The complex behaviour of Galton rank-order statistic

EUSTASIO DEL BARRIO 1,*, JUAN A. CUESTA-ALBERTOS2 and CARLOS MATRAN 1,†

1 Department of Statistics and Operational Research and IMUVA, University of Valladolid
E-mail: *eustasio.delbarrio@uva.es; †carlos.matran@uva.es
2 Department of Mathematics, Statistics and Computation, University of Cantabria
E-mail: cuestaj@unican.es

Galton’s rank-order statistic is one of the oldest statistical tools for two-sample comparisons. It is also a very natural index to measure departures from stochastic dominance. Yet, its asymptotic behaviour has been investigated only partially, under restrictive assumptions. This work provides a comprehensive study of this behaviour, based on the analysis of the so-called contact set (a modification of the set in which the quantile functions coincide).

We show that a.s. convergence to the population counterpart holds if and only if the contact set has zero Lebesgue measure. When this set is finite we show that the asymptotic behaviour is determined by the local behaviour of a suitable reparameterization of the quantile functions in a neighbourhood of the contact points. Regular crossings result in standard rates and Gaussian limiting distributions, but higher order contacts (in the sense introduced in this work) or contacts at the extremes of the supports may result in different rates and non-Gaussian limits.

Keywords: Relaxed stochastic dominance, asymptotics, consistency, Galton rank-order statistic, comparison of quantile functions, contact points, crossings, tangencies, contact intensity

1. Introduction and main results

The Introductory Remarks in Darwin’s report on the benefits of cross-fertilization to the propagation of vegetal species [12] include the following comment, by Galton: “The observations... have no primâ facie appearance of regularity. But as soon as we arrange them in order of their magnitudes,... We now see, with few exceptions, that... the largest plant on the crossed side... exceeds the largest plant on the self-fertilised side, that... the second exceeds the second,... and so on...”. With this argument, Galton opened a simple way of comparison of distributions, just by comparing the values with the same ranks in their respective settings.

Given two samples of equal size, $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively coming from the distribution functions (d.f.’s in the sequel) $F$ and $G$, let us denote’ by $F_n$ and $G_n$ the corresponding empirical d.f.’s. Galton’s solution consisted in reordering both data samples in increasing order: $X(1), \ldots, X(n)$ (coming from the control) and $Y(1), \ldots, Y(n)$ (from the treatment) and computing $G(F_n, G_n) := \# \{ i : X(i) > Y(i) \}$, concluding improvement under the treatment whenever $G(F_n, G_n)$ is small enough. When $F = G$ is continuous, the distribution of this ‘Galton Rank Order’ statistic is uniform on $\{0, 1, \ldots, n\}$ (see [9]; see also [21], [16] or [14] for alternative proofs). As explained in [16], in Darwin’s problem the sample sizes were 15 and $G(F_{15}, G_{15}) = 2$, thus the $p$-value associated to Galton’s approach is 3/16, which is not as rare as he suspected.

1 We have tried to use throughout standard or natural notation. However, a complete enough notation guide is included at the end of this section.
Galton’s strategy was related to the assessment of stochastic dominance of $G$ over $F$, $F \leq_{st} G$, being the alternative to the null hypothesis $F = G$. Recall that, by definition,

$$F \leq_{st} G \text{ whenever } F(x) \geq G(x) \text{ for every } x \in \mathbb{R}. $$

As noted in [17], this relation is better understood when it is stated in terms of the quantile functions: if $F^{-1}$ is the quantile function associated to $F$, defined by

$$F^{-1}(t) := \inf\{x : t \leq F(x)\}, \text{ for } t \in (0, 1),$$

then

$$F \leq_{st} G \text{ whenever } F^{-1}(t) \leq G^{-1}(t) \text{ for every } t \in (0, 1).$$

