Statistics + Probability Letters, to appear

# Estimating the marginal survival function in the presence of time dependent covariates

GLEN A. SATTEN
*National Center for HIV, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30333 USA*

SOMNATH DATTA
*Department of Statistics, University of Georgia, Athens, GA 30602 USA*

JAMES ROBINS
*Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115 USA*

**Abstract.** We propose a new estimator of the marginal (overall) survival function of failure times that is in the class of survival function estimators proposed by Robins (1993). These estimators are appropriate when, in addition to (right-censored) failure times, we also observe covariates for each individual that affect both the hazard of failure and the hazard of being censored. The observed data are re-weighted at each failure time $t$ according to Aalen's linear model for the cumulative hazard for being censored at some time greater than or equal to $t$ given each individual's covariates; then, a product-limit estimator is calculated using the weighted data. When covariates have no effect on censoring times, our estimator reduces to the ordinary Kaplan-Meier estimator. An expression for its asymptotic variance formula is obtained using martingale techniques.

**Keywords**: Aalen's linear hazard model, informative censoring, nonparametric estimation, right censoring; survival analysis

## 1. SURVIVAL FUNCTION ESTIMATION WITH TIME DEPENDENT COVARIATES

For the $i$th individual, let $t_i^*$ denote the (possibly unobserved) failure time, let $C_i$ denote the censoring time, $t_i = t_i^* \wedge C_i$ and $\delta_i = I[t_i^* \leq C_i]$. We denote the true survival function by $S(t) = E\{I[t^* > t]\}$, with hazard function $\lambda(t)$ and cumulative hazard function $\Lambda(t)$. Assume for the $i$th person, data is available on time dependent covariates $Z_{ij}(t)$, $1 \leq j \leq J$ that may affect both failure and survival times. Let $\overline{\boldsymbol{Z}}_i(t)$ denote the information on all values of $Z_{ij}(t)$ between 0 and $t$, *i.e.* $\overline{\boldsymbol{Z}}_i(t) =$

$\{Z_{ij}(s), 0 \leq s < t, j \leq J\}$. Rather than estimating the distribution of $t^*$ conditional on $\overline{\boldsymbol{Z}}_i(t)$ and then attempting to average over values of $\overline{\boldsymbol{Z}}_i(t)$, Robins (1993) introduced a class of direct estimators of $S(t)$. Estimators in this class assume a model in which the $i$th person's hazard of being censored at time $t$ does not depend on $t_i^*$ given $\overline{\boldsymbol{Z}}_i(t)$ and $t_i \geq t$, *i.e.*

$$\lambda_c[t|t_i^*, \overline{\boldsymbol{Z}}_i(t)] = \lambda_c[t|\overline{\boldsymbol{Z}}_i(t)] \tag{1}$$

where

$$\lambda_c[t|\cdot\,] = \lim_{dt \to 0} \frac{\Pr[C_i \in [t, t+dt), \delta_i = 0 | t_i \geq t, \cdot\,]}{dt} \quad ,$$

in continuous time and

$$\lambda_c[t|\cdot\,] = Pr[C_i = t, C_i \leq t_i^* | t_i \geq t, \cdot\,],$$

in discrete time. We assume data from individuals are independent and identically distributed.

To estimate the marginal distribution of failure times Robins (1993) proposed a class of estimators using a data-reweighting scheme. Let

$$K_i(t) = \prod_{s \leq t} [1 - d\Lambda_c[s|\overline{\boldsymbol{Z}}_i(s)]]$$

where $\Lambda_c[t|\overline{\boldsymbol{Z}}_i(t)] = \int_0^t \lambda_c[u|\overline{\boldsymbol{Z}}_i(u)]du$ ( or $\sum_{u \leq t} \lambda_c[u|\overline{\boldsymbol{Z}}_i(u)]$ in discrete time). Note that under our assumed censoring mechanism it is not necessarily the case that $K_i(t)$ is the probability that the $i$th person survived to time $t$ without being censored, and that $K_i(t)$ has this simple interpretation if $\lambda_c[t|t_i^*, \overline{\boldsymbol{Z}}_i(t)] = \lambda_c[t|t_i^*, \overline{\boldsymbol{Z}}_i(t_i^*), \overline{\boldsymbol{Z}}_i(t)]$ (Robins and Rotnitzky, 1992). This occurs, for example, if all $Z_{ij}(t)$ are time-independent.

