T E C H N I C A L
R E P O R T

0308

RELATIVE HAZARD RATE ESTIMATION FOR RIGHT CENSORED AND LEFT TRUNCATED DATA.

R. CAO, P. JANSSEN and N. VERAVERBEKE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Relative hazard rate estimation for right censored and left truncated data

Ricardo Cao[1]
Department of Mathematics, Universidade da Coruña, E-15071 A Coruña, Spain,
e-mail: rcao@udc.es

Paul Janssen[2] and Noël Veraverbeke[2]
Center for Statistics, Limburgs Universitair Centrum, B-3590 Diepenbeek,
Belgium, e-mail: paul.janssen@luc.ac.be, noel.veraverbeke@luc.ac.be

Abstract

The concept of relative hazard rate is introduced in a two sample problem. Some kernel estimator is proposed in the case where both samples are subject to left truncation and right censoring and an iid representation is obtained in this setup. As a consequence the asymptotic distribution of the estimator and the asymptotic mean squared error are found. An application to the famous Channing House data set illustrates the theory.

Keywords: asymptotic representation, kernel estimator, survival analysis, two-sample comparison.

Running heads: Relative hazard for censored and truncated data

1. Introduction

An important problem in survival analysis is to compare the lifetime of two populations (a control group and a treatment group) when no parametric assumptions are plausible for these populations. In the nonparametric two sample problem there exists a vast literature about either constructing test statistics to evaluate the significance of the observed differences or to construct some informative curves that

[1]Research supported in part by MCyT Grants PB98-0182-C02-01 and BFM2002-00265 (European FEDER support included) and XUGA Grant PGIDT00PXI10501PN.

[2]The last two authors acknowledge partial support by the Ministry of the Flemish Community (Project BIL00/28, International Scientific and Technological Cooperation) and by the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs).

provide a graphical idea of possible differences between the two lifetime distributions. Such graphical procedures go back to the well known Q-Q plots, P-P plots (see [4], [13] [9] and [12] among others), or, more recently, to relative distributions and relative densities (see [8] and [3] for instance). In the censored data case [6] and [1, 2] deal with the problems of nonparametric estimation of the relative distribution and the relative density, as well as bandwidth selection.
In survival analysis, however, exploring the risk of death is the key issue. Therefore the cumulative hazard and the hazard rate are the more suitable functions to study. In this context it seems very reasonable to compare two populations by means of the relative hazard rate of Y with respect to Y_0. The precise definition is as follows. Let F, Λ and λ (F_0, Λ_0 and λ_0) denote the distribution function (cdf), the cumulative hazard and the hazard rate of Y (Y_0). The relative cumulative hazard of Y with respect to Y_0 is Λ_R, the cumulative hazard of $\Lambda_0(Y)$.
We denote the corresponding hazard rate as λ_R. In this paper, we propose a nonparametric kernel estimator for λ_R for left truncated and right censored data: (Y, T, C) denotes a random vector; Y is the lifetime (the random variable of interest) with cdf F, T is a random left truncation time with cdf L and C is a random right censoring time with cdf G. Similar definitions hold for (Y_0, T_0, C_0). For simplicity we assume that the cdf's are continuous and that the random variables describing the lifetime, the censoring time and the truncation time are independent. In this model, we observe (T, Z, δ) if $T \leq Z$ where $Z = Y \wedge C$ and $\delta = \mathbf{1}\{Y \leq C\}$, with H the cdf of Z we have $1 - H = (1 - F)(1 - G)$. When $T > Z$ nothing is observed. The data consist of two left truncated right censored samples:

$$(T_1, Z_1, \delta_1), \ldots, (T_m, Z_m, \delta_m)$$

and

$$(T_{01}, Z_{01}, \delta_{01}), \ldots, (T_{0n}, Z_{0n}, \delta_{0n})$$

where the observed sample sizes m and n are random and the real sample sizes (say M and N) are unknown. With $\alpha = P(T \leq Z)$ and $\alpha_0 = P_0(T_0 \leq Z_0)$ (the probability of absence of truncation in the target and in the reference population) we have, from the SLLN, that $m/M \to \alpha$ and $n/N \to \alpha_0$.

From $P(\Lambda_0(Y) \leq t) = F(\Lambda_0^{-1}(t))$ we have that

$$\Lambda_R(t) = -\ln(1 - F(\Lambda_0^{-1}(t))) = \Lambda(\Lambda_0^{-1}(t)).$$

The relative hazard rate is defined as the derivate of the relative cumulative hazard, i.e.,

$$\lambda_R(t) = \Lambda_R^{'}(t) = \frac{\lambda(\Lambda_0^{-1}(t))}{\lambda_0(\Lambda_0^{-1}(t))}.$$

The relative hazard rate λ_R is a useful (graphical) tool for comparing the target distribution F (say the treatment distribution) with a reference distribition F_0 (say the control distribution). Under the hypothesis that F equals F_0 we have $\Lambda_R(t) \equiv t$ and $\lambda_R(t) \equiv 1$. Deviations from this line therefore provide visual information on how population Y differs from population Y_0.

If F is continuous and defining $a_F = \inf\{v : F(v) > 0\}$, we have

$$\Lambda(t) = \int_{a_F}^{t} \frac{dF(y)}{1 - F(y)} = \int_{a_F}^{t} \frac{dH^u(y)}{C(y)} \tag{1}$$

where

$$\begin{aligned} H^u(t) &= P(Z \leq t, \delta = 1 \mid T \leq Z) \\ &= \alpha^{-1} P(Y \leq t, T \leq Y \leq C) \\ &= \int_{a_F}^{t} \alpha^{-1} P(T \leq y \leq C) dF(y) \\ &= \int_{a_F}^{t} \alpha^{-1} L(y)(1 - G(y)) dF(y) \end{aligned} \tag{2}$$

and

$$\begin{aligned} C(t) &= P(T \leq t \leq Z \mid T \leq Z) \\ &= \alpha^{-1} P(T \leq t \leq Z) \\ &= \alpha^{-1} L(t)(1 - F(t))(1 - G(t)). \end{aligned} \tag{3}$$

Empirical versions of (2) and (3) are

$$H_m^u(t) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}\{Z_i \leq t, \delta_i = 1\}$$

and

$$C_m(t) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}\{T_i \leq t \leq Z_i\}.$$

Plugging in the empirical versions in (1) we obtain the TJW consistent estimator (see [14]):

$$\widehat{\Lambda}(t) = \sum_{i=1}^{m} \frac{\mathbf{1}\{Z_i \le t, \delta_i = 1\}}{mC_m(Z_i)}. \tag{4}$$

Some useful asymptotic representation for this estimator has been recently obtained in [15]. In a similar way, we can define $\widehat{\Lambda}_0(t)$ which is consistent for $\Lambda_0(t)$.

