Smoothing the Nelson-Aalen Estimtor

Biostat 277 presentation – Chi-hong Tseng

Reference:
1. Andersen, Borgan, Gill, and Keiding (1993). *Statistical Model Based on Counting Processes,* Springer-Verlag, p.229-255
2. Muller and Wang (1994). Hazard Rate Estimation Under Random Cnsoring with Varying Kernels and Bandwidths, *Biometrics* 50, 61-76

1 Survival Data and Nelson-Aalen Estimator

1. Survival time T
2. Censoring time C
3. Observational time $X = \min(T, C)$
4. Censoring indicator $D = I(T \leq C)$
5. Hazard function: $\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr(t \leq T < t + \Delta t | T \geq t)$
6. Cumulative hazard function : $A(t) = \int_0^t \alpha(u) du$
7. Nelson-Aalen estimator $\hat{A}(t) = \sum_{0 \leq T_i \leq t} \frac{D_i}{n_i}$
8. n_i = number of subject at risk at time T_i
9. at-risk process $Y_i(t) = I(X_i \geq t)$, $Y(t) = \sum Y_i(t)$
10. counting process $N_i(t) = I(X_i \leq t, D_i = 1)$, $N(t) = \sum N_i(t)$
11. Nelson-Aalen estimator $\hat{A}(t) = \int_0^t \frac{J(u)}{Y(u)} dN(u) = \sum_{i=1}^n \int_0^t \frac{J(u)}{Y(u)} dN_i(u)$
12. $J(s) = I(Y(s) > 0)$

2 Kernel Function Estimator

The kernel function estimator is given by,

$$\begin{aligned} \hat{\alpha}(t) &= b^{-1} \int_{\mathcal{T}} K(\frac{t-s}{b}) d\hat{A}(s) \\ &= b^{-1} \sum_j K(\frac{t-T_j}{b}) (Y(T_j))^{-1} \end{aligned}$$

1. K: Kernel function, $\int_{-1}^1 K(s) ds = 1$
   e.g. $K(x) = 0.75(1-x^2)$, $|x| \leq 1$.
2. b: bandwidth
3. at each t, $\{j, t-b \leq T_j \leq t+b\}$ contribute to the sum.

4. $\mathrm{E}\hat{\alpha}(t) = b^{-1}\int_{\mathcal{T}} K(\frac{t-s}{b})\Pr(Y(s)>0)dA(s)$

Consider

$$\begin{aligned}
\alpha^*(t) &= b^{-1}\int_{\mathcal{T}} K(\frac{t-s}{b})dA^*(s) \\
&= b^{-1}\int_{\mathcal{T}} K(\frac{t-s}{b})J(s)dA(s) \quad (\text{almost} = \mathrm{E}\hat{\alpha}(t)) \\
\tilde{\alpha}(t) &= b^{-1}\int_{\mathcal{T}} K(\frac{t-s}{b})dA(s)
\end{aligned}$$

we can see that $\hat{\alpha}(t) - \alpha^*(t)$ is a martingale:

$$\begin{aligned}
\hat{\alpha}(t) - \alpha^*(t) &= b^{-1}\int_{\mathcal{T}} K(\frac{t-s}{b})d(\hat{A} - A^*)(s) \\
&= b^{-1}\int_{\mathcal{T}} K(\frac{t-s}{b})\frac{J(s)dM(s)}{Y(s)}.
\end{aligned}$$

and its expected predictable variation process:

$$\begin{aligned}
\tilde{\tau}^2(t) &= \mathrm{E}(\hat{\alpha}(t) - \alpha^*(t))^2 \\
&= b^{-2}\int_{\mathcal{T}} K^2(\frac{t-s}{b})\mathrm{E}(\frac{J(s)}{Y(s)})\alpha(s)ds \\
&= b^{-1}\int_{-1}^{1} K^2(u)\mathrm{E}(\frac{J(t-bu)}{Y(t-bu)})\alpha(t-bu)du
\end{aligned}$$

therefore an estimator of variance of $\hat{\alpha}(t)$:

