

# Strong uniform consistency results of the weighted average of conditional artificial data points

Cédric Heuchenne<sup>1,2</sup>

*Institut de Statistique*

*Université catholique de Louvain*

August 24, 2007

## Abstract

In this paper, we study strong uniform consistency of a weighted average of artificial data points. This is especially useful when information is incomplete (censored data, missing data ...). In this case, reconstruction of the information is often achieved nonparametrically by using a local preservation of mean criterion for which the corresponding mean is estimated by a weighted average of new data points. The present approach enlarges the possible scope for applications beyond just the incomplete data context and can also be useful to treat the estimation of the conditional mean of specific functions of complete data points. As a consequence, we establish the strong uniform consistency of the Nadaraya-Watson (1964) estimator for general transformations of the data points. This result generalizes the one of Härdle, Janssen and Serfling (1988). In addition, the strong uniform consistency of a modulus of continuity will be obtained for this estimator. Applications of those two results are detailed for some popular estimators.

KEY WORDS: Kernel estimation; Nonparametric regression; Right censoring.

<sup>1</sup> This research was supported by ‘Projet d’Actions de Recherche Concertées’, No. 98/03-217, and by the IAP research network nr. P5/24 of the Belgian government.

<sup>2</sup> The first author is now at the University of Liège, HEC-Managment School of ULg.

# 1 Introduction

In many regression contexts where the data are incomplete, one has to reconstruct missing information by using other data points. In particular, if  $Z$  denotes a data point,  $X$  the covariate and  $\Delta$  is a binary variable equal to 1 if the data point  $Z$  is complete (in this case  $Z = Y$ , the true data point) and 0 if it is incomplete, a natural way to reconstruct a function  $\varphi_t(Y|x)$  at  $X = x$  and for  $t \in I \subseteq \mathbb{R}$  is to take  $\Gamma_t(Z, \Delta|x) = (\varphi_t(Y|x))^* = E[\varphi_t(Y|x)|x, Z, \Delta] = \varphi_t(Y|x)\Delta + E[\varphi_t(Y|x)|Y > Z, x](1 - \Delta)$ . (In the case of missing data,  $Z = -\infty$  and therefore  $E[\varphi_t(Y|x)|Y > Z, x] = E[\varphi_t(Y|x)|x]$ .) In censored regression, this scheme with  $\varphi_t(Y|x) = Y$  has been used by Buckley and James (1979), Koul, Susarla and Van Ryzin (1981), Leurgans (1987), Fan and Gijbels (1994) and Heuchenne and Van Keilegom (2004) among others. In estimation with missing data, this kind of new data points has been proposed by, e.g., Cheng (1994), Chu and Cheng (1995) and Cheng and Chu (1996). As explained in Heuchenne and Van Keilegom (2005) for nonparametric estimation with censored data,  $\varphi_t(Y|x)$  can be any function of  $x, t$  and  $Y$ : e.g.,  $Y$ ,  $Y^2$  or  $I(Y \leq t)$ , for fixed  $t \in I$ , if the objective is to estimate  $E[Y|x]$ ,  $E[Y^2|x]$  or  $E[I(Y \leq t)|x] = P(Y \leq t|x)$ , respectively. Therefore, there is a need to construct a general asymptotic theory for a nonparametric estimator of  $E[\varphi_t(Y|x)|x]$  ( $E[(\varphi_t(Y|x))^*|x]$ ) in the complete (incomplete) data case.

More precisely, let  $\{\Gamma_t, t \in I\}$  be a family of real valued measurable functions on  $\mathbb{R}$  and suppose we want to estimate

$$E[\Gamma_t(Z, \Delta|x)|x] = \sum_{\delta=0,1} \int \Gamma_t(z, \delta|x) dH_{\delta}(z|x), \quad (1.1)$$

where  $I$  is a possibly infinite or degenerate interval in  $\mathbb{R}$ ,  $x \in R_X$ , a compact interval in  $\mathbb{R}$  and  $H_{\delta}(y|x) = P(Z \leq y, \Delta = \delta|x)$  ( $\delta = 0, 1$ ). A natural nonparametric estimator for this conditional mean is given by

$$\frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n}) \Gamma_t(Z_i, \Delta_i|x)}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})}, \quad (1.2)$$

where  $K(\cdot)$  is a symmetric kernel density function and  $a_n$  is a sequence of nonnegative numbers such that  $a_n \rightarrow 0$  when  $n \rightarrow \infty$ . These quantities (in the framework of this paper) will be completely specified in Section 3. For easy reference to this estimator, we call it W.A.E. (weighted average estimator). In the case  $\Gamma_t(Z, \Delta|x) = Z$ , this estimator reduces to the usual Nadaraya-Watson (1964) estimator and in the case  $\Gamma_t(Z, \Delta|x) = I(Z \leq t)$ , we obtain the Stone (1977) estimator.

The objective of Section 3 is to provide the almost sure convergence of the W.A.E. uniformly in  $x, t$  with the rate  $(na_n)^{-1/2}(\log n)^{1/2}$ . Now, suppose  $s, t \in I$  with  $|t - s| \leq d_n$ , where  $d_n$  is a sequence of nonnegative numbers such that  $d_n \rightarrow 0$  when  $n \rightarrow \infty$ . This sequence will also be completely specified later (see assumption (A10) in Section 2). In Section 4, we aim to obtain the almost sure convergence of the modulus of continuity based on the W.A.E. uniformly in  $x, s, t$ ,  $|t - s| \leq d_n$ , with the rate  $(na_n)^{-1/2}(\log n)^{1/2}d_n^{1/2}$ . The utility of these results is illustrated for some typical examples in Section 2.

## 2 Examples and Assumptions

### Example 2.1 (Nonparametric estimation of conditional location and scale functions for complete data)

Suppose  $Y_1, \dots, Y_n$  are  $n$  i.i.d. random variables corresponding to  $X_1, \dots, X_n$ ,  $n$  i.i.d. covariates with distribution  $F_X(x) = P(X_1 \leq x)$ . Let  $F(t|x) = P(Y_1 \leq t|X_1 = x)$  be the conditional distribution of the response given the covariate. Standard location and scale estimators are given by

$$\hat{m}_{ST}(x) = \int_0^1 \hat{F}^{-1}(s|x) L(s) ds, \quad \hat{\sigma}_{ST}^2(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 L(s) ds - \hat{m}_{ST}^2(x), \quad (2.1)$$

where  $\hat{F}(t|x)$  is the Stone (1977) estimator (W.A.E. with  $\Gamma_t(Z, \Delta|x) = \Gamma_t(Y, 1|x) = I(Y \leq t)$ ),  $\hat{F}^{-1}(s|x) = \inf\{t : \hat{F}(t|x) \geq s\}$  and  $L(s)$  is a given score function satisfying  $\int_0^1 L(s) ds = 1$ . If the objective is to estimate

$$\int_0^1 F^{-1}(s|x) L(s) ds \quad (2.2)$$

and

$$\int_0^1 F^{-1}(s|x)^2 L(s) ds, \quad (2.3)$$

it is clear that  $\Gamma_{t1}(Y, 1|x) = Y L(F(Y|x))$  for (2.2) and  $\Gamma_{t2}(Y, 1|x) = Y^2 L(F(Y|x))$  for (2.3) since, for monotonic non-decreasing functions  $F(\cdot|x)$ ,  $E[\Gamma_{ti}(Y, 1|x)|x]$  equals the function to estimate (2.2) for  $i = 1$  and (2.3) for  $i = 2$ . Since the data points  $\Gamma_{ti}(Y, 1|x)$  depend themselves on  $F(Y|x)$ , they are estimated by  $Y L(\hat{F}(Y|x))$  and  $Y^2 L(\hat{F}(Y|x))$  so that the W.A.E. based on those data points corresponds to (2.1).

Note that when  $L(s) = I(0 \leq s \leq 1)$ ,  $\hat{m}_{ST}(x)$  and  $\hat{\sigma}_{ST}^2(x)$  reduce to estimators of the conditional mean and variance. Theorem 3.3 of the next section thus enables us to prove at the same time the strong uniform consistency of estimators of any location and scale functions defined by the score function  $L$ . This is achieved in two steps : first, an application of Theorem 3.3 for data points  $I(Y_i \leq t)$  ( $i = 1, \dots, n$ ) in order to delete the Stone estimators in the expressions  $YL(\hat{F}(Y|x))$  and  $Y^2L(\hat{F}(Y|x))$  and, second, an application of the same theorem on the functions  $\Gamma_{t1}(Y, 1|x)$  and  $\Gamma_{t2}(Y, 1|x)$ .

**Example 2.2 (Nonparametric estimation of conditional location and scale functions for censored data)**

Now, suppose  $Y_1, \dots, Y_n$  are possibly right censored by  $C_1, \dots, C_n$   $n$  i.i.d. random variables with distribution function  $G(t|x) = P(C_1 \leq t | X = x)$ . The observed random variable for the covariate  $X_i$  is therefore the pair  $(Z_i, \Delta_i)$ ,  $i = 1, \dots, n$ , with  $Z_i = Y_i \wedge C_i$  and  $\Delta_i = I(Y_i \leq C_i)$ . We will now assume independence of  $Y_i$  and  $C_i$  conditionally on  $X_i$ . Location and scale estimators are given by

$$\hat{m}_B(x) = \int_{-\infty}^{\tilde{T}} y L(\tilde{F}(y|x)) d\tilde{F}(y|x) \quad (2.4)$$

and

$$\hat{\sigma}_B^2(x) = \int_{-\infty}^{\tilde{T}} y^2 L(\tilde{F}(y|x)) d\tilde{F}(y|x) - \hat{m}_B^2(x), \quad (2.5)$$

where  $\tilde{F}(\cdot|\cdot)$  is the Beran (1981) estimator defined as

$$\tilde{F}(t|x) = 1 - \prod_{Z_i \leq t, \Delta_i=1} \left\{ 1 - \frac{W_i(x, a_n)}{\sum_{j=1}^n I(Z_j \geq Z_i) W_j(x, a_n)} \right\} I(t < Z_{(n)}), \quad (2.6)$$

with

$$W_i(x, a_n) = \frac{K\left(\frac{x-X_i}{a_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{a_n}\right)}, \quad (2.7)$$

the Nadaraya-Watson weights,  $K(\cdot)$  and  $a_n$  defined as in (1.2) for the W.A.E. and  $L(s)$  is a given score function satisfying  $\int_0^1 L(s) ds = 1$ . In order to avoid consistency problems in the right tails of the Beran estimator,  $\tilde{T}$  is chosen smaller than  $\inf_x \tau_{H(\cdot|x)}$ , where  $H(y|x) = P(Z \leq y|x)$  and  $\tau_{F(\cdot)} = \inf\{t : F(t) = 1\}$  for some  $F$ . Seeing that the objective is to

estimate  $E[YI(Y \leq \tilde{T})L(F(Y|x))|x]$  and  $E[Y^2I(Y \leq \tilde{T})L(F(Y|x))|x]$  with an estimator of the Nadaraya-Watson type, we rewrite (2.4) and (2.5) as

$$\hat{m}_B(x) = \sum_{i=1}^n W_i(x, a_n) \hat{\Gamma}_{t3}(Z_i, \Delta_i|x), \quad (2.8)$$

and

$$\hat{\sigma}_B^2(x) = \sum_{i=1}^n W_i(x, a_n) \hat{\Gamma}_{t4}(Z_i, \Delta_i|x) - \hat{m}_B^2(x), \quad (2.9)$$

where

$$\hat{\Gamma}_{t3}(Z_i, \Delta_i|x) = Z_i I(Z_i \leq \tilde{T}) L(\tilde{F}(Z_i|x)) \Delta_i + \frac{\int_{Z_i \wedge \tilde{T}}^{\tilde{T}} y L(\tilde{F}(y|x)) d\tilde{F}(y|x)}{1 - \tilde{F}(Z_i \wedge \tilde{T}|x)} (1 - \Delta_i),$$

and

$$\hat{\Gamma}_{t4}(Z_i, \Delta_i|x) = Z_i^2 I(Z_i \leq \tilde{T}) L(\tilde{F}(Z_i|x)) \Delta_i + \frac{\int_{Z_i \wedge \tilde{T}}^{\tilde{T}} y^2 L(\tilde{F}(y|x)) d\tilde{F}(y|x)}{1 - \tilde{F}(Z_i \wedge \tilde{T}|x)} (1 - \Delta_i).$$

