Local robust estimation of the Pickands dependence function

Mikael Escobar-Bach, Yuri Goegebeur, Armelle Guillou

Local robust estimation of the Pickands dependence function

Mikael Escobar-Bach[(1)], Yuri Goegebeur[(1)], Armelle Guillou[(2)]

[(1)] Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

[(2)] Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

Abstract

We consider the robust estimation of the Pickands dependence function in the covariate framework. Our estimator is based on local estimation with the minimum density power divergence criterion. We provide its main asymptotic properties, in particular the convergence of the stochastic process, correctly normalized, towards a tight centered Gaussian process. The finite sample performance of our estimator is illustrated on a small simulation study involving both uncontaminated and contaminated samples.

Keywords: Conditional Pickands dependence function, robustness, stochastic convergence.

1 Introduction

Modelling dependence among extremes is of primary importance in practical applications where extreme phenomena occur. To this aim, the copula function can be used as a margin-free description of the dependence structure. Indeed, according to the well-known result of Sklar (1959), the distribution function of a pair $(Y^{(1)}, Y^{(2)})$ can be represented in terms of the two margins F_1 and F_2 of $Y^{(1)}$ and $Y^{(2)}$ respectively, and a copula function C as follows:

$$\mathbb{P}\left(Y^{(1)} \leqslant y_1, Y^{(2)} \leqslant y_2\right) = C\left(F_1(y_1), F_2(y_2)\right).$$

This function C characterizes the dependence between $Y^{(1)}$ and $Y^{(2)}$ and is called an extreme value copula if and only if it admits a representation of the form

$$C(y_1, y_2) = \exp\left(\log(y_1 y_2) A\left(\frac{\log(y_2)}{\log(y_1 y_2)}\right)\right),$$

where A: $[0,1] \to [1/2, 1]$ is the Pickands dependence function, which is convex and satisfies $\max\{t, 1-t\} \leqslant A(t) \leqslant 1$, see Pickands (1981). Statistical inference on the bivariate function C is therefore equivalent to the statistical inference on the one-dimensional function A. Estimating this function A has been extensively studied in the literature. We can mention, among others, Capéraà *et al.* (1997), Fils-Villetard *et al.* (2008) or Bücher *et al.* (2011).
In this paper, we extend the above framework to the case where the pair $(Y^{(1)}, Y^{(2)})$ is recorded along with a random covariate $X \in \mathbb{R}^p$. In that context, the copula function together with the

marginal distribution functions depend on the covariate X. In the sequel, we denote by C_x, $F_1(.|x)$ and $F_2(.|x)$ the conditional copula function and the continuous conditional distribution functions of $Y^{(1)}$ and $Y^{(2)}$ given $X = x$. Our model can thus be written as

$$\mathbb{P}\left(F_1(Y^{(1)}|x) \leqslant y_1, F_2(Y^{(2)}|x) \leqslant y_2 \Big| X = x\right) = C_x(y_1, y_2), \tag{1}$$

where C_x admits a representation of the form

$$C_x(y_1, y_2) = \exp\left(\log(y_1 y_2) A\left(\frac{\log(y_2)}{\log(y_1 y_2)}\Big|x\right)\right),$$

with $A(.|.) : [0,1] \times \mathbb{R}^p \to [1/2, 1]$ is the conditional Pickands dependence function which is again a convex function satisfying $\max\{t, 1-t\} \leqslant A(t|x) \leqslant 1$ for all $x \in \mathbb{R}^p$. Note that recently, Portier and Segers (2015) also considered this model (1) but under the simplifying assumption that the dependence between $Y^{(1)}$ and $Y^{(2)}$ does not depend on the value taken by the covariate, i.e. $C_x = C$ (see also Gijbels *et al.*, 2015).

Moreover, in addition to the covariate context, we consider the case of contamination and we propose a robust estimator of the conditional Pickands dependence function $A(.|x)$. To reach this goal, we use the density power divergence method introduced by Basu *et al.* (1998). In particular, the density power divergence between two density functions g and h is defined as follows

$$\Delta_\alpha(g,h) := \begin{cases} \displaystyle\int_{\mathbb{R}} \left[h^{1+\alpha}(y) - \left(1+\frac{1}{\alpha}\right) h^\alpha(y) g(y) + \frac{1}{\alpha} g^{1+\alpha}(y)\right] dy, & \alpha > 0, \\ \displaystyle\int_{\mathbb{R}} \log\frac{g(y)}{h(y)} g(y) dy, & \alpha = 0. \end{cases}$$

Here the density function h is assumed to depend on a parameter vector θ, and if $Z_1, ..., Z_n$ is a sample of independent and identically distributed random variables according to the density function g, then the minimum density power divergence estimator (MDPDE) of θ is the point $\widehat{\theta}$ minimizing the empirical version

$$\widehat{\Delta}_\alpha(\theta) := \begin{cases} \displaystyle\int_{\mathbb{R}} h^{1+\alpha}(y) dy - \left(1+\frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n h^\alpha(Z_i), & \alpha > 0, \\ \displaystyle -\frac{1}{n} \sum_{i=1}^n \log h(Z_i) & \alpha = 0. \end{cases}$$

We can observe that for $\alpha = 0$ one recovers the log-likelihood function, up to the sign. A large value of α allows us to increase the robustness of the estimator, whereas a smaller value implies more efficiency. This parameter α can thus be selected in order to ensure a trade-off between these two antagonist concepts.

The remainder of the paper is organized as follows. In Section 2, we simplify the situation to the case where the two marginals $F_1(.|x)$ and $F_2(.|x)$ are known as standard exponential distribution functions, we propose a robust estimator for $A(.|x)$ and we prove its convergence in terms of a stochastic process. Then, in Section 3, we extend this result to the case of unknown margins. Finally, in Section 4, we illustrate the efficiency of our estimator in a small simulation study. All the proofs and technical results are postponed to the appendices.

2 Case of known margins

In this section, we restrict our interest to the case where the marginals $F_1(.|x)$ and $F_2(.|x)$ are standard exponential distribution functions and we denote by $A_0(.|x)$ the true conditional Pickands dependence function associated to this pair $(Y^{(1)}, Y^{(2)})$.

2.1 Construction of our estimator

Under the assumption of standard exponential margins, model (1) can be rewritten as follows

$$G(y_1, y_2|x) := \mathbb{P}\left(Y^{(1)} > y_1, Y^{(2)} > y_2 \Big| X = x\right) = \exp\left(-(y_1+y_2)A_0\left(\frac{y_2}{y_1+y_2}\Big|x\right)\right),$$

for all $y_1, y_2 > 0$. Let $t \in [0,1]$. Considering the univariate random variable

$$Z_t := \min\left(\frac{Y^{(1)}}{1-t}, \frac{Y^{(2)}}{t}\right),$$

it is clear that

$$\mathbb{P}(Z_t > z|X = x) = e^{-zA_0(t|x)}, \quad \forall z > 0 \text{ and } x \in \mathbb{R}^p.$$

Consequently, the conditional distribution of Z_t given $X = x$ is an exponential distribution with parameter $A_0(t|x)$.
Let $(Z_{t,i}, X_i), i = 1, \ldots, n$, be independent copies of the random pair (Z_t, X). In the present paper, we will develop a nonparametric robust estimator for $A_0(t|x)$ by fitting this exponential distribution function locally to the variables $Z_{t,i}$, $i = 1, ..., n$, by means of the MDPD criterion, adjusted to locally weighted estimation, i.e. we minimize for $\alpha > 0$

$$\begin{aligned}\widehat{\Delta}_{\alpha,x,t}(A(t|x)) &:= \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)\left\{\int_0^\infty \left(A(t|x)e^{-A(t|x)z}\right)^{1+\alpha} dz - \left(1+\frac{1}{\alpha}\right)\left(A(t|x)e^{-A(t|x)Z_{t,i}}\right)^{\alpha}\right\} \\ &= \frac{[A(t|x)]^\alpha}{n}\sum_{i=1}^{n} K_h(x-X_i)\left\{\frac{1}{1+\alpha} - \left(1+\frac{1}{\alpha}\right)e^{-\alpha A(t|x)Z_{t,i}}\right\}. \qquad (2)\end{aligned}$$

Here $K_h(.) := K(./h)/h^p$ where K is a joint density on $\mathbb{R}^p$ and $h = h_n$ is a positive non-random sequence satisfying $h_n \to 0$ as $n \to \infty$. The MDPDE for $A(t|x)$ satisfies the estimating equation

$$\widehat{\Delta}^{(1)}_{\alpha,x,t}(A(t|x)) = 0, \qquad (3)$$

where $\widehat{\Delta}^{(j)}_{\alpha,x,t}(.)$ denotes the derivative of order j of $\widehat{\Delta}_{\alpha,x,t}(.)$. We denote the estimator in the sequel as $\widehat{A}_{\alpha,n}(t|x)$.
Our aim in this paper is to show the weak convergence of the stochastic process

$$\left\{\sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right), t \in [0,1]\right\}, \qquad (4)$$

in the space of all continuous functions on $[0,1]$, denoted as $\mathcal{C}([0,1])$, when $n \to \infty$.

Our starting point is the estimating equation (3). By applying a Taylor series expansion of it around the true value $A_0(t|x)$, we get

$$0 = \widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)) + \left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right)\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) + \frac{1}{2}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right)^2 \widehat{\Delta}^{(3)}_{\alpha,x,t}(\widetilde{A}(t|x))$$

where $\widetilde{A}(t|x)$ is a random intermediate value between $A_0(t|x)$ and $\widehat{A}_{\alpha,n}(t|x)$. A straightforward rearrangement of the above display gives

$$\sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right) = \frac{-\sqrt{nh^p}\,\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x))}{\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,x,t}(\widetilde{A}(t|x))\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right)}. \quad (5)$$

Consequently, as a preliminary step to obtain the convergence of the stochastic process (4), we need to study the properties of the derivatives $\widehat{\Delta}^{(j)}_{\alpha,x,t}, j = 1, 2, 3$. According to Appendix 5.2, all of them can be expressed as a linear combination of a key statistic T_n, defined as

$$T_n(K, a, t, \lambda, \beta, \gamma|x) := \frac{a^\gamma}{n}\sum_{i=1}^{n} K_h(x - X_i) Z_{t,i}^{\beta} e^{-\lambda a Z_{t,i}}, \quad (6)$$

for $a \in [1/2, 1]$, $t \in [0, 1]$, $\lambda, \beta \geqslant 0$ and $\gamma \in \mathbb{R}$. Thus, we start in the next section by looking at the main asymptotic properties of this statistic T_n.

2.2 Asymptotic properties of T_n

Due to the regression context, we need some Hölder-type conditions on the density function f of the covariate X and on the conditional Pickands dependence function A_0. Let $\|.\|$ be some norm on $\mathbb{R}^p$.

Assumption $(\mathcal{D})$**.** *There exist $M_f > 0$ and $\eta_f > 0$ such that $|f(x) - f(z)| \leqslant M_f\|x - z\|^{\eta_f}$, for all $(x, z) \in \mathbb{R}^p \times \mathbb{R}^p$.*

Assumption $(\mathcal{A}_0)$**.** *There exist $M_{A_0} > 0$ and $\eta_{A_0} > 0$ such that $|A_0(t|x) - A_0(t|z)| \leqslant M_{A_0}\|x - z\|^{\eta_{A_0}}$, for all $(x, z) \in \mathbb{R}^p \times \mathbb{R}^p$ and $t \in [0, 1]$.*

Also a usual condition is assumed on the kernel K.

Assumption $(\mathcal{K}_1)$**.** *K is a bounded density function on $\mathbb{R}^p$ with support included in the unit hypersphere $\mathbb{S}^{p-1} \subset \mathbb{R}^p$.*

As a preliminary result, in Lemma 2.1 we prove the convergence in probability of our key statistic T_n.

Lemma 2.1 *Assume that for all $t \in [0, 1]$, $x \to A_0(t|x)$ and the density function f are both continuous. Under Assumption $(\mathcal{K}_1)$, if $h \to 0$ and $nh^p \to \infty$, then for $a \in [1/2, 1]$, $\lambda, \beta \geqslant 0$, $\gamma \in \mathbb{R}$ and $x \in \mathbb{R}^p$ where $f(x) > 0$, we have*

$$T_n(K, a, t, \lambda, \beta, \gamma|x) \xrightarrow{\mathbb{P}} a^\gamma \Gamma(\beta + 1)\frac{A_0(t|x)}{(\lambda a + A_0(t|x))^{\beta+1}} f(x),$$

as $n \to \infty$, where Γ is the gamma function defined as

$$\Gamma(r) := \int_0^\infty t^{r-1} e^{-t} dt, \quad \forall r > 0.$$

Now, our interest is in the rate of convergence in Lemma 2.1 when a is replaced by $A_0(t|x)$. More precisely, we want to show the weak convergence of the stochastic process

$$\left\{ \sqrt{nh^p} \left(T_n(K, A_0(t|x), t, \lambda, \beta, \gamma|x) - \Gamma(\beta+1) \frac{[A_0(t|x)]^{\gamma-\beta}}{(\lambda+1)^{\beta+1}} f(x) \right), t \in [0,1] \right\}.$$

To establish such a result, we use empirical processes arguments based on the theory of Vapnik-Červonenkis class (VC-class) of functions as formulated in van der Vaart and Wellner (1996). This allows us to show the following theorem.

Theorem 2.1 *Let $\gamma \in \mathbb{R}$ and $(\lambda, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+$ or $(\lambda, \beta) = (0,0)$. Under the assumptions of Lemma 2.1 and if $(\mathcal{D})$ and $(\mathcal{A}_0)$ hold with $\sqrt{nh^p} h^{\min(\eta_f, \eta_{A_0})} \to 0$, then the process*

$$\left\{ \sqrt{nh^p} \left(T_n(K, A_0(t|x), t, \lambda, \beta, \gamma|x) - \Gamma(\beta+1) \frac{[A_0(t|x)]^{\gamma-\beta}}{(\lambda+1)^{\beta+1}} f(x) \right), t \in [0,1] \right\}$$

weakly converges in $\mathcal{C}([0,1])$ towards a tight centered Gaussian process $\{B_t, t \in [0,1]\}$ with covariance structure given by

$$Cov(B_t, B_s) = [A_0(t|x)A_0(s|x)]^\gamma \|K\|_2^2 f(x) \times \left\{ \int_{\mathbb{R}_+^2} g(u,v) G_{t,s}(u,v|x) du dv + \frac{1-\lambda}{1+\lambda} \delta_0(\beta) \right\},$$

for all $(s,t) \in [0,1]^2$, where δ_0 is the Dirac measure on 0, and

$$\begin{aligned} g(u,v) &:= u^{\beta-1}(\beta - \lambda A_0(t|x)u) e^{-\lambda A_0(t|x)u} v^{\beta-1}(\beta - \lambda A_0(s|x)v) e^{-\lambda A_0(s|x)v}, \\ G_{t,s}(u,v|x) &:= G\left(\max((1-t)u, (1-s)v), \max(tu, sv) \middle| x \right). \end{aligned}$$

We now derive the limiting distribution of a vector of statistics of the form (6), when properly normalized. Let $\mathbb{T}_n$ be a $(m \times 1)$ vector defined as

$$\mathbb{T}_n := (T_n(K_1, a_1, t_1, \lambda_1, \beta_1, \gamma_1|x_1), T_n(K_2, a_2, t_2, \lambda_2, \beta_2, \gamma_2|x_2), ..., T_n(K_m, a_m, t_m, \lambda_m, \beta_m, \gamma_m|x_m))^T,$$

for some positive integer m and let Σ be a $(m \times m)$ covariance matrix with elements $(\sigma_{j,k})_{1 \leqslant j,k \leqslant m}$ defined as

$$\begin{aligned} \sigma_{j,k} := {} & a_j^{\gamma_j} a_k^{\gamma_k} \delta_0(x_j - x_k) \|K_j K_k\|_1 f(x_j) \times \left\{ \int_{\mathbb{R}_+^2} g_{j,k}(u,v) G_{t_j,t_k}(u,v|x_j) du dv \right. \\ & + \delta_0(\beta_j) \frac{\Gamma(\beta_k+1) A_0(t_k|x_k)}{[\lambda_k a_k + A_0(t_k|x_k)]^{\beta_k+1}} + \delta_0(\beta_k) \frac{\Gamma(\beta_j+1) A_0(t_j|x_j)}{[\lambda_j a_j + A_0(t_j|x_j)]^{\beta_j+1}} \left. - \delta_0(\beta_j)\delta_0(\beta_k) \right\} \quad (7) \end{aligned}$$

where

$$g_{j,k}(u,v) := u^{\beta_j - 1}(\beta_j - \lambda_j a_j u) e^{-\lambda_j a_j u} v^{\beta_k - 1}(\beta_k - \lambda_k a_k v) e^{-\lambda_k a_k v}.$$

Theorem 2.2 *Under the assumptions of Lemma 2.1 and with kernel functions $K_1, ..., K_m$ satisfying $(\mathcal{K}_1)$, we have*

$$\sqrt{nh^p} \left(\mathbb{T}_n - \mathbb{E}[\mathbb{T}_n] \right) \rightsquigarrow \mathcal{N}_m(0, \Sigma),$$

where $\mathcal{N}_m$ denotes a $m-$dimensional normal distribution.