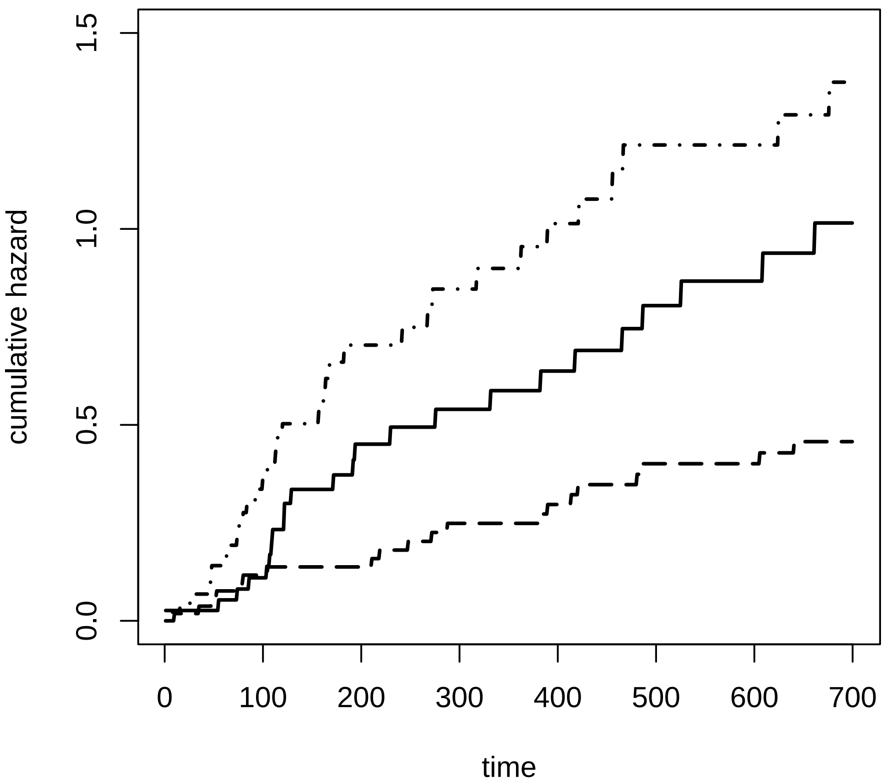
$$\hat{\tau}^2(t) = b^{-2}\int_{\mathcal{T}} K^2(\frac{t-s}{b})\frac{J(s)}{Y^2(s)}dN(s)$$

3 Mean Intergrated Squared Error and Optimal Bandwith

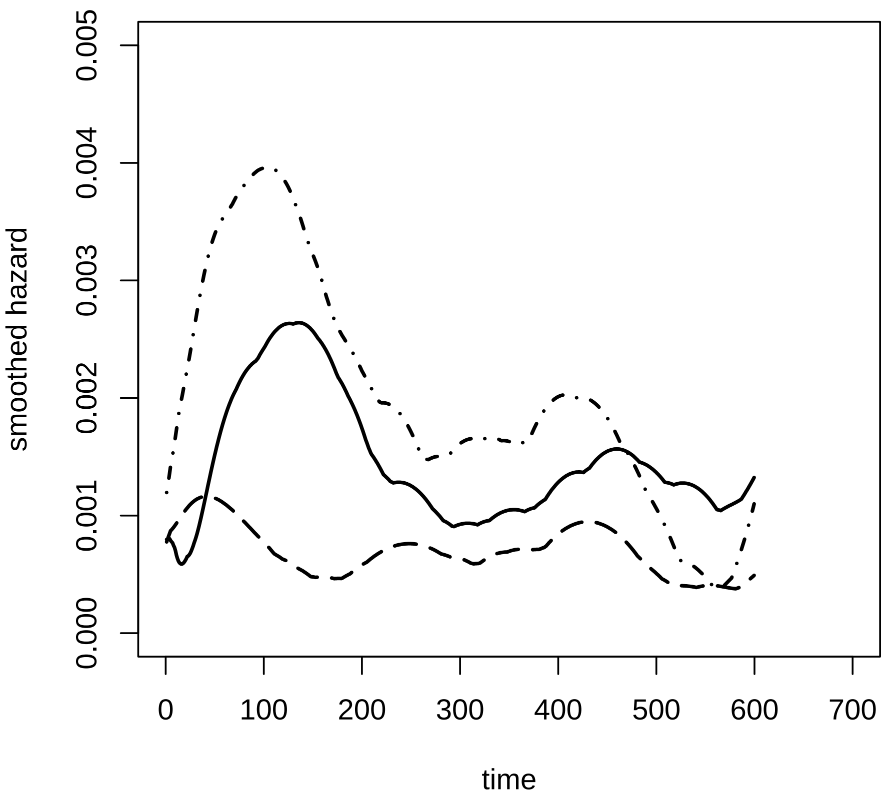
To obtain the optimal 'global' bandwidth, we choose the one minimizing the mean intergrated squared error(MISE), which is defined as,

