journal of statistical planning and inference

Journal of Statistical Planning and Inference 66 (1998) 253-279

On semiparametric random censorship models

Gerhard Dikta

Fachhochschule Aachen, Abteilung Jülich, Ginsterweg 1, D-52428 Jülich, Germany

Received 28 May 1996; received in revised form 29 April 1997

Abstract

We propose and study a semiparametric estimator of the distribution function in the random censorship model which generalizes the Cheng and Lin estimator in the proportional hazards model. Uniform consistency and a functional central limit result for this estimator are established. Some efficiency comparisons with the Kaplan–Meier estimator are also included.

AMS classification: primary 60F15, 60F17; secondary 62F10, 62G30

Keywords: Censored data; Binary data; Maximum likelihood estimation; Functional central limit theorem; Uniform consistency; Kaplan–Meier estimator; Proportional hazards model

1. Introduction

Assume that $X_1,\dots,X_n$ are independent and identically distributed (i.i.d.) positive random variables, defined on some probability space $(\Omega,\mathscr{A},\mathscr{P})$, with unknown continuous distribution function (d.f.) F. In the random censorship model these data are censored on the right in which case one observes

$$Z_i=\min(X_i,Y_i) \quad \text{and} \quad \delta_i=1_{[X_i\leqslant Y_i]}, \qquad 1\leqslant i\leqslant n,$$

where $Y_1,\dots,Y_n$ is another sequence of positive i.i.d. random variables with continuous d.f. G, being also independent of the X's. The variable δ_i indicates whether X_i is censored ($\delta_i=0$) or not ($\delta_i=1$).

As a fully nonparametric estimator of F the time-honoured Kaplan–Meier (1958) product limit estimator defined by

$$1-F_n^{\mathrm{km}}(t)=\prod_{i:Z_i\leqslant t}\left(\frac{n-R_i}{n-R_i+1}\right)^{\delta_i}$$

received much attention in the literature. Here R_i denotes the rank of Z_i within the Z-sample. Following Shorack and Wellner (1986, p. 295), one may motivate F_n^{km} by

0378-3758/98/$19.00
PII S0378-3758(97)00091-8

first looking at the cumulative hazard function Λ corresponding to F :

$$\Lambda(t)=\int_0^t \frac{1}{1-F(x)}F(\mathrm{d}x)=\int_0^t \frac{1}{1-H(x)}H^1(\mathrm{d}x). \tag{1}$$

Here H denotes the d.f. of Z and $H^1(x)=\mathscr{P}(\delta=1,\ Z\leqslant x)$. The Nelson–Aalen estimator of Λ is given by

$$\Lambda_n(t)=\int_0^t \frac{1}{1-H_n(x-)}\bar{H}_n(\mathrm{d}x)=\sum_{i:Z_i\leqslant t}\frac{\delta_i}{n-R_i+1},$$

where

$$H_n(x)=\frac{1}{n}\sum_{i=1}^n 1_{[Z_i\leqslant x]}$$

is the empirical d.f. of the Z-sample, $H_n(x-)=\lim_{z\uparrow x}H_n(z)$, and

$$\bar{H}_n(x)=\frac{1}{n}\sum_{i=1}^n 1_{[Z_i\leqslant x]}\delta_i$$

is the empirical version of $H^1(x)$. Note that $-\ln(1-F(t))=\Lambda(t)$. By using the approximation $\exp(-x)\approx 1-x$ for $x\approx 0$, we get

$$\exp(-\Lambda_n(t))=\prod_{i:Z_i\leqslant t}\exp\left(-\frac{1}{n-R_i+1}\right)^{\delta_i}\approx 1-F_n^{\mathrm{km}}(t),$$

the Kaplan–Meier estimator.

To motivate our semiparametric estimator, we observe that $H^1(x)=\mathscr{P}(\delta=1,\ Z\leqslant x)=\int_0^x m(z)H(\mathrm{d}z)$, where

$$m(x)=\mathscr{P}(\delta=1\mid Z=x)=\mathbb{E}(\delta\mid Z=x)$$

denotes the conditional expectation of δ given $Z=x$. The importance of $m(x)$ has been pointed out in Stute and Wang (1993) for proving consistency of Kaplan–Meier integrals $\int\varphi\,\mathrm{d}F_n^{\mathrm{km}}$. Clearly,

$$\Lambda(t)=\int_0^t \frac{1}{1-H(x)}H^1(\mathrm{d}x)=\int_0^t \frac{m(x)}{1-H(x)}H(\mathrm{d}x). \tag{2}$$

If δ is independent of Z we have $m(x)=\mathbb{E}(\delta)$ in which case we come up with the simple Proportional Hazards Model (PHM), see Koziol–Green (1976). A suitable estimate of $m(x)$ is obviously $c_n=1/n\sum_{i=1}^n\delta_i$. Plugging this into (2) we arrive at

$$1-F_n^{cl}(t)=\prod_{i:Z_i\leqslant t}\left(\frac{n-R_i}{n-R_i+1}\right)^{c_n}=(1-H_n(t))^{c_n}.$$

This estimator was proposed independently by Abdushukurov (1987) and Cheng and Lin (1987) who pointed out that under PHM F_n^{cl} is more efficient than F_n^{km} in terms of asymptotic variance. See Csörgő (1988b) for a review of fundamental properties of F_n^{cl}. He also developed a test for the PHM and applied it to seven data sets. See

Csörgő (1988a). In six cases the assumption of PHM was rejected, which indicates that PHM may not be appropriate in many practical situations.

Obviously, (2) can be used to define a great variety of semiparametric models by describing m in some parametric form. An appropriate estimator of F can then be obtained by estimating m parametrically. This approach generalizes the PHM and seems more flexible for practical purposes than PHM. Furthermore, the corresponding estimator of F, as in the case of F_n^{cl} and PHM, might also be more efficient than the Kaplan–Meier estimator.

In this article we assume that m belongs to a parametric family so that we can write

$$m(x) = m(x, \theta_0),$$

where $m(\cdot,\cdot)$ is a known continuous function and $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,k}) \in \Theta$ is an unknown parameter. If we interpret the indicators $\delta_1, \ldots, \delta_n$ as binary response variables to $Z_1, \ldots, Z_n$ possible candidates for m can be found in Cox and Snell (1989). Parametric forms for m can be motivated upon observing that

$$m(x) = \frac{\lambda_f(x)}{\lambda_h(x)} = \frac{\lambda_f(x)}{\lambda_g(x) + \lambda_f(x)} = \frac{\lambda_f(x)/\lambda_g(x)}{1 + \lambda_f(x)/\lambda_g(x)}, \tag{3}$$

where f, g, h denote the density functions of X, Y, Z and $\lambda_f, \lambda_g, \lambda_h$ the corresponding hazard functions, respectively. Since

$$\lambda_h(x) = \frac{h(x)}{1 - H(x)} = \lambda_g(x) + \lambda_f(x),$$

representation (3) is a consequence of

$$\int_0^t \frac{f(x)}{1 - F(x)} \mathrm{d}x = \Lambda(t) = \int_0^t \frac{m(x)h(x)}{1 - H(x)} \mathrm{d}x.$$

For the estimation of the parameter θ_0 we use a maximum likelihood approach. The function m may then be estimated parametrically through $m_n(x) = m(x, \hat{\theta}_n)$, where $\hat{\theta}_n$ is the MLE of θ_0. Λ_n and $\hat{F}_n$, respectively, are then given by

$$\Lambda_n(t) = \int_0^t \frac{m(x, \hat{\theta}_n)}{1 - H_n(x-)} H_n(\mathrm{d}x) = \sum_{i: Z_i \leqslant t} \frac{m(Z_i, \hat{\theta}_n)}{n - R_i + 1}$$

$$1 - \hat{F}_n(t) = \prod_{i: Z_i \leqslant t} \left(\frac{n - R_i}{n - R_i + 1} \right)^{m(Z_i, \hat{\theta}_n)}.$$

In the next section we state the main results of the paper and discuss some examples. The proofs of the results are given in the third section while in the Appendix we calculate the covariance structure of the limiting process of $n^{1/2}(\hat{F}_n - F)$. In particular, we show that our estimator is superior to the Kaplan–Meier estimator in terms of asymptotic variance under the stated model assumptions.

2. Main results and examples

We begin this section with some general results on the consistency and asymptotic normality of the MLE for the parameter θ_0. The approach for the investigation of consistency is similar to Stute (1992) who established the consistency of the MLE under random censorship for parametric densities.

Having observed $(\delta_1, Z_1), \dots, (\delta_n, Z_n)$, the likelihood function is given by

$$L_n(\theta) = \prod_{i=1}^{n} m(Z_i, \theta)^{\delta_i} [1 - m(Z_i, \theta)]^{1-\delta_i}.$$

Hence the corresponding normalized log likelihood function equals

$$l_n(\theta) = n^{-1} \sum_{i=1}^{n} [\delta_i w_1(Z_i, \theta) + (1 - \delta_i) w_2(Z_i, \theta)],$$

where

$$w_1(x, \theta) = \ln(m(x, \theta)) \quad \text{and} \quad w_2(x, \theta) = \ln(1 - m(x, \theta)).$$

Now, define $\hat{\theta}_n \in \Theta$ to be the maximizer of $l_n(\theta)$ if it exists. The SLLN yields

$$\begin{aligned} l_n(\theta) &\to L_H(\theta_0, \theta) \\ &= \int_0^{\infty} [w_1(x, \theta) m(x, \theta_0) + w_2(x, \theta)(1 - m(x, \theta_0))] H(\mathrm{d}x) \end{aligned}$$

with probability one, as $n \to \infty$.

We now state conditions for consistency and asymptotic normality of $\hat{\theta}_n$. Basically, these conditions are adaptations of those which can be found in the usual MLE theory. See, e.g., Perlman (1972) and Witting and Müller–Funk (1995).

To simplify the notation we write $D_r m(x, \theta_0)$ for $[\partial m(x, \theta)/\partial \theta_r]|_{\theta=\theta_0}$ and $\mathrm{Grad}(m(x, \theta_0))$ for $(D_1 m(x, \theta_0), \dots, D_k m(x, \theta_0))$. Furthermore, $\langle \cdot \mid \cdot \rangle$ denotes the inner product in $\mathbb{R}^k$.