We have now all the needed ingredients for proving the asymptotic properties of the MDPD estimator $\widehat{A}_{\alpha,n}(t|x)$.

2.3 Asymptotic properties of $\widehat{A}_{\alpha,n}(t|x)$

The first result states the existence and uniform consistency of a sequence of solutions to the estimating equation (3).

Theorem 2.3 *Let $\alpha > 0$ and $x \in \mathbb{R}^p$ such that $f(x) > 0$. Under the assumptions of Theorem 2.1, with probability tending to 1, there exists a sequence $\left(\widehat{A}_{\alpha,n}(t|x)\right)_{n\in\mathbb{N}}$ of solutions for the estimating equation (3) such that*

$$\sup_{t\in[0,1]} \left|\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right| = o_{\mathbb{P}}(1).$$

Now, we come back to our final goal which is the weak convergence of the stochastic process (4).

Theorem 2.4 *Let $\alpha > 0$. Under the assumptions of Theorem 2.1, the process*

$$\left\{\sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right), t \in [0,1]\right\}$$

weakly converges in $\mathcal{C}([0,1])$ towards a tight centered Gaussian process $\{N_t, t \in [0,1]\}$ with covariance structure given by

$$Cov\,(N_t, N_s) = \frac{\|K\|_2^2\, A_0(t|x)A_0(s|x)}{f(x)} \frac{(1+\alpha)^2}{(1+\alpha^2)^2}\, v_\alpha^T \Sigma(t,s) v_\alpha,$$

where

$$v_\alpha := \begin{pmatrix} \dfrac{\alpha}{1+\alpha} \\ -(1+\alpha) \\ 1+\alpha \end{pmatrix} \quad \textit{and} \quad \Sigma(t,s) := \begin{pmatrix} (1+\alpha)^2 & 1+\alpha & 1 \\ 1+\alpha & \Sigma_{2,2}(t,s) & \Sigma_{2,3}(t,s) \\ 1 & \Sigma_{2,3}(s,t) & \Sigma_{3,3}(t,s) \end{pmatrix}$$

with

$$\begin{aligned}
\Sigma_{2,2}(t,s) &:= (1-\alpha)(1+\alpha) + \alpha^2(1+\alpha)^2\, A_0(t|x)A_0(s|x) \int_{\mathbb{R}_+^2} e^{-\alpha[A_0(t|x)u + A_0(s|x)v]} G_{t,s}(u,v|x)dudv \\
\Sigma_{2,3}(t,s) &:= 1 - \alpha(1+\alpha)^2\, A_0(t|x)A_0(s|x) \int_{\mathbb{R}_+^2} (1-\alpha A_0(s|x)v) e^{-\alpha[A_0(t|x)u + A_0(s|x)v]} G_{t,s}(u,v|x)dudv \\
\Sigma_{3,3}(t,s) &:= (1+\alpha)^2\, A_0(t|x)A_0(s|x) \int_{\mathbb{R}_+^2} (1-\alpha A_0(t|x)u)(1-\alpha A_0(s|x)v) e^{-\alpha[A_0(t|x)u + A_0(s|x)v]} G_{t,s}(u,v|x)dudv.
\end{aligned}$$

In particular, for all $t \in [0,1]$, we have

$$\sqrt{nh^p}\left(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right) \rightsquigarrow \mathcal{N}_1\left(0, \frac{\|K\|_2^2 [A_0(t|x)]^2}{f(x)} \frac{(1+\alpha)^2(1+4\alpha+9\alpha^2+14\alpha^3+13\alpha^4+8\alpha^5+4\alpha^6)}{(1+2\alpha)^3(1+\alpha^2)^2}\right),$$

as $n \to \infty$.

3 Case of unknown margins

In this section, we consider the general framework where both $F_1(.|x)$ and $F_2(.|x)$ are unknown conditional distribution functions. We want to mimic what has been done in the previous section in case where these conditional distributions are assumed to be standard exponential distribution functions. To this aim, we consider the triplets

$$\left(-\log\left(F_{n,1}(Y_i^{(1)}|X_i)\right), -\log\left(F_{n,2}(Y_i^{(2)}|X_i)\right), X_i\right), i = 1, ..., n,$$

for suitable estimators $F_{n,j}$ of F_j, $j = 1, 2$, and we compute our univariate random variables

$$\breve{Z}_{n,t,i} := \min\left(\frac{-\log\left(F_{n,1}(Y_i^{(1)}|X_i)\right)}{1-t}, \frac{-\log\left(F_{n,2}(Y_i^{(2)}|X_i)\right)}{t}\right), i = 1, ..., n.$$

Then, similarly as in Section 2, the statistic

$$\breve{T}_n(K, a, t, \lambda, \beta, \gamma|x) := \frac{a^\gamma}{n}\sum_{i=1}^{n} K_h(x - X_i)\breve{Z}_{n,t,i}^{\beta} e^{-\lambda a \breve{Z}_{n,t,i}}, \tag{8}$$

is the cornerstone of our MDPDE which satisfies the estimating equation

$$\breve{\Delta}^{(1)}_{\alpha,x,t}(A(t|x)) = 0, \tag{9}$$

where

$$\breve{\Delta}_{\alpha,x,t}(A(t|x)) := \frac{[A(t|x)]^\alpha}{n}\sum_{i=1}^{n} K_h(x - X_i)\left\{\frac{1}{1+\alpha} - \left(1 + \frac{1}{\alpha}\right)e^{-\alpha A(t|x)\breve{Z}_{n,t,i}}\right\}.$$

We denote it by $\breve{A}_{\alpha,n}(t|x)$. Our final goal is still the same, that is the weak convergence of the stochastic process

$$\left\{\sqrt{nh^p}\left(\breve{A}_{\alpha,n}(t|x) - A_0(t|x)\right), t \in [0, 1]\right\}. \tag{10}$$

Again this result relies essentially on the asymptotic properties of our statistic $\breve{T}_n$, and so the idea will be to decompose

$$\sqrt{nh^p}\left(\breve{T}_n - \mathbb{E}[\breve{T}_n]\right)(K, a, t, \lambda, \beta, \gamma|x),$$

into the two terms

$$\left\{\sqrt{nh^p}\left(T_n - \mathbb{E}[T_n]\right)(K, a, t, \lambda, \beta, \gamma|x)\right\} + \left\{\sqrt{nh^p}\left([\breve{T}_n - T_n] - \mathbb{E}[\breve{T}_n - T_n]\right)(K, a, t, \lambda, \beta, \gamma|x)\right\}. \tag{11}$$

The first one can be dealt with using the results of Section 2.2 whereas we have to show that the second term is uniformly negligible.

To achieve this objective, let us introduce the following empirical kernel estimator of our unknown conditional distribution functions

$$F_{n,j}(y|x) := \frac{\sum_{i=1}^n K_c(x - X_i)\mathbb{1}_{\{Y_i^{(j)} \leqslant y\}}}{\sum_{i=1}^n K_c(x - X_i)},\ j = 1, 2,$$

where $c := c_n$ is a positive non-random sequence satisfying $c_n \to 0$ as $n \to \infty$. Here we kept the same kernel K as in the previous section, but of course any other kernel function can be used.

Before stating our main results, we need to impose again some assumptions, in particular a Hölder-type condition on each marginal conditional distribution function F_j similar to the one imposed on the density function of the covariate.

Assumption $(\mathcal{F})$**.** *There exist $M_{F_j} > 0$ and $\eta_{F_j} > 0$ such that $|F_j(y|x) - F_j(y|z)| \leqslant M_{F_j}\|x - z\|^{\eta_{F_j}}$, for all $y \in \mathbb{R}$ and all $(x, z) \in \mathbb{R}^p \times \mathbb{R}^p$ and $j = 1, 2$.*

Concerning the kernel K a stronger assumption than $(\mathcal{K}_1)$ is needed.

Assumption $(\mathcal{K}_2)$**.** *K satisfies Assumption $(\mathcal{K}_1)$ and belongs to the linear span (the set of finite linear combinations) of functions $k \geqslant 0$ satisfying the following property: the subgraph of k, $\{(s, u) : k(s) \geqslant u\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(s, u) : q(s, u) \geqslant \varphi(u)\}$, where q is a polynomial on $\mathbb{R}^p \times \mathbb{R}$ and φ is an arbitrary real function.*

This hypothesis has been already used in Giné and Guillou (2002) or Giné *et al.* (2004). In particular, it allows us to measure the discrepancy between the conditional distribution function F_j and its empirical kernel version $F_{n,j}$ under an additional assumption on the density of the covariate X, similar to the one already used in Portier and Segers (2015).

Lemma 3.1 *Assume that there exists $b > 0$ such that $f(x) \geqslant b, \forall x \in S_X \subset \mathbb{R}^p$, the support of f. If $(\mathcal{K}_2)$, $(\mathcal{D})$ and $(\mathcal{F})$ hold and for $q > 1$*

$$\frac{|\log c|^q}{nc^p} \longrightarrow 0,$$

as $n \to \infty$, then for any $0 < \eta < \min(\eta_f, \eta_{F_1}, \eta_{F_2})$, we have

$$\mathbb{E}\left[\sup_{(y,x)\in\mathbb{R}\times S_X} |F_{n,j}(y|x) - F_j(y|x)|\right] = o\left(\max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right)\right),\ \text{for } j = 1, 2.$$

We are now able to study the second term in (11).

Theorem 3.1 *Assume $(\mathcal{K}_2)$, $(\mathcal{D})$ and $(\mathcal{F})$ and that there exists $b > 0$ such that $f(x) \geqslant b, \forall x \in S_X \subset \mathbb{R}^p$. Consider two sequences h and c tending to 0, such that for $nh^p \to \infty$ and $q > 1$*

$$\sqrt{nh^p}\, r_n := \sqrt{nh^p} \max\left(\sqrt{\frac{|\log c|^q}{nc^p}}, c^\eta\right) \longrightarrow 0,$$

as $n \to \infty$. Then, for all $\gamma \in \mathbb{R}$ and $(\lambda, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+$ or $(\lambda, \beta) = (0, 0)$, and $x \in \mathbb{R}^p$, we have

$$\sup_{t\in[0,1], a\in[1/2,1]} \sqrt{nh^p}\left|\breve{T}_n - T_n - \mathbb{E}\left[\breve{T}_n - T_n\right]\right|(K, a, t, \lambda, \beta, \gamma|x) = o_{\mathbb{P}}(1).$$

Finally, the decomposition (11) combined with Theorem 3.1 and the results from Section 2.2, yields the desired final goal of this paper.

Theorem 3.2 *Let $\alpha > 0$. Under the assumptions of Theorem 3.1 and $(\mathcal{A}_0)$, with probability tending to 1, there exists a sequence $\left(\breve{A}_{\alpha,n}(t|x)\right)_{n\in\mathbb{N}}$ of solutions for the estimating equation (9) such that*

$$\sup_{t\in[0,1]} \left|\breve{A}_{\alpha,n}(t|x) - A_0(t|x)\right| = o_{\mathbb{P}}(1).$$

Moreover, if $\sqrt{nh^p}h^{\min(\eta_f,\eta_{A_0})} \to 0$, the process

$$\left\{\sqrt{nh^p}\left(\breve{A}_{\alpha,n}(t|x) - A_0(t|x)\right), t \in [0,1]\right\},$$

weakly converges in $\mathcal{C}([0,1])$ towards the tight centered Gaussian process $\{N_t, t \in [0,1]\}$ defined in Theorem 2.4.

4 A small simulation study

Our aim in this section is to illustrate the efficiency of our robust estimator of the conditional Pickands dependence function on a small simulation study.

We assume that the conditional distribution function of $(Y^{(1)}, Y^{(2)})$ given $X = x$ is the logistic distribution given by

$$F(y_1, y_2|x) := \exp\left\{-\left(y_1^{-1/x} + y_2^{-1/x}\right)^x\right\}, \text{ for } y_1, y_2 \geqslant 0$$

and

$$A_0(t|x) = \left(t^{1/x} + (1-t)^{1/x}\right)^x,$$

where the covariate X is a uniformly distributed random variable on $[0,1]$. For this model, the complete dependence is obtained in the limit as $x \downarrow 0$, whereas independence can be reached for $x = 1$. Note also that the conditional marginal distributions of $Y^{(j)}$ given X, $j = 1, 2$, under this logistic model are unit Fréchet distributions.
To compute our estimator $\breve{A}_{\alpha,n}$, two sequences, h and c, have to be chosen. Concerning c, we can use the following cross validation criterion introduced by Yao (1999), implemented by Gannoun *et al.* (2002), and already used in an extreme value context by Daouia *et al.* (2011, 2013) or Goegebeur *et al.* (2015)

$$c_j := \arg\min_{c\in\mathcal{C}} \sum_{i=1}^{n}\sum_{k=1}^{n}\left[\mathbb{1}_{\left\{Y_i^{(j)} \leqslant Y_k^{(j)}\right\}} - \widetilde{F}_{n,-i,j}(Y_k^{(j)}|X_i)\right]^2, \ j = 1, 2,$$

where $\mathcal{C}$ is a grid of values of c and $\widetilde{F}_{n,-i,j}(y|x) := \dfrac{\sum_{k=1,k\neq i}^{n} K_c(x - X_k)\mathbb{1}_{\{Y_k^{(j)} \leqslant y\}}}{\sum_{k=1,k\neq i}^{n} K_c(x - X_k)}$.

