



# Weak convergence of empirical Wasserstein type distances

Philippe Berthet, Jean Claude Fort

## ► To cite this version:

Philippe Berthet, Jean Claude Fort. Weak convergence of empirical Wasserstein type distances. 2018.  
hal-01838700v1

**HAL Id: hal-01838700**

**<https://hal.science/hal-01838700v1>**

Preprint submitted on 13 Jul 2018 (v1), last revised 28 Oct 2023 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Weak convergence of empirical Wasserstein type distances

Philippe Berthet<sup>\*1</sup> and Jean Claude Fort<sup>2</sup>

<sup>1</sup>Institut de mathématique, UMR5219; Université de Toulouse;  
CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France

<sup>2</sup>MAP5, Université Paris Descartes, France

## Abstract

We estimate the Wasserstein type distance between two continuous distributions  $F$  and  $G$  on  $\mathbb{R}$  such that the set  $\{F = G\}$  is a finite union of intervals, possibly empty or  $\mathbb{R}$ . The positive cost function  $\rho$  is not necessarily symmetric and the sample may come from any joint distribution  $H$  on  $\mathbb{R}^2$  having marginals  $F$  and  $G$  with light enough tails with respect to  $\rho$ . The rates of weak convergence and the limiting distributions are derived in a wide class of situations including the classical distances  $W_1$  and  $W_2$ . The key new assumption in the case  $F = G$  involves the behavior of  $\rho$  near 0, which we assume to be regularly varying with index ranging from 1 to 2. Rates are then also regularly varying with powers ranging from  $1/2$  to 1 also affecting the limiting distribution, in addition to  $H$ .

### *Keywords:*

Central limit theorems, Generalized Wasserstein distances, Empirical processes, Strong approximation, Dependent samples, Non-parametric statistics, Goodness of fit tests.

### *MSC Classification:*

62G30, 62G20, 60F05, 60F17.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Motivation . . . . .	2
1.2	Setting . . . . .	4
1.3	Overview . . . . .	5

---

<sup>\*</sup>Corresponding author [philippe.berthet@math.univ-toulouse.fr](mailto:philippe.berthet@math.univ-toulouse.fr)

<b>2</b>	<b>Assumptions</b>	<b>6</b>
2.1	Assumptions ( $FG$ ) . . . . .	6
2.2	Notation for regularity . . . . .	7
2.3	Assumptions ( $C$ ) . . . . .	8
2.4	Assumptions ( $CFG$ ) . . . . .	8
<b>3</b>	<b>Statement of the results</b>	<b>10</b>
<b>4</b>	<b>Applications</b>	<b>13</b>
4.1	Comparison and goodness-of-fit tests . . . . .	13
4.2	An application . . . . .	13
<b>5</b>	<b>Proofs</b>	<b>14</b>
5.1	The case $F = G$ . . . . .	14
5.2	The case $F < G$ . . . . .	27
5.3	The general case . . . . .	33
5.4	A special case : $F = G$ and $b = 2$ . . . . .	34

# 1 Introduction

## 1.1 Motivation

This article is the continuation of our previous paper [2]. In [2] we addressed the problem of estimating the distance between two different distributions with respect to a large class of Wasserstein costs. The framework was very simple: two samples of *i.i.d.* real random variables taking values in  $\mathbb{R}$  and having distinct continuous cumulative distribution function (*c.d.f.*)  $F$  and  $G$  are available. It has to be noticed that these samples are not assumed to be independent, for instance they may be formed from simultaneous experiments. From them we estimated Wasserstein type distances or costs between  $F$  and  $G$  by a natural and easily computed non-parametric plug-in estimator. The almost sure (*a.s.*) consistency of our estimator being easily established under minimal assumptions we mainly developed a sharp method of proof of the central limit theorem assuming that the tails of  $F$  and  $G$  are different enough and compatible with the cost. Our main contribution was to investigate rather deeply the latter relationship and to prove that the dependency only affects the limiting variance. However the special case of the distance  $W_1$  was not captured, and asymptotically non-symmetric costs or too close distributions were not allowed.

In the present paper the general framework remains exactly the same, but we investigate more situations, among which the important case  $F = G$ , the case where  $F \neq G$  on  $\mathbb{R}$  but may have arbitrarily close tails and the case where the two situations  $F = G$  and  $F \neq G$  are encountered, but alternate along a finite number of intervals. We now include the distance  $W_1$  and non-symmetric costs provided they are regularly varying at both sides of 0. New assumptions arise that again illustrate how delicate tail integrals of transforms of empirical quantile functions can be for heavy-tailed distributions, but also sharpen the

understanding of this estimation problem. Moreover, assumptions are kept as close as possible to necessary conditions.

More precisely, the motivation of our initial work was intimately related to the development of computer experiments. Many computer codes give as output not only a real multidimensional variable but the values of a function computed on so many points that it can be considered as a functional output. The case we are interested in is when this function is the *c.d.f.* of a real random variable. It turns out that Wasserstein distances are now commonly used to analyze such outputs. In view of defining new features for random *c.d.f.* such as median or quantiles, more general Wasserstein costs may be used as contrasts to compute these features by solving an optimization problem – see [12]. Nevertheless computer codes only provide samples of the underlying distributions and thus the efficient estimation of such distances between *c.d.f.* via random samples is a natural and challenging field of research.

In the framework of univariate probability distributions the distance usually called *p*-Wasserstein is simply the  $L^p$  distance of simulated random variables from a common and universal – uniform on  $[0, 1]$  – simulator  $U$ , namely

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

where  $F^{-1}$  is the generalized inverse of  $F$ . We 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  $G$  built through *i.i.d.* samples of  $F$  and  $G$ . According to [3]  $W_p$  could also be named the *p*-Kantorovitch distance.

Many authors were interested in the convergence of  $W_p^p(\mathbb{F}_n, F)$ , see *e.g.* the survey paper [3] or [10, 8, 9, 1]. Up to our knowledge there are only two recent works studying the convergence of  $W_2^2(\mathbb{F}_n, \mathbb{G}_n)$  [11, 13], for independent samples. The drawback of [11] is that the estimator is not explicit from the data and the centering in the CLT is the biased  $\mathbb{E}W_2^2(\mathbb{F}_n, \mathbb{G}_n)$  rather than  $W_2^2(F, G)$  itself, moreover the limiting variance has no closed form expression and seems not easy to estimate. The drawback of [13] is that only discrete distributions and the distance  $W_2$  are considered. Notice also that in the early work [7] a trimmed version of the Mallows distance  $W_2^2(\mathbb{F}_n, \mathbb{G}_n)$  is studied, however under an implicit assumption on the level of trimming which has to hold in probability. Our proofs indeed show that a sharp control of the trimmed integrals has to be made in a weaker sense, and we provide easily checked sufficient conditions.

As in [2] we investigate a larger class of convex costs than  $W_p$  and we assume that the samples are not necessarily independent but actually come from an *i.i.d.* sample of a joint two-dimensional distribution  $H$  with marginals  $F$  and  $G$ . We look for an explicit CLT under tractable conditions relating  $F$  and  $G$  to the cost function. The method worked out in [2] is extended to cover the cases announced above, and especially in the situation  $F = G$  to provide a new class of goodness-of-fit and comparison tests with rates faster than  $\sqrt{n}$ . The power of the test associated to a given Wasserstein type cost is also accessible since we establish the weak convergence at the slower rate  $\sqrt{n}$  under many alternatives

$F \neq G$ . Therefore the results and arguments below are a new step toward appropriate statistical tools for the above motivation.

## 1.2 Setting

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

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

where  $X \sim F, Y \sim G$  means that  $X$  and  $Y$  are joint real random variables having respective *c.d.f.*  $F$  and  $G$ . The minimum in (2) is (1). To any non negative function  $c(x, y)$  from  $\mathbb{R}^2$  to  $\mathbb{R}$  let associate the Wasserstein type cost

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

We are interested in triplets  $(c, F, G)$  such that  $W_c(F, G)$  is finite and can be estimated by using an explicit CLT. To guaranty that an analogue of (1) exists we consider cost functions defining a negative measure on  $\mathbb{R}^2$ , hence satisfying

$$c(x', y') - c(x', y) - c(x, y') + c(x, y) \leq 0, \quad x \leq x', y \leq y'. \quad (4)$$

If  $c$  satisfies (4) then for any functions  $a$  and  $b$ ,  $a(x) + b(y) + c(x, y)$  satisfies (4). In particular  $c(x, y) = -xy$  and  $(x - y)^2 = x^2 + y^2 - 2xy$  satisfy (4). More generally if  $\rho$  is a convex real function then  $c(x, y) = \rho(x - y)$  satisfies (4). Two important special cases are the symmetric power functions  $|x - y|^p$ ,  $p \geq 1$ , associated to  $W_p$  and the non-symmetric step functions  $c(x, y) = (x - y)(\alpha - \mathbf{1}_{x - y < 0})$  associated to the  $\alpha^{th}$  quantile,  $0 < \alpha < 1$ . The following result yields the analogue of (1).

**Theorem 1 (Cambanis, Simon, Stout [4])** *If  $c$  satisfies (4) and  $U$  is a random variable uniformly distributed on  $[0, 1]$ , then the minimum in (3) is*

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

Let  $(X_i, Y_i)_{1 \leq i \leq n}$  be an *i.i.d.* sample of a random vector with joint *c.d.f.*  $H$  on  $\mathbb{R}^2$  and marginal *c.d.f.*  $F$  and  $G$  on  $\mathbb{R}$ . Write  $\mathbb{F}_n$  and  $\mathbb{G}_n$  the random empirical *c.d.f.* built from the two marginal samples. Thus  $\mathbb{F}_n$  and  $\mathbb{G}_n$  are not independent in general. Consider a cost function  $c$  satisfying (4). Let  $X_{(i)}$  (resp.  $Y_{(i)}$ ) denote the  $i^{th}$  order statistic of the sample  $(X_i)_{1 \leq i \leq n}$  (resp.  $(Y_i)_{1 \leq i \leq n}$ ), i.e.  $X_{(1)} \leq \dots \leq X_{(n)}$ . By Theorem 1,

$$W_c(\mathbb{F}_n, \mathbb{G}_n) = \frac{1}{n} \sum_{i=1}^n c(X_{(i)}, Y_{(i)}). \quad (5)$$

is a natural estimator of  $W_c(F, G)$ . Notice that  $W_c(F, G)$  does not depend on the generally unknown  $H$  whereas  $W_c(\mathbb{F}_n, \mathbb{G}_n)$  strongly depends on  $H$  through its distribution. In [2] our main result was a central limit theorem,

$$\sqrt{n} (W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)) \rightarrow_{weak} \mathcal{N}(0, \sigma^2)$$

when the tails of  $F$  and  $G$  differ from at least  $\tau > 0$  and  $c(x, y)$  is asymptotically  $\rho(x - y)$  with  $\rho$  non-negative, symmetric, convex. The influence of  $H$  only appeared in the limiting variance  $\sigma^2 = \sigma^2(c, H)$  together with  $c$ . The main contribution was to work out almost minimal sufficient conditions relating  $(c, F, G)$  explicitly by controlling carefully the underlying joint quantile processes and their extremes. We now intend to complete the picture by extending this result to other natural cases, in particular  $\tau = 0$  and non symmetric costs  $\rho$ . The sufficient conditions relating  $\rho$  to the tails of  $F$  and  $G$  are then explored more deeply and kept rather close to the minimal requirement that the weak convergence of  $W_c(\mathbb{F}_n, \mathbb{G}_n)$  occur at an explicit rate and non degenerate limit distribution.

### 1.3 Overview

Hereafter we consider  $c(x, y) = \rho_c(x - y)$  for a non-negative real convex function  $\rho_c$  with  $\rho_c(0) = 0$ , and not assumed to be symmetric. In the spirit of [2] we separate out three sets of assumptions, labeled  $(FG)$ ,  $(C)$  and  $(CFG)$  respectively.

First,  $(FG)$  concerns the regularity and tails of  $F$  and  $G$ , and especially their density-quantile function. Conditions  $(FG)$  are satisfied by distributions having regular tails, among which all classical probability laws.

Second,  $(C)$  is about the rate of increase at infinity and the regular variation at 0 of  $\rho_c$ , without assuming differentiability at 0. Conditions  $(C)$  encompass a large class of Wasserstein costs  $c$  and the distance  $W_1$  is now allowed, together with non-symmetric variants of Wasserstein distances  $W_p$ ,  $p \geq 1$ , with slowly varying factors. The non-symmetry in particular impacts the case  $\tau = 0$  now included in  $(FG)$  and the situation just around the diagonal  $F = G$  where the random sign and amplitude of  $\mathbb{F}_n - \mathbb{G}_n$  can may oscillate.

The conditions  $(FG)$  and  $(C)$  are thus designed to separately select a class of probability distributions and admissible costs.

The third set  $(CFG)$  aims to mix the requirements on  $(c, F, G)$  making them compatible. We distinguish between  $(CFG_E)$ ,  $(CFG_D)$  and  $(CFG_{ED})$  depending on the situations  $\{F = G\} = \mathbb{R}$  or  $\{F \neq G\} = \mathbb{R}$  or  $\{F = G\} \neq \mathbb{R}$  and  $\{F \neq G\} \neq \mathbb{R}$ , respectively. The joint distribution  $H$  of the samples is not restricted and again only affects the limiting distributions. Under  $(CFG_E)$  and  $(CFG_{ED})$  the tail constraints on  $F$  and  $G$  also depend on the regular variation of  $\rho_c$  at 0, which determines the rate of convergence. This is to maintain the tails of  $\mathbb{F}_n$  and  $\mathbb{G}_n$  sufficiently close with high probability.

When dealing with empirical Wasserstein type integrals, in order to apply the usual delta method one would need to truncate and then to assume a convergence in probability of the extremal parts. This would be a restriction excluding many distributions  $F$  and  $G$ , depending of where the integral is non-adaptively cut. Moreover, the validity of the convergence of the extremal parts would bother the practitioner with variants of steps 1, 2, 3 of our proofs. In contrast, our key assumptions  $(CFG_E)$  and  $(CFG_D)$  explicitly relate the tails to the cost in such a way that truncation levels depending on the underlying distribution can be defined efficiently in the forthcoming proofs.

The paper is organized as follows. Assumptions are discussed in Section 2. In Section 3 we state our main results in the form of CLT for  $W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)$  at various rates. We propose some applications and perspectives in Section 4. All the results are proved in Section 5.

## 2 Assumptions

### 2.1 Assumptions $(FG)$

Consider a sequence  $(X_n, Y_n) \in \mathbb{R}^2$  of independent random vectors having the same *c.d.f.*  $H$  as  $(X, Y)$ . The distribution  $H$  may have a density or not. However we assume that the marginal *c.d.f.'s*  $F$  of  $X$  and  $G$  of  $Y$  have support  $\mathbb{R}$  and positive densities  $f = F'$  and  $g = G'$ . Let  $(E, D)$  be the partition of  $(0, 1)$  defined by

$$E = \{u : F^{-1}(u) = G^{-1}(u)\}, \quad D = \{u : F^{-1}(u) \neq G^{-1}(u)\}. \quad (6)$$

If  $u$  shifts infinitely many times between  $E$  and  $D$  it becomes difficult to control the stochastic integral  $W_c(\mathbb{F}_n, \mathbb{G}_n)$ . The case where  $|F^{-1}(u) - G^{-1}(u)| > \tau > 0$  as  $u \rightarrow 1$  and  $u \rightarrow 0$  has been treated in details in [2]. We would like to allow the diagonal  $|F^{-1}(u) - G^{-1}(u)| \leq \tau$  and thus encompass the case  $E = (0, 1)$  together with some tractable situations where  $E \neq \emptyset$  and  $D \neq \emptyset$ . Let assume that there exists a finite integer  $\kappa \geq 2$  and  $0 = u_0 < u_1 < \dots < u_\kappa = 1$  such that, writing  $A_k = (u_{k-1}, u_k)$ ,

$$(FG0) \quad F^{-1}(u_k) = G^{-1}(u_k) \text{ and } A_k \subset E \text{ or } A_k \subset D, \text{ for } k = 1, \dots, \kappa.$$

This covers three generic cases, namely  $E = (0, 1)$ ,  $D = (0, 1)$  and when at least one interval is included in  $D$  while  $E \neq \emptyset$ . The exponential rate of decrease of the right and left tails of  $F$  and  $G$  are defined to be, for  $x \in \mathbb{R}_+$ ,

$$\begin{aligned} \psi_X^+(x) &= -\log \mathbb{P}(X > x), & \psi_Y^+(x) &= -\log \mathbb{P}(Y > x), \\ \psi_X^-(x) &= -\log \mathbb{P}(X < -x), & \psi_Y^-(x) &= -\log \mathbb{P}(Y < -x). \end{aligned}$$

Only  $\psi_X^+$  and  $\psi_Y^+$  will be considered in subsequent proofs where arguments given for the right hand tail  $u \rightarrow 1$  in the integrals  $W_c(F, G)$  and  $W_c(\mathbb{F}_n, \mathbb{G}_n)$  work similarly for  $u \rightarrow 0$ . Lastly, define the density quantile functions

$$h_X = f \circ F^{-1}, \quad h_Y = g \circ G^{-1},$$

and assume that

$$(FG1) \quad F, G \in \mathcal{C}_2(\mathbb{R}), \quad f, g > 0 \text{ on } \mathbb{R}.$$

$$(FG2) \quad \sup_{0 < u < 1} \min(u, 1 - u) |(\log h(u))'| < +\infty \quad \text{for } h = h_X, h_Y.$$

$$(FG3) \quad \sup_{0 < u < 1} \frac{\min(u, 1 - u)}{(|\Gamma^{-1}(u)| + 1) h(u)} < +\infty \quad \text{for } (h, \Gamma) = (h_X, F) \text{ or } (h_Y, G).$$

**Remark 2** *Rewriting (FG2) and (FG3) we get*

$$\sup_{x \in \mathbb{R}} \frac{\min(F(x), 1 - F(x))}{f(x)} \left( \frac{1}{|x| + 1} + \frac{|f'(x)|}{f(x)} \right) < +\infty,$$

$$\sup_{x \in \mathbb{R}} \frac{\min(G(x), 1 - G(x))}{g(x)} \left( \frac{1}{|x| + 1} + \frac{|g'(x)|}{g(x)} \right) < +\infty.$$

In Section 5.2.1 of [2] we provided a simple sufficient condition for (FG1), (FG2), (FG3) based on the regular variation of  $\psi_X^\pm$  and  $\psi_Y^\pm$ .

