Supplementary Material

Estimation of extreme conditional quantiles under a general tail first order condition

Laurent Gardes   Armelle Guillou   Claire Roman

Proof of Lemma 1 – Let $\{t_{n,i},\ i=1,\ldots,n\}$ be a triangular array of real numbers satisfying

$$\min(t_{n,i};\ i=1,\ldots,n) \geq 0 \text{ and } \sum_{i=1}^{n} t_{n,i}^2 = 1. \tag{1}$$

Let $t_n := \max(t_{n,i};\ i=1,\ldots,n)$ and $\nu_n := \mathbb{E}(|V_n|^3)$. In a first step, let us show that if $\nu_n t_n \to 0$ as $n \to \infty$ then, for all $z \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sum_{i=1}^{n} t_{n,i} V_{n,i} \leq z\right) = \Phi(z), \tag{2}$$

where Φ is the cumulative distribution function of a $\mathcal{N}(0,1)$ distribution. Since the $V_{n,i}$ are independent and centered random variables, it suffices to prove that the Lindeberg condition is satisfied, i.e., that

$$\lim_{n\to\infty} \sum_{i=1}^{n} t_{n,i}^2 \mathbb{E}\left(V_{n,i}^2 \mathbb{I}_{\{t_{n,i}|V_{n,i}|>\varepsilon\}}\right) = 0,$$

for all $\varepsilon > 0$. Since $t_{n,i} \leq t_n$ for all $i \in \{1,\ldots,n\}$,

$$\sum_{i=1}^{n} t_{n,i}^2 \mathbb{E}\left(V_{n,i}^2 \mathbb{I}_{\{t_{n,i}|V_{n,i}|>\varepsilon\}}\right) \leq \sum_{i=1}^{n} t_{n,i}^2 \mathbb{E}\left(V_{n,i}^2 \mathbb{I}_{\{t_n|V_{n,i}|>\varepsilon\}}\right) = \mathbb{E}\left(V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}\right),$$

since the $V_{n,i}$ are identically distributed and under (1).

Hölder's inequality entails that $\mathbb{E}\left(V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}\right) \leq \nu_n^{2/3} \left[\mathbb{P}(t_n|V_n|>\varepsilon)\right]^{1/3}$. Chebyshev's inequality ensures that $\mathbb{P}(t_n|V_n|>\varepsilon) \leq t_n^2/\varepsilon^2$ and thus $\mathbb{E}\left(V_n^2 \mathbb{I}_{\{t_n|V_n|>\varepsilon\}}\right) \leq [\nu_n t_n/\varepsilon]^{2/3} \to 0$, as $n \to \infty$, by assumption. Convergence (2) is thus proved for all triangular array $\{t_{n,i},\ i=1,\ldots,n\}$ satisfying (1) with $\nu_n t_n \to 0$.

Now, remark that for all $\omega \in \{\nu_n T_n \to 0\}$, convergence (2) entails that

$$\begin{aligned}
& \left| \mathbb{P}\left(\sum_{i=1}^{n} T_{n,i} V_{n,i} \le z \middle| \{T_{n,i} = T_{n,i}(\omega);\ i=1,\ldots,n\} \right) - \Phi(z) \right| \\
= & \left| \mathbb{P}\left(\sum_{i=1}^{n} T_{n,i}(\omega) V_{n,i} \le z \right) - \Phi(z) \right| \to 0,
\end{aligned}$$

as $n \to \infty$. Note that the last equality is true since the $T_{n,i}$ are independent of the $V_{n,i}$. Hence, since $\mathbb{P}[\nu_n T_n \to 0] = 1$,

$$\lim_{n\to\infty} \left| \mathbb{P}\left(\sum_{i=1}^{n} T_{n,i} V_{n,i} \le z \middle| \{T_{n,i};\ i=1,\ldots,n\} \right) - \Phi(z) \right| = 0 \text{ a.s.} \tag{3}$$

To conclude the proof, let us remark that

$$\begin{aligned}
& \lim_{n\to\infty} \left| \mathbb{P}\left(\sum_{i=1}^{n} T_{n,i} V_{n,i} \le z \right) - \Phi(z) \right| \\
\le & \lim_{n\to\infty} \mathbb{E}\left[\left| \mathbb{P}\left(\sum_{i=1}^{n} T_{n,i} V_{n,i} \le z \middle| \{T_{n,i};\ i=1,\ldots,n\} \right) - \Phi(z) \right| \right] \\
\le & \mathbb{E}\left[\lim_{n\to\infty} \left| \mathbb{P}\left(\sum_{i=1}^{n} T_{n,i} V_{n,i} \le z \middle| \{T_{n,i};\ i=1,\ldots,n\} \right) - \Phi(z) \right| \right] = 0,
\end{aligned}$$

by the dominated convergence theorem and (3).

