Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 2, 954–982
https://doi.org/10.1214/19-AIHP990
© Association des Publications de l'Institut Henri Poincaré, 2020

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

A Central Limit Theorem for Wasserstein type distances between two distinct univariate distributions

Philippe Berthet[a], Jean-Claude Fort[b] and Thierry Klein[a,c]

[a] *Institut de Mathématiques de Toulouse; UMR5219, Université de Toulouse; CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France. E-mail: philippe.berthet@math.univ-toulouse.fr; thierry.klein@math.univ-toulouse.fr*
[b] *MAP5; UMR 8145; Université Paris Descartes; CNRS, 45 rue des Saints Pères 75006 Paris, France. E-mail: jean-claude.fort@parisdescartes.fr*
[c] *ENAC – Ecole Nationale de l'Aviation Civile, Université de Toulouse, France. E-mail: thierry01.klein@enac.fr*

Received 2 February 2018; revised 6 March 2019; accepted 29 March 2019

Abstract. In this article we study the natural nonparametric estimator of a Wasserstein type cost between two distinct continuous distributions F and G on $\mathbb{R}$. The estimator is based on the order statistics of a sample having marginals F, G and any joint distribution. We prove a central limit theorem under general conditions relating the tails and the cost function. In particular, these conditions are satisfied by Wasserstein distances of order $p > 1$ and compatible classical probability distributions.

Résumé. Dans cet article nous étudions l'estimateur non paramétrique naturel d'un coût de type Wasserstein entre deux lois F et G distinctes et continues sur $\mathbb{R}$. Cet estimateur est construit à partir des statistiques d'ordre d'un échantillon d'un couple quelconque de lois marginales F et G. Nous démontrons un théorème limite central sous des conditions générales reliant les queues de distribution à la fonction de coût. En particulier, ces conditions sont satisfaites par les distances de Wasserstein d'ordre $p > 1$ et les lois classiques compatibles.

MSC: 62G30; 62G20; 60F05; 60F17

Keywords: Central Limit Theorems; Generalized Wasserstein distances; Empirical processes; Strong approximation; Dependent samples

1. Introduction

1.1. *Motivation*

In this article we address the problem of estimating the distance between two different distributions with respect to a class of Wasserstein costs that we define in the sequel. The framework is very simple: two samples of independent and identically distributed (i.i.d.) real random variables taking values in $\mathbb{R}$ with continuous cumulative distribution functions (c.d.f.) F and G are available. These samples are not necessarily independent, for instance they may be issued from simultaneous experiments. From these samples we estimate the Wasserstein distances or costs between F and G and we prove a central limit theorem (CLT).

The motivation of this work is to be found in the fast development of computer experiments. Nowadays the output of many computer codes is not only a multidimensional variable but frequently a function computed on so many points that it can be considered as a functional output. In particular this function may be the density or the c.d.f. of a real random variable. To analyze such outputs one needs to choose a distance to compare various c.d.f. Among the possibilities offered by the literature the Wasserstein distances are now commonly used – for more details on general Wasserstein distances we refer to [19]. Since computer codes only provide samples of the underlying distributions, the estimation of such distances are of primordial importance. The p-Wasserstein distance between two univariate probability distributions simply is the L^p distance of simulated random variables from a universal simulator U, uniform on [0, 1], namely

$$W_p^p(F,G) = \int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right|^p du = \mathbb{E}\left|F^{-1}(U) - G^{-1}(U)\right|^p, \tag{1}$$

where F^{-1} is the generalized inverse of F. It is then natural to estimate $W_p^p(F,G)$ by its empirical counterpart that is $W_p^p(\mathbb{F}_n,\mathbb{G}_n)$ where $\mathbb{F}_n$ and $\mathbb{G}_n$ are the empirical c.d.f. of F and G build through i.i.d. samples of F and G, the two samples being possibly dependent.

Many authors were interested in the convergence of $W_p^p(\mathbb{F}_n,F)$, see *e.g.* the survey paper [4] or [1,7–9] in the i.i.d. case. Very few papers are devoted to the estimation of $W_p^p(F,G)$, actually only for $p=2$. In [14] then [11] the authors derive from the functional delta method a CLT for the trimmed version of Mallows distance $\int_\beta^{1-\beta}|F^{-1}(u)-G^{-1}(u)|^2\,du$ with $0<\beta<1/2$, in the independent case then in the dependent framework described above. We are concerned with the untrimmed case $\beta=0$ which so crucially involves the tails of F and G that it cannot be handled by classical direct arguments such as Hadamard differentiability and functional delta method. Two recent works study the convergence of $W_2^2(\mathbb{F}_n,\mathbb{G}_n)$ for two independent samples [10,18]. In [10] very general results are obtained in the multivariate setting, however the estimator is not explicit from the data, the centering in the CLT is $\mathbb{E}W_2^2(\mathbb{F}_n,\mathbb{G}_n)$ rather than $W_2^2(F,G)$ itself, and the limiting variance is also not explicit. In [18] multivariate but finite discrete distributions are considered, thus the trimming is unnecessary, and the CLT is explicit.

To investigate more deeply the univariate setting we consider a large class of convex or non convex costs $c(x,y)$, including in particular $W_p^p(F,G)$ for $p>1$. We study the natural and easily computed nonparametric plug-in estimator. In the main case of infinite support distributions our main contribution is to avoid trimming by working out a sharp analyse of the Wasserstein type stochastic integrals of the involved stochastic processes. We moreover allow dependent marginal samples from F and G, thanks to a new Brownian strong approximation result for k joint marginal quantile processes – Theorem 29 which is of independent interest. We restrict ourselves to strictly separated tails. Our main result is an explicit untrimmed CLT under almost minimal conditions relating the two continuous c.d.f. F and G to the cost $c(x,y)$.

1.2. *Setting*

Let F and G be two c.d.f. on $\mathbb{R}$. The p-Wasserstein distance between F and G is defined to be

$$W_p^p(F,G)=\min_{X\sim F,Y\sim G}\mathbb{E}|X-Y|^p, \tag{2}$$

where $X\sim F$, $Y\sim G$ means that X and Y are random variables with respective c.d.f. F and G. The minimum in (2) has the explicit expression (1). The Wassertein distances can be generalized to Wasserstein costs. Given a real non negative function $c(x,y)$ of two real variables, we consider the Wasserstein cost

$$W_c(F,G)=\min_{X\sim F,Y\sim G}\mathbb{E}c(X,Y). \tag{3}$$

We consider costs for which the minimum in (3) is finite and an analogue of (1) exists.

Definition 1. We call a good cost function any application c from $\mathbb{R}^2$ to $\mathbb{R}$ that defines a negative measure on $\mathbb{R}^2$ in the sense that it satisfies the "measure property" $\mathcal{P}$,

$$\mathcal{P}: c(x',y')-c(x',y)-c(x,y')+c(x,y)\le 0,\quad x\le x',\ y\le y'.$$

Remark 2. It is obvious that $c(x,y)=-xy$ satisfies the $\mathcal{P}$ property and if c satisfies $\mathcal{P}$ then any function of the form $a(x)+b(y)+c(x,y)$ also satisfies $\mathcal{P}$. In particular $(x-y)^2=x^2+y^2-2xy$ satisfies $\mathcal{P}$. More generally if ρ is a convex real function then $c(x,y)=\rho(x-y)$ satisfies $\mathcal{P}$. This is the case of $|x-y|^p$, $p\ge 1$ and for the cost associated to the α-quantile $c(x,y)=(x-y)(\alpha-\mathbf{1}_{x-y<0})$.

When the property $\mathcal{P}$ holds, the following theorem provides an explicit formula for W_c similar to (1).

Theorem 3 (Cambanis, Simon, Stout [5]). *Let c satisfy the "measure property" $\mathcal{P}$ and U be a random variable uniformly distributed on* $[0,1]$, *then*

$$W_c(F,G)=\int_0^1 c\big(F^{-1}(u),G^{-1}(u)\big)\,du=\mathbb{E}c\big(F^{-1}(U),G^{-1}(U)\big).$$

In view of Theorem 3, an estimator of $W_c(F,G)$ based on a sample from the joint distribution of $(F^{-1}(U),G^{-1}(U))$ seems the most natural one. However this optimal distribution is usually unknown, and such a sample is typically not available. Nevertheless, it is not necessary and one can sample from any coupling of the marginal c.d.f. as will be shown below. This is very interesting in practice, since experimental data can then be used without any assumption on the coupling structure. Moreover the joint distribution of the sample only affects the limiting variance in the CLT, not the rate of convergence.

Let $(X_i, Y_i)_{1\le i\le n}$ be an i.i.d. sample of a random vector with distribution H and marginal c.d.f. F and G. Write $\mathbb{F}_n$ and $\mathbb{G}_n$ the random empirical c.d.f. built from the two marginal samples. Let c be a good cost function. Denote by $X_{(i)}$ (resp. $Y_{(i)}$) the ith order statistic of the sample $(X_i)_{1\le i\le n}$ (resp. $(Y_i)_{1\le i\le n}$), i.e. $X_{(1)} \le \cdots \le X_{(n)}$. We have

$$W_c(\mathbb{F}_n, \mathbb{G}_n) = \frac{1}{n}\sum_{i=1}^{n} c(X_{(i)}, Y_{(i)}). \tag{4}$$

By Theorem 3, $W_c(\mathbb{F}_n, \mathbb{G}_n)$ is a natural estimator of $W_c(F, G)$. We study its asymptotic properties when $F \ne G$ and F and G are continuous, and establish the weak convergence of $\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G))$.

1.3. *Overview of the paper*

In order to control the integrals $W_c(F, G)$ and $W_c(\mathbb{F}_n, \mathbb{G}_n)$ we separate out three sets of assumptions. First, about the regularity of F and G and the separation of their tails, with the convention that G has a lighter right-hand tail than F and that we only provide notation for the right hand tails – with an easy adaptation for the left-hand tails. Second, on the rate of increase, the regularity, the asymptotic expansion of the cost $c(x, y)$ and its behavior close to the diagonal $y = x$. The first two sets are hereafter labeled (*FG*) and (C) respectively. They allow to separately select a class of probability distributions and an admissible cost. The third set of assumptions is labeled (*CFG*) and mixes the requirements on (F, G, c) making them compatible.

Conditions (C) encompass a large class of good Wasserstein costs c, however W_1 is not included – see Section 4. Conditions (*FG*) are satisfied by all classical probability distributions since the regular variation of tails is a sufficient condition. It is important to point out that conditions (*FG*) and (*CFG*) are unaffected by the joint distribution Π of the two samples. Given a cost c satisfying conditions (C), conditions (*FG*) and (*CFG*) provide sufficient regularity and tail conditions on F then regularity, tail and closeness conditions on G. The nice feature is that (*CFG*) are almost minimal to ensure that the limiting variance $\sigma^2(H, c)$ is finite whatever the joint distribution H, hence it is close to be minimal for our CLT.

A foreword about our method of proof. The (F, G, c)-dependent technique we propose consists in two major steps. At the first step we combine the assumptions to show that extreme tail terms and approximations taken at an appropriate level of truncation of the integral $W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)$ can be neglected in probability. Next, large quantiles can be centered on a larger scale and their deviation is led by the two marginal empirical quantile processes. All the assumptions (C), (*FG*) and (*CFG*) are required to control the outer integral error processes at the $\sqrt{n}$ rate in the asymptotic expansions below the first truncation. At the second step, since only the most central part of integrals eventually matters in the integral we prove its weak convergence to a Gaussian distribution by means of a new Brownian strong approximation of joint non extreme quantiles. It is noteworthy that most intermediate terms are actually not shown to vanish in probability due to the Brownian coupling, hence we work very closely to the weak convergence. Finally, the joint distribution naturally shows up together with the CLT rate $\sqrt{n}$.

The paper is organized as follows. Assumptions are discussed in Section 2. In Section 3 we state our main result in the form of a CLT for $\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G))$ and several Corollaries for $W_2^2(\mathbb{F}_n, \mathbb{G}_n)$. A discussion follows in Section 4. All the results are proved in Section 5. The Appendix contains the proofs of technical results used in Section 5 and useful complements on the assumptions.

2. Notation and assumptions

2.1. *Notation*

Let H denote the bivariate distribution function of the sample, thus $H(x, y) = \mathbb{P}(X \le x, Y \le y)$, $F(x) = H(x, +\infty)$ and $G(y) = H(+\infty, y)$. For the sake of clarity, we focus on the generic case where the c.d.f. F and G have positive densities $f = F'$ and $g = G'$ supported on the whole line $\mathbb{R}$. Write F^{-1} and G^{-1} their quantile functions. The tail exponential order of decay are defined to be

$$\psi_X(x) = -\log \mathbb{P}(X > x), \qquad \psi_Y(x) = -\log \mathbb{P}(Y > x), \quad x \in \mathbb{R}_+. \tag{5}$$

We introduce the density quantile functions $h_X = f \circ F^{-1}$, $h_Y = g \circ G^{-1}$, and their companion functions

$$H_X(u) = \frac{1-u}{F^{-1}(u)h_X(u)}, \qquad H_Y(u) = \frac{1-u}{G^{-1}(u)h_Y(u)}.$$

For $k \in \mathbb{N}_*$ denote $\mathcal{C}_k(I)$ the set of functions that are k times continuously differentiable on $I \subset \mathbb{R}$, and $\mathcal{C}_0(I)$ the set of continuous functions. Let $\mathcal{M}_2(m,+\infty)$ be the subset of functions $\varphi \in \mathcal{C}_2(\mathbb{R})$ such that φ'' is monotone on $(m,+\infty)$. Write $\mathrm{RV}(\gamma)$ the set of regularly varying functions at $+\infty$ with index $\gamma \geq 0$. We consider slowly varying functions L satisfying

$$L'(x) = \frac{\varepsilon_1(x)L(x)}{x}, \quad \varepsilon_1(x) \to 0 \text{ as } x \to +\infty. \tag{6}$$

This slight restriction is explained in Section A.4. Then for integrability reasons we impose

$$L'(x) \geq \frac{l_1}{x}, \quad l_1 \geq 1. \tag{7}$$

When $\gamma = 0$, we define $\mathrm{RV}_2^+(0,m) = \{L : L \in \mathcal{M}_2(m,+\infty) \text{ such that (6) and (7) hold}\}$. When $\gamma > 0$, we define $\mathrm{RV}_2^+(\gamma,m) = \{\varphi : \varphi \in \mathcal{M}_2(m,+\infty), \varphi(x) = x^\gamma L(x) \text{ such that } L' \text{ obeys (6)}\}$.

2.2. *Assumptions*

2.2.1. *Conditions* (*FG*)

We state this set of conditions for the right-hand tails only, under the convention that F has the heavier tail. Let $m > \max(0, F^{-1}(1/2), G^{-1}(1/2))$ be large enough to satisfy all the subsequent assumptions. Let $\overline{u} = \max(F(m), G(m)) > 1/2$. We assume that there exists $\tau_0 > 0$ such that

(*FG*1) $F, G \in \mathcal{C}_2(\mathbb{R}_+)$, $\quad f, g > 0$ on $\mathbb{R}_+$.

(*FG*2) $(1-u)\left|\left(\log h(u)\right)'\right|$ is bounded on $(\bar{u}, 1)$, $\quad h = h_X, h_Y$.

(*FG*3) H_X, H_Y are bounded on $(\bar{u}, 1)$.

(*FG*4) $\tau(u) = F^{-1}(u) - G^{-1}(u) \geq \tau_0$, $\quad u \geq \overline{u}$.

Remark 4. Assumption (*FG*4) means that the right tails of F and G are asymptotically well separated. In particular it allows translation models.

Rewriting (*FG*2) and (*FG*3) with the density functions we get the following equivalent conditions

$$(FG5) \quad \sup_{x>m} \frac{1-F(x)}{f(x)}\left(\frac{1}{x} + \frac{|f'(x)|}{f(x)}\right) < +\infty \quad \text{and} \quad \sup_{x>m} \frac{1-G(x)}{g(x)}\left(\frac{1}{x} + \frac{|g'(x)|}{g(x)}\right) < +\infty.$$

The following proposition provides a sufficient condition for (*FG*1), (*FG*2), (*FG*3). The proof is omitted and simply relies on (6), (7) and (A.11).

Proposition 5. *If $\psi_X \in \mathrm{RV}_2^+(0,m)$ then F satisfies (FG). If $\psi_X \in \mathrm{RV}_2(\gamma_1,m)$ for some $\gamma_1 > \gamma_0 > 0$ and, if $\gamma_1 = 1$ assume also that $\psi_X(x) = xL(x)$ with $L' \in \mathrm{RV}_1(-1,m)$ and (6), then F satisfies (FG) and moreover (FG3) can be improved into*

$$H_X(u) \leq \frac{1}{\gamma_0 \log(1/(1-u))}, \quad u > F(m). \tag{8}$$

Example 6. By the first part of Proposition 5 all classical probability distributions satisfy (*FG*) since they are smooth enough. An example of heavy tail is the Pareto distribution with parameter $p > 0$ for which

$$\psi_X(x) = p \log x, \qquad F^{-1}(u) = (1-u)^{-1/p}, \qquad H_X(u) = \frac{1}{p},$$

$$h_X(u) = p(1-u)^{1+1/p}, \qquad (1-u)\left|\left(\log h_X(u)\right)'\right| = \frac{1}{p}.$$

An example of light tail is the Weibull distribution with parameter $q > 0$ for which

$$\psi_X(x) = x^q, \qquad F^{-1}(u) = \big(\log\big(1/(1-u)\big)\big)^{1/q}, \qquad H_X(u) = \frac{1}{q\log(1/(1-u))},$$
$$h_X(u) = q(1-u)\big(\log\big(1/(1-u)\big)\big)^{1-1/q}, \qquad (1-u)\big|\big(\log h_X(u)\big)'\big| \sim \frac{1}{q} \quad \text{as } u \to 1,$$

and this distribution is log-convex if $q < 1$, log-concave if $q > 1$. If ψ_X is regularly varying with index $q > 0$ the previous functions are only modified by a slowly varying factor, as for the Gaussian distribution.

