

Kernel regression with Weibull-type tails

supporting information

Tertius de Wet *

Yuri Goegebeur †

Armelle Guillou ‡

Michael Osmann §

May 18, 2015

In order to prove Lemma 1, we need the following preliminary result that we state, for notational convenience, in the univariate framework.

Lemma S1. *Let U denote the tail quantile function of a distribution function F , i.e. $U(y) = \inf\{x : F(x) \geq 1 - 1/y\}$, $y > 1$. If for some positive function a*

$$\lim_{p \uparrow 1} \frac{U\left(\frac{u}{1-p}\right) - U\left(\frac{1}{1-p}\right)}{a\left(\frac{1}{1-p}\right)} = \ln u, \quad \forall u > 0, \quad (1)$$

then, with $V \sim U(0, 1)$,

$$\lim_{p \uparrow 1} \frac{U\left(\frac{1}{1-p}\right)}{a\left(\frac{1}{1-p}\right)} \mathbb{E} \left[\ln U\left(\frac{1}{1-V}\right) - \ln U\left(\frac{1}{1-p}\right) \mid V > p \right] = 1.$$

*Department of Statistics and Actuarial Science, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa (email: tdewet@sun.ac.za).

†Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: yuri.goegebeur@imada.sdu.dk).

‡Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France (email: armelle.guillou@math.unistra.fr).

§Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: mosma@imada.sdu.dk).

Proof of Lemma S1

We have

$$\begin{aligned} & \mathbb{E} \left[\ln U \left(\frac{1}{1-V} \right) - \ln U \left(\frac{1}{1-p} \right) \middle| V > p \right] \\ &= \frac{1}{1-p} \int_p^1 \left[\ln U \left(\frac{1-p}{1-u} \frac{1}{1-p} \right) - \ln U \left(\frac{1}{1-p} \right) \right] du \\ &= \int_1^\infty \left[\ln U \left(\frac{z}{1-p} \right) - \ln U \left(\frac{1}{1-p} \right) \right] \frac{dz}{z^2}. \end{aligned}$$

It is well-known that (1) implies

$$\lim_{p \uparrow 1} \frac{U \left(\frac{1}{1-p} \right)}{a \left(\frac{1}{1-p} \right)} \left[\ln U \left(\frac{u}{1-p} \right) - \ln U \left(\frac{1}{1-p} \right) \right] = \ln u, \quad \forall u > 0,$$

see for instance de Haan and Ferreira (2006), p 101. Now let $\tilde{a}(y) := a(y)/U(y)$, and consider a function $\tilde{a}_0 \sim \tilde{a}$. Thus

$$\begin{aligned} & \frac{U \left(\frac{1}{1-p} \right)}{a \left(\frac{1}{1-p} \right)} \mathbb{E} \left[\ln U \left(\frac{1}{1-V} \right) - \ln U \left(\frac{1}{1-p} \right) \middle| V > p \right] \\ &= \frac{\tilde{a}_0 \left(\frac{1}{1-p} \right)}{\tilde{a} \left(\frac{1}{1-p} \right)} \int_1^\infty \frac{\ln U \left(\frac{z}{1-p} \right) - \ln U \left(\frac{1}{1-p} \right)}{\tilde{a}_0 \left(\frac{1}{1-p} \right)} \frac{dz}{z^2} \\ &= \frac{\tilde{a}_0 \left(\frac{1}{1-p} \right)}{\tilde{a} \left(\frac{1}{1-p} \right)} \left\{ \int_1^\infty \ln z \frac{dz}{z^2} + \int_1^\infty \left[\frac{\ln U \left(\frac{z}{1-p} \right) - \ln U \left(\frac{1}{1-p} \right)}{\tilde{a}_0 \left(\frac{1}{1-p} \right)} - \ln z \right] \frac{dz}{z^2} \right\} \\ &= \frac{\tilde{a}_0 \left(\frac{1}{1-p} \right)}{\tilde{a} \left(\frac{1}{1-p} \right)} \left\{ 1 + \int_1^\infty \left[\frac{\ln U \left(\frac{z}{1-p} \right) - \ln U \left(\frac{1}{1-p} \right)}{\tilde{a}_0 \left(\frac{1}{1-p} \right)} - \ln z \right] \frac{dz}{z^2} \right\}. \end{aligned}$$

