

Goodness-of-fit tests for the cure rate in a mixture cure model

BY U.U. MÜLLER

Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A.
uschi@stat.tamu.edu

AND I. VAN KEILEGOM

Research Centre for Operations Research and Business Statistics, KU Leuven,
Naamsestraat 69, 3000 Leuven, Belgium
ingrid.vankeilegom@kuleuven.be

S. SUPPLEMENTARY ONLINE MATERIAL

S.1. Proof of Theorem 1

To derive the desired result for the Härdle–Mammen-type test statistic \mathcal{T}_n from equation (2) we write

$$\begin{aligned} \mathcal{T}_n &= nh^{1/2} \int_0^1 [\{\hat{p}(x) - p(x)\}^2 + \{p(x) - p_{\hat{\theta}}(x)\}^2 \\ &\quad + 2\{p(x) - p_{\hat{\theta}}(x)\}\{\hat{p}(x) - p(x)\}]\pi(x) dx. \end{aligned}$$

Consider local alternatives $p(x) = p_{\theta}(x) + n^{-1/2}h^{-1/4}\Delta_n(x)$, where θ denotes the true parameter. This covers the null hypothesis as a special case with $\Delta_n = 0$. We write \mathcal{T}_{0n} for the first term of \mathcal{T}_n . We will show that \mathcal{T}_{0n} determines the distribution of the test statistic under the null hypothesis and that $\mathcal{T}_{0n} - b_h$ converges in distribution,

$$\mathcal{T}_{0n} - b_h = nh^{1/2} \int_0^1 \{\hat{p}(x) - p(x)\}^2 \pi(x) dx - b_h \rightarrow N(0, V) \quad (n \rightarrow \infty), \quad (\text{S.1})$$

i.e., \mathcal{T}_n is under H_0 approximately normally distributed, with mean $b_h = O(h^{-1/2})$ tending to infinity and variance V . Before proving (S.1) we consider the second term of \mathcal{T}_n and, writing $p(x) = p_{\theta}(x) + n^{-1/2}h^{-1/4}\Delta_n(x)$, we can split it into three parts:

$$\begin{aligned} &nh^{1/2} \int_0^1 \{p_{\theta}(x) - p_{\hat{\theta}}(x) + n^{-1/2}h^{-1/4}\Delta_n(x)\}^2 \pi(x) dx \\ &= nh^{1/2} \int_0^1 \{p_{\theta}(x) - p_{\hat{\theta}}(x)\}^2 \pi(x) dx + nh^{1/2} \int_0^1 \{n^{-1/2}h^{-1/4}\Delta_n(x)\}^2 \pi(x) dx \\ &\quad + 2nh^{1/2} \int_0^1 \{p_{\theta}(x) - p_{\hat{\theta}}(x)\}n^{-1/2}h^{-1/4}\Delta_n(x) \pi(x) dx \\ &= \int_0^1 \Delta_n(x)^2 \pi(x) dx + o_p(1). \end{aligned}$$

The resulting integral comes from the middle part. To see that the first part involving $(p_{\theta} - p_{\hat{\theta}})^2$ is asymptotically negligible, use assumptions (A6) and (A7), which ensure that p_{θ} is Lipschitz

in θ , uniformly in x , and that $\hat{\theta}$ is root- n consistent. This gives the order $O_p(h^{1/2}) = o_p(1)$ as $h \rightarrow 0$. The third part in the above display vanishes with rate $O_p(h^{1/4}) = o_p(1)$, which follows
 20 by the same arguments, combined with the assumption that $\Delta_n(\cdot)$ is uniformly bounded.

Now consider the mixed third term of \mathcal{T}_n . We will show that this term is asymptotically negligible,

$$nh^{1/2} \int_0^1 \{p(x) - p_{\hat{\theta}}(x)\} \{\hat{p}(x) - p(x)\} \pi(x) dx = o_p(1). \quad (\text{S.2})$$

The proof requires some auxiliary results, which we provide when we verify (S.1), the asymptotic normality of \mathcal{T}_{0n} . It is therefore postponed to the end of this proof.