2.2.2. *Conditions* (C)

We consider smooth Wasserstein costs satisfying property $\mathcal{P}$. We impose (*wlog*) that $c(x,x) = 0$ and assume that, for $0 < \tau_1 < \tau_0$ and some $\gamma \geq 0$

$$(C1) \quad c(x,y) \geq 0, \quad c \in \mathcal{C}_1\big([-m,m] \times \mathbb{R} \cup \mathbb{R} \times [-m,m]\big).$$
$$(C2) \quad c(x,y) = \rho\big(|x-y|\big) = \exp\big(l\big(|x-y|\big)\big), \quad (x,y) \in (m,+\infty)^2, l \in \mathrm{RV}_2^+(\gamma,\tau_1).$$

Thus c is asymptotically smooth and symmetric. Moreover we need the following contraction of $c(x,y)$ along the diagonal $x = y$. We assume that there exists $d(m,\tau) \to 0$ as $\tau \to 0$ such that

$$(C3) \quad \big|c\big(x',y'\big) - c(x,y)\big| \leq d(m,\tau)\big(\big|x'-x\big| + \big|y'-y\big|\big) \quad \text{for } (x,y), \big(x',y'\big) \in D_m(\tau),$$

where $D_m(\tau) = \{(x,y) : \max(|x|,|y|) \leq m, |x-y| \leq \tau\}$.

Remark 7. Under $(C2)$ we have $\rho(|x-y|) \leq \rho(\max(x,y)) \leq \rho(x) + \rho(y)$, $(x,y) \in (m,+\infty)^2$, hence

$$\sup_{x>m,y>m} \frac{c(x,y)}{\rho(x)+\rho(y)} \leq 1. \tag{9}$$

Example 8. Typical costs satisfying the conditions (C) are, for $\alpha > 1$ and $\beta > 1$,

$$c_\alpha(x,y) = |x-y|^\alpha, \qquad c_\beta^-(x,y) = \exp\big(\big(\log\big(1+|x-y|\big)\big)^\beta\big) - 1, \qquad c_\beta^+(x,y) = \exp\big(|x-y|^\beta\big) - 1. \tag{10}$$

They satisfy $(C1)$, $(C2)$ with $\gamma = 0$, $\gamma = 0$ and $\gamma = \beta$ respectively, and $(C3)$.

2.2.3. *Conditions* (CFG)

Recall that if $(C2)$ holds then $l \in \mathrm{RV}_2^+(\gamma,\tau_1)$. Now when $\gamma = 0$ in order to compare the tail functions and the cost function we need

$$\limsup_{x\to+\infty} \frac{\log(xl'(x))}{\log l(x)} = 1 - \liminf_{x\to+\infty} \frac{\log(1/\varepsilon_1(x))}{\log l(x)} = \theta_1 \in [0,1], \tag{11}$$

where ε_1 is defined in (6). In the case $\gamma > 0$ we set $\theta_1 = 1$. The following crucial assumption (CFG) connects the distribution's tails with the cost function.

$$(CFG) \quad \text{There exists } \theta > 1 + \theta_1 \text{ such that } \big(\psi_X \circ l^{-1}\big)'(x) \geq 2 + \frac{2\theta}{x}, \quad x \geq l(\tau_1).$$

Remark 9. For Wasserstein distances given by $c_\alpha, \alpha > 1$, $l(x) = \alpha\log x$. We have $\gamma = 0$ and $\varepsilon_1(x) = \alpha/l(x)$ in (11) so that the restriction in (CFG) is $\theta > 1$.

Remark 10. If we have, for some $\zeta > 2$,

$$\mathbb{P}(X > x) \leq \frac{1}{\exp(l(x))^\zeta}, \quad x \in (m,+\infty), \tag{12}$$

then $\psi_X(x) \geq \zeta l(x)$ so that (CFG) holds with arbitrarily large θ.

We often make use of the following consequences of (CFG). Integrating (CFG) yields

$$\psi_X \circ l^{-1}(x) \geq 2x + 2\theta \log x + K, \quad x \geq l(\tau_1), \tag{13}$$

where the integrating constant K does not matter and may change from line to line. This also implies

$$\psi_X(x) \geq 2l(x) + 2\theta \log l(x) + K, \quad x \geq \tau_1, \tag{14}$$

and, more importantly for our needs, inverting (13) we obtain

$$l \circ \psi_X^{-1}(x) \leq \frac{x}{2} - \theta \log x + K, \quad x \geq \tau_1. \tag{15}$$

Now, (13) gives

$$\mathbb{P}\big(\rho(X) > x\big) = \mathbb{P}\big(l(X) > \log x\big) = \exp\big(-\psi_X \circ l^{-1}(\log x)\big) \leq \frac{K}{x^2(\log x)^{2\theta}}$$

and since $\theta > 1$ we have

$$\int_m^{+\infty} \sqrt{\mathbb{P}\big(\rho(X) > x\big)}\, dx < +\infty. \tag{16}$$

Remark 11. This is the same kind of condition that ensures the convergence of $W_1(\mathbb{F}_n, F)$ at rate $\sqrt{n}$ in (3.4) of [4]. So it turns out that (16) is almost a minimal assumption in proving Theorem 14. This is confirmed at Lemmas 20 and 21 establishing that the asymptotic variance of $\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G))$ is finite.

Example 12. For the Wasserstein cost c_α from (10), $\alpha > 1$, consider a Pareto distribution, $\psi_X(x) = p \log x$. Then (CFG) reads $\alpha x/p < x/2 - \theta \log x$, and holds if $p > 2\alpha$. Condition (CFG) is then obviously satisfied for all classical distributions with lighter tail than Pareto distributions.

For an over-exponential cost c_β^+ from (10), $\beta > 1$, (CFG) is satisfied if $\mathbb{P}(X > x) \leq \exp(-2x^\beta - \delta \log x)$ with $\delta > 4(1 - \beta)$.

Gaussian distributions are compatible without restriction with any cost less than $\rho(x) = \exp(ax^\gamma)$, $\gamma < 2$, $a > 0$. In the case $\gamma = 2$ the variance of X has to be less than $a/4$ for (CFG) to hold, and G may be any Gaussian distribution different from F with smaller variance or same variance but smaller expectation.

3. Statement of the results

3.1. *Consistency*

Under the following conditions, $W_c(\mathbb{F}_n, \mathbb{G}_n)$ is a consistent estimator of $W_c(F, G)$.

Theorem 13. *Assume that the good cost $c(x, y)$ is such that $0 \leq c(x, y) \leq V(x) + V(y)$ with V a non negative continuous function such that $\mathbb{E}V(X) < +\infty$ and $\mathbb{E}V(Y) < +\infty$. Then*

$$\lim_{n\to+\infty} W_c(\mathbb{F}_n, \mathbb{G}_n) = W_c(F, G) < +\infty \quad a.s.$$

3.2. *A central limit theorem*

Recall that our assumptions are stated for the right hand tails only, and the left hand tail of F and G should be reversed from $\mathbb{R}_-$ to $\mathbb{R}_+$ to obey the same conditions, so that if G has the heavier left tail then the couples (F, X) and (G, Y) are simply exchanged in (FG) and (CFG). In other words, we say that conditions (C), (FG) and (CFG) hold if they hold for $(c(x, y), X, Y)$ as stated above and also for $(c(-x, -y), -X, -Y)$ with possibly different functions ρ, l, ψ and F again denoting the heavier tail.

For $u, v \in (0, 1)$ define the covariance matrix

$$\Sigma(u, v) = \begin{pmatrix} \frac{\min(u,v)-uv}{h_X(u)h_X(v)} & \frac{\Pi(u,v)-uv}{h_X(u)h_Y(v)} \\ \frac{\Pi(v,u)-uv}{h_X(v)h_Y(u)} & \frac{\min(u,v)-uv}{h_Y(v)h_Y(u)} \end{pmatrix}, \quad \Pi(u, v) = \mathbb{P}\big(X \leq F^{-1}(u), Y \leq G^{-1}(v)\big), \tag{17}$$

and the gradient

$$\nabla(u)=\left(\frac{\partial}{\partial x}c\big(F^{-1}(u),G^{-1}(u)\big),\frac{\partial}{\partial y}c\big(F^{-1}(u),G^{-1}(u)\big)\right). \tag{18}$$

Let $\mathcal{N}$ stands for the normal distribution. Our main result is the following.

Theorem 14. *If* (C), (FG) *and* (CFG) *hold then*

$$\sqrt{n}\big(W_c(\mathbb{F}_n,\mathbb{G}_n)-W_c(F,G)\big)\to_{\text{weak}}\mathcal{N}\big(0,\sigma^2(H,c)\big)$$

with

$$\sigma^2(H,c)=\int_0^1\int_0^1\nabla(u)\Sigma(u,v)\nabla(v)\,du\,dv<+\infty. \tag{19}$$

Combining Theorem 14 and Lemma 22 we obtain the following trimmed version. The sequences k_n and k_n^- depend on (c,F,G) and are in $((\log n)/n,1/\sqrt{n})$. Trimming more than below may induce a bias since for any distribution and any cost one can find a sequence $\varepsilon_n\to 0$ such that $\sqrt{n}\int_{1-\varepsilon_n}^1 c(F^{-1}(u),G^{-1}(u))\,du\to+\infty$.

Corollary 15. *If* (C), (FG) *and* (CFG) *hold then, for any positive real sequences* $\varepsilon_n\le k_n/n$ *and* $\varepsilon_n^-\le k_n^-/n$ *with* k_n *and* k_n^- *defined at* (22) *for the right and left tails respectively,*

$$W_{c,n}(\mathbb{F}_n,\mathbb{G}_n)=\int_{\varepsilon_n^-}^{1-\varepsilon_n}c\big(\mathbb{F}_n^{-1}(u),\mathbb{G}_n^{-1}(u)\big)\,du$$

satisfies

$$\sqrt{n}\big(W_{c,n}(\mathbb{F}_n,\mathbb{G}_n)-W_c(F,G)\big)\to_{\text{weak}}\mathcal{N}\big(0,\sigma^2(H,c)\big).$$

Moreover, for any positive real sequences $\varepsilon_n,\varepsilon_n^-\to 0$ *we have*

$$\sqrt{n}\big(W_{c,n}(\mathbb{F}_n,\mathbb{G}_n)-W_{c,n}(F,G)\big)\to_{\text{weak}}\mathcal{N}\big(0,\sigma^2(H,c)\big).$$

Slight changes in the proof of Theorem 14 yield the following version with one known marginal.

Theorem 16. *If* (C), (FG) *and* (CFG) *hold then*

$$\sqrt{n}\big(W_c(\mathbb{F}_n,G)-W_c(F,G)\big)\to_{\text{weak}}\mathcal{N}\big(0,\sigma_x^2(F,c)\big),$$
$$\sqrt{n}\big(W_c(F,\mathbb{G}_n)-W_c(F,G)\big)\to_{\text{weak}}\mathcal{N}\big(0,\sigma_y^2(G,c)\big),$$

with

$$\sigma_x^2(F,c)=\int_0^1\int_0^1\frac{\partial}{\partial x}c\big(F^{-1}(u),G^{-1}(u)\big)\frac{\partial}{\partial x}c\big(F^{-1}(v),G^{-1}(v)\big)\frac{\min(u,v)-uv}{h_X(u)h_X(v)}\,du\,dv\le\sigma^2(H,c)<+\infty,$$
$$\sigma_y^2(G,c)=\int_0^1\int_0^1\frac{\partial}{\partial y}c\big(F^{-1}(u),G^{-1}(u)\big)\frac{\partial}{\partial y}c\big(F^{-1}(v),G^{-1}(v)\big)\frac{\min(u,v)-uv}{h_Y(u)h_Y(v)}\,du\,dv\le\sigma^2(H,c)<+\infty.$$

Theorem 14 can easily be specialized for $W_p^p(F,G)$, $p>1$, since c_p from (10) satisfies (C) and then (FG) and (CFG) provides compatible c.d.f. F and G. For instance, by applying Remarks 9 and 10 the following convergence holds for the square Wasserstein distance and two independent samples, in which case the copula function in (17) is $\Pi(u,v)=uv$ for $u,v\in(0,1)$.

Corollary 17. *Assume that the two samples are independent, that* (FG) *holds and that* $\mathbb{P}(X>x)\le 1/x^{4+\varepsilon}$ *for some* $\varepsilon>0$. *Then*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n(X_{(i)}-Y_{(i)})^2-W_2^2(F,G)\right)\to_{\text{weak}}\mathcal{N}\big(0,\sigma^2(H,c_2)\big)$$

with variance

$$\sigma^2(H,c_2)=4\int_0^1\int_0^1\left(\frac{\min(u,v)-uv}{h_X(u)h_X(v)}+\frac{\min(u,v)-uv}{h_Y(u)h_Y(v)}\right)\big(F^{-1}(u)-G^{-1}(u)\big)\big(F^{-1}(v)-G^{-1}(v)\big)\,du\,dv.$$

The next result could be useful for applications. We say that F is symmetric if $F(-x)=1-F(x)$, $x>0$.

Corollary 18. *Consider a family of c.d.f. defined for $a>0$ and $b\in\mathbb{R}$ by $F_{a,b}(x)=F((x-b)/a)$, $x\in\mathbb{R}$. Assume that F is symmetric and* $\operatorname{var}(Z)=1$, $V_4=\operatorname{var}(Z^2)<+\infty$ *for Z with c.d.f. F. If $(a,b)\neq(a',b')$ and $H(x,y)=F_{a,b}(x)F_{a',b'}(y)$ then it comes*

$$\sigma^2(H,c_2)=4\big(a^2+a'^2\big)\left(\big(b-b'\big)^2+\frac{V_4}{4}\big(a-a'\big)^2\right).$$

By Corollary 18, for two independent samples from two distinct Gaussian distributions $\mathcal{N}(\nu,\zeta^2)$ and $\mathcal{N}(\mu,\xi^2)$ we recover the limiting variance $4(\zeta^2+\xi^2)(\nu-\mu)^2+2(\zeta^2+\xi^2)(\zeta-\xi)^2$ of Theorem 2.2 in [15] which proves that our general non-parametric estimator performs as well as their plug-in parametric estimator in the univariate doubly Gaussian case.

Finally, it is easy to extend Theorem 14 to probability distributions supported by intervals. If (C) holds we have $c(x,y)\to+\infty$ as x or y tends to infinity while the other remains bounded so that $W_c(F,G)<+\infty$ means that the right hand and left hand supports of F and G are finite or infinite simultaneously.

Theorem 19. *Let F and G be supported by intervals and such that $W_c(F,G)<+\infty$. Assume that (FG), (C) and (CFG) hold, with $(FG4)$ discarded on each side – right or left – where the most lightly tailed distribution has bounded support. Then the conclusion of Theorem 14 holds true.*

4. Discussion

In this paper we have proved a CLT for the plug-in estimator of $W_c(F,G)$. This estimator is fast to compute and our easily verified assumptions are valid for a wide class of probability distributions and Wasserstein costs. Moreover our CLT is centered on $W_c(F,G)$ and the limiting variance has a closed form expression. It extends the results of [11] concerning trimmed Mallows distance to untrimmed Wasserstein distances W_p^p, $p>1$ and to more general cost functions, in the framework of a couple of samples having the same size but being possibly dependent. Due to the fact that we work beyond the case of W_2^2 for which an exact expansion exists, we require the marginal distributions to be distinct enough through $(FG4)$. In the limiting case $F=G$ a difficulty actually shows up in the untrimmed Mallows distance. In the one hand, in [11] when $F=G$ the limiting variance at the scale $\sqrt{n}$ is 0, which means that the rate of convergence is faster. In the other hand, in [9] $W_2^2(\mathbb{F}_n,F)$ has an exact rate of convergence n, but the Gaussian distributions are excluded since the expected limiting variance turns to be infinite, so F should have a sub-Gaussian tail. This suggests that it should be the same for $W_2^2(\mathbb{F}_n,\mathbb{G}_n)$.

Since $W_1(\mathbb{F}_n,\mathbb{G}_n)=\int_{\mathbb{R}}|\mathbb{F}_n(x)-\mathbb{G}_n(x)|\,dx$ the mathematical treatment of the distance W_1 is usually more direct. The known CLT type results concern $W_1(\mathbb{F}_n,F)$ as in [7]. Our present work does not catch the distance W_1 because it does not satisfy assumption $(C3)$ – the derivative of the absolute value does not vanish at 0. This is a meaningful border case since the limiting distribution of $\sqrt{n}(W_1(\mathbb{F}_n,\mathbb{G}_n)-W_1(F,G))$ should depends on the set $\{F=G\}$.

In our setting, if $(FG4)$ is not satisfied and the diagonal $\{F=G\}$ is allowed in extremes, the cost is evaluated randomly at 0 or infinity, which leads to a true difficulty.

In order to complete our work it remains to handle three main problems closely related to conditions (C3) and (FG4). Firstly, the case $F=G$, that clearly violates (FG4), for which the speed of weak convergence could be different from $\sqrt{n}$ and the limiting distribution could be non Gaussian, as discussed above. Secondly, the case of W_1 or close to W_1 up to a slowly varying function, for which (C3) does not hold, with $F\neq G$ and $F=G$. The third problem is to extend our results to samples of different sizes, still without assuming independence. We will hopefully achieve these three studies in forthcoming papers.