To develop an estimator for the survival function which is valid under censoring model (1), define

$$\overline{N}(t) = \sum_i \frac{I[t_i \le t, \delta_i = 1]}{K_i(t_i -)} \tag{2}$$

$$\widehat{N}(t) = \sum_i \frac{I[t_i \le t, \delta_i = 1]}{\widehat{K}_i(t_i -)} \;; \tag{3}$$

define the re-weighted risk set $\overline{Y}(t)$ by

$$\overline{Y}(t) = \sum_i \frac{I(t_i \ge t)}{K_i(t -)} \tag{4}$$

and let

$$\widehat{Y}(t) = \sum_i \frac{I(t_i \ge t)}{\widehat{K}_i(t -)}, \tag{5}$$

where $\widehat{K}_i(t)$ is a consistent estimator of $K_i(t)$. We define also the full-data counting processes $N^*(t) = \sum_i I(t_i^* \le t)$ and $Y^*(t) = \sum_i I(t_i^* \ge t)$. In the Appendix, we prove that

$$E[\overline{N}(t)] = E[N^*(t)] \tag{6}$$

and

$$E[\overline{Y}(t)] = E[Y^*(t)]. \tag{7}$$

Thus if $K_i(t)$ were known, the survival function $S(t)$ could be estimated using

$$\overline{S}(t) = \prod_{\tau_k \le t} \left[1 - \frac{\Delta \overline{N}(\tau_k)}{\overline{Y}(\tau_k)}\right]. \tag{8}$$

where $\tau_1 < \tau_2 < \cdots < \tau_m$ are the distinct observed failure times and for a right-continuous non-decreasing process $\{W(t)\}$, $\Delta W(t) = W(t) - W(t -)$ denotes its jump at time $t$. In the Appendix we show that $\widehat{S}(t)$ given by

$$\widehat{S}(t) = \prod_{\tau_k \le t} \left[1 - \frac{\Delta \widehat{N}(\tau_k)}{\widehat{Y}(\tau_k)}\right] \tag{9}$$

is also a consistent estimator of the survival function for an appropriate choice of $\widehat{K}_i$. Equations (6), (7) consistency of $\widehat{S}(t)$ were established by Robins (1993) but the proofs we give are simpler and more direct.

## 2. Our Proposed Estimator

Robins (1993) estimated $K_i$ using the proportional hazards model. We propose here use Aalen's additive hazard model for $\widehat{K}_i$ (Aalen 1980, 1989). Although Aalen's model is more flexible than the proportional hazards model, it has several difficulties that have limited its use in practical data analyses. First, the hazard estimates may be negative. Second, the estimator involves inverting a matrix that may have full rank; in this case the regression estimates are not defined for any covariables and estimates based on using a generalized inverse would depend on the particular choice of generalized inverse. We show here that these problems do not arise when the goal is estimation of $K_i$, and that an estimator of $S(t)$ using Aalen's linear hazard model will always be well defined and have desirable small sample properties.

Aalen's linear hazard model assumes that

$$\lambda_c[t \mid \overline{\boldsymbol{Z}}_i(t)] = \sum_{j=0}^{J} \beta_j(t) Z_{ij}(t) \tag{10}$$

where $\beta_j(t)$ is an unknown function, and $Z_{i0}(t) \equiv 1$. In writing (10) we assume the process $Z_{ij}(t)$ is predictable in time; i.e., the value of $Z_{ij}(t)$ is available at time $t-$. Defining $B_j(t) = \int_0^t \beta(s)ds$, Aalen's model estimates $\boldsymbol{B}(t) = (B_0(t), \cdots, B_J(t))$ by

$$\widehat{\boldsymbol{B}}(t) = \sum_{i=1}^{n} I(t_i \le t)(1 - \delta_i)\, \boldsymbol{A}^{-1}(t_i)\, \boldsymbol{Z}_i(t_i) \tag{11}$$

where $\boldsymbol{Z}_i(t)$ is the vector with components $Z_{i0}(t), Z_{i1}(t), \cdots, Z_{iJ}(t)$, and where

$$\boldsymbol{A}(t) = \sum_{i=1}^{n} I(t_i \ge t) \boldsymbol{Z}_i(t) \boldsymbol{Z}_i^{\mathrm{T}}(t) \ . \tag{12}$$