$$\begin{aligned}
\mathrm{MISE}(\hat{\alpha}(t)) &= \mathrm{E}\int_{t_1}^{t_2}(\hat{\alpha}(t) - \alpha(t))^2 dt \\
&= \mathrm{E}\int_{t_1}^{t_2}[(\hat{\alpha}(t) - \tilde{\alpha}(t)) + (\tilde{\alpha}(t) - \alpha(t))]^2 dt \\
&= \int_{t_1}^{t_2}\mathrm{E}(\hat{\alpha}(t) - \tilde{\alpha}(t))^2 dt + \int_{t_1}^{t_2}(\tilde{\alpha}(t) - \alpha(t))^2 dt \\
&\quad +2\int_{t_1}^{t_2}(\mathrm{E}\hat{\alpha}(t) - \tilde{\alpha}(t))(\tilde{\alpha}(t) - \alpha(t))dt \\
&= (\text{ variance term}) + (\text{squared bias term}) + R_2(t)
\end{aligned}$$

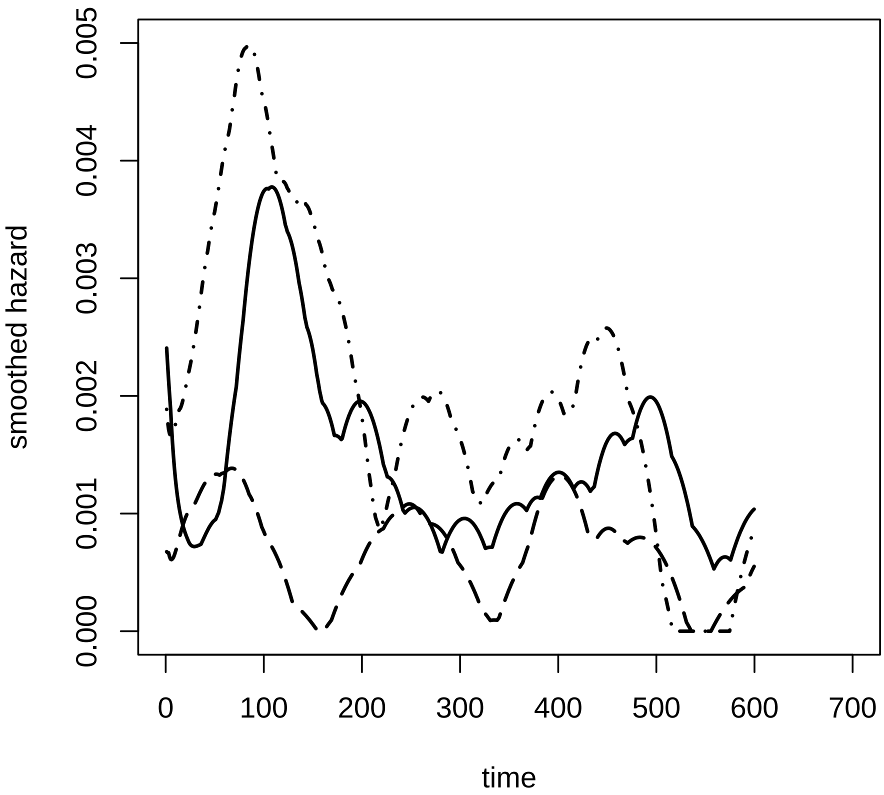
BMT data



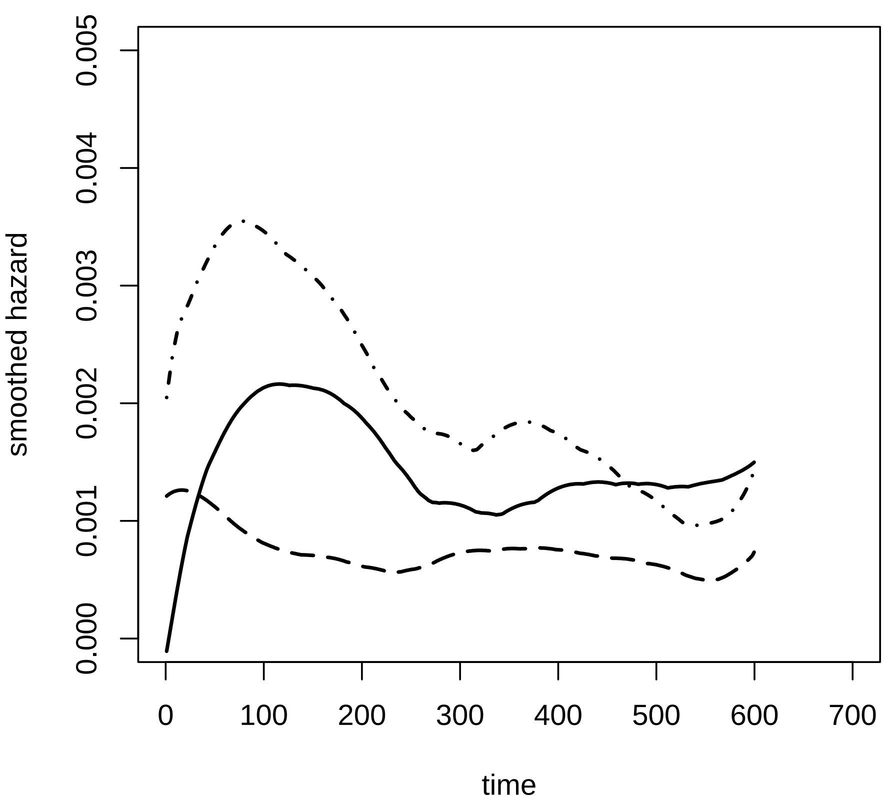
bandwidth=100



bandwidth=50



bandwidth=150



Let

$$R_2 = 2\int_{t_1}^{t_2} (\mathrm{E}\hat{\alpha}(t) - \tilde{\alpha}(t))(\tilde{\alpha}(t) - \alpha(t))dt = 2\int_{t_1}^{t_2} R_1(t)(\tilde{\alpha}(t) - \alpha(t))dt$$