25 Let Λ denote the cumulative conditional hazard function $\Lambda(t | x) = -\log S(t | x)$ and let $\hat{\Lambda}(t | x)$ be the nonparametric estimator provided in Xu and Peng (2014). These authors show that

$$(nh)^{1/2} \{\hat{p}(x) - p(x)\} = S(Y_{(n)}^1 | x) (nh)^{1/2} \{\hat{\Lambda}(Y_{(n)}^1 | x) - \Lambda(Y_{(n)}^1 | x)\} + o_p(1). \quad (\text{S.3})$$

They also show that $Y_{(n)}^1$ is weakly consistent for $\sup_x \tau_0(x)$, and that $(nh)^{1/2} \{S(Y_{(n)}^1 | x) - p(x)\} = o_p(1)$. This yields

$$(nh)^{1/2} \{\hat{p}(x) - p(x)\} = p(x) (nh)^{1/2} \{\hat{\Lambda}(Y_{(n)}^1 | x) - \Lambda(Y_{(n)}^1 | x)\} + o_p(1).$$

In order to derive asymptotic normality of \mathcal{T}_n , we will need an approximation of $\hat{\Lambda} - \Lambda$, which is, for example, provided in Du and Akritas (2002). They show in their Theorem 3.1 that, uniformly in x ,

$$\hat{\Lambda}(t | x) - \Lambda(t | x) = \sum_{i=1}^n w_{hi}(x) \zeta_i(t | x) + r(t, x)$$

30 with $\sup_{x, t \leq t_0} |r(t, x)| = O\{(nh)^{-3/4} (\log n)^{3/4}\}$ almost surely, where t_0 satisfies $\sup_x H(t_0 | x) < 1$, with w_{hi} as specified in Section 2, and with $\zeta_i(t | x)$ defined in (6). It follows from arguments in Xu and Peng that $\hat{\Lambda}(Y_{(n)}^1 | x) - \Lambda(Y_{(n)}^1 | x)$ is asymptotically equivalent to $\sum_{i=1}^n w_{hi}(x) \zeta_i\{\tau_0(x) | x\}$. Then,

$$\begin{aligned} \mathcal{T}_{0n} &= nh^{1/2} \int_0^1 \{\hat{p}(x) - p(x)\}^2 \pi(x) dx + o_p(1) \\ &= nh^{1/2} \int_0^1 p^2(x) \left[\sum_{i=1}^n w_{hi}(x) \zeta_i\{\tau_0(x) | x\} \right]^2 \pi(x) dx \\ &\quad + nh^{1/2} \int_0^1 p^2(x) \left[r\{\tau_0(x), x\} \sum_{i=1}^n w_{hi}(x) \zeta_i\{\tau_0(x) | x\} \right]^2 \pi(x) dx \\ &\quad + nh^{1/2} \int_0^1 p^2(x) [r\{\tau_0(x), x\}]^2 \pi(x) dx + o_p(1) \\ &= nh^{1/2} \int_0^1 p^2(x) \left[\sum_{i=1}^n w_{hi}(x) \zeta_i\{\tau_0(x) | x\} \right]^2 \pi(x) dx + o_p(1). \end{aligned}$$

The last two terms in the above decomposition of \mathcal{T}_{0n} (second equality) are of order $o_p(1)$, which
 35 can be explained as follows. Consider the approximation (S.3) of $\hat{p}(x) - p(x)$ by Du and Akritas (2002), which involves the difference $\hat{\Lambda} - \Lambda$; see the proof of Theorem 3.1 in Du and Akritas