5. Proofs

5.1. *Proof of Theorem 13*

First note that $\int_0^1 V(\mathbb{F}_n^{-1}(u))\,du=n^{-1}\sum_{i=1}^n V(X_i)$ converges to $\mathbb{E}V(X)=\int_0^1 V(F^{-1}(u))\,du$ on a set Ω_0 of probability one. We also have by Glivenko-Cantelli theorem that $\mathbb{F}_n$ almost surely weakly converges to the continuous F, say on the same Ω_0. In particular, given any ω in Ω_0 the sequence $V(\mathbb{F}_n^{-1}(u))$ converges to $V(F^{-1}(u))$ simultaneously for all u in $(0,1)$ where $F^{-1}(u)$ is continuous, hence almost everywhere on $(0,1)$. Next consider the Borel measurable functions $V(\mathbb{F}_n^{-1})$ and $V(F^{-1})$ as random variables on $(0,1)$ endowed with the uniform measure. By applying Vitali's theorem – see Theorem 5.5 in [17] with $r=1$ – the random variables $V(\mathbb{F}_n^{-1})$ and $V(F^{-1})$ are uniformly integrable. The same holds for $V(\mathbb{G}_n^{-1})$. As $c(\mathbb{F}_n^{-1}(u),\mathbb{G}_n^{-1}(u)) \le V(\mathbb{F}_n^{-1}(u)) + V(G_n^{-1}(u))$ we conclude that $c(\mathbb{F}_n^{-1},\mathbb{G}_n^{-1})$ is uniformly integrable on $(0,1)$ and converges in $L_1(0,1)$. This shows that it holds, on Ω_0,

$$\lim_{n\to+\infty}\int_0^1 c\big(\mathbb{F}_n^{-1}(u),\mathbb{G}_n^{-1}(u)\big)\,du = \int_0^1 c\big(F^{-1}(u),G^{-1}(u)\big)\,du.$$

5.2. *Proof of Theorem* 14

The proof of Theorem 14 is organized as follows. In Section 5.2.1 we prove (19). Section 5.2.2 is dedicated to the proof of the weak convergence of $\sqrt{n}(W_c(\mathbb{F}_n,\mathbb{G}_n) - W_c(F,G))$. As explained at Section 3.2 we only deal with the upper part of the integrals. For that purpose we split the interval $(1/2,1)$ into four parts, $(1/2,F(M))$, $(F(M),1-h_n/n)$, $(1-h_n/n,1-k_n/n)$, $(1-k_n/n,1)$, where $F(M)$, h_n, k_n will be specified further on. The first integral is the main term and the other ones will be proved to be small. We study the integral over $(1-k_n/n,1)$ at Step 1, the one over $(1-h_n/n,1-k_n/n)$ at Step 2 and the one over $(F(M),1-h_n/n)$ at Step 3. Finally, we deal with the main part at Step 4.

5.2.1. *The limiting variance*

In this section we establish that (C), (FG) and (CFG) imply that $\sigma^2(H,c)<+\infty$ in (19). The covariance matrix $\Sigma(u,v)$ and the gradient $\nabla(u)$ are defined at (17) and (18). It is sufficient to study the right hand tails, corresponding to the upper domain of integration $[1/2,1]^2$. As a matter of fact, this implies the same for $[0,1/2]^2$, then similar arguments hold for mixing both tails through $[1/2,1]\times[0,1/2]$ and $[0,1/2]\times[1/2,1]$ by separating the variables exactly as we show below. Hence by cutting $[1/2,1]=[1/2,\overline{u}]\cup[\overline{u},1]$ into mid quantiles and extremes we are reduced to control $\nabla(u)\Sigma(u,v)\nabla(v)$ on $[\overline{u},1]\times[\overline{u},1]$ then on $[1/2,\overline{u}]\times[1/2,1]$. The forthcoming two lemmas are then enough to conclude that (19) is true under (C), (FG) and (CFG).

Lemma 20. *Under* $(C2)$, $(FG1)$, $(FG4)$ *and* (CFG) *we have, for any* $\overline{u}>F(m)$,

$$\sigma^2(\overline{u}) = \int_{\overline{u}}^1\int_{\overline{u}}^1 \nabla(u)\Sigma(u,v)\nabla(v)\,du\,dv < +\infty.$$

Proof. By $(C2)$ we have, for $x\ge y\ge m$,

$$\frac{\partial}{\partial x}c(x,y) = -\frac{\partial}{\partial y}c(x,y) = \frac{\partial}{\partial x}\rho(x-y) = l'(x-y)\rho(x-y) = \rho'(x-y).$$

By $(FG4)$ it holds $F^{-1}(u)\ge\tau(u)=F^{-1}(u)-G^{-1}(u)\ge\tau_0>0$ for $u>F(m)$. Thus, for $u\in[\overline{u},1]$, $\nabla(u)=(\rho'\circ\tau(u),-\rho'\circ\tau(u))$. Let us split $\sigma^2(\overline{u})$ into

$$A_1 = \int_{\overline{u}}^1\int_{\overline{u}}^1 \rho'\circ\tau(u)\frac{\min(u,v)-uv}{h_X(u)h_X(v)}\rho'\circ\tau(v)\,du\,dv,$$

$$A_2 = -\int_{\overline{u}}^1\int_{\overline{u}}^1 \rho'\circ\tau(u)\frac{\Pi(v,u)-uv}{h_X(v)h_Y(u)}\rho'\circ\tau(v)\,du\,dv,$$

$$A_3 = -\int_{\overline{u}}^1\int_{\overline{u}}^1 \rho'\circ\tau(u)\frac{\Pi(u,v)-uv}{h_X(u)h_Y(v)}\rho'\circ\tau(v)\,du\,dv,$$

$$A_4 = \int_{\overline{u}}^1\int_{\overline{u}}^1 \rho'\circ\tau(u)\frac{\min(u,v)-uv}{h_Y(v)h_Y(u)}\rho'\circ\tau(v)\,du\,dv.$$

Observe that if $0<u<v<1$ then

$$0\le\frac{\min(u,v)-uv}{\sqrt{1-u}\sqrt{1-v}} = u\sqrt{\frac{1-v}{1-u}}\le 1,$$

so that we always have $0 \le \min(u,v) - uv \le \sqrt{1-u}\sqrt{1-v}$ and we get

$$|A_1| \le \left(\int_{\overline{u}}^1 \rho' \circ \tau(u) \frac{\sqrt{1-u}}{h_X(u)}\, du\right)^2, \qquad |A_4| \le \left(\int_{\overline{u}}^1 \rho' \circ \tau(u) \frac{\sqrt{1-u}}{h_Y(u)}\, du\right)^2.$$

Consider the bound of $|A_1|$ first. By $(C2)$, ρ' is $\mathcal{C}_1(m,+\infty)$ and positive. Now, as $u \to 1$, $\tau(u) \ge \tau_0 > 0$ is either unbounded or bounded. In both cases we have

$$0 < \rho'\big(\tau(u)\big) \le \max\Big(\rho' \circ F^{-1}(u), \sup_{\tau_0 < x \le l_2} \rho'(x)\Big) \le k_1 \rho' \circ F^{-1}(u)$$

for $k_1 \ge 1$ since by Proposition 31 the increasing function ρ is convex on $(l_2,+\infty)$ under $(C2)$. Observe that ρ is also invertible, so that $\rho(X)$ has quantile function, density function and density quantile function respectively given by

$$F_{\rho(X)}^{-1} = \rho \circ F^{-1}, \qquad f_{\rho(X)} = \frac{f \circ \rho^{-1}}{\rho' \circ \rho^{-1}}, \qquad h_{\rho(X)} = f_{\rho(X)} \circ F_{\rho(X)}^{-1} = \frac{h_X}{\rho' \circ F^{-1}}. \tag{20}$$

Recalling that (CFG) implies (16), the change of variable $x = \rho \circ F^{-1}(u)$ yields

$$\begin{aligned}
\frac{1}{k_1}\int_{\overline{u}}^1 \rho' \circ \tau(u) \frac{\sqrt{1-u}}{h_X(u)}\, du &\le \int_{F(m)}^1 \rho' \circ F^{-1}(u) \frac{\sqrt{1-u}}{h_X(u)}\, du \\
&= \int_{F(m)}^1 \frac{\sqrt{1-u}}{h_{\rho(X)}(u)}\, du = \int_{\rho(m)}^{+\infty} \sqrt{\mathbb{P}\big(\rho(X) > x\big)}\, dx < +\infty.
\end{aligned}$$

Having proved that $|A_1| < +\infty$ let us next study the upper bound of $|A_4|$. Under $(C2)$ and (11) we have, for some $\varepsilon_1(x) \to \gamma$,

$$\rho'(x) = l'(x)\rho(x) = \varepsilon_1(x)\frac{l(x)}{x}\rho(x) \le (1+\gamma)\frac{l(x)^{\theta_1'}}{x}\rho(x),$$

where $\theta_1' \in (\theta_1, \theta - 1)$ if $\gamma = 0$, and $\theta_1' = 1$ if $\gamma > 0$. It then follows from the change of variable $u = G(x)$ that, by setting $\phi = G^{-1} \circ F = \psi_Y^{-1} \circ \psi_X$,

$$\int_{\overline{u}}^1 \rho' \circ \tau(u) \frac{\sqrt{1-u}}{h_Y(u)}\, du \le (1+\gamma)\int_{G^{-1}(\overline{u})}^{+\infty} \frac{(l \circ \phi^{-1}(x))^{\theta_1'}}{\phi^{-1}(x)} \rho \circ \phi^{-1}(x)\sqrt{\mathbb{P}(Y > x)}\, dx. \tag{21}$$

Now, by $(FG4)$ we have

$$x \le \phi^{-1}(x) = F^{-1} \circ G(x) = \psi_X^{-1} \circ \psi_Y(x) = \psi_X^{-1}\left(\log\left(\frac{1}{\mathbb{P}(Y > x)}\right)\right),$$

thus by (15) it holds

$$l \circ \phi^{-1}(x) \le \frac{1}{2}\log\left(\frac{1}{\mathbb{P}(Y > x)}\right) - \theta \log\log\left(\frac{1}{\mathbb{P}(Y > x)}\right) + K.$$

We can bound (21) from above by

$$\begin{aligned}
&(1+\gamma)\int_{\phi(m)}^{+\infty} \frac{(l \circ \phi^{-1}(x))^{\theta_1'}}{\phi^{-1}(x)} \exp\big(l \circ \phi^{-1}(x)\big)\sqrt{\mathbb{P}(Y > x)}\, dx \\
&\quad \le K \int_{\phi(m)}^{+\infty} \frac{(\psi_Y(x))^{\theta_1' - \theta}}{\psi_X^{-1} \circ \psi_Y(x)}\, dx \\
&\quad \le K \int_{\phi(m)}^{+\infty} \frac{1}{x(\psi_Y(x))^{\theta - \theta_1'}}\, dx \le K \int_{\phi(m)}^{+\infty} \frac{1}{x(l(x))^{\theta - \theta_1'}}\, dx.
\end{aligned}$$

The last inequality comes from $\psi_Y(x) > \psi_X(x)$ by $(FG4)$. If $\gamma > 0$ then $\theta - \theta_1' = \theta - 1 > 0$ and $l(x) > x^{\gamma/2}$ hence the bounding integral is finite. If $\gamma = 0$ then $l(x) \ge \log x$ by (7) and having enforced $\theta - \theta_1 > \theta - \theta_1' > 1$ also makes the above integral finite. We have shown that $|A_4| < +\infty$. It remains to bound $A_2 = A_3$. Since F and G are continuous it holds

$$\Pi(u,v) \le \min\big(\mathbb{P}\big(X \le F^{-1}(u)\big), \mathbb{P}\big(Y \le G^{-1}(v)\big)\big) = \min(u,v),$$
$$\Pi(u,v) \ge \mathbb{P}\big(X \le F^{-1}(u)\big) + \mathbb{P}\big(Y \le G^{-1}(v)\big) - 1 = u + v - 1,$$

and thus

$$\Pi(u,v) - uv \le \min(u,v) - uv \le \sqrt{1-u}\sqrt{1-v},$$
$$\Pi(u,v) - uv \ge u + v - 1 - uv = -(1-u)(1-v),$$

which proves that $|\Pi(u,v) - uv| \le \sqrt{1-u}\sqrt{1-v}$. Hence A_2 and A_3 both satisfy

$$\begin{aligned}|A_2| &\le \int_{\overline{u}}^1 \rho' \circ \tau(v) \frac{\sqrt{1-v}}{h_X(v)}\, dv \int_{\overline{u}}^1 \rho' \circ \tau(u) \frac{\sqrt{1-u}}{h_Y(u)}\, du \\ &\le k_1^2 \int_{F(m)}^1 \rho' \circ F^{-1}(v) \frac{\sqrt{1-v}}{h_X(v)}\, du \int_{F(m)}^1 \rho' \circ F^{-1}(u) \frac{\sqrt{1-u}}{h_Y(u)}\, du\end{aligned}$$

where these integrals are already proved to be finite. Finally $\sigma^2(\overline{u}) = A_1 + A_2 + A_3 + A_4 < +\infty$. □

Lemma 21. *Under* $(C1)$, $(C2)$, $(FG1)$, $(FG4)$ *and* (CFG) *we have, for any* $\overline{u} > F(m)$,

$$\sigma_-^2(\overline{u}) = \int_{1/2}^{\overline{u}} \int_{1/2}^1 \nabla(u)\Sigma(u,v)\nabla(v)\, du\, dv < +\infty, \sigma_+^2(\overline{u}) = \int_{\overline{u}}^1 \int_{1/2}^{\overline{u}} \nabla(u)\Sigma(u,v)\nabla(v)\, du\, dv < +\infty.$$

Proof. Since F^{-1} and G^{-1} are bounded on $[1/2, \overline{u}]$ we have, by $(C1)$, that $\nabla(u)$ exists and is bounded on $[1/2, \overline{u}]$. Likewise $(FG1)$ ensures that h_X and h_Y are bounded on $[1/2, \overline{u}]$ hence $\Sigma(u,v)$ is bounded on $[1/2, \overline{u}]^2$. As a consequence,

$$A_0 = \int_{1/2}^{\overline{u}} \int_{1/2}^{\overline{u}} \nabla(u)\Sigma(u,v)\nabla(v)\, du\, dv, \quad |A_0| < +\infty.$$

By $(C2)$ we have $\nabla(u) = (\rho' \circ \tau(u), -\rho' \circ \tau(u))$ on $[\overline{u}, 1]$, thus

$$\begin{aligned}A_{01} &= \int_{1/2}^{\overline{u}} \int_{\overline{u}}^1 \frac{\partial}{\partial x} c\big(F^{-1}(u), G^{-1}(u)\big) \frac{\min(u,v) - uv}{h_X(u)h_X(v)} \rho' \circ \tau(u)\, du\, dv, \\ A_{02} &= -\int_{1/2}^{\overline{u}} \int_{\overline{u}}^1 \frac{\partial}{\partial y} c\big(F^{-1}(u), G^{-1}(u)\big) \frac{\Pi(v,u) - uv}{h_X(v)h_Y(u)} \rho' \circ \tau(u)\, du\, dv, \\ A_{03} &= -\int_{1/2}^{\overline{u}} \int_{\overline{u}}^1 \frac{\partial}{\partial x} c\big(F^{-1}(u), G^{-1}(u)\big) \frac{\Pi(u,v) - uv}{h_X(u)h_Y(v)} \rho' \circ \tau(u)\, du\, dv, \\ A_{04} &= \int_{1/2}^{\overline{u}} \int_{\overline{u}}^1 \frac{\partial}{\partial y} c\big(F^{-1}(u), G^{-1}(u)\big) \frac{\min(u,v) - uv}{h_Y(v)h_Y(u)} \rho' \circ \tau(u)\, du\, dv.\end{aligned}$$

Along the same arguments as in Lemma 20 we have

$$|A_{01}| \le I_X J_X, \qquad |A_{02}| \le I_Y J_X, \qquad |A_{03}| \le I_X J_Y, \qquad |A_{04}| \le I_Y J_Y,$$

where, by the previous boundedness argument on $[1/2, \overline{u}]$,

$$\begin{aligned}I_X &= \left(\int_{1/2}^{\overline{u}} \left|\frac{\partial}{\partial x} c\big(F^{-1}(u), G^{-1}(u)\big)\right| \frac{\sqrt{1-u}}{h_X(u)}\right) < +\infty, \\ I_Y &= \left(\int_{1/2}^{\overline{u}} \left|\frac{\partial}{\partial y} c\big(F^{-1}(u), G^{-1}(u)\big)\right| \frac{\sqrt{1-u}}{h_Y(u)}\right) < +\infty\end{aligned}$$

and by (CFG), (13), (14) and (16) on $[\overline{u}, 1]$,

$$\frac{J_X}{k_1} = \left(\int_{\overline{u}}^1 \rho' \circ F^{-1}(v) \frac{\sqrt{1-v}}{h_X(v)}\, dv\right) < +\infty, \qquad \frac{J_Y}{k_1} = \left(\int_{\overline{u}}^1 \rho' \circ F^{-1}(v) \frac{\sqrt{1-v}}{h_Y(v)}\, dv\right) < +\infty.$$

Therefore $\sigma_-^2(\overline{u}) = A_0 + A_{01} + A_{02} + A_{03} + A_{04} < +\infty$. In the same way the result holds for $\sigma_+^2(\overline{u})$. □

5.2.2. *Proof of the weak convergence*

*Step*1: *Extreme values.* In this first step we show that the contribution of extremes is negligible despite the rate $\sqrt{n}$. Without information on the joint distribution of extreme values we treat separately the upper tail of the integrals $W_c(\mathbb{F}_n, \mathbb{G}_n)$ and $W_c(F, G)$. Indeed the latter in not a centering of the former at the very end of tails so that the empirical quantile processes cannot help. Let K_n and k_n be positive increasing sequences, with K_n increasing and

$$K_n \to +\infty, \qquad \frac{K_n}{\log\log n} \to 0, \qquad k_n = \frac{\sqrt{n}}{K_n \exp(l \circ \psi_X^{-1}(\log n + K_n))}. \tag{22}$$

Under $(C2)$ and $(FG1)$ we have $l \circ \psi_X^{-1}(x) \to +\infty$ as $x \to +\infty$ thus $k_n = o(\sqrt{n}/K_n)$. Moreover, by (15) and (22) for any $\theta' \in (1, \theta)$ and all n large enough it holds

$$k_n \geq \frac{c}{K_n} \exp\left(-\frac{K_n}{2} + \theta \log(\log n + K_n)\right) > (\log n)^{\theta'}. \tag{23}$$

Hence we have $k_n / \log\log n \to +\infty$ and $k_n/\sqrt{n} \to 0$. Let us define

$$D_n = \int_{1-k_n/n}^1 c\big(F^{-1}(u), G^{-1}(u)\big)\, du,$$

$$S_n = \int_{1-k_n/n}^1 c\big(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)\big)\, du = \frac{1}{n} \sum_{i=n-[k_n]}^{n} c(X_{(i)}, Y_{(i)}).$$

Lemma 22.