Note that  $\hat{\Gamma}_{t3}(Z, \Delta|x)$  and  $\hat{\Gamma}_{t4}(Z, \Delta|x)$  actually estimate

$$\Gamma_{t3}(Z, \Delta|x) = Z I(Z \leq \tilde{T}) L(F(Z|x)) \Delta + \frac{\int_{Z \wedge \tilde{T}}^{\tilde{T}} y L(F(y|x)) dF(y|x)}{1 - F(Z \wedge \tilde{T}|x)} (1 - \Delta),$$

and

$$\Gamma_{t4}(Z, \Delta|x) = Z^2 I(Z \leq \tilde{T}) L(F(Z|x)) \Delta + \frac{\int_{Z \wedge \tilde{T}}^{\tilde{T}} y^2 L(F(y|x)) dF(y|x)}{1 - F(Z \wedge \tilde{T}|x)} (1 - \Delta),$$

respectively. It is easy to check that

$$E[\Gamma_{t3}(Z, \Delta|x)|x] = E[YI(Y \leq \tilde{T})L(F(Y|x))|x],$$

and

$$E[\Gamma_{t4}(Z, \Delta|x)|x] = E[Y^2I(Y \leq \tilde{T})L(F(Y|x))|x].$$

As for the complete data case, Theorem 3.3 enables us to prove the strong uniform consistency of estimators of any location and scale functions (truncated by  $\tilde{T}$ ) defined by the score function  $L$ . Note that in order to use Theorem 3.3 with the functions  $\Gamma_{t3}(Z, \Delta|x)$  and  $\Gamma_{t4}(Z, \Delta|x)$ , we first need to delete the Beran estimators that appear in  $\hat{\Gamma}_{t3}(Z, \Delta|x)$  and  $\hat{\Gamma}_{t4}(Z, \Delta|x)$ . This can be done by using Proposition 4.3 of Van Keilegom and Akritas

(1999).

**Example 2.3 (Estimation of a conditional distribution function under the heteroscedastic model)**

Now, suppose in the previous example that we want to estimate the conditional distribution function of the response given the covariate under the model  $Y = m(X) + \sigma(X)\varepsilon$  with  $\varepsilon$  independent of  $X$ . The corresponding preservation of means criterion is: construct new indicators for which the conditional mean equals the asked conditional distribution function and which use the above heteroscedastic model. More precisely, this estimator is a weighted sum of data points  $\hat{\Gamma}_{t5}(Z_i, \Delta_i|x)$ ,  $i = 1, \dots, n$ , that approximate

$$\hat{\Gamma}_{t5}(Z_i, \Delta_i|x) = I(Z_i \leq t)\Delta_i + \frac{F_\varepsilon(\frac{t-m(X)}{\sigma(X)} \wedge T) - F_\varepsilon(\frac{Z_i-m(X)}{\sigma(X)} \wedge T)}{1 - F_\varepsilon(\frac{Z_i-m(X)}{\sigma(X)} \wedge T)}(1 - \Delta_i), \quad (2.10)$$

where  $F_\varepsilon(y) = P(\varepsilon \leq y)$ ,  $T < \tau_{H_\varepsilon(\cdot)}$  and  $H_\varepsilon(y) = P(\frac{Z-m(X)}{\sigma(X)} \leq y)$ . We refer the reader to Heuchenne and Van Keilegom (2005) for a complete description and explanation of this estimator. The same paper also provides strong uniform consistency proofs for the estimator based on those new data points and a corresponding modulus of continuity. Those proofs largely use Theorems 3.3 and 4.3.

**Example 2.4 (Nonparametric regression with missing data)**

Suppose in Example 2.1 that some  $Y_i$ ,  $i = 1, \dots, n$ , are possibly missing. In this case,  $\Delta_i = 0$  if  $Y_i$  is a missing data and  $\Delta_i = 1$  otherwise. Moreover, the MAR (missing at random) assumption requires that

$$P(\Delta = 1|X, Y) = P(\Delta = 1|X) = p(X) \quad (2.11)$$

(see Little and Rubin, 1987, p.14). In this context, a simple idea (similar to the one developed by Chu and Cheng, 1995) to estimate a regression function is to construct a Nadaraya-Watson estimator with new data points given by

$$\hat{Y}_i^* = Y_i\Delta_i + \hat{m}_S(X_i)(1 - \Delta_i), \quad i = 1, \dots, n,$$

where  $\hat{m}_S(x)$  is the Nadaraya-Watson estimator based on the complete pairs:

$$\frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n})Y_i\Delta_i}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})\Delta_i}$$

with  $K(\cdot)$  and  $a_n$  as defined before. Therefore, two applications of Theorem 3.3 with data points  $\Gamma_{t6}(Z, \Delta|x) = Y\Delta$  and  $\Gamma_{t7}(Z, \Delta|x) = \Delta$  along with assumption (2.11) allow us to prove the uniform strong consistency of  $\hat{m}_S(x)$ . Next, if  $f_X(x) = \frac{dF_X(x)}{dx}$  and  $p(x)$  are uniformly Lipschitz continuous and  $m_S(x) = E[Y|X = x]$  is two times continuously differentiable, the uniform strong consistency of the W.A.E. with data points  $\Gamma_{t8}(Z, \Delta|x) = Y_i^* = Y_i\Delta_i + m_S(X_i)(1 - \Delta_i)$  is obtained in two steps. First, replace  $m_S(X_i)$  by  $m_S(x) + (X_i - x)m'_S(x) + O(a_n^2)$  (using appropriate assumptions on the support of  $K$  satisfied for example by the assumptions of Theorem 3.3). Then, by similar developments as in Corollary 1 (ii) of Theorem 2 in Masry (1996),  $m'_S(x) \sum_{i=1}^n W_i(x, a_n)(X_i - x)(1 - \Delta_i) = O(a_n^2)$  a.s. Second, a third application of Theorem 3.3 allows us to obtain the result.

The assumptions we need for the proofs of the results of Sections 3 and 4 are listed below.

(A1)  $\gamma_t(\cdot, \cdot|\cdot)$  is Lipschitz on  $R_X$  (compact) uniformly in  $t \in I$  :

$$\sup_{|x-x_j| \leq d, x, x_j \in R_X} \sup_{t \in I} |\gamma_t(z, \delta|x) - \gamma_t(z, \delta|x_j)| \leq L_0(z, \delta|x_j)d, \quad z \in \mathbb{R}, \delta = 0, 1,$$

where  $L_0(\cdot, \cdot|\cdot)$  is a (positive) function independent of  $t$  such that  $E[L_0(Z, \Delta|x)^6] \leq L_6 < \infty$  for all  $x \in R_X$ .

(A2)  $0 \leq \gamma_t(z, \delta|x) \leq \gamma_{t'}(z, \delta|x)$ ,  $t < t' \in I$ , for all  $x, z$  and  $\delta = 0, 1$ .

(A3)  $g(t|x) = E[\gamma_t(Z, \Delta|x)]$  is a continuous function of  $t \in I$  for all  $x$ .

(A4) For  $t_* = \inf\{t : t \in I\}$ ,  $t_* \geq -\infty$  and  $t^* = \sup\{t : t \in I\}$ ,  $t^* \leq \infty$ , the limit functions  $\gamma_{t^*} = \lim_{t \rightarrow t^*} \gamma_t$  and  $\gamma_{t_*} = \lim_{t \rightarrow t_*} \gamma_t$  exist and are finite a.s. (w.r.t.  $H(z) = P(Z \leq z)$ ) for all  $x$ .

(A5) There is a  $\lambda \in ]2, +\infty[$  such that, for all  $x$ ,  $E[\gamma_{t^*}(Z, \Delta|x)^{6\lambda}] \leq M_{6\lambda} < \infty$ ; in the case  $\lambda = +\infty$ ,  $\sup_{x, z, \delta} |\gamma_{t^*}(z, \delta|x)| < \infty$ .

(A6) Let  $\{c_n\}$  be a nonnegative sequence satisfying (i)  $0 \leq c_n \rightarrow 0$ , (ii)  $\Psi_n = nc_n/\log n \rightarrow \infty$ , (iii)  $c_n^{-1} \leq (n/\log n)^{1-2/\lambda}$ , for  $\lambda$  as in (A5).

(A7)(i)  $F_X(x)$  is differentiable with respect to  $x$  with derivative  $f_X(x)$ .

(ii)  $H_\delta(x, y) = P(X \leq x, Z \leq y, \Delta = \delta)$ ,  $x \in R_X$ ,  $y \in \mathbb{R}$ ,  $\delta = 0, 1$ , is differentiable with respect to  $(x, y)$ .

(iii)  $H_\delta(y) = P(Z \leq y, \Delta = \delta)$ ,  $y \in \mathbb{R}$ ,  $\delta = 0, 1$ , is differentiable with respect to  $y$ .

(iv) For the density  $f_{X|Z,\Delta}(x|z, \delta)$  of  $X$  given  $(Z, \Delta)$ ,  $\sup_{x,z} |f_{X|Z,\Delta}(x|z, \delta)| < \infty$ ,  $\sup_{x,z} |\dot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$  and  $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$  ( $\delta = 0, 1$ ), where  $\dot{f}_{X|Z,\Delta}(x|z, \delta)$  ( $\ddot{f}_{X|Z,\Delta}(x|z, \delta)$ ) denotes the first (second) derivative of  $f_{X|Z,\Delta}(x|z, \delta)$  with respect to  $x$ .