Because Aalen's model fits a function $\beta(t)$ for each covariate (including time-independent covariates) it provides very flexible estimates of $\Lambda_c[t \,|\, \overline{\boldsymbol{Z}}_i(t)]$. Specifically, we require $\widehat{\Lambda}_c[t \,|\, \overline{\boldsymbol{Z}}_i(t)]$ for values of $t \le t_i$. These can be estimated by writing

$$\widehat{\Lambda}_c[t \,|\, \overline{\boldsymbol{Z}}_i(t)] = \sum_{j=0}^{J} \int_0^t Z_{ij}(t) d\widehat{B}_j(t)$$

$$= \sum_{i'=1}^{n} I(t_{i'} \le t)(1 - \delta_{i'})\, \boldsymbol{Z}_i(t_{i'})\, \boldsymbol{R}^{-1}(t_{i'})\, \boldsymbol{Z}_{i'}(t_{i'}), \;\; t \le t_i \,. \tag{13}$$

A potential problem that occurs in Aalen's linear model when interpreting coefficients $\widehat{\beta}_j(t)$ or $\widehat{B}_j(t)$ calculated using (11) when $\boldsymbol{A}(t)$ does not have full rank is avoided when the goal is estimation of $\Lambda_c[t \,|\, \overline{\boldsymbol{Z}}_i(t)]$. If a generalized inverse matrix is used when $\boldsymbol{A}(t)$ does not have full rank, $\Lambda_c[t \,|\, \overline{\boldsymbol{Z}}_i(t)]$ is well defined, as it is easy to show using (12) that $\boldsymbol{Z}_i(t_j)$ for combinations of $i$ and $j$ such that $t_i \ge t_j$ and $\delta_j = 0$ is always contained in the column space of $\boldsymbol{A}(t_j)$. In our calculations, we used the spectral inverse $P \cdot Diag(E^{\ddagger}) \cdot P^T$ where $P$ is the matrix whose columns are the eigenvectors of $\boldsymbol{A}(t)$, $P^T$ is the transpose of $P$, $E$ is the vector of eigenvalues of $\boldsymbol{A}(t)$ and $E^{\ddagger}$ is the vector with entries $1/E_j$ if $E_j \ne 0$ and 0 otherwise. The computational convenience and flexibility of the Aalen model suggests that the resulting estimator of $S(t)$ should perform better than one using the proportional hazards model in practical applications as in Robins and Rotnitzky (1992) or Robins (1993).

Our proposed estimator $\widehat{S}(t)$ has desirable small-sample properties. Aalen's model may produce negative estimates of $\lambda_c[t \mid \overline{\boldsymbol{Z}}_i(t)]$, resulting in $\widehat{\Lambda}_c[t \mid \overline{\boldsymbol{Z}}_i(t)]$ not being monotone increasing. However, $\widehat{Y}(\tau_k) \geq \Delta\widehat{N}(\tau_k)$ even if $\widehat{\Lambda}_c[t \mid \overline{\boldsymbol{Z}}_i(t)]$ is not a monotone increasing function of $t$, so that $\widehat{S}(t)$ has the following properties: $\widehat{S}(t)$ decreases monotonically with jumps at the observed failure times; $\widehat{S}(0) = 1$; and $\widehat{S}(\tau_m+) = 0$ if the largest $t_i$ is from an uncensored observation and $0 < \widehat{S}(t_{(n)}+) \leq 1$ if the largest $t_i$ is from a censored observation. Finally, if $\widehat{K}_i(t) = \widehat{K}(t)$ (*i.e.*, if there are no differences in censoring hazard between individuals), the data-reweighted Kaplan-Meier (9) reduces to the usual Kaplan-Meier survival function estimator. Robins and Rotnitzky (1992) had originally proposed a consistent estimator of the survival function $S(t)$ under assumption (1). However their estimator, in contrast to $\widehat{S}(t)$ of this paper, was not a monotone function of $t$. Reweighting in survival analysis was first considered by Koul et al. (1981) in the context of regression analysis.