## 2.2 Notation for regularity

To specify the allowed cost functions  $c(x, y)$  the following definitions are required. As usual for  $k \in \mathbb{N}_*$  and  $I \subset \mathbb{R}$  let  $\mathcal{C}_k(I)$  denote the set of functions that are  $k$  times continuously differentiable on  $I$  and  $\mathcal{C}_0(I)$  the set of continuous functions on  $I$ . In forthcoming assumptions and proofs we consider functions defined either on  $(0, x_0)$  or on  $(y_0, +\infty)$  for some  $0 < x_0 < y_0$ . We distinguish the two domains by using a variable  $x \rightarrow 0$  and a variable  $y \rightarrow +\infty$ . In [2] only large values  $y \in (y_0, +\infty)$  played a role in terms of regular variation, so that we keep the same setting in (i) below. Unexpectedly, it turns out that the two domains interfere when  $|F - G|$  is arbitrarily small, and we need (ii).

(i) **Regularity on  $(y_0, +\infty)$ .** Let  $\mathcal{M}_2((y_0, +\infty))$  be the subset of functions  $l \in \mathcal{C}_2((y_0, +\infty))$  such that  $l''$  is monotone on  $(y_0, +\infty)$ . Write  $RV(+\infty, \gamma)$  the set of regularly varying functions at  $+\infty$  with index  $\gamma \geq 0$ . If  $\gamma = 0$  we restrict ourselves to slowly varying functions  $L$  at  $+\infty$  such that

$$L'(y) = \frac{\varepsilon(y)L(y)}{y}, \quad \lim_{y \rightarrow +\infty} \varepsilon(y) = 0. \quad (7)$$

This weak restriction is explained at Section 5 of [2]. In order to find distributions  $F$  and  $G$  compatible with the cost  $c$  we further impose

$$L'(y) \geq \frac{l_1}{y}, \quad l_1 \geq 1, \quad y \geq y_0. \quad (8)$$

For  $\gamma = 0$ , introduce

$$RV_2(+\infty, 0) = \{L : L \in \mathcal{M}_2((y_0, +\infty)) \text{ such that (7), (8) hold}\}$$

and for  $\gamma > 0$ ,

$$RV_2(+\infty, \gamma) = \{l : l \in \mathcal{M}_2((y_0, +\infty)), l(y) = y^\gamma L(y) \text{ such that } L' \text{ obeys (7)}\}.$$

(ii) **Regularity on  $(0, x_0)$ .** We consider positive slowly varying functions  $L$  at 0,

$$\lim_{x \searrow 0} \frac{L(\theta x)}{L(x)} = 1 \text{ for any } \theta > 0. \quad (9)$$



For  $b > 1$  let introduce

$$RV_2(0, b) = \{ \rho : L \in \mathcal{C}_2((0, x_0)), \rho(x) = x^b L(x) \text{ such that } L \text{ satisfies (9)} \}.$$

For  $b = 1$  let define

$$RV_2(0, 1) = \{ \rho : L \in \mathcal{C}_2((0, x_0)), \rho(x) = xL(x) \text{ such that } L \text{ satisfies (9), (10)} \}$$

where we impose the following finite limit

$$\lim_{x \searrow 0} L(x) = L(0) \in \mathbb{R}_+. \quad (10)$$

### 2.3 Assumptions (C)

We consider smooth Wasserstein costs such that, for some  $0 < x_0 < y_0 < +\infty$ ,

- (C0)  $c(z, z') = \rho_c(z - z') \geq 0$ ,  $z, z' \in \mathbb{R}$ ,  $c(0, 0) = 0$ ,  $\rho_c$  is convex.
- (C1)  $\rho_c(z) = \rho_-(-z)1_{z \leq 0} + \rho_+(z)1_{z \geq 0}$ ,  $z \in \mathbb{R}$ ,  $\rho_{\pm} \in \mathcal{C}_2((0, +\infty))$ .
- (C2)  $\rho_+(x) = x^{b_+} L_+(x) > 0$ ,  $0 < x \leq x_0$ ,  $\rho_+ \in RV_2(0, b_+)$ ,  $b_+ \geq 1$ ,  
 $\rho_-(x) = x^{b_-} L_-(x) > 0$ ,  $0 < x \leq x_0$ ,  $\rho_- \in RV_2(0, b_-)$ ,  $b_- \geq 1$ .
- (C3)  $\rho_+(y) = \exp(l_+(y))$ ,  $y \geq y_0$ ,  $l_+ \in RV_2(+\infty, \gamma_+)$ ,  $\gamma_+ \geq 0$ ,  
 $\rho_-(y) = \exp(l_-(y))$ ,  $y \geq y_0$ ,  $l_- \in RV_2(+\infty, \gamma_-)$ ,  $\gamma_- \geq 0$ .

Notice that  $\rho_{\pm}(0) = 0$  and  $\rho_{\pm}$  are positive, continuous, convex and increasing on  $\mathbb{R}_+$ . Define  $\rho(x) = \max(\rho_+(x), \rho_-(x))$  and  $b = \min(b_+, b_-)$ . For  $0 \leq x \leq x_0$  it holds

$$\rho(x) = x^b L(x), \quad L(x) = \begin{cases} L_+(x) & \text{if } b_+ < b_-, \\ L_-(x) & \text{if } b_- < b_+, \\ \max(L_+(x), L_-(x)) & \text{if } b_+ = b_-. \end{cases} \quad (11)$$

Further assume that

$$(C4) \quad \lim_{x \rightarrow 0} \frac{\rho_+(x)}{\rho(x)} = \pi_+, \quad \lim_{x \rightarrow 0} \frac{\rho_-(x)}{\rho(x)} \rightarrow \pi_-, \quad \pi_+, \pi_- \in [0, 1].$$

**Example 3** *Typical costs satisfying the conditions (C) are the following. For  $a = (a_-, a_+)$  with  $a_{\pm} > 0$  and  $b = (b_-, b_+)$  with  $b_{\pm} \geq 1$  define*

$$c_{a,b}(z, z') = a_- (z' - z)^{b_-} 1_{z < z'} + a_+ (z - z')^{b_+} 1_{z' < z}.$$

*This includes the case  $W_1$ .*

### 2.4 Assumptions (CFG)

The joint influence of  $l_{\pm}$ ,  $L_{\pm}$  and  $b_{\pm}$  on the allowed tails  $F^{-1}$  and  $G^{-1}$  is expressed as follows. Remind the sets  $E$  and  $D$  from (6). We need three different assumptions, each corresponding to the generic cases  $E = (0, 1)$ ,  $D = (0, 1)$  and when at least one interval is included in  $E$  and one in  $D$ .

**Assumption** ( $CFG_E$ ). In the case  $E = (0, 1)$  it is necessary that  $b_- < 2$  and  $b_+ < 2$  to handle most of the classical distributions. For some  $\theta_2 > 0$  and

$$(l, \psi) \in \{(l_+, \psi_X^+), (l_-, \psi_Y^+), (l_-, \psi_X^-), (l_+, \psi_Y^-)\} \quad (12)$$

we have, if  $1 < b < 2$ ,

$$l \circ \psi^{-1}(y) \leq \left(1 - \frac{b}{2}\right) y + \log L(\exp(-y/2)) - 2 \log \psi^{-1}(y) - \theta_2 \log y, \quad y > y_0, \quad (13)$$

and, if  $b = 1$ ,

$$l \circ \psi^{-1}(y) \leq \frac{y}{2} - 2 \log \psi^{-1}(y) - \theta_2 \log y, \quad y > y_0. \quad (14)$$

We shall also allow  $b_- = b_+ = 2$  for strictly sub-Gaussian distributions, by (20).

**Assumption** ( $CFG_D$ ). In the case  $D = (0, 1)$  let  $\theta_-, \theta_+ > 1$  be the parameter  $\theta$  of condition ( $CFG$ ) in [2] for the left and right tails respectively. For any  $(l, \psi)$  from (12) and  $\theta = \theta_+$  if  $l = l_+$  or  $\theta = \theta_-$  if  $l = l_-$  we have

$$(\psi \circ l^{-1})'(y) \geq 2 + \frac{2\theta}{y}, \quad y > y_0. \quad (15)$$

In other words, (15) is exactly ( $CFG$ ) in [2]. If

$$\liminf_{u \rightarrow 1} |F^{-1}(u) - G^{-1}(u)| = 0 \quad \text{or} \quad \liminf_{u \rightarrow 0} |F^{-1}(u) - G^{-1}(u)| = 0$$

we further impose, respectively to  $(l_+, \psi_X^+), (l_-, \psi_Y^+)$  or  $(l_-, \psi_X^-), (l_+, \psi_Y^-)$ , that for some  $\theta_2 > 0$  it holds

$$l \circ \psi^{-1}(y) \leq \frac{y}{2} - 2 \log \psi^{-1}(y) - \theta_2 \log y, \quad y > y_0. \quad (16)$$

**Assumption** ( $CFG_{ED}$ ). In the case  $D \neq \emptyset$  assume ( $CFG_D$ ) and, if moreover  $(A_1 \cup A_\kappa) \cap E \neq \emptyset$  then assume ( $CFG_E$ ) also.

**Remark 4** If  $\gamma_\pm > 0$  we have  $\theta_\pm > 2$  and, if  $\gamma_\pm = 0$  we have, as in [2],

$$\theta^\pm > 2 - \liminf_{y \rightarrow +\infty} \frac{\log(1/\varepsilon_\pm(y))}{\log l_\pm(y)}$$

where  $\varepsilon_\pm(y)$  corresponds to the function  $\varepsilon(y)$  of (7) applied to  $L(y) = l_\pm(y)$ .

**Remark 5** As will be seen in the proofs, ( $CFG_E$ ) and ( $FG3$ ) imply that we can find  $1 \leq b < b' < 2$  such that

$$\int_0^1 \left( \frac{\sqrt{u(1-u)}}{h_X(u)} \right)^{b'} du \leq \int_0^1 \left( \frac{|F^{-1}(u)|}{\sqrt{u(1-u)}} \right)^{b'} du < +\infty \quad (17)$$

which is a little stronger than the necessary condition that the left hand integral is finite for  $b' = b$ . By using  $F^{-1}(u) = \psi^{-1}(\log(1/(1-u)))$ , (13) also reads

$$(F^{-1}(u))^2 \rho(F^{-1}(u)) \leq \frac{L(\sqrt{1-u})}{(1-u)^{1-b/2}(\log(1/(1-u)))^{\theta_2}}, \quad u > u_0.$$

In particular, if  $L(x) = 1$  we deduce that  $(FG)$  and  $(CFG_E)$  imply

$$\mathbb{P}(X > y) \leq \left( \frac{1}{y^2 \rho(y)} \right)^{2/(2-b)}, \quad y > y_0.$$

**Example 6** For light tails of Weibull type  $\psi(y) = y^w$ ,  $w > 0$ , (17) is true and  $(CFG_E)$  requires that  $l(y^{1/w}) < Cy$  as  $y \rightarrow +\infty$  and hence a cost of type  $l(y) = y^\gamma$ ,  $y > y_0$  and  $l(x) = x^b$ ,  $x < x_0$ , is allowed provided that  $\gamma < w$  and  $1 \leq b < 2$ . For heavy tailed distributions such as Pareto  $\psi(y) = p \log y$  with index  $p > 2$  the conditions  $(CFG_E)$ ,  $(CFG_D)$  and  $(CFG_{ED})$  induce more constraints. For instance  $(CFG_E)$  applied with  $\rho_c(x) = x^b$ ,  $x < x_0$ , and  $l(y) = \alpha \log y$ ,  $y > y_0$ , implies that  $p > 4/(2-b)$  and  $1 \leq \alpha < p(1-b/2-2/p)$ , hence the minimal requirement on  $p$  is  $p \geq 6/(2-b)$ . Choosing  $\rho_c(x) = x^b$  on  $\mathbb{R}_+$  we have  $\alpha = b$  and the last constraint becomes  $p > 2(b+2)/(2-b)$ .

### 3 Statement of the results

Consider the joint Gaussian process  $\mathcal{G} = \{(\mathbb{B}^X(u), \mathbb{B}^Y(u)) : u \in (0, 1)\}$  with

$$\mathbb{B}^X(u) = \frac{B^X(u)}{h_X(u)}, \quad \mathbb{B}^Y(u) = \frac{B^Y(u)}{h_Y(u)}, \quad (18)$$

where  $(B^X, B^Y)$  are two standard Brownian bridges with covariance

$$\text{cov}(B^X(u), B^X(v)) = \text{cov}(B^Y(u), B^Y(v)) = \min(u, v) - uv, \quad u, v \in (0, 1),$$

and cross covariance

$$\text{cov}(B^X(u), B^Y(v)) = H(F^{-1}(u), G^{-1}(v)) - uv, \quad u, v \in (0, 1).$$

The existence of  $\mathcal{G}$  is proved in [2]. Let  $\mathbb{B}(u) = \mathbb{B}^X(u) - \mathbb{B}^Y(u)$ ,  $u \in (0, 1)$ .

We are now ready to state our main results. Remind (11) and set

$$v_n = \frac{1}{\rho(1/\sqrt{n})} = \frac{n^{b/2}}{L(1/\sqrt{n})} \quad (19)$$

hence, in our first statement we have  $K\sqrt{n} \leq v_n = o(n)$  for some  $K > 0$ . The constants  $\pi_-$  and  $\pi_+$  come from (C4).

**Theorem 7** Assume  $(FG)$ ,  $(C)$ ,  $E = (0, 1)$  and  $(CFG_E)$ . Then

$$v_n W_c(\mathbb{F}_n, \mathbb{G}_n) \rightarrow_{weak} \pi_- \int_0^1 1_{\{\mathbb{B}(u) < 0\}} |\mathbb{B}(u)|^{b-} du + \pi_+ \int_0^1 1_{\{\mathbb{B}(u) > 0\}} |\mathbb{B}(u)|^{b+} du$$

and the limiting random variable is positive and finite.

**Remark 8** As shown in [6], and since  $\mathbb{B}^X$  is a centered Gaussian process,

$$\mathbb{P}\left(\int_0^1 |\mathbb{B}^X(u)|^b du < +\infty\right) = 1 \text{ is equivalent to } \int_0^1 \left(\frac{\sqrt{u(1-u)}}{h_X(u)}\right)^b du < +\infty.$$

Since the latter bound is guaranteed by  $(CFG_E)$  and  $(FG3)$  which imply (17), the finiteness of the limiting random variable in Theorem 7 follows.

Our second statement is an extension of the main theorem of [2] which now allows  $F$  and  $G$  to have arbitrarily close tails.

**Theorem 9** Assume  $(FG)$ ,  $(C)$ ,  $D = (0, 1)$  and  $(CFG_D)$ . Then

$$\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)) \rightarrow_{weak} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \mathbb{E}\left(\left(\int_0^1 \rho'_c(F^{-1}(u) - G^{-1}(u))\mathbb{B}(u)du\right)^2\right) < +\infty.$$

**Remark 10** The finiteness and a closed form expression for  $\sigma^2 = \sigma^2(c, H)$  have been proved in [2]. We also refer to the latter paper for explicit examples in the independent samples case.

Our third result shows that if there exists a point, or equivalently an open interval by  $(FG)$ , where  $F \neq G$  then the rate is  $\sqrt{n}$ , whether  $E \neq \emptyset$  or not.