A useful feature of the quantile functions is that they provide a canonical representation of random variables (r.v.’s in that follows) with a given d.f.: if we consider the Lebesgue measure, $\ell$, on the unit interval $(0, 1)$, we set

$$\gamma(F, G) := \ell\{t : F^{-1}(t) > G^{-1}(t)\} \quad (2)$$

and observe that

$$G(F_n, G_m) = n\gamma(F_n, G_m). \quad (3)$$

Early work on Galton rank-order statistic mainly focused on the case $F = G$ and equal sample sizes. An exception is [6], that addressed the behaviour of a somewhat related statistic proposed by Hodges and Lehmann for robust estimation of location parameters. That statistic is based on the one sample Galton test and is analyzed in symmetric models. Concerning the two-sample statistic, special mention should be given to [10], which analyzes the joint behaviour of the Kolmogorov-Smirnov and Galton statistics (under $F = G$). Also for equal sample sizes, later, [15] considered the intermediate case with $F \neq G$ possibly, but $\ell\{F^{-1} = G^{-1}\} > 0$. Focusing on the dominance model $F = G$ vs $F \leq_{st} G$, [5] addressed the local asymptotic efficiency of $\gamma(F_n, G_m)$, noting that it is just a generalization of Galton’s statistic (recall (3)) and using empirical processes techniques to obtain the asymptotic distribution of $\gamma(F_n, G_m)$ under the null $F = G$ for independent samples with different sizes. Independently, looking for a feasible statistical way of relaxing the idea of “treatment improvement” underlying stochastic dominance, [1] used (2) as an index to measure deviation from stochastic dominance, $F \leq_{st} G$ and provided some asymptotic theory for the empirical index, for the case of d.f.’s with a single crossing point (the typical case in a location-scale family setting). In the same line, [22] adapted the theory to even cover the case of a finite number of crossings between the d.f.’s, under the additional assumption of an exponential density ratio model and using semiparametric estimates of the quantile functions.

Here, in a wide setting, we provide a complete set of distributional limit results for Galton rank-order statistic. In particular, we pursue on the goal of analyzing the scarcely treated case of a finite number of contact points between $F^{-1}$ and $G^{-1}$, leading to a sound study of the local behaviour at every isolated contact point between quantile functions. In summary, the paper uses the various properties of contact points (defined later) between $F^{-1}$ and $G^{-1}$ to provide a complete analysis of the complex asymptotic behaviour of the two sample Galton rank-order statistics $\gamma(F_n, G_m)$.

1.1. Main results

Here we describe the main results, whose proofs are deferred to Sections 3 and 4.
Intuitively, the asymptotic behaviour of $\gamma(F_n, G_m)$ depends on the size of the contact set, namely,

$$\Gamma := \{ t : F^{-1}(t) = G^{-1}(t) \}. \quad (4)$$

For equal sized samples this was already observed in [15]. We note that, since the index $\gamma$ is invariant with respect to strictly increasing transformations, the set $\Gamma$ could be equivalently expressed, in regular cases, as $\{ t : A(F^{-1}(t)) = 0 \}$, where $A(x) := G^{-1}(F(x)) - x$ is the shift function introduced in [13] as a richer alternative to the difference of means for comparing two continuous d.f.'s. The analysis of the Q-Q process associated to $A$ was done in [2], under smoothness assumptions, through strong approximations. Yet, intuition may fail without some regularity conditions and, as we show in this work, $\Gamma$ is not really the right set to look at. In fact, the asymptotic analysis of Galton’s rank statistic is better handled in terms of the alternative shift function $h(t) := F_G(t) - t$, underlying the associated P-P process considered in [3]. Here, and throughout this work, we denote $F_G := F \circ G^{-1}$ (similarly, $G_F = G \circ F^{-1}$) and

$$\hat{\Gamma} := \{ t : F_G(t) = t \}. \quad (5)$$

We observe that if $F$ and $G$ are continuous, then $\hat{\Gamma} = \Gamma$. However, these sets can be quite different: for $F = G$, a Bernoulli distribution with mean $p$, we have $\Gamma = [0,1]$, while $\hat{\Gamma} = \{1 - p, 1\}$. By focusing on the ‘right’ choice of contact set, our results go beyond the cases that could be treated from the analyses in [2] and [3]. In fact, we provide necessary and sufficient conditions for the a.s. consistency of $\gamma(F_n, G_m)$ without any smoothness assumption:

**Theorem 1.1.** Let $F, G$ be arbitrary d.f.'s. Then $\gamma(F_n, G_m) \overset{a.s.}{\longrightarrow} \gamma(F, G)$, as $n, m \to \infty$ if and only if $\ell(\hat{\Gamma}) = 0$.

A similar result holds for the one-sample statistic, $\gamma(F_n, G)$.

If $\ell(\hat{\Gamma}) > 0$, from Theorem 1.1, $\gamma(F_n, G_m)$ (or $\gamma(F_n, G)$) is not a consistent estimator of $\gamma(F, G)$.

In this case, we provide a completely general result about the asymptotic behaviour of $\gamma(F_n, G)$:

**Theorem 1.2.** Let $F, G$ be arbitrary d.f.'s, and $B$ a standard Brownian bridge on $[0,1]$. Then

$$\gamma(F_n, G) - \gamma(F, G) \overset{w}{\longrightarrow} \ell \left\{ t \in \hat{\Gamma} : B(t) > 0 \right\}, \text{ as } n \to \infty.$$

Still in the case $\ell(\hat{\Gamma}) > 0$, we prove weak convergence of the two-sample statistic under mild assumptions. This problem was also treated in [15] for equal sample sizes through combinatorial arguments and the method of moments. That combinatorial approach seems to be inappropriate to handle the case of unequal sample sizes. Additionally, our version yields a simple representation of the limit law.

**Theorem 1.3.** Let $F, G$ be d.f.’s such that $F_G$ is Lipschitz. If $B$ is a standard Brownian bridge on $[0,1]$ and $m, n \to \infty$ satisfy $0 < \lim inf \frac{n}{m+n} \leq \lim sup \frac{n}{m+n} < 1$, then

$$\gamma(F_n, G_m) - \gamma(F, G) \overset{w}{\longrightarrow} \ell \{ t \in \hat{\Gamma} : B(t) > 0 \}.$$

It should be noted that the limiting distribution in Theorems 1.2 or 1.3 is non-degenerate if and only if $\ell(\hat{\Gamma}) > 0$ (see Lemma 3.2 in Section 3). If $\hat{\Gamma} = (0, 1)$, a celebrated result by Paul Lévy (see Section 8, 2.0 in [18] or p. 85-86 in [7]) is that if $B$ is a standard Brownian bridge on $[0,1]$, then

$$P(\ell \{ t \in [0,1] : B(t) > 0 \} \leq x) = x, \text{ for every } x \in [0,1].$$
From this and Theorem 1.3 we recover, asymptotically, the classical result for the case $m = n$ and continuous $G = F$: in this case $\gamma(F_n, G_m)$ is uniformly distributed over $\{0, \frac{1}{n}, \ldots, \frac{n-1}{n}, 1\}$, continuity of $F$ ensures that $F(F^{-1}(t)) = t$ for every $t \in (0, 1)$ and Theorem 1.3 applies with $\Gamma = (0, 1)$.

When $\ell(\bar{\Gamma}) = 0$ the limiting distribution in Theorems 1.2 and 1.3 is degenerated at 0. In Section 4 we obtain non-degenerated limiting distributions, with different rates, when the set of contact points between $F^{-1}$ and $G^{-1}$ is finite. The key is the local asymptotic behaviour of $F_n^{-1} - G_m^{-1}$ around these influential points. To avoid unnecessary smoothness assumptions here, we must consider contact points between nondecreasing functions in a generalized sense, including virtual contact points: those corresponding to contacts between the vertical segments joining lateral limits at discontinuity points. Since quantile functions are left continuous, the following definition includes all these contact points.