(2002) for the definition of that difference. It depends on a quantity \hat{H} whose order is given in Lemma 4.2 of the same paper. This yields $\hat{p}(x) - p(x) = O_p\{(nh)^{-1/2}(\log n)^{1/2}\}$. Combining this order for the main term $\sum_{i=1}^n w_{hi}(x)\zeta_i\{\tau_0(x) | x\}$ and the order $O_p\{(nh)^{-3/4}(\log n)^{3/4}\}$ for the remainder term yields the order $O_p\{nh^{1/2}(nh)^{-5/4}(\log n)^{5/4}\}$ for the mixed second term. This is $o_p(1)$ since we assume $nh^3(\log n)^{-5} \rightarrow \infty$. The third term is $o_p(1)$ because it contains the squared remainder term and is therefore of smaller order. Inserting the definition of w_{hi} into the above gives

$$\mathcal{T}_{0n} = nh^{1/2} \int_0^1 p^2(x) \left[n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)}{f(x)} \zeta_i\{\tau_0(x) | x\} \right]^2 \pi(x) dx + o_p(1). \quad (\text{S.4})$$

Now write

$$\mathcal{T}_{0n} = \mathcal{T}_{1n} + \mathcal{T}_{2n} + o_p(1),$$

where

$$\begin{aligned} \mathcal{T}_{1n} &= \frac{h^{1/2}}{n} \int_0^1 p^2(x) \sum_{i=1}^n \frac{K_h^2(x - X_i)}{f^2(x)} \zeta_i^2\{\tau_0(x) | x\} \pi(x) dx, \\ \mathcal{T}_{2n} &= \frac{h^{1/2}}{n} \int_0^1 p^2(x) \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{K_h(x - X_i)K_h(x - X_j)}{f^2(x)} \zeta_i\{\tau_0(x) | x\} \zeta_j\{\tau_0(x) | x\} \pi(x) dx. \end{aligned}$$

Analogously as in the proof of Proposition 1 in Härdle and Mammen (1993), but now with ζ_j in place of heteroscedastic errors ε_j , we obtain the bias term as follows:

$$\begin{aligned} E(\mathcal{T}_{1n}) &= h^{1/2} \int_0^1 \frac{p^2(x)}{f^2(x)} E[K_h^2(x - X)\zeta^2\{\tau_0(x) | x\}] \pi(x) dx \\ &= h^{1/2} \int_0^1 \frac{p^2(x)}{f^2(x)} E\left[\frac{1}{h^2} K^2\left(\frac{x - X}{h}\right) E[\zeta^2\{\tau_0(x) | x\} | X]\right] \pi(x) dx \\ &= h^{1/2} \int_0^1 \frac{p^2(x)}{f^2(x)} \int \frac{1}{h} K^2(v) f(x + vh) \mu_{xx}(x + vh) dv \pi(x) dx \\ &= b_h + o(1), \end{aligned}$$

with, using assumptions (A1) and (A5),

$$b_h = h^{-1/2} R(K) \int_0^1 \frac{p^2(x)\pi(x)}{f(x)} \mu_{xx}(x) dx = O(h^{-1/2}).$$

We have $nh^3 \rightarrow \infty$, thanks to our assumptions on the bandwidth, and therefore $\text{var}(\mathcal{T}_{1n} | X_1, \dots, X_n) = O_p(n^{-1}h^{-3}) = o_p(1)$,

$$\mathcal{T}_{1n} = b_h + o_p(1).$$

The desired normal approximation (S.1), $\mathcal{T}_{0n} - b_h \approx N(0, V)$, now follows from this statement, combined with the asymptotic normality of \mathcal{T}_{2n} , $\mathcal{T}_{2n} \rightarrow N(0, V)$ in distribution as $n \rightarrow \infty$, which follows analogously to the arguments outlined in Härdle and Mammen (1993), using The-

orem 2.1 by de Jong (1987): write

$$\begin{aligned}\mathcal{T}_{2n} &= \frac{h^{1/2}}{n} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n V_{ij}, \\ V_{ij} &= \int_0^1 p^2(x) \frac{K_h(x - X_i)K_h(x - X_j)}{f^2(x)} \zeta_i\{\tau_0(x) | x\} \zeta_j\{\tau_0(x) | x\} \pi(x) dx.\end{aligned}$$