(A_1) For each $\theta \neq \theta_0$

$$\int 1_{[m(x,\theta)=0]} m(x, \theta_0) H(\mathrm{d}x) = 0 = \int 1_{[m(x,\theta)=1]} (1 - m(x, \theta_0)) H(\mathrm{d}x)$$

and

$$\int 1_B \, \mathrm{d}H > 0 \quad \text{where } B = \{m(\cdot, \theta) \neq m(\cdot, \theta_0)\}.$$

(A_2) There exists a measurable solution $\hat{\theta}_n \in \Theta$ of the equation $\mathrm{Grad}(l_n(\theta)) = 0$ which tends to θ_0 in probability.

(A_3) For $i = 1, 2$, $w_i(x, \theta)$ possesses continuous partial derivatives of second order with respect to θ at each $\theta \in \Theta$ and $x \geq 0$. Furthermore, $D_{r,s} w_i(\cdot, \theta)$ is measurable for each $\theta \in \Theta$ and there exists a neighborhood $V(\theta_0) \subset \Theta$ of θ_0 and a measurable function M such that for all $\theta \in V(\theta_0)$, $x \geq 0$, and $1 \leq r, s \leq k$

$$|D_{r,s} w_1(x, \theta)| + |D_{r,s} w_2(x, \theta)| \leq M(x) \quad \text{and} \quad \mathbb{E}(M(Z)) < \infty.$$

(A_4) For $1 \leqslant r \leqslant k$,

$$[(D_r m(Z,\theta_0))/m(Z,\theta_0)]^2 \quad \text{and} \quad [(D_r m(Z,\theta_0))/(1-m(Z,\theta_0))]^2$$

have finite expectation.

(A_5) The matrix $I(\theta_0)=(\sigma_{r,s})_{1\leqslant r,s\leqslant k}$, where

$$w(\delta,x,\theta)=\delta w_1(x,\theta)+(1-\delta)w_2(x,\theta)$$

and

$$\sigma_{r,s}=-\,\mathbb{E}(D_{r,s}w(\delta,Z,\theta_0))=\mathbb{E}\left(\frac{D_r(m(Z,\theta_0))D_s(m(Z,\theta_0))}{m(Z,\theta_0)(1-m(Z,\theta_0))}\right) \tag{4}$$

is positive definite.

(A_6) $m(x,\theta)$ possesses continuous partial derivatives of second order with respect to θ at each $\theta\in\Theta$ and $x\geqslant 0$. Furthermore, $D_{r,s}m(\cdot,\theta)$ is measurable for each $\theta\in\Theta$ and there exists a neighborhood $V(\theta_0)\subset\Theta$ of θ_0 and a measurable function M such that for all $\theta\in V(\theta_0)$, $x\geqslant 0$, and $1\leqslant r,s\leqslant k$

$$|D_{r,s}m(x,\theta)|\leqslant M(x) \quad \text{and} \quad \mathbb{E}(M(Z))<\infty,$$

and finally, $\mathrm{Grad}(m(\cdot,\theta_0))$ is bounded on $[0,T]$ with $H(T)<1$, i.e.

$$\sup_{0\leqslant x\leqslant T} \|\,\mathrm{Grad}(m(x,\theta_0))\,\| <\infty.$$

(A_7) For $1\leqslant r\leqslant k$, $D_r m(\cdot,\theta_0)$ is Lipschitz on $[0,T]$, i.e.

$$|D_r m(x,\theta_0)-D_r m(y,\theta_0)|\leqslant c|x-y|$$

for an appropriate constant c, with $H(T)<1$.

In our first theorem strong consistency of the MLE is given. It constitutes an adaptation of Theorem 2.4 in Perlman (1972) to the present situation.

Theorem 2.1. *Let $\Theta\subset\mathbb{R}^k$ be compact and $\theta_0\in\Theta$. Assume that $L_H(\theta_0,\theta_0)$ is finite and that $m(\cdot,\cdot)$ is continuous. If (A_1) is satisfied then, as $n\to\infty$, $\mathscr{P}_*(\hat\theta_n\to\theta_0)=1$, where $\mathscr{P}_*$ denotes the inner measure belonging to $\mathscr{P}$.*

Under some extra condition the assertion of Theorem 2.1 may be extended to a σ-compact parameter set Θ. For this, put

$$\bar w_i(x,\theta)=w_i(x,\theta)-w_i(x,\theta_0), \qquad i=1,2.$$

Corollary 2.2. *Let $\Theta=\cup\, C_r\subset\mathbb{R}^k$ be σ-compact. Assume that the assumptions of Theorem 2.1 are satisfied. Then $\mathscr{P}_*(\hat\theta_n\to\theta_0)=1$ if there exists a compact set $C=C_r$ such that*

$$\mathbb{E}\left(\sup_{\theta\in\Theta\setminus C}[\delta\bar w_1(Z,\theta)+(1-\delta)\bar w_2(Z,\theta)]\right)<0.$$

The next theorem states asymptotic normality of the MLE.

Theorem 2.3. *Let Θ be a connected, open subset of $\mathbb{R}^k$. If assumptions* $(A_2)-(A_5)$ *are satisfied, then $n^{1/2}(\hat\theta_n-\theta_0)$ is asymptotically normal, $\mathcal{N}_k(0,I^{-1}(\theta_0))$, where $I(\theta_0) = (\sigma_{r,s})_{1\leqslant r,\, s\leqslant k}$ is defined in* (4).

Our first large sample result on Λ_n and $\hat F_n$, their Glivenko–Cantelli convergence, is given in the next theorem.

Theorem 2.4. *Let $0<T<\infty$ with $H(T)<1$ and Θ be a connected, open subset of $\mathbb{R}^k$. Assume that $\hat\theta_n\in\Theta$ is a measurable solution of the equation* $\mathrm{Grad}(l_n(\theta))=0$ *such that $\hat\theta_n\to\theta_0$ with probability one, and that $m(x,\theta)$ possesses continuous partial derivatives with respect to θ at each $\theta\in\Theta$ and $x\geqslant 0$. Furthermore, $D_r(m(\cdot,\theta))$ is measurable for each $\theta\in\Theta$ and there exists a neighborhood $V(\theta_0)\subset\Theta$ of θ_0 and a measurable function M such that $|D_r(m(x,\theta))|\leqslant M(x)$ and $\mathbb{E}(M(Z))<\infty$ for all $\theta\in V(\theta_0)$, $x\geqslant 0$, and $1\leqslant r\leqslant k$. Then with probability one, as $n\to\infty$,*

$$\sup_{0\leqslant t\leqslant T} |\Lambda_n(t)-\Lambda(t)| \longrightarrow 0 \tag{5}$$

$$\sup_{0\leqslant t\leqslant T} |\hat F_n(t)-F(t)| \longrightarrow 0. \tag{6}$$

The next theorem presents the functional central limit result for the process $n^{1/2}(\Lambda_n(t) - \Lambda(t))$.

Theorem 2.5. *If H is continuous, $H(T)<1$, and (A_2)–(A_7) are satisfied, then $n^{1/2}(\Lambda_n(t)-\Lambda(t))$ tends weakly to a centered Gaussian process S, where the covariance structure of S is given for $0\leqslant s\leqslant t\leqslant T$ by*

$$\begin{aligned}\mathrm{Cov}(S(s),S(t)) = &\int_0^s \frac{m(x,\theta_0)}{(1-H(x))^2} H^1(\mathrm{d}y)\\ &+\int_0^s\int_0^t \frac{\alpha(x,y)}{(1-H(x))(1-H(y))} H(\mathrm{d}y)H(\mathrm{d}x),\end{aligned} \tag{7}$$

where

$$\alpha(x,y) = \langle \mathrm{Grad}(m(x,\theta_0)) | I^{-1}(\theta_0)(\mathrm{Grad}(m(y,\theta_0)))\rangle.$$

As a consequence of the preceding theorem we get the weak convergence of the process $n^{1/2}(\hat F_n - F)$.

Corollary 2.6. *Under the assumptions of Theorem* 2.5, *$n^{1/2}(\hat F_n(t)-F(t))$ tends weakly to the centered Gaussian process $(1-F)S$.*

The next corollary shows that for each $0\leqslant t\leqslant T$, the asymptotic variance of the Kaplan–Meier process exceeds that of $n^{1/2}(\hat F_n - F)$. For this, we denote the asymptotic variance of $n^{1/2}(F_n^{\mathrm{km}}(t)-F(t))$ by $v^{\mathrm{km}}(t)$ and that of $n^{1/2}(\hat{F}_n(t)-F(t))$ by $v(t)$, respectively.

Corollary 2.7. *If the assumptions of Theorem 2.5 are satisfied, then for* $0 \leqslant t \leqslant T$,

$$v^{\mathrm{km}}(t)-v(t)=(1-F(t))^2 r(t) \geqslant 0,$$

where

$$r(t)=\int_0^t \frac{1-m(x,\theta_0)}{(1-H(x))^2} H^1(\mathrm{d}x)-\int_0^t\int_0^t \frac{\alpha(x,y)}{(1-H(x))(1-H(y))} H(\mathrm{d}y)H(\mathrm{d}x).$$

The following example shows that $\hat{F}_n$ generalizes the Cheng and Lin (1987) estimator F_n^{cl} in the PHM.

Example 2.8. When $m(x,\theta)=\theta \in (0,1)$, we immediately obtain

$$m(x,\hat{\theta}_n)=\hat{\theta}_n=1/n\sum_{i=1}^n \delta_i$$

and therefore, as already pointed out in the introduction, $\hat{F}_n(t)=F_n^{cl}(t)$. Needless to say, that the covariance structure of our limiting process is identical to the covariance structure given by Cheng and Lin (1987).

In the next example we compare the asymptotic variance of the Kaplan–Meier estimator with the variance of our semiparametric estimator numerically.

