances, but to ignore the covariance. But the covariance is by definition far from 0. In fact, in our simulation setup, the results using these three estimators

Table 1: Simulation results for the mrl estimator.

Experiment		t=0.2	t=0.5	t=1.0	t=1.5
true mrl		0.6435	0.5692	0.4180	0.2293
$n = 200$, $U \simeq \exp(1)$	Min	0.4637	0.3561	0.1457	0.0020
	Mean	0.6376	0.5628	0.4087	0.2199
	Max	0.8173	0.7816	0.7404	0.4931
	NA	0	0	0	30
	Var ($\times 10^2$)	0.3130	0.4463	0.6818	0.7556
	as. Var	0.2769	0.3667	0.5123	0.5139
$n = 200$, $U \simeq \exp(0.2)$	Min	0.5315	0.4427	0.2859	0.0747
	Mean	0.6442	0.5703	0.4178	0.2296
	Max	0.7486	0.7047	0.6058	0.3622
	NA	0	0	0	0
	Var ($\times 10^2$)	0.1534	0.1847	0.2046	0.1588
	as. Var	0.1647	0.1823	0.1886	0.1401
$n = 1000$, $U \simeq \exp(1)$	Min	0.5718	0.4926	0.3319	0.1178
	Mean	0.6427	0.5688	0.4166	0.2278
	Max	0.7180	0.6598	0.5164	0.3521
	NA	0	0	0	0
	Var ($\times 10^3$)	0.5558	0.7852	1.0237	1.1719
	as. Var	0.5538	0.7334	1.0247	1.0277
$n = 1000$, $U \simeq \exp(0.2)$	Min	0.5837	0.5144	0.3590	0.1891
	Mean	0.6434	0.5693	0.4177	0.2298
	Max	0.7070	0.6351	0.4843	0.2824
	NA	0	0	0	0
	Var ($\times 10^3$)	0.3353	0.3854	0.3766	0.2688
	as. Var	0.3295	0.3646	0.3773	0.2802

had no clear relation with the observed variance of the mrl. We will therefore not present the results in our simulation reports.

A second approach uses the Kaplan–Meier estimator (and its estimated variance) in the denominator of the mrl and in its variance estimator. We will not pursue this possibility here, since we are mainly interested in the performance of Kaplan–Meier integrals. We will therefore also use only the Kaplan–Meier integrals in empirical versions of (5.1).

A third possibility is a direct application of the discrete representation of Stute. This means to compute the terms in (5.2) and use the empirical means, variances, and covariances of these terms in (5.4). We denote this estimator by $\hat{\sigma}^2_S$.

In $\hat{\sigma}^2_S$, one might replace the terms $\hat{A}_{kn}$ by the respective terms used in the mrl estimator. This should give estimators that are better centered and possibly be less variable. We denote this estimator by $\hat{\sigma}^2_{mrl}$.

Lastly, we use an empirical version of Yang's formula (5.1) where we use the estimated mrl for the inner expectation and the empirical Kaplan–Meier integrals for the outer integral. The denominator in the integrand is estimated by the Kaplan–

Meier estimator. Moreover, in (5.4), we use the mrl estimators for the $\hat{A}_{kn}$ terms. This version will be denoted by σ_Y^2.

Note that, from a computational point of view, $\hat{\sigma}_Y^2$ requires the computation of the mrl at all (unique) points z_i. In contrast, both $\hat{\sigma}_{mrl}^2$ and $\hat{\sigma}_S^2$ can easily be computed point-wise but are relatively more expensive when the variances are required at all observed points. All three variants, however, are computationally much cheaper than a bootstrap approach.

Table 2: Simulation results for variance estimators, $n = 200$.