De Jong's theorem requires that \mathcal{T}_{2n} is clean in the sense that the conditional expectations of the V_{ij} given X_i ($i = 1, \dots, n$) vanish. Using similar arguments as in Cao and González-Manteiga (2008; pages 183-184), we see that $\zeta_i\{\tau_0(x) | x\}$ and $\zeta_j\{\tau_0(x) | x\}$ in the expression of V_{ij} can be replaced by $\zeta_i\{\tau_0(X_i) | X_i\}$ and $\zeta_j\{\tau_0(X_j) | X_j\}$. Since $E[\zeta\{\tau_0(x) | x\} | X = x] = 0$, it follows that \mathcal{T}_{2n} is indeed clean up to an asymptotically negligible term (and the asymptotic mean is zero).

In order to obtain the asymptotic variance, it suffices to calculate the second moment of V_{ij} . We have

$$\begin{aligned}E(V_{ij}^2) &= E\left(\int_0^1 \int_0^1 \frac{p^2(x)}{f^2(x)} K_h(x - X_i)K_h(x - X_j) \zeta_i\{\tau_0(x) | x\} \zeta_j\{\tau_0(x) | x\} \pi(x) \right. \\ &\quad \times \left. \frac{p^2(z)}{f^2(z)} K_h(z - X_i)K_h(z - X_j) \zeta_i\{\tau_0(z) | z\} \zeta_j\{\tau_0(z) | z\} \pi(z) dx dz\right) \\ &= \int_0^1 \int_0^1 \frac{p^2(x)p^2(z)\pi(x)\pi(z)}{f^2(x)f^2(z)} E[K_h(x - X_i)K_h(z - X_i) \zeta_i\{\tau_0(x) | x\} \zeta_i\{\tau_0(z) | z\}] \\ &\quad \times E[K_h(x - X_j)K_h(z - X_j) \zeta_j\{\tau_0(x) | x\} \zeta_j\{\tau_0(z) | z\}] dx dz \\ &= \int_0^1 \int_0^1 \frac{p^2(x)p^2(z)\pi(x)\pi(z)}{f^2(x)f^2(z)} \\ &\quad \times \left(\int K_h(x - u)K_h(z - u) E[\zeta\{\tau_0(x) | x\} \zeta\{\tau_0(z) | z\} | X = u] f(u) du\right)^2 dx dz \\ &= \int_0^1 \int_0^1 \frac{p^2(x)p^2(z)\pi(x)\pi(z)}{f(x)f(z)} \mu_{xz}(x)\mu_{xz}(z) \left\{\frac{1}{h} K * K\left(\frac{z-x}{h}\right)\right\}^2 dx dz + o(1)\end{aligned}$$

as $h \rightarrow 0$. Hence, using the fact that \mathcal{T}_{2n} is clean, we obtain

$$\begin{aligned}\text{var}(\mathcal{T}_{2n}) &= \text{var}\left(\frac{h^{1/2}}{n} 2 \sum_{i < j}^n V_{ij}\right) = \frac{4h}{n^2} \sum_{i < j} \sum_{k < l} \text{cov}(V_{ij}, V_{kl}) = \frac{4h}{n^2} \sum_{i < j} E(V_{ij}^2) + o(1) \\ &= 2hE(V_{ij}^2) + o(1) = V_h + o(1),\end{aligned}$$

with

$$\begin{aligned}V_h &= 2h \int_0^1 \int_0^1 \frac{p^2(x)p^2(z)\pi(x)\pi(z)}{f(x)f(z)} \left\{\frac{1}{h} K * K\left(\frac{z-x}{h}\right)\right\}^2 \\ &\quad \times E[\zeta\{\tau_0(x) | x\} \zeta\{\tau_0(z) | z\} | X = x] E[\zeta\{\tau_0(x) | x\} \zeta\{\tau_0(z) | z\} | X = z] dx dz.\end{aligned}$$

Since $p(\cdot)$ is continuous and $\mu_{xz}(z)$ is continuous in z by assumption (A5), we can approximate V_h by V , as on page 1931 of Härdle and Mammen (1993), and obtain the desired formula,

$$V = 2K^{(4)}(0) \int \left\{ \frac{p^2(x)\pi(x) E[\zeta^2\{\tau_0(x) | x\} | X = x]}{f(x)} \right\}^2 dx.$$