**Theorem 11** Assume  $(FG)$ ,  $(C)$ ,  $D \neq \emptyset$  and  $(CFG_{ED})$ . If  $1 < b < 2$  then

$$\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)) \rightarrow_{weak} \mathcal{N}(0, \sigma_D^2)$$

where

$$\sigma_D^2 = \mathbb{E}\left(\left(\int_D \rho'_c(F^{-1}(u) - G^{-1}(u))\mathbb{B}(u)du\right)^2\right) < +\infty.$$

If  $b = 1$  then, for  $L_{\pm}(0)$  from (10),

$$\begin{aligned} \sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)) &\rightarrow_{weak} \int_D \rho'_c(F^{-1}(u) - G^{-1}(u))\mathbb{B}(u)du \\ &\quad + 1_{\{b_- = 1\}}L_-(0) \int_E 1_{\{\mathbb{B}(u) < 0\}} |\mathbb{B}(u)| du \\ &\quad + 1_{\{b_+ = 1\}}L_+(0) \int_E 1_{\{\mathbb{B}(u) > 0\}} |\mathbb{B}(u)| du. \end{aligned}$$

**Remark 12** In the second part of Theorem 11 the first term in the limiting random variable has distribution  $\mathcal{N}(0, \sigma_D^2)$  and is correlated in an explicit way to the other two terms. Theorem 11 also shows that whenever  $1 < b < 2$  Theorem 9 remains true if  $F$  and  $G$  are not stochastically ordered but cross each other at a finite number of points, since this implies  $\sigma_D^2 = \sigma^2$ .

The next corollary concerns the  $L_1$ -distance  $W_1(\mathbb{F}_n, \mathbb{G}_n) = \|\mathbb{F}_n - \mathbb{G}_n\|_{L_1}$ . Remind that  $c_{a,1}(z, z') = a_- (z' - z) 1_{z < z'} + a_+ (z - z') 1_{z' < z}$ .

**Corollary 13** *Assume  $(FG)$ ,  $(C)$  and  $(CFG_{ED})$ . Then*

$$\begin{aligned} \sqrt{n} (W_{c_{a,1}}(\mathbb{F}_n, \mathbb{G}_n) - W_{c_{a,1}}(F, G)) &\rightarrow_{weak} \int_D (a_- 1_{\{F(u) < G(u)\}} + a_+ 1_{\{F(u) > G(u)\}}) \mathbb{B}(u) du \\ &+ \int_E (a_- 1_{\{\mathbb{B}(u) < 0\}} + a_+ 1_{\{\mathbb{B}(u) > 0\}}) |\mathbb{B}(u)| du \end{aligned}$$

and, in particular for  $a_- = a_+ = 1$ ,

$$\sqrt{n} (W_1(\mathbb{F}_n, \mathbb{G}_n) - W_1(F, G)) \rightarrow_{weak} \int_D \mathbb{B}(u) du + \int_E |\mathbb{B}(u)| du.$$

In the case  $F = G$  our last result shows that for light tails one can handle the limiting case  $b = 2$  – here stated with  $L(x) = 1$  for  $x < x_0$  for sake of simplicity.

**Theorem 14** *Assume that  $E = \mathbb{R}$ ,  $(FG1)$ ,  $(FG2)$  and*

$$\lim_{u \rightarrow 0} \frac{u}{h(u)} = \lim_{u \rightarrow 1} \frac{1-u}{h(u)} = 0, \quad \int_0^1 \frac{u(1-u)}{h^2(u)} du < +\infty. \quad (20)$$

Moreover assume  $(C0)$  with  $\rho_c(x) = x^2$  for  $|x| \leq x_0$ . Then

$$nW_c(\mathbb{F}_n, \mathbb{G}_n) \rightarrow_{weak} \int_0^1 \mathbb{B}(u)^2 du.$$

**Remark 15** *Theorem 14 requires no assumption on the cost  $\rho(y)$  as  $y \rightarrow +\infty$  since only sub-Gaussian laws are allowed by (20). Therefore the tail part of  $nW_c(\mathbb{F}_n, \mathbb{G}_n)$  behaves the same as for compactly supported distributions. Namely, empirical extremes of both samples remain simultaneously stuck together very closely to their common deterministic counterpart  $F^{-1}$  that increases slowly.*

**Example 16** *For light tails of Weibull type it holds, for some  $w > 0$ ,*

$$h(u) = w(1-u) (\log(1/(1-u)))^{1-1/w}$$

and  $(1-u)/h^2(u) = 1/w ((1-u) \log(1/(1-u)))^{2(1-1/w)}$ . The first condition in (20) is then satisfied for  $w > 1$  and the second for  $w > 2$ , so that  $w > 2$  is required. This excludes Gaussian tails, as in Theorem 4.6 in [10].

It is easily seen that straightforward adaptations of the proof of Theorems 7 to 14 leads to analog results for  $\sqrt{n} (W_c(\mathbb{F}_n, G) - W_c(F, G))$  and  $v_n W_c(\mathbb{F}_n, F)$  by just replacing  $\mathbb{B}(u) = \mathbb{B}^X(u) - \mathbb{B}^Y(u)$  with  $\mathbb{B}^X(u)$ .

## 4 Applications

### 4.1 Comparison and goodness-of-fit tests

A straightforward consequence of Theorems 7 and 11 is the construction of a statistical test of the hypothesis  $\mathcal{H}_0 : F = G$  against  $\mathcal{H}_1 : F \neq G$ , based on two samples that may arise from correlated experiments. The distributions  $F$  and  $G$  are supposed to be  $C^2$  on  $\mathbb{R}$  or  $\mathbb{R}^+$  for instance and satisfy  $(CFG_{ED})$ . Let us choose the  $b$ -Wasserstein distance with  $1 < b < 2$ . By Theorem 7, under  $\mathcal{H}_0$  the statistic  $n^{b/2}W_c(\mathbb{F}_n, \mathbb{G}_n)$  converges to a positive finite random variable while by Theorem 11, under  $\mathcal{H}_1$  it converges almost surely to  $+\infty$  at the rate  $n^{b/2}W_c(F, G)$ . Mathematically this test is effectively valid when the set  $D = \{F^{-1} \neq G^{-1}\}$  is a finite union of non empty intervals, but we think that its validity could be extended to the more general case where  $D$  is of positive Lebesgue measure in  $(0, 1)$ . The use of  $W_2$ , with a rate  $n$  is more restrictive since it needs very light tails. Nevertheless if sub-Gaussian tails can be asserted, by Theorem 14 the previous test works with  $b = 2$ , which actually is a new test.

In each case the rather minimal  $(CFG)$  type conditions have to be checked. They are almost necessary in the proofs to overcome the difficulty of controlling how close the empirical tails of  $\mathbb{F}_n$  and  $\mathbb{G}_n$  must be under  $\mathcal{H}_0$ , and how far  $|\mathbb{F}_n - \mathbb{G}_n|$  can deviate from  $|F - G|$  in tails under  $\mathcal{H}_1$ . Interestingly the choice of  $\rho(x)$  may be with a locally polynomial shape as  $x \rightarrow 0$  and a different shape as  $x \rightarrow +\infty$  possibly linear, polynomial or exponential. This flexibility allows to test the tail or the mid-quantiles with more or less accuracy.

In the same vein, concerning the distribution functions, Corollary 1 yields

$$\begin{aligned} & \sqrt{n} \left( \int_{-\infty}^{+\infty} |\mathbb{F}_n(t) - \mathbb{G}_n(t)| dt - \int_{F^{-1}(D)} |F(t) - G(t)| dt \right) \\ & \rightarrow_{weak} \int_D \mathbb{B}(u) du + \int_E |\mathbb{B}(u)| du \end{aligned}$$

which seems not to have been already obtained. This provides weak limits for the power of the test under alternatives to  $\mathcal{H}_0 : F = G$  of the kind  $\mathcal{H}_1 : F = G_1$  where  $G_1^{-1}$  only differs from  $G^{-1}$  on an interval  $D$ , for instance with a slightly different right hand tail only. The test statistic  $\sqrt{n} \int_{-\infty}^{+\infty} |\mathbb{F}_n(t) - \mathbb{G}_n(t)| dt$  has an almost sure first order rate of escape  $\sqrt{n} \int_{G_1^{-1}(D)} |G_1(t) - G(t)| dt$ .

As a by-product of the results of Section 3 one can similarly build goodness-of-fit tests  $\mathcal{H}_0 : F = F_0$  against  $\mathcal{H}_1 : F \neq F_0$  by using one sample under  $F$  or by using an additional sample distributed as  $F_0$ . Notice that the test associated to  $b = 2$  was a consequence of [10].

### 4.2 An application

We conclude with a notion of quantile for a random variable taking values in the set of continuous *c.d.f.*'s. One important feature of this article is the first part

of Corollary 13 which is strongly related to the preprint [12]. Let  $0 < \alpha < 1$ . In [12] the  $\alpha$ -quantile  $F_\alpha$  of a random continuous *c.d.f.*  $\mathbb{F}$  is defined to be

$$F_\alpha = \underset{\theta \in \mathcal{F}}{\operatorname{Argmin}} \mathbb{E} W_{c_\alpha}(\mathbb{F}, \theta),$$

where  $c(x, y) = (x - y)(\alpha - \mathbf{1}_{x-y < 0})$  is the classical contrast for the  $\alpha$ -quantile of a real random variable and  $\mathcal{F}$  is the set of continuous *c.d.f.* As previously mentioned, in practice a realization  $\mathbb{F}(\omega)$  of  $\mathbb{F}$  is known through a  $n$ -sample of the distribution  $\mathbb{F}(\omega)$ . Hence we may assume that a  $N$ -sample  $\mathbb{F}_n^1, \dots, \mathbb{F}_n^N$  is available, where each  $\mathbb{F}_n^i$  is a  $n$ -empirical *c.d.f.* of  $\mathbb{F}^i$  and  $\mathbb{F}^1, \dots, \mathbb{F}^N$  are *i.i.d.* according to  $\mathbb{F}$ . Define

$$F_{N,n,\alpha} = \underset{\theta \in \mathcal{F}_n}{\operatorname{Argmin}} \frac{1}{N} \sum_{i=1}^N W_{c_\alpha}(\mathbb{F}_n^i, \theta),$$

where  $\mathcal{F}_n$  is the set of *c.d.f.* with at most  $n$  different values. Then one could use Corollary 13 to prove that  $F_{N,n,\alpha}$  is a consistent estimator of  $F_\alpha$  when  $N$  and  $n$  tend to  $+\infty$ , with rate of convergence  $\sqrt{Nn}$ .

## 5 Proofs

In the forthcoming proofs the high order quantiles are shown to have a secondary order impact compared to the mid-order quantiles that impose the rate as well as the limiting law under our sufficient conditions ensuring that the tails are not too heavy. For sake of simplicity we only work on the right hand tail, with quantiles of order  $u \in (\underline{u}, 1)$  for an arbitrary small  $\underline{u} > 0$ . The counterpart for the left hand tail is immediate by using the same arguments.

To help the reader the variable of frequently used deterministic functions defined on  $\mathbb{R}_+$  like  $\rho_\pm$ ,  $\rho_\pm^{-1}$ ,  $l_\pm$ ,  $l_\pm^{-1}$  or  $L_\pm$  is denoted  $x$  when considered as  $x \rightarrow 0$  and  $y$  when considered as  $y \rightarrow +\infty$ . In the subsequent proofs the constant  $K > 0$  may change at each appearance.

In steps numbered 0 we remind active hypotheses while introducing local notation. The non standard steps 1, 2 and 3 of the four proofs – including the one in [2] – are designed to address the non trivial problem of controlling the high order and extreme order quantiles under an explicit and almost minimal assumption on tails, namely  $(CFG_E)$ ,  $(CFG_D)$  or  $(CFG_{ED})$ . The secondary order terms in these conditions could be balanced slightly more sharply but at the price of adding technicalities to connect steps 1 and 2. Finally we point out that the convergence at steps 3 is weaker than in probability, due to the coupling approach.

### 5.1 The case $F = G$

We prove Theorem 7.

**Step 0.** In this section  $F = G$  and hence  $E = \mathbb{R}$ . For short, the key functions common to  $X, Y$  are denoted  $F^{-1}$ ,  $\psi$ ,  $H$  and  $h$ . Let assume  $(FG)$ ,  $(C)$  and  $(CFG_E)$  with  $1 \leq b_{\pm} < 2$  in  $(C2)$ . Hence  $\rho(x) = \max(\rho_+(x), \rho_-(x)) \geq \rho_c(x)$  and  $\rho_{\pm}(x)$  are positive convex increasing functions defined on  $\mathbb{R}_+^*$  with  $\rho_{\pm}(0) = 0$ . For  $0 \leq x \leq x_0$  we have  $\rho_{\pm}(x) = x^{b_{\pm}} L_{\pm}(x)$  and, whenever  $b_{\pm} = 1$  it is also assumed through (10) that  $\lim_{x \rightarrow 0} L_{\pm}(x) = L_{\pm}(0) < +\infty$ . Recall that  $b = \min(b_+, b_-)$  and, for  $0 \leq x \leq x_0$ ,  $\rho(x) = \max(\rho_+(x), \rho_-(x)) = x^b L(x)$  where  $L(x)$  is defined at (11) and is slowly varying as  $x \rightarrow 0$ . We then have

$$v_n = \frac{1}{\rho(1/\sqrt{n})}, \quad \lim_{n \rightarrow +\infty} \frac{\sqrt{n}}{v_n} = 1_{\{b=1\}} L(0).$$

Since  $L \in RV(0, 0)$  we have, by the Karamata representation theorem,

$$L(x) = \exp \left( \eta(x) + \int_B^{1/x} \frac{s(y)}{y} dy \right), \quad 0 < x \leq x_0, \quad (21)$$

with  $B > 0$ ,  $\eta(x)$  and  $s(y)$  are bounded measurable functions such that

$$\lim_{x \rightarrow 0} \eta(x) = \eta_{\infty} \in \mathbb{R}, \quad \lim_{y \rightarrow +\infty} s(y) = 0.$$

We can then define

$$\eta_0 = \sup_{0 < x \leq x_0} |\eta(x)| \in \mathbb{R}_+, \quad c_0 = e^{2\eta_0} \geq 1. \quad (22)$$

For  $y$  large it holds  $\rho_{\pm}(y) = \exp(l_{\pm}(y))$  where the functions  $l_{\pm}(y)$  are not asked to be in  $RV(+\infty, \gamma_{\pm})$  in this proof, but (7) does matter. However in practice if  $(C3)$  would not hold then  $(CFG_E)$  would be more difficult to translate in terms of admissible  $F$ . Hence, for some  $y_0 > x_0$ ,

$$\rho(y) = \exp(l(y)), \quad l(y) = \max(l_+(y), l_-(y)), \quad y \geq y_0.$$

Since  $\rho_{\pm}$  and  $\rho$  are convex, by (7) there exists  $d_{\pm} \geq 1$ ,  $d = \min(d_-, d_+)$  and  $d_{0,\pm}$ ,  $d_0$  such that

$$l_{\pm}(y) \geq d_{\pm} \log y + d_{0,\pm}, \quad l(y) \geq d \log y + d_0, \quad y \geq x_0. \quad (23)$$

By  $(CFG_E)$ , the joint influence of  $l$ ,  $L$  and  $b$  on the allowed tails  $F^{-1}$  is expressed at (13) if  $b > 1$  and (14) if  $b = 1$ .

We decompose the integral  $W_c(\mathbb{F}_n, \mathbb{G}_n)$  as follows, with the remainder term implicitly treated in a similar way with left hand tails. We will specify later two sequence  $i_n$  and  $j_n$  of positive integers such that  $n > j_n > i_n \rightarrow +\infty$ . The proof consists in four steps, each dealing with one of the four terms in

$$W_c^E(\mathbb{F}_n, \mathbb{G}_n) = I_{\mathcal{I}_n} + I_{\mathcal{J}_n} + I_{\mathcal{K}_n} + I_{\mathcal{L}}, \quad I_A = \int_A \rho_c(\mathbb{F}_n^{-1}(u) - \mathbb{G}_n^{-1}(u)) du, \quad (24)$$

where  $\mathcal{I}_n = (1 - i_n/n, 1]$ ,  $\mathcal{J}_n = (1 - j_n/n, 1 - i_n/n]$ ,  $\mathcal{K}_n = (\bar{u}, 1 - j_n/n]$ ,  $\mathcal{L} = [\underline{u}, \bar{u}]$  and  $0 < \underline{u} < 1/2 < \bar{u} < 1$ . In order to accurately choose  $i_n$  and  $j_n$  one has



to take into account two difficulties. First, the rate  $1/v_n$  is faster than  $1/\sqrt{n}$  so that  $\mathcal{I}_n \cup \mathcal{J}_n$  should be sufficiently small. Second, the empirical extreme quantile difference  $\mathbb{F}_n^{-1}(u) - \mathbb{G}_n^{-1}(u)$  may be either very large or very small as  $u \rightarrow 1$ , thus the cost function  $\rho_c(\mathbb{F}_n^{-1}(u) - \mathbb{G}_n^{-1}(u))$  is evaluated at 0 on some random subsets of  $\mathcal{I}_n \cup \mathcal{J}_n$  and at  $+\infty$  on some others. The later problem is the most difficult to address.