$U \simeq \exp(1)$		t=0.2	t=0.5	t=1.0	t=1.5
	sim. Var ($\times 10^2$)	0.3130	0.4463	0.6818	0.7556
$\hat{\sigma}_S^2$	Min	0.0012	0.0013	0.0009	0.0000
	Mean ($\times 10^2$)	0.4104	0.5134	0.6052	0.3152
	Max	0.0250	0.0299	0.0313	0.0241
$\hat{\sigma}_{mrl}^2$	Min	0.0012	0.0013	0.0009	0.0000
	Mean ($\times 10^2$)	0.4484	0.5801	0.7441	0.4086
	Max	0.0349	0.0514	0.0532	0.0380
$\hat{\sigma}_Y^2$	Min	0.0011	0.0012	0.0009	0.0000
	Mean ($\times 10^2$)	0.2858	0.3899	0.6235	1.6433
	Max	0.0072	0.0155	0.0512	0.3453
$U = \exp(0.2)$					
	sim. Var ($\times 10^2$)	0.1534	0.1847	0.2046	0.1588
$\hat{\sigma}_S^2$	Min	0.0011	0.0012	0.0010	0.0003
	Mean ($\times 10^2$)	0.1675	0.1853	0.1897	0.1438
	Max	0.0041	0.0049	0.0067	0.0060
$\hat{\sigma}_{mrl}^2$	Min	0.0011	0.0012	0.0010	0.0003
	Mean ($\times 10^2$)	0.1682	0.1863	0.1917	0.1474
	Max	0.0044	0.0053	0.0080	0.0097
$\hat{\sigma}_Y^2$	Min	0.0011	0.0012	0.0011	0.0004
	Mean ($\times 10^2$)	0.1686	0.1888	0.2032	0.1959
	Max	0.0023	0.0029	0.0047	0.0101

Looking first at the results for the case $n = 200$ and $U \simeq \exp(0.2)$ (Table 2), all three variance estimators are in rather close agreement with the simulated variances, even at $t = 1.5$. Moreover, $\hat{\sigma}_S^2 \leq \hat{\sigma}_{mrl}^2 \leq \hat{\sigma}_Y^2$ in the mean over all simulations, where also the variability of the estimators increases in this order.

The situation is less favourable in the case of heavy censoring and $n = 200$, also given in Table 2. Here, all estimators show a large variability. For $t \leq 1$, $\hat{\sigma}_Y^2$ seems to work best. However, at $t = 1.5$, none of the estimators is even close to the variability of the mrl estimator. But note that in this case the asymptotic variance is also not close to the observed variability of the mrl estimator.

Turning to the case $n = 1000$ with light censoring (Table 3), all three variance estimators are close together and close to the simulated variances. Once again, we find $\hat{\sigma}_S^2 \leq \hat{\sigma}_{mrl}^2 \leq \hat{\sigma}_Y^2$ in the mean over all simulations. We had expected to see $\hat{\sigma}_Y^2$

Table 3: Simulation results for variance estimators, $n = 1000$.

$U \simeq \exp(1)$		t=0.2	t=0.5	t=1.0	t=1.5
	sim. Var ($\times 10^3$)	0.5558	0.7852	1.0237	1.1719
$\hat{\sigma}^2_S$	Min	0.0004	0.0005	0.0006	0.0003
	Mean ($\times 10^3$)	0.6155	0.8133	1.1273	1.1074
	Max	0.0039	0.0052	0.0069	0.0041
$\hat{\sigma}^2_{mrl}$	Min	0.0004	0.0005	0.0006	0.0004
	Mean ($\times 10^3$)	0.6233	0.8281	1.1671	1.1989
	Max	0.0043	0.0061	0.0098	0.0092
$\hat{\sigma}^2_Y$	Min	0.0004	0.0005	0.0006	0.0003
	Mean ($\times 10^3$)	0.5579	0.7436	1.0644	1.2491
	Max	0.0007	0.0010	0.0017	0.0036
$U \simeq \exp(0.2)$					
	sim. Var ($\times 10^3$)	0.3353	0.3854	0.3766	0.2688
$\hat{\sigma}^2_S$	Min	0.0003	0.0003	0.0003	0.0002
	Mean ($\times 10^3$)	0.3302	0.3658	0.3781	0.2826
	Max	0.0004	0.0004	0.0005	0.0005
$\hat{\sigma}^2_{mrl}$	Min	0.0003	0.0003	0.0003	0.0002
	Mean ($\times 10^3$)	0.3305	0.3661	0.3787	0.2834
	Max	0.0004	0.0004	0.0005	0.0005
$\hat{\sigma}^2_Y$	Min	0.0003	0.0003	0.0003	0.0002
	Mean ($\times 10^3$)	0.3311	0.3676	0.3837	0.2978
	Max	0.0004	0.0004	0.0005	0.0005

perform less satisfactorily than the other two estimators since with a small amount of censoring the explicit use of the Kaplan–Meier estimator $1 - \hat{G}$ as a weight in (5.1) might result in unstable behaviour. But this seems not to be the case here.