**Definition 1.4.** We say that $t \in (0, 1)$ is a (generalized) contact point between $F^{-1}$ and $G^{-1}$ if either (i) $F^{-1}(t) = G^{-1}(t)$ or (ii) $F^{-1}(t) < G^{-1}(t) \leq F^{-1}(t+)$ or (iii) $G^{-1}(t) < F^{-1}(t) \leq G^{-1}(t+)$. The set of contact points will be denoted by $\Gamma^*$.

In Section 2 we analyze the points in $\Gamma^*$ in detail. We show in particular (Proposition 2.2) that these generalized contact points are exactly the generalized contact points between the identity and the transforms $F \circ G^{-1}$ or $G \circ F^{-1}$. Notice that these non-decreasing functions can present also left-jump discontinuities which apparently would lead to additional virtual contact points between them and the identity. However, as we show in Proposition 2.5, the contact points in the strict sense and those corresponding to right-jump discontinuities of $F_G$ and $G_F$ cover the interesting cases.

In some cases, and when it makes sense, quantile functions and the considered transforms can be extended by continuity to 0 and 1 and this, in turn, allows us to consider 0 or 1 as contact points (only in the strict sense). Hereafter we will distinguish extremal contact points (0 or 1 when they are contact points) and inner contact points (any other contact point).

If $\Gamma^*$ is finite, Corollary 4.2 in Section 4 shows that the analysis of the asymptotic behaviour of $\gamma(F_n, G_m)$ boils down to the analysis of the localized measure of the set where the first empirical quantile function exceeds the other, namely, of

$$
\ell_{n,m}^0 := \ell\left(\{F_n^{-1} > G_m^{-1}\} \cap (t_0 - \eta, t_0 + \eta)\right) - \ell\left(\{F^{-1} > G^{-1}\} \cap (t_0 - \eta, t_0 + \eta)\right),
$$

for $t_0 \in \Gamma^*$, assuming that $\eta > 0$ is small enough to ensure that $t_0$ is the only point in $\Gamma^* \cap (0 - \eta, 0 + \eta)$ (as we will see, asymptotically, $\ell_{n,m}^0$ does not depend on $\eta$, hence, we do not include it in our notation). We relate this behaviour to the character, position and intensity of the contact, in the following sense. For $t_0$ such that $F_G(t_0) = t_0$, set $H := \{h : t_0 + h \in [0, 1]\}$ and consider the function $\Delta : H \to \mathbb{R}$:

$$
\Delta(h) := F_G(t_0 + h) - t_0 - h.
$$

We will assume that $\Delta$ is locally Lipschitz at 0 (equivalently, that $F_G$ is locally Lipschitz at $t_0$) plus a higher order expansion, possibly on the positive and the negative sides (for extremal contact points only one expansion makes sense). More precisely, we will assume additionally to the Lipschitz property that there exist $\eta > 0$, $r_L = r_L(t_0), r_R = r_R(t_0) \geq 1$ and $C_L = C_L(t_0) \neq 0, C_R = C_R(t_0) \neq 0$ such that

$$
\Delta(h) = \begin{cases} 
C_L|h|^{r_L} + o(|h|^{r_L}), & \text{if } h \in (-\eta, 0), \\
C_R|h|^{r_R} + o(|h|^{r_R}), & \text{if } h \in (0, \eta). 
\end{cases}
$$

In these cases, we will say that $r_L$ (resp. $r_R$) is the intensity or order of the contact on the left (resp. on the right) between $F^{-1}$ and $G^{-1}$ at $t_0$. We observe that the assumptions imply that, for small enough
\[ \eta, \text{sgn}(F_G(t) - t) = \text{sgn}(C_L) \text{ on } (t_0 - \eta, t_0) \text{ and } \text{sgn}(F_G(t) - t) = \text{sgn}(C_R) \text{ on } (t_0, t_0 + \eta). \] A point \( t_0 \) satisfying these conditions will be called a regular contact point.