Proof of Proposition 7 – Remark that

$$\left(\frac{n_{x_0}}{\sigma_n^2(x_0)} \right)^{1/2} \left(\widehat{S}_n^{x_0}\left(y_n(x_0)\right) - S\left(y_n(x_0)|x_0\right) \right) = \sum_{i=1}^{n} T_{n,i}(x_0) V_{n,i}(x_0),$$

with $T_{n,i}(x_0) := (n_{x_0})^{1/2} \mathcal{W}_{n,i}(x_0)$ and

$$V_{n,i}(x_0) := [\sigma_n(x_0)]^{-1} \left(\mathbb{I}_{\{Y_i^{x_0} > y_n(x_0)\}} - S(y_n(x_0)|x_0) \right).$$

It thus suffices to apply Lemma 1 after remarking that $n_{x_0}/\sigma_n^2(x_0) \overset{a.s.}{\sim} v_n^2(x_0)$ and that

$$\mathbb{E}(|V_{n,1}(x_0)|^3) = \sigma_n^{-1}(x_0) \left\{ [S(y_n(x_0)|x_0)]^2 + [1 - S(y_n(x_0)|x_0)]^2 \right\} \sim \sigma_n^{-1}(x_0),$$

as $n \to \infty$, since $S(y_n(x_0)|x_0) \to 0$.

Proof of Proposition 8 – Let $U_1, \ldots, U_n$ be independent uniform random variables independent of the X_i. Since $Y_i^{x_0} = Q(U_i|x_0)$ and $Y_i \stackrel{d}{=} Q(U_i|X_i)$ for all $i \in \{1, \ldots, n\}$,

$$B_n(x_0) \stackrel{d}{=} v_n(x_0) \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \left[\mathbb{I}_{(-\infty, S(y_n(x_0)|X_i))} - \mathbb{I}_{(-\infty, S(y_n(x_0)|x_0))}\right](U_i).$$

From Owen (1987, Lemma 3.4.5), one has for all $\varepsilon > 0$,

$$\mathbb{P}(|B_n(x_0)| > \varepsilon) \quad \leq \quad \varepsilon + \mathbb{P}\left\{v_n(x_0) \sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) \mathbb{E}\left[\Delta_{n,i}(x_0) \middle| \mathbb{X}\right] > \varepsilon^2\right\},$$

where $\mathbb{X} := (X_1, \ldots, X_n)$ and

$$\Delta_{n,i}(x_0) := \left|\mathbb{I}_{(-\infty, S(y_n(x_0)|X_i))} - \mathbb{I}_{(-\infty, S(y_n(x_0)|x_0))}\right|(U_i).$$

Introducing the quantity $D_{n,i}(x_0) := |S(y_n(x_0)|X_i) - S(y_n(x_0)|x_0)|$, it is easy to check that $\mathbb{E}[\Delta_{n,i}(x_0)|\mathbb{X}] \leq 2D_{n,i}(x_0)$. Remarking that

$$\sum_{i=1}^{n} \mathcal{W}_{n,i}(x_0) D_{n,i}(x_0) = W_1\left(\mathcal{W}_{n,x_0}^*, \delta_{x_0}^*\right)$$

leads to $\mathbb{P}(|B_n(x_0)| > \varepsilon) \leq \varepsilon + \mathbb{P}\left[v_n(x_0) W_1\left(\mathcal{W}_{n,x_0}^*, \delta_{x_0}^*\right) > \varepsilon^2/2\right]$. The result is thus proved by using assumption (23) (or equivalently (22)) from the paper.

Proof of Lemma 3 – Let $\tilde{K} := K^2/\|K\|_2^2$ where $\|K\|_2^2 := \int_{\mathcal{U}_p} K^2(y)dy$. It is easy to check that $\tilde{K}$ also satisfy condition **(K)**. Hence, Lemma 2 entails that almost surely,

$$\lim_{n\to\infty} \frac{\|K\|_2^2}{nh_n^p} n_{x_0} = \lim_{n\to\infty} \hat{f}_n^2(x_0) \Bigg/ \left[\frac{1}{nh_n^p} \sum_{i=1}^{n} \tilde{K}\left(\frac{x_0 - X_i}{h_n}\right)\right] = f(x_0).$$

Hence, almost surely, $n_{x_0} \sim f(x_0) nh_n^p / \|K\|_2^2 =: m_n(x_0)$. It is easy to infer that, as soon as $nh_n^p S(y_n(x_0)|x_0) \to \infty$, we have

$$\frac{m_n(x_0)}{S(y_n(x_0)|x_0)} \left(\max_{1\leq i\leq n} \mathcal{W}_{n,i}^{\mathrm{NW}}(x_0, h_n)\right)^2 \leq \frac{f(x_0)}{\|K\|_2^2} \frac{1}{nh_n^p S(y_n(x_0)|x_0)} \frac{\|K\|_\infty^2}{\hat{f}_n^2(x_0)} \xrightarrow{a.s.} 0.$$

Similarly, using Assumption **(K)**, we have

$$\sum_{i=1}^{n} \mathcal{W}_{n,i}^{\mathrm{NW}}(x_0, h_n) \left|\frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)} - 1\right| \leq \sup_{\|x - x_0\| \leq h_n} \left|\frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1\right|$$

from which Lemma 3 follows according to Theorem 1.