**Step 1.** Let  $K_n$  be a positive sequence such that  $K_n \rightarrow +\infty$  and define

$$i_n = \frac{n}{v_n K_n \rho(\psi^{-1}(\log n + K_n))}. \quad (25)$$

Notice that (FG1) and (23) imply that  $\rho(\psi^{-1}(\log n + K_n)) \rightarrow +\infty$  and  $i_n = o(n^{1-b/2} L(1/\sqrt{n})/K_n)$  as  $n \rightarrow +\infty$ , so that  $i_n/\sqrt{n} \rightarrow 0$  even when  $b = 1$ , thanks to (10). The following lemma moreover ensures that  $i_n/\log \log n \rightarrow +\infty$ . Observe also that  $\psi^{-1}(\log n + K_n) = F^{-1}(1 - 1/ne^{K_n})$  is an extreme quantile just beyond the expected order  $F^{-1}(1 - 1/n)$  for  $X_{(n)}$  and  $Y_{(n)}$ , which is the key to Lemma 17. Let  $[y]$  denote the integer part of  $y$ . Consider the random variable

$$I_{\mathcal{I}_n} \leq \int_{\mathcal{I}_n} \rho(\mathbb{F}_n^{-1}(u) - \mathbb{G}_n^{-1}(u)) du = \frac{1}{n} \sum_{i=n-[i_n]}^n \rho(X_{(i)} - Y_{(i)}).$$

**Lemma 17** *Assume (FG1), (C) and (CFG<sub>E</sub>). There exists  $K_n$  such that we have*

$$K_n \rightarrow +\infty, \quad \lim_{n \rightarrow +\infty} \frac{K_n}{\log \log n} = 0, \quad \liminf_{n \rightarrow +\infty} \frac{\log i_n}{\log \log n} \geq \theta_2 > 0$$

and

$$\lim_{n \rightarrow +\infty} v_n I_{\mathcal{I}_n} = 0 \text{ in probability.}$$

**Proof. (i)** Let  $K_n \rightarrow +\infty$ ,  $K_n/\log \log n \rightarrow 0$  be as slow as needed later. By (FG1) we have  $F^{-1}(1 - 1/ne^{K_n}) \rightarrow +\infty$  as  $n \rightarrow +\infty$ , yet arbitrarily slowly. Thus, by (13) and (25) we have, for any  $\theta'' > 1 - b/2$ , any  $\theta' < \theta_2$  and all  $n$  large,

$$\begin{aligned} i_n &= \frac{n^{1-b/2} L(1/\sqrt{n})}{K_n \rho(\psi^{-1}(\log n + K_n))} \\ &\geq \frac{1}{K_n} \frac{L(1/\sqrt{n})}{L(1/\sqrt{ne^{K_n}})} \exp\left(-\left(1 - \frac{b}{2}\right) K_n + 2 \log \psi^{-1}(\log n + K_n) + \theta_2 \log(\log n + K_n)\right) \\ &\geq \frac{L(1/\sqrt{n})}{L(1/\sqrt{ne^{K_n}})} \frac{1}{e^{\theta'' K_n}} \left(F^{-1}\left(1 - \frac{1}{ne^{K_n}}\right)\right)^2 (\log n + K_n)^{\theta_2} \\ &\geq \frac{L(1/\sqrt{n})}{L(1/\sqrt{ne^{K_n}})} (\log n)^{\theta'}. \end{aligned}$$

Applying (21) and  $K_n \rightarrow +\infty$  we get

$$\frac{L\left(\frac{1}{\sqrt{ne^{K_n}}}\right)}{L(1/\sqrt{n})} = \exp\left(\eta\left(\frac{1}{\sqrt{ne^{K_n}}}\right) - \eta\left(\frac{1}{\sqrt{n}}\right) + \int_{\sqrt{n}}^{\sqrt{ne^{K_n}}} \frac{s(y)}{y} dy\right).$$

Since  $e^{K_n} < \log n$  we can furthermore choose  $K_n$  such that

$$K_n < \frac{1}{s_n}, \quad s_n = \sup_{\sqrt{n} \leq y \leq \sqrt{n \log n}} s(y),$$

where  $s_n \rightarrow 0$  as  $n \rightarrow +\infty$ . The slower is  $L$  the faster is  $1/s_n$  hence the resulting requirement is sometimes only the initial  $K_n/\log \log n \rightarrow 0$ . We readily obtain, by (22),

$$\limsup_{n \rightarrow +\infty} \frac{L\left(\frac{1}{\sqrt{ne^{K_n}}}\right)}{L(1/\sqrt{n})} \leq \limsup_{n \rightarrow +\infty} \exp\left(2\eta_0 + s_n \frac{K_n}{2}\right) < +\infty.$$

The claimed deterministic  $\liminf$  is proved by letting  $\theta' \rightarrow \theta_2$ . Notice that  $(CFG_E)$  was crucially required.

(ii) Concerning the stochastic integral  $I_{\mathcal{I}_n}$  the choice of  $i_n$  in (25) is minimal to guaranty the rate  $v_n$  and  $(CFG_E)$  is not required. Recall that  $F$  has support  $\mathbb{R}$ . Fix  $\varepsilon > 0$  and consider the events

$$A_n = \{v_n I_{\mathcal{I}_n} \geq 4\varepsilon\}, \quad B_n = \{X_{(n-[i_n])} > 0 \cap Y_{(n-[i_n])} > 0\}.$$

We have  $\mathbb{P}(A_n) \leq \mathbb{P}(A_n \cap B_n) + \mathbb{P}(B_n^c)$  and  $\mathbb{P}(B_n^c) \rightarrow 0$  as  $n \rightarrow +\infty$ . On  $B_n$  it holds

$$v_n I_{\mathcal{I}_n} \leq \frac{v_n}{n} \sum_{i=n-[i_n]}^n (\rho_+(X_{(i)}) + \rho_-(Y_{(i)})) \leq \frac{v_n}{n} (i_n+1) (\rho_+(X_{(n)}) + \rho_-(Y_{(n)}))$$

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

$$C_{n,X} = \left\{ \rho_+(X_{(n)}) \geq \varepsilon \frac{n}{v_n i_n} \right\}, \quad C_{n,Y} = \left\{ \rho_-(Y_{(n)}) \geq \varepsilon \frac{n}{v_n i_n} \right\}.$$

In order to evaluate  $\mathbb{P}(C_{n,X}) = 1 - (1 - \mathbb{P}(\rho_+(X) > \varepsilon n/v_n i_n))^n$  we combine  $\rho_+^{-1}(x) = l_+^{-1}(\log x)$ ,  $l^{-1} \leq l_+^{-1}$  and  $\psi_X = \psi$  with (25) to obtain, for  $n$  large enough to have  $K_n > 1/\varepsilon$ ,

$$\begin{aligned} & \mathbb{P}(\rho_+(X) > \varepsilon K_n \rho(\psi^{-1}(\log n + K_n))) \\ & \leq \exp(-\psi \circ l^{-1}(\log \varepsilon + \log K_n + l(\psi^{-1}(\log n + K_n)))) \\ & \leq \frac{1}{ne^{K_n}}. \end{aligned}$$

Therefore  $\mathbb{P}(C_{n,X}) \leq 1 - \exp(-\exp(-K_n)) \sim \exp(-K_n) \rightarrow 0$  as  $n \rightarrow +\infty$ , and similarly  $\mathbb{P}(C_{n,Y}) \rightarrow 0$ . This implies that  $v_n I_{\mathcal{I}_n} \rightarrow 0$  in probability.  $\square$

**Step 2.** Write  $\beta_n(u) = \beta_n^X(u) - \beta_n^Y(u)$  with

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

so that  $I_A = \int_A \rho_c(\beta_n(u)/\sqrt{n}) du$  in (24). Let  $\Delta_n = \mathcal{J}_n \cup \mathcal{K}_n \cup \mathcal{L} = [\underline{u}, 1 - i_n/n]$ . The next lemma shows that in the integral  $I_{\Delta_n}$  the cost function  $\rho$  is evaluated near 0 provided that  $n$  is large.

**Lemma 18** *Assume (FG) and (CFG<sub>E</sub>). For any  $0 < \xi < 1/2 - b/4$  it holds*

$$\lim_{n \rightarrow +\infty} (\log n)^\xi \sup_{u \in \Delta_n} \frac{|\beta_n(u)|}{\sqrt{n}} = 0 \quad a.s.$$

**Proof. (i)** Assuming (FG1), (FG2) and since  $i_n/\log \log n \rightarrow +\infty$  by Lemma 17 we can apply the classical hungarian results to  $|\beta_n(u)| \leq |\beta_n^X(u)| + |\beta_n^Y(u)|$  exactly as for Lemma 23 in [2] to get

$$\limsup_{n \rightarrow +\infty} \sup_{u \in \Delta_n} \frac{h(u)|\beta_n(u)|}{\sqrt{(1-u)\log \log n}} \leq 8 \quad a.s. \quad (27)$$

Next observe that (FG3) implies, for some  $0 < M < +\infty$  and  $u \in \Delta_n$ ,

$$\begin{aligned} \frac{1}{M} \frac{\sqrt{1-u}}{h(u)} \sqrt{\frac{\log \log n}{n}} &\leq \frac{F^{-1}(u)}{\sqrt{1-u}} \sqrt{\frac{\log \log n}{n}} \\ &\leq \varepsilon_n = \frac{F^{-1}(1 - i_n/n)}{\sqrt{i_n}} \sqrt{\log \log n}. \end{aligned} \quad (28)$$

**(ii)** Remind that  $e^{-K_n} < 1 < i_n$  for all  $n$  large, and  $F^{-1}(1 - i_n/n) \rightarrow +\infty$  as  $n \rightarrow +\infty$  with no obvious control on the rate. By (25) and the consequence (13) of (CFG<sub>E</sub>) we have already seen in the proof of Lemma 17 that if  $\theta'' < 1 - b/2$  and  $\theta'' < \theta' < \theta_2$  then it holds, for all  $n$  large enough,

$$\begin{aligned} i_n &\geq \frac{1}{e^{\theta'' K_n}} \left( F^{-1} \left( 1 - \frac{1}{ne^{K_n}} \right) \right)^2 (\log n + K_n)^{\theta_2} \\ &\geq \left( F^{-1} \left( 1 - \frac{1}{n} \right) \right)^2 (\log n)^{\theta'} \\ &\geq \left( F^{-1} \left( 1 - \frac{i_n}{n} \right) \right)^2 (\log \log n)(\log n)^{\theta''} \end{aligned}$$

hence for any  $0 < \xi < \theta''/2$  it holds  $\lim_{n \rightarrow +\infty} (\log n)^\xi \varepsilon_n = 0$ . The conclusion follows, by (27) and (28).  $\square$

Let  $j_n = n^\beta$  with  $1/2 < \beta < 1$ , so that  $i_n < \sqrt{n} < j_n$  for all  $n$  large. Remind  $\varepsilon_n$  from (28). Let introduce

$$\varepsilon_n(u) = 9 \frac{\sqrt{1-u}}{h(u)} \sqrt{\frac{\log \log n}{n}} \leq 9\varepsilon_n, \quad u \in \mathcal{J}_n. \quad (29)$$

**Lemma 19** Assume  $(FG)$ ,  $(C)$  and  $(CFG_E)$ . Then we have

$$\lim_{n \rightarrow +\infty} v_n I_{\mathcal{J}_n} = 0 \quad a.s.$$

**Proof.** (i) By Lemma 18, for all  $n$  large enough and any  $u \in \mathcal{J}_n$  it holds

$$\frac{1}{n} \leq \sup_{u \in \mathcal{J}_n} \sqrt{\frac{1-u}{n}} \leq \varepsilon_n(u) \leq \frac{1}{(\log n)^\xi}.$$

Consider  $L$  defined in (11). Using (21) and (22) we get

$$L_n = \sup_{u \in \mathcal{J}_n} \frac{L(\varepsilon_n(u))}{L(1/\sqrt{n})} \leq \exp \left( 2\eta_0 + \int_{(\log n)^\xi}^n \frac{|s(y)|}{y} dy \right) \quad (30)$$

hence

$$\lim_{n \rightarrow +\infty} \frac{\log L_n}{\log n} \leq \lim_{n \rightarrow +\infty} \frac{1}{\log n} \left( 2\eta_0 + \log n \sup_{(\log n)^\xi \leq y \leq n} |s(y)| \right) = 0. \quad (31)$$

(ii) Remind that  $\rho_\pm$  are increasing. By Lemma 18 and  $(C2)$  we almost surely have, for all  $n$  large,

$$I_{\mathcal{J}_n} \leq \int_{\mathcal{J}_n \cap \{\beta_n \geq 0\}} \rho_+(\varepsilon_n(u)) du + \int_{\mathcal{J}_n \cap \{\beta_n < 0\}} \rho_-(\varepsilon_n(u)) du$$

where, by (27), (28) and (29),  $\sup_{u \in \mathcal{J}_n} \varepsilon_n(u) \leq 9\varepsilon_n \rightarrow 0$ . Hence, recalling (11) we are reduced to study the bounding deterministic integral

$$I_{\mathcal{J}_n} \leq \int_{\mathcal{J}_n} \rho(\varepsilon_n(u)) du = \int_{\mathcal{J}_n} (\varepsilon_n(u))^b L(\varepsilon_n(u)) du.$$

By (11),  $L_n$  from (30) and  $(FG3)$  we further have

$$v_n I_{\mathcal{J}_n} \leq L_n (\log \log n)^{b/2} \int_{\mathcal{J}_n} \left( \frac{F^{-1}(u)}{\sqrt{1-u}} \right)^b du. \quad (32)$$

We next show that  $L_n (\log \log n)^{b/2}$  is a secondary order factor compared to the integral in (32), whatever the choice of  $1/2 < \beta < 1$  defining  $j_n$  in  $\mathcal{J}_n$ .

(iii) The fact that  $l(y) \geq \log y$  as  $y \rightarrow +\infty$  combined to  $(CFG_E)$  shows that for all  $u$  large enough, we have

$$\begin{aligned} F^{-1}(u) &= \psi^{-1} \left( \log \frac{1}{1-u} \right) \\ &\leq \exp \left( l \circ \psi^{-1} \left( \log \frac{1}{1-u} \right) \right) \\ &\leq \exp \left( \left( 1 - \frac{b}{2} \right) \log \frac{1}{1-u} + \log L(\sqrt{1-u}) - 2 \log F^{-1}(u) - \theta_2 \log \log \frac{1}{1-u} \right). \end{aligned}$$

Therefore we get

$$F^{-1}(u)^3 \leq \left( \frac{1}{1-u} \right)^{1-b/2} \frac{L(\sqrt{1-u})}{(\log(1/(1-u)))^{\theta_2}}$$

and

$$\begin{aligned} \int_{\mathcal{J}_n} \left( \frac{F^{-1}(u)}{\sqrt{1-u}} \right)^b du &\leq \int_{\mathcal{J}_n} \left( \frac{1}{1-u} \right)^{(1-b/2)b/3+b/2} \frac{L(\sqrt{1-u})^{b/3}}{(\log(1/(1-u)))^{\theta_2 b/3}} du \\ &\leq \int_{\mathcal{J}_n} \left( \frac{1}{1-u} \right)^{b(5-b)/6} \frac{L(\sqrt{1-u})^{b/3}}{(\log(1/(1-u)))^{\theta_2 b/3}} du. \end{aligned} \quad (33)$$

Since  $1 \leq b < 2$  we can find  $\gamma$  such that  $0 < b(5-b)/6 < \gamma < 1$ . The second factor in the integral (33) is slowly varying in  $1-u$  as  $u \rightarrow 1$  thus the whole integral is ultimately bounded from above by

$$(1-\gamma) \int_{\mathcal{J}_n} \left( \frac{1}{1-u} \right)^\gamma du = [-(1-u)^{1-\gamma}]_{1-j_n/n}^{1-i_n/n} \leq \frac{1}{n^{(1-\gamma)(1-\beta)}}. \quad (34)$$

We deduce from (31), (32), (33) and (34) the convergence

$$\lim_{n \rightarrow +\infty} v_n I_{\mathcal{J}_n} \leq \lim_{n \rightarrow +\infty} \frac{L_n(\log \log n)^{b/2}}{n^{(1-\gamma)(1-\beta)}} = 0 \quad a.s.$$

at a power rate.  $\square$

**Step 3.** Compared to  $\mathcal{J}_n$  the interval  $\mathcal{K}_n$  is so large that  $v_n I_{\mathcal{K}_n}$  can no more converge to zero. Instead it is made small with high probability by choosing  $\bar{u}$  and  $\beta$  properly, at Lemma 21. Moreover, in order to evaluate the integral of  $\rho_c(\beta_n(u)/\sqrt{n})$  over  $\mathcal{K}_n$  accurately enough it is no more sufficient to bound the process, therefore we approximate it at Lemma 20 by a Gaussian process which helps revealing the underlying deterministic integral to compute. Lastly the fact that  $\beta_n(u)$  itself may be very small or very large along  $\mathcal{K}_n$  makes a bit tedious the uniform control of the slowly varying part  $L(x)$  of  $\rho(x)$ .