with

$$\begin{array}{ll} R_1(t) & = (\mathrm{E}\hat{\alpha}(t) - \tilde{\alpha}(t)) = b^{-1}\int_{\mathcal{T}} K(\frac{t-s}{b})\mathrm{Pr}(Y(s)=0)dA(s) \\ & = \int_{-1}^{1} K(u)\mathrm{Pr}(Y(t-b_n u)=0)\alpha(t-b_n u)du \\ |R_1(t)| & \leq C_1 \sup_{t\in[t_1-c,t_2+c]} \mathrm{Pr}(Y(t)=0) \\ |R_2(t)| & \leq C_2 \sup_{t\in[t_1-c,t_2+c]} \mathrm{Pr}(Y(t)=0) \end{array}$$

Assume

1. $\int_{-1}^{1} K(u)du = 1, \quad \int_{-1}^{1} uK(u)du = 0, \int_{-1}^{1} u^2K(u)du = k_2 > 0$
2. there exist a sequence $\{a_n\}$, $a_n \to \infty$ as $n \to \infty$ (in most case, $a_n = \sqrt{n}$)
3. there exists a continuous function y such that $E(\frac{a_n^2 J(t)}{Y(t)}) \to \frac{1}{y(t)}$ uniformly on $[t_1 - c, t_2 + c]$, as $n \to \infty$
4. $\sup_{t\in[t_1-c,t_2+c]} \mathrm{Pr}(Y(t)=0) = o(a_n^{-2})$