## 3. VARIANCE CALCULATION

In this section, we report the asymptotic variance of $\sqrt{n}[\widehat{S}(t) - S(t)]$ when Aalen's linear model is used to model $K_i(t)$. The details are suppressed and can be found in Datta and Satten (2001). Remarkably, even if $K_i(t)$ were know, it is still advisable to use $\widehat{S}(t)$ in place of $\overline{S}(t)$, as we will show $\widehat{S}(t)$ has a smaller asymptotic variance than $\overline{S}(t)$. Similar results are found in Koul et al. (1981), Robins (1993) Robins and Rotnitzky (1995), Tsai and Crowley (1998) and Zhou (1999) for other estimation problems.

Let $\widehat{\Lambda}(t) = \int_0^t d\widehat{N}(u)/\widehat{Y}(u)$. Because the asymptotic variance of $\widehat{S}(t)$ is equal to $S^2(t)$ times the asymptotic variance of $\widehat{\Lambda}(t)$, it is sufficient to consider the asymptotic variance of $\widehat{\Lambda}(t)$. For simplicity we consider only the case where $t_i^*$ and $C_i$ are continuous. Let $M^*(t)$ denote the zero mean martingale associated with the uncensored data, *i.e.* $M^*(t) = N^*(t) - \int_0^t Y^*(u)d\Lambda(u)$. Define the counting process of censoring times $N_i^c(t) = I[t_i \le t, C_i \le t_i^*]$. Let $M_i^c(t) = N_i^c(t) - \int_0^t I[t_i \ge u]d\Lambda_c[u|\overline{\boldsymbol{Z}}_i(u)]$ be the zero mean martingale associate with $N_i^c(t)$.

It can be shown that up to negligible terms,

$$\widehat{\Lambda}(t) - \Lambda(t) = \int_0^t \frac{I[Y^*(s)>0]dM^*(s)}{Y^*(s)} + n^{-1}\int_0^t \zeta^{\mathrm{T}}(s,t)\Big[\mathbb{I} - \widetilde{\mathbb{Z}}\,(s)\boldsymbol{A}^{-1}(s)\widetilde{\mathbb{Z}}^{\,\mathrm{T}}(s)\Big]d\mathbb{M}^c(s) \tag{14}$$

where $\mathbb{M}^c(s)$ is the vector with $i$th component given by $M_i^c(t)$, $\widetilde{\mathbb{Z}}\,(s)$ is the matrix with $i$th row $I[t_i \ge s]\boldsymbol{Z}_i(s)$ and

$$\zeta_i(s,t) = \left[\frac{1}{S(s)} - \frac{1}{S(t \wedge t_i^*)} - \frac{I[t_i^* \le t]}{S(t_i^*)}\right]\frac{1}{K_i(s)}.$$

Note that the first term in (14) is based on the true failure times and hence is measurable with respect to $\mathcal{F}_0$, where $\mathcal{F}_t$ is the filtration generated by the true failure times, the covariate information up to time $t$ and knowledge of $I(t_i \le u, \delta_i = 0)$ up to time $t$. Because the second term in (14) is a martingale with respect to $\mathcal{F}_t$, it follows by conditioning that the two terms in (14) are orthogonal. Hence, by standard martingale methods (see e.g., Andersen *et al.*), the asymptotic variance of $\widehat{\Lambda}(t) - \Lambda(t)$ is given by the (in probability limit) sum of the predictable variation process of the martingales in (14)

$$A\text{-}Var\{\sqrt{n}\,[\widehat{\Lambda}(t) - \Lambda(t)]\} = \operatorname*{plim}_{n\to\infty}\left\{n\int_0^t \frac{J^*(s)\lambda(s)}{Y^*(s)}ds + n^{-1}\int_0^t \zeta^{\mathrm{T}}(s,t)\Big[\mathbb{I}\right.$$