Define  $\Delta'_n = (j_n/n, 1-j_n/n)$ . We first recall the strong approximation of the joint quantile processes

$$\mathcal{Q}_n(u) = (\beta_n^X(u), \beta_n^Y(u)), \quad u \in \Delta'_n,$$

by the joint Gaussian processes

$$\mathcal{G}_n(u) = (\mathbb{B}_n^X(u), \mathbb{B}_n^Y(u)), \quad \mathbb{B}_n^X(u) = \frac{B_n^X(u)}{h_X(u)}, \quad \mathbb{B}_n^Y(u) = \frac{B_n^Y(u)}{h_Y(u)}, \quad u \in \Delta'_n,$$

where  $B_n^X(u) = \mathbb{H}_n(H_X(u))$ ,  $B_n^Y(u) = \mathbb{H}_n(H^Y(u))$  and  $\mathbb{H}_n$  is a  $\mathbb{P}^{X,Y}$ -Brownian bridge indexed by the halfplanes

$$H_X(u) = \{(x, y) : x \leq F^{-1}(u)\}, \quad H^Y(u) = \{(x, y) : y \leq F^{-1}(u)\}.$$

Therefore  $B_n^X$  and  $B_n^Y$  are two standard Brownian bridges with cross covariance given for  $u, v \in (0, 1)$  by

$$\begin{aligned} \text{cov}(B_n^X(u), B_n^Y(v)) &= \mathbb{P}^{X,Y}(H_X(u) \cap H^Y(v)) - \mathbb{P}^{X,Y}(H_X(u))\mathbb{P}^{X,Y}(H^Y(v)) \\ &= \mathbb{P}(X \leq F^{-1}(u), Y \leq F^{-1}(v)) - uv \\ &= H(F^{-1}(u), F^{-1}(v)) - uv. \end{aligned}$$

Notice that  $H(F^{-1}(u), F^{-1}(v))$  is the copula function of  $(X, Y)$ . From now and for the remainder of the proof we work on the probability space of the following Lemma 20. The weak convergence finally established on this space at steps 4 and 5 remains valid on any probability space.

**Lemma 20** *Assume (FG). Then we can build on the same probability space versions of  $(X_n, Y_n)_{n \geq 1}$  and  $(\mathbb{H}_n)_{n \geq 1}$  such that  $\mathcal{Q}_n(u) = \mathcal{G}_n(u) + \mathcal{Z}_n(u)$  for all  $n \geq 1$  and  $u \in \Delta'_n$  where  $\mathcal{Z}_n(u) = (Z_n^X(u)/h_X(u), Z_n^Y(u)/h_Y(u))$  satisfies, for some  $v \in (0, 1/22)$ ,*

$$\lim_{n \rightarrow +\infty} n^v \sup_{u \in \Delta'_n} |Z_n^X(u)| = \lim_{n \rightarrow +\infty} n^v \sup_{u \in \Delta'_n} |Z_n^Y(u)| = 0 \quad a.s.$$

**Proof.** This follows from Theorem 28 in [2] with  $F = G$ .  $\square$

The joint strong approximation of Lemma 20 applied with  $F = G$  and  $h_X = h_Y = h$  combined to  $(CFG_E)$  provides a stochastic control of the deviations of  $v_n I_{\mathcal{K}_n}$  that is weaker than in probability but sufficient for the targeted weak convergence. Since it concerns the probability distribution of  $I_{\mathcal{K}_n}$  the following lemma remains true on any probability space.

**Lemma 21** *Assume (FG), (C) and  $(CFG_E)$ . There exists  $\beta \in (1/2, 1)$  such that for any choice of  $\lambda > 0$  and  $\varepsilon > 0$  one can find  $\bar{u} \in (1/2, 1)$  and  $n_0 > 0$  such that, for all  $n > n_0$ ,*

$$\mathbb{P}(v_n I_{\mathcal{K}_n} > \lambda) < \varepsilon.$$

**Proof.** Fix  $\lambda > 0$  and  $\varepsilon > 0$  then consider, with  $\beta_n$  as in (26) the event

$$\mathcal{C}_n^\lambda = \left\{ v_n \int_{\mathcal{K}_n} \rho_c \left( \frac{\beta_n(u)}{\sqrt{n}} \right) du > \lambda \right\}.$$

(i) For  $0 < \tau < \min(1, \lambda/2)$  define the random sets

$$\mathcal{K}_n^{<\tau} = \{u \in \mathcal{K}_n : |\beta_n(u)| < \tau\}, \quad \mathcal{K}_n^{>\tau} = \mathcal{K}_n \setminus \mathcal{K}_n^{<\tau}.$$

Recalling that the cost  $\rho_\pm$  is convex, positive and such that  $\rho_\pm(0) = 0$  we have  $\rho_\pm(\tau x) \leq \tau \rho_\pm(x)$  for all  $x \geq 0$ . It follows that

$$\begin{aligned} v_n I_{\mathcal{K}_n^{<\tau}} &\leq v_n \int_{\mathcal{K}_n^{<\tau} \cap \{\beta_n < 0\}} \tau \rho_+ \left( \frac{1}{\sqrt{n}} \right) du + v_n \int_{\mathcal{K}_n^{<\tau} \cap \{\beta_n \geq 0\}} \tau \rho_- \left( \frac{1}{\sqrt{n}} \right) du \\ &\leq \frac{\max(\rho_-(1/\sqrt{n}), \rho_+(1/\sqrt{n}))}{\rho(1/\sqrt{n})} \tau \int_{\mathcal{K}_n^{<\tau}} du \leq \tau. \end{aligned}$$

As a consequence,

$$\mathbb{P}(\mathcal{C}_n^\lambda) = \mathbb{P}(v_n(I_{\mathcal{K}_n^{<\tau}} + I_{\mathcal{K}_n^{>\tau}}) > \lambda) \leq \mathbb{P}(v_n I_{\mathcal{K}_n^{>\tau}} \geq \lambda - \tau) \leq \mathbb{P}\left(v_n I_{\mathcal{K}_n^{>\tau}} \geq \frac{\lambda}{2}\right).$$

(ii) For all  $n \geq n_0$  and  $n_0 = n_0(\varepsilon, \xi)$  large enough we have  $(\log n)^\xi < \sqrt{n}$  together with, by Lemma 18 and since  $\mathcal{K}_n^{>\tau} \subset \mathcal{K}_n \subset \Delta_n$ ,

$$\mathbb{P}(\mathcal{D}_n) > 1 - \frac{\varepsilon}{2}, \quad \mathcal{D}_n = \left\{ \sup_{u \in \mathcal{K}_n^{>\tau}} \frac{|\beta_n(u)|}{\sqrt{n}} \leq \frac{1}{(\log n)^\xi} \right\}.$$

Assume now that  $n \geq n_0$ . On the event  $\mathcal{D}_n$ , for any  $u \in \mathcal{K}_n^{>\tau}$  we have

$$\frac{\tau}{\sqrt{n}} \leq \min\left(\frac{1}{\sqrt{n}}, \frac{|\beta_n(u)|}{\sqrt{n}}\right) \leq \frac{1}{(\log n)^\xi}$$

which by (11), (21) and (22) yields

$$\begin{aligned} \frac{L(|\beta_n(u)|/\sqrt{n})}{L(1/\sqrt{n})} &\leq \exp\left(2\eta_0 + \int_{\min(\sqrt{n}, \sqrt{n}/|\beta_n(u)|)}^{\sqrt{n}/\tau} \frac{|s(y)|}{y} dy\right) \\ &\leq c_0 \exp\left(s_n \int_{\min(\sqrt{n}, \sqrt{n}/|\beta_n(u)|)}^{\sqrt{n}/\tau} \frac{1}{y} dy\right) \\ &= c_0 \exp(s_n (\max(0, \log(|\beta_n(u)|)) - \log \tau)) \\ &\leq c_0 \left| \frac{\beta_n(u)}{\tau} \right|^{q_n(u)} \end{aligned}$$

where the sequence  $s_n$  and the stochastic process  $q_n(u)$  are defined by

$$s_n = \sup_{(\log n)^\xi \leq y \leq \sqrt{n}/\tau} |s(y)|, \quad q_n(u) = s_n 1_{\{|\beta_n(u)| > 1\}}. \quad (35)$$

Since  $s(y) \rightarrow 0$  as  $y \rightarrow +\infty$  we have

$$\lim_{n \rightarrow +\infty} \sup_{u \in \mathcal{K}_n^{>\tau}} q_n(u) \leq \lim_{n \rightarrow +\infty} s_n = 0 \quad (36)$$

and this uniform convergence of  $q_n$  is certain, not almost sure. In other words, the uncertainty in the following inequality only comes from  $\mathbb{P}(\mathcal{D}_n)$ . We have shown that for all  $n$  large, on the event  $\mathcal{D}_n$ , it holds

$$v_n I_{\mathcal{K}_n^{>\tau}} \leq v_n \int_{\mathcal{K}_n^{>\tau}} \rho\left(\frac{|\beta_n(u)|}{\sqrt{n}}\right) du \leq \frac{c_0}{\tau^{s_n}} \int_{\mathcal{K}_n^{>\tau}} |\beta_n(u)|^{b+q_n(u)} du \quad (37)$$

where  $\tau^{s_n} \rightarrow 1$  as  $n \rightarrow +\infty$ . We are ready to bound  $\mathbb{P}(\mathcal{D}_n \cap \{v_n I_{\mathcal{K}_n^{>\tau}} \geq \lambda/2\})$ .

(iii) On the probability space of Lemma 20 we have

$$|\beta_n(u)| \leq \frac{|B_n^X(u)|}{h(u)} + \frac{|B_n^Y(u)|}{h(u)} + \frac{|Z_n^X(u)|}{h(u)} + \frac{|Z_n^Y(u)|}{h(u)}.$$

If  $\alpha \geq 1$  then  $(x+y)^\alpha \leq 2^{\alpha-1}(x^\alpha + y^\alpha)$  for all  $x, y \geq 0$ . Combining this fact with  $b + q_n(u) \geq b \geq 1$  and (36) thus implies that, for  $K > 1$  fixed and all  $n$  large enough,

$$\frac{1}{K4^{b-1}} \int_{\mathcal{K}_n^{\geq \tau}} |\beta_n(u)|^{b+q_n(u)} du \leq R_n^X + R_n^Y + S_n^X + S_n^Y$$

where

$$R_n^X = \int_{\mathcal{K}_n^{\geq \tau}} \left| \frac{B_n^X(u)}{h(u)} \right|^{b+q_n(u)} du, \quad S_n^X = \int_{\mathcal{K}_n^{\geq \tau}} \left| \frac{Z_n^X(u)}{h(u)} \right|^{b+q_n(u)} du.$$

It remains to prove that for an appropriate choice of  $\bar{u}$  and  $\beta$  we have

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \mathbb{P} \left( \mathcal{D}_n \cap \left\{ R_n^X \geq \frac{\lambda \tau^{s_n}}{8c_0} \right\} \right) &< \frac{\varepsilon}{8}, \\ \limsup_{n \rightarrow +\infty} \mathbb{P} \left( \mathcal{D}_n \cap \left\{ S_n^X \geq \frac{\lambda \tau^{s_n}}{8c_0} \right\} \right) &< \frac{\varepsilon}{8}, \end{aligned}$$

which ensures by (37) that  $\mathbb{P}(\mathcal{D}_n \cap \{v_n I_{\mathcal{K}_n^{\geq \tau}} \geq \lambda/2\}) \leq \varepsilon/2$ . For short, it is assumed below that  $1/9 < \tau^{s_n}/8$ .

(iv) The following integral  $T_n$  is crucial with respect to the integrability of the processes  $B_n^X$  and  $Z_n^X$ . Let  $b' > b$  be so close to  $b$  that  $0 < b'(5-b')/6 < \gamma < 1$ . Consider the random function  $q_n(u)$  from (35). For all  $n$  large enough we have  $b \leq b + q_n(u) < b'$  hence (33) and (34) entail

$$\begin{aligned} T_n &= \int_{\mathcal{K}_n} \left| \frac{\sqrt{1-u}}{h(u)} \right|^{b+q_n(u)} du \leq \int_{\mathcal{K}_n} \left| \frac{F^{-1}(u)}{\sqrt{1-u}} \right|^{b+q_n(u)} du \leq \int_{\mathcal{K}_n} \left| \frac{F^{-1}(u)}{\sqrt{1-u}} \right|^{b'} du \\ &\leq \left[ -\frac{(1-u)^{1-\gamma}}{1-\gamma} \right]_{\bar{u}}^{1-j_n/n} \leq \frac{(1-\bar{u})^{1-\gamma}}{1-\gamma}. \end{aligned}$$

(v) On the one hand we have, by Fubini-Tonelli and recalling that  $B_n^X$  is a standard Brownian bridge and the sequence  $s_n$  is defined at (36),

$$\begin{aligned} \mathbb{E}(R_n^X) &\leq T_n \sup_{u \in \mathcal{K}_n} \mathbb{E} \left( \left| \frac{B_n^X(u)}{\sqrt{u(1-u)}} \right|^{b+q_n(u)} \right) \\ &\leq T_n \sup_{0 \leq s \leq s_n} \mathbb{E}(|\mathcal{N}(0,1)|^{b+s}) = T_n \mathbb{E}(|\mathcal{N}(0,1)|^{b+s_n}). \end{aligned}$$

Assuming  $n$  so large that  $s_n < 2-b$  we get  $\mathbb{E}(R_n^X)/T_n < \mathbb{E}(|\mathcal{N}(0,1)|^2) = 1$  then choosing  $\bar{u}$  such that  $(1-\bar{u})^{1-\gamma} < 8(1-\gamma)\lambda/9c_0\varepsilon$  yields

$$\mathbb{P} \left( R_n^X \geq \frac{\lambda}{9c_0} \right) < \frac{9c_0}{\lambda} T_n < \frac{\varepsilon}{8}. \quad (38)$$



On the other hand we have  $\mathcal{K}_n^{>\tau} \subset \mathcal{K}_n \subset \Delta'_n$  and

$$S_n^X \leq \sup_{u \in \mathcal{K}_n^{>\tau}} \left| \frac{Z_n^X(u)}{\sqrt{1-u}} \right|^{b+q_n(u)} T_n.$$

By Lemma 20 it almost surely holds, for  $b' > b$  and all  $n$  large,

$$\sup_{u \in \mathcal{K}_n^{>\tau}} \left| \frac{Z_n^X(u)}{\sqrt{1-u}} \right| \leq \sup_{u \in \mathcal{K}_n} \frac{1}{n^v \sqrt{1-u}} = \frac{1}{n^v} \sqrt{\frac{n}{j_n}} \leq n^{(1-\beta)/2-v}$$

which vanishes provided  $1 - 2v < \beta < 1$ . Therefore, for this choice of  $\beta$ ,

$$\lim_{n \rightarrow +\infty} S_n^X = 0 \text{ a.s.}, \quad \lim_{n \rightarrow +\infty} \mathbb{P} \left( S_n^X \geq \frac{\lambda}{9c_0} \right) = 0. \quad (39)$$

(vi) Putting together the conclusions of (i)-(v), and especially (37), (38) and (39), implies

$$\mathbb{P}(\mathcal{C}_n^\lambda) \leq 1 - \mathbb{P}(\mathcal{D}_n) + \mathbb{P} \left( \mathcal{D}_n \cap \left\{ v_n I_{\mathcal{K}_n^{>\tau}} \geq \frac{\lambda}{2} \right\} \right) < \frac{\varepsilon}{2} + 4\frac{\varepsilon}{8} = \varepsilon.$$

Finally notice that the same  $\beta$  works whatever the choice of  $\lambda, \varepsilon$ .  $\square$

**Step 4.** Now  $\mathcal{L} = [\underline{u}, \bar{u}]$  is fixed. By Lemmas 18 and 20 there almost surely exists  $n_0(\omega)$  such that, for all  $n \geq n_0(\omega)$ ,  $\varepsilon_n(u)$  from (29),  $B_n(u) = B_n^X(u) - B_n^Y(u)$  and  $Z_n(u) = Z_n^X(u) - Z_n^Y(u)$ ,

$$\left| \frac{\beta_n(u)}{\sqrt{n}} \right| \leq \varepsilon_n(u) \leq x_0, \quad \beta_n(u) = \frac{B_n(u) + Z_n(u)}{h(u)}, \quad u \in \mathcal{L}.$$

As a consequence, the cost  $\rho_c$  is evaluated at 0 all along this step. Let  $\alpha > 0$  and consider  $I_{\mathcal{L}} = I_{\mathcal{L}_{1,n}} + I_{\mathcal{L}_{2,n}} + I_{\mathcal{L}_{3,n}}$  where, for  $n \geq n_0(\omega)$ ,

$$I_{\mathcal{L}_{k,n}} = \int_{\mathcal{L}_{k,n}} \rho_c \left( \frac{B_n(u) + Z_n(u)}{\sqrt{n}h(u)} \right) du, \quad k = 1, 2, 3, \quad (40)$$

and  $\mathcal{L} = \mathcal{L}_{1,n} \cup \mathcal{L}_{2,n} \cup \mathcal{L}_{3,n}$  with  $\mathcal{L}_{1,n} = \mathcal{L} \cap \{|B_n(u)| \leq \alpha\}$ ,  $\mathcal{L}_{2,n} = \mathcal{L} \cap \{|B_n(u)| \geq 1/\alpha\}$  and  $\mathcal{L}_{3,n} = \mathcal{L} \cap \{\alpha < |B_n(u)| < 1/\alpha\}$ . Also define

$$0 < \underline{h} = \min_{u \in \mathcal{L}} h(u) \leq \bar{h} = \max_{u \in \mathcal{L}} h(u) < +\infty.$$

(4.1) Choose  $\alpha \in (0, 1)$  arbitrarily small. In view of the almost sure rate  $1/n^v$  from Lemma 20 and (11) we have, given  $\underline{u}, \bar{u}$  then  $\underline{h}$ ,

$$\begin{aligned} \lim_{n \rightarrow +\infty} v_n I_{\mathcal{L}_{1,n}} &\leq \lim_{n \rightarrow +\infty} \frac{1}{\rho(1/\sqrt{n})} \int_{\mathcal{L}_{1,n}} \rho \left( \frac{\alpha + 1/n^v}{\sqrt{n}\underline{h}} \right) du \\ &\leq \lim_{n \rightarrow +\infty} \frac{\rho(2\alpha/\sqrt{n}\underline{h})}{\rho(1/\sqrt{n})} = \frac{(2\alpha)^b}{\underline{h}^b} \quad \text{a.s.} \end{aligned} \quad (41)$$

The last equality holds by definition of  $\rho \in RV(0, b)$ .