It remains to select the sequence h, which can be done from the condition

$$\sqrt{nh^p}\sqrt{\frac{|\log c|^q}{nc^p}} \to 0 \tag{12}$$

by taking $h = c/|\log c|^\xi$, where $\xi p > q$ and $c := \min(c_1, c_2)$.

The procedure of contamination used in this paper is the following.

1. we simulate n triplets $(Y_1^{(1)}, Y_1^{(2)}, X_1), ..., (Y_n^{(1)}, Y_n^{(2)}, X_n)$, independently according to the distributions previously mentioned;

2. we simulate independently $n_0 := \lfloor n\varepsilon \rfloor$ variables $\widetilde{Y}^{(1)}$ and $\widetilde{Y}^{(2)}$, from a unit Fréchet distribution and also n_0 variables $\widetilde{X}$, independently from the same distribution as X;

3. we consider the $(n + n_0)$ triplets:

$$\left(\overline{Y}_i^{(1)}, \overline{Y}_i^{(2)}, \overline{X}_i\right) := \begin{cases} \left(Y_i^{(1)}, Y_i^{(2)}, X_i\right) & \text{if } i = 1, ..., n, \\ \left(Y_{1,n}^{(1)}, \widetilde{Y}_i^{(2)}, \widetilde{X}_i\right) & \text{if } i = n+1, ..., n + \lfloor n_0/2 \rfloor, \\ \left(\widetilde{Y}_i^{(1)}, Y_{1,n}^{(2)}, \widetilde{X}_i\right) & \text{if } i = n + \lfloor n_0/2 \rfloor + 1, ..., n + n_0, \end{cases}$$

   where $Y_{1,n}^{(j)} = \min_{1 \leqslant i \leqslant n} Y_i^{(j)}$, $j = 1, 2$;

4. we apply our methodology to the resulting univariate variables

$$\breve{Z}_{n,t,i} = \min\left(\frac{-\log\left(F_{n+n_0,1}\left(\overline{Y}_i^{(1)}|\overline{X}_i\right)\right)}{1-t}, \frac{-\log\left(F_{n+n_0,2}\left(\overline{Y}_i^{(2)}|\overline{X}_i\right)\right)}{t}\right), i = 1, ..., n + n_0.$$

This procedure of contamination is illustrated in Figure 1, where the non-contaminated sample of 1000 pairs is represented as circles whereas the contaminated pairs are represented as crosses. Here, the percentage of contamination is 5%. This scatterplot is obtained before the empirical transformation of the margins into unit exponential distributions on the left panel and after this transformation on the right panel.

The percentage of contamination is set to $\varepsilon = 0\%$, 5% and 10%, while $n = 1000$ and the procedure is repeated $N = 100$ times. Concerning the kernel, each time we use the bi-quadratic function

$$K(x) := \frac{15}{16}(1 - x^2)^2 \mathbb{1}_{[-1,1]}(x).$$

In all the settings, $\mathcal{C} = \{0.06, 0.12, 0.18, 0.24, 0.3\}$ and $\xi = 1.1$ are used, since an extensive simulation study indicates that these choices seem to give always reasonable results.

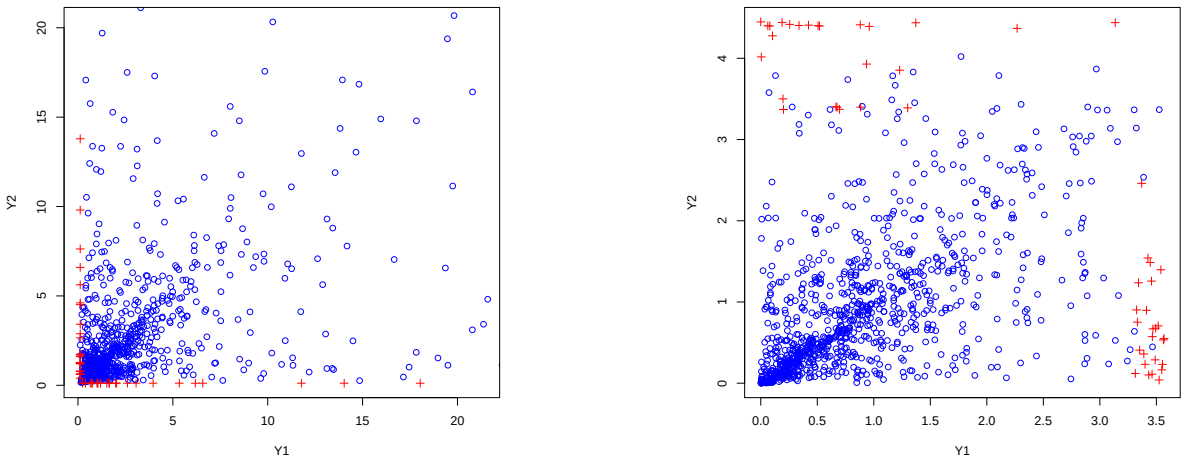

Figure 1: Procedure of contamination: on the left before step 4, on the right after the empirical transformation of the margins into unit exponential distributions. The non-contaminated sample of 1000 pairs is represented as circles whereas the contaminated pairs are represented as crosses.

As indicators of efficiency we compute the bias and the mean squared error

$$\begin{aligned}\text{Bias}(\breve{A}_{\alpha,n}(t|.)) &:= \frac{1}{MN}\sum_{m=1}^{M}\sum_{i=1}^{N}\left|\breve{A}^{(i)}_{\alpha,n}(t|x_m) - A_0(t|x_m)\right|,\\ \text{MSE}(\breve{A}_{\alpha,n}(t|.)) &:= \frac{1}{MN}\sum_{m=1}^{M}\sum_{i=1}^{N}\left[\breve{A}^{(i)}_{\alpha,n}(t|x_m) - A_0(t|x_m)\right]^2,\end{aligned}$$

where $\breve{A}^{(i)}_{\alpha,n}(t|x_m)$ is the estimate of $A_0(t|x_m)$ obtained with the i^{th} sample evaluated at points $t \in [0,1]$ and $x_m = m/(M+1)$, $m = 1, \cdots, M$. Our method is implemented for $M = 9$ and $t = i/50$, $i = 1, \cdots, 49$.

Figures 2 till 4 represent the boxplots of our estimator $\breve{A}_{\alpha,n}(.|x)$ based on N samples for three positions: $x = 0.3, 0.5$ and 0.7, respectively. Three values of α have been reported: 0.1, 0.5, and 1 corresponding to the different rows on each figure, whereas three different percentages of contamination have been used: from the left to the right of each figure: $0\%, 5\%$ and 10%. Based on these simulations, we can draw the following conclusions:

- in case of uncontaminated datasets, the best estimator is achieved with $\alpha = 0.1$, although there are no big differences. This is not surprising since when $\alpha = 0$, one recovers the maximum likelihood estimator which is well-known to be efficient, but not robust;
- in case of contamination, a larger value of α (0.5 or 1) is needed;
- as expected, increasing the percentage of contamination negatively affects the estimator, whatever the distribution and value of α;

- for x close to 0, which corresponds to the complete dependence situation, the estimation of the conditional Pickands dependence function seems to be slightly more difficult in terms of bias than in case x close to 1, that is, the independent case.

Then, in Figure 5, we show from the top to the bottom, the bias and the mean squared error of our estimator $\check{A}_{\alpha,n}(t|.)$ based on N samples for the three percentages of contamination, from the left to the right $0\%, 5\%$ and 10%. Again the three same values of α have been tried. The same conclusions can be made, i.e. increasing the percentage of contamination deteriorates the performance of the estimator. Also, the estimator is not too much sensitive on the value of α in the uncontaminated case, whereas in the other cases a larger value of α increases significantly the performance of the estimator. In fact $\alpha = 0.5$ seems to be almost the best value in all the settings considered.

Note that we have tried other types of distributions, positions, values of α and kernels, but still the same conclusions remain. To keep the length of the paper under control, they are not included.

5 Appendix A: Technical results

5.1 Asymptotic covariance matrix of the finite dimensional vector $\mathbb{T}_n$

Our aim in this section is to compute the explicit expression of the elements of the covariance matrix $\Sigma = (\sigma_{j,k})_{1\leqslant j,k\leqslant m}$ given in (7). According to (17), we have

$$\mathbb{E}\left[K_{j,h}(x_j - X_1)Z_{t_j,1}^{\beta_j}e^{-\lambda_j a_j Z_{t_j,1}}\right] = \frac{f(x_j)\Gamma(\beta_j+1)A_0(t_j|x_j)}{(\lambda_j a_j + A_0(t_j|x_j))^{\beta_j+1}}(1+o(1)),$$

for $1 \leqslant j \leqslant m$. In order to compute the cross expectation, we need to derive the conditional distribution function of the pair (Z_{t_j}, Z_{t_k}) given $X = x$. Let $u, v > 0$

$$\begin{aligned}\mathbb{P}\left(Z_{t_j} > u, Z_{t_k} > v | X = x\right) &= \mathbb{P}\left(Y^{(1)} > \max\left((1-t_j)u, (1-t_k)v\right), Y^{(2)} > \max\left(t_j u, t_k v\right)\Big| X = x\right)\\ &= G\left(\max\left((1-t_j)u, (1-t_k)v\right), \max\left(t_j u, t_k v\right)\Big| x\right).\end{aligned}$$

Hence, for $j, k \in \{1, ..., m\}^2$, we have

$$\begin{aligned}&\mathbb{E}\left[K_{j,h}(x_j - X_1)Z_{t_j,1}^{\beta_j}e^{-\lambda_j a_j Z_{t_j,1}}K_{k,h}(x_k - X_1)Z_{t_k,1}^{\beta_k}e^{-\lambda_k a_k Z_{t_k,1}}\right] \qquad (13)\\ &\quad = \mathbb{E}\left[K_{j,h}(x_j - X_1)K_{k,h}(x_k - X_1)\mathbb{E}\left[Z_{t_j,1}^{\beta_j}e^{-\lambda_j a_j Z_{t_j,1}}Z_{t_k,1}^{\beta_k}e^{-\lambda_k a_k Z_{t_k,1}}\Big| X_1\right]\right].\end{aligned}$$

We focus now on the conditional expectation. Using the fact that

$$z^{\beta}e^{-a\lambda z} - \delta_0(\beta) = \int_{\mathbb{R}_+} \mathbb{1}_{\{z>u\}}u^{\beta-1}(\beta - a\lambda u)e^{-a\lambda u}du, \qquad (14)$$

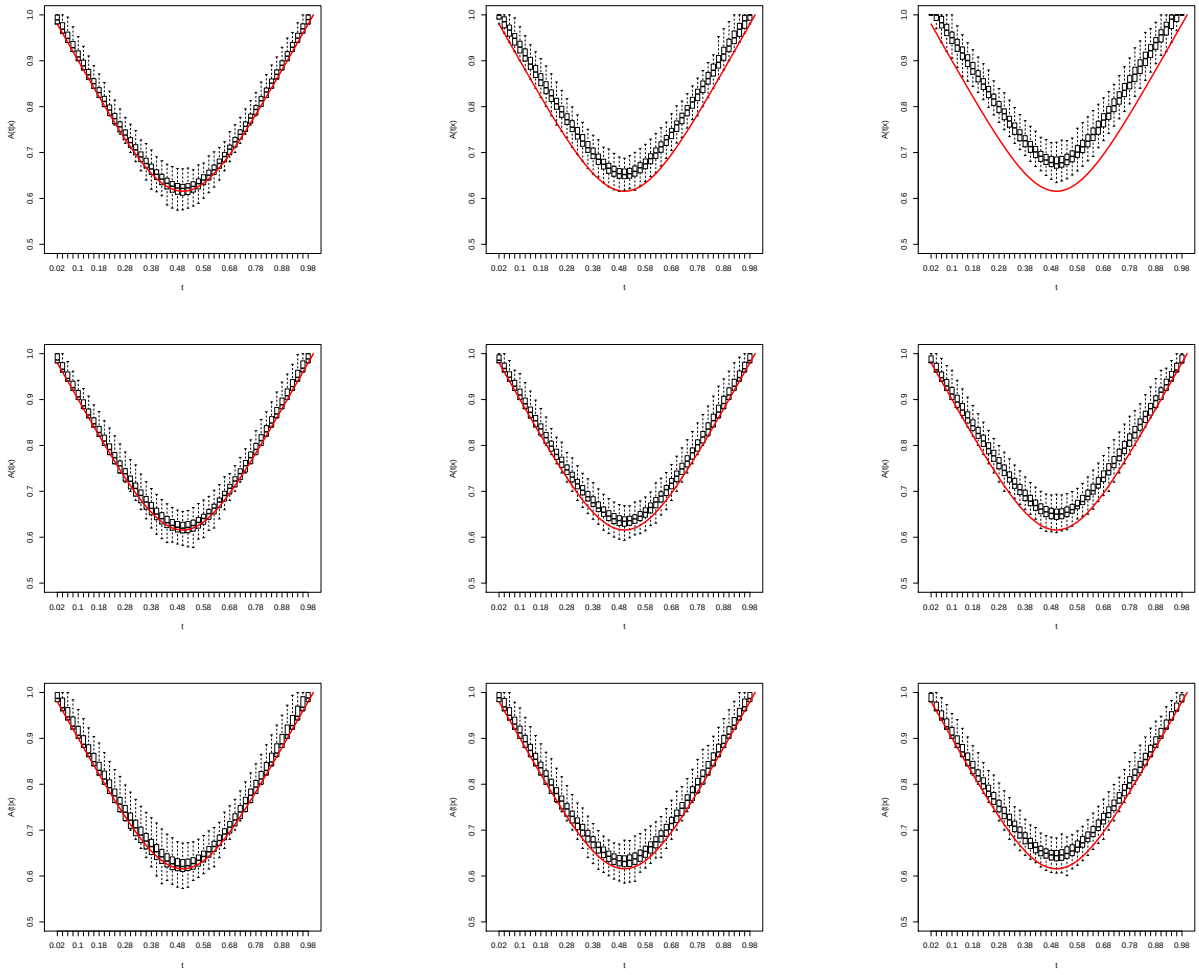

Figure 2: Estimation of $A_0(.|0.3)$ (full line) for the logistic distribution. From the top to the bottom: $\alpha = 0.1, 0.5, 1$ and from the left to the right: $0\%, 5\%$ and 10% of contamination.