$$- \widetilde{\mathbb{Z}}(s)\boldsymbol{A}^{-1}(s)\widetilde{\mathbb{Z}}^{\mathrm{T}}(s)\Big]\mathrm{Diag}\Big([t_i \geq s]\lambda_i^c(s)\Big)\Big[\mathbb{I} - \widetilde{\mathbb{Z}}(s)\boldsymbol{A}^{-1}(s)\widetilde{\mathbb{Z}}^{\mathrm{T}}(s)\Big]\zeta(s,t)\,ds\Big\},$$

$$= \operatorname*{plim}_{n\to\infty} \Big\{ n\int_0^t \frac{J^*(s)\lambda(s)}{Y^*(s)}ds + n^{-1}\int_0^t \eta^{\mathrm{T}}(s,t)\eta(s,t)\,ds$$

$$- n^{-1}\int_0^t \eta^{\mathrm{T}}(s,t)\widetilde{\mathbb{Z}}(s)\boldsymbol{A}^{-1}(s)\widetilde{\mathbb{Z}}^{\mathrm{T}}(s)\eta(s,t)\,ds\Big\}, \qquad (15)$$

where $\eta^{\mathrm{T}}(s,t) = ([t_1 \geq s](\lambda_1^c(u))^{1/2}\zeta_1(s,t), \cdots, [t_n \geq s](\lambda_n^c(u))^{1/2}\zeta_n(s,t))$,

$$= V_1 + V_2 - V_{3,J}, \text{ say.} \qquad (16)$$

We may identify the first term of (16) as the asymptotic variance we would obtain if there were no censoring. The term

$$V_2 = \operatorname*{plim}_{n\to\infty} \Big\{ n^{-1}\int_0^t \eta^{\mathrm{T}}(s,t)\,\eta(s,t)\,ds \Big\}$$

is the added variance in the Nelson-Aalen estimator due to right-censoring if the true $K_i$ were used. Finally, the term $V_{3,J}$ is the *decrease* in variance that is due to estimating the coefficients in Aalen's linear hazard model for $\Lambda_i^c$. Note that if we denote by $\mathbb{Z}_J(s)$ and $\boldsymbol{A}_J(s)$ the values of $\mathbb{Z}(s)$ and $\boldsymbol{A}(s)$ obtained using $J$ regressors in our model for $d\Lambda_i^c(s)$, then $\mathbb{Z}_{J+1}(s)\boldsymbol{A}_{J+1}^{-1}(s)\mathbb{Z}_{J+1}^{\mathrm{T}}(s) - \mathbb{Z}_J(s)\boldsymbol{A}_J^{-1}(s)\mathbb{Z}_J^{\mathrm{T}}(s)$ is a non-negative definite matrix, indicating that increasing the number of regressors in the model for the censoring process *lowers* the asymptotic variance of $\widehat{\Lambda}(t) - \Lambda(t)$.

**References**

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes, In *Lecture Notes on Mathematical Statistics and Probability*, 2, W. Klonecki, A. Kozek and J. Rosiski, editors, pp 1-25, New York, Springer-Verlag.

Aalen, O. O. (1989). A linear regression model for the analysis of lifetimes, *Statistics in Medicine,* 8, 907-925.

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based On Counting Processes,* New York, Springer-Verlag.

Datta S, Satten GA (2001). Variance calculation of an estimator of marginal survival. University of Georgia Technical Report 2001:XY.

Koul, H., V. Susarla, and J. Van Ryzin (1981). Regression analysis with randomly right censored data. *Ann. Statist.* 9, 1276-1288.

Robins, J. M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers, In *Proceedings of the American Statistical Association - Biopharmaceutical Section*, 24-33.

Robins, J. M. and A. Rotnitzky (1992). Recovery of information and adjustment for dependent censoring using surrogate markers, In *AIDS Epidemiology – Methodological Issues*, N. Jewell, K. Dietz and V. Farewell, editors, pp. 297-331, Boston, Birkhauser.