1.1. An estimator for the relative cumulative hazard

If F_0 (or Λ_0) would be known, we could easily obtain from (4) an estimator for Λ_R:

$$\widetilde{\Lambda}_m(t) = \sum_{i=1}^{m} \frac{\mathbf{1}\{\Lambda_0(Z_i) \le t, \delta_i = 1\}}{\sum_{j=1}^{m} \mathbf{1}\{\Lambda_0(T_j) \le \Lambda_0(Z_i) \le \Lambda_0(Z_j)\}}. \tag{5}$$

Observe that $\widetilde{\Lambda}_m$ is nothing but (4) based on the *pseudovalues* $(\Lambda_0(T_i), \Lambda_0(Z_i), \delta_i)$, $i = 1, \ldots, m$. Actually, F_0 is unknown and we replace the pseudovalues by the relative data $(\widehat{\Lambda}_0(T_i), \widehat{\Lambda}_0(Z_i), \delta_i)$, $i = 1, \ldots, m$, to obtain

$$\widehat{\Lambda}_R(t) = \sum_{i=1}^{m} \frac{\mathbf{1}\{\widehat{\Lambda}_0(Z_i) \le t, \delta_i = 1\}}{\sum_{j=1}^{m} \mathbf{1}\{\widehat{\Lambda}_0(T_j) \le \widehat{\Lambda}_0(Z_i) \le \widehat{\Lambda}_0(Z_j)\}}.$$

1.2. An estimator for the relative hazard rate

Let us consider a kernel function, K, (a known density function) and $h = h_m$, a nonnegative bandwidth sequence. Since

$$\frac{1}{h}\int K\Big(\frac{t-v}{h}\Big)d\Lambda_R(v) = \frac{1}{h}\int K\Big(\frac{t-\Lambda_0(y)}{h}\Big)d\Lambda(y)$$

is close to $\lambda_R(t)$, a natural estimator is

$$\widehat{\lambda}_R(t) = \frac{1}{h}\int K\Big(\frac{t-\widehat{\Lambda}_0(y)}{h}\Big)d\widehat{\Lambda}(y). \tag{6}$$

In this paper, we will focus on the asymptotic distributional behaviour of this relative hazard rate estimator. In Section 2 we will collect some useful asymptotic representations. They are needed in Section 3 to obtain an iid representation for

$\widehat{\lambda}_R(t)$ that will be the key to obtain the limit distribution of $\lambda_R(t)$ and to study its mean squared error. An application to the well known Channing House data set is given in Section 4. Finally, Section 5 contains the proofs of the results stated in Section 3.

2. Some useful representations

As notation we use, with W an arbitrary cdf, $a_W = \inf\{t : W(t) > 0\}$ and $b_W = \sup\{t : W(t) < 1\}$. From now on, for conciseness, we also skip in our notation the conditioning event $T \le Z$ when writting expectations and probabilities.
Asymptotic representations for $\widehat{\Lambda}(t)$ - $\Lambda(t)$ have been studied in [7] and [15]. The following important lemma is Theorem 2.1 in [15].

Lemma 1. Assume $a_L \le a_H$ and for some $b < b_H$

$$\int_{a_H}^{b} \frac{dH^u(t)}{C^3(t)} < \infty.$$

Then, we have, uniformly in $a_H \le t \le b$,

$$\widehat{\Lambda}(t) - \Lambda(t) = \widehat{L}(t) + \widehat{R}(t)$$

where $\widehat{L}(t) = m^{-1} \sum_{i=1}^{m} \widehat{\xi}_i(t)$ with

$$\widehat{\xi}_i(t) = \frac{\mathbf{1}\{Z_i \le t, \delta_i = 1\}}{C(Z_i)} - \int_{a_H}^{t} \frac{\mathbf{1}\{T_i \le v \le Z_i\}}{C^2(v)} dH^u(v)$$

and

$$\sup_{a_H \le t \le b} |\widehat{R}(t)| = O\big(\frac{\log\log m}{m}\big) \qquad a.s. \qquad \blacksquare$$

To obtain the appropriate appoximation for the relative hazard rate estimator given by (6), we need asymptotic representations for the TJW product limit estimator based on the pseudosample $(\Lambda_0(T_i), \Lambda_0(Z_i), \delta_i), i = 1, \ldots, m$; and for the TJW product limit estimator based on the pseudosample $(\Lambda_0(T_{0i}), \Lambda_0(Z_{0i}), \delta_{0i}), i = 1, \ldots, n$. We denote these estimators as $\widetilde{\Lambda}_m(t)$ (see also (5)) and $\widetilde{\Lambda}_{0n}(t)$. Further note that $\widetilde{\Lambda}_m(t) = \widehat{\Lambda}(\Lambda_0^{-1}(t))$ estimates $\widetilde{\Lambda}(t) \equiv \Lambda_R(t) = \Lambda(\Lambda_0^{-1}(t))$ and that

$\widetilde{\Lambda}_{0n}(t) = \widehat{\Lambda}_0(\Lambda_0^{-1}(t))$ estimates $\Lambda_0(\Lambda_0^{-1}(t)) \equiv t$.

For the pseudodata $(\Lambda_0(T_i), \Lambda_0(Z_i), \delta_i), i = 1, \ldots, n$, we need:

$$\widetilde{H}(t) = P(\Lambda_0(Z_i) \leq t) = H(\Lambda_0^{-1}(t))$$

$$\widetilde{C}(t) = P\big(\Lambda_0(T_i) \leq t \leq \Lambda_0(Z_i) | \Lambda_0(T_i) \leq \Lambda_0(Z_i)\big) = C(\Lambda_0^{-1}(t))$$

$$\widetilde{H}^u(t) = P\big(\Lambda_0(Z_i) \leq t, \delta_i = 1 | \Lambda_0(T_i) \leq \Lambda_0(Z_i)\big) = H^u(\Lambda_0^{-1}(t))$$

Note that $a_{\widetilde{H}} = a_{H \circ \Lambda_0^{-1}}$ and $b_{\widetilde{H}} = b_{H \circ \Lambda_0^{-1}}$; and let $\widetilde{L}$ be the cdf of $\Lambda_0(T_i)$.

The following two lemmas are a straightforward consequence of Lemma 1.