Since

$$\begin{array}{ll} \tilde{\alpha}(t) - \alpha(t) & = \int_{-1}^{1} K(u)(\alpha(t-b_n u) - \alpha(t))du \\ & = -b_n\alpha'(t)\int_{-1}^{1} uK(u)du + \frac{1}{2}b_n^2\alpha''(t)\int_{-1}^{1} u^2K(u)du + o(b_n^2) \\ & = \frac{1}{2}b_n^2\alpha''(t)k_2 + o(b_n^2) \end{array}$$

the squared bias term $= \int_{t_1}^{t_2}(\tilde{\alpha}(t) - \alpha(t))^2dt = \frac{1}{4}b_n^4k_2^2\int_{t_1}^{t_2}(\alpha''(t))^2dt + o(b_n^4)$
For the variance term:

$$\begin{array}{ll} \mathrm{E}(\hat{\alpha}(t) - \tilde{\alpha}(t))^2 & = \mathrm{E}(\hat{\alpha}(t) - \alpha^*(t))^2 + \mathrm{E}(\alpha^*(t) - \tilde{\alpha}(t))^2 \\ & \quad +2\mathrm{E}(\hat{\alpha}(t) - \alpha^*(t))(\alpha^*(t) - \tilde{\alpha}(t)) \end{array}$$

the first term on the right-hand side (see $\hat{\tau}^2(t)$) can be expressed by,

$$(a_n^2 b_n)^{-1}\frac{\alpha(t)}{y(t)}\int_{-1}^{1} K^2(u)du + o((a_n^2 b_n)^{-1})$$

and the second term on the right-hand side:

$$|\alpha^*(t) - \tilde{\alpha}(t)| = \int_{-1}^{1} K(u)I(Y(t-b_n u)=0)\alpha(t-b_n u)du = o(a_n^{-2})$$

by Cauchy -Schwarz inequality, the last term on the right-hand side is of the form $o((a_n^2 b_n^{1/2})^{-1})$. so the 'variance term' can be written as,

$$\int_{t_1}^{t_2} \mathrm{E}(\hat{\alpha}(t) - \tilde{\alpha}(t))^2 dt = (a_n^2 b_n)^{-1} \int_{-1}^{1} K^2(t)dt \int_{t_1}^{t_2} \frac{\alpha(t)}{y(t)} dt + o((a_n^2 b_n)^{-1})$$

All together,

$$\mathrm{MISE}(\hat{\alpha}) = \frac{1}{4} b_n^4 k_2^2 \int_{t_1}^{t_2} (\alpha''(t))^2 dt + a_n^{-2} b_n^{-1} \int_{-1}^{1} K^2(t)dt \int_{t_1}^{t_2} \frac{\alpha(t)}{y(t)} dt + o(b_n^4) + o((a_n^2 b_n)^{-1})$$

to minimize the sum of the first two terms, the optimal bandwidth is

$$b_n = a_n^{-2/5} k_2^{-2/5} \left[\int_{-1}^{1} K^2(t)dt \int_{t_1}^{t_2} \frac{\alpha(t)}{y(t)} dt \right]^{1/5} \left[\int_{t_1}^{t_2} (\alpha''(t))^2 dt \right]^{-1/5}$$

Two quantities are remained to estimated: $\alpha''(t)$ and $\int_{t_1}^{t_2} \frac{\alpha(t)}{y(t)} dt$

$$\hat{\alpha''}(t) = b^{-3} \int_{\mathcal{T}} K''(\frac{t-s}{b}) d\hat{A}(s)$$

and, an estimator of $\int_{t_1}^{t_2} \frac{\alpha(t)}{y(t)} dt$ is

$$a_n^2 \int_{t_1}^{t_2} \frac{d\hat{A}(t)}{Y(t)} dt.$$

Because the bias is

$$\mathrm{E}(\hat{\alpha}(t)) - \alpha(t) \approx \frac{1}{2} b^2 \alpha''(t) k_2$$

the bias-corrected estimate is $\hat{\alpha}(t) + \frac{1}{2} b_n^2 \hat{\alpha}''(t) k_2$

3.1 Cross Validation Method

We can expressed the MISE,

$$\mathrm{MISE}(b) = \mathrm{E} \int_{t_1}^{t_2} (\hat{\alpha}(t))^2 dt - 2\mathrm{E} \int_{t_1}^{t_2} \hat{\alpha}(t)\alpha(t) dt + \int_{t_1}^{t_2} (\alpha(t))^2 dt,$$

it's a function of b; the first term is easy to caluclate, the third term deos not depend on b, and a consistent estimator of the second term is the 'cross -validation' estimator:

$$-2 \sum_{i \neq j} \frac{1}{b} K(\frac{T_i - T_j}{b}) \frac{\Delta N(T_i)}{Y(T_i)} \frac{\Delta N(T_j)}{Y(T_j)}$$

and plotting MISE(b) against b will find the optimal b

3.2 Boundary Correction and Local Bandwidth

Now we allow both bandwidth $b = b(t)$ and kernel $K = K_t$ depend on t:

$$\hat{\alpha}(t, b(t)) = b(t)^{-1} \sum_j K_t(\frac{t - T_j}{b(t)})(Y(T_j))^{-1}$$

Assuming the support of the survival time distribution is on $[0, R]$.