Example 2.9. Consider the two parameter model

$$m(x,\theta)=\frac{\theta_1}{\theta_1+x^{\theta_2}}, \quad \theta_1>0, \quad \theta_2 \in \mathbb{R}$$

which arises from Weibull distributed variables X and Y. In particular, we get from (3)

$$\theta_1=\frac{\alpha_1^{\beta_1}\beta_1}{\alpha_2^{\beta_2}\beta_2}, \quad \theta_2=\beta_2-\beta_1,$$

where $F(x)=1-\exp(-(\alpha_1 x)^{\beta_1})$ and $G(x)=1-\exp(-(\alpha_2 x)^{\beta_2})$. The MLE of θ_0 is the solution of

$$\sum_{i=1}^n (m(Z_i,\theta)-\delta_i)=0$$

and

$$\sum_{i=1}^n \ln(Z_i)(m(Z_i,\theta)-\delta_i)=0.$$

In Tables 1 and 2 we list $v^{\mathrm{km}}(t)/v(t)$ for several degrees of censoring, i.e. $\mathscr{P}(\delta=1)$. In each case we chose $t=H^{-1}(0.10)$, $H^{-1}(0.25)$, $H^{-1}(0.50)$, $H^{-1}(0.75)$, and $H^{-1}(0.90)$

Table 1
$v^{km}(t)/v(t)$ with $\theta_{0,2}=1$, i.e. m is decreasing

Percentage of uncensored data	90%	80%	70%	60%	50%	25%
$t=H^{-1}(0.10)$	1.01	1.02	1.03	1.05	1.08	1.17
$t=H^{-1}(0.25)$	1.01	1.03	1.05	1.07	1.09	1.17
$t=H^{-1}(0.50)$	1.02	1.04	1.07	1.10	1.13	1.23
$t=H^{-1}(0.75)$	1.04	1.09	1.13	1.17	1.20	1.25
$t=H^{-1}(0.90)$	1.07	1.13	1.17	1.21	1.24	1.31

Table 2
$v^{km}(t)/v(t)$ with $\theta_{0,2}=-3$, i.e. m is increasing

Percentage of uncensored data	90%	80%	70%	60%	50%	25%
$t=H^{-1}(0.10)$	1.09	1.23	1.43	1.76	2.26	4.60
$t=H^{-1}(0.25)$	1.05	1.10	1.16	1.25	1.36	1.92
$t=H^{-1}(0.50)$	1.03	1.07	1.11	1.16	1.22	1.41
$t=H^{-1}(0.75)$	1.02	1.04	1.07	1.11	1.17	1.38
$t=H^{-1}(0.90)$	1.02	1.03	1.06	1.08	1.12	1.29

where $H^{-1}(q)$ is the q-quantile of H. For F and G we used

(a) $F(x)=1-\exp(-(4x))$, $G(x)=1-\exp(-(\alpha x)^2)$,

(b) $F(x)=1-\exp(-(0.2x)^6)$, $G(x)=1-\exp(-(\alpha x)^3)$,

with varying α in order to cope with different degrees of censoring.

As Tables 1 and 2 show, our semiparametric estimator is more efficient than the Kaplan–Meier estimator in terms of asymptotic variance under the given model assumptions. One observes a substantial gain in efficiency under heavy censoring especially when t is in a region where m is small.

In the next example we illustrate our semiparametric approach under a parametric model assumption.

Example 2.10. Consider a clinical study in which a therapy is investigated. Assume that the time individuals enter the study is uniformly distributed over the whole period of the study, say $[0,T]$. Since G is mainly determined by the time of entry into the study, $\lambda_g(x)=1/(T-x)$. Now, if we take for λ_f an appropriate parametric model, for example a bathtub-shaped or a polynomial hazard function, representation (3) leads to a corresponding parametric model for $m(x,\theta)$. However, in this setup there is a parametric way to estimate f, which is based on the modified likelihood function

$$\bar{L}_n(\theta)=\prod_{i=1}^{n} f^{\delta_i}(Z_i,\theta)[1-F(Z_i,\theta)]^{1-\delta_i}. \tag{8}$$

The preceding example might give the impression that there is no need for our method. But this impression is misleading, as the following example shows.

Example 2.11. If the hazard functions λ_g and λ_f are linked up by a multiplicative effect, i.e.

$$\lambda_f(x) = \Psi(x, \theta_0)\lambda_g(x), \tag{9}$$

where $\Psi(\cdot,\cdot)$ belongs to a parametric family, but in addition no parametric form for f can be given, then the ordinary MLE for censored data cannot be used. However, we can use representation (3) to get

$$m(x,\theta) = \frac{\Psi(x,\theta)}{1+\Psi(x,\theta)}$$

and apply our method to this parametric form of m. The case $\Psi(x,\theta)=\theta$ again leads to PHM.

Remark 2.12. If instead of (9) we assume that G is a parametric function of F, and in addition that no parametric form for f is assumed, then again the ordinary MLE for censored data cannot be applied.

Beirlant et al. (1992) introduced an approach, the Long Run Proportional Hazards Model (LRPHM), to generalize the PHM which can be used in this situation. They assumed that

$$1 - H = (1-G)(1-F) = \Psi(1-F, \theta),$$

where

$$\Psi(u,\theta) = \frac{\theta_2 u^{\theta_1}}{1+(\theta_2 - 1)u^{\theta_3}}, \quad u \in [0,1]$$

and $\theta \in R^3$, $\theta_1 = \lim_{x\to\infty}(1-H(x))/\mathscr{P}(\delta=1, Z>x)$. If we apply representation (3), a straightforward calculation shows that

$$m(x,\theta) = \bar{m}(1-H(x),\theta),$$

where $\bar{m}$ belongs to a parametric family.

In this situation we cannot apply our MLE since m depends on the unknown d.f. H. However, we can estimate H by H_n and take $m_n(x,\theta)=\bar{m}(1-H_n(x),\theta)$ instead of $m(x,\theta)$ for MLE to get an estimate $\bar{m}_n(x,\hat{\theta}_n)$ which finally leads to $\hat{F}_n(x)$. Nevertheless, our preceding theorems have to be adapted to this situation in order to get the correct asymptotic theory.

Remark 2.13. Finally, we want to remark that diagnostics to check model assumptions can be based on methods which are used in the analysis of binary data, since $\delta_1, \delta_2, \dots, \delta_n$ are n independent binary variables with conditional probability of success given by $m(Z_1), m(Z_2), \dots, m(Z_n)$. For these methods we refer to Cox and Snell (1989).

3. Proofs

To discriminate between two different parameters in our setup we need a proper version of the Kullback–Leibler information.

Definition 3.1. For two functions r, s defined on $[0,\infty)$ with values in $[0,1]$, which satisfy $\int 1_{[r=0]} s \, \mathrm{d}H = 0 = \int 1_{[r=1]}(1-s)\, \mathrm{d}H$, define

$$K_H(s,r) = \int_0^\infty \ln(s/r) s \, \mathrm{d}H + \int_0^\infty \ln[(1-s)/(1-r)](1-s)\, \mathrm{d}H,$$

where $0 \ln(0/0)$ and $0/0$ are taken to be zero.

With this definition we get

Lemma 3.2. *K_H is well defined with $0 \leqslant K_H(s,r) \leqslant \infty$ and*

$$K_H(s,r) > 0 \quad \Leftrightarrow \quad \int 1_{\{r \neq s\}} \, \mathrm{d}H > 0.$$

The proof is similar to the corresponding arguments in usual MLE theory and therefore omitted. See, e.g. Stute (1992).

As a consequence of the preceding lemma we get

Lemma 3.3. *Assume that $L_H(\theta_0, \theta_0)$ is finite. Then $L_H(\theta_0, \cdot)$ has a unique maximum at θ_0 if (A_1) is satisfied.*

Proof. Since $K_H(m(\cdot,\theta_0), m(\cdot,\theta)) = L_H(\theta_0,\theta_0) - L_H(\theta_0,\theta)$, the assertion follows from Lemma 3.2. □

Proof of Theorem 2.1. Since $0 \leqslant m(x,\theta) \leqslant 1$ we get

$$\sup_{\theta \in \Theta} w_i(x,\theta) \leqslant 0, \qquad i = 1,2,$$

and arguments similar to those used in the proof of Theorem 2.4 in Perlman (1972) or Theorem 3.1 in Stute (1992) can be applied to prove the present theorem. □

Proof of Theorem 2.3. As in the proof of Satz 6.35 in Witting and Müller–Funk (1995), Taylor's expansion yields

$$n^{1/2}(A_n + o_p(1))(\hat{\theta}_n - \theta_0) + o_p(1) = -n^{-1/2} \sum_{i=1}^n \mathrm{Grad}(w(\delta_i, Z_i, \theta_0)), \tag{10}$$

as $n \to \infty$, where $A_n = (a^n_{r,s})_{1 \leqslant r,s \leqslant k}$ is a $k \times k$ matrix with $a^n_{r,s} = D_{r,s}(l_n(\theta_0))$.

Since

$$\mathbb{E}(\mathrm{Grad}(w(\delta, Z, \theta_0))) = 0$$

and

$$
\begin{aligned}
&\mathbb{E}(D_r(w(\delta,Z,\theta_0))D_s(w(\delta,Z,\theta_0)))\\
&\quad=\mathbb{E}(\delta^2 D_r(\ln(m(Z,\theta_0)))D_s(\ln(m(Z,\theta_0))))\\
&\qquad+\mathbb{E}((1-\delta)^2 D_r(\ln(1-m(Z,\theta_0)))D_s(\ln(1-m(Z,\theta_0))))\\
&\quad=\mathbb{E}(m^{-1}(Z,\theta_0)D_r(m(Z,\theta_0))D_s(m(Z,\theta_0)))\\
&\qquad+\mathbb{E}((1-m(Z,\theta_0))^{-1}D_r(m(Z,\theta_0))D_s(m(Z,\theta_0)))\\
&\quad=\mathbb{E}\left(\frac{D_r(m(Z,\theta_0))D_s(m(Z,\theta_0))}{m(Z,\theta_0)(1-m(Z,\theta_0))}\right)
\end{aligned}
$$

the right-hand side of (10) is asymptotically normal with covariance matrix $I(\theta_0)$ according to the multivariate CLT.