It remains to verify equation (S.2), that the third mixed term of \mathcal{T}_n is asymptotically negligible. To see this use the arguments by Xu and Peng (2014) and Du and Akritas (2002), which we used above to derive approximation (S.4), namely

$$\hat{p}(x) - p(x) = p(x)n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)}{f(x)} \zeta_i\{\tau_0(x) \mid x\} + O\{(nh)^{-3/4}(\log n)^{3/4}\}.$$

almost surely. Inserting the first term on the right-hand side into (S.2) gives

$$nh^{1/2} \int_0^1 \{p(x) - p_{\hat{\theta}}(x)\} p(x) n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)}{f(x)} \zeta_i\{\tau_0(x) \mid x\} \pi(x) dx. \quad (\text{S.5})$$

To determine the order of the part involving the remainder term remember that $p(x) = p_{\theta}(x) + n^{-1/2}h^{-1/4}\Delta_n(x)$ with Δ_n uniformly bounded. Now use the assumptions on the bandwidth and assumptions (A6) and (A7) to obtain $p(x) - p_{\hat{\theta}}(x) = p_{\theta}(x) - p_{\hat{\theta}}(x) + n^{-1/2}h^{-1/4}\Delta_n(x) = O_p(n^{-1/2}h^{-1/4})$. Hence the second part of the statistic has the rate

$$O_p\{nh^{1/2}(nh)^{-3/4}(\log n)^{3/4}n^{-1/2}h^{-1/4}\} = O_p\{n^{-1/4}h^{-1/2}(\log n)^{3/4}\} = o_p(1),$$

so it suffices to study statistic (S.5). Using a Taylor expansion we write it as a sum of three terms, $T_{31} + T_{32} + T_{33}$, where

$$T_{31} = nh^{1/2}(\hat{\theta} - \theta)^\top \int_0^1 p(x)n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)}{f(x)} \zeta_i\{\tau_0(x) \mid x\} \dot{p}_{\theta}(x) \pi(x) dx,$$

$$T_{32} = nh^{1/2}(\hat{\theta} - \theta)^\top \int_0^1 p(x)n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)}{f(x)} \zeta_i\{\tau_0(x) \mid x\} \{\dot{p}_{\xi}(x) - \dot{p}_{\theta}(x)\} \pi(x) dx,$$

$$T_{33} = nh^{1/2}n^{-1/2}h^{-1/4} \int_0^1 p(x)n^{-1} \sum_{i=1}^n \frac{K_h(x - X_i)}{f(x)} \zeta_i\{\tau_0(x) \mid x\} \Delta_n(x) \pi(x) dx.$$

Here \dot{p}_t denotes the vector of partial derivatives with respect to the parameter evaluated at t , and ξ is an intermediate value between θ and $\hat{\theta}$. 60

Consider the second term first and use the Lipschitz assumption (A6) on the gradient \dot{p}_{θ} and the root- n consistency of $\hat{\theta}$, assumption (A7), to obtain $T_{32} = O_p(nh^{1/2}n^{-1/2}n^{-1/2}) = o_p(1)$. The first and the third term, T_{31} and T_{33} are similar. We consider only T_{33} , which is more difficult to handle since it contains the additional factor $h^{-1/4}$ and therefore converges at a slower rate. To get the desired $o_p(1)$ rate note that the order of the integral in T_{33} is $O(h^3)$ using the assumptions on the kernel K , and the formula of $E\{\zeta_i(t \mid x) \mid X_i\}$ given in equation (3.3) on page 474 in Van Keilegom and Veraverbeke (1997). Our function ζ is the function g in that paper. This gives