- interior points $I = t : b(t) \leq t \leq R - b(t)$
- left boundary region $B_L = t : 0 \leq t < b(t)$
- right boundary region $B_R = t : R - b(t) < t \leq R$

the kernels K_t are polynomials

$$K_t(z) = \{ \begin{array}{ll} K_+(1, z) & t \in I, \\ K_+(t/b(t), z) & t \in B_L, \\ K_-((R - t)/b(t), z) & t \in B_R, \end{array}$$

where $K_\pm : [0, 1] \times [-1, 1] \to \mathcal{R}$ are bounded kernel function with

$$\int K_\pm(q, z)dz = 1, \quad \int K_\pm(q, z)zdz = 0, \quad \int K_\pm(q, z)z^2dz \neq 0,$$

$$K_-(q, z) = K_+(q, -z)$$

and some differentiability conditions. e.g,

$$K_+(q, z) = \frac{12}{(1 + q)^4}(z + 1)[z(1 - 2q) + (3q^2 - 2q + 1)/2]$$

and $K_+(1, z) = \frac{3}{4}(1 - x^2)$ is the Epanechnikov kernel

For local bandwidth, we consider to minimize local mean square error to obtain bandwidth $b(t)$

$$\text{MSE}(t, b(t)) = \hat{v}(t, b(t)) + \hat{\beta}^2(t, b(t))$$

$\hat{v}$ and $\hat{\beta}$ are variance and bias estimatrs,

$$\hat{v}(t, b(t)) = \frac{1}{nb(t)} \int K_t^2(u)(\frac{\hat{\alpha}(t - b(t)u)}{\hat{y}(t - b(t)u)})du$$

$$\hat{\beta}(t, b(t)) = \int \hat{\alpha}(t - b(t)u)K_t(u)du - \hat{\alpha}(t)$$

The following is the recommended algorithm to obtain the $\alpha(t)$ with $b(t)$.

1 choose $K_+(q,z)$ and initial bandwidth b_0 to obtain $\hat{\alpha}(t)$

1a alternatively, choose a parametric model and obtain MLE $\hat{\alpha}(t)$

2 choose equidistant grid of m_l points $t_i, i = 1 \cdots m_l$, between 0 and R. choose a grid of bandwidth $\bar{b}_j, j = 1, \cdots, l$, between some b_1 and b_2. compute $\hat{v}(t_i)$ and $\hat{\beta}(t_i)$ with all bandwidth $\bar{b}_j$'s and obtain minimizer $\bar{b}(t_i)$ of MSE(t_i).

3 obtain final bandwidth using boundary -modified smoother with bandwidth b_0 or $\frac{3}{2}b_0$

$$\hat{b}(t) = \sum K_t(\frac{t-t_i}{b_0})\bar{b}(t_i) / \sum K_t(\frac{t-t_i}{b_0})$$

4 obtain final estimate $\hat{\alpha}(t)$ with $b(t) = \hat{b}(t)$

4 Large Sample Properties

Now we discuss the large sample properties of the kernel function estimator with 'global' bandwidth.

Theoremn IV.2.1 pointwise consistency

Assume

1. t be an interior point of $\mathcal{T}$
2. α is continuous at t
3. bandwith $b_n \to 0$ as $n \to \infty$
4. there exist an $\epsilon > 0$ such that $\inf_{s \in [t-\epsilon, t+\epsilon]} b_n Y(s) \to_p \infty$ as $n \to \infty$

Then $\hat{\alpha} \to_p \alpha$ as $n \to \infty$

Proof.