Lemma 2. Assume $a_{\widetilde{L}} \leq a_{\widetilde{H}}$ and for some $\widetilde{b} < b_{\widetilde{H}}$

$$\int_{a_{\widetilde{H}}}^{\widetilde{b}} \frac{d\widetilde{H}^u(t)}{\widetilde{C}^3(t)} \equiv \int_{a_H}^{\Lambda_0^{-1}(\widetilde{b})} \frac{dH^u(t)}{C^3(t)} < \infty.$$

Then we have, uniformly in $a_{\widetilde{H}} \leq t \leq \widetilde{b}$,

$$\widetilde{\Lambda}_m(t) - \Lambda_R(t) = \widetilde{L}_m(t) + \widetilde{R}_m(t)$$

where $\widetilde{L}_m(t) = m^{-1} \sum_{i=1}^{m} \widetilde{\xi}_i(t)$ with $\widetilde{\xi}_i(t) = \widehat{\xi}_i(\Lambda_0^{-1}(t))$ and

$$\sup_{a_{\widetilde{H}} \leq t \leq \widetilde{b}} |\widetilde{R}_m(t)| = O\Big(\frac{\log \log m}{m}\Big) \qquad a.s.$$ ■

For the pseudodata $(\Lambda_0(T_{0i}), \Lambda_0(Z_{0i}), \delta_{0i}), i = 1, \ldots, n$, we use the parallel notation $\widetilde{H}_0(t) = H_0(\Lambda_0^{-1}(t)), \widetilde{C}_0(t) = C_0(\Lambda_0^{-1}(t)), \widetilde{H}_0^u(t) = H_0^u(\Lambda_0^{-1}(t)), a_{\widetilde{H}_0} = a_{H_0 \circ \Lambda_0^{-1}}, b_{\widetilde{H}_0} = b_{H_0 \circ \Lambda_0^{-1}}$; and $\widetilde{L}_0$ denotes the cdf of $\Lambda_0(T_{0i})$.

Lemma 3. Assume $a_{\widetilde{L}_0} \leq a_{\widetilde{H}_0}$ and for some $\widetilde{b}_0 < b_{\widetilde{H}_0}$

$$\int_{a_{\widetilde{H}_0}}^{\widetilde{b}_0} \frac{d\widetilde{H}_0^u(t)}{\widetilde{C}_0^3(t)} \equiv \int_{a_{H_0}}^{\Lambda_0^{-1}(\widetilde{b}_0)} \frac{dH_0^u(t)}{C_0^3(t)} < \infty.$$

Then we have, uniformly in $a_{\widetilde{H}_0} \le t \le \widetilde{b}_0$

$$\widetilde{\Lambda}_{0n}(t) - t = \widetilde{L}_{0n}(t) + \widetilde{R}_{0n}(t)$$

where $\widetilde{L}_{0n}(t) = n^{-1} \sum_{i=1}^{n} \widetilde{\xi}_{0i}(t)$ with

$$\widetilde{\xi}_{0i}(t) = \frac{\mathbf{1}\{Z_{0i} \le \Lambda_0^{-1}(t), \delta_{0i} = 1\}}{C_0(Z_{0i})} - \int_{a_{H_0}}^{\Lambda_0^{-1}(t)} \frac{\mathbf{1}\{T_{0i} \le v \le Z_{0i}\}}{C_0^2(v)} dH_0^u(v)$$

and

$$\sup_{a_{\widetilde{H}_0} \le t \le \widetilde{b}_0} |\widetilde{R}_{0n}(t)| = O\Big(\frac{\log \log n}{n}\Big) \qquad a.s. \qquad \blacksquare$$

3. A representation for the relative hazard rate estimator

Proceeding as in [8] we have, by a Taylor expansion,

$$\widehat{\lambda}_R(t) = A_m(t) + B_{n,m}(t) + R_{n,m}(t) \tag{7}$$

where

$$A_m(t) = \frac{1}{h} \int K\Big(\frac{t - \Lambda_0(y)}{h}\Big) d\widehat{\Lambda}(y) \tag{8}$$

$$B_{n,m}(t) = \frac{1}{h^2} \int \big(\Lambda_0(y) - \widehat{\Lambda}_0(y)\big) K'\Big(\frac{t - \Lambda_0(y)}{h}\Big) d\widehat{\Lambda}(y) \tag{9}$$

$$R_{n,m}(t) = \frac{1}{2h^3} \int \big(\Lambda_0(y) - \widehat{\Lambda}_0(y)\big)^2 K''\big(\Delta_{ty}\big) d\widehat{\Lambda}(y) \tag{10}$$

with Δ_{ty} a value between $\big(t - \Lambda_0(y)\big)/h$ and $\big(t - \widehat{\Lambda}_0(y)\big)/h$.

For further discussion we require some conditions. We first define some functions that will appear in the moment calculations of the terms in the iid representation (see the proof of Corollary 1): $\Psi_0(s) = \varphi_0(\Lambda_0^{-1}(s))$, with

$$\varphi_0(s) = \int_{a_{H_0}}^{s} \frac{dH_0^u(y)}{C_0^2(y)} = \int_{a_{H_0}}^{s} \frac{\lambda_0(y)}{C_0^2(y)} dy$$

and $\Psi(s) = \varphi(\Lambda_0^{-1}(s))$, with

$$\varphi(s) = \int_{a_H}^{s} \frac{dH^u(y)}{C^2(y)} = \int_{a_H}^{s} \frac{\lambda(y)}{C^2(y)} dy.$$

Let us now state the assumptions that will be used along the paper.

(A1) K is a twice differentiable density function on $[-1, 1]$ with K'' bounded.
(A2) As the samples sizes tend to infinity, $\lim_{n\to\infty} \frac{m}{n} = \kappa^2$, with $0 < \kappa^2 < \infty$.
(A3) The bandwidth h satisfies $h \to 0$ and $nh^3 \to \infty$.
(A4) The endpoints of the supports satisfy $a_L < a_{H_0} < b_H < b_{H_0}$, $a_{\widetilde{L}} \leq a_{\widetilde{H}}$ and $a_{\widetilde{L}_0} \leq a_{\widetilde{H}_0}$.
(A5) There exist some $\widetilde{b} < b_{\widetilde{H}}$ and $\widetilde{b}_0 < b_{\widetilde{H}_0}$ such that

$$\int_{a_{\widetilde{H}}}^{\widetilde{b}} \frac{d\widetilde{H}^u(s)}{\widetilde{C}^3(s)} < \infty$$

and

$$\int_{a_{\widetilde{H}_0}}^{\widetilde{b}_0} \frac{d\widetilde{H}_0^u(s)}{\widetilde{C}_0^3(s)} < \infty.$$

(A6) The relative hazard function, Λ_R, is Lipschitz continuous, with Lipschitz constant L_{Λ_R}, in a neighbourhood of t.
(A7) The relative hazard rate, λ_R, is twice continuously differentiable at t.
(A8) The function $F \circ \Lambda_0^{-1}$ is Lipschitz continuous in a neighbourhood of t.
(A9) Ψ_0 is differentiable in a neighbourhood of t, with Ψ_0' continuous at t.
(A10) The functions Ψ_0 and Ψ are twice continuously differentiable at t.