1. *If* $(C2)$, $(FG1)$, $(FG4)$ *and* (CFG) *hold then* $\sqrt{n} D_n \to 0$.
2. *If* $(C2)$ *and* (CFG) *hold then* $\sqrt{n} S_n \to 0$ *in probability.*

Proof. 1. By C_2 and FG_4 we have

$$D_n = \int_{1-k_n/n}^1 \rho\big(F^{-1}(u) - G^{-1}(u)\big)\, du \leq \int_{1-k_n/n}^1 w(u)\, du$$

where

$$w(u) = \exp\big(l \circ F^{-1}(u)\big) = \exp\left(l \circ \psi_X^{-1}\left(\log\left(\frac{1}{1-u}\right)\right)\right).$$

Under (CFG), for $\theta > 1$ it holds, by (15),

$$l \circ F^{-1}(u) \leq \frac{1}{2} \log\left(\frac{1}{1-u}\right) - \theta \log\log\left(\frac{1}{1-u}\right) + K$$

thus, as $n \to +\infty$,

$$\int_{1-k_n/n}^1 w(u)\, du \leq \left[-\frac{K\sqrt{1-u}}{(\log(1/(1-u)))^\theta}\right]_{1-k_n/n}^1 = \frac{K\sqrt{k_n/n}}{(\log(n/k_n))^\theta} \to 0$$

so that $w(u)$ is integrable on $(\overline{u}, 1)$. By (CFG) $\varphi = \psi_X \circ l^{-1}$ satisfies

$$\big(\varphi^{-1}\big)'(x) = \frac{1}{\varphi' \circ \varphi^{-1}(x)} \leq \frac{1}{2 + 2\theta/\varphi^{-1}(x)}$$

and for x large enough,

$$\big(\varphi^{-1}\big)'(x) = \big(l \circ \psi_X^{-1}\big)'(x) \le \frac{1}{2 + 2\theta/(x/2 - \theta \log x + K)} < \frac{1}{2} - \frac{\theta}{x}. \tag{24}$$

We then have

$$\big(-(1-u)w(u)\big)' = w(u)\left(1 - \big(l \circ \psi_X^{-1}\big)'\left(\log\left(\frac{1}{1-u}\right)\right)\right) > \frac{w(u)}{2},$$

which gives

$$\int_{1-k_n/n}^{1} w(u)\,du \le 2\big[-(1-u)w(u)\big]_{1-k_n/n}^{1} \le \frac{2k_n}{n} w\left(1 - \frac{k_n}{n}\right),$$

since $\lim_{u\to 1}(1-u)w(u) = 0$. Recalling (22) it follows that for n large enough,

$$\sqrt{n} D_n \le \frac{2k_n}{\sqrt{n}} \exp\left(l \circ \psi_X^{-1}\left(\log\left(\frac{n}{k_n}\right)\right)\right) \le \frac{2}{K_n} \exp\left(l \circ \psi_X^{-1}\left(\log\left(\frac{n}{k_n}\right)\right) - l \circ \psi_X^{-1}(\log n + K_n)\right).$$

By (22), (23) and (15) with $\theta > 1$ we get

$$\log\left(\frac{n}{k_n}\right) \sim \frac{\log n}{2} + \log K_n + l \circ \psi_X^{-1}(\log n + K_n) \le \log n + \frac{K_n}{2} + \log K_n - \theta \log(\log n + K_n) + K$$

hence $\sqrt{n} D_n \le 2/K_n \to 0$ as $n \to +\infty$ since $l \circ \psi_X^{-1}$ is increasing.

2. Next we control S_n the stochastic sum of extreme values. Fix $\delta > 0$ and consider the events

$$A_n = \{\sqrt{n} S_n \ge 4\delta\}, \qquad B_{n,X} = \{X_{(n-[k_n])} > m\}, \qquad B_{n,Y} = \{Y_{(n-[k_n])} > m\}.$$

We have $\mathbb{P}(A_n) \le \mathbb{P}(A_n \cap B_{n,X} \cap B_{n,Y}) + \mathbb{P}(B_{n,X}^c) + \mathbb{P}(B_{n,X}^c)$. Since F and G are strictly increasing it obviously holds, for $\xi > 0$ and $u_0 = F(m + \xi)$, as $n \to +\infty$,

$$\mathbb{P}\big(B_{n,X}^c\big) = \mathbb{P}\left(\mathbb{F}_n^{-1}\left(1 - \frac{k_n}{n}\right) < m\right) \le \mathbb{P}\big(\mathbb{F}_n^{-1}(u_0) < F^{-1}(u_0) - \xi\big) \to 0$$

and likewise, $\mathbb{P}(B_{n,Y}^c) \to 0$. By (9) we can write, under $B_{n,X} \cap B_{n,Y}$,

$$\sqrt{n} S_n \le \frac{1}{\sqrt{n}} \sum_{i=n-[k_n]}^{n} \big(\rho(X_{(i)}) + \rho(Y_{(i)})\big) \le \frac{k_n + 1}{\sqrt{n}} \big(\rho(X_{(n)}) + \rho(Y_{(n)})\big),$$

hence $\mathbb{P}(A_n \cap B_{n,X} \cap B_{n,Y}) \le \mathbb{P}(C_{n,X}) + \mathbb{P}(C_{n,Y})$ where

$$C_{n,X} = \left\{\rho(X_{(n)}) \ge \delta \frac{\sqrt{n}}{k_n}\right\}, \qquad C_{n,Y} = \left\{\rho(Y_{(n)}) \ge \delta \frac{\sqrt{n}}{k_n}\right\}.$$

Now we have, by $(FG4)$ and since $X_1, \dots, X_n$ are independent,

$$\mathbb{P}(C_{n,Y}) \le \mathbb{P}(C_{n,X}) = 1 - \left(1 - \mathbb{P}\left(\rho(X) > \delta \frac{\sqrt{n}}{k_n}\right)\right)^n,$$

then combining $\rho^{-1}(x) = l^{-1}(\log x)$ with (22) gives,

$$\mathbb{P}\left(\rho(X) > \delta \frac{\sqrt{n}}{k_n}\right) = \exp\big(-\psi_X \circ l^{-1}\big(\log \delta + l \circ \psi_X^{-1}(\log n + \log K_n) + \log K_n\big)\big)$$

Now by (CFG) $\psi_X \circ l^{-1}$ is increasing. As soon as $\log K_n > |\log \delta|$ we get

$$\mathbb{P}\left(\rho(X) > \delta \frac{\sqrt{n}}{k_n}\right) \le \exp\big(-\psi_X \circ l^{-1}\big(l \circ \psi_X^{-1}(\log n + \log K_n)\big)\big) = \frac{1}{nK_n},$$

which yields $\mathbb{P}(C_{n,X}) \le 1 - \exp(-K/K_n) \to 0$. From

$$\mathbb{P}(A_n) \le \mathbb{P}(A_n \cap B_{n,X} \cap B_{n,Y}) + \mathbb{P}\big(B_{n,X}^c\big) + \mathbb{P}\big(B_{n,X}^c\big) \le \mathbb{P}(C_{n,X}) + \mathbb{P}(C_{n,X}) + \mathbb{P}\big(B_{n,X}^c\big) + \mathbb{P}\big(B_{n,X}^c\big)$$

we conclude that $\mathbb{P}(A_n) \to 0$. □

*Step*2: *Centered high order quantiles*. This section ends the part of the proof of Theorem 14 devoted to the secondary order. We split the arguments into the three lemmas below. Remind that k_n is defined at (22). Let introduce

$$h_n = n^{\beta}, \qquad \beta \in \left(\frac{1}{2}, 1\right), \qquad I_n = \left[1 - \frac{h_n}{n}, 1 - \frac{k_n}{n}\right], \tag{25}$$

and define the centered random integral of non extreme tail quantiles to be

$$T_n = \int_{1-h_n/n}^{1-k_n/n} \Big(c\big(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)\big) - c\big(F^{-1}(u), G^{-1}(u)\big)\Big)\, du.$$

Lemma 23. *If* $(C2)$, (FG) *and* (CFG) *hold then* $\sqrt{n}T_n \to 0$ *almost surely.*

The proof of this lemma is based on the two following lemmas whose proof are postponed to the appendix. In order to bound T_n we first evaluate the quantile empirical processes

$$\beta_n^X(u) = \sqrt{n}\big(\mathbb{F}_n^{-1}(u) - F^{-1}(u)\big), \qquad \beta_n^Y(u) = \sqrt{n}\big(\mathbb{G}_n^{-1}(u) - G^{-1}(u)\big). \tag{26}$$

Lemma 24. *Define* $\Delta_n = [\overline{u}, 1 - k_n/n]$. *Under* $(FG1)$ *and* $(FG2)$ *we have*

$$\limsup_{n\to+\infty} \sup_{u\in\Delta_n} \frac{|\beta_n(u)|h(u)}{\sqrt{(1-u)\log\log n}} \le 4 \quad a.s.$$

where $(\beta_n, h) = (\beta_n^X, h_X)$ *or* $(\beta_n, h) = (\beta_n^Y, h_Y)$.

In the next key lemma we have to carefully check that the conditions given at Proposition 30 are almost surely met on $I_n \subset \Delta_n$. For $u \in I_n$ and $n \ge 3$ define

$$\varepsilon_n(u) = \varepsilon_n^X(u) - \varepsilon_n^Y(u), \qquad \varepsilon_n^X(u) = \frac{\beta_n^X(u)}{\sqrt{n}}, \qquad \varepsilon_n^Y(u) = \frac{\beta_n^Y(u)}{\sqrt{n}}. \tag{27}$$

In the third lemma, the condition $(FG4)$ is crucial.

Lemma 25. *Assume that* $(C2)$, (FG) *and* (CFG) *hold. Then there exists* $K_2 > 0$ *such that*

$$\limsup_{n\to+\infty} \sup_{u\in I_n} \frac{|c(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)) - c(F^{-1}(u), G^{-1}(u))|}{\rho' \circ F^{-1}(u)\ |\varepsilon_n(u)|} \le K_2 \quad a.s.$$

Proof of Lemma 23. Remind notation from (22), (25) and (27). By Lemma 25 it holds, with probability one, for all n large enough

$$|T_n| \le K \int_{1-h_n/n}^{1-k_n/n} \rho' \circ F^{-1}(u)\big|\varepsilon_n(u)\big|\, du.$$

We proceed as in the proof of Lemma 20 where similar integrable functions show up, however they have now to be integrated to sharply evaluate $\sqrt{n}|T_n|$. From Lemma 24 it follows, with probability one, that for all n large and all $u \in I_n \subset \Delta_n$,

$$\big|\varepsilon_n(u)\big| \le \left|\frac{\beta_n^X(u)}{\sqrt{n}}\right| + \left|\frac{\beta_n^Y(u)}{\sqrt{n}}\right| \le 5\sqrt{\frac{\log\log n}{n}}\left(\frac{\sqrt{1-u}}{h_X(u)} + \frac{\sqrt{1-u}}{h_Y(u)}\right). \tag{28}$$

We then compute separately the following two integrals

$$\sqrt{n}|T_n| \leq 5K\sqrt{\log\log n}\left(\int_{1-h_n/n}^{1-k_n/n} t_X(u)\,du + \int_{1-h_n/n}^{1-k_n/n} t_Y(u)\,du\right), \tag{29}$$

where, for $Z = X, Y$ we write $t_Z(u) = \rho' \circ F^{-1}(u)\sqrt{1-u}/h_Z(u)$.

Consider the first integral in (29). Since ρ is convex by Proposition 31 we can use (20) as in the proof of Lemma 20 to justify the change of variable $u = F \circ \rho^{-1}(x)$ then apply (5) to $\rho^{-1}(x) = l^{-1}(\log x)$ and rewrite the first integral as

$$\begin{aligned}\int_{1-h_n/n}^{1-k_n/n} t_X(u)\,du &= \int_{1-h_n/n}^{1-k_n/n} \frac{\sqrt{1-u}}{h_{\rho(X)}(u)}\,du = \int_{b(n/h_n)}^{b(n/k_n)} \sqrt{\mathbb{P}\big(\rho(X) > x\big)}\,dx \\ &= \int_{b(n/h_n)}^{b(n/k_n)} \exp\left(-\frac{1}{2}\psi_X \circ l^{-1}(\log x)\right)dx\end{aligned}$$

where, by (*CFG*) reformulated into (15),

$$b(x) = \rho \circ F^{-1}\left(1 - \frac{1}{x}\right) = \exp\big(l \circ \psi_X^{-1}(\log x)\big) \leq \frac{K\sqrt{x}}{(\log x)^\theta}. \tag{30}$$

Equation (16) justifies that t_X is integrable since $\theta > 1$ and, by (13),

$$\exp\left(-\frac{1}{2}\psi_X \circ l^{-1}(\log x)\right) \leq \frac{K}{x(\log x)^\theta}.$$

Now observe that $\varphi = \psi_X \circ l^{-1}$ satisfies $\varphi' = (\psi_X'/l') \circ l^{-1}$ and (*CFG*) reads

$$\varphi'(x) \geq 2 + \frac{2\theta}{x}, \quad x > l(\tau_1),$$

so that we have, for all $x > b(n/h_n) > l(\tau_1)$,

$$\left(-x\exp\left(-\frac{1}{2}\varphi(\log x)\right)\right)' = \left(\frac{1}{2}\varphi'(\log x) - 1\right)\exp\left(-\frac{1}{2}\varphi(\log x)\right) \geq \frac{\theta}{\log x}\exp\left(-\frac{1}{2}\varphi(\log x)\right).$$

Therefore it holds, thanks to the upper bound (30) and since $b(x)$ is increasing,

$$\begin{aligned}\int_{1-h_n/n}^{1-k_n/n} t_X(u)\,du &\leq \frac{\log b(n/k_n)}{\theta}\int_{b(n/h_n)}^{b(n/k_n)} \frac{\theta}{\log x}\exp\left(-\frac{1}{2}\psi_X \circ l^{-1}(\log x)\right)dx \\ &\leq \frac{\log b(n)}{\theta}\left[-x\exp\left(-\frac{1}{2}\psi_X \circ l^{-1}(\log x)\right)\right]_{b(n/h_n)}^{b(n/k_n)} \\ &\leq \frac{K\log n}{\theta}\frac{b(n/h_n)}{\sqrt{n/h_n}} = \frac{K}{\theta(1-\beta)^\theta(\log n)^{\theta-1}}\end{aligned}$$

since $h_n = n^\beta$. This proves that

$$\lim_{n\to+\infty} \sqrt{\log\log n}\int_{1-h_n/n}^{1-k_n/n} t_X(u)\,du = 0.$$

We now turn to the second integral in (29),

$$J_n = \int_{1-h_n/n}^{1-k_n/n} t_Y(u)\,du = \int_{1-h_n/n}^{1-k_n/n} l' \circ F^{-1}(u)\frac{\sqrt{1-u}}{h_Y(u)}\rho \circ F^{-1}(u)\,du.$$

By (A.10) and (6), under ($C2$) we have $l'(x) = \varepsilon_1(x)l(x)/x$ with $\varepsilon_1(x) \to \gamma$ as $x \to +\infty$. If $\gamma = 0$ the rate of $\varepsilon_1(x)$ is given by (11) and we pick $\theta_1' \in (\theta_1, \theta - 1)$. If $\gamma > 0$ let $\theta_1' = 1$. Recall that $\phi^{-1} = F^{-1} \circ G = \psi_X^{-1} \circ \psi_Y$. Start with

$$J_n \leq (1+\gamma)\int_{1-h_n/n}^{1-k_n/n} \frac{(l \circ F^{-1}(u))^{\theta_1'}}{F^{-1}(u)}\frac{\sqrt{1-u}}{h_Y(u)}\rho \circ F^{-1}(u)\,du$$

$$= (1+\gamma)\int_{G^{-1}(1-h_n/n)}^{G^{-1}(1-k_n/n)} \frac{(l\circ\phi^{-1}(x))^{\theta_1'}}{\phi^{-1}(x)} \exp\big(l\circ\phi^{-1}(x)\big)\sqrt{\mathbb{P}(Y>x)}\,dx.$$

Observe that (CFG) and (15) imply

$$l\circ\phi^{-1}(x) = l\circ\psi_X^{-1}\circ\psi_Y(x) \le \frac{\psi_Y(x)}{2} - \theta\log\psi_Y(x) + K.$$

Since $\psi_X^{-1}\circ\psi_Y(x)\ge x$ by ($FG4$) and $\psi_Y'(x)\ge K/x$ by ($FG5$) it readily follows, for $\theta-\theta_1'>1$ and $K>0$,

$$\begin{aligned} J_n &\le (1+\gamma)\int_{G^{-1}(1-h_n/n)}^{G^{-1}(1-k_n/n)} \frac{(\psi_Y(x))^{\theta_1'-\theta}}{\psi_X^{-1}\circ\psi_Y(x)}\,dx \le K\int_{G^{-1}(1-h_n/n)}^{G^{-1}(1-k_n/n)} \frac{\psi_Y'(x)}{(\psi_Y(x))^{\theta-\theta_1'}}\,dx \\ &= K\left[\frac{-1}{(\psi_Y(x))^{\theta-\theta_1'-1}}\right]_{\psi_Y^{-1}(\log(n/h_n))}^{\psi_Y^{-1}(\log(n/k_n))} \le \frac{K}{((1-\beta)\log n)^{\theta-\theta_1'-1}} \end{aligned}$$

therefore

$$\lim_{n\to+\infty}\sqrt{\log\log n}\int_{1-h_n/n}^{1-k_n/n} t_Y(u)\,du = 0.$$

As a conclusion, the almost sure upper bound of $\sqrt{n}|T_n|$ tends to zero. □

Step 3: *Upper middle order quantiles*. At (25) we have defined $h_n=n^\beta$ with $\beta\in(1/2,1)$ to be chosen. Let us introduce

$$I_{M,n} = \left(F(M), 1-\frac{h_n}{n}\right), \quad M>m. \tag{31}$$

Since $F(M)>F(m)=\overline{u}$ and (A.2) in Section A.2 holds we have by ($C2$)

$$U_{M,n} = \int_{F(M)}^{1-h_n/n} \big(c\big(\mathbb{F}_n^{-1}(u),\mathbb{G}_n^{-1}(u)\big) - c\big(F^{-1}(u),G^{-1}(u)\big)\big)\,du = \int_{F(M)}^{1-h_n/n} \rho\big(\big|\tau(u)+\varepsilon_n(u)\big|\big) - \rho\big(\tau(u)\big)\,du,$$

where $\varepsilon_n(u)$ is as in (27). In order to control the last integral, we expand ρ and make use of a distribution free Brownian approximation of the joint quantile processes.

Lemma 26. *Assume* ($C2$), (FG) *and* (CFG). *For any* $\varepsilon>0$ *and* $\lambda>0$ *we can find* $M>m$ *such that*, *for all* n *large enough*, $\mathbb{P}(\sqrt{n}|U_{M,n}|>\lambda)<\varepsilon$.