As for the integral in the right-hand side, we use Proposition B.2.17 of de Haan and Ferreira (2006) p 382, so that for all $\varepsilon, \delta > 0$ and for p sufficiently large

$$\begin{aligned} \int_1^\infty \left| \frac{\ln U \left(\frac{z}{1-p} \right) - \ln U \left(\frac{1}{1-p} \right)}{\tilde{a}_0 \left(\frac{1}{1-p} \right)} - \ln z \right| \frac{dz}{z^2} &\leq \varepsilon \int_1^\infty z^{\delta-2} dz \\ &= \frac{\varepsilon}{1-\delta}, \end{aligned}$$

provided $\delta < 1$. Collecting the terms establishes the result of Lemma S1. \square

A note about Assumption (\mathcal{F})

In this section we elaborate on the smoothness condition (\mathcal{F}) . We will focus on the case $t = 0$ and show that the assumption

$$\lim_{n \rightarrow \infty} \sup_{z \in \Omega} \left| \frac{\bar{F}(\omega_n; x - h_n z)}{\bar{F}(\omega_n; x)} - 1 \right| = 0, \quad (2)$$

can be satisfied by imposing some more structure on \bar{F} and some additional conditions on h_n and ω_n . In particular, assume that $\ell^*(\cdot; x)$ is normalized, i.e.

$$\ell^*(y; x) = c(x) \exp \left(\int_1^y \frac{\varepsilon(v; x)}{v} dv \right),$$

where $c(x) > 0$ and $\varepsilon(v; x) \rightarrow 0$ as $v \rightarrow \infty$, and impose the following Hölder continuity conditions

$$\begin{aligned} |\theta(x) - \theta(y)| &\leq M_\theta \|x - y\|^{\eta_\theta}, \\ |c(x) - c(y)| &\leq M_c \|x - y\|^{\eta_c}, \\ \sup_{v \geq 1} |\varepsilon(v; x) - \varepsilon(v; y)| &\leq M_\varepsilon \|x - y\|^{\eta_\varepsilon}, \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm on \mathbb{R}^q , and $M_\theta, M_c, M_\varepsilon, \eta_\theta, \eta_c, \eta_\varepsilon$ are positive constants.

Using the well known inequality $|\exp(x) - 1| \leq \exp(|x|)|x|$ we obtain

$$\begin{aligned} \left| \frac{\bar{F}(\omega_n; x - h_n z)}{\bar{F}(\omega_n; x)} - 1 \right| &\leq \exp \left(\omega_n^{1/\theta(x)} \ell^*(\omega_n; x) \left| \omega_n^{1/\theta(x-h_n z) - 1/\theta(x)} \frac{\ell^*(\omega_n; x - h_n z)}{\ell^*(\omega_n; x)} - 1 \right| \right) \\ &\quad \times \omega_n^{1/\theta(x)} \ell^*(\omega_n; x) \left| \omega_n^{1/\theta(x-h_n z) - 1/\theta(x)} \frac{\ell^*(\omega_n; x - h_n z)}{\ell^*(\omega_n; x)} - 1 \right|. \end{aligned}$$

Furthermore,

$$\begin{aligned} \left| \omega_n^{1/\theta(x-h_n z) - 1/\theta(x)} \frac{\ell^*(\omega_n; x - h_n z)}{\ell^*(\omega_n; x)} - 1 \right| &\leq \left| \omega_n^{1/\theta(x-h_n z) - 1/\theta(x)} - 1 \right| \frac{\ell^*(\omega_n; x - h_n z)}{\ell^*(\omega_n; x)} \\ &\quad + \left| \frac{\ell^*(\omega_n; x - h_n z)}{\ell^*(\omega_n; x)} - 1 \right|. \end{aligned} \quad (3)$$

We now examine the two terms in the right hand side of (3) separately.