In the case with heavy censoring ($U \simeq \exp(1)$) we see that $\hat{\sigma}^2_Y$ compares favourably with the other two estimators: It is somewhat closer to the simulated variances and has less variability. To look closer at the problem with the other two estimators, we sampled one of the A_{ki} from each of the 1000 simulation runs. Density estimates of the numerator and denominator variables are given in Figure 2. The distributions of the empirical A_k terms are far from normal. They are multimodal with one of the modes close to 0 and they have rather heavy tails. Moreover, the empirical versions of the A_{ki} are dependent so that the variances of the A_{ki} are difficult to estimate accurately. This might explain the larger variability of $\hat{\sigma}^2_S$ and $\hat{\sigma}^2_{mrl}$ compared to $\hat{\sigma}^2_Y$. Moreover, with a larger proportion of censored observations the estimator of the distribution of the censoring variable stabilises and thus also Yang's estimator stabilises.

In conclusion, our limited experience suggests that the variance estimator $\hat{\sigma}^2_Y$ is to be preferred in cases of heavy censoring, while with light censoring all estimators behave similarly.

Lastly, we look at a borderline case where the expectation in (5.1) is finite but where the bias condition (1.6) of Stute (1995, p. 425) is violated. Suppose F is exponential with expectation 1 and G is exponential with expectation 5. In this case, the mrl of F is constant 1. The function $C(x)$ is given by

$$C(x) = \int_0^{x-} \frac{1}{(1-H(u))(1-G(u))}\, dG(u) = \frac{0.2}{1.2}\left(e^{1.2x} - 1\right)$$

and thus

$$\int \phi_k(x)\sqrt{C(x)}\, dF(x)$$

diverges for $k = 1, 2$. With a sample size of $n = 1000$ (Table 4), the variance estimators $\hat{\sigma}^2_{mrl}$ and $\hat{\sigma}^2_S$ are rather larger than the simulated variances while $\hat{\sigma}^2_Y$ is still quite close. Moreover, the latter is much less variable than the other two. In fact, looking at the behaviour of the A_{ki}, they show very heavy tails with occasional huge values.

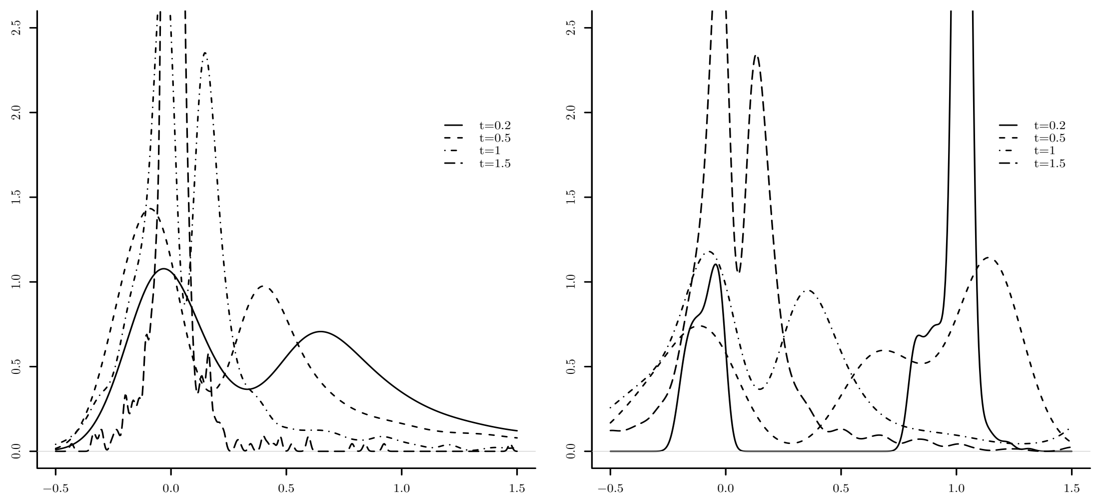

Figure 2: Density estimates for A_{1i} (left) and A_{2i} (right). Solid line: $t = 0.2$, short dashed line: $t = 0.5$, dashed and dotted line: $t = 1.0$, long dashed line: $t = 1.5$.