Proof. 1. Under ($C2$) we have $l'(x)=\varepsilon_1(x)l(x)/x$ where $\varepsilon_1(x)\to\gamma$ as $x\to+\infty$ thus ε_1 is bounded on $(M,+\infty)$. Moreover, (CFG) ensures that $l\circ\psi_Y^{-1}(x)\le l\circ\psi_X^{-1}(x)<x$ whereas (14) and (7) entail that $\psi_X(x)>2l(x)\ge 2\log x$ thus

$$F^{-1}(u) = \psi_X^{-1}\left(\log\left(\frac{1}{1-u}\right)\right) < \frac{1}{\sqrt{1-u}}$$

for all $u\in I_{M,n}$ and $x\in F(I_{M,n})$. Under ($FG4$) we have $\tau(u)=F^{-1}(u)-G^{-1}(u)\ge\tau_0$ for $u\in I_{M,n}$. Hence by choosing $M>m$ and $K>0$ sufficiently large, (28) and ($FG3$) imply that it almost surely eventually holds

$$\begin{aligned} &\sup_{u\in I_{M,n}} \varepsilon_1\circ\tau(u)\frac{l\circ\tau(u)}{\tau(u)}\big|\varepsilon_n(u)\big| \\ &\quad\le K\sup_{u\in I_{M,n}} l\circ F^{-1}(u)\big(H_X(u)F^{-1}(u)+H_Y(u)G^{-1}(u)\big)\sqrt{\frac{\log\log n}{n(1-u)}} \\ &\quad\le K\sqrt{\log\log n}\sup_{u\in I_{M,n}} \frac{l\circ\psi_X^{-1}(\log(1/(1-u)))}{\sqrt{n(1-u)}}F^{-1}(u) \end{aligned}$$

$$\leq K\sqrt{n\log\log n}\sup_{u\in I_{M,n}}\frac{\log(1/(1-u)}{n(1-u)}$$

$$\leq K\frac{\log n}{h_n}\sqrt{n\log\log n}$$

which vanishes since $\beta > 1/2$ in (25). We have shown that

$$\lim_{n\to+\infty}\sup_{u\in I_{M,n}}\left|\varepsilon_n(u)\right|l'\circ\tau(u)=0 \quad \text{a.s.} \tag{32}$$

2. By (32), the second part of Proposition 30 can be applied for all large n. It says that

$$\rho\big(\left|\tau(u)+\varepsilon_n(u)\right|\big)-\rho\big(\tau(u)\big)=k_0\big(\tau(u),\varepsilon_n(u)\big)\rho'\circ\tau(u)\varepsilon_n(u)$$

where, by (A.12),

$$\lim_{\delta_0\to 0}\sup_{\tau(u)>\tau_0}\sup_{|\varepsilon_n(u)|l'\circ\tau(u)\leq\delta_0}\left|k_0\big(\tau(u),\varepsilon_n(u)\big)-1\right|=0$$

which can be reformulated through (32) into $k_1(u)=k_0(\tau(u),\varepsilon_n(u))$ and

$$\lim_{n\to+\infty}\sup_{u\in I_{M,n}}\left|k_1(u)-1\right|=0 \quad \text{a.s.} \tag{33}$$

Thus, given any $\vartheta\in(0,1)$ the random function $k_1(u)$ is such that $k_1(u)\in(1-\vartheta,1+\vartheta)$ for all $u\in I_{M,n}$ and

$$\sqrt{n}U_{M,n}=\int_{F(M)}^{1-h_n/n}k_1(u)\rho'\circ\tau(u)\big(\beta_n^X(u)+\beta_n^Y(u)\big)\,du.$$

From now on we make use of notation introduced at Section A.3 and we work on the probability space of Theorem 29. This allows us to write

$$\sqrt{n}U_{M,n}=\int_{F(M)}^{1-h_n/n}k_1(u)\rho'\circ\tau(u)\left(\frac{B_n^X(u)+Z_n^X(u)}{h_X(u)}+\frac{B_n^Y(u)+Z_n^Y(u)}{h_Y(u)}\right)du$$

where $(U_{M,n},B_n^X,Z_n^X,B_n^Y,Z_n^Y,k_1)$ are built together on Ω^* in such a way that for some small $\xi>0$ independent of the distribution P_H,

$$\lim_{n\to+\infty}n^\xi\sup_{u\in I_{M,n}}\left|Z_n^X(u)\right|=\lim_{n\to+\infty}n^\xi\sup_{u\in I_{M,n}}\left|Z_n^Y(u)\right|=0 \quad \text{a.s.} \tag{34}$$

and B_n^X, B_n^Y are Brownian bridges defined at (A.7). Let $\sqrt{n}U_{M,n}=N_{M,n}+R_{M,n}+S_{M,n}$ with

$$N_{M,n}=\int_{F(M)}^{1-h_n/n}\rho'\circ\tau(u)\left(\frac{B_n^X(u)}{h_X(u)}+\frac{B_n^Y(u)}{h_Y(u)}\right)du,$$
$$R_{M,n}=\int_{F(M)}^{1-h_n/n}k_1(u)\rho'\circ\tau(u)\left(\frac{Z_n^X(u)}{h_X(u)}+\frac{Z_n^Y(u)}{h_Y(u)}\right)du,$$
$$S_{M,n}=\int_{F(M)}^{1-h_n/n}\big(k_1(u)-1\big)\rho'\circ\tau(u)\left(\frac{B_n^X(u)}{h_X(u)}+\frac{B_n^Y(u)}{h_Y(u)}\right)du.$$

3. We first deal with $R_{M,n}$. Since $\rho'(x)$ is increasing by Proposition 31, $(C2)$ implies $l'(x)<Kl(x)/x$ with $K>\gamma$ and (CFG) entails $l\circ\psi_Y^{-1}(x)\leq l\circ\psi_X^{-1}(x)\leq x/2-\theta\log x$ by (13) we readily have

$$\left|\int_{F(M)}^{1-h_n/n}\frac{\rho'\circ\tau(u)}{h_X(u)}Z_n^X(u)\,du\right|\leq\frac{K}{n^\xi}\int_{F(M)}^{1-h_n/n}\frac{l\circ\psi_X^{-1}(\log(1/(1-u))}{F^{-1}(u)h_X(u)}\exp(l\circ\psi_X^{-1}\big(\log\big(1/(1-u)\big)\big)\,du$$
$$\leq\frac{K}{n^\xi}\int_{F(M)}^{1-h_n/n}\frac{\log(1/(1-u))}{(\log(1/(1-u)))^\theta}\frac{H_X(u)}{(1-u)^{3/2}}\,du$$

which is, by using $(FG3)$ and $\theta > 1$ then choosing $\beta \in (1-\xi, 1)$, less than

$$\frac{K}{n^{\xi}} \int_{F(M)}^{1-h_n/n} \frac{1}{(1-u)^{3/2}}\, du < K n^{-\xi/2}.$$

The same bound holds for h_Y since $F^{-1} > G^{-1}$ and

$$\left| \int_{F(M)}^{1-h_n/n} \frac{\rho' \circ \tau(u)}{h_Y(u)} Z_n^Y(u)\, du \right| \leq \frac{K}{n^{\xi}} \int_{F(M)}^{1-h_n/n} \frac{G^{-1}(u) \log(1/(1-u)) H_Y(u)}{F^{-1}(u)(\log(1/(1-u)))^{\theta}(1-u)^{3/2}}\, du.$$

By (33), (34) and the above bounds we have almost surely for n large enough $|R_{M,n}| \leq 2Kn^{-\xi/2} \to 0$.

4. As $N_{M,n}$ is the sum of two linear functionals of Brownian bridges it is a mean zero Gaussian random variable with variance

$$\sigma^2(M,n) = \int_{F(M)}^{1-h_n/n} \int_{F(M)}^{1-h_n/n} \rho' \circ \tau(u) \rho' \circ \tau(v) \Xi(u,v)\, du\, dv$$

where

$$\begin{aligned} \Xi(u,v) &= \operatorname{cov}\left(\frac{B_n^X(u)}{h_X(u)} + \frac{B_n^Y(u)}{h_Y(u)}, \frac{B_n^X(v)}{h_X(v)} + \frac{B_n^Y(v)}{h_Y(v)} \right) \\ &= \frac{\min(u,v) - uv}{h_X(u)h_X(v)} + \frac{\Pi(v,u) - uv}{h_X(v)h_Y(u)} + \frac{\Pi(u,v) - uv}{h_X(u)h_Y(v)} + \frac{\min(u,v) - uv}{h_Y(v)h_Y(u)}. \end{aligned}$$

Therefore by Lemma 20 taken in $\overline{u} = F(M)$ we see that $\sigma^2(M,n) \to \sigma^2(M)$ as $n \to \infty$ and $\sigma^2(M) \to 0$ as $M \to +\infty$. On the other hand,

$$|S_{M,n}| \leq \sup_{u \in I_{M,n}} \left| k_1(u) - 1 \right| \int_{F(M)}^{1-h_n/n} \rho' \circ \tau(u) \left| \frac{B_n^X(u)}{h_X(u)} + \frac{B_n^Y(u)}{h_Y(u)} \right| du,$$

which tends to 0 in probability when $n \to +\infty$ since the integral is bounded in probability.

5. As a conclusion, for any $\varepsilon > 0$ and $\lambda > 0$ we can find $M = M(\varepsilon, \lambda) > m$ such that

$$\mathbb{P}\left(\sqrt{n}|U_{M,n}| > \lambda\right) \leq \mathbb{P}\left(|N_{M,n}| > \frac{\lambda}{3}\right) + \mathbb{P}\left(|R_{M,n}| > \frac{\lambda}{3}\right) + \mathbb{P}\left(|S_{M,n}| > \frac{\lambda}{3}\right) \leq \frac{\sigma^2(M,n)}{(\lambda/3)^2} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} < \varepsilon,$$

for all $n > n(\varepsilon, \lambda, M)$. □

Step 4: *Centered middle order quantiles.* For $M > m$ define $I_M = (F(-M), F(M))$ and consider the centered random integral

$$\mathbb{M}_{M,n} = \int_{F(-M)}^{F(M)} \left(c\left(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)\right) - c\left(F^{-1}(u), G^{-1}(u)\right) \right) du.$$

In order to conclude the proof of Theorem 14 it remains to exploit the Brownian approximation of the joint quantile processes β_n^X and β_n^Y defined at (26) to accurately approximate $\sqrt{n}\mathbb{M}_{M,n}$. Recalling (18) write

$$\nabla_x(u) = \frac{\partial}{\partial x} c\left(F^{-1}(u), G^{-1}(u)\right), \qquad \nabla_y(u) = \frac{\partial}{\partial y} c\left(F^{-1}(u), G^{-1}(u)\right)$$

and

$$\sqrt{n}\mathbb{N}_{M,n} = \int_{F(-M)}^{F(M)} \left(\nabla_x(u) \beta_n^X(u) + \nabla_y(u) \beta_n^Y(u) \right) du.$$

In the following lemma the condition $(C3)$ is essential, which excludes the distance W_1.

Lemma 27. *Assume* (C), (FG) *and* (CFG). *Then for any* $\delta > 0$, *any* $\varepsilon > 0$ *and any* $M > m' > m$ *there exists* $n(\varepsilon, \delta, M)$ *such that for all* $n > n(\varepsilon, \delta, M)$,

$$\mathbb{P}\big(|\sqrt{n}\mathbb{M}_{M,n} - \sqrt{n}\mathbb{N}_{M,n}| > \varepsilon\big) \le \delta.$$

Proof. 1. Under $(FG1)$, h_X and h_Y are away from 0 on I_M and we write

$$\eta_M = \min\Big(\inf_{u\in I_M} h_X(u), \inf_{u\in I_M} h_Y(u)\Big) > 0.$$

We keep working on the probability space of Theorem 29. In particular, since $I_M \subset \mathcal{I}_n$ we can apply again Theorem 29 and get the analogue of (34)

$$\mathbb{P}\bigg(\sup_{u\in I_M}\bigg|\frac{Z_n^X(u)}{h_X(u)}\bigg| > \frac{1}{n^{\xi}}\bigg) = o(1), \qquad \mathbb{P}\bigg(\sup_{u\in I_M}\bigg|\frac{Z_n^Y(u)}{h_Y(u)}\bigg| > \frac{1}{n^{\xi}}\bigg) = o(1). \tag{35}$$

Introduce the event

$$A_n(M, C) = \bigg\{\sup_{u\in I_M}\big|\mathbb{F}_n^{-1}(u) - F^{-1}(u)\big| + \big|\mathbb{G}_n^{-1}(u) - G^{-1}(u)\big| \le \frac{4C}{\sqrt{n}}\bigg\}.$$

By (35), for any $\delta > 0$ one can find $C_\delta > 0$ so large that, for all n large enough,

$$\begin{aligned}
&\mathbb{P}\big(A_n(M, C_\delta)^c\big) \\
&\quad = \mathbb{P}\Big(\sup_{u\in I_M} \sqrt{n}\big|\mathbb{F}_n^{-1}(u) - F^{-1}(u)\big| + \sqrt{n}\big|\mathbb{G}_n^{-1}(u) - G^{-1}(u)\big| > 4C_\delta\Big) \\
&\quad \le \mathbb{P}\bigg(\sup_{u\in I_M}\bigg|\frac{B_n^X(u)}{h_X(u)}\bigg| > C_\delta\bigg) + \mathbb{P}\bigg(\sup_{u\in I_M}\bigg|\frac{B_n^Y(u)}{h_Y(u)}\bigg| > C_\delta\bigg) + o(1) \\
&\quad \le 2\mathbb{P}\Big(\sup_{u\in I_M}\big|B(u)\big| > \eta_M C_\delta\Big) + \frac{\delta}{2} \le \delta,
\end{aligned}$$

where B denotes a standard Brownian bridge.

2. Since $F \neq G$ and F, G are continuous, for any $\tau_1 \in (0, \tau_0)$ there exists an open interval $I(\tau_1) \subset I_M$ such that $|\tau(u)| > \tau_1$ for $u \in I(\tau_1)$, provided that $m > 0$ is chosen large enough. By taking $M > m$, by $(FG4)$ we further have $\tau_M = \sup_{u\in I_M} |\tau(u)| \ge \tau_0 > \tau_1$. Thus

$$D_M^+(\tau_1) = \big\{u : \tau_1 < \big|\tau(u)\big| \le \tau_M\big\} \cap I_M, D_M^-(\tau_1) = \big\{u : \big|\tau(u)\big| \le \tau_1\big\} \cap I_M,$$

are such that $I(\tau_1) \subset D_M^+(\tau_1) \neq \varnothing$ and $D_M^-(\tau_1) \subset I_M$ is possibly empty, and $D_M^+(\tau_1) \cup D_M^-(\tau_1) = I_M$. By $(C3)$, for any (x, y), $(x', y') \in D_m(\tau)$,

$$\big|c\big(x', y'\big) - c(x, y)\big| \le d(m, \tau)\big(\big|x' - x\big| + \big|y' - y\big|\big)$$

with $d(m, \tau) \to 0$ as $\tau \to 0$ and m is fixed. Observe that $u \in D_M^-(\tau_1) = D_m^-(\tau_1)$ if, and only if, $(F^{-1}(u), G^{-1}(u)) \in D_m(\tau_1)$. Let $\tau_1' \in (\tau_1, \tau_0)$ and $m' \in (m, M)$.

Now, given M and C_δ, if $A_n(M, C_\delta)$ is true for a large enough n then $(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)) \in D_{m'}(\tau_1')$ whenever $(F^{-1}(u), G^{-1}(u)) \in D_m(\tau_1) \subset D_{m'}(\tau_1')$ and $u \in I_M$. Thus, under the event $A_n(M, C_\delta)$ it holds

$$\begin{aligned}
\sqrt{n}\mathbb{M}_{M,n}^-(\tau_1) &:= \sqrt{n}\int_{u\in D_M^-(\tau_1)} \big|c\big(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)\big) - c\big(F^{-1}(u), G^{-1}(u)\big)\big|\,du \\
&\le \sqrt{n}\int_{u\in D_M^-(\tau_1)} d\big(m', \tau_1'\big)\big(\big|\mathbb{F}_n^{-1}(u) - F^{-1}(u)\big| + \big|\mathbb{G}_n^{-1}(u) - G^{-1}(u)\big|\big)\,du \\
&\le 4C_\delta\, d\big(m', \tau_1'\big).
\end{aligned}$$

3. The main term is

$$\sqrt{n}\mathbb{M}^+_{M,n}(\tau_1) := \sqrt{n}\int_{u\in D^+_M(\tau_1)} \big(c\big(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)\big) - c\big(F^{-1}(u), G^{-1}(u)\big)\big)\, du.$$

Under the event $A_n(M, C_\delta)$ the Taylor expansion of $c(F^{-1}(u), G^{-1}(u))$ is justified on $D^+_M(\tau_1)$, that is away from the diagonal. As a matter of fact, under $(C1)$ we have, for x, y in $(-M, M)$ such that $|x-y| \geq \tau$,

$$\big|c(x+\varepsilon_x, y+\varepsilon_y) - c(x,y) - \nabla_x(x,y)\varepsilon_x - \nabla_y(x,y)\varepsilon_y\big| \leq \lambda(M,\tau)\Theta\big(|\varepsilon_x| + |\varepsilon_y|\big),$$

where $\Theta(s)/s \to 0$ as $s \to 0$ for M and τ_1 fixed. Then the expansion of $c(F^{-1}(u), G^{-1}(u))$ on $u \in D^+_M(\tau_1)$ can be written as

$$c\big(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)\big) - c(F^{-1}(u), G^{-1}(u) = \big(\nabla_x(u)\beta_n^X(u) + \nabla_y(u)\beta_n^Y(u)\big) + \mathcal{R}_n(u).$$

We have

$$\begin{aligned}
&\left|\sqrt{n}\mathbb{M}^+_{M,n}(\tau_1) - \int_{u\in D^+_M(\tau_1)} \big(\nabla_x(u)\beta_n^X(u) + \nabla_y(u)\beta_n^Y(u)\big)\, du\right| \\
&\quad\leq \sqrt{n}\left|\int_{u\in D^+_M(\tau_1)} \mathcal{R}_n(u)\, du\right| \\
&\quad\leq \lambda(M,\tau_1)\sqrt{n}\Theta\left(\frac{1}{\sqrt{n}}\sup_{u\in I_M}\big|\beta_n^X(u)\big| + \big|\beta_n^Y(u)\big|\right).
\end{aligned}$$