Robins, J. M. and A. Rotnitzky (1995). Semiparametric regression estimation in the presence of dependent censoring, *Biometrika,* 82, 805-820.

Tsai, W.-Y. and J. A. Crowley (1998). A note on nonparametric estimators of the bivariate survival function under univariate censoring, *Biometrika*, 85, 573-580.

Zhou, M. J. (1999). Regression analysis with censored data: extensions of the Koul-Susarla-Van Ryzin approach, *Statistics and Probability Letters*, 41, 229-236.

**Appendix**

We first prove two lemmas which will be used in the proof of consistency of our estimator. For $t \geq 0$, let $\mathcal{F}_t = \sigma\big(\{t_i^*, Z_{ij}(u), I(t_i \leq u, \delta_i = 0), 0 \leq u \leq t, j \leq J, i = 1, \cdots, n\}\big)$ be the sigma algebra generated by the failure time, the covariate information up to time $t$ and knowledge of $I(t_i \leq u, \delta_i = 0)$ up to time $t$.

Recall $S(t) = Pr[t_1^* > t]$ and let $\Lambda(t)$ be the cumulative hazard function corresponding to $S(t)$; also $\lambda_c[t|\overline{\boldsymbol{Z}}_i(t)]$ is the hazard of being censored, with cumulative hazard $\Lambda_c[t|\overline{\boldsymbol{Z}}_i(t)]$. Also recall the following definitions from Section 4:

$$\widehat{\Lambda}(t) = \int_0^t \frac{d\widehat{N}(s)}{\widehat{Y}(s)}, \tag{A.1}$$

$$M^*(t) = N^*(t) - \int_0^t Y^*(u)\, d\Lambda(u), \tag{A.2}$$

and

$$M_i^c(t) = N_i^c(t) - \int_0^t I[t_i \geq u] d\Lambda_c[u|\overline{\boldsymbol{Z}}_i(u)]. \tag{A.3}$$

Also let

$$\widehat{M}_i^c(t) = N_i^c(t) - \int_0^t I[t_i \geq u] d\widehat{\Lambda}_c[u|\overline{\boldsymbol{Z}}_i(u)]. \tag{A.4}$$

To simplify our notation let $\Lambda_i^c(u)$ denote $\Lambda_i^c[u|\overline{\boldsymbol{Z}}_i(u)]$, $0 \leq u \leq t_i$; $1 \leq i \leq n$.

***Proofs of Equations (6) and (7):***

Note that $I[C_i \geq t_i^*] = \prod_{s<t_i^*} [1 + dX_i(s)]$ where $X_i(s) = -I[C_i \leq s]$. We also have on the set $\{C_i \geq t_i^*\}$, $K_i(t_i^* - ) = \prod_{s<t_i^*} [1 - I[t_i \geq s] d\Lambda_i^c(s)]$. The Duhamel equation (see, e.g., Andersen et al., 1993) then gives

$$\frac{I[C_i \geq t_i^*]}{K_i(t_i^* - )} = 1 + \int_0^{t_i^*-} \frac{I[C_i \geq s]}{K_i(s)} \Big\{ dX_i(s) - I[t_i \geq s] d\Lambda_i^c(s) \Big\}.$$

Note for $s < t_i^*$ we may write $X_i(s) = -I[C_i \leq s, C_i \leq t_i^*] = -N_i^c(s)$. Hence,

$$\frac{I[C_i \geq t_i^*]}{K_i(t_i^* - )} = 1 - \int_0^{\infty} \frac{I[t_i^* > s, C_i \geq s]}{K_i(s)} dM_i^c(s) = 1 - \int_0^{\infty} \frac{I[t_i^* > s]}{K_i(s)} dM_i^c(s) \quad \text{(A.5)}$$

since the martingale $M_i^c(s)$ vanishes for $\{C_i < s\}$. Equation (6) follows on multiplying both sides of (A.5) by $I[t_i^* \leq t]$ and noting that $I[t \geq t_i^* > s]/K_i(s)$ is $\mathcal{F}_s$-predictable, so that the second term in (A.5) is a zero mean martingale; equation (6) follows on taking expectations.