Looking at a sample size of $n = 200$ (Table 5), the mrl has a somewhat larger downward bias. The variance estimators based on the A_{ki} are now rather far from the simulated variances especially at larger t. The estimator $\hat{\sigma}^2_Y$ is closer to the simulated variances. In conclusion, there seems to be some leeway to improve on variance estimators based on the A_{ki}, possibly also in the case of light censoring.

Table 4: Simulation results, $F \simeq \exp(1)$, $G \simeq \exp(0.2)$, $n = 1000$.

		t=0.2	t=0.5	t=1.0	t=1.5
mrl	Min	0.8742	0.8424	0.7749	0.7495
	Mean	0.9950	0.9936	0.9906	0.9836
	Max	1.1390	1.1382	1.2160	1.2711
	Var ($\times 10^3$)	1.5778	2.2273	4.0604	7.5099
$\hat{\sigma}^2_S$	Min	0.0010	0.0013	0.0019	0.0025
	Mean ($\times 10^3$)	1.8755	2.7261	5.0650	9.3061
	Max	0.0148	0.0233	0.0551	0.1101
$\hat{\sigma}^2_{mrl}$	Min	0.0010	0.0013	0.0019	0.0025
	Mean ($\times 10^3$)	1.8863	2.7482	5.1377	9.5365
	Max	0.0152	0.0242	0.0587	0.1216
$\hat{\sigma}^2_Y$	Min	0.0009	0.0012	0.0019	0.0025
	Mean ($\times 10^3$)	1.5599	2.2317	4.0548	7.3789
	Max	0.0047	0.0075	0.0196	0.0442

Table 5: Simulation results, $F \simeq \exp(1)$, $G \simeq \exp(0.2)$, $n = 200$.

		t=0.2	t=0.5	t=1.0	t=1.5
mrl	Min	0.7533	0.6608	0.6291	0.4774
	Mean	0.9857	0.9819	0.9772	0.9663
	Max	1.2871	1.3645	1.4910	1.7176
	Var ($\times 10^2$)	0.8320	1.2447	2.1862	4.1630
$\hat{\sigma}^2_S$	Min	0.0032	0.0036	0.0040	0.0043
	Mean ($\times 10^2$)	1.0050	1.4446	2.6152	4.5671
	Max	0.1212	0.1998	0.3797	0.6479
$\hat{\sigma}^2_{mrl}$	Min	0.0032	0.0036	0.0041	0.0043
	Mean ($\times 10^2$)	1.0286	1.4921	2.7655	5.0157
	Max	0.1368	0.2372	0.5054	0.8394
$\hat{\sigma}^2_Y$	Min	0.0033	0.0037	0.0042	0.0044
	Mean ($\times 10^2$)	0.7962	1.1502	2.1417	4.0573
	Max	0.0429	0.0661	0.1197	0.4599

References

[1] Fleming, T.R. and Harrington, D.P. (1991): *Counting Processes & Survival Analysis*. New York: Wiley.

[2] Gijbels, I. and Veraverbeke, N. (1991): Almost sure asymptotic representation for a class of functionals of the Kaplan–Meier estimator. *Ann. Statist.*, **19**, 1457-1470.

[3] Gill, R. (1983): Large sample behaviour of the product–limit estimator on the whole line. *Ann. Statist.*, **11**, 49–58.

[4] Gill, R. (1994): Lectures on survival analysis. In D. Bakry, R. D. Gill, and S. A. Molchanov (Eds.): *Lectures on Probability Theory*. Ecole d'Eté de Probabilités de Saint–Flour XXII. Berlin: Springer, 115-241.

[5] Miller, R.G. (1983): What price Kaplan–Meier?. *Biometrics*, **39**, 1077-1082.

[6] Shaked, M. and Shanthikumar, J.G. (1991): Dynamic multivariate mean residual life functions. *J. Appl. Prob.*, **28**, 613-629.

[7] Stute, W. (1995): The central limit theorem under random censorship. *Ann. Statist.*, **23**, 422-439.

[8] Stute, W. and Wang, J.-L. (1993): The strong law under random censorship. *Ann. Statist.*, **21**, 1591-1607.

[9] Yang, S. (1994): A central limit theorem for functionals of the Kaplan–Meier estimator. *Stat. and Prob. Letters*, **21**, 337-345.