As $|\mathbb{M}_{M,n} - \mathbb{M}^+_{M,n}(\tau_1)| \leq \mathbb{M}^-_{M,n}(\tau_1)$, whenever $A_n(M, C_\delta)$ is true, it holds

$$\begin{aligned}
&\left|\sqrt{n}\mathbb{M}_{M,n} - \int_{u\in D^+_M(\tau_1)} \big(\nabla_x(u)\beta_n^X(u) + \nabla_y(u)\beta_n^Y(u)\big)\, du\right| \\
&\quad\leq \sqrt{n}\mathbb{M}^-_{M,n}(\tau_1) + \lambda(M,\tau_1)\sqrt{n}\Theta\left(\frac{4C_\delta}{\sqrt{n}}\right),
\end{aligned}$$

where $\sqrt{n}\Theta(4C_\delta\sqrt{n}) \to 0$ as $n \to +\infty$. We also have $D^-_M(\tau_1) = I_M \setminus D^+_M(\tau_1) \subset I_M$ and ∇_x, ∇_y are bounded on I_M thus

$$\left|\int_{u\in D^-_M(\tau_1)} \big(\nabla_x(u)\beta_n^X(u) + \nabla_y(u)\beta_n^Y(u)\big)\, du\right| \leq 2m\frac{4C_\delta}{\sqrt{n}}\sup_{u\in I_M}\big|\nabla_x(u)\big| + \big|\nabla_y(u)\big|.$$

Hence under $A_n(M, C_\delta)$ $|\sqrt{n}\mathbb{M}_{M,n} - \sqrt{n}\mathbb{N}_{M,n}|$ is bounded by

$$4C_\delta\, d\big(m', \tau_1'\big) + \lambda(M,\tau_1)\sqrt{n}\Theta\left(\frac{4C_\delta}{\sqrt{n}}\right) + 2m\frac{4C_\delta}{\sqrt{n}}\sup_{u\in I_M}\big|\nabla_x(u)\big| + \big|\nabla_y(u)\big|.$$

Therefore, for any $\delta > 0$, any $\varepsilon > 0$ and any triplet $M > m' > m$ we can choose τ_1 and $\tau_1' > \tau_1$ so small that $4C_\delta d(m', \tau_1') \leq \varepsilon/2$. Then there exists $n(\varepsilon, \delta, M)$ such that for all $n > n(\varepsilon, \delta, M)$,

$$\mathbb{P}\big(|\sqrt{n}\mathbb{M}_{M,n} - \sqrt{n}\mathbb{N}_{M,n}| > \varepsilon\big) \leq \mathbb{P}\big(A_n(M, C_\delta)^c\big) \leq \delta. \qquad \square$$

Step 5: *Conclusion.* Now recall that $\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F,G)) = \sqrt{n}D_n + \sqrt{n}S_n + \sqrt{n}T_n + \sqrt{n}U_{M,n} + \sqrt{n}\mathbb{M}_{M,n}$. By Steps 1 and 2, $\sqrt{n}D_n + \sqrt{n}S_n + \sqrt{n}T_n$ converges to zero in probability. Hence, we only need to prove the weak convergence of $\sqrt{n}U_{M,n} + \sqrt{n}\mathbb{M}_{M,n}$. Let $\mathbb{X}_\infty$ be a centered Gaussian random variable with variance $\sigma^2(H,c)$. For any B-bounded r-Lipschitz function Φ, we have

$$\begin{aligned}
&\mathbb{E}\big|\Phi\big(\sqrt{n}(U_{M,n} + \mathbb{M}_{M,n})\big) - \Phi(\mathbb{X}_\infty)\big| \\
&\quad\leq \mathbb{E}\big|\Phi\big(\sqrt{n}(U_{M,n} + \mathbb{M}_{M,n})\big) - \Phi(\sqrt{n}\mathbb{M}_{M,n})\big| + \mathbb{E}\big|\Phi(\sqrt{n}\mathbb{M}_{M,n}) - \Phi(\mathbb{X}_\infty)\big|
\end{aligned}$$

Dealing with the first right hand term we have

$$\begin{aligned}
&\mathbb{E}\big|\Phi\big(\sqrt{n}(U_{M,n}+\mathbb{M}_n)\big)-\Phi(\sqrt{n}\mathbb{M}_{M,n})\big| \\
&\quad =\mathbb{E}\big(\big|\Phi\big(\sqrt{n}(U_{M,n}+\mathbb{M}_{M,n})\big)-\Phi(\sqrt{n}\mathbb{M}_{M,n})\big|\mathbb{1}_{|\sqrt{n}U_{M,n}|>\lambda}\big) \\
&\qquad +\mathbb{E}\big(\big|\Phi\big(\sqrt{n}(U_{M,n}+\mathbb{M}_{M,n})\big)-\Phi(\sqrt{n}\mathbb{M}_{M,n})\big|\mathbb{1}_{|\sqrt{n}U_{M,n}|\le\lambda}\big) \\
&\quad \le 2B\mathbb{P}\big(|\sqrt{n}U_{M,n}|>\lambda\big)r\lambda
\end{aligned}$$

By Lemma 26 we can make $2B\mathbb{P}(|\sqrt{n}U_{M,n}|>\lambda)r\lambda$ as small as we want by choosing λ small enough and M large enough. We now consider the second right hand term

$$\mathbb{E}\big|\Phi(\sqrt{n}\mathbb{M}_{M,n})-\Phi(\mathbb{X}_\infty)\big|\le\mathbb{E}\big|\Phi(\sqrt{n}\mathbb{M}_{M,n})-\Phi(\sqrt{n}\mathbb{N}_{M,n})\big|+\mathbb{E}\big|\Phi(\sqrt{n}\mathbb{N}_{M,n}n)-\Phi(\mathbb{X}_\infty)\big|$$

By Lemma 27 the term $\mathbb{E}|\Phi(\sqrt{n}\mathbb{M}_{M,n})-\Phi(\sqrt{n}\mathbb{N}_{M,n})|$ can be made as small as desired. As $\sqrt{n}\mathbb{N}_{M,n}$ is a Gaussian random variable with variance

$$\sigma^2(M,H,c)=\int_{F(-M)}^{F(M)}\int_{F(-M)}^{F(M)}\nabla(u)\Sigma(u,v)\nabla(v)\,du\,dv$$

that converges to $\sigma^2(H,c)$, the term $\mathbb{E}|\Phi(\sqrt{n}\mathbb{N}_{M,n})-\Phi(\mathbb{X}_\infty)|$ is small enough for large enough M. This achieves the proof of Theorem 14.

Appendix

A.1. *Proof of Lemma* 24

Remind that $\Delta_n=[\overline{u},1-k_n/n]$ where $k_n/\log\log n\to+\infty$ and $k_n/n\to 0$ comes from (22) and (23). Let us study $(\beta_n,h)=(\beta_n^X,h_X)$ in Lemma 24. Under $(FG1)$ we have $f>0$ on $\mathbb{R}$ thus the random variables $U_i=F(X_i)$ are independent, uniformly distributed on [0, 1] and such that $X_{(i)}=F^{-1}(U_{(i)})$. Let $\mathbb{F}_{U,n}$ and $\mathbb{F}_{U,n}^{-1}$ denote the empirical cdf and quantile functions associated to $U_1,\dots,U_n$ so that $\mathbb{F}_n=\mathbb{F}_{U,n}\circ F$ and $\mathbb{F}_n^{-1}=F^{-1}\circ\mathbb{F}_{U,n}^{-1}$. Write $q_n(u)=\mathbb{F}_{U,n}^{-1}(u)-u$. By [6] we have

$$\limsup_{n\to\infty}\sup_{u\in\Delta_n}\frac{\sqrt{n}q_n(u)}{\sqrt{(1-u)\log\log n}}\le 4\quad\text{a.s.}\tag{A.1}$$

Since $(FG1)$ ensures that h_X is $\mathcal{C}_1$ on Δ_n the following expansion almost surely asymptotically holds,

$$\begin{aligned}
&\sup_{u\in\Delta_n}\big|\big(F^{-1}\big(u+q_n(u)\big)-F^{-1}(u)\big)h_X(u)-q_n(u)\big| \\
&\quad =\sup_{u\in\Delta_n}\left|\left(\frac{q_n(u)}{h_X(u)}+\frac{q_n^2(u)}{2}\left(\frac{1}{h_X(u)}\right)'_{u=u^*}\right)h_X(u)-q_n(u)\right| \\
&\quad \le A_nB_n
\end{aligned}$$

where $|u-u^*|\le|q_n(u)|$ and, by (A.1),

$$A_n=\sup_{u\in\Delta_n}\frac{q_n^2(u)}{2(1-u)}\le K\frac{\log\log n}{n}$$

whereas, by $(FG2)$,

$$\begin{aligned}
B_n&=\sup_{u\in\Delta_n}(1-u)h_X(u)\left|\left(\frac{1}{h_X(u)}\right)'_{u=u^*}\right| \\
&\le\sup_{u\in\Delta_n}\big(1-u^*\big)h_X\big(u^*\big)\left|\left(\frac{1}{h_X(u)}\right)'_{u=u^*}\right|\sup_{u\in\Delta_n}\frac{1-u}{1-u^*}\frac{h_X(u)}{h_X(u^*)} \\
&\le K\sup_{u\in\Delta_n}\frac{1-u}{1-u^*}\sup_{u\in\Delta_n}\frac{h_X(u)}{h_X(u^*)}.
\end{aligned}$$

Now, (A.1) shows that the random sequence

$$\sup_{u\in\Delta_n}\left|\frac{1-u^*}{1-u}-1\right| \le \sup_{u\in\Delta_n}\frac{1}{\sqrt{1-u}}\sup_{u\in\Delta_n}\left|\frac{q_n(u)}{\sqrt{1-u}}\right| \le 5\sqrt{\frac{n}{k_n}}\sqrt{\frac{\log\log n}{n}}$$

almost surely tends to 0. Moreover $(FG2)$ implies that

$$\left|\left(\log h_X(u)\right)'\right| \le K\left(\log\frac{1}{1-u}\right)'$$

so that $|\log h_X(u_2)-\log h_X(u_1)| \le K(\log(1-u_1)-\log(1-u_2))$ for any $u_1<u_2$ in Δ_n. Therefore, the random sequence

$$\sup_{u\in\Delta_n}\frac{h_X(u)}{h_X(u^*)} \le \sup_{u\in\Delta_n}\max\left(\frac{1-u^*}{1-u},\frac{1-u}{1-u^*}\right)^K$$

almost surely tends to 1. We have shown that it almost surely ultimately holds

$$\sup_{u\in\Delta_n}\left|\frac{\beta_n^X(u)h_X(u)-\sqrt{n}q_n(u)}{\sqrt{(1-u)\log\log n}}\right| \le A_nB_n\sqrt{\frac{n}{\log\log n}}\sup_{u\in\Delta_n}\frac{1}{\sqrt{1-u}} \le 10K\sqrt{\frac{\log\log n}{k_n}},$$

which proves Lemma 24, by (A.1) again.

A.2. *Proof of Lemma 25*

In this proof the condition $(FG4)$ allows to make use of Proposition 30. In view of (23) and (25) we eventually have $I_n\subset\Delta_n$. Hence Lemma 24 and $(FG3)$ imply that, almost surely, for all n large

$$\begin{aligned}\sup_{u\in I_n}\frac{|\mathbb{F}_n^{-1}(u)-F^{-1}(u)|}{F^{-1}(u)} &\le 2K_0\sup_{u\in I_n}\frac{\sqrt{1-u}}{F^{-1}(u)h_X(u)}\sqrt{\frac{\log\log n}{n}}\\ &= 2K_0\sup_{u\in I_n}H_X(u)\sqrt{\frac{\log\log n}{n(1-u)}} \le K\sqrt{\frac{\log\log n}{k_n}}.\end{aligned}$$

The same bound holds for $|\mathbb{G}_n^{-1}(u)-G^{-1}(u)|/G^{-1}(u)$. By (23) we then get

$$\lim_{n\to+\infty}\sup_{u\in I_n}\frac{|\varepsilon_n^Y(u)|}{F^{-1}(u)} \le \lim_{n\to+\infty}\sup_{u\in I_n}\frac{|\varepsilon_n^Y(u)|}{G^{-1}(u)} = \lim_{n\to+\infty}\sup_{u\in I_n}\frac{|\varepsilon_n^X(u)|}{F^{-1}(u)} = 0 \quad \text{a.s.}$$

so that $\sup_{u\in I_n}|\varepsilon_n(u)|/F^{-1}(u)$ almost surely vanishes. Under $(FG1)$ the law of large numbers for $\mathbb{F}_n$ and $\mathbb{G}_n$ readily implies

$$\lim_{n\to+\infty}\mathbb{F}_n^{-1}\left(1-\frac{h_n}{n}\right) = \lim_{n\to+\infty}\mathbb{G}_n^{-1}\left(1-\frac{h_n}{n}\right) = +\infty \quad \text{a.s.}$$

Therefore for any $q_0>0$, all n large enough and all $u\in I_n$, it holds

$$\min\left(\mathbb{F}_n^{-1}(u),F^{-1}(u),\mathbb{G}_n^{-1}(u),G^{-1}(u)\right)>m, \qquad \left|\varepsilon_n(u)\right|<q_0F^{-1}(u) \tag{A.2}$$

which implies, by $(C2)$ and for $\tau(u)=F^{-1}(u)-G^{-1}(u)$,

$$c\left(\mathbb{F}_n^{-1}(u),\mathbb{G}_n^{-1}(u)\right)-c\left(F^{-1}(u),G^{-1}(u)\right)=\rho\left(\left|\tau(u)+\varepsilon_n(u)\right|\right)-\rho\left(\tau(u)\right).$$

Case 1. Assume that $\gamma=0$ in $(C2)$. By Proposition 31 ρ' is increasing and

$$\left|\rho\left(\left|\tau(u)+\varepsilon_n(u)\right|\right)-\rho\left(\tau(u)\right)\right| \le \rho'\left(\tau(u)+\left|\varepsilon_n(u)\right|\right)\left|\varepsilon_n(u)\right|.$$

Observe that if

$$\liminf_{u\to1}\frac{G^{-1}(u)}{F^{-1}(u)}=q_1>0$$

then the result follows with $K_2 = 1$ since by taking $0 < q_0 < q_1 \le 1$ in (A.2) we ultimately have, with probability one,

$$\rho'\big(\tau(u) + |\varepsilon_n(u)|\big) = \rho'\left(F^{-1}(u)\left(1 - \frac{G^{-1}(u)}{F^{-1}(u)} + \frac{|\varepsilon_n(u)|}{F^{-1}(u)}\right)\right) \le \rho'\big(F^{-1}(u)\big).$$

If $q_1 = 0$, let us control $\rho'(\tau(u) + |\varepsilon_n(u)|) \le \rho'(F^{-1}(u)(1 + |\varepsilon_n(u)|/F^{-1}(u)))$. Remind that l is increasing and slowly varying whereas l' is decreasing, by (6) and (7). As a consequence, for $y > x$, $x \to +\infty$, $y \sim x$ we have $l(x) \le l(y) \le l(2x) \sim l(x)$ and

$$\frac{\rho'(y)}{\rho'(x)} = \frac{l'(y)}{l'(x)}\frac{\rho(y)}{\rho(x)} \le \frac{\rho(y)}{\rho(x)} = \exp\big(l(y) - l(x)\big) \le \exp\big(l'(x)(y - x)\big).$$

Therefore, by (6), (11) and (*FG*3), taking $\theta_1' \in (\theta_1, \theta - 1)$ yields

$$\begin{aligned}
1 &\le \frac{1}{\rho' \circ F^{-1}(u)}\rho'\left(F^{-1}(u)\left(1 + \frac{|\varepsilon_n(u)|}{F^{-1}(u)}\right)\right) \\
&\le \exp\big(l' \circ F^{-1}(u)|\varepsilon_n(u)|\big) = \exp\left(\varepsilon_1 \circ F^{-1}(u) l \circ F^{-1}(u)\frac{|\varepsilon_n(u)|}{F^{-1}(u)}\right) \\
&\le \exp\left(\left(l \circ F^{-1}\left(1 - \frac{k_n}{n}\right)\right)^{\theta_1'} K\sqrt{\frac{\log\log n}{k_n}}\right)
\end{aligned}$$

provided n is large enough and $u \in I_n$. Moreover (13) implies

$$l \circ F^{-1}\left(1 - \frac{k_n}{n}\right) = l \circ \psi_X^{-1}\left(\log\left(\frac{n}{k_n}\right)\right) \le l \circ \psi_X^{-1}(\log n) \le \log n. \tag{A.3}$$

By choosing θ' in (23) such that $\theta > \theta' > 1 + \theta_1' \ge \max(1, 2\theta_1')$ we get

$$\lim_{n\to+\infty} \sup_{u \in I_n} \frac{\rho'(\tau(u) + |\varepsilon_n(u)|)}{\rho' \circ F^{-1}(u)} \le 1 \quad \text{a.s.}$$

which yields the result with $K_2 = 1$ again.