Consider now the left-hand side of (10) and fix $1\leqslant r,s\leqslant k$. Then the SLLN yields

$$
D_{r,s}(l_n(\theta_0))=\frac{1}{n}\sum_{i=1}^{n}D_{r,s}w(\delta_i,Z_i,\theta_0)\longrightarrow\mathbb{E}(D_{r,s}(w(\delta,Z,\theta_0))).
$$

But

$$
\begin{aligned}
&\mathbb{E}(D_{r,s}(w(\delta,Z,\theta_0)))\\
&\quad=\mathbb{E}\left(m(Z,\theta_0)\frac{\partial}{\partial\theta_r}\left(\frac{1}{m(Z,\theta)}\frac{\partial}{\partial\theta_s}m(Z,\theta)\right)_{\theta=\theta_0}\right.\\
&\qquad\left.-(1-m(Z,\theta_0))\frac{\partial}{\partial\theta_r}\left(\frac{1}{1-m(Z,\theta)}\frac{\partial}{\partial\theta_s}m(Z,\theta)\right)_{\theta=\theta_0}\right)\\
&\quad=\mathbb{E}\left(m(Z,\theta_0)\left(-\frac{D_r(m(Z,\theta_0))D_s(m(Z,\theta_0))}{m^2(Z,\theta_0)}+\frac{D_{r,s}(m(Z,\theta_0))}{m(Z,\theta_0)}\right)\right.\\
&\qquad\left.-(1-m(Z,\theta_0))\left(\frac{D_r(m(Z,\theta_0))D_s(m(Z,\theta_0))}{(1-m(Z,\theta_0))^2}+\frac{D_{r,s}(m(Z,\theta_0))}{1-m(Z,\theta_0)}\right)\right)\\
&\quad=-\mathbb{E}\left(\frac{D_r(m(Z,\theta_0))D_s(m(Z,\theta_0))}{m(Z,\theta_0)(1-m(Z,\theta_0))}\right)\\
&\quad=-\sigma_{r,s}
\end{aligned}
$$

This, together with a Cramér–Slutsky argument and the regularity if $I(\theta_0)$ completes the proof. □

As the proof of the theorem shows we have

Corollary 3.4. *If the assumptions of the Theorem 2.3 are satisfied, then*

$$n^{1/2}(\hat{\theta}_n - \theta_0) = n^{-1/2} \sum_{i=1}^{n} I^{-1}(\theta_0)(\mathrm{Grad}(w(\delta_i, Z_i, \theta_0))) + o_p(1), \quad n \to \infty. \tag{11}$$

In the following lemma (11) leads to a representation of $n^{1/2}(m(x, \hat{\theta}_n) - m(x, \theta_0))$ which will be used later on in the investigation of the process $n^{1/2}(\Lambda_n - \Lambda)$.

Lemma 3.5. *Let* $0 < T < \infty$ *with* $H(T) < 1$. *If the assumptions of Theorem* 2.3 *and* (A_6) *are satisfied then, as* $n \to \infty$,

$$n^{1/2}(m(x, \hat{\theta}_n) - m(x, \theta_0)) = n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i - m(Z_i, \theta_0)}{m(Z_i, \theta_0)(1 - m(Z_i, \theta_0))} \alpha(x, Z_i) + R_n(x),$$

where

$$\alpha(x, y) = \langle \mathrm{Grad}(m(x, \theta_0)) \mid I^{-1}(\theta_0)(\mathrm{Grad}(m(y, \theta_0))) \rangle$$

and $|R_n(x)| \leqslant M(x) \cdot O_p(n^{-1/2}) + o_p(1)$ *uniformly for* $0 \leqslant x \leqslant T$.

Proof. Taylor's expansion yields

$$\begin{aligned} n^{1/2}(m(x, \hat{\theta}_n) - m(x, \theta_0)) &= n^{1/2} \langle \mathrm{Grad}(m(x, \theta_0)) \mid \hat{\theta}_n - \theta_0 \rangle \\ &\quad + \frac{n^{1/2}}{2} \sum_{1 \leqslant r,s \leqslant k} D_{r,s}(m(x, \theta^*))(\hat{\theta}_{n,r} - \theta_{0,r})(\hat{\theta}_{n,s} - \theta_{0,s}), \end{aligned}$$

where $\theta^* \in \Theta$ lies in the interior of the line segment connecting $\hat{\theta}_n$ and θ_0. Furthermore, Satz 6.7 in Witting and Müller–Funk (1995) shows that θ^* can be chosen to be measurable. Since $n^{1/2}(\hat{\theta}_n - \theta_0)$ is asymptotically normal, (A_2) and (A_6) yields

$$\frac{n^{1/2}}{2} \left| \sum_{1 \leqslant r,s \leqslant k} D_{r,s}(m(x, \theta^*))(\hat{\theta}_{n,r} - \theta_{0,r})(\hat{\theta}_{n,s} - \theta_{0,s}) \right| \leqslant M(x)\, O_p(n^{-1/2}) + o_p(1),$$

where $o_p(1)$ is not depending on x. Now, apply (A_6) and (11) to get for $0 \leqslant x \leqslant T$

$$\begin{aligned} & n^{1/2} \langle \mathrm{Grad}(m(x, \theta_0)) \mid \hat{\theta}_n - \theta_0 \rangle \\ &\quad = n^{-1/2} \sum_{i=1}^{n} \langle \mathrm{Grad}(m(x, \theta_0)) \mid I^{-1}(\theta_0)(\mathrm{Grad}(w(\delta_i, Z_i, \theta_0))) \rangle + o_p(1). \end{aligned}$$

Since

$$\mathrm{Grad}(w(\delta_i, Z_i, \theta_0)) = \frac{\delta_i - m(Z_i, \theta_0)}{m(Z_i, \theta_0)(1 - m(Z_i, \theta_0))} \mathrm{Grad}(m(Z_i, \theta_0))$$

the proof is complete. □

The motivation of $\hat{F}_n$ shows that $\exp(-\Lambda_n(t)) \approx 1-\hat{F}_n(t)$. To simplify the argumentation in the forthcoming proofs we give in the following lemma a bound for this approximation which is based on Lemma 7.1 in Breslow and Crowley (1974).

Lemma 3.6. *Let* $0<T<\infty$ *with* $H(T)<1$. *Then we have with probability one for large* n

$$\sup_{0\leqslant t\leqslant T} |(1-\hat{F}_n(t))-\exp(-\Lambda_n(t))| \leqslant \frac{H_n(T)}{n(1-H_n(T))}.$$

Proof. The mean-value theorem yields

$$\begin{aligned}\sup_{0\leqslant t\leqslant T} |(1-\hat{F}_n(t))-\exp(-\Lambda_n(t))| &= \sup_{0\leqslant t\leqslant T} |\exp(r(t))[\ln(1-\hat{F}_n(t))+\Lambda_n(t)]| \\ &\leqslant \sup_{0\leqslant t\leqslant T} |\ln(1-\hat{F}_n(t))+\Lambda_n(t)|,\end{aligned}$$

where $r(t)$ lies between $\ln(1-\hat{F}_n(t))$ and $-\Lambda_n(t)$. Now, apply Lemma 7.1 of Breslow and Crowley (1974) to complete the proof. □

Proof of Theorem 2.4. Taylor's expansion yields

$$m(x,\hat{\theta}_n)=m(x,\theta_0)+\langle \mathrm{Grad}(m(x,\theta^*)) \mid \hat{\theta}_n-\theta_0\rangle,$$

where θ^* lies in the line segment connecting $\hat{\theta}_n$ and θ_0 and can be chosen to be measurable. Now, strong consistency of $\hat{\theta}_n$ guarantees that $\theta^*\in V(\theta_0)$ for large n so that

$$|\langle \mathrm{Grad}(m(x,\theta^*)) \mid \hat{\theta}_n-\theta_0\rangle| \leqslant kM(x)\|\hat{\theta}_n-\theta_0\|.$$

Therefore, we get with probability one for large n

$$\begin{aligned}|\Lambda_n(t)-\Lambda(t)| \leqslant{}& \left|\int_0^t \frac{m(x,\theta_0)}{1-H_n(x-)}H_n(\mathrm{d}x)-\int_0^t \frac{m(x,\theta_0)}{1-H(x)}H_n(\mathrm{d}x)\right| \\ &+\left|\int_0^t \frac{m(x,\theta_0)}{1-H(x)}H_n(\mathrm{d}x)-\int_0^t \frac{m(x,\theta_0)}{1-H(x)}H(\mathrm{d}x)\right| \\ &+k\|\hat{\theta}_n-\theta_0\| \int_0^t \frac{M(x)}{1-H_n(x-)}H_n(\mathrm{d}x).\end{aligned}$$

Since $H(t)<1$, the first term on the right-hand side tends to zero according to the Glivenko–Cantelli theorem and the second according to the SLLN. For the third term we have with probability one for large n and small ε that $1-H_n(x-)>1-H(t)-\varepsilon>0$

for all $0 \leqslant x \leqslant t$. Therefore, the third term is bounded from above by

$$k\|\hat{\theta}_n - \theta_0\| \frac{1}{1 - H(t) - \varepsilon} \int_0^\infty M(x) H_n(\mathrm{d}x)$$

which tends to zero with probability one by the SLLN and the assumed consistency of $\hat{\theta}_n$. Since $\Lambda_n(\cdot)$ and $\Lambda(\cdot)$ are nondecreasing a standard argument completes the proof of (5).

Now, (6) follows from the continuity of exp together with (10) and Lemma 3.6. □

The remaining part of this section focuses on weak convergence of the processes $n^{1/2}(\Lambda_n - \Lambda)$ and $n^{1/2}(\hat{F}_n - F)$ in $D[0, T]$, the space of right-continuous functions on $[0, T]$ with left-hand limits. See Billingsley (1968, p. 109).

In the next lemmata we develop a representation for $n^{1/2}(\Lambda_n - \Lambda)$ which easily leads to the corresponding weak limit.