**(4.2)** Write  $\mathcal{L}_{2,n}^+ = \mathcal{L} \cap \{B_n(u) \geq 1/\alpha\}$  and  $\mathcal{L}_{2,n}^- = \mathcal{L} \cap \{B_n(u) \leq -1/\alpha\}$ . By Lemma 20 we have, for  $n$  large enough,

$$v_n I_{\mathcal{L}_{2,n}^+} = \frac{1}{\rho(1/\sqrt{n})} \int_{\mathcal{L}_{2,n}^+} \rho_+ \left( \frac{\beta_n(u)}{\sqrt{n}} \right) du \leq \frac{1}{\rho(1/\sqrt{n})} \int_{\mathcal{L}_{2,n}^+} \rho_+ \left( \frac{2B_n(u)}{h(u)\sqrt{n}} \right) du$$

then similar arguments as for (ii) in the proof of Lemma 21 yield

$$v_n I_{\mathcal{L}_{2,n}^+} \leq c_0 \frac{\rho_+(1/\sqrt{n})}{\rho(1/\sqrt{n})} \int_{\mathcal{L}_{2,n}^+} \left( \frac{2B_n(u)}{h(u)} \right)^{b_+ + s_n} du$$

where  $s_n \rightarrow 0$  is defined at (35) with  $\tau = 2/\alpha$ . By replacing  $\min(u, 1-u)$  with  $u(1-u) \leq \min(u, 1-u)$  in (CFG3) it follows that

$$v_n I_{\mathcal{L}_{2,n}^+} \leq K \int_{\mathcal{L}} 1_{\{B_n(u) \geq 1/\alpha\}} \left| \frac{F^{-1}(u)}{\sqrt{u(1-u)}} \right|^{b_+ + s_n} \left( \frac{B_n(u)}{\sqrt{u(1-u)}} \right)^{b_+ + s_n} du$$

where  $K > 0$ . As a consequence of (CFG<sub>E</sub>) we obtain exactly as for (33) and (34) that if  $b' \in (b, 2)$  is chosen sufficiently close to  $b$  then

$$\int_{(0,1)} \left| \frac{F^{-1}(u)}{\sqrt{u(1-u)}} \right|^{b'} du = K' < +\infty. \quad (42)$$

Since  $2u - 1 \leq H(u, u) \leq u$  for  $u \in (0, 1)$  we have

$$-(1-u)^2 \leq \text{Cov}(B_n^X(u), B_n^Y(u)) = H(u, u) - u^2 \leq u(1-u)$$

hence

$$0 \leq \text{Var}(B_n(u)) \leq 2u(1-u) + 2(1-u)^2 = 2(1-u)$$

and the random variable  $B_n(u)/\sqrt{u(1-u)}$  is centered Gaussian with variance bounded above by  $2/u$ . Let denote  $\mathcal{N}(0, 1)$  the standard normal distribution. By Hölder inequality we have, for  $u \in \mathcal{L}$  and  $n$  large,

$$\mathbb{E} \left( 1_{\{B_n(u) \geq 1/\alpha\}} \left| \frac{B_n(u)}{\sqrt{u(1-u)}} \right|^{b_+ + s_n} \right) \leq K'' \mathbb{P} \left( \sup_{\underline{u} \leq u \leq \bar{u}} |B_n(u)| \geq \frac{1}{\alpha} \right)^{1/2}$$

where  $K'' = (3/\underline{u}) \sup_{b_+ \leq s \leq b} \left( \mathbb{E} |\mathcal{N}(0, 1)|^{2s} \right)^{1/2} < +\infty$  only depends on  $b$ . We conclude that it asymptotically holds

$$\mathbb{E} \left( v_n I_{\mathcal{L}_{2,n}^+} \right) \leq K K' K'' \mathbb{P} \left( \sup_{\underline{u} \leq u \leq \bar{u}} |B_n(u)| > \frac{1}{\alpha} \right)^{1/2} \leq C \exp \left( -\frac{1}{\alpha^2} \right) \quad (43)$$

where  $C$  depends on  $M, b, F$  and  $\alpha$  was left arbitrary from the beginning. Clearly  $\mathbb{E}(v_n I_{\mathcal{L}_{2,n}^-})$  also obeys (43) by the same arguments. Notice that for the left hand

tail  $\underline{u}$  and  $1 - \bar{u}$  play a symmetric role in the previous control of the variance of  $B_n(u)$  by  $u(1 - u)$ .

**(4.3)** Let  $\mathcal{L}_{3,n}^+ = \mathcal{L} \cap \{\alpha < B_n(u) < 1/\alpha\}$  and  $\mathcal{L}_{3,n}^- = \mathcal{L} \cap \{-1/\alpha < B_n(u) < -\alpha\}$ . By Lemma 20 again we almost surely ultimately have

$$\text{sign}(B_n(u) + Z_n(u))1_{\mathcal{L}_{3,n}}(u) = \text{sign}(B_n(u))1_{\mathcal{L}_{3,n}}(u)$$

where  $\text{sign}(x) = 1_{x>0} - 1_{x<0}$ . Therefore, (C2) implies, for all  $n$  large enough,

$$\begin{aligned} & 1_{\mathcal{L}_{3,n}}(u)\rho_c\left(\frac{B_n(u) + Z_n(u)}{\sqrt{n}h(u)}\right) \\ &= 1_{\mathcal{L}_{3,n}^+}(u)\rho_+\left(\frac{B_n(u) + Z_n(u)}{\sqrt{n}h(u)}\right) + 1_{\mathcal{L}_{3,n}^-}(u)\rho_-\left(\frac{|B_n(u) + Z_n(u)|}{\sqrt{n}h(u)}\right). \end{aligned}$$

Now assume that  $\alpha < 2/\bar{h}$  and  $\mathcal{L}_{3,n} \neq \emptyset$ , so that

$$\begin{aligned} v_n I_{\mathcal{L}_{3,n}} &= \frac{1}{\rho(1/\sqrt{n})} \left( \int_{\mathcal{L}_{3,n}^+} \rho_+\left(\frac{|B_n(u)|}{\sqrt{n}h(u)}\right) du + R_n^+ \right) \\ &\quad + \frac{1}{\rho(1/\sqrt{n})} \left( \int_{\mathcal{L}_{3,n}^-} \rho_-\left(\frac{|B_n(u)|}{\sqrt{n}h(u)}\right) du + R_n^- \right) \end{aligned}$$

where we have, by convexity and differentiability of  $\rho_{\pm}$  on  $(0, +\infty)$ ,

$$\begin{aligned} R_n^{\pm} &= \int_{\mathcal{L}_{3,n}^{\pm}} \left( \rho_{\pm}\left(\frac{|B_n(u) + Z_n(u)|}{\sqrt{n}h(u)}\right) - \rho_{\pm}\left(\frac{|B_n(u)|}{\sqrt{n}h(u)}\right) \right) du \\ &\leq \sup_{u \in \mathcal{L}_{3,n}^{\pm}} \rho'_{\pm}\left(\frac{|B_n(u)| + |Z_n(u)|}{\sqrt{n}h(u)}\right) \frac{|Z_n(u)|}{\sqrt{n}h(u)} \end{aligned}$$

The regular variation (C2) further implies  $x\rho'_{\pm}(x)/\rho_{\pm}(x) \rightarrow 1$  as  $x \rightarrow 0$ . As a consequence, with probability one, for all  $n$  large it holds

$$\begin{aligned} \frac{R_n^{\pm}}{\rho(1/\sqrt{n})} &\leq \frac{1}{\rho(1/\sqrt{n})} \rho_{\pm}\left(\frac{|B_n(u)| + |Z_n(u)|}{\sqrt{n}h(u)}\right) \sup_{u \in \mathcal{L}_{3,n}} \frac{|Z_n(u)|}{|B_n(u)| + |Z_n(u)|} \\ &\leq \frac{\rho_{\pm}(2/\sqrt{n}h\alpha)}{\rho(1/\sqrt{n})} \frac{2}{\alpha n^v} \leq \frac{\rho_{\pm}(2/\sqrt{n}h\alpha)}{\rho_{\pm}(1/\sqrt{n})} \frac{2}{\alpha n^v} \leq \left(\frac{2}{h\alpha}\right)^{b_{\pm}} \frac{3}{\alpha n^v} \end{aligned}$$

which vanishes as  $n \rightarrow +\infty$ . Here we have used that  $\rho_{\pm}(\theta x)/\rho_{\pm}(x) \rightarrow \theta^{b_{\pm}}$  as  $x \rightarrow 0$  for any fixed  $\theta > 0$ , and Lemma 20. Finally we see that

$$\frac{1}{\rho_{\pm}(1/\sqrt{n})} \int_{\mathcal{L}_{3,n}^{\pm}} \rho_{\pm}\left(\frac{|B_n(u)|}{\sqrt{n}h(u)}\right) du = \int_{\mathcal{L}_{3,n}^{\pm}} \left(\frac{|B_n(u)|}{h(u)}\right)^b du + R_{3,n}^{\pm}$$

with

$$R_{3,n}^{\pm} = \int_{\mathcal{L}_{3,n}^{\pm}} L_n^{\pm}(u) \left(\frac{|B_n(u)|}{h(u)}\right)^b du, \quad L_n^{\pm}(u) = \frac{L_{\pm}(|B_n(u)|/\sqrt{n}h(u))}{L_{\pm}(1/\sqrt{n})} - 1.$$

Clearly, it follows

$$|R_{3,n}^\pm| \leq \left(\frac{1}{h\alpha}\right)^b \sup_{u \in \mathcal{L}_{3,n}} |L_n^\pm(u)| \leq \left(\frac{1}{h\alpha}\right)^b \left( \sup_{\alpha/\bar{h}\sqrt{n} \leq x \leq 1/\alpha h\sqrt{n}} \frac{L_\pm(x)}{L_\pm(1/\sqrt{n})} - 1 \right)$$

thus, by (21) and (22) we get  $|R_{3,n}^\pm| \rightarrow 0$  as  $n \rightarrow +\infty$ . We conclude that

$$I_{\mathcal{L}_{3,n}}^* = \rho_+(1/\sqrt{n}) \int_{\mathcal{L}_{3,n}^+} \left( \frac{|B_n(u)|}{h(u)} \right)^{b_+} du + \rho_-(1/\sqrt{n}) \int_{\mathcal{L}_{3,n}^-} \left( \frac{|B_n(u)|}{h(u)} \right)^{b_-} du \quad (44)$$

almost surely satisfies  $\lim_{n \rightarrow +\infty} v_n |I_{\mathcal{L}_{3,n}} - I_{\mathcal{L}_{3,n}}^*| = 0$ .

**Step 5.** Consider  $W_c^E(\mathbb{F}_n, \mathbb{G}_n) = \int_0^1 \rho_c(\mathbb{F}_n^{-1}(u), \mathbb{G}_n^{-1}(u)) du$ . As we assumed that

$$\lim_{n \rightarrow +\infty} \frac{\rho_+(1/\sqrt{n})}{\rho(1/\sqrt{n})} = \pi_+, \quad \lim_{n \rightarrow +\infty} \frac{\rho_-(1/\sqrt{n})}{\rho(1/\sqrt{n})} = \pi_-$$

by (C4) and  $E = \mathbb{R}$  we have established that  $v_n W_c^E(\mathbb{F}_n, \mathbb{G}_n) \rightarrow_{weak} W$  with

$$W = \pi_+ \int_0^1 1_{\{B(u) > 0\}} \left( \frac{|B(u)|}{h(u)} \right)^{b_+} du + \pi_- \int_0^1 1_{\{B(u) < 0\}} \left( \frac{|B(u)|}{h(u)} \right)^{b_-} du$$

and  $B$  is a standard Brownian bridge. To see this write  $W_c^E(\mathbb{F}_n, \mathbb{G}_n) = I_{\mathcal{I}_n} + I_{\mathcal{J}_n} + I_{\mathcal{K}_n} + I_{\mathcal{L}_{1,n}} + I_{\mathcal{L}_{2,n}} + I_{\mathcal{L}_{3,n}}$  where each of the first three integrals is indeed the sum of its left hand tail and right hand tail version, likewise for  $I_{\mathcal{L}_{3,n}}^*$  defined at (44). We have shown that  $v_n(I_{\mathcal{I}_n} + I_{\mathcal{J}_n}) \rightarrow 0$  in probability. Let  $\Psi$  be a real valued  $k$ -Lipschitz function on  $\mathbb{R}$ , bounded by  $m$ . Given arbitrarily small constants  $\lambda > 0$ ,  $\varepsilon > 0$  and  $\alpha > 0$  then an appropriate choice of  $0 < \underline{u}, \bar{u} < 1$  and thus  $\underline{h}$  it holds, for all  $n$  large enough, by Lemma 21 and step 4,

$$\begin{aligned} & \mathbb{E} \left( \Psi(v_n(I_{\mathcal{K}_n} + I_{\mathcal{L}_{1,n}} + I_{\mathcal{L}_{2,n}} + I_{\mathcal{L}_{3,n}})) - \Psi(v_n I_{\mathcal{L}_{3,n}}^*) \right) \\ & \leq 4m\mathbb{P}(v_n I_{\mathcal{K}_n} > \lambda) + 4m\mathbb{P}(v_n |I_{\mathcal{L}_{3,n}} - I_{\mathcal{L}_{3,n}}^*| > \lambda) \\ & + 4m\mathbb{P}\left(v_n I_{\mathcal{L}_{1,n}} > \frac{(5\alpha)^b}{\underline{h}^b}\right) + k\mathbb{E}\left(4\lambda + \frac{(5\alpha)^b}{\underline{h}^b} + v_n I_{\mathcal{L}_{2,n}}\right) \\ & \leq 12m\varepsilon + 4k\lambda + \frac{k(5\alpha)^b}{\underline{h}^b} + 2kC \exp\left(-\frac{1}{\alpha^2}\right) \end{aligned}$$

which is as small as desired. Finally it is easily seen that  $v_n I_{\mathcal{L}_{3,n}}^* \rightarrow_{weak} W$  as  $(\underline{u}, \bar{u}) \rightarrow (0, 1)$  and  $\alpha \rightarrow 0$  so that  $\mathbb{E}(\Psi(W))$  can replace  $\mathbb{E}(\Psi(v_n I_{\mathcal{L}_{3,n}}^*))$  above with an asymptotically arbitrarily small error.  $\square$

## 5.2 The case $F < G$

We establish Theorem 9.