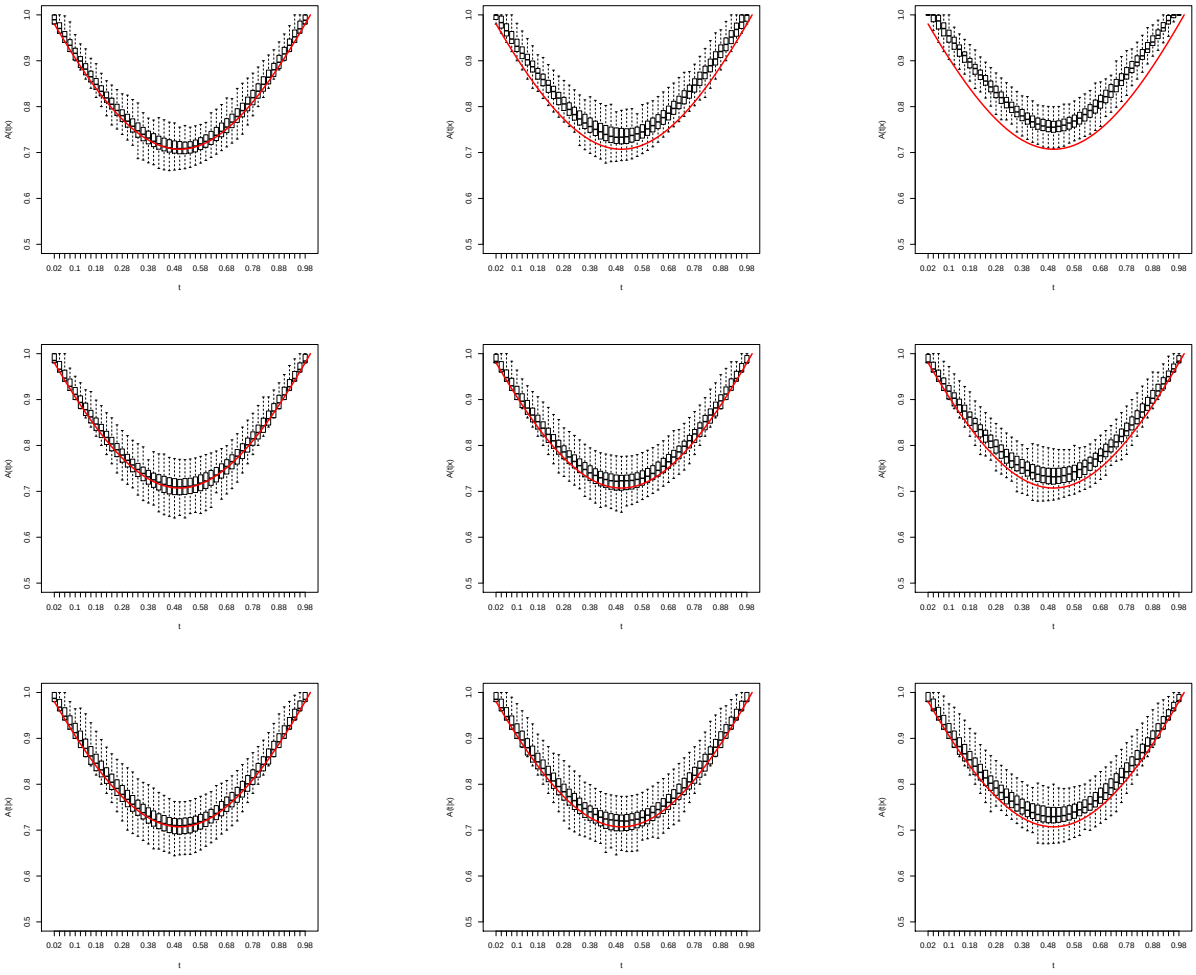

Figure 3: Estimation of $A_0(.|0.5)$ (full line) for the logistic distribution. From the top to the bottom: $\alpha = 0.1, 0.5, 1$ and from the left to the right: $0\%, 5\%$ and 10% of contamination.

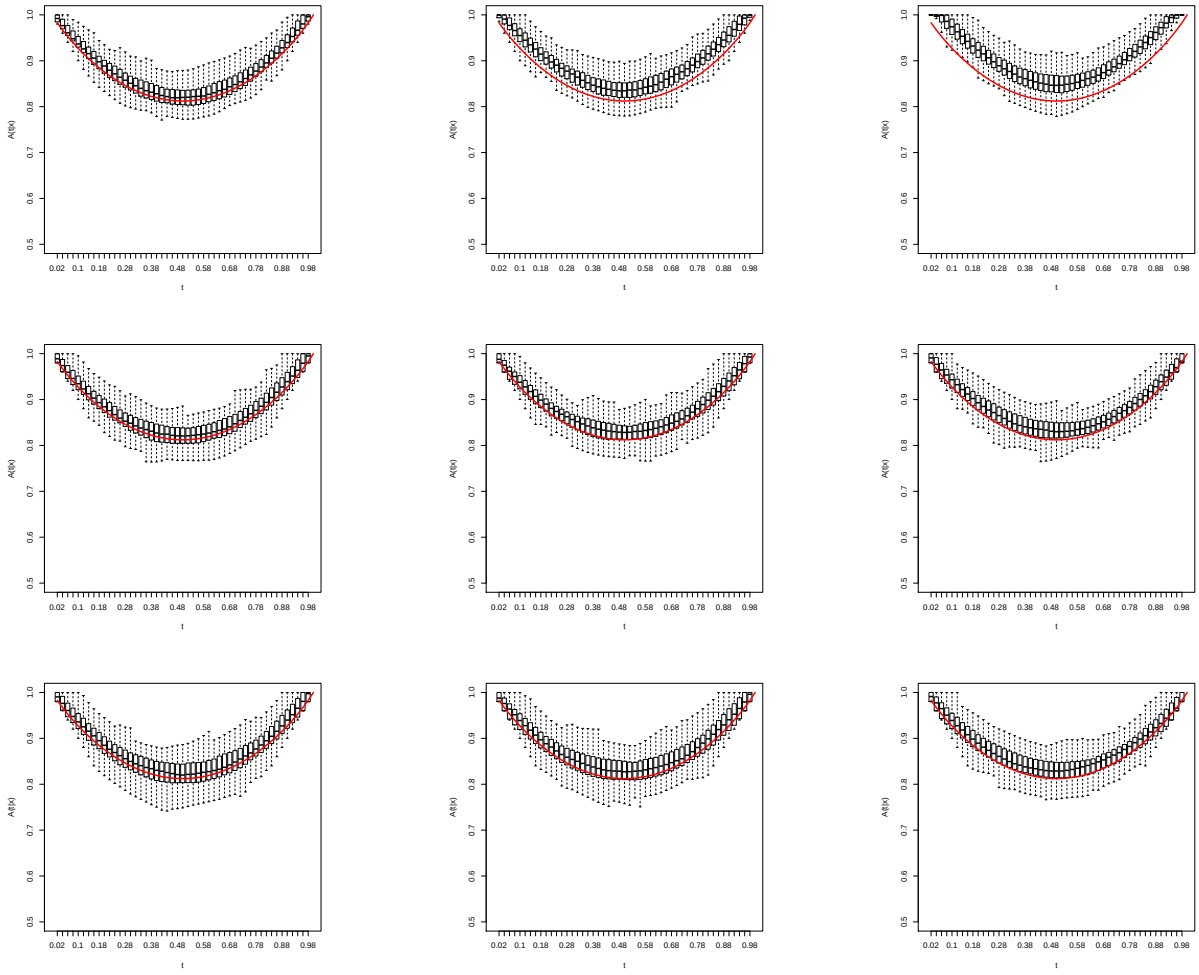

Figure 4: Estimation of $A_0(.|0.7)$ (full line) for the logistic distribution. From the top to the bottom: $\alpha = 0.1, 0.5, 1$ and from the left to the right: $0\%, 5\%$ and 10% of contamination.

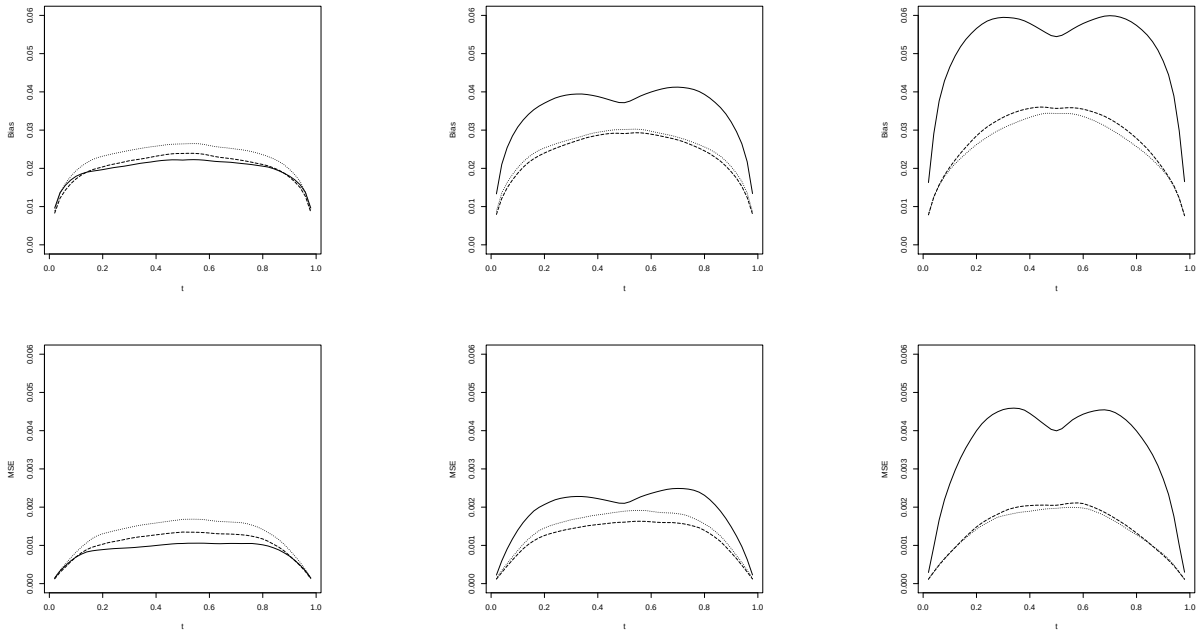

Figure 5: Estimation of $A_0(.|.)$ for the logistic distribution. From the top to the bottom: bias and MSE and from the left to the right: $0\%, 5\%$ and 10% of contamination. Three values of α: 0.1 (full line), 0.5 (dashed line) and 1 (dotted line).

we have

$$\begin{aligned}
&\mathbb{E}\left[Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}} Z_{t_k,1}^{\beta_k} e^{-\lambda_k a_k Z_{t_k,1}} \Big| X_1 \right] \\
&\quad = \mathbb{E}\left[\left(\delta_0(\beta_j) + \int_{\mathbb{R}_+} \mathbb{1}_{\{Z_{t_j}>u\}} u^{\beta_j-1}(\beta_j - \lambda_j a_j u) e^{-\lambda_j a_j u} du \right) \right. \\
&\qquad \left. \times \left(\delta_0(\beta_k) + \int_{\mathbb{R}_+} \mathbb{1}_{\{Z_{t_k}>v\}} v^{\beta_k-1}(\beta_k - \lambda_k a_k v) e^{-\lambda_k a_k v} dv \right) \Big| X_1 \right] \\
&\quad = \int_{\mathbb{R}_+^2} g_{j,k}(u,v) G_{t_j,t_k}\left(u, v \Big| X_1\right) du dv \\
&\qquad + \delta_0(\beta_j) \frac{\Gamma(\beta_k+1) A_0(t_k|X_1)}{[\lambda_k a_k + A_0(t_k|X_1)]^{\beta_k+1}} + \delta_0(\beta_k) \frac{\Gamma(\beta_j+1) A_0(t_j|X_1)}{[\lambda_j a_j + A_0(t_j|X_1)]^{\beta_j+1}} - \delta_0(\beta_j)\delta_0(\beta_k). \quad (15)
\end{aligned}$$

In case $x_j = x_k$, combining the continuity and boundedness of the functions f, $A_0(t|.)$ and $G(u,v|.)$, the expression of $\sigma_{j,k}$ in (7) follows.
In case $x_j \neq x_k$, it is sufficient to observe that, for n large enough,

$$\mathbb{E}\left[K_{j,h}(x_j - X_1) K_{k,h}(x_k - X_1) \mathbb{E}\left[Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}} Z_{t_k,1}^{\beta_k} e^{-\lambda_k a_k Z_{t_k,1}} \Big| X_1 \right] \right] = 0,$$

since for $h < \dfrac{\|x_j - x_k\|}{2}$, we clearly have $\left\{ u : \dfrac{\|x_j - u\|}{h} \leqslant 1 \right\} \bigcap \left\{ u : \dfrac{\|x_k - u\|}{h} \leqslant 1 \right\} = \varnothing$.

5.2 The derivatives of $\widehat{\Delta}_{\alpha,x,t}$ and their asymptotic properties

Straightforward computations for $a \in [1/2, 1]$, $\alpha > 0$ and $x \in \mathbb{R}^p$, give

$$\begin{aligned}
\widehat{\Delta}^{(1)}_{\alpha,x,t}(a) &= \alpha a^{-1} \widehat{\Delta}_{\alpha,x,t}(a) + a^{\alpha}(1+\alpha) \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i) Z_{t,i} e^{-\alpha a Z_{t,i}}, \\
\widehat{\Delta}^{(2)}_{\alpha,x,t}(a) &= \alpha a^{-1} \widehat{\Delta}^{(1)}_{\alpha,x,t}(a) - \alpha a^{-2} \widehat{\Delta}_{\alpha,x,t}(a) + \alpha(\alpha+1) a^{\alpha-1} \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(1 - a Z_{t,i}) Z_{t,i} e^{-\alpha a Z_{t,i}}, \\
\widehat{\Delta}^{(3)}_{\alpha,x,t}(a) &= \alpha \left(2a^{-3} \widehat{\Delta}_{\alpha,x,t}(a) + a^{-1} \widehat{\Delta}^{(2)}_{\alpha,x,t}(a) - 2a^{-2} \widehat{\Delta}^{(1)}_{\alpha,x,t}(a) \right) \\
&\quad + (\alpha-1)\alpha(\alpha+1) \frac{a^{\alpha-2}}{n} \sum_{i=1}^{n} K_h(x - X_i)(1 - a Z_{t,i}) Z_{t,i} e^{-\alpha a Z_{t,i}} \\
&\quad - \alpha(\alpha+1) \frac{a^{\alpha-1}}{n} \sum_{i=1}^{n} K_h(x - X_i)(\alpha(1 - a Z_{t,i}) + 1) Z_{t,i}^2 e^{-\alpha a Z_{t,i}}.
\end{aligned}$$

The convergence in probability of the three first derivatives of $\widehat{\Delta}_{\alpha,x,t}$ and of $\widehat{\Delta}_{\alpha,x,t}$ itself is

therefore a direct application of Lemma 2.1, which yields, as $n \to \infty$

$$\begin{aligned}
\widehat{\Delta}_{\alpha,x,t}(a) \xrightarrow{\mathbb{P}} \ell^{(0)}_{\alpha,x,t}(a) &:= a^{\alpha} f(x) \left(\frac{1}{\alpha+1} - \left(1+\frac{1}{\alpha}\right) \frac{A_0(t|x)}{\alpha a + A_0(t|x)} \right), \\
\widehat{\Delta}^{(1)}_{\alpha,x,t}(a) \xrightarrow{\mathbb{P}} \ell^{(1)}_{\alpha,x,t}(a) &:= \alpha a^{-1} \ell^{(0)}_{\alpha,x,t}(a) + (1+\alpha) a^{\alpha} f(x) \frac{A_0(t|x)}{(\alpha a + A_0(t|x))^2}, \\
\widehat{\Delta}^{(2)}_{\alpha,x,t}(a) \xrightarrow{\mathbb{P}} \ell^{(2)}_{\alpha,x,t}(a) &:= \alpha a^{-1} \ell^{(1)}_{\alpha,x,t}(a) - \alpha a^{-2} \ell^{(0)}_{\alpha,x,t}(a) \\
&\quad + \alpha(\alpha+1) a^{\alpha-1} f(x) \left(\frac{A_0(t|x)}{(\alpha a + A_0(t|x))^2} - \frac{2a A_0(t|x)}{(\alpha a + A_0(t|x))^3} \right), \\
\widehat{\Delta}^{(3)}_{\alpha,x,t}(a) \xrightarrow{\mathbb{P}} \ell^{(3)}_{\alpha,x,t}(a) &:= \alpha \left(2a^{-3} \ell^{(0)}_{\alpha,x,t}(a) + a^{-1} \ell^{(2)}_{\alpha,x,t}(a) - 2a^{-2} \ell^{(1)}_{\alpha,x,t}(a) \right) \\
&\quad + (\alpha-1)\alpha(\alpha+1) a^{\alpha-2} f(x) \left[\frac{A_0(t|x)}{(\alpha a + A_0(t|x))^2} - \frac{2a A_0(t|x)}{(\alpha a + A_0(t|x))^3} \right] \\
&\quad - \alpha(\alpha+1) a^{\alpha-1} f(x) \left[2(\alpha+1) \frac{A_0(t|x)}{(\alpha a + A_0(t|x))^3} - \alpha \frac{6a A_0(t|x)}{(\alpha a + A_0(t|x))^4} \right].
\end{aligned}$$

Now the rate of convergence of $\widehat{\Delta}^{(j)}_{\alpha,x,t}(A_0(t|x)), j \in \{1,2,3\}$, to its limit is also useful to study (5) and thus to reach our final goal. The aim of the next corollary is to provide such a rate.