Lemma 3.7. *If H is continuous, $H(T) < 1$, and the assumptions of Lemma 3.5 are satisfied then, as $n \to \infty$,*

$$\begin{aligned} n^{1/2}&(\Lambda_n(t) - \Lambda(t)) \\ &= n^{1/2} \int_0^t \frac{m(x,\theta_0)}{1 - H(x)} \mathrm{d}(H_n(x) - H(x)) \\ &\quad + n^{1/2} \int_0^t \frac{m(x,\theta_0)(H_n(x) - H(x))}{(1 - H(x))^2} H_n(\mathrm{d}x) \\ &\quad + n^{-1/2} \sum_{i=1}^n \frac{\delta_i - m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1 - m(Z_i,\theta_0))} \int_0^t \frac{\alpha(x,Z_i)}{1 - H(x)} H_n(\mathrm{d}x) + \mathrm{o_p}(1) \end{aligned} \tag{12}$$

uniformly on $[0, T]$.

Proof. A straightforward calculation shows

$$\begin{aligned} \frac{m(x,\hat{\theta}_n)}{1 - H_n(x-)} &= \frac{m(x,\theta_0)}{1 - H(x)} + \frac{m(x,\theta_0)(H_n(x-) - H(x))}{(1 - H(x))^2} + \frac{m(x,\hat{\theta}_n) - m(x,\theta_0)}{1 - H(x)} \\ &\quad + \frac{m(x,\hat{\theta}_n)(H_n(x-) - H(x))^2}{(1 - H_n(x-))(1 - H(x))^2} \\ &\quad + \frac{m(x,\hat{\theta}_n) - m(x,\theta_0)}{(1 - H(x))^2}(H_n(x-) - H(x)) \\ &= I_1(x) + I_2(x) + I_3(x) + I_4(x) + I_5(x). \end{aligned}$$

Since $|H_n(x-) - H_n(x)| \leqslant 1/n$ we get

$$n^{1/2} \sup_{0\leqslant t\leqslant T}\left|\int_0^t I_5(x)H_n(\mathrm{d}x)\right| \leqslant n^{1/2}\frac{\|H_n - H\| + 1/n}{(1-H(T))^2} \times \int_0^T |m(x,\hat{\theta}_n) - m(x,\theta_0)|H_n(\mathrm{d}x).$$

Taylor's expansion together with (A_2) and (A_6) yields

$$\int_0^T |m(x,\hat{\theta}_n) - m(x,\theta_0)|H_n(\mathrm{d}x) \leqslant \sup_{0\leqslant x\leqslant T} \|\mathrm{Grad}(m(x,\theta_0))\| \cdot \|\hat{\theta}_n - \theta_0\| + \frac{\|\hat{\theta}_n - \theta_0\|^2}{2}k^2 \int_0^T M(x)H_n(\mathrm{d}x) + \mathrm{o_p}(1).$$

Since $n^{1/2}\|H_n - H\|$ is bounded in probability, the consistency of $\hat{\theta}_n$ together with the SLLN guarantee that

$$n^{1/2} \sup_{0\leqslant t\leqslant T}\left|\int_0^t I_5(x)H_n(\mathrm{d}x)\right| = \mathrm{o_p}(1).$$

Now, take $\varepsilon > 0$ such that $H(T) + \varepsilon < 1$. Then we get with probability one for large n

$$n^{1/2} \sup_{0\leqslant t\leqslant T}\left|\int_0^t I_4(x)H_n(\mathrm{d}x)\right| \leqslant n^{1/2}\frac{(\|H_n - H\| + 1/n)^2}{(1-H(T)-\varepsilon)^3}.$$

Since $n^{1/2}\|H_n - H\|$ is bounded in probability the right-hand side tends to zero in probability according to the Glivenko–Cantelli theorem.

For the third term we apply Lemma 3.5 to get

$$\sup_{0\leqslant t\leqslant T}\left| n^{1/2}\int_0^t I_3(x)H_n(\mathrm{d}x) - n^{-1/2}\sum_{i=1}^n \frac{\delta_i - m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))} \times \int_0^t \frac{\alpha(x,Z_i)}{1-H(x)}H_n(\mathrm{d}x)\right|$$

$$\leqslant \int_0^T \frac{|R_n(x)|}{1-H(x)}H_n(\mathrm{d}x) = \mathrm{o_p}(1).$$

Since $|H_n(x-) - H_n(x)| \leqslant 1/n$ we can replace $H_n(x-)$ by $H_n(x)$ in the I_2 term which completes the proof of the lemma. □

Our next lemma simplifies the second term on the right-hand side of (12).

Lemma 3.8. *If $H(T)<1$ then, as $n\to\infty$,*

$$n^{1/2}\sup_{0\leqslant t\leqslant T}\left|\int_0^t \frac{m(x,\theta_0)(H_n(x)-H(x))}{(1-H(x))^2}H_n(\mathrm{d}x) - \int_0^t \frac{m(x,\theta_0)(H_n(x)-H(x))}{(1-H(x))^2}H(\mathrm{d}x)\right|$$

tends to 0 in probability.

Proof. First observe that

$$\int_0^t \frac{m(x,\theta_0)(H_n(x)-H(x))}{(1-H(x))^2}H_n(\mathrm{d}x)=U_n(t)-\frac{1}{n}U_n(t)+\frac{1}{n^2}\sum_{i=1}^n \frac{m(Z_i,\theta_0)}{1-H(Z_i)}1_{[Z_i\leqslant t]},$$

where

$$U_n(t)=\frac{1}{n(n-1)}\sum_{1\leqslant i\neq j\leqslant n} h(Z_i,Z_j)\cdot 1_{[Z_i\leqslant t]}$$

is a U-statistic process as studied in Stute (1994) with kernel

$$h(x,y)=\frac{m(x,\theta_0)(1_{[y\leqslant x]}-H(x))}{(1-H(x))^2}\cdot 1_{[x\leqslant T]}.$$

Obviously,

$$\sup_{0\leqslant t\leqslant T}\frac{1}{n^2}\sum_{i=1}^n \frac{m(Z_i,\theta_0)}{1-H(Z_i)}1_{[Z_i\leqslant t]}\leqslant \frac{1}{n(1-H(T))}=\mathrm{O}(n^{-1})$$

and

$$\sup_{0\leqslant t\leqslant T}\left|\frac{1}{n}U_n(t)\right|\leqslant \frac{1}{n(1-H(T))^2}=\mathrm{O}(n^{-1}).$$

Note that $\int_0^\infty h(x,y)H(\mathrm{d}y)=0$ and $h\in L_2(H\otimes H)$. Corollary 1.1 in Stute (1994) then yields

$$\mathbb{E}\left(\sup_{0\leqslant t\leqslant T}\left|U_n(t)-\int_0^\infty\int_0^t h(x,y)H(\mathrm{d}x)H_n(\mathrm{d}y)\right|^2\right)=\mathrm{O}(n^{-2}).$$

Since

$$\int_0^\infty\int_0^t h(x,y)H(\mathrm{d}x)H_n(\mathrm{d}y)=\int_0^t \frac{m(x,\theta_0)(H_n(x)-H(x))}{1-H(x)}H(\mathrm{d}x)$$

the proof is complete. □

The next three lemmata simplify the third term on the right-hand side of (12). In particular, we show that the integrating measure H_n appearing in the third term on the right-hand side of (12) can be replaced by H. Basically, this is achieved by another application of Stute's (1994) results on U-Statistic processes.

Lemma 3.9. *If H is continuous, $H(T)<1$, and the assumptions of Lemma 3.5 are satisfied, then with probability one, as $n \to \infty$,*

$$n^{-1/2}\sum_{i=1}^{n}\frac{\delta_i - m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))}\int_0^t \frac{\alpha(x,Z_i)}{1-H(x)}H_n(\mathrm{d}x)$$
$$= n^{1/2}(U_{1,n}(t) - U_{2,n}(t)) + \mathrm{O}(n^{-1/2})$$

uniformly on $[0,T]$, where

$$U_{1,n}(t) = \frac{1}{n(n-1)}\sum_{1\leqslant i\neq j\leqslant n}\frac{\delta_i\alpha(Z_j,Z_i)}{m(Z_i,\theta_0)(1-H(Z_j))}1_{[Z_j\leqslant t]}$$

$$U_{2,n}(t) = \frac{1}{n(n-1)}\sum_{1\leqslant i\neq j\leqslant n}\frac{(1-\delta_i)\alpha(Z_j,Z_i)}{(1-m(Z_i,\theta_0))(1-H(Z_j))}1_{[Z_j\leqslant t]}.$$

Proof. Since

$$\frac{\delta_i - m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))} = \frac{\delta_i}{m(Z_i,\theta_0)} - \frac{1-\delta_i}{1-m(Z_i,\theta_0)}$$

a straightforward calculation shows that the left-hand side equals

$$n^{1/2}(U_{1,n}(t) - U_{2,n}(t)) - n^{-1/2}(U_{1,n}(t) - U_{2,n}(t)) + n^{-1/2}R_n(t),$$

where

$$R_n(t) = n^{-1}\sum_{i=1}^{n}\frac{\delta_i - m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))}\cdot\frac{\alpha(Z_i,Z_i)}{1-H(Z_i)}1_{[Z_i\leqslant t]}.$$

According to (A_6) we have for an appropriate constant $c>0$

$$\sup_{0\leqslant x\leqslant T}\left|\frac{\alpha(x,Z_i)}{m(Z_i,\theta_0)}\right| \leqslant c\sum_{s=1}^{k}\frac{|D_s(m(Z_i,\theta_0))|}{m(Z_i,\theta_0)}.$$

From (A_4) and the SLLN we then get with probability one, as $n\to\infty$,

$$\sup_{0\leqslant t\leqslant T}|U_{1,n}(t)| \leqslant \frac{c}{n(1-H(T))}\sum_{i=1}^{n}\sum_{s=1}^{k}\frac{|D_s(m(Z_i,\theta_0))|}{m(Z_i,\theta_0)} = \mathrm{O}(1).$$

A similar approximation also holds for $U_{2,n}(t)$ and $R_n(t)$ which finally proves the lemma. □

Lemma 3.10. *If H is continuous, $H(T)<1$, and the assumptions of Lemma 3.5 are satisfied then, as $n \to \infty$,*

$$n^{1/2} \sup_{0\leqslant t\leqslant T} \left| U_{1,n}(t) - \int_0^\infty \int_0^t h_1(x,y)H(\mathrm{d}x)H_{1,n}(\mathrm{d}y) \right.$$
$$\left. - \int_0^\infty \int_0^t h_1(x,y)H_n(\mathrm{d}x)H_1(\mathrm{d}y) + \int_0^\infty \int_0^t h_1(x,y)H(\mathrm{d}x)H_1(\mathrm{d}y) \right|$$

tends to 0 in probability, where

$$h_1(x,y) = \frac{1_{[y>0]} \cdot \alpha(x,y)}{m(y,\theta_0)(1-H(x))} \cdot 1_{[x\leqslant T]},$$

H_1 is the d.f. of $\bar{Z} = \delta \cdot Z$, and $H_{1,n}$ the empirical d.f. of the $\bar{Z}$-sample.