It is worth mentioning that part of the constraints in (A4) come from identifiability conditions to be imposed to the left truncation scheme. The inequalities $b_H < b_{H_0}$ is a technical condition needed to avoid problems with the right tails of the cumulative hazard function estimator. It essentially requires that one should choose, between the two populations, the one with larger right endpoint to be the baseline.
Recall the term $A_m(t)$ in (8). A standard change of variable gives

$$A_m(t) = \frac{1}{h} \int K\Big(\frac{t-v}{h}\Big) d\widetilde{\Lambda}_m(v).$$

Now apply Lemma 2 to obtain:

$$A_m(t) = A_m^{(1)}(t) + A_m^{(2)}(t) + O\Big(\frac{\log\log m}{m}\Big) \qquad a.s. \tag{11}$$

where

$$A_m^{(1)}(t) = \int K(u)\lambda_R(t-hu)du$$

$$A_m^{(2)}(t) = \frac{1}{mh}\sum_{i=1}^{m}\int \widetilde{\xi}_i(t-hu)K'(u)du. \tag{12}$$

We now consider $B_{n,m}(t)$ defined in (9),

$$\begin{aligned} B_{n,m}(t) &= \frac{1}{h^2}\int \big(v-\widetilde{\Lambda}_{0n}(v)\big)K'\Big(\frac{t-v}{h}\Big)d\widetilde{\Lambda}_m(v) \\ &= B_{n,m}^{(1)}(t) + B_{n,m}^{(2)}(t) \end{aligned} \tag{13}$$

where

$$B_{n,m}^{(1)}(t) = \frac{1}{h^2}\int \big(v-\widetilde{\Lambda}_{0n}(v)\big)K'\Big(\frac{t-v}{h}\Big)d\widetilde{\Lambda}(v)$$

$$B_{n,m}^{(2)}(t) = \frac{1}{h^2}\int \big(v-\widetilde{\Lambda}_{0n}(v)\big)K'\Big(\frac{t-v}{h}\Big)d\Big(\widetilde{\Lambda}_m(v)-\widetilde{\Lambda}(v)\Big).$$

An application of Lemma 3 leads to the further decomposition:

$$B_{n,m}^{(1)}(t) = B_{n,m}^{(1.1)}(t) + B_{n,m}^{(1.2)}(t) \tag{14}$$

where

$$\begin{aligned} B_{n,m}^{(1.1)}(t) &= -\frac{1}{nh^2}\sum_{i=1}^{n}\int \widetilde{\xi}_{0i}(v)K'\Big(\frac{t-v}{h}\Big)\lambda_R(v)dv \\ &= -\frac{1}{nh}\sum_{i=1}^{n}\int \widetilde{\xi}_{0i}(t-uh)K'(u)\lambda_R(t-hu)du \end{aligned}$$

$$B_{n,m}^{(1.2)}(t) = -\frac{1}{h^2}\int \widetilde{R}_{0n}(v)K'\Big(\frac{t-v}{h}\Big)\lambda_R(v)dv$$

In Section 5 we will show that

$$B_{n,m}^{(1.2)}(t) = O\big(\frac{\log\log n}{n}\big) \qquad a.s. \tag{15}$$

and

$$B_{n,m}^{(2)}(t) = O\Bigg(\frac{\log\log n}{nh^2}\Big[h + \Big\{\Big(\frac{h\log m}{m}\Big)^{\frac{1}{2}} \vee \frac{(\log\log m)^{\frac{1}{2}}}{m}\Big\}\Big]\Bigg) \qquad a.s. \tag{16}$$

We further show in Section 5 that

$$R_{n,m}(t) = o_P\big((mh)^{-1/2}\big). \tag{17}$$

Based on the previous discussion we obtain the following theorem.

Theorem 1. Assume conditions (A1)-(A6) and (A8)-(A9) and let $t < \min\{\widetilde{b}_0, \widetilde{b}\}$, then

$$\widehat{\lambda}_R(t) = A_m^{(1)}(t) + A_m^{(2)}(t) + B_{n,m}^{(1.1)}(t) + C_{n,m}(t),$$

with $C_{n,m}(t) = o_P\big((mh)^{-1/2}\big)$ and $E(|C_{n,m}(t)|^d) = o((mh)^{-\frac{d}{2}})$ for any $d > 0$.

Corollary 1. Assume conditions (A1)-(A10) and let $t < \min\{\widetilde{b}_0, \widetilde{b}\}$, then

$$E[\widehat{\lambda}_R(t)] = \lambda_R(t) + \frac{1}{2}\lambda_R''(t)D_K h^2 + o(h^2) + o\big((mh)^{-1/2}\big)$$

$$\mathrm{Var}\big(\widehat{\lambda}_R(t)\big) = \frac{\sigma^2}{mh} + o\big(\frac{1}{mh}\big)$$

where

$$\sigma^2 = C_K\Big\{\frac{\lambda_R(t)}{C\big(\Lambda_0^{-1}(t)\big)} + \kappa^2\frac{\lambda_R^2(t)}{C_0\big(\Lambda_0^{-1}(t)\big)}\Big\}$$

with $C_K = \int K^2(u)du$ and $D_K = \int u^2K(u)du$.

Remark. As a straightforward consequence of Corollary 1, an asymptotic formula for the mean squared error of the estimator is

$$MSE\left(\widehat{\lambda}_R(t)\right) = AMSE\left(\widehat{\lambda}_R(t)\right) + o(h^4) + o\big(\frac{1}{mh}\big),$$

with

$$AMSE\left(\widehat{\lambda}_R(t)\right) = \frac{1}{4}\lambda''_R(t)^2 D_K^2 h^4 + \frac{\sigma^2}{mh}.$$

The smoothing parameter that minimizes this criterion is

$$h_{AMISE} = \frac{\sigma^2}{\lambda''_R(t)^2 D_K^2} m^{-\frac{1}{5}}.$$

Observe that plug-in bandwidth selectors for h_{AMISE} would require the estimation of many underlying functions. Smoothing techniques are required to estimate some of these functions.

The following result gives the limit distribution of the estimator either using some asymptotically undersmoothing bandwidth (that kills the bias) or with the asymptotical optimal rate for the smoothing parameter.

Corollary 2. Assume the conditions of Corollary 1 and let $t < \min\{\widetilde{b}_0, \widetilde{b}\}$.
If $mh^5 \to 0$, then

$$(mh)^{\frac{1}{2}}\{\widehat{\lambda}_R(t) - \lambda_R(t)\} \xrightarrow{d} N(0, \sigma^2).$$

If $mh^5 \to c$, for some $c > 0$, then

$$(mh)^{\frac{1}{2}}\{\widehat{\lambda}_R(t) - \lambda_R(t)\} \xrightarrow{d} N(\frac{1}{2}\lambda''_R(t) D_K c^{1/2}, \sigma^2).$$

4. Example: Channing House data

We have applied the kernel relative hazard estimator to the well known Channing House data set (see [10] or [11]). Channing House is a retirement centre in Palo Alto, California. The data were collected between the opening of the house in 1964 and July 1, 1975. In that time 97 men and 365 women passed through the centre. For each of these, their age on entry and also on leaving or death was recorded. A large number of the observations were censored mainly due to the resident being alive on July 1, 1975, the end of the study. An individual must survive to a sufficient age, T_i, to enter the retirement center and all individuals who died prior to entering Channing House are not included in the study. Therefore the lifetimes in the study are left truncated. Over the time of the study 130 women and 46 men died at Channing House. Differences between the survival of the sexes was one of the primary concerns of that study.