$$E(T_{33}) = nh^{1/2}n^{-1/2}h^{-1/4}O(h^3) = O(n^{1/2}h^{13/4}) = o(1).$$

The second moment, $E(T_{33}^2)$, is

$$\begin{aligned} & nh^{1/2}n^{-2} \sum_{i=1}^n E \left(\int_0^1 \int_0^1 p(x)p(y) \frac{K_h(x-X_i)}{f(x)} \frac{K_h(y-X_i)}{f(y)} \right. \\ & \quad \times E[\zeta_i\{\tau_0(x) \mid x\} \zeta_i\{\tau_0(y) \mid y\} \mid X_i] \Delta_n(x) \Delta_n(y) \pi(x) \pi(y) dx dy \Big) \\ & + nh^{1/2}n^{-2} \sum_{\substack{i,j=1 \\ j \neq i}}^n E \left(\int_0^1 \int_0^1 p(x)p(y) \frac{K_h(x-X_i)}{f(x)} \frac{K_h(y-X_j)}{f(y)} \right. \\ & \quad \times E[\zeta_i\{\tau_0(x) \mid x\} \mid X_i] E[\zeta_j\{\tau_0(y) \mid y\} \mid X_j] \Delta_n(x) \Delta_n(y) \pi(x) \pi(y) dx dy \Big) \end{aligned}$$

Now use again the order of the integral in the expression of T_{33} to obtain

$$E(T_{33}^2) = O(h^{1/2}) + O(nh^{1/2}h^6) = o(1) + O(nh^{6.5}) = o(1).$$

Since both the mean and the variance of T_{33} tend to zero in probability, we can apply Chebyshev's inequality and obtain the desired $T_{33} = o_p(1)$. This completes the proof of (S.2) \square

S.2. Proof of equation (8)

65 We sketch the proof of the asymptotic equivalence of \mathcal{T}_n and $\tilde{\mathcal{T}}_n$, that is, $\tilde{\mathcal{T}}_n - \mathcal{T}_n = o_p(1)$. Consider

$$\begin{aligned} |\tilde{\mathcal{T}}_n - \mathcal{T}_n| &= nh^{1/2} \left| \int \{\hat{p}(x) - p_{\hat{\theta}}(x)\}^2 d\{\hat{F}(x) - F(x)\} \right| \\ &\leq nh^{1/2} \left| \int \{\hat{p}(x) - p_{\hat{\theta}}(x)\}^2 d\{\hat{F}(x) - \tilde{F}(x)\} \right| \\ &\quad + nh^{1/2} \left| \int \{\hat{p}(x) - p_{\hat{\theta}}(x)\}^2 d\{\tilde{F}(x) - F(x)\} \right|, \end{aligned} \tag{S.6}$$

where $\tilde{F}(x) = \int \hat{F}(x - vs) dL(v)$, $L(\cdot) = \int_{-\infty}^{\cdot} \ell(u) du$, ℓ is a symmetric kernel density function with mean zero and bounded support, and $s = s_n$ is a bandwidth sequence controlling the smoothness of $\tilde{F}(x)$. We first show

$$\sup_x |\hat{F}(x) - \tilde{F}(x)| = o_p(n^{-1/2}), \tag{S.7}$$

70 assuming $ns^4 \rightarrow 0$ as $n \rightarrow \infty$, using arguments similar, but in fact much simpler, than those used in the proof of Lemma A.3 in a 2017 preprint by Neumeyer and Van Keilegom on distribution functions of residuals entitled *Bootstrap of residual processes in regression: to smooth or not to smooth ?*. Write

$$\begin{aligned} \tilde{F}(x) - \hat{F}(x) &= \int \{\hat{F}(x - vs) - \hat{F}(x)\} dL(v) \\ &= \int \{\hat{F}(x - vs) - \hat{F}(x) - F(x - vs) + F(x)\} dL(v) \\ &\quad + \int \{F(x - vs) - F(x)\} dL(v) \\ &= n^{-1/2} \int \{E_n(x - vs) - E_n(x)\} dL(v) + \int \{F(x - vs) - F(x)\} dL(v), \end{aligned}$$

where E_n denotes the empirical process $n^{-1/2} \sum_{i=1}^n \{1(X_i \leq \cdot) - F(\cdot)\}$. Since E_n is asymptotically equicontinuous, and since ℓ has a bounded support, it follows that the first term has the order $o_p(n^{-1/2})$, uniformly in x . To see that the second term has the order $O_p(s^2) = o_p(n^{-1/2})$, use the fact that f' is bounded, that $ns^4 = o(1)$ and that the kernel ℓ has mean zero. This proves (S.7). 75