Corollary 5.1 *Under the assumptions of Theorem 2.1, then for any $j \in \{1,2,3\}$, the process*

$$\left\{ \sqrt{nh^p} \left(\widehat{\Delta}^{(j)}_{\alpha,x,t}(A_0(t|x)) - \ell^{(j)}_{\alpha,x,t}(A_0(t|x)) \right), t \in [0,1] \right\}$$

weakly converges in $\mathcal{C}([0,1])$ towards a tight centered Gaussian process. In particular, we have

$$\sup_{t \in [0,1]} \left| \widehat{\Delta}^{(j)}_{\alpha,x,t}(A_0(t|x)) - \ell^{(j)}_{\alpha,x,t}(A_0(t|x)) \right| = o_{\mathbb{P}}(1).$$

Proof of Corollary 5.1. As usual, it is sufficient to show the finite dimensional convergence and the tightness of the process. Using Theorem 2.2 we directly solve the finite dimensional convergence issue. Next, Theorem 2.1 combined with (18) implies tightness for any process $t \to \sqrt{nh^p}(T_n - \mathbb{E}[T_n])(K, A_0(t|x), t, \lambda, \beta, \gamma|x)$ and similarly as in Lemma 1 in Bai and Taqqu (2013), we have tightness for any multivariate process with similar coordinates. Corollary 5.1 then follows.

6 Appendix B: Proofs of the main results

6.1 Proof of Lemma 2.1

Using the fact that the conditional distribution function of Z_t given $X = x$ is an exponential distribution with parameter $A_0(t|x)$ and since $\lambda a + A_0(t|x) > 0$, we have

$$\mathbb{E}\left[Z_t^{\beta} e^{-\lambda a Z_t} \middle| X = x \right] = \Gamma(\beta+1) \frac{A_0(t|x)}{(\lambda a + A_0(t|x))^{\beta+1}}. \tag{16}$$

Then

$$\begin{aligned}
\mathbb{E}\left[K_h(x-X)Z_t^{\beta}e^{-\lambda a Z_t}\right] &= \mathbb{E}\left[K_h(x-X)\Gamma(\beta+1)\frac{A_0(t|X)}{(\lambda a+A_0(t|X))^{\beta+1}}\right] \\
&= \Gamma(\beta+1)\int_{\mathbb{R}^p} K_h(x-y)\frac{A_0(t|y)}{(\lambda a+A_0(t|y))^{\beta+1}}f(y)dy \\
&= \Gamma(\beta+1)\int_{\mathbb{S}^{p-1}} K(z)\frac{A_0(t|x-zh)}{(\lambda a+A_0(t|x-zh))^{\beta+1}}f(x-hz)dz \\
&= \Gamma(\beta+1)\frac{A_0(t|x)}{(\lambda a+A_0(t|x))^{\beta+1}}f(x)(1+o(1)), \qquad (17)
\end{aligned}$$

by the dominated convergence theorem using the continuity of $A_0(t|.)$ and f on $x-\mathbb{S}^{p-1}$ together with the boundedness of $A_0(t|.)$. Consequently

$$\mathbb{E}[T_n(K,a,t,\lambda,\beta,\gamma|x)] = a^{\gamma}\Gamma(\beta+1)\frac{A_0(t|x)}{(\lambda a+A_0(t|x))^{\beta+1}}f(x)(1+o(1)).$$

Also, similar arguments yield

$$\mathrm{Var}(T_n(K,a,t,\lambda,\beta,\gamma|x)) = \frac{1}{nh^p}\,\frac{\|K\|_2^2\,A_0(t|x)\,a^{2\gamma}\,\Gamma(2\beta+1)\,f(x)}{(2\lambda a+A_0(t|x))^{2\beta+1}}(1+o(1)) = o(1),$$

from which the convergence in probability simply follows.

6.2 Proof of Theorem 2.1

First, remark that to show Theorem 2.1, it is sufficient to look at the weak convergence of the process

$$\left\{\sqrt{nh^p}\left(T_n(K,A_0(t|x),t,\lambda,\beta,\gamma|x)-\mathbb{E}\left[T_n(K,A_0(t|x),t,\lambda,\beta,\gamma|x)\right]\right), t\in[0,1]\right\}, \qquad (18)$$

since

$$\lim_{n\to\infty}\sup_{t\in[0,1]}\sqrt{nh^p}\left|\mathbb{E}\left[T_n(K,A_0(t|x),t,\lambda,\beta,\gamma|x)\right]-\Gamma(\beta+1)\frac{[A_0(t|x)]^{\gamma-\beta}}{(\lambda+1)^{\beta+1}}f(x)\right| = 0.$$

Indeed, according to (17), we have

$$\begin{aligned}
&\left|\mathbb{E}\left[T_n(K,A_0(t|x),t,\lambda,\beta,\gamma|x)\right]-\Gamma(\beta+1)\frac{[A_0(t|x)]^{\gamma-\beta}}{(\lambda+1)^{\beta+1}}f(x)\right| \\
&\quad\leqslant \Gamma(\beta+1)A_0^{\gamma}(t|x)\int_{\mathbb{S}^{p-1}}K(y)\left|\frac{A_0(t|x-yh)}{(\lambda A_0(t|x)+A_0(t|x-yh))^{\beta+1}}f(x-hy)-\frac{A_0^{-\beta}(t|x)}{(\lambda+1)^{\beta+1}}f(x)\right|dy.
\end{aligned}$$

Now, using Assumptions $(\mathcal{D})$ and $(\mathcal{A}_0)$, we deduce that

$$\begin{aligned}
&\left|\frac{A_0(t|x-yh)}{(\lambda A_0(t|x)+A_0(t|x-yh))^{\beta+1}}f(x-hy)-\frac{A_0^{-\beta}(t|x)}{(\lambda+1)^{\beta+1}}f(x)\right| \\
&\leqslant \frac{A_0(t|x-yh)}{(\lambda A_0(t|x)+A_0(t|x-yh))^{\beta+1}}|f(x-yh)-f(x)| + \left|\frac{A_0(t|x-yh)}{(\lambda A_0(t|x)+A_0(t|x-yh))^{\beta+1}}-\frac{A_0^{-\beta}(t|x)}{(\lambda+1)^{\beta+1}}\right|f(x) \\
&= O(h^{\min(\eta_f,\eta_{A_0})})
\end{aligned}$$

for n large enough such that $h \leqslant 1$, with a bound which is uniform in t.

Then, to show the weak convergence of the stochastic process (18), we will use Theorem 19.28 in van der Vaart (1998). To apply this result, we need to introduce some notations. Define the covering number $N(\mathcal{F}, L_2(Q), \tau)$ as the minimal number of $L_2(Q)$-balls of radius τ needed to cover the class of functions $\mathcal{F}$ and the uniform entropy integral as

$$J(\delta, \mathcal{F}, L_2) := \int_0^\delta \sqrt{\log \sup_{\mathcal{Q}} N(\mathcal{F}, L_2(Q), \tau \|F\|_{Q,2})}\, d\tau,$$

where $\mathcal{Q}$ is the set of all probability measures Q for which $0 < \|F\|_{Q,2}^2 := \int F^2 dQ < \infty$ and F is an envelope function of the class $\mathcal{F}$.

Let P denote the law of the vector $(Y^{(1)}, Y^{(2)}, X)$ and define the expectation under P, the empirical version and empirical process as follows

$$Pf := \int f dP, \quad \mathbb{P}_n f := \frac{1}{n} \sum_{i=1}^n f\left(Y_i^{(1)}, Y_i^{(2)}, X_i\right), \quad \mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n - P)f,$$

for any real-valued measurable function f.
For any $\gamma \in \mathbb{R}$ and $(\lambda, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+$ or $(\lambda, \beta) = (0, 0)$, we introduce our sequence of classes $\mathcal{F}_n$ as

$$\begin{aligned} \mathcal{F}_n &:= \{(y_1, y_2, z) \to f_{n,t}(y_1, y_2, z), t \in [0,1]\} \\ &:= \left\{(y_1, y_2, z) \to \sqrt{h^p} K_h(x - z)[A_0(t|x)]^{\gamma-\beta}[A_0(t|x) Z_t(y_1, y_2)]^\beta e^{-\lambda A_0(t|x) Z_t(y_1, y_2)}, t \in [0,1]\right\}, \end{aligned}$$

where $Z_t(y_1, y_2) := \min\left(\frac{y_1}{1-t}, \frac{y_2}{t}\right)$. Remark that $Z_t = Z_t\left(Y^{(1)}, Y^{(2)}\right)$. Denote now by F_n an envelope function of the class $\mathcal{F}_n$ and for any $y \in \mathbb{R}^p$, define the bivariate function $\rho_y : [0,1]^2 \to \mathbb{R}_+$ as

$$\rho_y(t, s) := \mathbb{E}\left[\left(A_0^\gamma(t|x) Z_t^\beta e^{-\lambda A_0(t|x) Z_t} - A_0^\gamma(s|x) Z_s^\beta e^{-\lambda A_0(s|x) Z_s}\right)^2 \middle| X = y\right].$$

Naturally, ρ_y defines a semimetric on $[0,1]^2$ and since it is bi-continuous, it makes $[0,1]$ totally bounded.

Now, according to Theorem 19.28 in van der Vaart (1998), the weak convergence of the stochastic process (18) follows from the four following conditions

$$\sup_{\rho_x(t,s) \leqslant \delta_n} P(f_{n,t} - f_{n,s})^2 \longrightarrow 0 \text{ for every } \delta_n \searrow 0, \tag{19}$$

$$PF_n^2 = O(1), \tag{20}$$

$$PF_n^2\{F_n > \varepsilon\sqrt{n}\} \longrightarrow 0 \text{ for every } \varepsilon > 0, \tag{21}$$

$$J(\delta_n, \mathcal{F}_n, L_2) \longrightarrow 0 \text{ for every } \delta_n \searrow 0. \tag{22}$$

We start to prove (19). By definition, we have

$$\begin{aligned} P(f_{n,t}-f_{n,s})^2 &= \int_{\mathbb{R}^p} h^{-p}K^2\left(\frac{x-u}{h}\right)\rho_u(t,s)f(u)du \\ &= \int_{\mathbb{S}^{p-1}} K^2(u)\rho_{x-hu}(t,s)f(x-hu)du \\ &= \|K\|_2^2 f(x)\rho_x(t,s) + \int_{\mathbb{S}^{p-1}} K^2(u)f(x-hu)[\rho_{x-hu}(t,s)-\rho_x(t,s)]du \\ &\qquad +\rho_x(t,s)\int_{\mathbb{S}^{p-1}} K^2(u)[f(x-hu)-f(x)]du. \end{aligned}$$

Due to Assumptions $(\mathcal{D})$, $(\mathcal{K}_1)$ and since ρ_x is bounded, it remains to show that

$$\sup_{\rho_x(t,s)\leqslant\delta_n} |\rho_{x-hu}(t,s)-\rho_x(t,s)| \to 0. \tag{23}$$

Recall that

$$\begin{aligned} \rho_y(t,s) &= [A_0(t|x)]^{2\gamma}\,\mathbb{E}\left[Z_t^{2\beta}e^{-2\lambda A_0(t|x)Z_t}\middle|X=y\right] + [A_0(s|x)]^{2\gamma}\,\mathbb{E}\left[Z_s^{2\beta}e^{-2\lambda A_0(s|x)Z_s}\middle|X=y\right] \\ &\quad -2\left[A_0(t|x)A_0(s|x)\right]^{\gamma}\mathbb{E}\left[Z_t^{\beta}e^{-\lambda A_0(t|x)Z_t}Z_s^{\beta}e^{-\lambda A_0(s|x)Z_s}\middle|X=y\right]. \end{aligned}$$

All these expectations have been already computed in (15) and (16). Using the mean value theorem combined with the boundedness of $A_0(.|.)$ and Assumption $(\mathcal{A}_0)$, we can easily infer that for all $(y,y')\in\mathbb{R}^p\times\mathbb{R}^p$, we have

$$\sup_{(t,s)\in[0,1]^2} \left|\rho_y(t,s)-\rho_{y'}(t,s)\right| \leqslant C\|y-y'\|^{\eta_{A_0}},$$

for some positive constant C. This implies (23) and thus (19) is established.

Now, we move to the proof of (20) and (21). Since the function $x\to x^\beta e^{-\lambda x}$ is bounded over $\mathbb{R}_+$ by $(\beta/\lambda)^\beta e^{-\beta}$ and $A_0(t|x)\in[1/2,1]$, $\mathcal{F}_n$ admits the natural envelope function

$$(y_1,y_2,z)\to F_n(y_1,y_2,z) := \sqrt{h^p}K_h(x-z)M, \tag{24}$$

where $M := \left(\frac{\beta}{\lambda}\right)^\beta e^{-\beta}\max(1,2^{\beta-\gamma})$. Consequently

$$\begin{aligned} &PF_n^2 = M^2\int_{\mathbb{R}^p} h^{-p}K^2\left(\frac{x-u}{h}\right)f(u)du = M^2\int_{\mathbb{S}^{p-1}} K^2(u)f(x-hu)du = M^2\|K\|_2^2 f(x)(1+o(1)), \\ &PF_n^2\{F_n>\varepsilon\sqrt{n}\} = M^2\int_{\{K(u)>M^{-1}\varepsilon\sqrt{nh^p}\}} K^2(u)f(x-hu)du = o(1), \end{aligned}$$

for all $\varepsilon>0$, since $nh^p\to\infty$, K satisfies Assumption $(\mathcal{K}_1)$ and f is continuous.