Proof of Lemma 4 – First, remark that since $k_n \to \infty$ as $n \to \infty$,

$$\frac{(\ell+1)^2}{2\ell+1}\frac{n_{x_0}}{k_n} = \frac{(\ell+1)^2}{k_n(2\ell+1)}\left(\sum_{i=1}^{k_n} i^\ell\right)^2 \Big/ \sum_{i=1}^{k_n} i^{2\ell} \to 1,$$

as $n \to \infty$. Thus, $n_{x_0} \sim m_n(x_0)$ with $m_n(x_0) = (2\ell+1)/(\ell+1)^2 k_n$. As soon as $k_n S(y_n(x_0)|x_0) \to \infty$, we have

$$\frac{m_n(x_0)}{S(y_n(x_0)|x_0)}\left(\max_{1\le i\le n} \mathcal{W}_{n,i}^{\rm NN}(x_0,k_n)\right)^2 = \frac{2\ell+1}{k_n S(y_n(x_0)|x_0)} \to 0.$$

Using the bound

$$\sum_{i=1}^n \mathcal{W}_{n,i}^{\rm NN}(x_0,k_n)\left|\frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)} - 1\right| \le \sup_{\|x-x_0\|\le D_{(k_n)}(x_0)}\left|\frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1\right|,$$

we prove Lemma 4 by applying Theorem 1.

Proof of Lemma 5 – We start by remarking that

$$\sum_{i=1}^n \mathbb{I}_{[0,1]}\left(\left\|\frac{X_i - x_0}{h_n}\right\|_\infty\right)\mathbb{I}_{[0,1]}\left(\frac{r(i)}{k_n}\right) = k_n \wedge M_n.$$

Then, straightforward calculation shows that

$$n_{x_0}^{-1} = \frac{\tau^2}{M_n} + \frac{2\tau(1-\tau)}{k_n \vee M_n} + \frac{(1-\tau)^2}{k_n}.$$

Next, since by assumption $nh_n^p/\log\log n \to \infty$ and since the uniform kernel satisfies condition **(K)**, Lemma 2 ensures that $(2h_n)^{-p}n^{-1}M_n \xrightarrow{a.s.} f(x_0)$ as $n \to \infty$. Hence, as a first conclusion, $n_{x_0} \sim \ell_n 2^p f(x_0) C^{-2}(\kappa) =: m_n(x_0)$ almost surely. Furthermore,

$$\max_{1\le i\le n} \mathcal{W}_{n,i}^{\rm LC}(x_0,\tau,h_n,k_n) \le \frac{\tau}{M_n} + \frac{1-\tau}{k_n}.$$

Hence, using again the almost sure convergence $(2h_n)^{-p}n^{-1}M_n \to f(x_0)$,

$$\lim_{n\to\infty} \ell_n \max_{1\le i\le n} \mathcal{W}_{n,i}^{\rm LC}(x_0,\tau,h_n,k_n) = \frac{\tau(\kappa\wedge 1)}{2^p f(x_0)} + (1-\tau)(\kappa^{-1}\wedge 1),$$

almost surely for all $\kappa \in [0,\infty]$. As a consequence, since $\ell_n S(y_n(x_0)|x_0) \to \infty$, condition (21) from the paper is satisfied. Finally, using the bounds obtained in the proofs of Lemmas 3 and 4, one has

$$\sum_{i=1}^n \mathcal{W}_{n,i}^{\rm LC}(x_0,\tau,h_n,k_n)\left|\frac{S(y_n(x_0)|X_i)}{S(y_n(x_0)|x_0)} - 1\right| \le \sup_{\|x-x_0\|\le h_n \vee D_{(k_n)}(x_0)}\left|\frac{S(y_n(x_0)|x)}{S(y_n(x_0)|x_0)} - 1\right|,$$

and thus condition (22) from the paper holds. Theorem 1 concludes the proof.

Proof of Lemma 6 – Since (X, Y) and Z are independent

$$\begin{aligned}\mathbb{E}[g(X,Y,Z)] &= \int\int\left(\int g(x,y,z)\mathbb{P}_Y(dy|X=x)\right)\mathbb{P}_X(dx)\mathbb{P}_Z(dz)\\ &= \int\int \Psi(x,z)\mathbb{P}_X(dx)\mathbb{P}_Z(dz).\end{aligned}$$

The conclusion follows since X and Z are independent.

Reference

Owen, A.B. (1987). *Nonparametric conditional estimation.* Ph.D Dissertation, Stanford University.