The kernel relative hazard function estimation, given in (6), for the lifetime of men with respect to women was computed. The Gaussian kernel was used to compute the estimator with different bandwidths to cover differents amount of smoothing. These estimators (for $h = 0.2, 0.3, 0.4$) are plotted in Figure 1(a). Figures 1(b)-1(d) contain the kernel relative hazard estimator for each of these three bandwidths with some 95% pointwise confidence intervals for certain selected points. These confidence limits have been computed using some estimators of the bias and variance appearing in Corollary 1. Empirical estimators are used to estimate the functions C_0, C and Λ_0 in σ^2. The function λ_R appearing in σ^2 is estimated by means of (6) with the bandwidth already used for the estimation itself. The term λ_R'' in the bias has been estimated using the second derivative of the estimator in (6). The bandwidth, g, for such an estimator has been selected as twice the original bandwidth, h. It is already known in classical settings that the optimal smoothing parameter for the second derivative estimation should be asymptotically larger than that for the estimator itself.

Put Figures 1(a)-1(d) about here.

Looking at Figures 1(a)-1(d) it is clearly seen that the estimated relative risk of death for men with respect to women is between 2 and 2.5 for $t = 0.25$. This means that men have slightly more than twice the risk of death than women with a cumulative risk of 0.25, that corresponds to an age of about 77 years for the women population. Moreover, the confidence limits at $t = 0.25$ (no matter which of three bandwidths is used) is always above 1, meaning that this higher risk of death for men is statistically significant.

Figures 1(b)-1(d) also show that the risk of death for both populations is not significantly different for $t \in [0.5, 1.8]$. This interval of cumulative hazard for women corresponds to an age interval, in years, of $[83, 92]$.

5. Proofs

5.1 Proof of (15)

$$\begin{aligned}
|B^{(1.2)}_{n,m}(t)| &\le \frac{1}{h^2}\|K'\|_\infty \sup_{t-h\le v\le t+h}|\widetilde{R}_{0n}(v)|\int_{t-h}^{t+h}\lambda_R(v)dv \\
&= \frac{1}{h^2}\|K'\|_\infty \sup_{t-h\le v\le t+h}|\widetilde{R}_{0n}(v)|[\Lambda_R(t+h)-\Lambda_R(t-h)] \\
&= O\big(\frac{\log\log n}{nh}\big) \qquad a.s.
\end{aligned}$$

where the order relation follows from the Lipschitz continuity of Λ_R and Lemma 3.

5.2 Proof of (16)

From Lemma 3 we have

$$B^{(2)}_{n,m}(t) = B^{(2.1)}_{n,m}(t) + B^{(2.2)}_{n,m}(t)$$

with

$$B^{(2.1)}_{n,m}(t) = -\frac{1}{h^2}\int \widetilde{L}_{0n}(v)K'\Big(\frac{t-v}{h}\Big)d(\widetilde{\Lambda}_m(v)-\widetilde{\Lambda}(v))$$
$$B^{(2.2)}_{n,m}(t) = -\frac{1}{h^2}\int \widetilde{R}_{0n}(v)K'\Big(\frac{t-v}{h}\Big)d(\widetilde{\Lambda}_m(v)-\widetilde{\Lambda}(v))$$

To deal with $B^{(2.1)}_{n,m}(t)$ we will find an order bound for $E[(B^{(2.1)}_{n,m}(t))^2]$. Assume $u < v$ and recall Lemma 3 and the definition of Ψ_0 in Section 3.

$$E[(B^{(2.1)}_{n,m}(t))^2]$$

$$= \frac{2}{nh^4}\int_{t-h}^{t+h}\Big[\int_u^{t+h} K'\Big(\frac{t-v}{h}\Big)d(\widetilde{\Lambda}_m(v)-\widetilde{\Lambda}(v))\Big]\Psi_0(u)K'\Big(\frac{t-u}{h}\Big)d(\widetilde{\Lambda}_m(u)-\widetilde{\Lambda}(u)).$$

Rewriting the inner integral using integration by parts, we obtain

$$E[(B^{(2.1)}_{n,m}(t))^2] = \frac{2}{nh^4}[E(I_1)-E(I_2)]$$

where

$$I_1 = \int_{t-h}^{t+h} \Big[\int_{u}^{t+h} \frac{1}{h} K''\Big(\frac{t-v}{h}\Big)(\widetilde{\Lambda}_m(v) - \widetilde{\Lambda}(v))dv \Big] \Psi_0(u) K'\Big(\frac{t-u}{h}\Big) d(\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u))$$

$$I_2 = \int_{t-h}^{t+h} \Psi_0(u) \bigg(K'\Big(\frac{t-u}{h}\Big)\bigg)^2 (\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u)) d(\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u)).$$

For I_2 we use further integration by parts to obtain

$$\begin{aligned} I_2 &= \frac{1}{2} \int_{t-h}^{t+h} \Psi_0(u) \bigg(K'\Big(\frac{t-u}{h}\Big)\bigg)^2 d(\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u))^2 \\ &= \frac{1}{2} \int_{t-h}^{t+h} (\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u))^2 \Big[\Psi_0'(u) \bigg(K'\Big(\frac{t-u}{h}\Big)\bigg)^2 - \frac{2}{h} \Psi_0(u) K'\Big(\frac{t-u}{h}\Big) K''\Big(\frac{t-u}{h}\Big)\Big] du. \end{aligned}$$

Similar to the derivation in the appendix, we can derive an exponential inequality for

$$\Delta_m = \sup_{a_{\widetilde{F}_0} \le y \le b_{\widetilde{H}}} |\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u)|$$

from which we obtain

$$E(\Delta_m^2) = 2 \int_0^\infty u P(\Delta_m > u) du = O(m^{-1}). \tag{18}$$

Hence, using assumptions (A1) and (A9), we obtain that $E(I_2) = O(m^{-1})$.
For I_1 we also use integration by parts and find

$$\begin{aligned} I_1 &= - \int_{t-h}^{t+h} \bigg[- \frac{1}{h} K''\big(\frac{t-u}{h}\big)(\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u)) \Psi_0(u) K'\big(\frac{t-u}{h}\big) \\ &\quad + \Big\{ \int_{u}^{t+h} \frac{1}{h} K''\big(\frac{t-v}{h}\big)(\widetilde{\Lambda}_m(v) - \widetilde{\Lambda}(v)) dv \Big\} \\ &\quad \times \Big\{ \Psi_0'(u) K'\big(\frac{t-u}{h}\big) - \frac{1}{h} \Psi_0(u) K''\big(\frac{t-u}{h}\big) \Big\} \bigg] (\widetilde{\Lambda}_m(u) - \widetilde{\Lambda}(u)) du. \end{aligned}$$

Using Cauchy-Schwarz followed by conditions (A1), (A9) and equation (18), we get $E(I_1) = O(m^{-1})$.