Proof. Since H is continuous

$$\frac{\delta_i \alpha(x,Z_i)}{m(Z_i,\theta_0)} = \frac{1_{[\bar{Z}_i>0]}\alpha(x,\bar{Z}_i)}{m(\bar{Z}_i,\theta_0)}$$

with probability one. Therefore,

$$U_{1,n}(t) = \frac{1}{n(n-1)} \sum_{1\leqslant i\neq j\leqslant n} h_1(Z_j,\bar{Z}_i) \cdot 1_{[Z_j\leqslant t]}$$

is a U-statistic process with

$$\int_0^\infty \int_0^T h_1^2(x,y)H(\mathrm{d}x)H_1(\mathrm{d}y) = \mathbb{E}\left(\left(\frac{\delta_2\alpha(Z_1,Z_2)}{m(Z_2,\theta_0)(1-H(Z_1))}1_{[Z_1\leqslant T]}\right)^2\right)$$
$$\leqslant c\mathbb{E}\left(\left(\sum_{s=1}^k \frac{D_s(m(Z_2,\theta_0))}{m(Z_2,\theta_0)}\right)^2\right) < \infty$$

according to (A_4) and (A_6) for an appropriate constant c. An application of Theorem 1.5 in Stute (1994) then completes the proof. □

A similar result holds for $U_{2,n}(t)$.

Lemma 3.11. *If H is continuous, $H(T)<1$, and the assumptions of Lemma 3.5 are satisfied then, as $n \to \infty$,*

$$n^{1/2} \sup_{0\leqslant t\leqslant T} \left| U_{2,n}(t) - \int_0^\infty \int_0^t h_2(x,y)H(\mathrm{d}x)H_{2,n}(\mathrm{d}y) \right.$$
$$\left. - \int_0^\infty \int_0^t h_2(x,y)H_n(\mathrm{d}x)H_2(\mathrm{d}y) + \int_0^\infty \int_0^t h_2(x,y)H(\mathrm{d}x)H_2(\mathrm{d}y) \right|$$

tends to 0 *in probability, where*

$$h_2(x,y)=\frac{1_{[y>0]}\cdot\alpha(x,y)}{(1-m(y,\theta_0))(1-H(x))}\cdot 1_{[x\leqslant T]},$$

H_2 *is the d.f. of* $\bar{\bar{Z}}=(1-\delta)Z$, *and* $H_{2,n}$ *is the empirical d.f. of the* $\bar{\bar{Z}}$*-sample.*

Straightforward calculation shows that

$$\int_0^\infty\int_0^t h_1(x,y)H(\mathrm{d}x)H_{1,n}(\mathrm{d}y)-\int_0^\infty\int_0^t h_2(x,y)H(\mathrm{d}x)H_{2,n}(\mathrm{d}y)$$
$$=\frac{1}{n}\sum_{i=1}^n\frac{\delta_i-m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))}\cdot\int_0^t\frac{\alpha(x,Z_i)}{1-H(x)}H(\mathrm{d}x)$$

and, since $\mathbb{E}(\alpha(x,Z)(\delta-m(Z,\theta_0))/(m(Z,\theta_0)(1-m(Z,\theta_0))))=0$,

$$\int_0^\infty\int_0^t h_1(x,y)H_n(\mathrm{d}x)H_1(\mathrm{d}y)-\int_0^\infty\int_0^t h_2(x,y)H_n(\mathrm{d}x)H_2(\mathrm{d}y)=0$$

and

$$\int_0^\infty\int_0^t h_1(x,y)H(\mathrm{d}x)H_1(\mathrm{d}y)-\int_0^\infty\int_0^t h_2(x,y)H(\mathrm{d}x)H_2(\mathrm{d}y)=0.$$

Therefore, the last lemmata yield

Lemma 3.12. *If* H *is continuous,* $H(T)<1$, *and the assumptions of Lemma* 3.5 *are satisfied, then, as* $n\to\infty$,

$$n^{1/2}(\Lambda_n(t)-\Lambda(t))=n^{1/2}S_n(t)+o_p(1)$$

uniformly on $[0,T]$, *where*

$$S_n(t)=\frac{H_n^1(t)-H^1(t)}{1-H(t)}-\int_0^t\frac{H_n^1(x)-H^1(x)}{(1-H(x))^2}H(\mathrm{d}x)+\int_0^t\frac{H_n(x)-H(x)}{(1-H(x))^2}H^1(\mathrm{d}x)$$
$$+\frac{1}{n}\sum_{i=1}^n\frac{\delta_i-m(Z_i,\theta_0)}{m(Z_i,\theta_0)(1-m(Z_i,\theta_0))}\int_0^t\frac{\alpha(x,Z_i)}{1-H(x)}H(\mathrm{d}x),\tag{13}$$

$H^1(t)=\int_0^t m(x,\theta_0)H(\mathrm{d}x)$ *and* $H_n^1(t)=\int_0^t m(x,\theta_0)H_n(\mathrm{d}x)$.

Proof. Integrating by parts yields

$$\int_0^t\frac{m(x,\theta_0)}{1-H(x)}\mathrm{d}(H_n(x)-H(x))=\frac{H_n^1(t)-H^1(t)}{1-H(t)}-\int_0^t\frac{H_n^1(x)-H^1(x)}{(1-H(x))^2}H(\mathrm{d}x)$$

so that (13) follows from (12) and the preceding lemmata. □

The representation (13) incorporates the processes

$$\alpha_n(x) = n^{1/2}(H_n(x) - H(x)),$$

$$\beta_n(x) = n^{1/2}(H_n^1(x) - H^1(x)),$$

and

$$\gamma_n(x) = n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i - m(Z_i, \theta_0)}{m(Z_i, \theta_0)(1 - m(Z_i, \theta_0))} \alpha(x, Z_i),$$

which are in $D[0, T]$.

Lemma 3.13. *If H is continuous, $H(T) < 1$, the assumptions of Lemma 3.5 and* (A_7) *are satisfied, then $(\alpha_n, \beta_n, \gamma_n) \in (D[0, T])^3$ converges weakly to a centered Gaussian process (α, β, γ).*

Proof. Obviously, each of the processes is centered. Since

$$(\alpha_n(x), \beta_n(x), \gamma_n(x)) = n^{-1/2} \sum_{i=1}^{n} \Bigg(1_{[Z_i \leqslant x]} - H(x), m(Z_i, \theta_0) 1_{[Z_i \leqslant x]} - H^1(x), \frac{\delta_i - m(Z_i, \theta_0)}{m(Z_i, \theta_0)(1 - m(Z_i, \theta_0))} \alpha(x, Z_i) \Bigg)$$

the weak convergence of the finite dimensional distributions to a multivariate normal distribution follows from the multivariate CLT. Hence it remains to prove tightness. But tightness of α_n is well known (see Billingsley (1968, Theorem 16.4)) and with the same argumentation tightness of β_n can be shown. We will prove here only the tightness of γ_n which together with the tightness of the two other processes then implies the tightness of $(\alpha_n, \beta_n, \gamma_n)$.

Now,

$$\gamma_n(x) = \sum_{1 \leqslant r,s \leqslant k} \gamma_n^{r,s}(x)$$

where

$$\gamma_n^{r,s}(x) = \left(n^{-1/2} \sum_{i=1}^{n} \frac{\delta_i - m(Z_i, \theta_0)}{m(Z_i, \theta_0)(1 - m(Z_i, \theta_0))} D_s m(Z_i, \theta_0) \right) \bar{\sigma}_{r,s} D_r m(x, \theta_0)$$

and $I^{-1}(\theta_0) = (\bar{\sigma}_{r,s})_{1 \leqslant r,s \leqslant k}$.

Observe that $\gamma_n^{r,s} \in C[0, T]$ the space of continuous functions on $[0, T]$. Since

$$\mathbb{E}((\gamma_n^{r,s}(x_2) - \gamma_n^{r,s}(x_1))^2) = \bar{\sigma}_{r,s}^2 (D_r m(x_2, \theta_0) - D_r m(x_1, \theta_0))^2 \times \mathbb{E}\left(\left(\frac{\delta - m(Z, \theta_0)}{m(Z, \theta_0)(1 - m(Z, \theta_0))} D_s m(Z, \theta_0) \right)^2 \right)$$

tightness in $C[0, T]$ of $\gamma_n^{r,s}$ follows from (A_7) according to Theorem 12.3 in Billingsley (1968). Therefore, γ_n is tight in $C[0, T]$. Since tightness in C implies tightness in D, the proof is complete. □

Proof of Theorem 2.5. According to Lemma 3.12 we have to prove weak convergence of $n^{1/2}S_n$ in $D[0,T]$. Since

$$n^{1/2}S_n(t)=\frac{\beta_n(t)}{1-H(t)}-\int_0^t\frac{\beta_n(x)}{(1-H(x))^2}H(\mathrm{d}x)+\int_0^t\frac{\alpha_n(x)}{(1-H(x))^2}H^1(\mathrm{d}x)$$
$$+\int_0^t\frac{\gamma_n(x)}{1-H(x)}H(\mathrm{d}x)$$

the weak convergence of $n^{1/2}S_n$ follows from Lemma 3.13 and the continuous mapping theorem. In Appendix we show that the covariance structure of S has the form given in (7). □

Corollary 2.6 is a consequence of Theorem 2.5 and Lemma 3.6.