(A8) Define new data points as  $\Gamma_t(z, \delta|x) = \sum_{i=1}^{i_0} q_i \gamma_{ti}(z, \delta|x)$ ,  $z \in \mathbb{R}$ ,  $t \in I$ ,  $x \in R_X$ ,  $\delta = 0, 1$ , with fixed and finite  $i_0, q_1, \dots, q_{i_0}$  and with families  $\{\gamma_{ti}, t \in I\}$ ,  $1 \leq i \leq i_0$ , satisfying assumptions (A1)-(A5), with common  $\lambda$  in (A5).

(A9)(i) Consider kernel sequences of step-function form,  $K_n(u) = \sum_{j=1}^{j_n} m_{nj} I(-b_{nj} \leq u \leq b_{nj})$ ,  $u \in \mathbb{R}$ , with  $\{j_n\}$ ,  $\{m_{nj}\}$  and  $\{b_{nj}\}$ , some sequences of constants ( $\{j_n\}$  and  $\{b_{nj}\}$  nonnegative) such that  $|2 \sum_{j=1}^{j_n} m_{nj} b_{nj} - 1| = O(\max(\Psi_n^{-1/2}, a_n^2))$ , with  $j_n = O(n^s)$ ,  $s > 0$  and  $\Psi_n = na_n / \log n$ .

(ii)  $\sup_n \sum_{j=1}^{j_n} |m_{nj}| b_{nj}^{1/2} < \infty$ .

(iii)  $\sup_n \sum_{j=1}^{j_n} |m_{nj}| b_{nj}^3 < \infty$ .

(A10) Let  $\{c_n\}$  and  $\{d_n\}$  two nonnegative sequences that satisfy (i)  $0 \leq c_n, d_n \rightarrow 0$ , (ii)  $\Psi_n = nc_n / \log n \rightarrow \infty$ , (iii)  $c_n^{-1} \leq d_n(n/\log n)^{1-2/\lambda}$  for  $\lambda$  as in (A5).

(A11) The data points  $\gamma_t(Z_i, \Delta_i|x)$ ,  $t \in I$ ,  $x \in R_X$ ,  $i = 1, \dots, n$ , have the following mean-Lipshitz properties when  $d_n \rightarrow 0$ :

(i)  $\sup_{\{x \in R_X, |t-s| \leq d_n, s, t \in I\}} |E[\gamma_t(Z, \Delta|x) - \gamma_s(Z, \Delta|x)]| \leq C_L d_n$ ,

(ii)  $\sup_{\{x \in R_X, |t-s| \leq d_n, s, t \in I\}} E[(\gamma_t(Z, \Delta|x) - \gamma_s(Z, \Delta|x))^2] \leq C_{L_2} d_n$ , for  $n$  sufficiently large.

(A12)(i) – (iii) Consider kernel sequences of the same form and with the same assumptions as in (A9) except that  $|2 \sum_{j=1}^{j_n} m_{nj} b_{nj} - 1| = O(\max(\Psi_n^{-1/2} d_n^{-1/2}, a_n^2))$  in (A9) (i) with  $d_n$  as in (A10).

### 3 Strong uniform consistency of the weighted average of artificial data points

Strong uniform consistency and modulus of continuity proofs are achieved in three steps. First, we consider new data points  $\gamma_t(Z_i, \Delta_i|x)$ ,  $i = 1, \dots, n$ ,  $t \in I$ ,  $x \in R_X$ , and kernels that are defined by indicators. Second, we combine those data points to obtain the  $\Gamma_t(Z_i, \Delta_i|x)$  used in the previous section and we sum indicators to construct kernels of step-function form. Third, by using a number of indicators that tends to infinity in the

step-function kernel, we show the announced results for the usual smooth kernels.

**Proposition 3.1** *Assume (A6), (A7). Then,*

$$P(M_n(c_n) > C_0 \Psi_n^{-1/2} + C_1 c_n^2) = O(n^{-2})$$

for some  $C_0, C_1 > 0$ , where

$$\begin{aligned} M_n(c_n) &= \sup_{x \in R_X} \sup_{t \in I} \left| \frac{1}{2nc_n} \sum_{i=1}^n \gamma_t(Z_i, \Delta_i | x) I(x - c_n < X_i \leq x + c_n) \right. \\ &\quad \left. - \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t(z, \delta | x) h_{\delta}(x, z) dz \right|, \end{aligned}$$

$h_{\delta}(x, z)$  is the joint density of  $X$  and  $Z$  for  $\delta = 0, 1$  and  $\gamma_t(z, \delta | x)$ ,  $t \in I$ ,  $x \in R_X$ ,  $z \in \mathbb{R}$ ,  $\delta = 0, 1$ , satisfy assumptions (A1)-(A5).

**Proposition 3.2** *Assume (A7)-(A9) and that  $a_n$  satisfies (i)  $a_n B_n \rightarrow 0$ , (ii)  $na_n b_n / \log n \rightarrow \infty$  and (iii)  $a_n^{-1} \leq b_n (n / \log n)^{1-2/\lambda}$ , where  $b_n = \min_{j \leq j_n} b_{nj}$ ,  $B_n = \max_{j \leq j_n} b_{nj}$  and  $\lambda$  is given as in (A5). Then,*

$$\sup_{x \in R_X} \sup_{t \in I} |d_{tn}(x) - d_t(x)| = O(\max(\Psi_n^{-1/2}, a_n^2)) \text{ a.s.},$$

where

$$d_{tn}(x) = \frac{1}{na_n} \sum_{i=1}^n \Gamma_t(Z_i, \Delta_i | x) K_n\left(\frac{x - X_i}{a_n}\right),$$

and

$$d_t(x) = \sum_{\delta=0,1} \int_{-\infty}^{\infty} \Gamma_t(z, \delta | x) h_{\delta}(x, z) dz.$$

**Theorem 3.3** *Assume (A7), (A8). For the sequence  $a_n$ , we suppose (i)  $a_n \rightarrow 0$ , (ii)  $na_n^{5/2} / \log n \rightarrow \infty$ , (iii)  $a_n^{-5/2} \leq (n / \log n)^{1-2/\lambda}$ , where  $\lambda$  is given as in (A5) and (iv)  $na_n^4 \rightarrow 0$ . The kernel  $K$  is assumed to be symmetric with bounded support, bounded first derivative and  $\int K(u) du = 1$ . Then,*

$$\sup_{x \in R_X} \sup_{t \in I} |d_{tn}(x) - d_t(x)| = O(\Psi_n^{-1/2}) \text{ a.s.},$$

where  $d_{tn}(x)$  and  $d_t(x)$  are defined with kernel  $K$  and  $\Psi_n = na_n / \log n$ . Moreover, if  $\inf_{x \in R_X} |f_X(x)| > 0$ ,

$$\sup_{x \in R_X} \sup_{t \in I} \left| \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{a_n}\right) \Gamma_t(Z_i, \Delta_i | x)}{\sum_{i=1}^n K\left(\frac{x - X_i}{a_n}\right)} - \frac{d_t(x)}{f_X(x)} \right| = O(\Psi_n^{-1/2}) \text{ a.s.}$$

**Remark 3.4 (density estimator)** If we denote  $f_{nX}(x) = (1/na_n) \sum_{i=1}^n K(\frac{x-X_i}{a_n})$ , the classical kernel density estimator, we have using Theorem 3.3 with  $\Gamma(Z_i, \Delta_i|x) = 1$  that  $\sup_{x \in R_X} |f_{nX}(x) - f_X(x)| = O(\Psi_n^{-1/2})$  a.s., since  $\sup_{x \in R_X} |f_X(x)| < \infty$ .

**Remark 3.5 (moment conditions)** For a number of artificial data points, the moment conditions in (A1) and (A5) are not used. Indeed, those data points can often be of the form  $\gamma_{t^*}(Z_i, \Delta_i|x) \leq \gamma_{t^*}^*(Z_i, \Delta_i)$  and such that  $L_0(Z_i, \Delta_i|x) \leq L_0^*(Z_i, \Delta_i)$ . In this case, the strong law of large numbers can be immediately used with  $(1/n) \sum_{i=1}^n \gamma_{t^*}^{*\lambda}(Z_i, \Delta_i) - E[\gamma_{t^*}^{*\lambda}(Z, \Delta)]$  and  $(1/n) \sum_{i=1}^n L_0^*(Z_i, \Delta_i) - E[L_0^*(Z, \Delta)]$  in the appendix. The terms  $V_n \Psi_{nj}^{-1/2}$  and  $2W_n \Psi_{nj}^{-1/2}$  can then be treated outside Proposition 3.1 and be directly introduced in (A.21) in the proof of Proposition 3.2 (see the appendix) such that the final result of Theorem 3.3 is preserved.

**Remark 3.6 (boundary effects)** The degree of smoothing of  $f_{X|Z,\Delta}(x|z, \delta)$  allows us via (A7) (iv) to obtain the artificial order  $O(\Psi_n^{-1/2})$  near the boundaries of  $R_X$ . If we suppose for instance the weaker condition

$$\sup_{|x-x'| \leq d, x, x' \in R_X} \sup_{t \in I} \left| \sum_{\delta=0,1} \int \gamma_t(z, \delta|x) (h_\delta(x', z) - h_\delta(x, z)) dz \right| \leq Cd,$$

instead of (A7) (iv), then the more realistic rate  $O(a_n)$  can be obtained near the boundaries.

**Remark 3.7 (bandwidth assumptions)** The bandwidth parameter  $a_n$  could tend to zero more slowly. Indeed, the condition  $na_n^4 \rightarrow 0$  of Theorem 3.3 can be written with another power on  $a_n$ . By example, if  $na_n^5(\log n)^{-1} = O(1)$ , Theorem 3.3 also holds if  $na_n^3/\log n \rightarrow \infty$  and  $a_n^{-3} \leq (n/\log n)^{1-2/\lambda}$ .

**Remark 3.8 (artificial data representation)** The representation

$$\Gamma_t(z, \delta|x) = \sum_{i=1}^{i_0} q_i \gamma_{ti}(z, \delta|x),$$

needed in the above proofs, requires nonnegative  $\gamma_{ti}(z, \delta|x)$ ,  $i = 1, \dots, i_0$ . This assumption is not restrictive since any random variable  $X$  with real values can be represented by  $X = \max(X, 0) - (-\min(X, 0))$ , where the two terms of this difference are nonnegative.

**Remark 3.9 (Extension to local linear estimator with conditional new data points)** The extension of Theorem 3.3 to local linear estimator is easily obtained by similar developments as in Corollary 1 (ii) of Theorem 2 in Masry (1996) and if  $f_X(x)$  is uniformly Lipschitz continuous. Indeed, using those arguments the local linear estimator reduces to the classical weighted sum of conditional new data points discussed above.