From the previous discussion we can conclude

$$E[(B_{n,m}^{(2.1)}(t))^2] = O(m^{-2}h^{-4})$$

which implies $B_{n,m}^{(2.1)}(t) = O_P(m^{-1}h^{-2})$. Now using conditions (A2) and (A3), $(mh)^{1/2}B_{n,m}^{(2.1)}(t) = o_P(1)$.
The term $B_{n,m}^{(2.2)}(t)$ can be easily bounded:

$$\begin{aligned}
&|B_{n,m}^{(2.2)}(t)| \le \frac{1}{h^2}\|K'\|_\infty \sup_{t-h\le v\le t+h} |\widetilde{R}_{0n}(v)| \int_{t-h}^{t+h} |d(\widetilde{\Lambda}_m(v) - \widetilde{\Lambda}(v))| \\
&\le \frac{1}{h^2}\|K'\|_\infty \sup_{t-h\le v\le t+h} |\widetilde{R}_{0n}(v)| \Big\{ \big[\widetilde{\Lambda}_m(t+h) - \widetilde{\Lambda}_m(t-h)\big] + \big[\widetilde{\Lambda}(t+h) - \widetilde{\Lambda}(t-h)\big] \Big\} \\
&\le \frac{1}{h^2}\|K'\|_\infty \sup_{t-h\le v\le t+h} |\widetilde{R}_{0n}(v)| \Big\{ 2\big[\widetilde{\Lambda}(t+h) - \widetilde{\Lambda}(t-h)\big] + \big[\widetilde{\Lambda}_m(t+h) - \widetilde{\Lambda}_m(t-h)\big] \\
&-\big[\widetilde{\Lambda}(t+h) - \widetilde{\Lambda}(t-h)\big] \Big\}.
\end{aligned}$$

Using the Lipschitz-continuity of $\widetilde{\Lambda} \equiv \Lambda_R$ and the modulus of continuity result (that we develop in the proof of (17)) we have that

$$B_{n,m}^{(2.2)}(t) = O\Big(\frac{\log\log n}{nh^2}\Big\{h + \Big(\Big(\frac{h\log m}{m}\Big)^{\frac{1}{2}} \vee \frac{(\log\log m)^{\frac{1}{2}}}{m}\Big)\Big\}\Big) \qquad a.s.$$

5.3 Proof of (17)

$$\big|R_{n,m}(t)\big| \le \frac{1}{2h^3}\|K''\|_\infty \Delta_{0n}^2 \int \mathbf{1}\{|\Delta_{ty}| \le 1\} d\widehat{\Lambda}(y)$$

where $\Delta_{0n} = \sup_{a_{F_0}\le y\le b_H} \|\widehat{\Lambda}_0(y) - \Lambda_0(y)\|$. Note that in the definition of Δ_{0n} the value of y stays below b_H. Because $b_H < b_{H_0}, a_L \le a_{H_0}$ and

$$\int_{a_{H_0}}^{b_H} \frac{dH_0^u(y)}{C_0^3(y)} < \infty,$$

we have by Lemma 2.5 in [15] that $\Delta_{0n} = O\big(n^{-1/2}(\log\log n)^{1/2}\big)$ a.s. We further have that

$$\mathbf{1}\{|\Delta_{ty}| \le 1\} \le \mathbf{1}\{t - h - \Delta_{0n} \le \Lambda_0(y) \le t + h + \Delta_{0n}\}.$$

This implies

$$\begin{aligned}\left|R_{n,m}(t)\right| &\leq \frac{1}{2h^3}\|K''\|_\infty \Delta_{0n}^2\{\widehat{\Lambda}(\Lambda_0^{-1}(t+h+\Delta_{0n})) - \widehat{\Lambda}(\Lambda_0^{-1}(t-h-\Delta_{0n}))\}\\ &\leq R_{n,m}^{(1)}(t) + R_{n,m}^{(2)}(t)\end{aligned}$$

where

$$\begin{aligned}R_{n,m}^{(1)}(t) = &\frac{1}{2h^3}\|K''\|_\infty \Delta_{0n}^2\Big[\Big\{\widehat{\Lambda}(\Lambda_0^{-1}(t+h+\Delta_{0n})) - \Lambda(\Lambda_0^{-1}(t+h+\Delta_{0n}))\Big\}\\ &-\Big\{\widehat{\Lambda}(\Lambda_0^{-1}(t-h-\Delta_{0n})) - \Lambda(\Lambda_0^{-1}(t-h-\Delta_{0n}))\Big\}\Big]\end{aligned}$$

$$R_{n,m}^{(2)}(t) = \frac{1}{2h^3}\|K''\|_\infty \Delta_{0n}^2\Big\{\Lambda(\Lambda_0^{-1}(t+h+\Delta_{0n})) - \Lambda(\Lambda_0^{-1}(t-h-\Delta_{0n}))\Big\}.$$

The identity $\Lambda_R = \Lambda \circ \Lambda_0^{-1}$ and condition (A6) imply

$$\left|R_{n,m}^{(2)}(t)\right| \leq \frac{1}{h^3}\|K''\|_\infty \Delta_{0n}^2 L_{\Lambda_R}(h+\Delta_{0n}).$$

Using the order relation for Δ_{0n} and the condition $mh^3/(\log\log m)^2 \to \infty$, we have

$$\sqrt{mh}R_{n,m}^{(2)}(t) = o_P(1).$$

To handle the term $R_{n,m}^{(1)}(t)$, we use a Taylor expansion and apply Theorem 2.2 in [15] to obtain

$$R_{n,m}^{(1)}(t) = \frac{1}{2h^3}\|K''\|_\infty \Delta_{0n}^2\Big[\Big\{T_{n,m}^{(1)}(t) + T_{n,m}^{(2)}(t)\Big\} + O\big(\frac{\log\log m}{m}\big)\Big] \qquad a.s.$$

where

$$T_{n,m}^{(1)}(t) = \frac{\Gamma_m(t+h+\Delta_{0n}) - \Gamma_m(t-h-\Delta_{0n})}{1-F(\Lambda_0^{-1}(t+h+\Delta_{0n}))}$$

$$T_{n,m}^{(2)}(t) = \Gamma_m(t-h-\Delta_{0n})\left\{\frac{1}{1-F(\Lambda_0^{-1}(t+h+\Delta_{0n}))} - \frac{1}{1-F(\Lambda_0^{-1}(t-h-\Delta_{0n}))}\right\},$$

with $\Gamma_m(v) = \widehat{F}(\Lambda_0^{-1}(v)) - F(\Lambda_0^{-1}(v))$.