Proof of Corollary 2.7. Since

$$v^{\mathrm{km}}(t)=(1-F(t))^2\int_0^t\frac{1}{(1-H(x))^2}H^1(\mathrm{d}x),$$

Corollary 2.6 yields

$$v^{\mathrm{km}}(t)-v(t)=(1-F(t))^2r(t).$$

It remains to prove that $r(t)\geqslant 0$ for $0\leqslant t\leqslant T$. But

$$\int_0^t\int_0^t\frac{\alpha(x,y)}{(1-H(x))(1-H(y))}H(\mathrm{d}y)H(\mathrm{d}x)=\langle b|I^{-1}(\theta_0)(b)\rangle,$$

where $b\in\mathbb{R}^k$ is defined by

$$b=\left(\mathbb{E}\left(\frac{D_1m(Z,\theta_0)}{1-H(Z)}1_{[Z\leqslant t]}\right),\cdots,\mathbb{E}\left(\frac{D_km(Z,\theta_0)}{1-H(Z)}1_{[Z\leqslant t]}\right)\right).$$

Since $I^{-1}(\theta_0)$ is positive definite we get by a well known argument (see Rao (1973, 1f.1.1))

$$\sup_{h\in\mathbb{R}^k\setminus\{0\}}\frac{\langle h|b\rangle^2}{\langle h|I(\theta_0)(h)\rangle}=\langle b|I^{-1}(\theta_0)(b)\rangle.$$

Therefore, the proof is complete if we show that for all $h\in\mathbb{R}^k\setminus\{0\}$

$$\langle h|b\rangle^2\leqslant\int_0^t\frac{1-m(x,\theta_0)}{(1-H(x))^2}H^1(\mathrm{d}x)\times\langle h|I(\theta_0)(h)\rangle.$$

Fix $h=(h_1,\cdots,h_k)\in\mathbb{R}^k\setminus\{0\}$ and define

$$\tilde{H}(t)=\int_0^t m(x,\theta_0)(1-m(x,\theta_0))H(\mathrm{d}x).$$

The Cauchy–Schwarz inequality yields

$$\langle h|b\rangle^2 = \left(\int_0^t \frac{1}{1-H(x)} \sum_{r=1}^k \frac{h_r D_r m(x,\theta_0)}{m(x,\theta_0)(1-m(x,\theta_0))} \tilde{H}(\mathrm{d}x)\right)^2$$

$$\leqslant \int_0^t \frac{1}{(1-H(x))^2} \tilde{H}(\mathrm{d}x) \int_0^t \left(\sum_{r=1}^k \frac{h_r D_r m(x,\theta_0)}{m(x,\theta_0)(1-m(x,\theta_0))}\right)^2 \tilde{H}(\mathrm{d}x)$$

$$\leqslant \int_0^t \frac{1}{(1-H(x))^2} \tilde{H}(\mathrm{d}x) \times \langle h|I(\theta_0)(h)\rangle$$

which completes the proof. □

Appendix

We now calculate the covariance structure of the limiting process S given in Theorem 2.5.

First observe that

$$n^{1/2} S_n(t) = n^{-1/2} \sum_{i=1}^n (A(Z_i,t) + B(Z_i,t) + C(Z_i,t) + D(Z_i,t))$$

where

$$A(t) = A(Z,t) = \frac{m(Z,\theta_0) 1_{[Z \leqslant t]} - H^1(t)}{1-H(t)}$$

$$B(t) = B(Z,t) = -\int_0^t \frac{m(Z,\theta_0) 1_{[Z \leqslant x]} - H^1(x)}{(1-H(x))^2} H(\mathrm{d}x)$$

$$C(t) = C(Z,t) = \int_0^t \frac{1_{[Z \leqslant x]} - H(x)}{(1-H(x))^2} H^1(\mathrm{d}x)$$

$$D(t) = D(Z,t) = \frac{\delta - m(Z,\theta_0)}{m(Z,\theta_0)(1-m(Z,\theta_0))} \int_0^t \frac{\alpha(x,Z)}{1-H(x)} H(\mathrm{d}x)$$

and each of the terms is centered. Recall that

$$H^1(t) = \int_0^t m(x,\theta_0) H(\mathrm{d}x)$$

and define

$$H^1_*(t) = \int_0^t m(x,\theta_0) H^1(\mathrm{d}x).$$

The calculation is done as in the Appendix of Breslow and Crowley (1974) using $\mathrm{Cov}(S(s),S(t)) = \mathrm{Var}(S(s)) + \mathrm{Cov}(S(s),S(t)-S(s))$ with $0 \leqslant s \leqslant t \leqslant T$.

$$(a_1)\quad \mathrm{Var}(A(s)) = \frac{H^1_*(s) - H^1(s)H^1(s)}{(1-H(s))^2}$$

$$(a_2)\quad \mathrm{Var}(B(s)) = 2\int_0^s\int_0^x \frac{H^1_*(y) - H^1(x)H^1(y)}{(1-H(x))^2(1-H(y))^2} H(\mathrm{d}y)H(\mathrm{d}x)$$

$$(a_3)\quad \mathrm{Var}(C(s)) = 2\int_0^s\int_0^x \frac{H(y) - H(x)H(y)}{(1-H(x))^2(1-H(y))^2} H^1(\mathrm{d}y)H^1(\mathrm{d}x)$$

$$\begin{aligned}(a_4)\quad \mathrm{Var}(D(s)) &= \int_0^s\int_0^s \mathbb{E}\left(\left[\frac{\delta - m(Z,\theta_0)}{m(Z,\theta_0)(1-m(Z,\theta_0))}\right]^2 \right.\\ &\qquad \left.\times \frac{\alpha(x,Z)\alpha(y,Z)}{(1-H(x))(1-H(y))}\right) H(\mathrm{d}y)H(\mathrm{d}x)\\ &= \int_0^s\int_0^s \mathbb{E}\left(\left[\frac{\alpha(x,Z)\alpha(y,Z)}{m(Z,\theta_0)(1-m(Z,\theta_0))}\right]\right)\\ &\qquad \times \frac{1}{(1-H(x))(1-H(y))} H(\mathrm{d}y)H(\mathrm{d}x)\\ &= \int_0^s\int_0^s \frac{\alpha(x,y)}{(1-H(x))(1-H(y))} H(\mathrm{d}y)H(\mathrm{d}x)\end{aligned}$$

$$(a_5)\quad 2\mathrm{Cov}(A(s),B(s)) = -\frac{2}{1-H(s)}\int_0^s \frac{H^1_*(x) - H^1(x)H^1(s)}{(1-H(x))^2} H(\mathrm{d}x)$$

$$(a_6)\quad 2\mathrm{Cov}(A(s),C(s)) = \frac{2}{1-H(s)}\int_0^s \frac{H^1(x) - H^1(s)H(x)}{(1-H(x))^2} H^1(\mathrm{d}x)$$

$$\begin{aligned}(a_7)\quad 2\mathrm{Cov}(B(s),C(s)) &= -2\int_0^s\int_0^x \frac{H^1(y) - H(y)H^1(x)}{(1-H(x))^2(1-H(y))^2} H^1(\mathrm{d}y)H(\mathrm{d}x)\\ &\quad -2\int_0^s\int_0^x \frac{H^1(y) - H^1(y)H(x)}{(1-H(x))^2(1-H(y))^2} H(\mathrm{d}y)H^1(\mathrm{d}x).\end{aligned}$$

Conditioning on Z yields for the remaining terms

$$\mathrm{Cov}(A(s),D(s)) = \mathrm{Cov}(B(s),D(s)) = \mathrm{Cov}(C(s),D(s)) = 0.$$

Now, $\mathrm{Var}(S(s))$ equals the sum of the terms (a_1)–(a_7). If we compare these terms with those given in the Appendix of Breslow and Crowley (1974), abbreviated there by $(A.1),\ldots,(A.6)$, we find $(a_3)=(A.1)$, $(a_6)=(A.4)$, $(a_7)=(A.5)$. Furthermore,

$$(a_1)=(A.2)\ -\frac{H^1(s)-H_*^1(s)}{(1-H(s))^2},$$

$$(a_2)=(A.3)\ -2\int_0^s\int_0^x \frac{H^1(y)-H_*^1(y)}{(1-H(x))^2(1-H(y))^2}H(\mathrm{d}y)H(\mathrm{d}x),$$

$$(a_5)=(A.6)\ +\frac{2}{1-H(s)}\int_0^s \frac{H^1(x)-H_*^1(x)}{(1-H(x))^2}H(\mathrm{d}x).$$

According to the Appendix in Breslow and Crowley (1974)

$$(A.1)+\cdots+(A.6)=\int_0^s \frac{1}{(1-H(x))^2}H^1(\mathrm{d}x).$$

Therefore,

$$\begin{aligned}\mathrm{Var}(S(s)) = &\int_0^s \frac{1}{(1-H(x))^2}H^1(\mathrm{d}x)\\ &+\int_0^s\int_0^s \frac{\alpha(x,y)}{(1-H(x))(1-H(y))}H(\mathrm{d}y)H(\mathrm{d}x)\\ &-R(s)\end{aligned}$$

where

$$\begin{aligned}R(s) = &\frac{H^1(s)-H_*^1(s)}{(1-H(s))^2}+2\int_0^s\int_0^x \frac{H^1(y)-H_*^1(y)}{(1-H(x))^2(1-H(y))^2}H(\mathrm{d}y)H(\mathrm{d}x)\\ &-\frac{2}{1-H(s)}\int_0^s \frac{H^1(x)-H_*^1(x)}{(1-H(x))^2}H(\mathrm{d}x).\end{aligned}$$