For integer \( r_L \) and \( r_R \), expression (8) is a kind of left- and right- Taylor expansion. However, \( r_L \) and \( r_R \) are not necessarily integer in the definition above. We can classify regular contact points as crossing points (the case \( \text{sgn}(C_L) \neq \text{sgn}(C_R) \)) or tangency points (if \( \text{sgn}(C_L) = \text{sgn}(C_R) \)). Notice also that under a proper Taylor expansion, \( t_0 \) is a crossing or tangency point depending only on whether \( r_L = r_R \) is odd or even, while decomposition (8) allows to have a crossing point with odd \( r_L = r_R \) or a tangency point with even \( r_L = r_R \).

We must stress that (8) does not necessarily imply smoothness conditions on \( F^{-1} \) or \( G^{-1} \). As an example, consider the case when \( F^{-1}(t_0) < F^{-1}(t_0) \leq G^{-1}(t_0) < F^{-1}(t_0) \). Then \( t_0 \) is a discontinuity point of \( F^{-1} \) (maybe also of \( G^{-1} \)), but \( F_G(t) = t_0 \) for \( t \) close enough to \( t_0 \) and (8) holds with \( r_L = r_R = 1 \), \( C_R = -C_L = -1 \). We should also note that, while (8) excludes discontinuity points for \( F_G \), in particular, virtual contact points between \( F_G \) and the identity, our approach allows to handle these points in a rather straightforward way (see (43), (44) and Theorem 4.10). Finally, we note that while (8) requires the contact orders to be at least 1, lower orders can also be considered. If, for instance, \( \Delta(h) = \text{sgn}(h) |h|^r \), with \( 0 < r < 1 \), then \( \Delta \) is not Lipschitz around 0, but \( G_F \) is and, under some additional assumptions, the local behaviour can be studied through \( \rho_{n,m}^{t_0} \), the version of \( \rho_{n,m} \) in which the roles of the \( X \) and \( Y \) samples are exchanged (see the comments before the proof of Theorem 1.5).

For a compact description of the limit distribution for the terms \( \ell_{n,m}^{t_0} \) we consider independent random elements \( B_1, B_2, W_0, W_1, \{ \xi_1, n \}_{n \geq 1}, \{ \xi_2, n \}_{n \geq 1}, \{ \xi_3, n \}_{n \geq 1}, \{ \xi_4, n \}_{n \geq 1} \), where \( B_1 \) are Brownian bridges on \([0, 1]\), \( W_i \) are Brownian motions on \([0, \infty)\) and \( \{ \xi_i, n \}_{n \geq 1} \) sequences of i.i.d. exponential r.v.’s with unit mean. We set \( S_i := \xi_{i,1} + \ldots + \xi_{i,k}, k \geq 1, i = 1, \ldots, 4 \). We fix \( \lambda \in (0, 1) \) and set

\[ B_\lambda := \frac{1}{\sqrt{1-\lambda}} B_1 - \frac{1}{\sqrt{1-\lambda}} B_2. \]

We consider \( r_L, r_R \geq 1 \) and denote \( r_0 := \max(r_L, r_R) \). Also, for real numbers \( a, b \), we will use the notation \( a^{+}(b) \) for \( a^+ \) (the positive part of \( a \)) either \( a^- \) (the negative part) depending on whether \( b > 0 \) or \( b < 0 \). For \( t_0 \in (0, 1) \) we define

\[ T_{r_L,r_R}(t_0; C_L, C_R) := \text{sgn}(C_L) \left( \frac{(B_\lambda(t_0))^+}{C_L} \right)^{1/r_0} I(r_L = r_0) + \text{sgn}(C_R) \left( \frac{(B_\lambda(t_0))^+}{C_R} \right)^{1/r_0} I(r_R = r_0), \tag{9} \]