Finally, it remains to prove (22). First, we introduce the class of functions $\mathcal{W} := \{(y_1, y_2) \to A_0(t|x)Z_t(y_1, y_2), t \in [0,1]\}$ and its subgraph σ_t in $\mathbb{R}_+^2 \times \mathbb{R}$ as

$$\begin{aligned}\sigma_t &:= \{(u,v,w) : A_0(t|x)Z_t(u,v) > w\} \\ &= \left\{(u,v,w) : \frac{A_0(t|x)}{1-t}u > w\right\} \cap \left\{(u,v,w) : \frac{A_0(t|x)}{t}v > w\right\}.\end{aligned}$$

We can show that $\{\sigma_t : t \in [0,1]\}$ is a *VC*-class of sets. Indeed, if we look more generally, at the collection of sets $\mathcal{C} := \{\{(x,y) : \delta x > y\}, \delta > 0\}$ in $\mathbb{R}_+ \times \mathbb{R}$ and if we define two points (x_1, y_1) and (x_2, y_2) such that, without loss of generality, $\frac{y_1}{x_1} \leqslant \frac{y_2}{x_2}$. Then, for any $\delta > 0$, $\delta x_2 \geqslant y_2$ implies that $\delta x_1 \geqslant y_1$. Thus, $\mathcal{C}$ cannot shatter the set $\{(x_1, y_1), (x_2, y_2)\}$ and by consequence it is a *VC*-class of sets. Now, the collection of one set $\mathbb{R}_+$ is naturally a *VC*-class of sets. According to Lemma 2.6.17 (vii) in van der Vaart and Wellner (1996), $\mathcal{C} \times \mathbb{R}_+$ is a *VC*-class of sets as well. Invoking Lemma 2.6.17 (ii), $\{\sigma_t : t \in [0,1]\}$ belongs to a *VC*-class and as such is *VC*. Define now for all $z \in \mathbb{R}_+$

$$\phi_{\lambda,\beta}(z) := z^\beta e^{-\lambda z}.$$

We can easily check that $\phi_{\lambda,\beta}$ is of bounded variation. This implies that $\phi_{\lambda,\beta}$ can be decomposed as the sum of two monotone functions, say $\phi_{\lambda,\beta}^{(1)}$ and $\phi_{\lambda,\beta}^{(2)}$. Thus, according to Lemma 2.6.18 (viii) in van der Vaart and Wellner (1996), $\phi_{\lambda,\beta}^{(1)} \circ \mathcal{W}$ and $\phi_{\lambda,\beta}^{(2)} \circ \mathcal{W}$ are *VC*. Now, according to Theorem 2.6.7 in van der Vaart and Wellner (1996), there exists a universal constant C such that for any $j = 1, 2$ and $0 < \tau < 1$

$$\sup_{\mathcal{Q}} N(\phi_{\lambda,\beta}^{(j)} \circ \mathcal{W}, L_2(Q), \tau \|\mathcal{W}_j\|_{Q,2}) \leqslant CV_j(16e)^{V_j}\left(\frac{1}{\tau}\right)^{2(V_j-1)},$$

where V_j is the *VC*-index of $\phi_{\lambda,\beta}^{(j)} \circ \mathcal{W}$ and $\mathcal{W}_j$ its envelope function. Now, consider the sequence of class of functions

$$\mathcal{F}_{n,j} := \{z \to \sqrt{h^p}K_h(x-z)\} \otimes \phi_{\lambda,\beta}^{(j)} \circ \mathcal{W},$$

for $j = 1, 2$. Since we only update the previous sets with one single function and only one ball is needed to recover the class $\{z \to \sqrt{h^p}K_h(x-z)\}$ whatever the measure Q, we have

$$\sup_{\mathcal{Q}} N(\mathcal{F}_{n,j}, L_2(Q), \tau \|\kappa F_n\|_{Q,2}) \leqslant CV_j(16e)^{V_j}\left(\frac{1}{\tau}\right)^{2(V_j-1)},$$

where κ is a suitable constant. Moreover, the class of (constant) functions $\{[A_0(t|x)]^{\gamma-\beta}, t \in [0,1]\}$ is naturally a *VC*-class and $\sup_{t\in[0,1]}[A_0(t|x)]^{\gamma-\beta} = \max(1, 2^{\beta-\gamma})$. Thus for any $0 < \tau < 1$ we can divide the interval $\left[0, \max(1, 2^{\beta-\gamma})\right]$ in at least $\lceil 1/2\tau \rceil$ balls of radius $\tau \max(1, 2^{\beta-\gamma})$. Hence

$$\sup_{\mathcal{Q}} N\left(\{[A_0(t|x)]^{\gamma-\beta}, t \in [0,1]\}, L_2(Q), \tau \max(1, 2^{\beta-\gamma})\right) \leqslant \left\lceil \frac{1}{2\tau} \right\rceil \leqslant \frac{3}{2}\left(\frac{1}{\tau}\right)^2.$$

Consequently, we have

$$\sup_{\mathcal{Q}} N(\{[A_0(t|x)]^{\gamma-\beta}, t \in [0,1]\} \otimes \mathcal{F}_{n,j}, L_2(Q), \tau \max(1, 2^{\beta-\gamma})\|\kappa F_n\|_{Q,2}) \leqslant \frac{3C}{2} V_j (16e)^{V_j} \left(\frac{1}{\tau}\right)^{2V_j}.$$

Finally, since our class of interest $\mathcal{F}_n$ is included in the class of functions

$$\widetilde{\mathcal{F}}_n := \{[A_0(t|x)]^{\gamma-\beta}, t \in [0,1]\} \otimes \mathcal{F}_{n,1} + \{[A_0(t|x)]^{\gamma-\beta}, t \in [0,1]\} \otimes \mathcal{F}_{n,2},$$

with envelope function $2\max(1, 2^{\beta-\gamma})\kappa F_n$, using Lemma 16 in Nolan and Pollard (1987), we have

$$\begin{aligned}
\sup_{\mathcal{Q}} N(\mathcal{F}_n, L_2(Q), 2\tau \max(1, 2^{\beta-\gamma})\|\kappa F_n\|_{Q,2}) &\leqslant \sup_{\mathcal{Q}} N(\widetilde{\mathcal{F}}_n, L_2(Q), 2\tau \max(1, 2^{\beta-\gamma})\|\kappa F_n\|_{Q,2}) \\
&\leqslant \frac{9C^2}{4} V_1 V_2 (16e)^{V_1+V_2} \left(\frac{4}{\tau}\right)^{2(V_1+V_2)} \\
&=: L \left(\frac{1}{\tau}\right)^{V}.
\end{aligned}$$

Thus, (22) is established since for any sequence $\delta_n \searrow 0$ and n large enough, we have

$$J(\delta_n, \mathcal{F}_n, L_2) \leqslant \int_0^{\delta_n} \sqrt{\log([2\kappa \max(1, 2^{\beta-\gamma})]^V L) - V \log(\tau)} d\tau = o(1).$$

This achieves the proof of Theorem 2.1 since the covariance structure follows from (7).

6.3 Proof of Theorem 2.2

To prove this theorem, we will make use of the Cramér-Wold device (see, e.g., Severini, 2005, p. 337), according to which it is sufficient to show that

$$\Lambda_n := \xi^T \sqrt{nh^p} \left(\mathbb{T}_n - \mathbb{E}[\mathbb{T}_n]\right) \rightsquigarrow \mathcal{N}_1\left(0, \xi^T \Sigma \xi\right),$$

for all $\xi \in \mathbb{R}^m$. A straightforward rearrangement of the terms leads to

$$\begin{aligned}
\Lambda_n &= \frac{1}{n} \sum_{i=1}^{n} \sqrt{nh^p} \left\{ \sum_{j=1}^{m} \xi_j a_j^{\gamma_j} K_{j,h}(x_j - X_i) Z_{t_j,i}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,i}} - \mathbb{E}\left[\sum_{j=1}^{m} \xi_j a_j^{\gamma_j} K_{j,h}(x_j - X_i) Z_{t_j,i}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,i}} \right] \right\} \\
&=: \frac{1}{n} \sum_{i=1}^{n} W_i.
\end{aligned}$$

Since $W_1, ..., W_n$ are independent and identically distributed random variables, $\mathrm{Var}(\Lambda_n) = \dfrac{\mathrm{Var}(W_1)}{n}$ with

$$\mathrm{Var}(W_1) = nh^p \sum_{j=1}^{m} \sum_{k=1}^{m} \xi_j \xi_k \mathbb{C}_{j,k},$$

where

$$\begin{aligned}\mathbb{C}_{j,k} &:= \mathbb{E}\left[a_j^{\gamma_j} K_{j,h}(x_j - X_1) Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}} a_k^{\gamma_k} K_{k,h}(x_k - X_1) Z_{t_k,1}^{\beta_k} e^{-\lambda_k a_k Z_{t_k,1}}\right] \\ &\quad -\mathbb{E}\left[a_j^{\gamma_j} K_{j,h}(x_j - X_1) Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}}\right] \mathbb{E}\left[a_k^{\gamma_k} K_{k,h}(x_k - X_1) Z_{t_k,1}^{\beta_k} e^{-\lambda_k a_k Z_{t_k,1}}\right].\end{aligned}$$

According to Appendix 5.1, $Var(\Lambda_n) = \xi^T \Sigma \xi (1 + o(1))$. Hence, to ensure the convergence in distribution of Λ_n to a normal random variable, we have to verify the Lyapounov condition for triangular arrays of random variables (Billingsley, 1995, p. 362). In the present context this simplifies to verifying $\frac{1}{n^2}\mathbb{E}(|W_1|^3) \to 0$. We have

$$\begin{aligned}\mathbb{E}(|W_1|^3) &\leqslant n^{3/2} h^{3p/2} \left\{ \mathbb{E}\left[\left(\sum_{j=1}^m |\xi_j| a_j^{\gamma_j} K_{j,h}(x_j - X_1) Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}}\right)^3\right] \right. \\ &\quad +3\mathbb{E}\left[\left(\sum_{j=1}^m |\xi_j| a_j^{\gamma_j} K_{j,h}(x_j - X_1) Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}}\right)^2\right] \\ &\quad \times \mathbb{E}\left[\sum_{j=1}^m |\xi_j| a_j^{\gamma_j} K_{j,h}(x_j - X_1) Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}}\right] \\ &\quad \left. + 4\left[\mathbb{E}\left(\sum_{j=1}^m |\xi_j| a_j^{\gamma_j} K_{j,h}(x_j - X_1) Z_{t_j,1}^{\beta_j} e^{-\lambda_j a_j Z_{t_j,1}}\right)\right]^3 \right\}.\end{aligned}$$

A similar treatment as in Appendix 5.1 yields for all $q \in \mathbb{N}^*$

$$\mathbb{E}\left[\left(\sum_{j=1}^m |\xi_j| a_j^{\gamma_j} K_{j,h}(x_j - X_1) Z_{t_j,1}^{\beta_j} e^{-\lambda_j a Z_{t_j,1}}\right)^q\right] = \mathcal{O}((h^p)^{-q+1}),$$

and hence $\frac{1}{n^2}\mathbb{E}(|W_1|^3) = \mathcal{O}\left((\sqrt{nh^p})^{-1}\right) = o(1)$.

6.4 Proof of Theorem 2.3

To prove the theorem we will adjust the arguments used to prove existence and consistency of solutions of the likelihood estimating equation, see e.g. Theorem 3.7 and Theorem 5.1 in Chapter 6 of Lehmann and Casella (1998), to the MDPD framework. Let $\zeta, b > 0$, $C(.|.) : [0,1] \times \mathbb{R}^p \to [1/2 - \zeta, 1 + \zeta]$ and $\forall t \in [0,1]$, $r(t) := |A_0(t|x) - C(t|x)|$. Define in addition the b-level of r as

$$T_b := \{t \in [0,1],\ \ r(t) > b\}.$$

We firstly show that for any $b > 0$

$$\mathbb{P}\left(\forall t \in T_b,\ \ \widehat{\Delta}_{\alpha,x,t}(A_0(t|x)) < \widehat{\Delta}_{\alpha,x,t}\left(C(t|x)\right)\right) \to 1, \tag{25}$$

as $n \to \infty$, for any function $C(.|x)$ different from but close enough to $A_0(.|x)$. By applying a Taylor series expansion, we have

$$\begin{aligned}\widehat{\Delta}_{\alpha,x,t}(C(t|x)) - \widehat{\Delta}_{\alpha,x,t}(A_0(t|x)) &= (C(t|x) - A_0(t|x))\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)) + \frac{1}{2}(C(t|x) - A_0(t|x))^2\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) \\ &\quad + \frac{1}{6}(C(t|x) - A_0(t|x))^3\widehat{\Delta}^{(3)}_{\alpha,x,t}(\widetilde{C}(t|x)),\end{aligned}$$

where $\widetilde{C}(t|x)$ is an intermediate value between $C(t|x)$ and $A_0(t|x)$. According to Appendix 5.2, as $n \to \infty$

$$\sup_{t\in[0,1]} \left|\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x))\right| = \sup_{t\in[0,1]} \left|\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)) - \ell^{(1)}_{\alpha,x,t}(A_0(t|x))\right| \xrightarrow{\mathbb{P}} 0.$$

This convergence implies, that for all $0 < \varepsilon \leqslant b^2$

$$\begin{aligned}\mathbb{P}\left(\forall t \in T_b,\ r(t)|\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x))| \leqslant r^3(t)\right) &\geqslant \mathbb{P}\left(\forall t \in T_b,\ |\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x))| \leqslant r^2(t), \sup_{t\in[0,1]} \left|\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)\right| \leqslant \varepsilon\right) \\ &= \mathbb{P}\left(\sup_{t\in[0,1]} \left|\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)\right| \leqslant \varepsilon\right) \longrightarrow 1,\end{aligned}$$

as $n \to \infty$. Now, concerning $\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x))$, we have

$$\sup_{t\in[0,1]} \left|\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) - f(x)[A_0(t|x)]^{\alpha-2}\frac{1+\alpha^2}{(1+\alpha)^2}\right| \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$. Consequently, there exists $\delta_1 > 0$ such that

$$\forall t \in [0,1],\ \frac{r^2(t)}{2}\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) > \delta_1 r^2(t),$$

with probability tending to 1.
Finally, since $x \to x^\lambda e^{-x}$ is bounded $\forall \lambda \geqslant 1$ on $\mathbb{R}^+$ and by Lemma 2.1

$$T_n(K, a, t, 0, 0, 0|x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \xrightarrow{\mathbb{P}} f(x),$$

as $n \to \infty$, we have for any $\varepsilon > 0$, $n^{-1}\sum_{i=1}^n K_h(x - X_i) \leqslant f(x) + \varepsilon$ with probability tending to 1. This implies that

$$\sup_{a\in[1/2,1],t\in[0,1]} \left|\widehat{\Delta}^{(3)}_{\alpha,x,t}(a)\right| =: M < \infty \tag{26}$$

with probability tending to 1. We can therefore conclude that

$$\forall t \in [0,1],\ \frac{r^3(t)}{6}\left|\widehat{\Delta}^{(3)}_{\alpha,x,t}(\widetilde{C}(t|x))\right| \leqslant \frac{M}{6}r^3(t),$$

with probability tending to 1.
Overall, we have shown that

$$\mathbb{P}\left(\forall t \in T_b,\ \ \widehat{\Delta}_{\alpha,x,t}(C(t|x)) - \widehat{\Delta}_{\alpha,x,t}(A_0(t|x)) > \delta_1 r^2(t) - \left(1 + \frac{M}{6}\right) r^3(t)\right) \longrightarrow 1,$$

as $n \to \infty$, where the right-hand side of the inequality is positive for $r(t) < \delta_1/(1 + M/6)$. Thus, setting

$$\sup_{t\in[0,1]} r(t) < \delta_1/(1 + M/6),$$

(25) follows.