Condition (A8) implies

$$\sqrt{mh}\frac{1}{h^3}\|K''\|_\infty \Delta_{0n}^2 T_{n,m}^{(2)}(t) = o_P(1).$$

To obtain an appropriate order bound for $T_{n,m}^{(1)}(t)$ we rely on the following inequality:

$$T_{n,m}^{(1)}(t) \le c\Big[\sup_{|u|\le 2Lh} |\Gamma_m(t) - \Gamma_m(t+u)| + \mathbf{1}\{\Delta_{0n} > h\}\Big], \tag{19}$$

with L the Lipschitz constant for $F \circ \Lambda_0^{-1}$ and c some positive constant, in combination with Lemma 3.1 in [15] and an exponential probability inequality for Δ_{0n}.

Lemma 3.1 in [15] yields

$$\sup_{|u|\le 2Lh} |\Gamma_m(t) - \Gamma_m(t+u)| = O\left(\frac{(\log\log m)^{1/2}}{m} \vee \left(\frac{h\log m}{m}\right)^{1/2}\right) \qquad a.s.$$

Apply the DKW-inequality (see the appendix) to obtain

$$\mathbf{1}\{\Delta_{0n} > h\} = O_P\big(e^{-nch^2}\big),$$

for some constant $c > 0$. Using (19), the orders obtained for the two terms in the right hand side of that expression and condition (A3), we obtain

$$\sqrt{mh}\frac{1}{2h^3}\|K''\|_\infty \Delta_{0n}^2 T_{n,m}^{(1)}(t) = o_P(1).$$

As a final conclusion, we have

$$\sqrt{mh}R_{n,m}^{(1)}(t) = o_P(1).$$

5.4 Proof of Theorem 1

The proof simply consists of collecting expressions (7), (8), (9), (10), (11), (12), (13), (14), (15), (16) and (17). This gives the iid representation with the order in probability for the remainder term. To prove the moment order for $C_{n,m}(t)$ one may revise all the steps in the proof using equation (1.17) in Theorem 1 in [7].

■

5.5 Proof of Corollary 1

Using the moment order for the remainder term of the iid representation given in Theorem 1 and Cauchy-Schwarz inequality, the bias and the variance of the estimator results in

$$E[\widehat{\lambda}_R(t) - \lambda_R(t)] = \tfrac{1}{2}\lambda''_R(t) D_K h^2 + o(h^2) + o\big((mh)^{-1/2}\big).$$

$$\begin{aligned} &\mathrm{Var}\big(\widehat{\lambda}_R(t)\big) = \mathrm{Var}\big(A_m^{(2)}(t)\big) + \mathrm{Var}\big(B_{n,m}^{(1.1)}(t)\big) + o\big(\frac{1}{mh}\big) \\ &+ o\big(mh^{-\frac{1}{2}}\big(\mathrm{Var}\big(A_m^{(2)}(t)\big)\big)^{\frac{1}{2}}\big) + o\big(mh^{-\frac{1}{2}}\big(\mathrm{Var}\big(B_{n,m}^{(1.1)}(t)\big)\big)^{\frac{1}{2}}\big). \end{aligned} \tag{20}$$

The statement in Corollary 1 yields then on obtaining expressions for the variances of the terms in (20). Let us consider the first one,

$$\mathrm{Var}\big(A_m^{(2)}(t)\big) = \frac{2}{mh^2}\int_{-1}^{1}\int_{-1}^{v} E[\widetilde{\xi}_i(t-hu)\widetilde{\xi}_i(t-hv)]K'(u)K'(v)dudv.$$

For $u < v$ we have (see (1.9) in [7]) that

$$E[\widetilde{\xi}_i(t-hu)\widetilde{\xi}_i(t-hv)] = \Psi(t-hv),$$

where the function Ψ has been defined in Section 3.
Note that $\varphi'(s) = \dfrac{\lambda(s)}{C(s)}$ and $\Psi'(s) = \dfrac{\lambda_R(s)}{C(\Lambda_0^{-1}(s))}$. Therefore using (A10), with Δ_{1tv} a value between $t - hv$ and t,

$$\begin{aligned} \mathrm{Var}\big(A_m^{(2)}(t)\big) &= \frac{2}{mh^2}\int_{-1}^{1}\int_{-1}^{v} \Psi(t-hv)K'(u)K'(v)dudv \\ &= \frac{2}{mh^2}\int_{-1}^{1}\int_{-1}^{v} [\Psi(t) - hv\Psi'(t) + \frac{h^2v^2}{2}\Psi''(\Delta_{1tv})]K'(u)K'(v)dudv \\ &= \frac{C_K}{mh}\frac{\lambda_R(t)}{C\big(\Lambda_0^{-1}(t)\big)} + O\big(\frac{1}{m}\big), \end{aligned}$$

since

$$\int_{-1}^{1}\int_{-1}^{v} K'(u)K'(v)dudv = \frac{K^2(v)}{2}\Bigg|_{-1}^{1} = 0,$$

$$\int_{-1}^{1}\int_{-1}^{v} vK'(u)K'(v)dudv = -\frac{C_K}{2},$$

and

$$\left|\int_{-1}^{1}\int_{-1}^{v} v^2K'(u)K'(v)dudv\right| \leq \frac{1}{2}\|K'\|_\infty.$$

The variance of the second term in (20) is

$$\mathrm{Var}(B_{n,m}^{(1.1)}(t)) = \frac{2}{nh^2}\int_{-1}^{1}\int_{-1}^{v} E[\widetilde{\xi}_0(t-hu)\widetilde{\xi}_0(t-hv)]\lambda_R(t-hu)\lambda_R(t-hv)K'(u)K'(v)dudv.$$

Recall the definition of Ψ_0. Using again (1.9) in [7] we have for $u < v$ that

$$E[\widetilde{\xi}_0(t-hu)\widetilde{\xi}_0(t-hv)] = \Psi_0(t-hv).$$

Note that $\varphi_0'(s) = \dfrac{\lambda_0(s)}{C_0(s)}$ and $\Psi_0'(s) = \dfrac{1}{C_0(\Lambda_0^{-1}(s))}$.