A similar argument yields

$$\frac{I[C_i \geq t]}{K_i(t - )} = 1 + \int_0^{t-} \frac{I[C_i \geq s]}{K_i(s)} \Big\{ dX_i(s) - I[t_i \geq s] d\Lambda_i^c(s) \Big\} = 1 - \int_0^{t-} \frac{dM_i^c(s)}{K_i(s)}. \quad \text{(A.6)}$$

Equation (7) then follows upon multiplying by $I[t_i^* \geq t]$, noting that $I[t_i^* \geq t]/K_i(s)$ is $\mathcal{F}_s$-predictable, and then taking expectations.

***Consistency of $\widehat{S}(t)$:***

Assume $t$ is such that $y(t) \equiv Pr(t_i^* \geq t) > 0$ and $E[K_i^{-2}(t)] < \infty$. For simplicity of presentation we assume all variables have (absolutely) continuous distributions. Using the uniform (on compact set of $t$) Ceasaro consistency of $\widehat{K}_i(t)$ under Aalen's linear model, it is straightforward to show that

$$\frac{1}{n}\widehat{Y}(t) = \frac{1}{n}\overline{Y}(t) + o_p(1) \xrightarrow{p} y(t), \tag{A.7}$$

where the last equality follows from the laws of large number for i.i.d. random variables and equation (7); also it is straightforward to note by the laws of large numbers that

$$\frac{1}{n}Y^*(t) \xrightarrow{p} y(t); \tag{A.8}$$

By similar arguments

$$\frac{1}{n}N^*(t) \xrightarrow{p} 1 - S(t); \; \frac{1}{n}\widehat{N}(t) \xrightarrow{p} 1 - S(t); \tag{A.9}$$

and

$$\int_0^t \frac{1}{y(u)} d\{n^{-1}\widehat{N}(u)\} - \int_0^t \frac{1}{y(u)} d\{n^{-1}N^*(u)\} = o_p(1). \tag{A.10}$$

Let $J^*(t) = I[Y^*(t) > 0]$, let $\widetilde{\Lambda}(t) = \int_0^t J^*(s)d\Lambda(s)$ and note $\widetilde{\Lambda}(t) = \Lambda(t) + o_p(1)$ (since $y(t) > 0$). Note

$$\widehat{\Lambda}(t) - \widetilde{\Lambda}(t) = \int_0^t J^*(s)\left\{\frac{d\widehat{N}(s)}{\widehat{Y}(s)} - d\Lambda(s)\right\}$$

$$= \int_0^t \frac{J^*(s)}{Y^*(s)} dM^*(s) + \int_0^t J^*(s)\left\{\frac{d\widehat{N}(s)}{\widehat{Y}(s)} - \frac{dN^*(s)}{Y^*(s)}\right\} \tag{A.11}$$

The first term of (A.11) is $o_p(1)$ by standard martingale methods. The second term in (A.11) is bounded by

$$\int_0^t \left| \frac{J^*(u)}{n^{-1}\widehat{Y}(u)} - \frac{1}{y(u)} \right| \frac{d\widehat{N}(u)}{n} + \int_0^t \left| \frac{J^*(u)}{n^{-1}Y^*(u)} - \frac{1}{y(u)} \right| \frac{dN^*(u)}{n}$$

$$+ \left| \int_0^t \frac{1}{y(u)} \left\{ \frac{d\widehat{N}(u)}{n} - \frac{d\overline{N}(u)}{n} \right\} \right|. \tag{A.12}$$

The first two terms in (A.12) are $o_p(1)$ using (A.7)-(A.8) and because it can be shown using monotonicity that the convergence in (A.7)-(A.8) is in fact uniform on compact intervals $[0, t]$. The second term in (A.12) converges to zero in probability by (A.9). Hence, $\widehat{\Lambda}(t)$ is a consistent estimator of $\Lambda(t)$.

Consistency of $\widehat{S}(t)$ for $S(t)$ now follows from the continuity of product integrals (see proposition II.8.7 in Andersen *et al.*, 1993 and the remarks following). □