To complete the proof we adjust the line of argumentation of Theorem 3.7 in Chapter 6 of Lehmann and Casella (1998). Let δ be a positive function such that $\forall t \in [0,1]$, $A_0(t|x) \pm \delta(t) \in [1/2 - \zeta, 1 + \zeta]$. Define $\forall b > 0$, the event

$$S_n(b,\delta) := \left\{\upsilon \in \Omega : \forall t \in T_b,\ \ \widehat{\Delta}_{\alpha,x,t}(A_0(t|x)) < \widehat{\Delta}_{\alpha,x,t}(A_0(t|x) \pm \delta(t))\right\},$$

where we define as previously $T_b := \{t \in [0,1],\ \ \delta(t) > b\}$. Since $A_0(t|x) \in [1/2, 1]$ and $\zeta > 0$, we can always find a pair (b,δ) such that

$$0 < b \leqslant \inf_{t\in[0,1]} \delta(t).$$

Thus, assume that $b := \inf_{t\in[0,1]} \delta(t)$. For all $t \in T_b = [0,1]$ and $\upsilon \in S_n(b,\delta)$, since $\widehat{\Delta}_{\alpha,x,t}(a)$ is differentiable with respect to a, there exists $\widetilde{A}_{\alpha,n,\delta(t)}(t|x) \in (A_0(t|x) - \delta(t), A_0(t|x) + \delta(t))$ where $\widehat{\Delta}_{x,\alpha,t}(a)$ achieves a local minimum, so $\widehat{\Delta}^{(1)}_{\alpha,x,t}(\widetilde{A}_{\alpha,n,\delta(t)}(t|x)) = 0$.

By (25), $\mathbb{P}(S_n(b,\delta)) \to 1$ for any $b > 0$ and $\|\delta\|_\infty$ small enough, and hence there exists a sequence δ_n with $\|\delta_n\|_\infty \searrow 0$ and $b_n := \inf_{t\in[0,1]} \delta_n(t) > 0$, such that $\mathbb{P}(S_n(b_n,\delta_n)) \to 1$, as $n \to \infty$. Now, let $\widehat{A}_{\alpha,n}(t|x) := \widetilde{A}_{\alpha,n,\delta_n(t)}(t|x)$ if $\upsilon \in S_n(b_n,\delta_n)$ and arbitrary otherwise. Since $\upsilon \in S_n(b_n,\delta_n)$ implies $\widehat{\Delta}^{(1)}_{\alpha,x,t}(\widehat{A}_{\alpha,n}(t|x)) = 0$, we have that

$$\mathbb{P}\left(\widehat{\Delta}^{(1)}_{\alpha,x,t}(\widehat{A}_{\alpha,n}(t|x)) = 0\right) \geqslant \mathbb{P}\left(S_n(b_n,\delta_n)\right) \to 1,$$

as $n \to \infty$, which establishes the existence part.

Concerning the uniform consistency of the solution sequence, note that for any $\varepsilon > 0$ and n large enough such that $\|\delta_n\|_\infty \leqslant \varepsilon$, we have

$$\begin{aligned}\mathbb{P}\left(\sup_{t\in[0,1]} \left|\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right| \leqslant \varepsilon\right) &\geqslant \mathbb{P}\left(\sup_{t\in[0,1]} \left|\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right| \leqslant \|\delta_n\|_\infty\right) \\ &= \mathbb{P}\left(\sup_{t\in T_{b_n}} \left|\widehat{A}_{\alpha,n}(t|x) - A_0(t|x)\right| \leqslant \|\delta_n\|_\infty\right) \geqslant \mathbb{P}\left(S_n(b_n,\delta_n)\right) \to 1,\end{aligned}$$

as $n \to \infty$, whence the uniform consistency of the estimator sequence.

6.5 Proof of Theorem 2.4

The starting point is (5). According to Corollary 5.1, $\left\{\sqrt{nh^p}\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)), t\in[0,1]\right\}$ weakly converges, as $n\to\infty$, towards a tight centered Gaussian process and

$$\left\{\widehat{\Delta}^{(2)}_{\alpha,x,t}\left(A_0(t|x)\right), t\in[0,1]\right\} \xrightarrow{\mathbb{P}} \left\{\ell^{(2)}_{\alpha,x,t}(A_0(t|x)) = f(x)[A_0(t|x)]^{\alpha-2}\frac{1+\alpha^2}{(1+\alpha)^2}, t\in[0,1]\right\}.$$

Combining these results with (26), we have, as $n\to\infty$,

$$\left\{\left[\widehat{\Delta}^{(2)}_{\alpha,x,t}(A_0(t|x)) + \frac{1}{2}\widehat{\Delta}^{(3)}_{\alpha,x,t}(\widetilde{A}(t|x))(\widehat{A}_{\alpha,n}(t|x) - A_0(t|x))\right]^{-1}, t\in[0,1]\right\}$$
$$\xrightarrow{\mathbb{P}} \left\{[f(x)]^{-1}[A_0(t|x)]^{2-\alpha}\frac{(1+\alpha)^2}{1+\alpha^2}, t\in[0,1]\right\}.$$

Concerning the covariance structure, it follows from Theorem 2.2 and the fact that

$$\widehat{\Delta}^{(1)}_{\alpha,x,t}(A_0(t|x)) = v_\alpha^T T_n^{(3)}(t|x),$$

where

$$T_n^{(3)}(t|x) := \begin{pmatrix} T_n(K, A_0(t|x), t, 0, 0, \alpha-1|x) \\ T_n(K, A_0(t|x), t, \alpha, 0, \alpha-1|x) \\ T_n(K, A_0(t|x), t, \alpha, 1, \alpha|x)\end{pmatrix}.$$

6.6 Proof of Lemma 3.1

We use the following decomposition

$$\begin{aligned} F_{n,j}(y|x) - F_j(y|x) &= [f(x)]^{-1}\left[n^{-1}\sum_{i=1}^n K_c(x-X_i)\mathbb{1}_{\{Y_i^{(j)}\leqslant y\}} - \mathbb{E}\left(K_c(x-X_1)\mathbb{1}_{\{Y_1^{(j)}\leqslant y\}}\right)\right] \\ &\quad + [f(x)]^{-1}\left[\mathbb{E}\left(K_c(x-X_1)\mathbb{1}_{\{Y_1^{(j)}\leqslant y\}}\right) - f(x)F_j(y|x)\right] \\ &\quad - \frac{n^{-1}\sum_{i=1}^n K_c(x-X_i)\mathbb{1}_{\{Y_i^{(j)}\leqslant y\}}}{n^{-1}\sum_{i=1}^n K_c(x-X_i)}\frac{1}{f(x)}\left[n^{-1}\sum_{i=1}^n K_c(x-X_i) - \mathbb{E}\left(K_c(x-X_1)\right)\right] \\ &\quad - \frac{n^{-1}\sum_{i=1}^n K_c(x-X_i)\mathbb{1}_{\{Y_i^{(j)}\leqslant y\}}}{n^{-1}\sum_{i=1}^n K_c(x-X_i)}\frac{1}{f(x)}\left[\mathbb{E}\left(K_c(x-X_1)\right) - f(x)\right] \\ &=: \sum_{k=1}^4 T_k(y|x). \end{aligned}$$

An intermediate result to study $T_1(y|x)$ and $T_3(y|x)$ consist in showing that, for $q>1$,

$$\mathbb{E}\left[\sup_{(y,x)\in\mathbb{R}\times\mathbb{R}^p}\left|n^{-1}\sum_{i=1}^n K_c(x-X_i)\mathbb{1}_{\{Y_i^{(j)}\leqslant y\}} - \mathbb{E}\left[K_c(x-X_1)\mathbb{1}_{\{Y_1^{(j)}\leqslant y\}}\right]\right|\right] = o\left(\sqrt{\frac{|\log c|^q}{nc^p}}\right) \quad (27)$$

$$\mathbb{E}\left[\sup_{x\in\mathbb{R}^p}\left|n^{-1}\sum_{i=1}^n K_c(x-X_i) - \mathbb{E}\left[K_c(x-X_1)\right]\right|\right] = o\left(\sqrt{\frac{|\log c|^q}{nc^p}}\right) \quad (28)$$

To this aim, let us introduce the class

$$\begin{aligned}\mathcal{G} &:= \left\{(u,v) \to K\left(\frac{x-v}{d}\right)\mathbb{1}_{\{u\leqslant y\}};\, y\in\mathbb{R},\, x\in\mathbb{R}^p, d>0\right\}\\ &= \left\{K\left(\frac{x-\cdot}{d}\right);\, x\in\mathbb{R}^p, d>0\right\}\otimes\left\{\mathbb{1}_{\{\cdot\leqslant y\}};\, y\in\mathbb{R}\right\}\\ &=: \mathcal{G}_1\otimes\mathcal{G}_2.\end{aligned}$$

Under Assumption $(\mathcal{K}_2)$, $\mathcal{G}_1$ is a uniformly bounded VC-class of measurable functions (see e.g. Giné and Guillou, 2002). Next, since the collection of all cells $\{(-\infty,a], a\in\mathbb{R}\}$ is a VC-class of sets, it follows that $\mathcal{G}_2$ is also a uniformly bounded VC-class of measurable functions. Now, using the fact that the covering number of the direct product of two VC-classes is bounded by the product of the respective covering numbers,

$$\mathcal{G}_n := \left\{(u,v) \to K\left(\frac{x-v}{c}\right)\mathbb{1}_{\{u\leqslant y\}};\, y\in\mathbb{R},\, x\in\mathbb{R}^p, c=c_n>0\right\},$$

admits the same bound for the covering number as $\mathcal{G}$, that is

$$N(\mathcal{G}_n, L_2(Q), \tau\|K\|_\infty) \leqslant C\, V_{\mathcal{G}}(16e)^{V_{\mathcal{G}}}\left(\frac{1}{\tau}\right)^{2(V_{\mathcal{G}}-1)} =: \left(\frac{A_{\mathcal{G}}}{\tau}\right)^{\nu_{\mathcal{G}}},$$

where C is a universal constant, $\tau\in(0,1)$ and $V_{\mathcal{G}}$ is the VC-index of $\mathcal{G}$ (see Theorem 2.6.7 in van der Vaart and Wellner, 1996). Now, according to Proposition 2.1 in Giné and Guillou (2001) (see also Theorem 2.1 in Giné and Guillou, 2002) for $\sigma^2 \geqslant \sup_{g\in\mathcal{G}_n}\mathrm{Var}(g)$, $U\geqslant\|K\|_\infty$ and $0<\sigma\leqslant U$, there exists a universal constant B such that

$$\begin{aligned}&\mathbb{E}\left[\sup_{(y,x)\in\mathbb{R}\times\mathbb{R}^p}\left|n^{-1}\sum_{i=1}^n K_c(x-X_i)\mathbb{1}_{\{Y_i^{(j)}\leqslant y\}} - \mathbb{E}\left[K_c(x-X_1)\mathbb{1}_{\{Y_1^{(j)}\leqslant y\}}\right]\right|\right]\\ &\qquad\leqslant [nc^p]^{-1}B\left[U\nu_{\mathcal{G}}\log\left(\frac{UA_{\mathcal{G}}}{\sigma}\right)+\sqrt{\nu_{\mathcal{G}}n\sigma^2\log\left(\frac{UA_{\mathcal{G}}}{\sigma}\right)}\right].\end{aligned}$$

Since

$$\mathrm{Var}\left(K\left(\frac{x-X_1}{c}\right)\mathbb{1}_{\{Y_1^{(j)}\leqslant y\}}\right)\leqslant c^p\int_{\mathbb{S}^{p-1}}K^2(u)\,f(x-cu)du\leqslant c^p\|f\|_\infty\|K\|_2^2,$$

the choices $\sigma^2=\sigma_n^2:=c^p\|f\|_\infty\|K\|_2^2$ and $U=\|K\|_\infty$ imply that $\sigma_n^2\leqslant U^2$ for n large enough. This yields (27). Similar arguments can be used in order to show (28). Since $f(x)\geqslant b, \forall x\in S_X$

$$\mathbb{E}\left[\sup_{(y,x)\in\mathbb{R}\times S_X}|T_k(y|x)|\right]=o\left(\sqrt{\frac{|\log c|^q}{nc^p}}\right),\ \text{for } k=1,3.$$

Concerning $T_2(y|x)$, remark that

$$\begin{aligned}\mathbb{E}\left[K_c(x-X_1)\mathbb{1}_{\{Y_1^{(j)}\leqslant y\}}\right]-f(x)F_j(y|x) &= \mathbb{E}\left[K_c(x-X_1)F_j(y|X_1)\right]-f(x)F_j(y|x)\\ &= \int_{\mathbb{R}^p}K_c(x-u)\left[F_j(y|u)-F_j(y|x)\right]f(u)du\\ &\quad +F_j(y|x)\int_{\mathbb{R}^p}K_c(x-u)[f(u)-f(x)]du\\ &= o(c^\eta),\end{aligned}$$

by Assumptions $(\mathcal{D})$ and $(\mathcal{F})$. The same bound can also be obtained for $T_4(y|x)$. This achieves the proof of Lemma 3.1.