## 4 Modulus of continuity for the weighted average of conditional synthetic data points

The development of this section is similar to Section 3. The strong uniform consistency of the modulus of continuity is established via two preliminary results.

**Proposition 4.1** *Assume (A7), (A10). Then,*

$$P(M_n(c_n) > C_0 \Psi_n^{-1/2} d_n^{1/2} + C_1 c_n^2 d_n) = O(n^{-2}),$$

for some  $C_0, C_1 > 0$ , where

$$\begin{aligned} M_n(c_n) &= \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} \left| \frac{1}{2nc_n} \sum_{i=1}^n (\gamma_t(Z_i, \Delta_i|x) - \gamma_s(Z_i, \Delta_i|x)) I(x - c_n < X_i \leq x + c_n) \right. \\ &\quad \left. - \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta|x) - \gamma_s(z, \delta|x)) h_{\delta}(x, z) dz \right|, \end{aligned}$$

and  $\gamma_t(z, \delta|x)$ ,  $t \in I$ ,  $x \in R_X$ ,  $z \in \mathbb{R}$ ,  $\delta = 0, 1$ , satisfies assumptions (A1)-(A5) and (A11).

**Proposition 4.2** *Assume (A7), (A8), (A11), (A12) and that  $a_n$  and  $d_n$  satisfy (i)  $a_n B_n \rightarrow 0$ ,  $d_n \rightarrow 0$ , (ii)  $na_n b_n / \log n \rightarrow \infty$  and (iii)  $a_n^{-1} \leq b_n d_n (n / \log n)^{1-2/\lambda}$ , where  $b_n = \min_{j \leq j_n} b_{nj}$ ,  $B_n = \max_{j \leq j_n} b_{nj}$ , and  $\lambda$  is given as in (A5). Then,*

$$\sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} |d_{stn}(x) - d_{st}(x)| = O(\max(\Psi_n^{-1/2} d_n^{1/2}, a_n^2 d_n)) \text{ a.s.},$$

where

$$d_{stn}(x) = \frac{1}{na_n} \sum_{i=1}^n (\Gamma_t(Z_i, \Delta_i|x) - \Gamma_s(Z_i, \Delta_i|x)) K_n\left(\frac{x - X_i}{a_n}\right),$$

$$d_{st}(x) = \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\Gamma_t(z, \delta|x) - \Gamma_s(z, \delta|x)) h_{\delta}(x, z) dz.$$

**Proof.** The proof is along the same lines as the proof of Proposition 3.2.

**Theorem 4.3** *Assume (A7), (A8), (A11) and that  $a_n$  and  $d_n$  satisfy (i)  $a_n \rightarrow 0$ ,  $d_n \rightarrow 0$ , (ii)  $na_n^{5/2}d_n^{-1/2}/\log n \rightarrow \infty$ , (iii)  $a_n^{-5/2} \leq d_n^{1/2}(n/\log n)^{1-2/\lambda}$ , where  $\lambda$  is given as in (A5), (iv)  $\log n/na_n d_n = O(1)$  and (v)  $na_n^4 \rightarrow 0$ . The kernel  $K$  is symmetric with bounded support, bounded first derivative and  $\int K(u)du = 1$ . Then,*

$$\sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} |d_{stn}(x) - d_{st}(x)| = O(\Psi_n^{-1/2} d_n^{1/2}) \text{ a.s.},$$

where  $d_{stn}(x)$  and  $d_{st}(x)$  are defined with kernel  $K$  and  $\Psi_n = na_n/\log n$ . Moreover, if  $na_n^{5/2}/\log n \rightarrow \infty$  and  $\inf_{x \in R_X} |f_X(x)| > 0$ ,

$$\sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} \left| \frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n}) \Gamma_{ts}(Z_i, \Delta_i | x)}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})} - \frac{d_{st}(x)}{f_X(x)} \right| = O(\Psi_n^{-1/2} d_n^{1/2}) \text{ a.s.},$$

where  $\Gamma_{ts}(Z_i, \Delta_i | x) = \Gamma_t(Z_i, \Delta_i | x) - \Gamma_s(Z_i, \Delta_i | x)$ .

**Remark 4.4 (bandwidth assumptions)** If  $na_n^5(\log n)^{-1} = O(1)$ , Theorem 4.3 also holds if  $na_n^3/\log n \rightarrow \infty$  and  $a_n^{-3} \leq d_n^{1/2}(n/\log n)^{1-2/\lambda}$ .

## Appendix : Proofs of main results

**Proof of Proposition 3.1.** Let  $f_n = \Psi_n^{-1/2} c_n$ . We have, for  $M_{0n}(c_n) = M_n(c_n)$  in Proposition 3.1,

$$\begin{aligned} M_{0n}(c_n) &= \sup_{x \in R_X} \sup_{t \in I} \left| \frac{1}{2nc_n} \sum_{i=1}^n \gamma_t(Z_i, \Delta_i | x) I(x - c_n < X_i \leq x + c_n) \right. \\ &\quad \left. - \frac{1}{2c_n} \int_{x-c_n}^{x+c_n} \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t(z, \delta | x) h_{\delta}(u, z) dz du \right| \\ &\quad + \sup_{x \in R_X} \sup_{t \in I} \left| \frac{1}{2c_n} \int_{x-c_n}^{x+c_n} \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t(z, \delta | x) h_{\delta}(u, z) dz du \right. \\ &\quad \left. - \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t(z, \delta | x) h_{\delta}(x, z) dz \right| \\ &= \sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x)| + \sup_{x \in R_X} \sup_{t \in I} |M_{2tn}(x)|. \end{aligned}$$

First, we treat the term  $M_{2tn}(x)$ . It is given by

$$\sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t(z, \delta | x) \left\{ \frac{1}{2c_n} \int_{x-c_n}^{x+c_n} h_{\delta}(u, z) du - h_{\delta}(x, z) \right\} dz.$$

Using two Taylor developments of order three around  $x$ , we get

$$\frac{1}{2c_n} \int_{x-c_n}^{x+c_n} h_\delta(u, z) du - h_\delta(x, z) = (c_n^2/12)[\ddot{f}_{X|Z,\Delta}(\theta_1|z, \delta) + \ddot{f}_{X|Z,\Delta}(\theta_2|z, \delta)]h_\delta(z),$$

where  $\theta_1$  ( $\theta_2$ ) is between  $x + c$  and  $x$  ( $x$  and  $x - c$ ). Since  $\sup_{x,z} |\ddot{f}_{X|Z,\Delta}(x|z, \delta)| < \infty$  ( $\delta = 0, 1$ ) and  $\sup_{\{x \in R_X, t \in I\}} E[\gamma_t(Z, \Delta|x)] < \infty$  with  $\gamma_t(Z, \Delta|x) \geq 0$ ,

$$\sup_{\{x \in R_X, t \in I\}} |M_{2tn}(x)| \leq C_1 c_n^2. \quad (\text{A.1})$$

Let  $L_X$  be the length of  $R_X$  and divide  $R_X$  into  $[\frac{2L_X}{f_n}]$  intervals of length smaller than or equal to  $f_n$  ( $[x]$  denotes the integer part of  $x$ ). Define  $x_0 = \inf\{x : x \in R_X\}$  and let  $I_X$  be the set of points  $\{x_k = x_0 + k[\frac{2L_X}{f_n}]^{-1}L_X, 1 \leq k \leq [\frac{2L_X}{f_n}] - 1 = L_X^n\}$  and  $x_{L_X^n+1} = \sup\{x : x \in R_X\}$  which limit the intervals. Using the Lipschitz condition (A1), we can rewrite for  $1 \leq j \leq L_X^n$ ,

$$\begin{aligned} & \sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x)| \\ & \leq \max_{x_j \in I_X} \sup_{x \in [x_{j-1}, x_{j+1}]} \sup_{t \in I} \left| \frac{1}{2nc_n} \sum_{i=1}^n \gamma_t(Z_i, \Delta_i|x_j) I(x - c_n < X_i \leq x + c_n) \right. \\ & \quad \left. - \frac{1}{2c_n} \int_{x-c_n}^{x+c_n} \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t(z, \delta|x_j) h_\delta(u, z) dz du \right| + \max_{x_j \in I_X} (E[L_0(Z, \Delta|x_j)] + |V_n(x_j)|) \Psi_n^{-1/2}, \end{aligned} \quad (\text{A.2})$$

where  $V_n(x_j) = (1/2n) \sum_{i=1}^n L_0(Z_i, \Delta_i|x_j) - (E[L_0(Z, \Delta|x_j)]/2)$ . For simplicity, we rewrite the term on the right hand side of the above equality as

$$(2c_n)^{-1} \max_{x_j \in I_X} \sup_{x \in [x_{j-1}, x_{j+1}]} \sup_{t \in I} |M_{3tn}(x_j, x)| + \max_{x_j \in I_X} (E[L_0(Z, \Delta|x_j)] + |V_n(x_j)|) \Psi_n^{-1/2}.$$

We then have

$$\begin{aligned} & P(\max_{x_j \in I_X} (E[L_0(Z, \Delta|x_j)] + |V_n(x_j)|) > 2C_2) \\ & \leq \sum_j \{P(|2V_n(x_j)| > 2C_2) + P(E[L_0(Z, \Delta|x_j)] > C_2)\}, \end{aligned}$$

where the second term on the right hand side of the above expression is zero when

$$C_2 > L_6^{1/6}. \quad (\text{A.3})$$

For the first term, we use an extension of Chebyshev's inequality :

$$\begin{aligned} & P\left(|n^{-1} \sum_{i=1}^n L_0(Z_i, \Delta_i|x_j) - E[L_0(Z, \Delta|x_j)]| > 2C_2\right) \\ & \leq \frac{1}{(2nC_2)^6} E[\{\sum_{i=1}^n (L_0(Z_i, \Delta_i|x_j) - E[L_0(Z, \Delta|x_j)])\}^6] = O(n^{-3}), \end{aligned}$$

since  $E[L_0(Z, \Delta|x_j)|^6] \leq L_6 < \infty$ . Then, with  $L_X^n O(n^{-3}) = o(n^{-2})$ ,

$$P(\max_{x_j \in I_X} (E[L_0(Z, \Delta|x_j)] + |V_n(x_j)|) > 2C_2) = o(n^{-2}),$$

for which, using the Borel-Cantelli Lemma, we obtain

$$V_n = \max_{x_j \in I_X} (E[L_0(Z, \Delta|x_j)] + |V_n(x_j)|) = O(1) \text{ a.s.} \quad (\text{A.4})$$