The Hölder continuity of $\theta(x)$ and the inequality $|\exp(x) - 1| \leq \exp(|x|)|x|$ lead for n large to

$$\left| \omega_n^{1/\theta(x-h_n z) - 1/\theta(x)} - 1 \right| \leq \exp \left(\frac{M_\theta h_n^{\eta_\theta} \ln \omega_n}{\theta(x) \underline{\theta}(x)} \right) \frac{M_\theta h_n^{\eta_\theta} \ln \omega_n}{\theta(x) \underline{\theta}(x)}$$

with $\underline{\theta}(x) := \inf_{z \in \Omega} \theta(x - z)$. Thus, by assuming that $h_n^{\eta_\theta} \ln \omega_n \rightarrow 0$ as $n \rightarrow \infty$, we get for n large the inequality

$$\left| \omega_n^{1/\theta(x-h_n z) - 1/\theta(x)} - 1 \right| \leq C_1 h_n^{\eta_\theta} \ln \omega_n, \quad (4)$$

where C_1 is a positive constant.

Concerning ℓ^* we get the following decomposition

$$\begin{aligned} \left| \frac{\ell^*(\omega_n; x - h_n z)}{\ell^*(\omega_n; x)} - 1 \right| &\leq \left| \frac{c(x - h_n z)}{c(x)} - 1 \right| \exp \left(\int_1^{\omega_n} \frac{\varepsilon(v; x - h_n z) - \varepsilon(v; x)}{v} dv \right) \\ &\quad + \left| \exp \left(\int_1^{\omega_n} \frac{\varepsilon(v; x - h_n z) - \varepsilon(v; x)}{v} dv \right) - 1 \right|. \end{aligned} \quad (5)$$

Again by the Hölder continuity conditions we obtain

$$\left| \frac{c(x - h_n z)}{c(x)} - 1 \right| \leq \frac{M_c h_n^{\eta_c}}{c(x)}, \quad (6)$$

and, for n large

$$\left| \exp \left(\int_1^{\omega_n} \frac{\varepsilon(v; x - h_n z) - \varepsilon(v; x)}{v} dv \right) - 1 \right| \leq C_2 h_n^{\eta_\varepsilon} \ln \omega_n, \quad (7)$$

for C_2 a positive constant.

Combining (3), (4), (5), (6), (7), we get thus for n large that

$$\begin{aligned} \omega_n^{1/\theta(x)} \ell^*(\omega_n; x) \left| \omega_n^{1/\theta(x-h_n z) - 1/\theta(x)} \frac{\ell^*(\omega_n; x - h_n z)}{\ell^*(\omega_n; x)} - 1 \right| &\leq \\ &C_3 h_n^{\eta_c} \omega_n^{1/\theta(x)} \ell^*(\omega_n; x) + C_4 h_n^{\eta_\theta \wedge \eta_\varepsilon} \omega_n^{1/\theta(x)} \ell^*(\omega_n; x) \ln \omega_n, \end{aligned}$$

with $C_3, C_4 > 0$.

Thus (2) will be satisfied if one assumes additionally that

$$h_n^{\eta_c} \omega_n^{1/\theta(x)} \ell^*(\omega_n; x) \rightarrow 0$$

and

$$h_n^{\eta_\theta \wedge \eta_\varepsilon} \omega_n^{1/\theta(x)} \ell^*(\omega_n; x) \ln \omega_n \rightarrow 0,$$

as $n \rightarrow \infty$.

Proof of Lemma 2

From the rule of repeated expectations we obtain

$$\begin{aligned}
\tilde{m}_n(K, t; x) &= \mathbb{E}[K_{h_n}(x - X)m(\omega_n, t; X)] \\
&= \int_{\mathbb{R}^q} K_{h_n}(x - u)m(\omega_n, t; u)g(u)du \\
&= \int_{\Omega} K(z)m(\omega_n, t; x - h_n z)g(x - h_n z)dz.
\end{aligned}$$