$$|\hat{\alpha}(t) - \alpha(t)| \leq |\hat{\alpha}(t) - \alpha^*(t)| + |\alpha^*(t) - \tilde{\alpha}(t)| + |\tilde{\alpha}(t) - \alpha(t)|$$

we have to show

$$|\hat{\alpha}(t) - \alpha^*(t)| \to_p 0$$
$$|\alpha^*(t) - \tilde{\alpha}(t)| \to_p 0$$
$$|\tilde{\alpha}(t) - \alpha(t)| \to_p 0$$

$$\begin{aligned}
&\Pr(|\hat{\alpha}(t) - \alpha^*(t)| > \eta) \qquad \text{by Lenglart's inequality} \\
&\le \Pr(\sup_{t-b_n \le s \le t+b_n} \left| b_n^{-1} \int_{t-b_n}^{s} K(\frac{t-u}{b_n}) \frac{J(u)dM(u)}{Y(u)} \right| > \eta) \\
&\le \frac{\delta}{\eta^2} + \Pr(b_n^{-1} \int_{-1}^{1} K^2(u) \frac{J(t-b_n u)\alpha(t-b_n u)du}{Y(t-b_n u)}) > \delta)
\end{aligned}$$

since α and K are bounded, and $b_n Y(t) \to_p \infty$ in a neighborhood of t, the last term on the right-hand side can be arbitrary small. so $|\hat{\alpha}(t) - \alpha^*(t)| \to_p 0$

$$|\alpha^*(t) - \tilde{\alpha}(t)| \le \int_{-1}^{1} |K(u)|\{1 - J(t-b_n u)\}\alpha(t-b_n u)du \to_p 0$$

$$|\tilde{\alpha}(t) - \alpha(t)| \le \int_{-1}^{1} |K(u)||\alpha(t-b_n u) - \alpha(t)|du \to_p 0$$

□

Theoremn IV.2.2 uniform conisitency
Assume

1. t be an interior point of $\mathcal{T}$
2. $0 < t_1 < t_2 < t$ be fixed numbers
3. α is continuous on $[0, t]$
4. the kernel K is of bounded variation
5. bandwith $b_n \to 0$ as $n \to \infty$
6. $b_n^{-2} \int_0^t \frac{J(s)\alpha(s)ds}{Y(s)} \to_p 0$
7. $\int_0^t (1 - J(s))\alpha(s)ds \to_p 0$

Then as $n \to \infty$,

$$\sup_{s \in [t_1, t_2]} |\hat{\alpha}(s) - \alpha(s)| \to_p 0$$

Proof.

$$\begin{aligned}
|\hat{\alpha}(t) - \alpha^*(t)| &= |b^{-1} \int_{\mathcal{T}} K(\frac{t-s}{b}) d(\hat{A} - A^*)(s)| \\
&\le 2b^{-1} V(K) \sup_{t \in [0,t]} |\hat{A}(s) - A^*(s)|
\end{aligned}$$

where $V(K)$ is the toal variation of $K(t)$.
similar to the proof of the consistency of Nelson-Aalen estimator, with assumption (6) the last term converges to zero.

$\sup_{t\in[t_1,t_2]}|\alpha^*(t)-\tilde{\alpha}(t)| \to_p 0$ with assumption (7)
$\sup_{t\in[t_1,t_2]}|\tilde{\alpha}(t)-\alpha(t)| \to_p 0$ follows by the boundedness of K and continuity of α. □

Remark Sufficient conditions for consistency

- uniform consistency of Nelson Aalen estimator : $\inf_{s\in[0,t]} Y(s) \to_p \infty$
- pointwise consistency of kernel function estimator : $\inf_{s\in[0,t]} bY(s) \to_p \infty$
- uniform consistency of kernel function estimator : $\inf_{s\in[0,t]} b^2Y(s) \to_p \infty$

Theorem IV.2.4 Asymptotic normality
Assume

1. t be an interior point of $\mathcal{T}$
2. α is continuous at t
3. there exits positive constants $\{a_n\}$, increasing to ∞ as $n\to\infty$,
4. $b_n \to 0$, $a_n^2 b_n \to \infty$ as $n\to\infty$
5. there exists a function y, positive and continuous at t such that
$$\sup_{s\in[t-\epsilon,t+\epsilon]} |a_n^{-2}Y(s)-y(s)| \to_p 0$$
as $n\to\infty$, for an ϵ

Then

1. $a_n b_n^{1/2}(\hat{\alpha}(t)-\tilde{\alpha}(t)) \to_d \mathcal{N}(0,\tau^2(t))$, $\tau^2(t)=\dfrac{\alpha(t)}{y(t)}\displaystyle\int_{-1}^{1} K^2(u)du$
2. $a_n^2 b_n \hat{\tau}^2(t) \to_p \tau^2$
3. for $t_1 \neq t_2$, $\hat{\alpha}(t_1)$ and $\hat{\alpha}(t_2)$ are asymptotic independent.