Case 2. Assume that $\gamma > 1$ in (*C*2). Since l' is now increasing the above argument fails to guaranty that $\rho'(x) \sim \rho'(y)$ as $y \sim x$ are sufficiently close. Instead we check the sufficient condition in Proposition 30. The function $l(x)/x$ is increasing as it is regularly varying with index $\gamma - 1 > 0$. Recall also that (*CFG*) yields (A.3) and that $H_X + H_Y$ is bounded thanks to (*FG*3). As a consequence of $I_n \subset \Delta_n$ and Lemma 24 we almost surely have, for all n large,

$$\begin{aligned}
\sup_{u \in I_n} \frac{l \circ \tau(u)}{\tau(u)}|\varepsilon_n(u)| &\le 2K_0 \sup_{u \in I_n} l \circ F^{-1}(u)\big(H_X(u) + H_Y(u)\big)\sqrt{\frac{\log\log n}{n(1-u)}} \\
&\le 2K_0 l \circ F^{-1}\left(1 - \frac{k_n}{n}\right)\sqrt{\frac{\log\log n}{k_n}} \sup_{u \in I_n}\big(H_X(u) + H_Y(u)\big) \\
&\le K\frac{\log n}{\sqrt{k_n}}\sqrt{\log\log n} \sup_{u \in I_n}\big(H_X(u) + H_Y(u)\big).
\end{aligned} \tag{A.4}$$

Since $\theta > 2$ in (*CFG*) choosing $\theta' \in (2, \theta)$ in (23) makes the upper bound in (A.4) vanish. Therefore, under (*CFG*) the requirements of Proposition 30 are almost surely ultimately fulfilled with

$$x_0 = \tau_0, \qquad x = \tau(u), \qquad |\varepsilon| = |\varepsilon_n(u)| \le \frac{\delta_0}{l'(x)} = \frac{\delta_0}{l' \circ \tau(u)} = \frac{\delta_0 \tau(u)}{\gamma l \circ \tau(u)}, \quad u \in I_n,$$

which entails that, for all n large enough and $K_2 = k_0$,

$$\big|\rho\big(|\tau(u) + \varepsilon_n(u)|\big) - \rho\big(\tau(u)\big)\big| \le k_0 \rho' \circ \tau(u)|\varepsilon_n(u)| \le K_2 \rho' \circ F^{-1}(u)|\varepsilon_n(u)|. \tag{A.5}$$

Case 3. Assume that $0<\gamma\leq 1$ in $(C2)$. Since $l(x)/x$ is either decreasing or, if $\gamma=1$, not even monotone, $l\circ\tau(u)/\tau(u)$ cannot be compared to the worse case $\tau(u)\sim F^{-1}(u)$ directly. However, by Proposition 30, if $u\in I_n$ is such that $|\varepsilon_n(u)|\leq \delta_0/l'\circ\tau(u)$ then (A.5) holds. Consider

$$I_n^-=\left\{u\in I_n:\left|\varepsilon_n(u)\right|>\frac{\delta_0}{l'\circ\tau(u)}\right\}.$$

Since $l'(x)\sim\gamma l(x)/x$ and $\rho(x)\sim x\rho'(x)/\gamma l(x)$ as $x\to+\infty$, for $0<x_0<\tau_0$ we can find $\xi_0>1/\gamma$ such that

$$\rho(x)\leq\xi_0\rho'(x)\frac{x}{l(x)},\quad x\geq x_0. \tag{A.6}$$

Let $\xi_1>\gamma/\delta_0$ and assume n so large that $l\circ\tau(u)>1/\xi_1$ and $\tau(u)\geq\tau_0$ for $u\in I_n$. Any $u\in I_n^-$ then satisfies

$$\tau_0\leq\max\bigl(\tau(u),\left|\varepsilon_n(u)\right|\bigr)\leq\max\left(\delta_0\xi_1\frac{l\circ\tau(u)}{l'\circ\tau(u)},\left|\varepsilon_n(u)\right|\right)\leq\frac{x_n(u)}{2}:=\xi_1 l\circ\tau(u)\left|\varepsilon_n(u)\right|.$$

By (A.6) and the fact that $l(x)$ is increasing it follows that

$$\left|\rho\bigl(\left|\tau(u)+\varepsilon_n(u)\right|\bigr)-\rho\bigl(\tau(u)\bigr)\right|\leq\rho\bigl(\tau(u)+\left|\varepsilon_n(u)\right|\bigr)\leq\xi_0\rho'\bigl(x_n(u)\bigr)\frac{x_n(u)}{l\circ\tau(u)}=2\xi_0\xi_1\rho'\bigl(x_n(u)\bigr)\left|\varepsilon_n(u)\right|.$$

Using (A.3) as for (A.4) we almost surely eventually have

$$\frac{1}{2\xi_1}\sup_{u\in I_n^-}\frac{x_n(u)}{F^{-1}(u)}=\sup_{u\in I_n^-}l\circ\tau(u)\frac{|\varepsilon_n(u)|}{F^{-1}(u)}\leq K\frac{\log n}{\sqrt{k_n}}\sqrt{\log\log n}\sup_{u\in I_n^-}\bigl(H_X(u)+H_Y(u)\bigr)$$

and the upper bound tends to 0 provided that $2<\theta'<\theta$ from (23). As a conclusion, $x_n(u)\leq F^{-1}(u)$ on I_n^- even if $|\varepsilon_n(u)|$ is large and it asymptotically holds, for $K_2=\max(k_0,2\xi_0\xi_1)$,

$$\left|\rho\bigl(\left|\tau(u)+\varepsilon_n(u)\right|\bigr)-\rho\bigl(\tau(u)\bigr)\right|\leq K_2\rho'\bigl(F^{-1}(u)\bigr)\left|\varepsilon_n(u)\right|,\quad u\in I_n.$$

A.3. *Strong approximation of the joint quantile processes*

Let P_H denote the probability distribution associated to the c.d.f. H. In this section $(FG1)$ and $(FG2)$ are crucially required to justify the key approximation used at steps 4 and of the main proof. Let k_n be defined as in (22), thus $k_n/n\to 0$, $k_n/\log\log n\to+\infty$. Consider $\mathcal{I}_n=(k_n/n,1-k_n/n)$ which contains both $I_{M,n}$ from (31) and Δ_n from (25). As in (26) write $\beta_n^X=\sqrt{n}(\mathbb{F}_n^{-1}-F^{-1})$ and $\beta_n^Y=\sqrt{n}(\mathbb{G}_n^{-1}-G^{-1})$ the quantile processes associated to each sample. Our goal is to derive a coupling of

$$\bigl\{\bigl(\beta_n^X(u),\beta_n^Y(u)\bigr):u\in\mathcal{I}_n\bigr\}\quad\text{and}\quad\left\{\left(\frac{B_n^X(u)}{h_X(u)},\frac{B_n^Y(u)}{h_Y(u)}\right):u\in\mathcal{I}_n\right\},$$

where (B_n^X,B_n^Y) are two marginal standard Brownian Bridges indexed by $u\in[0,1]$,

$$B_n^X(u)=\mathbb{B}_n(\mathcal{H}_{F^{-1}(u)})\quad\text{and}\quad B_n^Y(u)=\mathbb{B}_n\bigl(\mathcal{H}^{G^{-1}(u)}\bigr), \tag{A.7}$$

driven by a sequence $\mathbb{B}_n$ of P_H-Brownian Bridge indexed by the set $\mathcal{C}$ of half planes $\mathcal{H}_{x_0}=\{(x,y):x\leq x_0\}$ or $\mathcal{H}^{y_0}=\{(x,y):y\leq y_0\}$. In other words, $\mathbb{B}_n$ is a zero mean Gaussian process indexed by $\mathcal{C}$ having covariance

$$\operatorname{cov}\bigl(\mathbb{B}_n(A),\mathbb{B}_n(B)\bigr)=P_H(A\cap B)-P_H(A)P_H(B),$$

for $A,B\in\mathcal{C}$, and B_n^X are centered Gaussian processes with covariance

$$\begin{aligned}
\operatorname{cov}\bigl(B_n^X(u),B_n^X(v)\bigr)&=P_H(\mathcal{H}_{F^{-1}(u)}\cap\mathcal{H}_{F^{-1}(v)})-uv=\min(u,v)-uv,\\
\operatorname{cov}\bigl(B_n^Y(u),B_n^Y(v)\bigr)&=P_H\bigl(\mathcal{H}^{G^{-1}(u)}\cap\mathcal{H}^{G^{-1}(v)}\bigr)-uv=\min(u,v)-uv,\\
\operatorname{cov}\bigl(B_n^X(u),B_n^Y(v)\bigr)&=P_H\bigl(\mathcal{H}_{F^{-1}(u)}\cap\mathcal{H}^{G^{-1}(v)}\bigr)-uv=L(u,v)-uv,
\end{aligned}$$

for $u, v \in [0,1]$, where the usual copula function $L(u,v) = H(F^{-1}(u), G^{-1}(v))$ measures the distortion between H and $F \otimes G$ on all quadrants, half spaces and then rectangles.

The coupling is achieved at Theorem 29 simply by combining the strong approximation of the empirical process (see [3])

$$\Lambda_n(A) = \sqrt{n}\big(P_{H_n}(A) - P_H(A)\big), \quad A \in \mathcal{C}, P_{H_n} = \frac{1}{n}\sum_{i \le n} \delta_{(X_i, Y_i)},$$

with the usual quantile transform and classical results for real quantiles. This result has an interest by itself as it is valid whatever the joint distribution Π satisfying the marginal conditions $(FG1)$ and $(FG2)$.

Remark 28. Theorem 29 remains valid for the d marginal quantile processes of a distribution in $\mathbb{R}^d$ provided each marginal distribution obeys $(FG1)$ and $(FG2)$, with no change in the proof for $d = 2$.

Theorem 29. *Assume that F, G satisfy $(FG1)$ and $(FG2)$. One can built on the same probability space the sequence $\{(X_n, Y_n)\}$ and a sequence of versions of $\{(B_n^X(u), B_n^Y(u)) : u \in \mathcal{I}_n\}$ such that*

$$\beta_n^X(u) = \frac{B_n^X(u) + Z_n^X(u)}{h_X(u)}, \qquad \beta_n^Y(u) = \frac{B_n^Y(u) + Z_n^Y(u)}{h_Y(u)},$$

satisfy, for some $\xi > 0$,

$$\lim_{n\to+\infty} n^\xi \sup_{u\in\mathcal{I}_n} \big|Z_n^X(u)\big| = \lim_{n\to+\infty} n^\xi \sup_{u\in\mathcal{I}_n} \big|Z_n^Y(u)\big| = 0 \quad \textit{a.s.}$$

Moreover we can take $(B_n^X(u), B_n^Y(u)) = \sum_{k=1}^n (G_k^X(u), G_k^Y(u))/\sqrt{n}$ where $\{(G_k^X(u), G_k^Y(u)) : u \in (0,1)\}$ is a sequence of independent versions of Brownian Bridges (G^X, G^Y) with $\operatorname{cov}(G^X(u), G^Y(v)) = L(u,v) - uv$.

Proof. Define the two marginal empirical processes to be, for $x \in \mathbb{R}$,

$$\alpha_n^X(x) = \sqrt{n}\big(\mathbb{F}_n(x) - F(x)\big) = \Lambda_n(\mathcal{H}_x), \alpha_n^Y(x) = \sqrt{n}\big(\mathbb{G}_n(x) - G(x)\big) = \Lambda_n\big(\mathcal{H}^x\big).$$

Under $(FG1)$ the random variables $U_i = F(X_i)$ and $V_i = G(Y_i)$ are uniform on $(0,1)$. Write $\alpha_n^{X,U}$ and $\alpha_n^{Y,V}$ the uniform empirical process associated to $U_1, \ldots, U_n$ and $V_1, \ldots, V_n$ respectively. Also write $\mathbb{F}_{X,U,n}$ and $\mathbb{F}_{X,U,n}^{-1}$ the empirical c.d.f. and quantile functions then $\beta_n^{X,U}(u) = \sqrt{n}(\mathbb{F}_{X,U,n}^{-1}(u) - u)$. Likewise write $\mathbb{F}_{Y,V,n}$, $\mathbb{F}_{Y,V,n}^{-1}$ and $\beta_n^{Y,V}$. Clearly $\alpha_n^{X,U}$ and $\alpha_n^{Y,V}$ are not independent, neither are $\beta_n^{X,U}$ and $\beta_n^{Y,V}$. What is next obtained for X is also valid for Y. Under $(FG1)$ and $(FG2)$ the arguments given in Section A.1 yield that

$$\lim_{n\to+\infty} \frac{\sqrt{n}}{\log\log n} \sup_{u\in\mathcal{I}_n} \big|h_X(u)\beta_n^X(u) - \beta_n^{X,U}(u)\big| = 0 \quad \text{a.s.} \tag{A.8}$$

since $\beta_n^{X,U} = \sqrt{n} q_n$ and the supremum is showed to be less than $\sqrt{n} a_n b_n$ with the almost sure bounds such that $a_n < K(\log\log n)/n$ and $b_n \to 0$ as $n \to +\infty$. By [2] and [12] we also have

$$\limsup_{n\to+\infty} \frac{n^{1/4}}{\sqrt{\log n}(\log\log n)^{1/4}} \sup_{u\in\mathcal{I}_n} \big|\beta_n^{X,U}(u) + \alpha_n^{X,U}(u)\big| \le \frac{1}{2^{1/4}} \quad \text{a.s.} \tag{A.9}$$

thus for any $\xi < 1/4$ it holds

$$\lim_{n\to+\infty} n^\xi \sup_{u\in\mathcal{I}_n} \big|h_X(u)\beta_n^X(u) + \alpha_n^{X,U}(u)\big| = 0 \quad \text{a.s.}$$

It is important here that (A.8) and (A.9) holds true for $\beta_n^{X,U}$ and $\beta_n^{Y,U}$ simultaneously with probability one whatever the underlying probability space. Hence, recalling that $\alpha_n^{X,U} = \alpha_n^X \circ F^{-1}$, $P_{H_n}(\mathcal{H}_{F^{-1}(u)}) = \mathbb{F}_n(F^{-1}(u))$ and $P_H(\mathcal{H}_{F^{-1}(u)}) = u$ it follows that

$$\lim_{n\to+\infty} n^\xi \sup_{u\in\mathcal{I}_n} \big|h_X(u)\beta_n^X(u) + \Lambda_n(\mathcal{H}_{F^{-1}(u)})\big| = 0 \quad \text{a.s.}$$

$$\lim_{n\to+\infty} n^\xi \sup_{u\in\mathcal{I}_n} \big|h_Y(u)\beta_n^Y(u) + \Lambda_n\big(\mathcal{H}^{G^{-1}(u)}\big)\big| = 0 \quad \text{a.s.}$$

on any probability space. It remains to approximate Λ_n uniformly on $\mathcal{C}$. The collection of sets $\mathcal{C}$ is a VC-class of order 3 thus satisfies the uniform entropy condition (VC) used in [3] with $v_0 = 2(3-1) = 4$. By their Proposition 1 taken with $\theta = 2$ there exists a probability space on which the sequence $\{(X_n, Y_n)\}$ can be built together with a sequence $\mathbb{B}_n$ of P_H-Brownian Bridges indexed by $\mathcal{C}$ such that

$$\mathbb{P}\left(\sup_{A\in\mathcal{C}}\left|\Lambda_n(A) + \mathbb{B}_n(A)\right| \geq \frac{K}{n^{\beta_2}}\right) \leq \frac{1}{n^2}$$

where we take $\beta_2 < 1/22$ to avoid the $\log n$ factor. Note that since $\mathbb{B}_n$ and $-\mathbb{B}_n$ have the same distribution, we choose to approximate with $-\mathbb{B}_n$. Consider in particular $\mathcal{H}_n^X = \{\mathcal{H}_{F^{-1}(u)} : u \in \mathcal{I}_n\} \subset \mathcal{C}$ and define $B_n^X(u) = \mathbb{B}_n(\mathcal{H}_{F^{-1}(u)})$. On the previous probability space it holds

$$\limsup_{n\to+\infty} n^{\beta_2} \sup_{u\in\mathcal{I}_n}\left|\alpha_n^X \circ F^{-1}(u) + B_n^X(u)\right| = \limsup_{n\to+\infty} n^{\beta_2} \sup_{A\in\mathcal{H}_n^X}\left|\Lambda_n(A) + \mathbb{B}_n(A)\right| \leq K \quad \text{a.s.}$$

the above comparison between $h_X(u)\beta_n^X(u)$ and $\alpha_n^X \circ F^{-1}(u)$ gives in turn, for $\xi < \min(1/4, \beta_2) = \beta_2$ and $Z_n^X(u) = h_X(u)\beta_n^X(u) - B_n^X(u)$,

$$\limsup_{n\to+\infty} n^{\xi} \sup_{u\in\mathcal{I}_n}\left|Z_n^X(u)\right| = 0 \quad \text{a.s.}$$

In the same way we simultaneously obtain, for $Z_n^Y(u) = h_Y(u)\beta_n^Y(u) - B_n^Y(u)$,

$$\limsup_{n\to+\infty} n^{\xi} \sup_{u\in\mathcal{I}_n}\left|Z_n^Y(u)\right| = 0 \quad \text{a.s.}$$

The processes B_n^X and B_n^Y are joint through the leading process $\mathbb{B}_n$ and the covariance $\operatorname{cov}(B_n^X(u), B_n^Y(v)) = L(u,v) - uv$. The second statement to be proved follows by applying Theorem 1 of [3] in place of Proposition 1. If $\beta_2 > 0$ is chosen small enough the approximating process can be built in the form $\mathbb{B}_n = \sum_{k=1}^n \mathbb{B}_k^*/\sqrt{n}$ where $\{\mathbb{B}_k^* : k \geq 1\}$ is a sequence of independent P_H-Brownian Bridges. Since $\mathbb{B}_n$ is again a P_H-Brownian Bridge, $G_k^X(u) = \mathbb{B}_k^*(\mathcal{H}_{F^{-1}(u)})$ and $G_k^Y(u) = \mathbb{B}_k^*(\mathcal{H}^{G^{-1}(u)})$ are standard Brownian Bridges with the desired correlation structure. □

A.4. *Consequences of* ($C2$)

In this section, we establish the deterministic expansion that is required to stochastically control the centered extreme values of the cost function evaluated at large quantiles. We start by presented some regular and slow variation properties – for more details we refer to [13,16]. For $k \in \mathbb{N}_*$ let $\mathcal{M}_k(m, +\infty)$ be the subset of functions $\varphi \in \mathcal{C}_k$ such that $\varphi^{(k)}$ is monotone on $(m, +\infty)$, and hence $\varphi, \varphi', \varphi'', \ldots, \varphi^{(k)}$ are also monotone on $(m, +\infty)$ by changing m. Let $\mathcal{M}_0(m, +\infty)$ denote the set of continuous functions monotone on $(m, +\infty)$. Write $\mathrm{RV}(\gamma)$ the set of regularly varying functions at $+\infty$ with index $\gamma \in \mathbb{R}$. They are of the form $x^\gamma L(x)$ with $L \in \mathrm{RV}(0)$, which means that given any $\lambda > 0$, $L(\lambda x)/L(x) \to 1$ as $x \to +\infty$. If $L \in \mathrm{RV}(0)$ is monotone on $(m, +\infty)$ then L is equivalent at $+\infty$ to a function in $\mathcal{C}_\infty(m, +\infty) \cap \mathrm{RV}(0)$. Therefore, at the first order, it is not a restriction to assume that functions of $\mathrm{RV}(\gamma)$ are in $\mathcal{M}_k(m, +\infty)$ as well. Problems however arise with respect to differentiation. In particular, two apparently close slowly varying functions may have very different local variations. First consider the smooth regular variation. Let introduce

$$\mathrm{RV}_k(\gamma, m) = \mathrm{RV}(\gamma) \cap \mathcal{M}_k(m, +\infty), \quad \gamma \neq 0.$$