Straightforward calculations lead to

$$\begin{aligned}
&|\tilde{m}_n(K, t; x) - m(\omega_n, t; x)g(x)| \\
&\leq m(\omega_n, t; x) \int_{\Omega} K(z)|g(x - h_n z) - g(x)|dz \\
&+ g(x) \int_{\Omega} K(z)|m(\omega_n, t; x - h_n z) - m(\omega_n, t; x)|dz \\
&+ \int_{\Omega} K(z)|m(\omega_n, t; x - h_n z) - m(\omega_n, t; x)||g(x - h_n z) - g(x)|dz \\
&=: T_1 + T_2 + T_3.
\end{aligned}$$

Using (\mathcal{G}) and (\mathcal{K}) we obtain

$$T_1 \leq m(\omega_n, t; x)c_g h_n \int_{\Omega} K(z)d(0, z)dz = m(\omega_n, t; x)g(x)O(h_n).$$

Concerning T_2 , using (\mathcal{F}) and (\mathcal{K})

$$\begin{aligned}
T_2 &= g(x)m(\omega_n, t; x) \int_{\Omega} K(z) \left| \frac{m(\omega_n, t; x - h_n z)}{m(\omega_n, t; x)} - 1 \right| dz \\
&= g(x)m(\omega_n, t; x)O(\Phi(\omega_n, h_n; x)).
\end{aligned}$$

Similar arguments lead to $T_3 = m(\omega_n, t; x)g(x)O(h_n\Phi(\omega_n, h_n; x))$. Now, collecting all the terms establishes Lemma 2. \square

Proof of Lemma 4

Let $r_n := \sqrt{nh_n^q \bar{F}(\omega_n; x)g(x)}$. By straightforward calculations we can write

$$\begin{aligned}
r_n \left(\frac{\hat{\bar{F}}(\omega_n; x)}{\bar{F}(\omega_n; x)} - 1 \right) &= \frac{g(x)}{\hat{g}_n(x)} r_n \frac{T_n^{(0)}(x, K) - \mathbb{E}(T_n^{(0)}(x, K))}{\bar{F}(\omega_n; x)g(x)} \\
&\quad + \frac{g(x)}{\hat{g}_n(x)} r_n \frac{\mathbb{E}(T_n^{(0)}(x, K)) - \bar{F}(\omega_n; x)g(x)}{\bar{F}(\omega_n; x)g(x)} \\
&\quad - \frac{g(x)}{\hat{g}_n(x)} r_n \frac{\hat{g}_n(x) - g(x)}{g(x)} \\
&=: T_4 + T_5 - T_6.
\end{aligned}$$

From Lemma 3 we obtain that

$$r_n \left[\frac{T_n^{(0)}(x, K)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(0)}(x, K)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \xrightarrow{\mathcal{D}} N(0, \|K\|_2^2),$$

and hence, also using Parzen's results on kernel density estimation (Parzen, 1962), summarized in Lemma 3 in Daouia *et al.* (2013), $T_4 \xrightarrow{\mathcal{D}} N(0, \|K\|_2^2)$. Concerning T_5 , we use Lemma 3 in Daouia *et al.* (2013) and Lemmas 1 and 2 from the present paper to obtain $T_5 = O_{\mathbb{P}}(r_n h_n) + O_{\mathbb{P}}(r_n \Phi(\omega_n, h_n; x))$. Finally, $T_6 = O_{\mathbb{P}}(r_n h_n) + O_{\mathbb{P}}(\sqrt{\bar{F}(\omega_n; x)})$ by Lemma 3 in Daouia *et al.* (2013). Combining these results establishes our Lemma 4. \square

Proof of Theorem 2

First consider

$$\tilde{\theta}_n^{(2)}(x; t, K_1, K_2) := \frac{(-\ln \bar{F}(\omega_n; x))T_n^{(t+1)}(x, K_1)}{(t+1)T_n^{(t)}(x, K_2)}$$