Appendix

For a function ϕ with $\mathbb{E}(|\phi(T)|) < \infty$ which also meets the appropriately modified moment conditions (1.5) and (1.6) in Stute (1995, p. 425), Stute (1995) gives a representation of the Kaplan–Meier integral $\int_0^\tau \phi(u)\, d\hat{F}_n(u)$ in terms of sums of independent random variables up to $o_P(n^{-1/2})$. We specialise to absolutely continuous distributions F and G and assume $\tau = \infty$. This rather special case leads to a transparent derivation of the main variance formula and allows to compare the results of Stute (1995) with those of Yang (1994). In particular, Stute's moment condition (1.5) and Yang's condition (ii) (Yang 1994, p. 339) simply reads

$$\int_0^\infty \frac{\phi(u)^2}{1-G(u)}\, dF(u) < \infty$$

Stute's representations in the special case can be written

$$\frac{\delta\phi(Z)}{1-G(Z)} + (1-\delta)\frac{\mathbb{E}(\phi(T)\,|\,T > Z)}{1-G(Z)} - \int\int \frac{\phi(w)\mathbf{1}(v < Z)\mathbf{1}(v < w)}{1-H(v)}\, dF(w)\, d\Lambda^G(v) \tag{6.1}$$

where $\Lambda^G(t) := \int_0^t \frac{dG(u)}{1-G(u)}$ is the integrated hazard function of G. The expectation of the above expression with respect to (δ, Z) is easily seen to be $\mathbb{E}(\phi(T))$: Compute the conditional expectation of the first term given $\{T = t\}$ to see that the expectation of that term is $\mathbb{E}(\phi(T))$. For the last two terms, simply write out the expectation with respect to (δ, Z).

Subtracting the expectation $\mathbb{E}(\phi(T))$ and evaluating (6.1) at a fixed argument t

we get

$$\delta\left[\frac{\phi(t)}{1-G(t)}-\frac{\mathbb{E}(\phi(T)\,|\,T>t)}{1-G(t)}\right] \tag{6.2}$$
$$+\frac{\mathbb{E}(\phi(T)\,|\,T>t)}{1-G(t)}-\int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{1-H(v)}\,dF(w)\,d\Lambda^G(v)$$
$$-\mathbb{E}(\phi(T))$$

We will start by re-expressing the double integral. For this, note that $\Lambda^H=\Lambda^F+\Lambda^G$ from the definition of the distribution function of Z. Thus,

$$\begin{aligned}
&\int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{1-H(v)}\,dF(w)\,d\Lambda^G(v)\\
&= \int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{1-H(v)}\,dF(w)\,d(\Lambda^H(v)-\Lambda^F(v))\\
&= \int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{(1-H(v))^2}\,dH(v)\,dF(w)\\
&\quad -\int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{(1-H(v))(1-F(v))}\,dF(v)\,dF(w)\\
&= \int\left(\frac{1}{1-H(\min(t,w))}-1\right)\phi(w)\,dF(w)\\
&\quad -\int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{(1-H(v))(1-F(v))}\,dF(v)\,dF(w)
\end{aligned}$$

where we used Fubini's theorem in the second equation, and where the first term in the third equation results from a transform of variables.

The last three terms in (6.2) can thus be written as

$$\begin{aligned}
&\frac{\mathbb{E}(\phi(T)\,|\,T>t)}{1-G(t)}-\int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{1-H(v)}\,dF(w)\,d\Lambda^G(v)-\mathbb{E}(\phi(T))\\
&= \int_t^\infty\frac{\phi(w)}{1-H(t)}\,dF(w)-\int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{1-H(v)}\,dF(w)\,d\Lambda^G(v)\\
&\quad -\int_0^\infty\phi(w)\,dF(w)\\
&= \int_t^\infty\frac{\phi(w)}{1-H(t)}\,dF(w)-\int\left(\frac{1}{1-H(\min(t,w))}-1\right)\phi(w)\,dF(w)\\
&\quad +\int\!\!\int\frac{\phi(w)\mathbf{1}(v<t)\mathbf{1}(v<w)}{(1-H(v))(1-F(v))}\,dF(v)\,dF(w)-\int_0^\infty\phi(w)\,dF(w)\\
&= -\int_0^t\frac{\phi(w)}{1-H(w)}\,dF(w)+\int_0^t\frac{\mathbb{E}(\phi(T)\,|\,T>v)}{1-H(v)}\,dF(v)\\
&= -\int_0^t\frac{\phi(w)}{1-G(w)}\,d\Lambda^F(w)+\int_0^t\frac{\mathbb{E}(\phi(T)\,|\,T>v)}{1-G(v)}\,d\Lambda^F(v)\\
&= -\int\frac{\phi(w)-\mathbb{E}(\phi(T)\,|\,T>w)}{1-G(w)}\mathbf{1}(w<t)\,d\Lambda^F(w)
\end{aligned}$$