Therefore, with $\Delta_{2tv}, \Delta_{3tu}$ and Δ_{4tv} intermediate points,

$$\begin{aligned}
\mathrm{Var}(B_{n,m}^{(1.1)}(t)) &= \frac{2}{nh^2}\int_{-1}^{1}\int_{-1}^{v}[\Psi_0(t) - hv\Psi_0'(t) + \frac{h^2v^2}{2}\Psi_0''(\Delta_{2tv})] \\
&\quad \times[\lambda_R(t) - hu\lambda_R'(t) + \frac{h^2u^2}{2}\lambda_R''(\Delta_{3tu})] \\
&\quad \times[\lambda_R(t) - hv\lambda_R'(t) + \frac{h^2u^2}{2}\lambda_R''(\Delta_{4tv})]K'(u)K'(v)dudv \\
&= \frac{C_K}{nh}\frac{\lambda_R^2(t)}{C_0(\Lambda_0^{-1}(t))} + O\Big(\frac{1}{n}\Big),
\end{aligned}$$

since

$$\int_{-1}^{1}\int_{-1}^{v} uK'(u)K'(v)dudv = \frac{C_K}{2}. \qquad \blacksquare$$

5.6 Proof of Corollary 2

It is a straightforward consequence of Theorem 1, Corollary 1 and the Central Limit Theorem for triangular arrays. Using Lyapunov condition, we only need to check that there exist some $c > 0$ such that

$$E\left(\left|\int \widetilde{\xi}_i(t-hu)K'(u)du\right|^{2+c}\right) < \infty \tag{21}$$

$$E\left(\left|\int \widetilde{\xi}_{0i}(t-hu)K'(u)du\right|^{2+c}\right) < \infty. \tag{22}$$

Standard arguments give some bound for the term in (22):

$$\left|\int \widetilde{\xi}_i(t-hu)K'(u)du\right| \leq \frac{\delta_{0i}}{\widetilde{C}_0(\Lambda_0(Z_{0i}))} + \int_{a_{\widetilde{H}_0}}^{\widetilde{b}_0} \frac{d\widetilde{H}_0^u(s)}{\widetilde{C}_0^2(s)}.$$

The last term (non random) can be directly bounded by a constant using (A5). On the other hand

$$E\left(\left|\frac{\delta_{0i}}{\widetilde{C}_0(\Lambda_0(Z_{0i}))}\right|^{2+c}\right) < \int_{a_{\widetilde{H}_0}}^{\widetilde{b}_0} \frac{d\widetilde{H}_0^u(s)}{\widetilde{C}_0^{2+c}(s)} < \infty,$$

using, once more, condition (A5). Now, the inequality $(a+b)^{2+c} \leq 2^{1+c}(a^{2+c}+b^{2+c})$ (valid for any pair of positive real numbers, a and b) and condition (A1) lead to (22). The proof of (21) is completely parallel to that of (22). For this reason we skip it. ■

Appendix: a Dvoretzky-Kiefer-Wolfowitz-type inequality for Δ_{0n}

Simple algebra gives

$$\begin{aligned} |\widehat{\Lambda}_0(y) - \Lambda_0(y)| \;\; \leq \;\; & \int_{a_{F_0}}^{y} \frac{|C_0(v) - C_{0n}(v)|}{C_{0n}(v)C_0(v)} dH_{0n}^u(v) + \frac{|H_{0n}^u(y) - H_0^u(y)|}{C_0(y)} \\ & + \int_{a_{F_0}}^{y} \frac{|H_{0n}^u(v) - H_0^u(v)|}{C_0^2(v)} d|C_0(v)|. \end{aligned}$$

Define $A_n = \left\{ \sup_{a_{F_0} \leq y \leq b_H} |C_{0n}(y) - C_0(y)| \leq \frac{d}{2} \right\}$ for some $d > 0$. On A_n we have $C_{0n}(y) \geq C_0(y) - \frac{d}{2}$.

Recall condition (A4) and take $d = \alpha_0^{-1} L_0(a_{H_0})(1 - H_0(b_H))$. We then have that $C_0(y) = \alpha_0^{-1} L_0(y)(1 - H_0(y)) \geq d$. From this discussion we arrive at the conclusion:

on A_n we have $C_{0n}(y) \geq d/2$. Now use the inequality above to obtain

$$P(\Delta_{0n} > h) \leq P\left(\frac{2}{d^2}C_{0n} > \frac{h}{3}\right) + P\left(\frac{1}{d}H_{0n} > \frac{h}{3}\right)$$
$$+P\left(\frac{1}{d^2}H_{0n}\int_{a_{F_0}}^{b_H}|dC_0(y)| > \frac{h}{3}\right) + P(A_n^c),$$

where

$$C_{0n} = \sup_{a_{F_0}\leq y\leq b_H}|C_{0n}(y) - C_0(y)|$$
$$H_{0n} = \sup_{a_{F_0}\leq y\leq b_H}|H_{0n}^u(y) - H_0^u(y)|.$$

Since H_0^u and H_{0n}^u are monotone we can apply the DKW-inequality to H_{0n} (see [5]). Following [7], p 224, we also have a DKW-inequality for C_{0n}. The function C_0 is of bounded variation. We therefore obtain

$$P(\Delta_{0n} > h) \leq C_1 \exp(-D_1 n h^2)$$

for some constants $C_1 > 0$ and $D_1 > 0$.

References

[1] Cao, R., Janssen, P., and Veraverbeke, N. (2000). Relative density esimation with censored data. *Canad. J. Statist.* **28** 97–111.

[2] Cao, R., Janssen, P., and Veraverbeke, N. (2001). Relative density esimation and local bandwith selection with censored data. *Comput. Statist. Data Anal.* **36** 497–510.

[3] Ćwik, J., Mielniczuk, J. (1993). Data-dependent bandwidth choice for a grade density kernel esimate. *Statist. Probab. Lett.* **16** 397–405.

[4] Doksum, K. A. and Sievers, G. L. (1976). Plotting with confidence: graphical comparisons of two populations. *Biometrika* **63** 421–434.

[5] Dvoretzky, A., Kiefer, J.C. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of a classical multinomial estimator. *Ann. Math. Statist.* **33** 642–669.

[6] Gastwirth, J.L. and Wang, J.-L. (1988). Control percentile test procedures for censored data. *J. Statist. Plann. Inference* **18** 267–276.

[7] Gijbels, I. and Wang, J.-L. (1993). Strong representations of the survival function estimator for truncated and censored data with applications. *J. Multivariate Anal.* **47** 210–229.

[8] Handcock, M. and Janssen, P. (2002). Statistical inference for the relative density. *Sociol Methods and Res.* **30** 394–424.

[9] Holmgren, E. B. (1995). The P-Pplot as a method of comparing treatment effects. *J. Amer. Statist. Assoc.* **90** *360–365.*

[10] Hyde, J. (1977). Testing survival under right censoring and left truncation. *Biometrika* **64** 225–230.

[11] Klein, J.P and Moeschberger, M.L. (1997) *Survival Analysis, techniques for censored and truncated survival data*. Springer, New York.

[12] Li, G., Tiwari, R. C. and Wells, M. T. (1996). Quantile comparison functions in two-sample problems, with application to comparisons of diagnostic markers. *J. Amer. Statist. Assoc.* **91** 689–698.

[13] Switzer, P. (1976). Confidence procedures for two-sample problems. *Biometrika* **63** 13–25.

[14] Tsai, W.Y., Jewell, N.P. and Wang, M.C. (1987). A note on the product limit estimator under right censoring and left truncation. *Biometrika* **74** 883–886.

[15] Zhou, Y. and Yip, P. (1999). A strong representation of the product-limit estimator for left truncated and right censored data. *J. Multivariate Anal.* **69** 261–280.