Proof. A.
$$a_n b_n^{1/2}(\hat{\alpha}(t)-\alpha^*(t)) = \int_{\mathcal{T}} H(s)dM(s),$$
with $H(s)=a_n b_n^{-1/2}K(\dfrac{t-s}{b_n})\dfrac{J(s)}{Y(s)}$.
Using Rebolledo's martingale CLT,
$$\begin{aligned}&\inf\nolimits_{\mathcal{T}} H^2(s)Y(s)\alpha(s)ds = \textstyle\int_{\mathcal{T}} a_n^2 b_n^{-1} K^2(\frac{t-s}{b_n})\frac{J(s)\alpha(s)ds}{Y(s)}\\ &= \textstyle\int_{-1}^{1} a_n^2 K^2(u)\frac{J(t-b_nu)\alpha(t-b_nu)du}{Y(t-b_nu)}\\ &\to_p \textstyle\frac{\alpha(t)}{y(t)}\int_{-1}^{1} K^2(u)du, \quad \text{as } n\to\infty\end{aligned}$$

With any $\varepsilon > 0$,

$$I(|H(s)| > \varepsilon) = I(|K(\frac{t-s}{b_n})\frac{J(s)a_n^2}{Y(s)}| > \varepsilon a_n b_n^{1/2})$$

converges to zero uniformly because $a_n b_n^1/2 \to \infty$, therefore

$$\inf_{\mathcal{T}} H^2(s)Y(s)\alpha(s)I(|H(s)| > \varepsilon)ds \to 0$$

Also

$$a_n b_n^{1/2}(\alpha^*(t) - \tilde{\alpha}(t)) \to 0$$

B. since

$$a_n b_n^{1/2}(\hat{\alpha}(t_i) - \alpha^*(t_i)) = \int_{\mathcal{T}} H_i(s)dM(s),$$

with $H_i(s) = a_n b_n^{-1/2} K(\dfrac{t_i - s}{b_n})\dfrac{J(s)}{Y(s)}$ therefore,

$$\begin{aligned}
&\textstyle\int_{\mathcal{T}} H_1(s)H_2(s)Y(s)\alpha(s)ds \\
&\quad = \textstyle\int_{\mathcal{T}} a_n^2 b_n^{-1} K(\frac{t_1-s}{b_n})K(\frac{t_2-s}{b_n})\frac{J(s)\alpha(s)ds}{Y(s)} \\
&\quad \to 0
\end{aligned}$$

□

Theorem IV.2.5 Assume all conditions in Theorem IV.2.4, and

1. α is twice continuously differentiable in a neighbourhood of t
2. $\int_{-1}^{1} K(u)du = 1, \quad \int_{-1}^{1} uK(u)du = 0, \int_{-1}^{1} u^2K(u)du = k_2 > 0$
3. $\limsup_{n\to\infty} a_n^{2/5} b_n < \infty$

Then

$$a_n b_n^{1/2}(\hat{\alpha}(t) - \alpha(t) - 2^{-1}b_n^2\alpha''(t)k_2) \to_d \mathcal{N}(0, \tau^2(t))$$

Proof. With Taylor expansion (t^* is between $t - b_n u$ and t)

$$\begin{aligned}
&a_n b_n^{1/2}(\tilde{\alpha}(t) - \alpha(t) - 2^{-1}b_n^2\alpha''(t)k_2) \\
&\quad = a_n b_n^{1/2}\left[\textstyle\int_{-1}^{1} K(u)\alpha(t - b_n u)du - \alpha(t) - 2^{-1}b_n^2\alpha''(t)k_2\right] \\
&\quad = \tfrac{1}{2}a_n b_n^{2/5}\left[u^2K(u)\alpha(t^*)du - \alpha''(t)k_2\right] \\
&\quad \to 0
\end{aligned}$$

□