To treat the first term on the right hand side of (A.2), we introduce some additional notation. Let

$$G_{tn}(x_j, x) = n^{-1} \sum_{i=1}^n \gamma_t(Z_i, \Delta_i|x_j) I(X_i \leq x),$$

and

$$G_t(x_j, x) = E[G_{tn}(x_j, x)] = \int_{x_0}^x \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t(z, \delta|x_j) h_{\delta}(u, z) dz du.$$

Therefore,

$$\begin{aligned} |M_{3tn}(x_j, x)| &= |G_{tn}(x_j, x + c_n) - G_{tn}(x_j, x - c_n) - [G_t(x_j, x + c_n) - G_t(x_j, x - c_n)]| \\ &\leq 2 \sup_{|z| \leq c_n} |G_{tn}(x_j, x + z) - G_{tn}(x_j, x) - [G_t(x_j, x + z) - G_t(x_j, x)]| \\ &= 2 \sup_{|z| \leq c_n} M_{4tn}(x_j, x, z). \end{aligned} \quad (\text{A.5})$$

By conditions (A2) – (A5), the functions  $g(t|x_j)$ ,  $j = 1, \dots, L_X^n$ , are nondecreasing, continuous in  $t$  with finite limits  $g(t^*|x_j)$  and  $g(t_*|x_j)$  as  $t \rightarrow t^*$  and  $t_*$ . For each  $x_j$ ,  $j = 1, \dots, L_X^n$ , define  $I_{nj}$  as the grid of values of  $t$ ,  $\{t_*, t_1, \dots, t_{N_{nj}}, t^*\}$  which is such that  $|g(t_1|x_j) - g(t_*|x_j)| \leq f_n$ ,  $|g(t_{k+1}|x_j) - g(t_k|x_j)| \leq f_n$ , for  $k = 1, \dots, N_{nj} - 1$ ,  $|g(t^*|x_j) - g(t_{N_{nj}}|x_j)| \leq f_n$ . Clearly,  $I$  is divided into  $O(f_n^{-2})$  intervals. Next, let  $I_{nj}^*$  be the set  $\{(t_*, t_1), (t_1, t_2), \dots, (t_{N_{nj}}, t^*)\}$ . Clearly, the cardinality  $N_{nj} + 1$  of  $I_{nj}^*$  is bounded by

$$\frac{2(g(t^*|x_j) - g(t_*|x_j))}{f_n}. \quad (\text{A.6})$$

Also, for fixed  $j, x, z, n$ , the functions  $G_{tn}(x_j, x + z) - G_{tn}(x_j, x)$  and  $G_t(x_j, x + z) - G_t(x_j, x)$  are monotone in  $t$  and have finite limits as  $t \rightarrow t_*$ ,  $t^*$ . We therefore have

$$\begin{aligned} M_{4tn}(x_j, x, z) &\leq \max_{t \in I_{nj}} M_{4tn}(x_j, x, z) \\ &+ \max_{(s,t) \in I_{nj}^*} |G_t(x_j, x + z) - G_s(x_j, x + z) - [G_t(x_j, x) - G_s(x_j, x)]|. \end{aligned} \quad (\text{A.7})$$

It is easily shown that the second term on the right hand side of the above expression is bounded by

$$\begin{aligned}
& 2 \int_{R_X} \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta|x_j) - \gamma_s(z, \delta|x_j)) h_{\delta}(u, z) dz du \\
&= 2 \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta|x_j) - \gamma_s(z, \delta|x_j)) h_{\delta}(z) dz \\
&\leq 2(g(t|x_j) - g(s|x_j)) \leq 2f_n,
\end{aligned}$$

using monotonicity of  $\gamma_t$  with respect to  $t$ . Therefore,

$$\begin{aligned}
& \max_{x_j \in I_X} \sup_{x \in [x_{j-1}, x_{j+1}]} \sup_{t \in I} \sup_{|z| \leq c_n} M_{4tn}(x_j, x, z) \\
&\leq \max_{x_j \in I_X} \sup_{x \in [x_{j-1}, x_{j+1}]} \max_{t \in I_{nj}} \sup_{|z| \leq c_n} M_{4tn}(x_j, x, z) + 2f_n. \tag{A.8}
\end{aligned}$$

We have

$$\begin{aligned}
& \max_{x_j \in I_X} \sup_{x \in [x_{j-1}, x_{j+1}]} \max_{t \in I_{nj}} \sup_{|z| \leq c_n} M_{4tn}(x_j, x, z) \\
&\leq \max \left( \max_{j \in I_X} \sup_{x \in [x_{j-1}, x_{j+1}]} \max_{t \in I_{nj}} \sup_{x+z \in [x_{j-1}, x_{j+1}]} M_{4tn}(x_j, x, z), \right. \\
&\quad \max_{x_j \in I_X \setminus \{x_{L_X^n}\}} \sup_{x \in [x_j, x_{j+1}]} \max_{t \in I_{nj}} \sup_{x+z \in [x_{j+1}, x+c_n]} M_{4tn}(x_j, x, z), \\
&\quad \max_{x_j \in I_X \setminus \{x_1\}} \sup_{x \in [x_j, x_{j+1}]} \max_{t \in I_{nj}} \sup_{x+z \in [x-c_n, x_{j-1}]} M_{4tn}(x_j, x, z), \\
&\quad \max_{x_j \in I_X \setminus \{x_{L_X^n}\}} \sup_{x \in [x_{j-1}, x_j]} \max_{t \in I_{nj}} \sup_{x+z \in [x_{j+1}, x+c_n]} M_{4tn}(x_j, x, z), \\
&\quad \left. \max_{x_j \in I_X \setminus \{x_1\}} \sup_{x \in [x_{j-1}, x_j]} \max_{t \in I_{nj}} \sup_{x+z \in [x-c_n, x_{j-1}]} M_{4tn}(x_j, x, z) \right). \tag{A.9}
\end{aligned}$$

Introducing  $G_{tn}(x_j, x_{j+k})$  and  $G_t(x_j, x_{j+k})$  for  $k = -1, 0$  or  $1$ , it is easily shown that

$$\begin{aligned}
& \max_{x_j \in I_X} \sup_{x \in [x_{j-1}, x_{j+1}]} \max_{t \in I_{nj}} \sup_{|z| \leq c_n} M_{4tn}(x_j, x, z) \\
&\leq 2 \max_{x_j \in I_X} \max_{t \in I_{nj}} \max_{k \in \{-1, 0, 1\}} \sup_{|z| \leq c_n} M_{4tn}(x_j, x_{j+k}, z). \tag{A.10}
\end{aligned}$$

Now, put  $Q_n = M_{\lambda} f_n^{-1/(\lambda-1)}$ , where, for all  $x$ ,  $(E[(\gamma_{t^*}(Z, \Delta|x))^\lambda])^{1/\lambda} \leq M_{\lambda} < \infty$  for some  $\lambda$ ,  $2 < \lambda < \infty$ . In the case  $\lambda = \infty$ ,  $M_{\infty}$  denotes then  $\sup_{x, z, \delta} |\gamma_{t^*}(z, \delta|x)|$ . Also, put

$$H_{tn}(x_j, x) = n^{-1} \sum_{i=1}^n \gamma_t(Z_i, \Delta_i|x_j) I(\gamma_t(Z_i, \Delta_i|x_j) \leq Q_n) I(X_i \leq x),$$

and define  $M_{5tn}(x_j, x, z)$  by substitution of  $H_{tn}$  for  $G_{tn}$  and  $E[H_{tn}]$  for  $G_t$  in  $M_{4tn}(x_j, x, z)$ . Recalling (A.2), (A.4), (A.5), (A.8) and (A.10), this is seen to yield

$$\begin{aligned} \sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x)| &\leq 2c_n^{-1} \max_{x_j \in I_X} \max_{t \in I_{nj}} \max_{k \in \{-1, 0, 1\}} \sup_{|z| \leq c_n} M_{5tn}(x_j, x_{j+k}, z) \\ &\quad + 2f_n c_n^{-1} (1 + V_n/2 + W_n + \theta_n), \end{aligned} \quad (\text{A.11})$$

where

$$W_n = f_n^{-1} \max_{x_j \in I_X} \max_{t \in I_{nj}} \max_{k \in \{-1, 0, 1\}} \sup_{|z| \leq c_n} |M_{6tn}(x_j, x_{j+k}, z)|,$$

$$M_{6tn}(x_j, x, z) = G_{tn}(x_j, x+z) - G_{tn}(x_j, x) - [H_{tn}(x_j, x+z) - H_{tn}(x_j, x)],$$

and

$$\theta_n = f_n^{-1} \max_{x_j \in I_X} \max_{t \in I_{nj}} \max_{k \in \{-1, 0, 1\}} \sup_{|z| \leq c_n} |E[M_{6t}(x_j, x_{j+k}, z)]|.$$

Using (A2), (A4) and the fact that  $f_n^{-1} = (Q_n/M_\lambda)^{\lambda-1}$ , we have

$$\begin{aligned} M_\lambda^{\lambda-1} W_n &\leq Q_n^{\lambda-1} \max_{x_j \in I_X} n^{-1} \sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i | x_j) I(\gamma_{t^*}(Z_i, \Delta_i | x_j) > Q_n) \\ &\leq \max_{x_j \in I_X} n^{-1} \sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i | x_j)^\lambda, \end{aligned}$$

if  $\lambda < \infty$  and  $W_n = 0$  if  $\lambda = \infty$ . Next, for  $\lambda < \infty$ ,

$$\begin{aligned} &P\left(\max_{x_j \in I_X} n^{-1} \sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i | x_j)^\lambda > C_3\right) \\ &\leq \sum_j \{P(n^{-1} \sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i | x_j)^\lambda - E[\gamma_{t^*}(Z, \Delta | x_j)^\lambda] > C_3/2) \\ &\quad + P(E[\gamma_{t^*}(Z, \Delta | x_j)^\lambda] > C_3/2)\}, \end{aligned}$$

where the second term on the right hand side of the above expression is zero when

$$C_3/2 > M_\lambda^\lambda. \quad (\text{A.12})$$

For the first term, we also use the extension of Chebyshev's inequality :

$$\begin{aligned} &P\left(n^{-1} \sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i | x_j)^\lambda - E[\gamma_{t^*}(Z, \Delta | x_j)^\lambda] > C_3/2\right) \\ &\leq \frac{64}{(nC_3)^6} E[\{\sum_{i=1}^n (\gamma_{t^*}(Z_i, \Delta_i | x_j)^\lambda - E[\gamma_{t^*}(Z, \Delta | x_j)^\lambda])\}^6] = O(n^{-3}), \end{aligned}$$

since  $E[\gamma_{t^*}(Z, \Delta|x_j)^{6\lambda}] \leq M_{6\lambda} < \infty$ . Then, with  $L_X^n O(n^{-3}) = o(n^{-2})$ ,

$$P\left(\max_{x_j \in I_X} n^{-1} \sum_{i=1}^n \gamma_{t^*}(Z_i, \Delta_i|x_j)^\lambda > C_3\right) = o(n^{-2}),$$

for which, using the Borel-Cantelli Lemma, we obtain

$$W_n = O(1) \text{ a.s.} \quad (\text{A.13})$$

We also see that

$$\theta_n \leq \frac{\max_{x_j \in R_X} E[\gamma_{t^*}(Z, \Delta|x_j)^\lambda]}{M_\lambda^{\lambda-1}} \leq M_\lambda, \quad (\text{A.14})$$

if  $\lambda < \infty$  and  $\theta_n = 0$  if  $\lambda = \infty$ .