**Step 0.** In this section  $D = \mathbb{R}$ . Without loss of generality, assume that  $F^{-1} > G^{-1}$  everywhere. We again focus on arguments for the right hand tail, thus we write  $\psi_X = \psi_X^+$  and  $\psi_Y = \psi_Y^+$  on  $(y_0, +\infty)$ . Therefore  $\psi_X^{-1} > \psi_Y^{-1}$  and  $\psi_X^{-1} > 0$  on  $(u_0, 1)$  where  $u_0 = F^{-1}(y_0)$ . We need this stochastic ordering only to simplify the control of extremes without imposing  $(CFG_E)$ . Let assume  $(FG)$ ,  $(C)$  with  $b \in [1, 2)$  and  $(CFG_D)$ . For  $y$  large it holds  $\rho_{\pm}(y) = \exp(l_{\pm}(y))$  with  $l_{\pm} \in RV_2^+(\gamma_{\pm}, +\infty)$ . By (15), for  $y_0 > 0$  and  $\theta_+, \theta_- > 1$  playing exactly the role of  $\theta$  in (CFG) of [2] we have

$$(\psi_X \circ l_+^{-1})'(y) \geq 2 + \frac{2\theta_+}{y}, \quad (\psi_Y \circ l_-^{-1})'(y) \geq 2 + \frac{2\theta_-}{y}, \quad y > y_0. \quad (45)$$

In particular, this implies

$$l_+ \circ \psi_X^{-1}(y) \leq \frac{y}{2} - \theta_+ \log y + K, \quad y > y_0. \quad (46)$$

By (16), whenever  $F^{-1}(u) - G^{-1}(u) > 0$  is not asymptotically away from 0 as  $u \rightarrow 1$  we further ask that, for some  $\theta_2 > 0$ ,

$$l_+ \circ \psi_X^{-1}(y) \leq \frac{y}{2} - 2 \log \psi_X^{-1}(y) - \theta_2 \log y, \quad y > y_0. \quad (47)$$

Notice that if  $F$  is logconvex then  $\log \psi_X^{-1}(y) > \log y$  and (47) already implies (46) with  $\theta_+ > 2$  whereas if  $F$  is logconcave then  $\log \psi_X^{-1}(y) < \log y$  and (46) implies (47) with  $\theta_2 > 1$ . Since  $(CFG_D)$  implies  $(CFG)$  of [2] through (16) hence (45), we are allowed to use most results of the latter paper. In particular Theorem 9 is true when  $F^{-1}(u) - G^{-1}(u) > \delta$  for some  $\delta > 0$  and  $b > 1$  to ensure  $(C3)$  in [2]. We thus focus on the case  $F^{-1}(u) - G^{-1}(u) \rightarrow 0$  as  $u \rightarrow 1$  which requires (47) whatever  $b$ , and we isolate out the case  $b = 1$  only when necessary to extend the main result of [2], at step 4. We often use  $F^{-1}(u) = \psi_X^{-1}(\log(1/(1-u)))$ . A consequence is that (47) also reads

$$\rho \circ F^{-1}(u) = \rho_+ \circ F^{-1}(u) \leq \frac{1}{F^{-1}(u)^2 \sqrt{1-u} |\log(1-u)|^{\theta_2}}, \quad u > u_0.$$

Let study  $W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G) = I_{\mathcal{I}_n} + I_{\mathcal{J}_n} + I_{\mathcal{K}_n} + I_{\mathcal{L}}$  with the notation

$$I_A = \int_A (\rho_c(\tau(u) + \tau_n(u)) - \rho_c(\tau(u))) du, \quad A \subset (0, 1), \quad (48)$$

$$\tau(u) = F^{-1}(u) - G^{-1}(u), \quad \tau_n(u) = \frac{\beta_n(u)}{\sqrt{n}} = \frac{\beta_n^X(u) - \beta_n^Y(u)}{\sqrt{n}},$$

and  $\mathcal{I}_n = (1 - i_n/n, 1]$ ,  $\mathcal{J}_n = (1 - j_n/n, 1 - i_n/n]$ ,  $\mathcal{K}_n = (\underline{u}, 1 - j_n/n]$ ,  $\mathcal{L} = [\underline{u}, \bar{u}]$  with  $0 < \underline{u} < 1/2 < \bar{u} < 1$ .

**Step 1.** Consider a non negative increasing sequence  $K_n \rightarrow +\infty$  to be chosen later in such a way that  $K_n / \log \log n \rightarrow 0$ . Define

$$i_n = \frac{\sqrt{n}}{K_n \exp(l \circ \psi_X^{-1}(\log n + K_n))}. \quad (49)$$

We have  $l \circ \psi_X^{-1}(y) = l_+ \circ \psi_X^{-1}(y) \rightarrow +\infty$  as  $y \rightarrow +\infty$  thus  $i_n = o(\sqrt{n}/K_n)$ . When (47) is enforced then for any  $\theta' \in (0, \theta_2)$  and all  $n$  large enough,

$$\begin{aligned} i_n &\geq \frac{K}{K_n} \left( F^{-1} \left( 1 - \frac{1}{ne^{K_n}} \right) \right)^2 \exp \left( -\frac{K_n}{2} + \theta_+ \log(\log n + K_n) \right) \\ &> \left( F^{-1} \left( 1 - \frac{1}{n} \right) \right)^2 (\log n)^{\theta'}. \end{aligned} \quad (50)$$

Otherwise, when only (46) holds then for  $\theta' \in (1, \theta_+)$ ,

$$i_n \geq \frac{K}{K_n} \exp \left( -\frac{K_n}{2} + \theta_+ \log(\log(n + K_n)) \right) > (\log n)^{\theta'}. \quad (51)$$

Hence in both case we have  $i_n / \log \log n \rightarrow +\infty$  and  $i_n / \sqrt{n} \rightarrow 0$ . Let us define

$$\begin{aligned} I_{\mathcal{I}_n}^1 &= \int_{\mathcal{I}_n} \rho_+(\tau(u)) du, \\ I_{\mathcal{I}_n}^2 &= \int_{\mathcal{I}_n} \rho_c(\mathbb{F}_n^{-1}(u) - \mathbb{G}_n^{-1}(u)) du = \frac{1}{n} \sum_{i=n-[i_n]}^n \rho_c(X_{(i)} - Y_{(i)}). \end{aligned}$$

**Lemma 22** Assume that (C), (FG) and (CFG<sub>D</sub>) hold. Then  $\sqrt{n}I_{\mathcal{I}_n}^1 \rightarrow 0$  and  $\sqrt{n}I_{\mathcal{I}_n}^2 \rightarrow 0$  in probability.

**Proof.** This readily follows from Lemma 21 in [2]. For  $\sqrt{n}I_{\mathcal{I}_n}^1$  the mentioned proof only needed  $\theta > 0$  hence  $\theta_+, \theta_- > 0$ . For  $\sqrt{n}I_{\mathcal{I}_n}^2$  the initial expansion

$$\sum_{i=n-[i_n]}^n \rho_c(X_{(i)} - Y_{(i)}) \leq \sum_{i=n-[i_n]}^n \rho_+(X_{(i)}) + \sum_{i=n-[i_n]}^n \rho_-(Y_{(i)})$$

almost surely holds for  $n$  large enough, when  $\min(X_{(n-[i_n])}, Y_{(n-[i_n])}) > 0$ .  $\square$

**Step 2.** We now study  $I_{\mathcal{J}_n}$  with  $j_n = n^\beta$ ,  $\beta \in (1/2, 1)$ . Recall that  $\Delta_n = \mathcal{J}_n \cup \mathcal{K}_n \cup \mathcal{L}$  and  $\tau(u) = F^{-1}(u) - G^{-1}(u) > 0$  for all  $u \in \Delta_n$ .

(i) Define  $\varepsilon_n = \sup_{u \in \Delta_n} \varepsilon_n(u)$  where  $\varepsilon_n(u) = \varepsilon_n^X(u) + \varepsilon_n^Y(u)$  and

$$\varepsilon_n^X(u) = \frac{\sqrt{\log \log n}}{\sqrt{n}} \frac{\sqrt{1-u}}{h_X(u)}, \quad \varepsilon_n^Y(u) = \frac{\sqrt{\log \log n}}{\sqrt{n}} \frac{\sqrt{1-u}}{h_Y(u)}.$$

The current  $\varepsilon_n$  is bounded by the one of (28). By combining (27) and (28) with (50) as in Lemma 18 we get, for some  $\zeta > 0$ ,

$$\lim_{n \rightarrow +\infty} (\log n)^\zeta \sup_{u \in \Delta_n} \tau_n(u) \leq 9 \lim_{n \rightarrow +\infty} (\log n)^\zeta \varepsilon_n = 0 \quad a.s.$$

Let  $m_n \rightarrow +\infty$  be a non negative sequence so slow that  $m_n \varepsilon_n \rightarrow 0$ . Consider  $\mathcal{J}_n = \mathcal{J}_n^< \cup \mathcal{J}_n^>$  where

$$\begin{aligned} \mathcal{J}_n^< &= \{u \in \mathcal{J}_n : 0 < \tau(u) \leq m_n \varepsilon_n(u)\}, \\ \mathcal{J}_n^> &= \{u \in \mathcal{J}_n : 0 < m_n \varepsilon_n(u) < \tau(u)\}. \end{aligned}$$

By (27) again we almost surely ultimately have

$$-9\varepsilon_n(u) < \tau_n(u) = \frac{\beta_n^X(u)}{\sqrt{n}} - \frac{\beta_n^Y(u)}{\sqrt{n}} < 9\varepsilon_n(u), \quad u \in \mathcal{J}_n.$$

Notice that if  $u \in \mathcal{J}_n^>$  then

$$0 < (m_n - 9)\varepsilon_n(u) < \tau(u) + \tau_n(u) < \tau(u) + 9\varepsilon_n(u) < \tau(u) \left(1 + \frac{9}{m_n}\right) \quad (52)$$

whereas if  $u \in \mathcal{J}_n^<$  then it is possible that  $\tau(u) + \tau_n(u) < 0$  since

$$-9\varepsilon_n(u) < \tau_n(u) < \tau(u) + \tau_n(u) < (m_n + 9)\varepsilon_n(u). \quad (53)$$

Let us control  $|I_{\mathcal{J}_n}| \leq |I_{\mathcal{J}_n^<}| + |I_{\mathcal{J}_n^>}|$ , starting with the first term.

(ii) Recall that  $\sup_{u \in \mathcal{J}_n^<} m_n \varepsilon_n(u) \rightarrow 0$  as  $n \rightarrow +\infty$ . By (53) we have, for  $u \in I_{\mathcal{J}_n^<}$  and  $m_n > 9$ ,

$$\begin{aligned} |\rho_c(\tau(u) + \tau_n(u)) - \rho_c(\tau(u))| &\leq \rho_c(\tau(u) + \tau_n(u)) + \rho_+(\tau(u)) \\ &\leq \rho_-(9\varepsilon_n(u)) + 2\rho_+(2m_n\varepsilon_n(u)). \end{aligned}$$

hence  $\sqrt{n}|I_{\mathcal{J}_n^<}| \leq R_{1,n} + R_{2,n}$  for all  $n$  large enough, with

$$\begin{aligned} R_{1,n} &= K\sqrt{n} \int_{\mathcal{J}_n^<} \varepsilon_n(u)^{b_-} L_-(9\varepsilon_n(u)) du, \\ R_{2,n} &= K\sqrt{n} \int_{\mathcal{J}_n^<} (m_n \varepsilon_n(u))^{b_+} L_+(2m_n \varepsilon_n(u)) du. \end{aligned}$$

**Lemma 23** Assume (C), (FG) and (CFG). We have  $R_{1,n} \rightarrow 0$  and  $R_{2,n} \rightarrow 0$ .

**Proof.** If  $F^{-1}(u) - G^{-1}(u) > \delta$  then the set  $\mathcal{J}_n^<$  is ultimately empty. Otherwise (47) holds. We have  $\sqrt{1-u}(1/h_X(u) + 1/h_Y(u)) \leq 2F^{-1}(u)/\sqrt{1-u}$  for  $u \in \mathcal{J}_n$  in view of  $F^{-1}(u) > G^{-1}(u)$  and (FG3). If  $\min(b_+, b_-) - 1 > 0$  this extra power cancels the slowly varying functions and we asymptotically have

$$R_{1,n} + R_{2,n} \leq K\sqrt{n} \int_{\mathcal{J}_n^<} m_n \varepsilon_n(u) du \leq Km_n \sqrt{\log \log n} \int_{\mathcal{J}_n} \frac{F^{-1}(u)}{\sqrt{1-u}} du.$$

If  $b_+ = 1$  then  $L_+(x)$  is bounded on  $[0, x_0]$  since  $xL_+(x)$  is convex non negative and starts from 0. Hence  $L_+(2m_n \varepsilon_n(u))$  is bounded on  $\mathcal{J}_n$ , and the above upper bound remains true. Likewise if  $b_- = 1$  then  $L_-(9\varepsilon_n(u))$  is bounded on  $\mathcal{J}_n$ . Observe that (47) and  $l(y) > \log y$  imply

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

thus  $\psi_X^{-1}(y)^6 \leq e^y$  and  $F^{-1}(u) < 1/(1-u)^{1/6}$ . Therefore

$$\int_{\mathcal{J}_n} \frac{F^{-1}(u)}{\sqrt{1-u}} du \leq K \left(\frac{j_n}{n}\right)^{1/3} = Kn^{(\beta-1)/3}$$

with  $\beta < 1$  and the conclusion follows since  $m_n \rightarrow +\infty$  is arbitrarily slow.  $\square$

We have shown that  $\sqrt{n}I_{\mathcal{J}_n^<} \rightarrow 0$  almost surely.

(iii) By (52) we ultimately have, for all  $u \in \mathcal{J}_n$ ,

$$|\rho_c(\tau(u) + \tau_n(u)) - \rho_c(\tau(u))| = |\rho_+(\tau(u) + \tau_n(u)) - \rho_+(\tau(u))|.$$

Consider now  $\mathcal{J}_n^> = \mathcal{J}_n^{<\delta} \cup \mathcal{J}_n^{>\delta}$  with

$$\mathcal{J}_n^{<\delta} = \{u \in \mathcal{J}_n : m_n \varepsilon_n(u) < \tau(u) < \delta\}, \quad \mathcal{J}_n^{>\delta} = \{u \in \mathcal{J}_n : \tau(u) > \delta\}.$$

Since  $\tau(u) > \delta$  on  $\mathcal{J}_n^{>\delta}$  and Proposition 31 and Lemma 24 of [2] are satisfied by  $\rho_+$  – thanks to (7) and (8) – we readily deduce from Lemma 22 of [2] that

$$\lim_{n \rightarrow +\infty} \sqrt{n} \int_{\mathcal{J}_n^{>\delta}} |\rho_+(\tau(u) + \tau_n(u)) - \rho_+(\tau(u))| du = 0 \quad a.s.$$

Concerning  $\mathcal{J}_n^{<\delta}$  observe that by (52) again  $0 < \tau(u) + \tau_n(u) < 2\delta$  for all  $n$  large. Since  $\rho_+$  is convex it ensues

$$\begin{aligned} |\rho_+(\tau(u) + \tau_n(u)) - \rho_+(\tau(u))| &\leq \max(\rho'_+(\tau(u) + \tau_n(u)), \rho'_+(\tau(u))) |\tau_n(u)| \\ &\leq K_\delta |\tau_n(u)| \end{aligned}$$

with  $K_\delta = \rho'_+(2\delta)$ . Therefore, with probability one, for all  $n$  large enough

$$\sup_{u \in \mathcal{J}_n^{<\delta}} |\rho_+(\tau(u) + \tau_n(u)) - \rho_+(\tau(u))| \leq K_\delta \sup_{u \in \mathcal{J}_n^{<\delta}} |\tau_n(u)| \leq K \sup_{u \in \mathcal{J}_n^{<\delta}} |\varepsilon_n(u)|.$$

As already seen, (16) implies  $F^{-1}(u) < 1/(1-u)^{1/6}$ . As a consequence, with probability one it ultimately holds

$$\begin{aligned} \sqrt{n} \int_{\mathcal{J}_n^{<\delta}} |\rho_+(\tau(u) + \tau_n(u)) - \rho_+(\tau(u))| du &\leq K \sqrt{n} \int_{\mathcal{J}_n} \varepsilon_n(u) du \\ &\leq K \sqrt{\log \log n} \int_{\mathcal{J}_n} \frac{F^{-1}(u)}{\sqrt{1-u}} du \leq K n^{(\beta-1)/3} \sqrt{\log \log n} \end{aligned}$$

which vanishes as  $n \rightarrow +\infty$ . We conclude that  $\sqrt{n}I_{\mathcal{J}_n^>} \rightarrow 0$  almost surely.