Consider the first term of (S.6). Using integration by parts, we can write it as the sum of two integrals. The first integral can be bounded by

$$2nh^{1/2} \int |\hat{F}(x) - \tilde{F}(x)| |\hat{p}(x) - p_{\hat{\theta}}(x)| |\hat{p}'(x) - p'_{\hat{\theta}}(x)| dx.$$

To determine the rate of $\hat{p}(x) - p_{\hat{\theta}}(x)$, we only need to consider $\hat{p}(x) - p(x)$, because it is of larger order than $p_{\hat{\theta}}(x) - p(x)$, which converges with the parametric root- n rate, thanks to Assumptions (A6) and (A7). As explained in the paragraph after equation (S.4), we have $\hat{p}(x) - p(x) = O_p\{(nh)^{-1/2}(\log n)^{1/2}\}$. Further it can be shown that $\hat{p}'(x) - p'_{\hat{\theta}}(x) = O_p\{(nh^3)^{-1/2}(\log n)^{1/2}\}$, uniformly in x . This combined with (S.7) gives the order

$$\begin{aligned} & nh^{1/2} o_p(n^{-1/2}) O_p\{(nh)^{-1/2}(\log n)^{1/2}\} O_p\{(nh^3)^{-1/2}(\log n)^{1/2}\} \\ & = o_p\{(nh^3)^{-1/2} \log n\} = o_p(1) \end{aligned}$$

for the above term. For the last step we used the assumptions on the bandwidth $h = h_n$ in Theorem 1: we assume that $nh^3(\log n)^{-5} \rightarrow \infty$, which implies $(nh^3)^{-1/2} \log n = O(1)$. The same arguments yield for the second integral the order

$$nh^{1/2} o_p(n^{-1/2}) [O_p\{(nh)^{-1/2}(\log n)^{1/2}\}]^2,$$

which is $o_p(1)$, since it is of smaller order than the first one. This yields the desired order $o_p(1)$ for the first term on the right-handside of (S.6). 80

The second term of (S.6) is

$$nh^{1/2} \left| \int \{\hat{p}(x) - p_{\hat{\theta}}(x)\}^2 \{\tilde{f}(x) - f(x)\} dx \right|,$$

where $\tilde{f}(x) = (ns)^{-1} \sum_{i=1}^n \ell\{(X_i - x)/s\}$ is the standard kernel density estimator with $\tilde{f}(x) - f(x) = O_p\{(ns)^{-1/2}(\log n)^{1/2}\}$, uniformly in x . It is of order

$$nh^{1/2}(nh)^{-1} \log n (ns)^{-1/2}(\log n)^{1/2} = n^{-1/2}h^{-1/2}s^{-1/2}(\log n)^{3/2} = o(1),$$

if s is chosen such that $nhs(\log n)^{-3} \rightarrow \infty$. Hence both terms of (S.6) are of order $o_p(1)$, which completes the proof of $\tilde{\mathcal{T}}_n - \mathcal{T}_n = o_p(1)$. □

REFERENCES

- DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields.*, 75, 261-277. 85
- DU, Y. AND AKRITAS, M.G. (2002). Uniform strong representation of the conditional Kaplan–Meier process. *Math. Methods Statist.*, 11, 152-182.
- HÄRDLE, W. AND MAMMEN, E. (1993) Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21, 1926-1947.
- VAN KEILEGOM, I. AND VEREVERBEKE, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, 49, 467-491. 90
- XU, J. AND PENG, Y. (2014). Nonparametric cure rate estimation with covariates. *Canad. J. Statist.*, 42, 1-17.

~~[Received 2 January 2017. Editorial decision on 1 April 2017]~~