Now, define

$$w_n = \left\lceil \frac{2Q_n c_n}{f_n} + 1 \right\rceil$$

and

$$\eta_{njk} = x_{j+k} + \frac{rc_n}{w_n}, \quad \text{for } r = -w_n, -w_n + 1, \dots, w_n.$$

We then have

$$\begin{aligned} \sup_{|z| \leq c_n} M_{5tn}(x_j, x_{j+k}, z) &\leq \max_{-w_n \leq r \leq w_n} M_{5tn}(x_j, x_{j+k}, rc_n/w_n) \\ &\quad + \max_{-w_n \leq r \leq w_n - 1} |E[H_{tn}(x_j, \eta_{njk(r+1)})] - E[H_{tn}(x_j, \eta_{njk}r)]|. \end{aligned}$$

The second term of the right hand side of the above expression is bounded by

$$\begin{aligned} &Q_n \max_{-w_n \leq r \leq w_n - 1} \int_{\eta_{njk}}^{\eta_{njk(r+1)}} \sum_{\delta=0,1} \int_{-\infty}^{\infty} h_\delta(u, z) dz du \\ &\leq Q_n \max_{-w_n \leq r \leq w_n - 1} \int_{\eta_{njk}}^{\eta_{njk(r+1)}} f_X(u) du \leq Q_n C_4 (\eta_{njk(r+1)} - \eta_{njk}) \leq C_4 f_n / 2, \end{aligned}$$

where  $C_4$  is the Lipschitz constant of  $F_X(\cdot)$ . The goal is therefore to calculate, using (A.11), (A.14) and the above last inequality,

$$\begin{aligned} &P\left(\sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x)| > C_0 \Psi_n^{-1/2}\right) \\ &\leq P(2 \max_{j,t,k,r} M_{5tn}(x_j, x_{j+k}, rc_n/w_n) + f_n(2W_n + V_n) > (C_0 - 2 - C_4 - 2M_\lambda)f_n) \\ &\leq \sum_{j,t,k,r} P(M_{5tn}(x_j, x_{j+k}, rc_n/w_n) > (1/6)(C_0 - 2 - C_4 - 2M_\lambda)f_n) \\ &\quad + P(W_n > (1/6)(C_0 - 2 - C_4 - 2M_\lambda)) \\ &\quad + P(V_n > (1/3)(C_0 - 2 - C_4 - 2M_\lambda)), \end{aligned}$$

where  $C_0$ ,  $C_2$  and  $C_3$  can be chosen to satisfy  $2M_\lambda < (C_3/M_\lambda^{\lambda-1}) \leq (1/6)(C_0 - C_4 - 2M_\lambda - 2)$  and  $L_6^{1/6} < C_2 \leq (1/6)(C_0 - C_4 - 2M_\lambda - 2)$ . In this way, using (A.3), (A.4), (A.12) and (A.13), we only have to treat the first term on the right hand side of the above expression.

Defining  $C'_0 = (1/6)(C_0 - 2 - C_4 - 2M_\lambda)$ , we have by Bernstein's inequality,

$$P(M_{5tn}(x_j, x_{j+k}, rc_n/w_n) > C'_0 f_n) \leq 2 \exp(-\nu_{tjknr}),$$

where

$$\nu_{tjknr} = \frac{C'^2_0 n^2 f_n^2}{2n\sigma_{tjknr}^2 + \frac{2}{3}nC'_0 f_n Q_n},$$

and  $\sigma_{tjknr} = \text{Var}[D_{tjknr}]$  for

$$D_{tjknr} = \gamma_t(Z, \Delta | x_j) I(\gamma_t(Z, \Delta | x_j) \leq Q_n) (I(X \leq \eta_{njkr}) - I(X \leq x_{j+k})).$$

We have

$$\begin{aligned} \sigma_{tjknr}^2 \leq E[D_{tjknr}^2] &\leq \int_{x_{j+k} \wedge \eta_{njkr}}^{\eta_{njkr} \vee x_{j+k}} \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_t^2(z, \delta | x_j) I(\gamma_t(z, \delta | x_j) \leq Q_n) h_\delta(u, z) dz du \\ &\leq C_5 M_\lambda^2 c_n, \end{aligned} \quad (\text{A.15})$$

using (A5), condition (A7) (iv) and where  $C_5 = \max_{\delta} \sup_{x,z} |f_{X|Z,\Delta}(x|z, \delta)|$ . Using (A6) (iii),

$$Q_n f_n = M_\lambda f_n^{(\lambda-2)/(\lambda-1)} = M_\lambda \left( \frac{c_n \log n}{n} \right)^{(\lambda-2)/2(\lambda-1)} \leq M_\lambda c_n. \quad (\text{A.16})$$

We thus have by (A.15) and (A.16) that

$$\nu_{tjknr} \geq C''_0 \log n,$$

with

$$C''_0 = \frac{C'^2_0}{2M_\lambda(C_5 M_\lambda + \frac{1}{3}C'_0)} > \max\left(\frac{6}{3C_5 + 2}, \frac{3L_6^{1/3}}{M_\lambda(6M_\lambda C_5 + 2L_6^{1/6})}\right).$$

Therefore,

$$\sum_{j,t,k,r} P(M_{5tn}(x_j, x_{j+k}, rc_n/w_n) > C'_0 f_n) \leq 6 \sum_{j=1}^{L_X^n} (N_{nj} + 2)(2w_n + 1) n^{-C''_0}, \quad (\text{A.17})$$

where  $C''_0$  has to be chosen large enough so that the right hand side of (A.17) tends to zero sufficiently fast. Thus, the highest order term on the right hand side of (A.17) is

$$96 \left( \frac{L_X M_\lambda Q_n c_n}{f_n^3} \right) n^{-C''_0}, \quad (\text{A.18})$$

where

$$\frac{Q_n c_n}{f_n} = M_\lambda c_n \left( \frac{n}{c_n \log n} \right)^{\frac{\lambda}{2(\lambda-1)}}$$

and

$$N_{nj} \leq \frac{2g(t^*|x_j)}{f_n} \leq 2 \frac{M_\lambda}{f_n},$$

using (A.6) and the fact that  $E[\gamma_{t^*}(Z, \Delta|x_j)] \leq M_\lambda$ . Using (A6) (iii), (A.18) is bounded by

$$L_\lambda \left( \frac{n}{c_n \log n} \right)^{\frac{3\lambda-2}{2(\lambda-1)}} c_n n^{-C_0''} \leq L_\lambda \left( \frac{n}{\log n} \right)^2 n^{-C_0''},$$

where  $L_\lambda = 96M_\lambda^2 L_X$ . Therefore, choosing  $C_0'' \geq 4$  allows us to write

$$P\left(\sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x)| > C_0 \Psi_n^{-1/2}\right) = O(n^{-2}). \quad (\text{A.19})$$

By (A.1) and (A.19), we finally obtain

$$P\left(\sup_{x \in R_X} \sup_{t \in I} |M_{1tn}(x) + M_{2tn}(x)| > C_0 \Psi_n^{-1/2} + C_1 c_n^2\right) = O(n^{-2}). \quad (\text{A.20})$$

**Proof of Proposition 3.2.** First, define new data points as  $\Gamma_t(z, \delta|x) = \sum_{i=1}^{i_0} q_i \gamma_{ti}(z, \delta|x)$ ,  $z \in \mathbb{R}$ ,  $t \in I$ ,  $x \in R_X$ ,  $\delta = 0, 1$ , with fixed and finite  $i_0, q_1, \dots, q_{i_0}$  and with families  $\{\gamma_{ti}, t \in I\}$ ,  $1 \leq i \leq i_0$ , satisfying assumptions (A1)-(A5), with common  $\lambda$  in (A5). If we consider kernel sequences of step-function form,  $K_n(u) = \sum_{j=1}^{j_n} m_{nj} I(-b_{nj} \leq u \leq b_{nj})$ ,  $u \in \mathbb{R}$ , with  $\{j_n\}$ ,  $\{m_{nj}\}$ ,  $\{b_{nj}\}$  sequences of constants characterized in assumption (A9), the expression of Proposition 3.2 can be bounded by

$$\sup_{x \in R_X} \sup_{t \in I} |d_{tn}(x) - d_t(x)| \leq \sum_{i=1}^{i_0} |q_i| (S_{ni}^{(1)} + S_{ni}^{(2)}),$$

where

$$S_{ni}^{(1)} = 2 \sum_{j=1}^{j_n} |m_{nj}| b_{nj} M_{0n}^{(i)}(c_{nj}), \quad (\text{A.21})$$

$$\begin{aligned} M_{0n}^{(i)}(c_{nj}) &= \sup_{x \in R_X} \sup_{t \in I} \left| \frac{1}{2nc_{nj}} \sum_{k=1}^n \gamma_{ti}(Z_k, \Delta_k|x) I(x - c_{nj} < X_k \leq x + c_{nj}) \right. \\ &\quad \left. - \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_{ti}(z, \delta|x) h_\delta(x, z) dz \right|, \quad i = 1, \dots, i_0, \end{aligned}$$

$(M_{0n}^{(i)}(c_{nj}))$  is simply  $M_{0n}(c_n)$  that appeared in the proof of Proposition 3.1 with  $\gamma_t$  and  $c_n$  replaced by  $\gamma_{ti}$  and  $c_{nj}$ , respectively,  $i = 1, \dots, i_0$ ,  $j = 1, \dots, j_n$ ,  $c_{nj} = a_n b_{nj}$  and

$$\begin{aligned} S_{ni}^{(2)} &= \sup_{x \in R_X} \sup_{t \in I} \left| \left( 2 \sum_{j=1}^{j_n} m_{nj} b_{nj} - 1 \right) \sum_{\delta=0,1} \int_{-\infty}^{\infty} \gamma_{ti}(z, \delta|x) h_{\delta}(x, z) dz \right| \\ &\leq C_5 M_{\lambda} \left| \left( 2 \sum_{j=1}^{j_n} m_{nj} b_{nj} - 1 \right) \right|. \end{aligned}$$

Next, define

$$\varepsilon_n = 2C_0 \Psi_n^{-1/2} \sum_{j=1}^{j_n} |m_{nj}| b_{nj}^{1/2} + 2C_1 a_n^2 \sum_{j=1}^{j_n} |m_{nj}| b_{nj}^3.$$

Then,

$$P(S_{ni}^{(1)} > \varepsilon_n) \leq \sum_{j=1}^{j_n} P(M_{0n}^{(i)}(c_{nj}) > C_0 \Psi_{nj}^{-1/2} + C_1 c_{nj}^2),$$

with  $\Psi_{nj} = \Psi_n b_{nj}$  and  $c_{nj} = a_n b_{nj}$ . By using (A.20), we thus obtain

$$P(S_{ni}^{(1)} > \varepsilon_n) \leq O(j_n n^{-2}).$$

For  $s$  in (A9) (i) smaller than 1 ( $j_n = O(n^s)$ ,  $s > 0$ ) and using the Borel-Cantelli Lemma, we obtain

$$S_{ni}^{(1)} = O(\varepsilon_n) \text{ a.s.},$$

for which  $\varepsilon_n = O(\max(\Psi_n^{-1/2}, a_n^2))$ .