6.7 Proof of Theorem 3.1

Let

$$\mathcal{I}_n := \{g_{\theta,\delta,n} : \theta\in\Theta, \delta\in H\}$$

where for $\theta\in\Theta := [0,1]\times[1/2,1]$, and $\delta\in H := \{\delta=(\delta_1,\delta_2);\ \delta:\mathbb{R}\times\mathbb{R}\times S_X\to\mathbb{R}\}$,

$$\begin{aligned}g_{\theta,\delta,n}(y_1,y_2,u) &:= \sqrt{h^p}K_h(x-u)q_{\theta,\delta}(y_1,y_2,u)\\ &:= \sqrt{h^p}K_h(x-u)a^\gamma[Z_{\theta,\delta}(y_1,y_2,u)]^\beta\exp\left(-\lambda a Z_{\theta,\delta}(y_1,y_2,u)\right)\end{aligned}$$

with

$$Z_{\theta,\delta}(y_1,y_2,u) := \min\left(\frac{-\log\left(|\delta_1(y_1,y_2,u)|\right)}{1-t},\frac{-\log\left(|\delta_2(y_1,y_2,u)|\right)}{t}\right).$$

For convenience, denote $\delta_n := (F_{n,1},F_{n,2})$ and $\delta_0 := (F_1,F_2)$. According to Lemma 3.1, $r_n^{-1}|\delta_n-\delta_0|$ converges in probability towards the null function $H_0 := \{0\}$ in H, endowed with the norm $\|\delta\|_H := \|\delta_1\|_\infty+\|\delta_2\|_\infty$ for any $\delta\in H$. In order to apply Theorem 2.3 in van der Vaart and Wellner (2007), we have now to show
Assertion 1: $\sup_{\theta\in\Theta}\sqrt{n}PG_n(\theta,b_n)\longrightarrow 0$ for every $b_n\to 0$
and
Assertion 2: $\sup_{\theta\in\Theta}|\mathbb{G}_nG_n(\theta,b)|\xrightarrow{\mathbb{P}}0$, for every $b>0$,
where $G_n(\theta,b)$ is the minimal envelope function for the class

$$\mathcal{E}_n(\theta,b) := \{g_{\theta,\delta_0+r_n\delta,n}-g_{\theta,\delta_0,n} : \delta\in H, \|\delta\|_H\leqslant b\},$$

i.e.

$$\begin{aligned}G_n(\theta,b) &:= \sup_{\|\delta\|_H\leqslant b}|g_{\theta,\delta_0+r_n\delta,n}-g_{\theta,\delta_0,n}|\\ &= \sqrt{h^p}K_h(x-\cdot)\sup_{\|\delta\|_H\leqslant b}|q_{\theta,\delta_0+r_n\delta}-q_{\theta,\delta_0}|.\end{aligned}\tag{29}$$

Now, remark that $\forall(y_1,y_2,u)\in\mathbb{R}\times\mathbb{R}\times S_X$

$$\sup_{\|\delta\|_H\leqslant b}|q_{\theta,\delta_0+r_n\delta}-q_{\theta,\delta_0}|(y_1,y_2,u) = \sup_{(\delta_1(y_1,y_2,u),\delta_2(y_1,y_2,u))\in B}|q_{\theta,\delta_0+r_n\delta}-q_{\theta,\delta_0}|(y_1,y_2,u),$$

where $B := \{(x,y) \in \mathbb{R}^2 : |x| + |y| \leqslant b\}$. Since B is compact and $\delta \to q_{\theta,\delta}(y_1, y_2, u)$ is continuous, (29) reaches its supremum on at least one position $\delta^*_{\theta,b}(y_1,y_2,x) = (\delta^*_{1,\theta,b}(y_1,y_2,x), \delta^*_{2,\theta,b}(y_1,y_2,x))$ in B. Thus, according to Theorem 18.19 in Aliprantis and Border (2006), one can find a measurable function $\delta^*_{\theta,b}$ bounded by b in H such that

$$G_n(\theta, b) = |g_{\theta,\delta_0 + r_n\delta^*_{\theta,b},n} - g_{\theta,\delta_0,n}|.$$

Proof of Assertion 1. For any positive sequence $b_n \to 0$, we have

$$\sqrt{n}PG_n(\theta, b_n) = \sqrt{nh^p} \int_{\mathbb{S}^{p-1}} K(u)\mathbb{E}\left[|q_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}} - q_{\theta,\delta_0}| \Big| X = x - hu \right] f(x - hu)du.$$

Note that for any $(\delta, \delta') \in H \times H$, using (14)

$$|q_{\theta,\delta} - q_{\theta,\delta'}| \quad \leqslant \quad a^\gamma \int_0^{+\infty} |\beta - \lambda as| s^{\beta-1} e^{-\lambda as} \mathbb{1}_{\{s \in [\min(Z_{\theta,\delta}, Z_{\theta,\delta'}), \max(Z_{\theta,\delta}, Z_{\theta,\delta'})]\}} ds. \tag{30}$$

Consequently

$$\begin{aligned}
&\mathbb{E}\left[|q_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}} - q_{\theta,\delta_0}| \Big| X = x - hu \right] \\
&\leqslant a^\gamma \int_0^{+\infty} |\beta - \lambda as| s^{\beta-1} e^{-\lambda as}\, \mathbb{P}\left(s \in [\min(Z_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0}), \max(Z_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0})] \Big| X = x - hu \right) ds.
\end{aligned}$$

Remark now that

$$\begin{aligned}
&\left\{ s \in [\min(Z_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0}), \max(Z_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0})] \right\} \\
= &\left\{ e^{-s} \in \left[\min\left(\max\left(\left|F_1 + r_n\delta^*_{1,\theta,b_n}\right|^{\frac{1}{1-t}}, \left|F_2 + r_n\delta^*_{2,\theta,b_n}\right|^{\frac{1}{t}} \right), \max\left(F_1^{\frac{1}{1-t}}, F_2^{\frac{1}{t}} \right) \right), \right.\right. \\
&\qquad \left.\left. \max\left(\max\left(\left|F_1 + r_n\delta^*_{1,\theta,b_n}\right|^{\frac{1}{1-t}}, \left|F_2 + r_n\delta^*_{2,\theta,b_n}\right|^{\frac{1}{t}} \right), \max\left(F_1^{\frac{1}{1-t}}, F_2^{\frac{1}{t}} \right) \right) \right] \right\} \\
\subset &\left\{ e^{-s} \in \left[\min\left(\left|F_1 + r_n\delta^*_{1,\theta,b_n}\right|^{\frac{1}{1-t}}, F_1^{\frac{1}{1-t}} \right), \max\left(\left|F_1 + r_n\delta^*_{1,\theta,b_n}\right|^{\frac{1}{1-t}}, F_1^{\frac{1}{1-t}} \right) \right] \right\} \\
&\cup \left\{ e^{-s} \in \left[\min\left(\left|F_2 + r_n\delta^*_{2,\theta,b_n}\right|^{\frac{1}{t}}, F_2^{\frac{1}{t}} \right), \max\left(\left|F_2 + r_n\delta^*_{2,\theta,b_n}\right|^{\frac{1}{t}}, F_2^{\frac{1}{t}} \right) \right] \right\} \\
\subset &\left\{ e^{-(1-t)s} \in \left[\min(\left|F_1 + r_n\delta^*_{1,\theta,b_n}\right|, F_1), \max(\left|F_1 + r_n\delta^*_{1,\theta,b_n}\right|, F_1) \right] \right\} \\
&\cup \left\{ e^{-ts} \in \left[\min(\left|F_2 + r_n\delta^*_{2,\theta,b_n}\right|, F_2), \max(\left|F_2 + r_n\delta^*_{2,\theta,b_n}\right|, F_2) \right] \right\} \\
\subset &\left\{ e^{-(1-t)s} \in [F_1 - r_nb_n, F_1 + r_nb_n] \right\} \cup \left\{ e^{-ts} \in [F_2 - r_nb_n, F_2 + r_nb_n] \right\} \\
=: &\; A_{n,1}(s) \cup A_{n,2}(s).
\end{aligned}$$

Since for any subsets A and B we have $\mathbb{1}_{\{A \cup B\}} \leqslant \mathbb{1}_{\{A\}} + \mathbb{1}_{\{B\}}$, we can deduce that

$$\begin{aligned}
&\mathbb{P}\left(s \in [\min(Z_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0}), \max(Z_{\theta,\delta_0 + r_n\delta^*_{\theta,b_n}}, Z_{\theta,\delta_0})] | X = x - hu \right) \\
&\qquad \leqslant \mathbb{P}\left(A_{n,1}(s) | X = x - hu \right) + \mathbb{P}\left(A_{n,2}(s) | X = x - hu \right) \\
&\qquad = \int_0^1 \mathbb{1}_{\{e^{-(1-t)s} \in [v - r_nb_n, v + r_nb_n]\}} dv + \int_0^1 \mathbb{1}_{\{e^{-ts} \in [v - r_nb_n, v + r_nb_n]\}} dv \\
&\qquad \leqslant 2r_nb_n + 2r_nb_n = 4r_nb_n.
\end{aligned} \tag{31}$$

This implies that

$$\sqrt{nh^p}\mathbb{E}\left[|q_{\theta,\delta_0+r_n\delta^*_{\theta,b_n}} - q_{\theta,\delta_0}|\Big|X = x - hu\right] \leqslant 4\sqrt{nh^p}r_n b_n \sup_{a\in[1/2,1]}\int_0^\infty a^\gamma|\beta - \lambda as|s^{\beta-1}e^{-\lambda as}ds.$$

This achieves the proof of Assertion 1 since K is bounded, $\sup_{a\in[1/2,1]}\int_0^\infty a^\gamma|\beta-\lambda as|s^{\beta-1}e^{-\lambda as}ds < +\infty$, $\sqrt{nh^p}r_n \to 0$ and $b_n \to 0$.

Proof of Assertion 2. The idea is to apply Lemma 2.2 in van der Vaart and Wellner (2007). To this aim, first observe that the class $\mathcal{E}_n(\theta, b)$ admits an envelope function E_n of the same form as F_n in (24), for some suitable constant $M > 0$. Thus E_n satisfies the conditions (20) and (21), with F_n replaced by E_n. Consequently, it remains to show the two following convergences

$$\sup_{\theta\in\Theta} PG_n^2(\theta, b) \longrightarrow 0, \tag{32}$$

$$J(d_n, \{G_n(\theta, b) : \theta \in \Theta\}, L_2) \longrightarrow 0 \text{ for all } d_n \searrow 0. \tag{33}$$

We start to show (32). Since

$$PG_n^2(\theta, b) = \int_{\mathbb{S}^{p-1}} K^2(u)\mathbb{E}\left(|q_{\theta,\delta_0+r_n\delta^*_{\theta,b}} - q_{\theta,\delta_0}|^2\Big|X = x - hu\right) f(x - hu)du,$$

and (30), (32) follows from the proof of Assertion 1.

Now, to deal with the uniform entropy integral, we can adjust the lines of proof of Theorem 2.1 by considering the classes of functions defined on $\mathbb{R}\times\mathbb{R}\times S_X$

$$\phi^{(j)}_{\lambda,\beta}\circ\mathcal{W}\circ\Psi,\ j = 1, 2,$$

where Ψ is either the function

$$(y_1, y_2, u) \to (-\log(F_1(y_1|u)), -\log(F_2(y_2|u))$$

or

$$(y_1, y_2, u) \to \left(-\log\left(\left|F_1(y_1|u) + r_n\delta^*_{1,\theta,b}(y_1, y_2, u)\right|\right), -\log\left(\left|F_2(y_2|u) + r_n\delta^*_{2,\theta,b}(y_1, y_2, u)\right|\right)\right)$$

which are VC-classes. This allows us to prove that there exist positive constants C and V such that

$$\sup_{\mathcal{Q}} N(\{G_n(\theta, b) : \theta \in \Theta\}, L_2(Q), \tau\|E_n\|_{Q,2}) \leqslant C\left(\frac{1}{\tau}\right)^V,$$

from which (33) follows. This achieves the proof of Theorem 3.1.

6.8 Proof of Theorem 3.2

One can check that the proof of Theorems 2.3 and 2.4 are mainly due to the asymptotic properties of $\widehat{\Delta}^{(j)}_{\alpha,x,t}, j = 1, 2$ and 3. Thus, if we are able to prove that the two key statistics T_n and $\breve{T}_n$ are sufficiently close enough, in the sense that

$$\sup_{t\in[0,1],a\in[1/2,1]} \sqrt{nh^p} \left|\breve{T}_n - T_n\right| (K, a, t, \lambda, \beta, \gamma|x) = o_{\mathbb{P}}(1), \tag{34}$$

and

$$\sup_{t\in[0,1],a\in[1/2,1]} \sqrt{nh^p}\mathbb{E}\left[\left|\breve{T}_n - T_n\right|\right] (K, a, t, \lambda, \beta, \gamma|x) = o(1), \tag{35}$$

then we can swap $\widehat{\Delta}^{(j)}_{\alpha,x,t}$ by $\breve{\Delta}^{(j)}_{\alpha,x,t}$, $j = 1, 2$ and 3. According to Theorem 3.1, (34) is a direct consequence of (35). So it remains to prove (35). Note that

$$\begin{aligned}
\sqrt{nh^p}\,\mathbb{E}\left[\left|\breve{T}_n - T_n\right|\right] (K, a, t, \lambda, \beta, \gamma|x) &= \sqrt{n}\,\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n}\left[\sqrt{h^p}K_h(x - X_i)a^\gamma \breve{Z}^\beta_{n,t,i} e^{-\lambda a \breve{Z}_{n,t,i}}\right.\right.\right. \\
&\qquad \left.\left.\left. -\sqrt{h^p}K_h(x - X_i)a^\gamma Z^\beta_{t,i} e^{-\lambda a Z_{t,i}}\right]\right|\right] \\
&\leqslant \sqrt{n}\,\mathbb{E}\left[\left|g_{\theta,\delta_n,n}(Y^{(1)}_1, Y^{(2)}_1, X_1) - g_{\theta,\delta_0,n}(Y^{(1)}_1, Y^{(2)}_1, X_1)\right|\right] \\
&\leqslant \sqrt{n}\,PG_n(\theta, b),
\end{aligned}$$

since $\delta_n \in \delta_0 + r_n\mathcal{B}(0, b)$ where $\mathcal{B}(0, b) := \{\delta : \|\delta\|_H \leqslant b\}$. This implies that

$$\sup_{t\in[0,1],a\in[1/2,1]} \sqrt{nh^p}\,\mathbb{E}\left[\left|\breve{T}_n - T_n\right|\right] (K, a, t, \lambda, \beta, \gamma|x) \leqslant \sup_{t\in[0,1],a\in[1/2,1]} \sqrt{n}PG_n(\theta, b) = o(1)$$

by Assertion 1 since it is clear from its proof that $b_n \to 0$ can be replaced by any fixed value b in (31) without changing the conclusion. This achieves the proof of Theorem 3.2.

Acknowledgement

This work was supported by a research grant (VKR023480) from VILLUM FONDEN and an international project for scientific cooperation (PICS-6416).

References

Aliprantis, C. D. and Border, K. C. (2006). *Infinite dimensional analysis, A hitchhiker's guide*, third edition, Springer, Berlin.

Bai, S. and Taqqu, M. S. (2013). Multivariate limit theorems in the context of long-range dependence. *Journal of Time Series Analysis*, 34, 717–743.

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85, 549–559.

Billingsley, P. (1995). *Probability and measure*, third edition, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

Bücher, A., Dette, H. and Volgushev, S. (2011). New estimators of the Pickands dependence function and a test for extreme-value dependence. *Annals of Statistics*, 39, 1963–2006.

Capéraà, P., Fougères, A.L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84, 567–577.

Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557-2589.

Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. *TEST*, 20, 311–333.

Fils-Villetard, A., Guillou, A. and Segers, J. (2008). Projection estimators of Pickands dependence functions. *Canadian Journal of Statistics*, 36, 369–382.

Gannoun, A., Girard, S., Guinot, C. and Saracco, J. (2002). Reference ranges based on nonparametric quantile regression. *Statistics in Medicine*, 21, 3119–3135.

Gijbels, I., Omelka, M. and Veraverbeke, N. (2015). Estimation of a copula when a covariate affects only marginal distributions. *Scandinavian Journal of Statistics*, 42, 1109–1126.

Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, 37, 503–522.

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, 38, 907–921.

Giné, E., Koltchinskii, V. and Zinn, J. (2004). Weighted uniform consistency of kernel density estimators. *The Annals of Probability*, 32, 2570–2605.

Goegebeur, Y., Guillou, A. and Rietsch, T. (2015). Robust conditional Weibull-type estimation. *Annals of the Institute of Statistical Mathematics*, 67, 479–514.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*, second edition, Springer-Verlag, New York.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Annals of Statistics*, 15, 780–799.

Pickands, J. (1981). Multivariate extreme value distributions. *Bulletin de l'Institut International de Statistique*, 49, 859–878.

Portier, F. and Segers, J. (2015). On the weak convergence of the empirical conditional copula under a simplifying assumption. *ArXiv:1511.06544*.

Severini, T. A. (2005). *Elements of distribution theory*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8, 229–231.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, 3, Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes, with applications to statistics*, Springer Series in Statistics. Springer-Verlag, New York.

van der Vaart, A. W. and Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *Asymptotics: Particles, Processes and Inverse Problems*, IMS Lecture Notes Monograph Series, 55, 234–252.

Yao, Q. (1999). Conditional predictive regions for stochastic processes. *Technical report*, University of Kent at Canterbury.