and write

$$\begin{aligned}
& r_n \left(\tilde{\theta}_n^{(2)}(x; t, K_1, K_2) - \theta(x) \right) \\
&= \frac{\bar{F}(\omega_n; x)g(x)}{(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_2)} \\
&\quad \times \left\{ r_n (-\ln \bar{F}(\omega_n; x))^{t+1} \left[\frac{T_n^{(t+1)}(x, K_1)}{(t+1)\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(t+1)}(x, K_1)}{(t+1)\bar{F}(\omega_n; x)g(x)} \right) \right] \right. \\
&\quad - \theta(x) r_n (-\ln \bar{F}(\omega_n; x))^t \left[\frac{T_n^{(t)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} - \mathbb{E} \left(\frac{T_n^{(t)}(x, K_2)}{\bar{F}(\omega_n; x)g(x)} \right) \right] \\
&\quad \left. + r_n \left[\frac{\mathbb{E}((-\ln \bar{F}(\omega_n; x))^{t+1} T_n^{(t+1)}(x, K_1)) - \theta(x)(t+1)\mathbb{E}((-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_2))}{(t+1)\bar{F}(\omega_n; x)g(x)} \right] \right\} \\
&=: \frac{\bar{F}(\omega_n; x)g(x)}{(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_2)} \{T_7 + T_8 + T_9\}.
\end{aligned}$$

From Lemma 3, we have

$$T_7 + T_8 \xrightarrow{\mathcal{D}} N \left(0, \frac{\theta^{2t+2}(x)\Gamma(2t+1)}{t+1} [2(2t+1)\|K_1\|_2^2 + (t+1)\|K_2\|_2^2 - 2(2t+1)\|K_1 K_2\|_1] \right).$$

Concerning T_9 , using Lemmas 1 and 2, we have

$$\begin{aligned}
T_9 &= r_n \left\{ \theta^t(x)\Gamma(t+1)b(-\ln \bar{F}(\omega_n; x); x)(1 + o(1)) + O \left(\frac{1}{(-\ln \bar{F}(\omega_n; x))^{1-\varepsilon}} \right) \right. \\
&\quad \left. + O(h_n) + O(\Phi(\omega_n, h_n; x)) \right\},
\end{aligned}$$

and hence under our assumptions, we have $T_9 \rightarrow \lambda\sqrt{g(x)}\theta^t(x)\Gamma(t+1)$.

According to Lemma 3, $(-\ln \bar{F}(\omega_n; x))^t T_n^{(t)}(x, K_2)/(\bar{F}(\omega_n; x)g(x)) = \theta^t(x)\Gamma(t+1) + o_{\mathbb{P}}(1)$. Collecting the above results gives

$$\begin{aligned}
& r_n(\tilde{\theta}_n^{(2)}(x; t, K_1, K_2) - \theta(x)) \xrightarrow{\mathcal{D}} N \left(\lambda\sqrt{g(x)}, \right. \\
&\quad \left. \frac{\theta^2(x)\Gamma(2t+1)}{(t+1)\Gamma^2(t+1)} [2(2t+1)\|K_1\|_2^2 + (t+1)\|K_2\|_2^2 - 2(2t+1)\|K_1 K_2\|_1] \right).
\end{aligned}$$

To complete the proof we consider

$$\begin{aligned}
r_n(\hat{\theta}_n^{(2)}(x; t, K_1, K_2) - \theta(x)) &= r_n(\tilde{\theta}_n^{(2)}(x; t, K_1, K_2) - \theta(x)) \\
&\quad + r_n \left(-\ln \frac{\hat{F}(\omega_n; x)}{\bar{F}(\omega_n; x)} \right) \frac{T_n^{(t+1)}(x, K_1)}{(t+1)T_n^{(t)}(x, K_2)}.
\end{aligned}$$

Again using Lemmas 1, 2, and 3 we obtain that

$$r_n(\hat{\theta}_n^{(2)}(x; t, K_1, K_2) - \theta(x)) = r_n(\tilde{\theta}_n^{(2)}(x; t, K_1, K_2) - \theta(x)) + O_{\mathbb{P}}\left(\frac{1}{-\ln \bar{F}(\omega_n; x)}\right). \quad \square$$

References

- [1] Daouia, A., Gardes, L., Girard, S., 2013. On kernel smoothing for extremal quantile regression. *Bernoulli*, 19, 2557-2589.
- [2] de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer.
- [3] Parzen, E., 1962. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33, 1065–1076.