The following statements are taken as $x \to +\infty$. Assuming that $k \geq 1$ and $\gamma \neq 0$, if $\varphi \in \mathrm{RV}_k(\gamma, m)$ then φ' is monotone, so that it holds, by the monotone density theorem,

$$\varphi'(x) \sim \frac{\gamma\varphi(x)}{x}. \tag{A.10}$$

This implies that $\varphi' \in \mathrm{RV}_{k-1}(\gamma - 1, m)$ and, whenever $k \geq 2$ and $\gamma \neq 1$, φ'' in turns satisfies $\varphi'' \in \mathrm{RV}_{k-2}(\gamma - 2, m)$ and

$$\varphi''(x) \sim \frac{(\gamma - 1)\varphi'(x)}{x} \sim \frac{\gamma(\gamma - 1)\varphi(x)}{x^2}. \tag{A.11}$$

For $L \in \mathrm{RV}(0)$ it holds, by (6) and Karamata's theorem,

$$\frac{\int_m^x L'(t)(\frac{1}{\varepsilon_1(t)})\,dt}{\int_m^x L'(t)\,dt} = \frac{\int_m^x L'(t)(\frac{L(t)}{tL'(t)})\,dt}{\int_m^x L'(t)\,dt} = \frac{1}{L(x)}\int_m^x \frac{L(t)}{t}\,dt \to +\infty.$$

Hence the function $1/\varepsilon_1(t)$ is unbounded and, if $L \in \mathcal{C}_1(m,+\infty)$, continuous on $(m,+\infty)$. It is not very restrictive to exclude such functions that are asymptotically oscillating and not going to infinity. If $L(x) = \varphi(L_1(x))$ where $\varphi \in \mathrm{RV}_2(\gamma, m)$ and $\gamma > 0$ then we get $\varepsilon_1(x) \sim \gamma x L_1'(x)/L_1(x)$. For instance, if $L(x) = \varphi(\log x)$ where $\varphi \in \mathrm{RV}_2(\gamma, m)$ and $\gamma > 0$ then $\varepsilon_1(x) \sim \gamma/\log x$. Also remind the well known representation, for $x \in (m,+\infty)$, $L(x) = d_0(x)\exp(\int_m^x \varepsilon_0(t)/t\,dt)$, $d_0(x) \to d_0 > 0$, $\varepsilon_0(x) \to 0$. If $d_0(x)$ is constant then $d_0 = L(m)$ and $\varepsilon_0(x) = \varepsilon_1(x)$ from (6). More generally, (6) is equivalent to $xd_0'(x) \to 0$ and we have $\varepsilon_1(x) = \varepsilon_0(x) + xd_0'(x)$.

Proposition 30. *Assume* $(C2)$. *Then it holds, for any* $x_0 > \tau_1$,

$$\rho\big(|x+\varepsilon|\big) - \rho(x) = k_0(x,\varepsilon)\rho'(x)\varepsilon \quad \text{and} \quad \lim_{\delta_0\to 0}\sup_{x>x_0}\sup_{|\varepsilon|l'(x)\le\delta_0}\big|k_0(x,\varepsilon)-1\big| = 0. \tag{A.12}$$

In particular, there exists $\delta_0 > 0$ *and* $k_0 > 0$ *such that, for all* $x > x_0$ *and* $|\varepsilon| \le \delta_0/l'(x)$ *we have* $|\rho(|x+\varepsilon|) - \rho(x)| \le k_0\rho'(x)|\varepsilon|$.

Proof. Fix $x_0 > \tau_1 > 0$ and let $M > x_0$ be as large as needed below. If $\varepsilon = 0$ then (A.12) requires that $k_0(x,0) = 1$ for $x > x_0$. For $\varepsilon \neq 0$ we distinguish between $x \in (x_0, M)$ and $x \ge M$. In the first case, since $\rho \in \mathcal{C}_2(\mathbb{R}_+)$ under $(C2)$ the Taylor expansion of ρ holds uniformly on (x_0, M). Namely, for any δ_0 small enough, $x \in (x_0, M)$ and $|\varepsilon| \le \varepsilon_0 = \delta_0/\inf\{l'(x) : x \in (x_0, M)\} < x_0 - \tau_1$ we have

$$\rho\big(|x+\varepsilon|\big) - \rho(x) = k_0(x,\varepsilon)\rho'(x)\varepsilon, \qquad k_0(x,\varepsilon) = 1 + \frac{\rho''(x^*)}{2\rho'(x)}\varepsilon,$$

with $x^* \in (x_0 - \varepsilon_0, M + \varepsilon_0)$ and $|k_0(x,\varepsilon) - 1| \le K\delta_0$ where $K < +\infty$ depends on x_0, M, ρ. We deduce that, for any $M > x_0$,

$$\lim_{\delta_0\to 0}\sup_{x_0<x<M}\sup_{|\varepsilon|l'(x)\le\delta_0}\big|k_0(x,\varepsilon)-1\big| = 0. \tag{A.13}$$

If $x \ge M$ then $l'(x) > 0$ and we intend to expand

$$\rho\big(|x+\varepsilon|\big) - \rho(x) = \rho(x)\big(\exp\big(l\big(|x+\varepsilon|\big) - l(x)\big) - 1\big). \tag{A.14}$$

Case 1. Assume $(C2)$ with $\gamma > 0$. Write $l(x) = x^\gamma L(x)$ where $L \in \mathcal{RV}_2(0, \tau_1)$ satisfies (6). For any $\delta_0 \in (0, \gamma l(M)/4)$ define

$$\Delta_0 = \big\{(x,\varepsilon) : x \ge M, |\varepsilon|l'(x) \le \delta_0\big\}. \tag{A.15}$$

By (A.10), for M large enough and $(x,\varepsilon) \in \Delta_0$ it holds $l'(x) > \gamma l(x)/2x$, which implies $|\varepsilon|/x \le 2\delta_0/\gamma l(x) < 1/2$ and $|x+\varepsilon| = x+\varepsilon > M/2$. Therefore $\sup_{(x,\varepsilon)\in\Delta_0}|\varepsilon|/x \to 0$ as $\delta_0 \to 0$ and we have, for $(x,\varepsilon) \in \Delta_0$,

$$\frac{l(x+\varepsilon) - l(x)}{x^\gamma} = \left(1 + \frac{\varepsilon}{x}\right)^\gamma L(x+\varepsilon) - L(x) = \frac{\gamma\varepsilon}{x}\big(1 + \delta_1(x,\varepsilon)\big)L(x+\varepsilon) + L(x+\varepsilon) - L(x) \tag{A.16}$$

where $\sup_{(x,\varepsilon)\in\Delta_0}|\delta_1(\varepsilon,x)| \to 0$ as $\delta_0 \to 0$. By (6) we also have, for $(x,\varepsilon) \in \Delta_0$,

$$\big|L(x+\varepsilon) - L(x)\big| \le \sup_{|y-x|\le|\varepsilon|}\big|L'(y)\big||\varepsilon| = \sup_{|y-x|\le|\varepsilon|}\big|\varepsilon_1(y)\big|\frac{L(y)}{y}|\varepsilon|$$

where $\varepsilon_1(y) \to 0$ as $y > x - |\varepsilon| > M/2 \to +\infty$. Moreover, for $\delta = 2\delta_0/\gamma l(M)$,

$$\frac{1}{L(x)}\sup_{|y-x|\le|\varepsilon|} L(y) = \sup_{1-|\varepsilon|/x<\lambda<1+|\varepsilon|/x}\frac{L(\lambda x)}{L(x)} \le \sup_{1-\delta<\lambda<1+\delta}\frac{L(\lambda x)}{L(x)}$$

and the second term has limit 1 as $x \to +\infty$ since $L \in \mathrm{RV}(0)$. Hence for any $\eta > 0$, assuming M so large that $\sup_{y>M/2} |\varepsilon_1(y)| < \eta/4$ and δ_0 small ensures that, for $(x,\varepsilon) \in \Delta_0$,

$$\big|L(x+\varepsilon) - L(x)\big| \le \frac{\eta}{3} \frac{L(x)}{x-|\varepsilon|}|\varepsilon| \le \frac{\eta}{2}\frac{|\varepsilon|}{x}L(x)$$

and (A.16) reads

$$l(x+\varepsilon) - l(x) = \big(1+\delta_2(x,\varepsilon)\big)\frac{\gamma x^\gamma L(x)}{x}\varepsilon = \big(1+\delta_3(x,\varepsilon)\big)l'(x)\varepsilon.$$

Then (A.14) gives

$$\frac{\rho(x+\varepsilon)-\rho(x)}{l'(x)\rho(x)\varepsilon} = \frac{\exp(l(x+\varepsilon)-l(x))-1}{l'(x)\varepsilon} = 1+\delta_4(x,\varepsilon) = k_0(x,\varepsilon)$$

with $\sup_{(x,\varepsilon)\in\Delta_0} |\delta_k(\varepsilon,x)| < \eta$ for $k=2,3,4$. We have proved that for any $\eta > 0$ there exists M such that

$$\lim_{\delta_0\to 0} \sup_{x\ge M} \sup_{|\varepsilon| l'(x)\le\delta_0} \big|k_0(x,\varepsilon)-1\big| \le \eta,$$

which yields (A.12) when combined to (A.13).

Case 2. Assume $(C2)$ with $\gamma = 0$. Since $l \in \mathcal{RV}_2^+(0,\tau_1)$, (6) and (7) give $\varepsilon_1(x) = xl'(x)/l(x) \ge l_1/l(x) > 0$ where $\varepsilon_1(x) \to 0$. Thus $l'(x) \to 0$ as $x \to +\infty$ and $l' \in \mathcal{RV}_2(-1,\tau_1)$, and l' is decreasing on $(M,+\infty)$ since $l' \in \mathcal{M}_2(\tau_1,+\infty)$. Consider $\delta_0 \in (0, 1/2l_1)$ and define Δ_0 as in (A.15). For $(x,\varepsilon) \in \Delta_0$ it holds

$$\frac{|\varepsilon|}{x} \le l_1\frac{|\varepsilon|}{x} \le l(x)\varepsilon_1(x)\frac{|\varepsilon|}{x} = |\varepsilon| l'(x) \le \delta_0,$$

hence $|x+\varepsilon| = x+\varepsilon > M/2$ again, and

$$l'\big(x+|\varepsilon|\big)|\varepsilon| \le \big|l(x+\varepsilon)-l(x)\big| \le l'\big(x-|\varepsilon|\big)|\varepsilon|$$

where, since $l''(x) \sim -l'(x)/x$ by (A.11),

$$0 \le \frac{l'(x-|\varepsilon|)-l'(x)}{l'(x-|\varepsilon|)} \le \sup_{x-|\varepsilon|\le y\le x} \frac{|l''(y)\varepsilon|}{l'(y)} \le \frac{|\varepsilon|}{x-|\varepsilon|} \le 2\delta_0,$$

$$0 \le \frac{l'(x)-l'(x+|\varepsilon|)}{l'(x)} \le \sup_{x\le y\le x+|\varepsilon|} \frac{|l''(y)\varepsilon|}{l'(y)} \le \frac{|\varepsilon|}{x} \le \delta_0.$$

We deduce that for $k=1,2$ and $\sup_{(x,\varepsilon)\in\Delta_0} |\delta_k(\varepsilon,x)| \to 0$ as $\delta_0 \to 0$ it holds

$$l(x+\varepsilon)-l(x) = \big(1+\delta_1(x,\varepsilon)\big)l'(x)\varepsilon$$

for all $(x,\varepsilon) \in \Delta_0$ and, by (A.14),

$$\frac{\rho(x+\varepsilon)-\rho(x)}{l'(x)\rho(x)\varepsilon} = \frac{\exp(l(x+\varepsilon)-l(x))-1}{l'(x)\varepsilon} = 1+\delta_2(x,\varepsilon) = k_0(x,\varepsilon)$$

thus (A.12) follows. □

Several arguments exploit the asymptotic convexity of ρ which follows from $(C2)$.

Proposition 31. *Under $(C2)$ the function $\rho(x)$ is convex on $(l_2,+\infty)$ for some $l_2 > 0$. If moreover $l_1 > 1$ it is strictly convex.*

Proof. We have to show that $\rho''(x) = (l''(x)+l'(x)^2)\rho(x) \ge 0$ if $(C2)$ holds. In the case $1 \ne \gamma > 0$ we have, by (A.10) and (A.11), as $x \to +\infty$,

$$l'(x) \sim \frac{\gamma l(x)}{x}, \qquad l''(x) \sim \frac{\gamma(\gamma-1)l(x)}{x^2} \ll l'(x), \qquad \frac{l''(x)}{l'(x)} \sim \frac{\gamma-1}{x},$$

thus there exists $l_2 > l^{-1}(1/\gamma)$ such that all $x > l_2$ satisfy $l'(x) > 0$ and

$$l''(x) + l'(x)^2 \sim l'(x)\left(\frac{\gamma - 1}{x} + \frac{\gamma l(x)}{x}\right) \geq \frac{l'(x)}{x}\big(\gamma l(x) - 1\big) > 0.$$

If $\gamma = 1$ then $l(x) = xL(x)$ and $l''(x) = 2L'(x) + xL''(x) \sim L'(x)$ whereas $l'(x)^2 \sim (L(x) + xL'(x))^2 \sim L^2(x)$. Since $L'(x)/L^2(x) = \varepsilon_1(x)/xL(x) \to 0$ we have $l''(x) + l'(x)^2 > 0$ for $x > l_2$. If $\gamma = 0$ in $(C2)$ then by (6), (7) and (A.11) we have $l_1/x \leq l'(x) \leq l(x)/x$ and $l' \in \mathrm{RV}_1^+(-1, 0)$ and we get, as $x \to +\infty$,

$$l''(x) + l'(x)^2 \sim \frac{l'(x)}{x}\big(\varepsilon_1(x)l(x) - 1\big) \geq \frac{l'(x)}{x}(l_1 - 1).$$

Therefore if $l_1 > 1$, $\rho(x)$ is strictly convex on $(l_2, +\infty)$ for l_2 large enough. It remains convex for $l_1 = 1$. □

Acknowledgements

We thank the referee who suggested the Vitali's argument for an elegant alternative proof of Theorem 13 and pointed [11] to our attention.

References

[1] P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos and C. Matrán. Uniqueness and approximate computation of optimal incomplete transportation plans. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** (2) (2011) 358–375. MR2814414 https://doi.org/10.1214/09-AIHP354
[2] R. R. Bahadur. A note on quantiles in large samples. *Ann. Math. Stat.* **37** (1966) 577–580. MR0189095 https://doi.org/10.1214/aoms/1177699450
[3] P. Berthet and D. M. Mason. Revisiting two strong approximation results of Dudley and Philipp. In *High Dimensional Probability* 155–172. *IMS Lecture Notes Monogr. Ser.* **51**. Inst. Math. Statist, Beachwood, OH, 2006. MR2387767 https://doi.org/10.1214/074921706000000824
[4] S. G. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics and kantorovich transport distances. Preprint, 2016. To appear in *Mem. Amer. Math. Soc.*
[5] S. Cambanis, G. Simons and W. Stout. Inequalities for $Ek(X, Y)$ when the marginals are fixed. *Z. Wahrsch. Verw. Gebiete* **36** (4) (1976) 285–294. MR0420778 https://doi.org/10.1007/BF00532695
[6] M. Csörgö and P. Révész. Strong approximations of the quantile process. *Ann. Statist.* **6** (4) (1978) 882–894. MR0501290
[7] E. del Barrio, E. Giné and C. Matrán. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27** (2) (1999) 1009–1071. MR1698999 https://doi.org/10.1214/aop/1022677394
[8] E. del Barrio, E. Giné and C. Matrán. Correction: "Central limit theorems for the Wasserstein distance between the empirical and the true distributions" [Ann. Probab. **27** (1999), no. 2, 1009–1071; MR1698999 (2000g:60034)]. *Ann. Probab.* **31** (2) (2003) 1142–1143. MR1964961
[9] E. del Barrio, E. Giné and F. Utzet. Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli* **11** (1) (2005) 131–189. MR2121458 https://doi.org/10.3150/bj/1110228245
[10] E. del Barrio and J.-M. Loubès. Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.* **47** (2) (2019) 926–951. MR3916938 https://doi.org/10.1214/18-AOP1275
[11] G. Freitag, C. Czado and A. Munk. A nonparametric test for similarity of marginals – with applications to the assessment of population bioequivalence. *J. Statist. Plann. Inference* **137** (3) (2007) 697–711. MR2301710 https://doi.org/10.1016/j.jspi.2006.06.003
[12] J. Kiefer. Deviations between the sample quantile process and the sample df. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)* 299–319. Cambridge Univ. Press, London, 1970. MR0277071
[13] T. Mikosch. Regular variation, subexponentiality and their applications in probability theory. Operations Research EURANDOM European Institute for Statistics, Probability, and their Applications. EURANDOM report, Eindhoven University of Technology, 1999.
[14] A. Munk and C. Czado. Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** (1) (1998) 223–241. MR1625620 https://doi.org/10.1111/1467-9868.00121
[15] T. Rippl, A. Munk and A. Sturm. Limit laws of the empirical Wasserstein distance: Gaussian distributions. *J. Multivariate Anal.* **151** (2016) 90–109. MR3545279 https://doi.org/10.1016/j.jmva.2016.06.005
[16] E. Seneta. *Regularly Varying Functions*. *Lecture Notes in Mathematics* **508**. Springer-Verlag, Berlin–New York, 1976. MR0453936
[17] G. R. Shorack. *Probability for Statisticians*. Springer, 2000. MR1762415
[18] M. Sommerfeld and A. Munk. Inference for empirical Wasserstein distances on finite spaces. ArXiv e-prints, 2016. MR3744719 https://doi.org/10.1111/rssb.12236
[19] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003. MR1964483 https://doi.org/10.1007/b12016