**Step 3.** The convergence of  $I_{\mathcal{K}_n}$  is weaker than in probability.

**Lemma 24** *Assume (FG), (C) and (CFG<sub>D</sub>). There exists  $\beta \in (1/2, 1)$  such that for any choice of  $\lambda > 0$  and  $\varepsilon > 0$  one can find  $\bar{u} \in (1/2, 1)$  and  $n_0 > 0$  such that, for all  $n > n_0$ ,*

$$\mathbb{P}(\sqrt{n}I_{\mathcal{K}_n} > \lambda) < \varepsilon.$$

**Proof.** Fix  $\delta > 0$  and consider

$$\mathcal{K}_n^{<\delta} = \{u \in \mathcal{K}_n : 0 < \tau(u) < \delta\}, \quad \mathcal{K}_n^{>\delta} = \{u \in \mathcal{K}_n : \tau(u) > \delta\}.$$



The claimed result holds for  $I_{\mathcal{K}_n^{>\delta}}$  by applying Lemma 25 from [2] with  $\delta = \tau_0$  and  $\bar{u} = F(M)$ . Let us apply Lemma 20 to get, for  $K > \sup_{|x| < 2\delta} \rho'_c(x)$ ,

$$\begin{aligned} \sqrt{n} I_{\mathcal{K}_n^{<\delta}} &= \sqrt{n} \int_{\mathcal{K}_n^{<\delta}} |\rho_c(\tau(u) + \tau_n(u)) - \rho_c(\tau(u))| du \leq K \int_{\mathcal{K}_n^{<\delta}} |\beta_n(u)| du \\ &\leq K \int_{\mathcal{K}_n} \left( \frac{|B_n^X(u)|}{h_X(u)} + \frac{|B_n^Y(u)|}{h_Y(u)} \right) du + \int_{\mathcal{K}_n} \left( \frac{|Z_n^X(u)|}{h_X(u)} + \frac{|Z_n^Y(u)|}{h_Y(u)} \right) du. \end{aligned}$$

The first two terms satisfy

$$\begin{aligned} \mathbb{E} \left( \int_{\mathcal{K}_n} \frac{|B_n^X(u)|}{h_X(u)} du \right) &\leq \int_{\mathcal{K}_n} \frac{\sqrt{1-u} \mathbb{E}(|B_n^X(u)|)}{h_X(u) \sqrt{u(1-u)}} du \\ &\leq \int_{\bar{u}}^1 \frac{F^{-1}(1-u)}{\sqrt{1-u}} du \leq 3(1-\bar{u})^{1/3} \end{aligned}$$

and the last two terms obey, with probability one as  $n \rightarrow +\infty$ ,

$$\begin{aligned} \int_{\mathcal{K}_n} \frac{|Z_n^X(u)|}{h_X(u)} du &\leq \sup_{u \in \mathcal{K}_n} |Z_n^X(u)| \int_{\mathcal{K}_n} \frac{F^{-1}(1-u)}{1-u} du \\ &\leq \frac{1}{n^v} \int_{\bar{u}}^{1-j_n/n} \frac{F^{-1}(1-u)}{1-u} du \\ &\leq \frac{1}{n^v} \int_{\bar{u}}^{1-j_n/n} \frac{1}{(1-u)^{7/6}} du \leq \frac{6}{n^v} n^{(1-\beta)/6} \end{aligned}$$

which vanishes if  $\beta > 1 - 6v$  is chosen close enough to 1.  $\square$

**Step 4.** Here we recall that (C2) with  $b_{\pm} > 1$  and (15) respectively imply (C3) and (CFG) in [2]. Clearly steps 4 and 5 of [2] remain true in the current framework and lead to the same conclusion as the main theorem in the latter paper, whence Theorem 9. The new case to conclude with is  $b = 1$ . Since we almost surely have

$$0 \leq |\tau_n(u)| < \underline{\tau} = \min_{u \in \mathcal{L}} \tau(u)$$

for all  $n$  large enough, we only deal with  $\rho_+$ . Assuming that  $b_+ = 1$  and  $\rho_+(x) = xL_+(x)$  we have, for some  $\varepsilon > 0$  such that  $\mathcal{L}_\varepsilon \subset (0, 1)$  is an  $\varepsilon$ -neighborhood of  $\mathcal{L}$ ,

$$\begin{aligned} &\left| \sqrt{n} \int_{\mathcal{L}} (\rho_+(\tau(u) + \tau_n(u)) - \rho_+(\tau(u))) du - \sqrt{n} \int_{\underline{u}}^{\bar{u}} \rho'_+(\tau(u)) \tau_n(u) du \right| \\ &\leq \frac{\sqrt{n}}{2} \sup_{u \in \mathcal{L}_\varepsilon} |\rho''_+(\tau(u))| \int_{\underline{u}}^{\bar{u}} \tau_n^2(u) du \leq \frac{K}{\sqrt{n}} \int_{\underline{u}}^{\bar{u}} \beta_n^2(u) du \end{aligned}$$

which almost surely vanishes by the law of the iterated logarithm. Thus we can conclude as in [2] by combining this with the previous steps 1, 2, 3. In particular, the limiting variance is finite as a consequence of (15).

### 5.3 The general case

We now prove Theorem 11. Recall that  $(FG0)$  implies the existence of  $0 = u_0 < u_1 < \dots < u_\kappa = 1$  such that  $F^{-1}(u_k) = G^{-1}(u_k)$  and  $A_k = (u_{k-1}, u_k) \subset E$  or  $A_k \subset D$  for  $k = 1, \dots, \kappa$ . We now study the mixed case where at least one of these intervals is included in  $E$  and one in  $D$ , so that  $\kappa \geq 2$ . Consider, using notation (48),

$$\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)) = \sqrt{n} \sum_{k=1}^{\kappa} I_{A_k}.$$

Let  $0 \leq \lambda < \min_{1 \leq k \leq \kappa} (u_k - u_{k-1})/2$ . Define the intervals  $A_{k,\lambda}^+ = (u_{k-1}, u_{k-1} + \lambda) \subset A_k$  for  $2 \leq k \leq \kappa$  and  $A_{k,\lambda}^- = (u_k - \lambda, u_k) \subset A_k$  for  $1 \leq k \leq \kappa - 1$ . If  $A_k \subset D$  we have  $F^{-1}(u) \neq G^{-1}(u)$  for  $u \in A_{k,\lambda}^+ \cup A_{k,\lambda}^-$ . If  $A_k \subset E$  the intervals  $A_{k,\lambda}^+$  and  $A_{k,\lambda}^-$  are assumed to be empty instead. Consider first the intervals  $A_{k,\lambda}^+$  for  $2 \leq k \leq \kappa$  and set  $0 < u_- < u_1 < u_{\kappa-1} < u_+ < 1$ . Since

$$\lim_{n \rightarrow +\infty} \sup_{u_- < u < u_+} \left| \frac{\beta_n(u)}{\sqrt{n}} \right| = 0 \quad a.s.$$

we have, by (C2), for  $K = \sup_{u_- < u < u_+} (\rho'_-(2|\tau(u)|), \rho'_+(2|\tau(u)|)) < +\infty$ ,

$$\lim_{n \rightarrow +\infty} \sup_{u_- < u < u_+} \frac{\rho_c(\beta_n(u)/\sqrt{n})}{|\beta_n(u)/\sqrt{n}|} \leq K \quad a.s.$$

Therefore, in view of step 4 in the previous proof for  $F \neq G$  we get

$$\sqrt{n} |I_{A_{k,\lambda}^+}| \leq K \int_{A_{k,\lambda}^+} |\beta_n(u)| du \leq \frac{K}{\underline{h}} \left( \int_{A_{k,\lambda}^+} |B_n(u)| du + \int_{A_{k,\lambda}^+} |Z_n(u)| du \right)$$

where  $\underline{h} = \min_{u_- < u < u_+} \min(h_X(u), h_X(u)) > 0$ . Lemma 20 further yields

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left( \sqrt{n} |I_{A_{k,\lambda}^+}| > \alpha \right) \leq \mathbb{P} \left( \int_{A_{k,\lambda}^+} |B(u)| du > \frac{2\alpha \underline{h}}{K} \right)$$

for any  $\alpha > 0$  and all  $2 \leq k \leq \kappa$ , where  $B$  has the same law as  $B_n = B_n^X - B_n^Y$ . The latter upper bound vanishes as  $\lambda \rightarrow 0$ . A similar conclusion holds for  $A_{k,\lambda}^-$  and  $1 \leq k \leq \kappa - 1$ . Write  $A_{1,\lambda}^* = A_1 \setminus A_{1,\lambda}^-$ ,  $A_{\kappa,\lambda}^* = A_\kappa \setminus A_{\kappa,\lambda}^+$  and  $A_{k,\lambda}^* = A_k \setminus (A_{k,\lambda}^+ \cup A_{k,\lambda}^-)$  for  $2 \leq k \leq \kappa - 1$ .

(i) Consider the case  $1 < b < 2$ . Fix  $\lambda > 0$  arbitrarily small and write

$$\sqrt{n}(W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)) = \sqrt{n}I_E + \sqrt{n}I_{D,\lambda}^* + \sqrt{n}I_{D,\lambda}^\pm \quad (54)$$

where

$$I_E = \sum_{A_k \subset E} I_{A_k^*} = \sum_{A_k \subset E} I_{A_k}, \quad I_{D,\lambda}^* = \sum_{A_k \subset D} I_{A_{k,\lambda}^*}, \quad I_{D,\lambda}^\pm = \sum_{A_k \subset D} I_{A_{k,\lambda}^+ \cup A_{k,\lambda}^-}.$$

We just proved that

$$\lim_{\lambda \rightarrow 0} \lim_{n \rightarrow +\infty} \mathbb{P} \left( \sqrt{n} I_{D,\lambda}^\pm > \alpha \right) = 0.$$

Since  $b > 1$  we have  $v_n/\sqrt{n} \rightarrow 0$  as  $n \rightarrow +\infty$ . Therefore step 1 to 4 of the proof 4.1 when  $F = G$  show that

$$\lim_{n \rightarrow +\infty} \sqrt{n} I_E = \lim_{n \rightarrow +\infty} \frac{\sqrt{n}}{v_n} v_n I_E = 0 \quad \text{in probability.}$$

In the case  $\kappa \geq 3$  then for all  $2 \leq k \leq \kappa - 1$  with  $A_k \subset D$  we have  $\delta_k = \inf_{u \in A_k^*} |\tau(u)| > \delta > 0$  and  $\tau(u)$  has constant sign on  $A_k$ . It follows from step 1 to 4 of the proof 4.2 when  $F \neq G$  that the weak limit of  $\sqrt{n} I_{D,\lambda}^*$  is  $\int_{D_\lambda} \rho'_c(\tau(u)) \mathbb{B}(u) du$  where  $D_\lambda = \bigcup_{A_k \subset D} I_{A_{k,\lambda}}^*$  and  $\mathbb{B}(u) = B^X(u)/h_X(u) - B^Y(u)/h_Y(u)$ . By letting  $\lambda \rightarrow 0$  we conclude that

$$\sqrt{n} (W_c(\mathbb{F}_n, \mathbb{G}_n) - W_c(F, G)) \rightarrow_{weak} \int_D \rho'_c(\tau(u)) \frac{B(u)}{h(u)} du$$

which is easily seen to have the normal distribution  $\mathcal{N}(0, \sigma_D^2)$ .

(ii) Assume that  $b = 1$ . Starting again from (54) we again obtain that

$$\sqrt{n} I_{D,\lambda}^* \rightarrow_{weak} \int_{D_\lambda} \rho'_c(\tau(u)) \mathbb{B}(u) du$$

while the steps 1 to 4 of the proof 4.1 now entails, for  $v_n$  from (19),

$$v_n I_E \rightarrow_{weak} \pi_+ \int_E 1_{\{\mathbb{B}(u) > 0\}} |\mathbb{B}(u)| du + \pi_- \int_E 1_{\{\mathbb{B}(u) < 0\}} |\mathbb{B}(u)| du.$$

Finally observe that (C4) implies  $\sqrt{n}/v_n \rightarrow L_+(0)/\pi_+$  and  $\sqrt{n}/v_n \rightarrow L_-(0)/\pi_-$  as  $n \rightarrow +\infty$ . As previously we conclude by letting  $\lambda \rightarrow 0$ .

## 5.4 A special case : $F = G$ and $b = 2$

We establish Theorem 14.

**Step 0.** Assume (C0),  $\rho_c(x) = x^2$  for  $|x| < x_0$ ,  $E = \mathbb{R}$ , (FG1), (FG2) and

$$\lim_{u \rightarrow 0} \frac{u}{h(u)} = \lim_{u \rightarrow 1} \frac{1-u}{h(u)} = 0, \quad \int_0^1 \frac{u(1-u)}{h^2(u)} du < +\infty.$$

This proof partially follows the line of the proof of Lemma 2.4 of [10].

**Step 1.** We show that  $\sup_{1/n \leq u \leq 1-1/n} |\mathbb{F}_n^{-1}(u) - \mathbb{G}_n^{-1}(u)| \rightarrow 0$  in probability, so that the behaviour of  $\rho_c$  near 0 only matters. Write  $h = h_X$ . Define  $U_i = F(X_i)$

and  $V_i = F(Y_i)$ ,  $i = 1, \dots, n$ . Consider  $nI_{\mathcal{I}_n}$  with  $i_n = 1$  and

$$\begin{aligned} I_{\mathcal{I}_n} &= \int_{1-1/n}^1 (\mathbb{F}_n^{-1}(u) - \mathbb{G}_n^{-1}(u))^2 du \\ &= \int_{1-1/n}^1 (F^{-1}(U_{(n)}) - F^{-1}(V_{(n)}))^2 du \\ &\leq \frac{2}{n} \left( F^{-1}(U_{(n)}) - F^{-1}(1 - \frac{1}{n}) \right)^2 \\ &\quad + \frac{2}{n} \left( F^{-1}(V_{(n)}) - F^{-1}(1 - \frac{1}{n}) \right)^2. \end{aligned}$$

By the mean theorem, for some random  $U_{(n)}^*$  between  $U_{(n)}$  and  $1 - 1/n$ ,

$$F^{-1}(U_{(n)}) - F^{-1}(1 - \frac{1}{n}) = \frac{U_{(n)} - 1 + 1/n}{h(U_{(n)}^*)} = \frac{U_{(n)} - 1 + 1/n}{h(U_{(n)})} \frac{h(U_{(n)})}{h(U_{(n)}^*)}.$$

By a classical argument – see [BFK17] – we have, thanks to (FG2),

$$\max \left( \frac{h(U_{(n)})}{h(U_{(n)}^*)}, \frac{h(U_{(n)}^*)}{h(U_{(n)})} \right) \leq \max \left( \frac{1 - U_{(n)}}{1 - U_{(n)}^*}, \frac{1 - U_{(n)}^*}{1 - U_{(n)}} \right)^K.$$

Now recall that  $U_{(n)} - 1 + 1/n = O_P(1/n)$  and  $d_{(n)} = n(1 - U_{(n)}) \rightarrow_{weak} d_{(\infty)}$  where  $d_{(\infty)}$  is a positive finite random variable. Hence

$$\begin{aligned} \left( F^{-1}(U_{(n)}) - F^{-1}(1 - \frac{1}{n}) \right)^2 &\leq \frac{(U_{(n)} - 1 + 1/n)^2}{h^2(U_{(n)})} \max \left( \frac{1}{d_{(n)}}, d_{(n)} \right)^{2K} \\ &= \frac{(1 - U_{(n)})^2}{h^2(U_{(n)})} \left( 1 - \frac{1}{d_{(n)}} \right)^2 \max \left( \frac{1}{d_{(n)}}, d_{(n)} \right)^{2K} \end{aligned}$$

where  $(1 - U_{(n)})^2/h^2(U_{(n)}) \rightarrow 0$  almost surely and  $(1 - 1/d_{(n)})^2 \max(1/d_{(n)}, d_{(n)})^{2K} = O_P(1)$ . Hence  $nI_{\mathcal{I}_n} = o_P(1)$ .

**Step 2.** Now consider, for  $j_n = n^\beta$ ,

$$nI_{\mathcal{J}_n} = \int_{1-j_n/n}^{1-1/n} (\beta_n^X(u) - \beta_n^Y(u))^2 du.$$

**Lemma 25** *There exists a sequence of processes  $\mathbb{B}_n^X$  having the same law as  $\mathbb{B}^X$  of (18) such that*

$$\Xi_n = \sup_{1/n \leq u \leq 1-1/n} |\beta_n^X(u) - \mathbb{B}_n^X(u)| \frac{h(u)}{\sqrt{1-u}} = O_P(1).$$

*Proof.* It is an immediate extension of Corollary 4.2.1. page 382 of [5] starting from (4.2.2) of Theorem 4.2.1 of [5].  $\square$

As a consequence,

$$\begin{aligned} \mathbb{P}(nI_{\mathcal{J}_n} > 3\alpha) &\leq \mathbb{P}\left(\int_{1-j_n/n}^{1-1/n} (\mathbb{B}_n^X(u) - \mathbb{B}_n^Y(u))^2 du > \alpha\right) \\ &\quad + 2\mathbb{P}\left(\Xi_n^2 \int_{1-j_n/n}^{1-1/n} \frac{1-u}{h^2(u)} du > \alpha\right) \end{aligned}$$

hence  $nI_{\mathcal{J}_n} \rightarrow 0$  in probability. We conclude the proof by applying the steps 3 to 5 in Section 5.1 with many simplifications since  $L(x) = 1$  now.

## 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):358–375, 2011.
- [2] P. Berthet, J.C. Fort, and T. Klein. A central limit theorem for wasserstein type distances between two different laws. *Preprint, hal-01526879*, 2017.
- [3] S. G. Bobkov and M. Ledoux. One-dimensional empirical measures, order statistics and kantorovich transport distances. *To appear in: Memoirs of the AMS*, Preprint 2016.
- [4] S. Cambanis, G. Simons, and W. Stout. Inequalities for  $Ek(X, Y)$  when the marginals are fixed. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 36(4):285–294, 1976.
- [5] M. Csörgö and L. Horváth. *Weighted approximations in probability and statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. 1993.
- [6] M. Csörgö, L. Horváth, and Q.-M. Shao. Convergence of integrals of uniform empirical and quantile processes. *Stochastic Processes and their Applications*, 45(2):283–294, 1993.
- [7] C. Czado and A. Munk. Assessing the similarity of distributions—finite sample performance of the empirical Mallows distance. *J. Statist. Comput. Simulation*, 60(4):319–346, 1998.
- [8] 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):1009–1071, 1999.
- [9] 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):1142–1143, 2003.

- [10] 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):131–189, 2005.
- [11] E. del Barrio and J.-M. Loubès. Central limit theorems for empirical transportation cost in general dimension. *arXiv:1705.01299v2*, may 2017.
- [12] J.C. Fort and T. Klein. New fréchet features for random distributions and associated sensitivity indices. *Preprint, hal-01137577*, 2015.
- [13] M. Sommerfeld and A. Munk. Inference for Empirical Wasserstein Distances on Finite Spaces. *ArXiv e-prints*, October 2016.