Defining

$$M_i(t) := \mathbf{1}[t_i \le t, \delta = 1] - \int_0^t \frac{\mathbf{1}[z_i \ge u]}{1 - F(u)} \, dF(u)$$

to be the martingale for the counting process $\mathbf{1}[t_i \le t, \delta_i = 1]$ with respect to the standard filtration, we can finally write (6.1) for the i–th observation as

$$\int \frac{\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u)}{1 - G(u)} \, dM_i(u) \tag{6.3}$$

In sum, we have the representation

$$\int_0^\infty \phi(u) \, d\hat{F}_n(u) - \int_0^\infty \phi(u) \, dF(u) = \tag{6.4}$$
$$\frac{1}{n} \sum_{i=1}^n \int_0^\infty \frac{\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u)}{1 - G(u)} \, dM_i(u) + o_P(n^{-1/2})$$

which is just (5.1).

With the last representation at hand it is easy to derive a variance expression using standard martingale arguments:

$$\begin{aligned}
&\mathrm{Var}_{Z,\delta}\left(\int_0^\infty \phi(u) \, d\hat{F}_n(u) - \int_0^\infty \phi(u) \, dF(u)\right) \\
&= \mathbb{E}_{Z,\delta}\left(\left(\int_0^\infty \phi(u) \, d\hat{F}_n(u) - \int_0^\infty \phi(u) \, dF(u)\right)^2\right) \\
&\approx \mathbb{E}_{Z,\delta}\left(\left(\int_0^\infty \frac{\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u)}{1 - G(u)} \, dM(u)\right)^2\right) \\
&= \mathbb{E}_{Z,\delta}\left(\int_0^\infty \left(\frac{\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u)}{1 - G(u)}\right)^2 d\left\langle M, M\right\rangle(u)\right) \\
&= \mathbb{E}_{Z,\delta}\left(\int \mathbf{1}(Z > u) \left(\frac{\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u)}{1 - G(u)}\right)^2 d\Lambda^F(u)\right) \\
&= \int\int \mathbf{1}(z > u) \left(\frac{\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u)}{1 - G(u)}\right)^2 \frac{1}{1 - F(u)} \, dF(u) \, dH(z) \\
&= \int (1 - H(u)) \frac{(\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u))^2}{(1 - H(u))(1 - G(u))} \, dF(u) \\
&= \int \frac{(\phi(u) - \mathbb{E}(\phi(T) \,|\, T > u))^2}{1 - G(u)} \, dF(u)
\end{aligned}$$

This is also Yang's (1994) variance formula valid for arbitrary G. We will also need the covariances of the representations for ϕ_1 and ϕ_2. Shortening $\phi_k(u) -$

$\mathbb{E}(\phi_k(T) \,|\, T > u)$ to $R(\phi_k)(u)$ we have by a similar reasoning as above:

$$\begin{aligned}
&\operatorname{Cov}\left(\int_0^\infty \frac{R(\phi_1)(u)}{1-G(u)}\, dM(u), \int_0^\infty \frac{R(\phi_2)(u)}{1-G(u)}\, dM(u)\right) \\
&= \mathbb{E}_{Z,\delta}\left(\int_0^\infty \frac{R(\phi_1)(u)R(\phi_2)(u)}{(1-G(u))^2}\, d\langle M, M\rangle(u)\right) \\
&= \mathbb{E}_{Z,\delta}\left(\int_0^\infty \mathbf{1}(Z > u) \frac{R(\phi_1)(u)R(\phi_2)(u)}{(1-G(u))^2}\, d\Lambda^F(u)\right) \\
&= \int\int \mathbf{1}(z > u) \frac{R(\phi_1)(u)R(\phi_2)(u)}{(1-G(u))^2}\, d\Lambda^F(u)\, dH(z) \\
&= \int_0^\infty \frac{R(\phi_1)(u)R(\phi_2)(u)}{1-G(u)}\, dF(u)
\end{aligned}$$