**Proof of Theorem 3.3.** Let  $b_{nj} = j a_n^{3/2}$  and  $m_{nj} = K(j a_n^{3/2}) - K((j+1) a_n^{3/2})$  in (A9).

Then (A9) (i) becomes

$$\left| \left( 2 \sum_{j=1}^{j_n} m_{nj} b_{nj} - 1 \right) \right| \leq \int |K_n(u) - K(u)| du \leq C a_n^{3/2},$$

for some  $C > 0$  and since  $\int K_n(u) du = 2 \sum_{j=1}^{j_n} m_{nj} b_{nj}$ , (A9) (ii) and (A9) (iii) become

$$\sup_n a_n^{9/4} \sum_{j=1}^{j_n} j^{1/2} |K'(\theta_{nj})| < \infty$$

and

$$\sup_n a_n^6 \sum_{j=1}^{j_n} j^3 |K'(\theta_{nj})| < \infty,$$

where  $\theta_{nj}$  is between  $ja_n^{3/2}$  and  $(j+1)a_n^{3/2}$ . Therefore, we can choose  $0 < s < 1$  such that  $j_n = O(a_n^{-3/2})$ .

Next, if we denote  $d_{tn}(x)$  using kernel  $K$  by  $d_{tn}(x, K)$ , it is clear that

$$\begin{aligned} & \sup_{x \in R_X} \sup_{t \in I} |d_{tn}(x, K) - d_{tn}(x, K_n)| \\ & \leq \sup_{x \in R_X} \sup_{t \in I} \left( \frac{Da_n^{3/2}}{na_n} \sum_{i=1}^n |\Gamma_t(Z_i, \Delta_i | x) I(x - a_n < X_i \leq x + a_n)| \right) = O(a_n^{3/2}) \text{ a.s.}, \end{aligned} \quad (\text{A.22})$$

for some constant  $D > 0$ , a kernel support equal to  $[-1, 1]$  and where Proposition 3.1 is used with  $c_n = a_n$  (with  $c_n = (L/2)a_n$  if  $L$  is the length of the support).

Finally, write

$$\begin{aligned} \sup_{x \in R_X} \sup_{t \in I} \left| \frac{\sum_{i=1}^n K(\frac{x-X_i}{a_n}) \Gamma_t(Z_i, \Delta_i | x)}{\sum_{i=1}^n K(\frac{x-X_i}{a_n})} - \frac{d_t(x)}{f_X(x)} \right| & \leq \sup_{x \in R_X} \sup_{t \in I} \left| \frac{d_{tn}(x, K) - d_t(x)}{f_{nX}(x)} \right| \\ & + \sup_{x \in R_X} \sup_{t \in I} \left| \frac{d_t(x)(f_X(x) - f_{nX}(x))}{f_X(x)f_{nX}(x)} \right|. \end{aligned}$$

If we use the fact that  $\inf_{x \in R_X} |f_X(x)| > 0$  in addition to the obtained results for  $d_{tn}(x, K)$  in (A.22) and  $d_{tn}(x, K_n)$  in Proposition 3.2, both terms on the right hand side of the above expression are  $O(\Psi_n^{-1/2})$  a.s. since  $\sup_{x \in R_X} \sup_{t \in I} |d_t(x)|$  is bounded (using the definition (A8) of the points  $\Gamma(\cdot, \cdot | \cdot)$ ).

**Proof of Proposition 4.1.** For  $M_{7n}(c_n) = M_n(c_n)$  in Proposition 4.1, let

$$\begin{aligned} M_{7n}(c_n) &= \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} \left| \frac{1}{2nc_n} \sum_{i=1}^n (\gamma_t(Z_i, \Delta_i | x) \right. \\ &\quad \left. - \gamma_s(Z_i, \Delta_i | x)) I(x - c_n < X_i \leq x + c_n) \right. \\ &\quad \left. - \frac{1}{2c_n} \int_{x-c_n}^{x+c_n} \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta | x) - \gamma_s(z, \delta | x)) h_{\delta}(u, z) dz du \right| \\ &\quad + \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} \left| \frac{1}{2c_n} \int_{x-c_n}^{x+c_n} \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta | x) - \gamma_s(z, \delta | x)) h_{\delta}(u, z) dz du \right. \\ &\quad \left. - \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta | x) - \gamma_s(z, \delta | x)) h_{\delta}(x, z) dz \right| \\ &= \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} |M_{8stn}(x)| + \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} |M_{9stn}(x)|. \end{aligned}$$

First,  $M_{9stn}(x)$  is treated as  $M_{2tn}(x)$  in Proposition 3.1 such that

$$\sup_{\{x \in R_X, |t-s| \leq d_n, s, t \in I\}} |M_{9stn}(x)| \leq C_{m1} d_n c_n^2, \quad (\text{A.23})$$

using (A11). This inequality corresponds to (A.1) in Proposition 3.1.

Divide  $R_X$  into  $[\frac{2L_X}{f_n d_n^{1/2}}]$  intervals of length smaller than or equal to  $f_n d_n^{1/2}$ . Denote by  $J_X$  the set of points  $\{x_k = x_0 + k[\frac{2L_X}{f_n d_n^{1/2}}]^{-1}L_X, 1 \leq k \leq [\frac{2L_X}{f_n d_n^{1/2}}] - 1 = L_X^{d_n}\}$  and  $x_{L_X^{d_n}+1} = \sup\{x : x \in R_X\}$  which limit the intervals. Then,  $M_{8stn}(x)$  is treated like (A.2) in Proposition 3.1, where  $\gamma_t(\cdot, \cdot | \cdot)$  is replaced by  $\gamma_t(\cdot, \cdot | \cdot) - \gamma_s(\cdot, \cdot | \cdot)$ ,  $V_n(x_j)$  by  $V_n^d(x_j) = (1/n) \sum_{i=1}^n L_0(Z_i, \Delta_i | x_j) - E[L_0(Z, \Delta | x_j)]$  and  $V_n$  by  $V_n^d = \max_{x_j \in J_X} (2E[L_0(Z, \Delta | x_j)] + |V_n^d(x_j)|)$ . Using Chebyshev's inequality,  $P(V_n^d > C_{m2}) = o(n^{-2})$  with  $C_{m2}$  chosen larger than  $4L_6^{1/6}$ . A development similar to (A.5) and (A.9) is used to obtain

$$\begin{aligned} & \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} |M_{8stn}(x)| \\ & \leq 2c_n^{-1} \max_{x_j \in J_X} \max_{|t-s| \leq d_n, s, t \in I} \max_{k \in \{-1, 0, 1\}} \sup_{|z| \leq c_n} M_{10stn}(x_j, x_{j+k}, z) + V_n^d \Psi_n^{-1/2} d_n^{1/2}, \end{aligned} \quad (\text{A.24})$$

where

$$\begin{aligned} M_{10stn}(x_j, x, z) &= |G_{stn}(x_j, x+z) - G_{stn}(x_j, x) - [G_{st}(x_j, x+z) - G_{st}(x_j, x)]|, \\ G_{stn}(x_j, x) &= n^{-1} \sum_{i=1}^n (\gamma_t(Z_i, \Delta_i | x_j) - \gamma_s(Z_i, \Delta_i | x_j)) I(X_i \leq x), \end{aligned}$$

and

$$G_{st}(x_j, x) = E[G_{stn}(x_j, x)] = \int_{x_0}^x \sum_{\delta=0,1} \int_{-\infty}^{\infty} (\gamma_t(z, \delta | x_j) - \gamma_s(z, \delta | x_j)) h_{\delta}(u, z) dz du.$$

Partition  $I$  into  $O(f_n^{-1} d_n^{-3/2})$  intervals such that for each  $x_j$ ,  $j = 0, \dots, L_X^{d_n} + 2$ ,  $g(t^* | x_j) - g(t_* | x_j)$  is divided into  $m_j = [\frac{g(t^* | x_j) - g(t_* | x_j)}{C_L d_n}]$  intervals of length  $C_{m3}(x_j) C_L d_n$ ,  $1 \leq C_{m3}(x_j) \leq 2$ . In this way we can construct  $|g(t_* | x_j) - g(t_1 | x_j)| = C_{m3}(x_j) C_L d_n$ ,  $|g(t_{\alpha+1} | x_j) - g(t_{\alpha} | x_j)| = C_{m3}(x_j) C_L d_n$ ,  $\alpha = 1, \dots, m_j - 2$ ,  $|g(t^* | x_j) - g(t_{m_j-1} | x_j)| = C_{m3}(x_j) C_L d_n$ ,  $t_0 = t_*$ ,  $t_{m_j} = t^*$  for all  $j$ . Let  $I_{j\alpha} = [g(t_{\alpha-1} | x_j), g(t_{\alpha+1} | x_j)]$ ,  $\alpha = 1, \dots, m_j - 1$ . For each  $s, t \in I$  with  $|t-s| \leq d_n$ , there exists an interval  $I_{j\alpha}$  such that  $g(s | x_j), g(t | x_j) \in I_{j\alpha}$ . Partition each  $I_{j\alpha}$  by a grid  $g(t_{\alpha\beta} | x_j) = g(t_{\alpha} | x_j) + \beta \frac{C_{m3}(x_j) C_L d_n}{p_n}$ ,  $\beta = -p_n, \dots, p_n$ , where  $p_n = [\Psi_n^{1/2} d_n^{1/2} + 1]$ . Using (A7) (iv), (A4), (A5) and the monotonicity of  $\gamma_t(Z, \Delta | x)$ , (A.24) is majorized by

$$\begin{aligned} & 2c_n^{-1} \max_{x_j \in J_X} \max_{k \in \{-1, 0, 1\}} \max_{1 \leq \alpha \leq m_j - 1} \max_{-p_n \leq \beta, \zeta \leq p_n} \sup_{|z| \leq c_n} M_{10t_{\alpha\zeta} t_{\alpha\beta} n}(x_j, x_{j+k}, z) \\ & + 4c_n^{-1} \max_{x_j \in J_X} \max_{1 \leq \alpha \leq m_j - 1} \max_{-p_n \leq \beta \leq p_n - 1} C_5 c_n |g(t_{\alpha(\beta+1)} | x_j) - g(t_{\alpha\beta} | x_j)| + V_n^d \Psi_n^{-1/2} d_n^{1/2}, \end{aligned} \quad (\text{A.25})$$

where  $C_5$  is defined as in (A.15) and the second term of the above expression equals  $4C_5 \frac{C_{m3}(x_j) C_L d_n}{p_n} \leq C_{m4} \Psi_n^{-1/2} d_n^{1/2}$ , where  $C_{m4} = 8C_5 C_L$ .