Since $H^1(y)\geqslant H_*^1(y)$

$$H_{**}(x)=\int_0^x \frac{H^1(y)-H_*^1(y)}{(1-H(y))^2}H(\mathrm{d}y)$$

is a sub-distribution function. Integration by parts then yields

$$\begin{aligned}&\int_0^s\int_0^x \frac{H^1(y)-H_*^1(y)}{(1-H(x))^2(1-H(y))^2}H(\mathrm{d}y)H(\mathrm{d}x)\\ &=\int_0^s \frac{H_{**}(x)}{(1-H(x))^2}H(\mathrm{d}x)=\frac{H_{**}(s)}{1-H(s)}-\int_0^s \frac{1}{1-H(x)}H_{**}(\mathrm{d}x)\\ &=\frac{H_{**}(s)}{1-H(s)}-\int_0^s \frac{H^1(x)-H_*^1(x)}{(1-H(x))^3}H(\mathrm{d}x).\end{aligned}$$

Therefore,

$$R(s)=\frac{H^1(s)-H_*^1(s)}{(1-H(s))^2}-2\int_0^s \frac{H^1(x)-H_*^1(x)}{(1-H(x))^3}H(\mathrm{d}x).$$

Using integration by parts again we get

$$2\int_0^s \frac{H^1(x)-H^1_*(x)}{(1-H(x))^3}H(\mathrm{d}x)=\frac{H^1(s)-H^1_*(s)}{(1-H(s))^2}-\int_0^s \frac{1-m(x,\theta_0)}{(1-H(x))^2}H^1(\mathrm{d}x).$$

Hence

$$R(s)=\int_0^s \frac{1-m(x,\theta_0)}{(1-H(x))^2}H^1(\mathrm{d}x)$$

and finally

$$\begin{aligned}\mathrm{Var}(S(s)) &= \int_0^s \frac{m(x,\theta_0)}{(1-H(x))^2}H^1(\mathrm{d}x)\\ &\quad+\int_0^s\int_0^s \frac{\alpha(x,y)}{(1-H(x))(1-H(y))}H(\mathrm{d}x)H(\mathrm{d}y).\end{aligned} \tag{14}$$

The calculation of $\mathrm{Cov}(S(s),S(t)-S(s))$ is done in the same way. We calculate all the terms and compare them with the corresponding ones in Breslow and Crowley (1974), denoted there by $(B.1),\dots,(B.9)$.

(b_1) $\mathrm{Cov}(A(s),A(t)-A(s))$

$$=\frac{H^1_*(s)-H^1(s)H^1(t)}{(1-H(s))(1-H(t))}-\frac{H^1_*(s)-H^1(s)H^1(s)}{(1-H(s))^2}$$

(b_2) $\mathrm{Cov}(A(s),B(t)-B(s))$

$$=-\frac{1}{1-H(s)}\int_s^t \frac{H^1_*(s)-H^1(s)H^1(x)}{(1-H(x))^2}H(\mathrm{d}x)$$

(b_3) $\mathrm{Cov}(A(s),C(t)-C(s))$

$$=\frac{H^1(s)}{1-H(s)}\int_s^t \frac{1}{1-H(x)}H^1(\mathrm{d}x)$$

(b_4) $\mathrm{Cov}(B(s),A(t)-A(s))$

$$\begin{aligned}&=\frac{1}{1-H(s)}\int_0^s \frac{H^1_*(x)-H^1(x)H^1(s)}{(1-H(x))^2}H(\mathrm{d}x)\\ &\quad-\frac{1}{1-H(t)}\int_0^s \frac{H^1_*(x)-H^1(x)H^1(t)}{(1-H(x))^2}H(\mathrm{d}x)\end{aligned}$$

(b_5) $\mathrm{Cov}(B(s),B(t)-B(s))$

$$=\int_0^s\int_s^t \frac{H^1_*(x)-H^1(x)H^1(y)}{(1-H(x))^2(1-H(y))^2}H(\mathrm{d}y)H(\mathrm{d}x)$$

(b_6) $\operatorname{Cov}(B(s), C(t) - C(s))$

$$= -\int_0^s \int_s^t \frac{H^1(x)}{(1-H(x))^2(1-H(y))} H^1(\mathrm{d}y)H(\mathrm{d}x)$$

(b_7) $\operatorname{Cov}(C(s), A(t) - A(s))$

$$= \left(\frac{1}{1-H(t)} - \frac{1}{1-H(s)}\right) \int_0^s \frac{H^1(x)}{(1-H(x))^2} H^1(\mathrm{d}x)$$

$$- \left(\frac{H^1(t)}{1-H(t)} - \frac{H^1(s)}{1-H(s)}\right) \int_0^s \frac{H(x)}{(1-H(x))^2} H^1(\mathrm{d}x)$$

(b_8) $\operatorname{Cov}(C(s), B(t) - B(s))$

$$= -\int_0^s \int_s^t \frac{H^1(x) - H(x)H^1(y)}{(1-H(x))^2(1-H(y))^2} H(\mathrm{d}y)H^1(\mathrm{d}x)$$

(b_9) $\operatorname{Cov}(C(s), C(t) - C(s))$

$$= \int_0^s \int_s^t \frac{H(x)}{(1-H(y))(1-H(x))^2} H^1(\mathrm{d}y)H^1(\mathrm{d}x)$$

(b_{10}) $\operatorname{Cov}(D(s), D(t) - D(s))$

$$= \int_0^s \int_s^t \frac{\alpha(x,y)}{(1-H(y))(1-H(x))} H(\mathrm{d}y)H(\mathrm{d}x).$$

Conditioning on Z shows that all the other terms which have to be added to $(b_1) + \cdots + (b_{10})$ to get $\operatorname{Cov}(S(s), S(t) - S(s))$ are zero.

Comparing the terms with those given in the Appendix of Breslow and Crowley (1974), denoted there by $(B.1), \ldots, (B.9)$, we find $(b_3) = (B.4)$, $(b_6) = (B.7)$, $(b_7) = (B.2)$, $(b_8) = (B.3)$, and $(b_9) = (B.1)$.

A comparison of the other terms shows

$$(b_1) = (B.5) + \frac{H^1_*(s) - H^1(s)}{(1-H(s))(1-H(t))} - \frac{H^1_*(s) - H^1(s)}{(1-H(s))^2}$$

$$(b_2) = (B.6) - \frac{H^1_*(s) - H^1(s)}{1-H(s)} \int_s^t \frac{1}{(1-H(x))^2} H(\mathrm{d}x)$$

$$(b_4) = (B.8) + \left(\frac{1}{1-H(s)} - \frac{1}{1-H(t)}\right) \int_0^s \frac{H^1_*(x) - H^1(x)}{(1-H(x))^2} H(\mathrm{d}x)$$

$$(b_5) = (B.9) + \int_0^s \int_s^t \frac{H^1_*(x) - H^1(x)}{(1-H(x))^2(1-H(y))^2} H(\mathrm{d}y)H(\mathrm{d}x).$$

According to Breslow and Crowley (1974) the sum of $(B.1)$–$(B.3)$ is zero as is the sum of $(B.4)$–$(B.6)$ and $(B.7)$–$(B.9)$. Therefore,

$$
\begin{aligned}
(b_1)+(b_2)+(b_3) &= \frac{H_*^1(s)-H^1(s)}{(1-H(s))(1-H(t))} - \frac{H_*^1(s)-H^1(s)}{(1-H(s))^2} \\
&\quad - \frac{H_*^1(s)-H^1(s)}{1-H(s)} \int_s^t \frac{1}{(1-H(x))^2} H(\mathrm{d}x) \\
&= 0 \\
(b_4)+(b_5)+(b_6) &= \left(\frac{1}{1-H(s)} - \frac{1}{1-H(t)}\right) \int_0^s \frac{H_*^1(x)-H^1(x)}{(1-H(x))^2} H(\mathrm{d}x) \\
&\quad + \int_0^s \int_s^t \frac{H_*^1(x)-H^1(x)}{(1-H(x))^2(1-H(y))^2} H(\mathrm{d}y)H(\mathrm{d}x) \\
&= 0
\end{aligned}
$$

and finally

$$
\mathrm{Cov}(S(s), S(t)-S(s)) = \int_0^s \int_s^t \frac{\alpha(x,y)}{(1-H(x))(1-H(y))} H(\mathrm{d}y)H(\mathrm{d}x) \qquad (15)
$$

References

Abdushukurov, A.A., 1987. Nonparametric estimation in the proportional hazards model of random censorship. Akad. Nauk Uz Tashkent. VINITI No. 3448-V (in Russian).

Beirlant, J., Carbonez, A., van der Meulen, E., 1992. Long run proportional hazards models of random censorship. J. Statist. Plann. Inference 32, 25–44.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Breslow, N., Crowley, J., 1974. A large sample study of the life table and product-limit estimates under random censorship. Ann. Statist. 2, 437–453.

Cheng, P.E., Lin, G.D., 1987. Maximum likelihood estimation of a survival function under the Koziol–Green proportional hazards model. Statist. Probab. Lett. 5, 75–80.

Cox, D.R., Snell, E.J., 1989. Analysis of Binary Data, 2nd ed. Chapman & Hall, London.

Csörgő, S., 1988a. Testing for the proportional hazards model of random censorship. In: Mandl, P., Hušková, M. (Eds.), Proc. 4th Prague Symp. Asympt. Statist. Charles Univ. Prague.

Csörgő, S., 1988b. Estimation in the proportional hazards model of random censorship. Statistics, 19, 437–463.

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457–481.

Koziol, J.A., Green, S.B., 1976. A Cramér-von Mises statistic for randomly censored data. Biometrika 63, 465–474.

Perlman, M.D., 1972. On the strong consistency of approximate maximum likelihood estimates. Proc. Sixth Berk. Symp. Math. Statist. Probab., vol. 1, University of California Press, Berkeley, CA, pp. 263–281.

Rao, C.R., 1973. Linear statistical inference and its applications, 2nd ed. Wiley, New York.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Stute, W., 1992. Strong consistency of the MLE under random censorship. Metrika 39, 257–267.

Stute, W., 1994. U-statistic processes: a martingale approach. Ann. Probab. 22, 1725–1744.

Stute, W., Wang, J.-L., 1993. The strong law under random censorship. Ann. Statist. 21, 1591–1607.

Witting, H., Müller-Funk, U., 1995. Mathematische Statistik II. B.G. Teubner, Stuttgart.