Now, put  $T_n = M_\lambda(f_n^{-1}d_n^{-1/2})^{\frac{1}{\lambda-1}}$ . Define  $M_\lambda$  for  $2 < \lambda \leq \infty$  as in the proof of Proposition 3.1,  $H_{stn}(x_j, x)$  by

$$n^{-1} \sum_{i=1}^n (\gamma_t(Z_i, \Delta_i|x_j) - \gamma_s(Z_i, \Delta_i|x_j)) I(|\gamma_t(Z_i, \Delta_i|x_j) - \gamma_s(Z_i, \Delta_i|x_j)| \leq T_n) I(X_i \leq x),$$

and  $M_{11stn}(x_j, x, z)$  by substitution of  $H_{stn}$  for  $G_{stn}$  and  $E[H_{stn}]$  for  $G_{st}$ . Then, (A.25) is majorized by

$$\begin{aligned} & 2c_n^{-1} \max_{x_j \in J_X} \max_{k \in \{-1, 0, 1\}} \max_{1 \leq \alpha \leq m_j - 1} \max_{-p_n \leq \beta, \zeta \leq p_n} \sup_{|z| \leq c_n} M_{11t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, z) \\ & + 2c_n^{-1} f_n d_n^{1/2} (W_n^d + \theta_n^d) + (V_n^d + C_{m4}) \Psi_n^{-1/2} d_n^{1/2}, \end{aligned} \quad (\text{A.26})$$

where  $W_n^d$  and  $\theta_n^d$  are defined similarly to  $W_n$  and  $\theta_n$  in the proof of Proposition 3.1. It is easy to check that  $P(2W_n^d > C_{m5}) = o(n^{-2})$  and  $2\theta_n^d < C_{m6}$ , where  $C_{m5}$  and  $C_{m6}$  are chosen such that  $C_{m5} > 2^{\lambda+2}M_\lambda$  and  $C_{m6} = 2^{\lambda+1}M_\lambda$ .

Next, consider

$$\kappa_{njk} = x_{j+k} + \frac{rc_n}{p_n}, \quad \text{for } r = -p_n, -p_n + 1, \dots, p_n.$$

For fixed  $j, k, \alpha, \beta, \zeta, n$ ,  $H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k} + z) - H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k})$  and  $E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k} + z)] - E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k})]$  are monotone with respect to  $z$  and have finite limits in  $x_{j+k} + c_n$  and  $x_{j+k} - c_n$ . Therefore,

$$\begin{aligned} & \sup_{|z| \leq c_n} M_{11t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, z) \\ & \leq \max_{-p_n \leq r \leq p_n} M_{11t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, rc_n/p_n) \\ & \quad + \max_{-p_n \leq r \leq p_n - 1} |E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k} + \frac{(r+1)c_n}{p_n})] - E[H_{t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k} + \frac{rc_n}{p_n})]|, \end{aligned}$$

where the second term on the right hand side of the above expression is bounded by

$$4C_L C_5 c_n \Psi_n^{-1/2} d_n^{1/2}.$$

Therefore, (A.26) is majorized by

$$\begin{aligned} & 2c_n^{-1} \max_{x_j \in J_X} \max_{k \in \{-1, 0, 1\}} \max_{1 \leq \alpha \leq m_j - 1} \max_{-p_n \leq \beta, \zeta, r \leq p_n} M_{11t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, rc_n/p_n) \\ & + (2W_n^d + V_n^d + C_{m4} + C_{m6} + C_{m7}) \Psi_n^{-1/2} d_n^{1/2}, \end{aligned} \quad (\text{A.27})$$

where  $C_{m7} = 8C_L C_5$ .

Next,

$$\begin{aligned}
& P\left(\sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} |M_{8stn}(x)| > C_{m0} \Psi_n^{-1/2} d_n^{1/2}\right) \\
& \leq \sum_{j, k, \alpha, \beta, \zeta, r} P(M_{11t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, rc_n/p_n) > C'_{m0} f_n d_n^{1/2}) \\
& \quad + P(W_n^d > C'_{m0}) + P(V_n^d > 2C'_{m0}),
\end{aligned} \tag{A.28}$$

where  $C'_{m0} = (1/6)(C_{m0} - 16C_5C_L - 2^{\lambda+1}M_\lambda)$  and  $C_{m0}$ ,  $C_{m2}$  and  $C_{m5}$  can be chosen such that  $C'_{m0}$  is larger than  $C_{m2}/2$  and  $C_{m5}/2$  in order to satisfy

$$\max(2^{\lambda+1}M_\lambda, 2L_6^{1/6}) < C'_{m0}.$$

By Bernstein's inequality,

$$P(M_{11t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, rc_n/p_n) > C'_{m0} f_n d_n^{1/2}) \leq 2 \exp(-\phi_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}),$$

where

$$\phi_{nt_{\alpha\zeta}t_{\alpha\beta}jkr} = \frac{C'^2_{m0} n^2 f_n^2 d_n}{2n\sigma_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}^2 + \frac{2}{3}nC'_{m0} f_n d_n^{1/2} T_n},$$

$\sigma_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}^2 = \text{Var}[\Omega_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}]$  and

$$\begin{aligned}
\Omega_{nt_{\alpha\zeta}t_{\alpha\beta}jkr} &= (\gamma_{t_{\alpha\beta}}(Z, \Delta|x_j) - \gamma_{t_{\alpha\zeta}}(Z, \Delta|x_j)) \\
&\quad \times I(|\gamma_{t_{\alpha\beta}}(Z, \Delta|x_j) - \gamma_{t_{\alpha\zeta}}(Z, \Delta|x_j)| \leq T_n) (I(X \leq \kappa_{njkr}) - I(X \leq x_{j+k})).
\end{aligned}$$

Using (A11) (ii),  $\sigma_{nt_{\alpha\zeta}t_{\alpha\beta}jkr}^2 \leq C_{L2}C_5c_n d_n$ , and (A10) (iii),  $T_n \Psi_n^{-1/2} \leq d_n^{1/2}$ . Therefore,

$$\phi_{nt_{\alpha\zeta}t_{\alpha\beta}jkr} \geq C''_{m0} \log n,$$

with

$$C''_{m0} = \frac{C'^2_{m0}}{2(C_{L2}C_5 + \frac{1}{3}C'_{m0})}.$$

Finally,

$$\sum_{j, k, \alpha, \beta, \zeta, r} P(M_{11t_{\alpha\zeta}t_{\alpha\beta}n}(x_j, x_{j+k}, rc_n/p_n) > C'_{m0} f_n d_n^{1/2}) \leq 2 \sum_{j=1}^{L_X^{d^n}} \sum_{k=-1}^1 \sum_{\alpha=1}^{m_j-1} \sum_{-p_n \leq \beta, \zeta, r \leq p_n} n^{-C''_{m0}},$$

for which the highest order term on the right hand side is

$$96 \frac{L_X M_\lambda}{C_L} \Psi_n^2 c_n^{-1} n^{-C''_{m0}} \leq 96 \frac{L_X M_\lambda}{C_L} \frac{n^2}{(\log n)^2} n^{-C''_{m0}}.$$

Choosing  $C''_{m0}$  sufficiently large finishes the proof.

**Proof of Theorem 4.3.** Let  $b_{nj} = ja_n^{3/2}d_n^{-1/2}$  and  $m_{nj} = K(ja_n^{3/2}d_n^{-1/2}) - K((j+1)a_n^{3/2}d_n^{-1/2})$  in (A12). Then, (A12) (i) becomes

$$|(2 \sum_{j=1}^{j_n} m_{nj} b_{nj} - 1)| \leq \int |K_n(u) - K(u)| du \leq C a_n^{3/2} d_n^{-1/2},$$

for some  $C > 0$ . Assumptions (A12) (ii) and (A12) (iii) are easily satisfied using  $j_n = O(a_n^{-3/2}d_n^{1/2})$  such that  $s$  can then be chosen between 0 and 1.

Next, let us denote  $d_{stn}(x)$  using the kernel  $K$  by  $d_{stn}(x, K)$ . It is clear that

$$\begin{aligned} & \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} |d_{stn}(x, K) - d_{stn}(x, K_n)| \\ & \leq \sup_{x \in R_X} \sup_{|t-s| \leq d_n, s, t \in I} \left( \frac{D a_n^{3/2} d_n^{-1/2}}{n a_n} \sum_{i=1}^n |\Gamma_{ts}(Z_i, \Delta_i | x)| I(x - a_n < X_i \leq x + a_n) \right) \\ & = O(a_n^{3/2} d_n^{1/2}) \text{ a.s.}, \end{aligned}$$

for which we use Proposition 4.1 with  $c_n = a_n$  (for a kernel support equal to  $[-1, 1]$ ). Finally, the proof of the second expression in Theorem 4.3 is along the same lines as in the proof of Theorem 3.3.

## References

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.

Buckley, J. and James, I. R. (1979). Linear regression with censored data. *Biometrika*, **66**, 429–436.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.*, **89**, 81–87.

Cheng, P. E. and Chu, C. K. (1996). Kernel estimation of distribution functions and quantiles with missing data. *Statist. Sinica*, **6**, 63–78.

Chu, C. K. and Cheng, P. E. (1995). Nonparametric regression estimation with missing data. *J. Statist. Plann. Infer.*, **48**, 85–99.

Fan, J. and Gijbels, I. (1994). Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc.*, **89**, 560–570.

Härdle, W., Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.*, **16**, 1428–1449.

Heuchenne, C. and Van Keilegom, I. (2004). Polynomial regression with censored data based on preliminary nonparametric estimation (conditionally accepted by *Ann. Inst. Statist. Math.*).

Heuchenne, C. and Van Keilegom, I. (2005). Mean preservation in nonparametric regression with censored data (work in progress).

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

Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika*, **74**, 301-309.

Little, R. J. A. and Rubin, D. B. (1987). Statistical Analysis with Missing Data. Wiley, New York.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Series Analysis*, **17**, 571-599.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.*, **9**, 141–142.

Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595–645.

Van Keilegom, I. and Akritas, M. G. (1999). Transfer of tail information in censored regression models. *Ann. Statist.*, **27**, 1745–1784.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A.*, **26**, 359–372.