when \( r_0 > 1 \) or \( r_0 = 1 \) and \( C_R C_L > 0 \), while

\[ T_{1,1}(t_0; C_L, C_R) := \frac{(B_\lambda(t_0))^+}{C_L} + \frac{(B_\lambda(t_0))^+}{C_R} + \text{sgn}(C_L) B_2(t_0) \frac{B_2(t_0)}{\sqrt{1-\lambda}}, \tag{10} \]

when \( C_L C_R < 0 \). Additionally, for \( r_0 > 1 \) and \( t_0 = 0, 1 \) we define

\[ T_{r_0,t_0}(t_0; C, C) := \text{sgn}(C) \ell \{ y \in (0, \infty) : \text{sgn}(C) W_{t_0}(y) > (\lambda(1-\lambda))^1/2 |C|y^r \}, \tag{11} \]

while in the case \( r_0 = 1 \) (although with obvious redundancies to keep a simple notation) we set

\[ T_{1,1}(0; C, C) := \text{sgn}(C) \lambda(1-\lambda) \int_0^\infty I\{ \text{sgn}(C) \lambda S^2_{(1-\lambda)y} > \text{sgn}(C)(1-\lambda)(1+C)S^2_{(1-\lambda)y} \} \, dy \]

\[ T_{1,1}(1; C, C) := \text{sgn}(C) \lambda(1-\lambda) \int_0^\infty I\{ \text{sgn}(C) \lambda S^2_{(1-\lambda)y} > \text{sgn}(C)(1-\lambda)(1+C)S^2_{(1-\lambda)y} \} \, dy. \tag{12} \]

We are ready to present the results describing the asymptotic behaviour of \( \ell_{n,m}^{t_0} \) for regular contact points. Theorem 1.5 deals with inner contact points, while extremal contact points are considered in

\[ \ldots \]
Theorem 1.6 (in fact, Theorem 1.5 remains valid for extremal contact points, but the limit distribution is Dirac’s measure on 0 in that case).

Theorem 1.5. Assume $t_0$ is a regular inner contact point with contact orders $r_L = r_L(t_0), r_R = r_L(t_0) \geq 1$ and constants $C_L = C_L(t_0), C_R = C_R(t_0)$. If $r = \max(r_L, r_R)$ and $n, m \to \infty$ with $\frac{n}{n+m} \to \lambda \in (0,1)$, then, for every small enough $\eta > 0$,

$$(n + m)^{\frac{1}{r}} t_{n,m}^{\frac{1}{r}} \frac{u}{m} \to T_{rL,rR}(t_0; C_L, C_R).$$

Theorem 1.6. Assume $t_0 \in \{0, 1\}$ is regular with contact order $r \geq 1$ and constant $C$. If $m, n \to \infty$ with $\frac{n}{n+m} \to \lambda \in (0,1)$, then, for every small enough $\eta > 0$,

$$(n + m)^{\frac{1}{r}} t_{n,m}^{\frac{1}{r}} \frac{u}{m} \to T_{r}(t_0; C,C).$$

We see from Theorems 1.5 and 1.6 that, with the same contact intensities, $t_{n,m}^{\frac{1}{r}}$ vanishes faster for extremal contact points. In Subsection 4.1 we provide examples of extremal contact points for which $t_{n,m}^{\frac{1}{r}}$ converges at rate $(n + m)^{-c}$ for every $c \in (0,1]$ and of inner contact points for which the rate is $(n + m)^{-c}$, $c \in (0, \frac{1}{2}]$. Another distinctive feature of the limiting distributions for inner contact points is that for crossing points (those with $C_L(t_0)C_R(t_0) < 0$) the limiting distribution takes positive and negative values with positive probabilities. If $t_0$ is a tangency point ($C_L(t_0)C_R(t_0) > 0$) then the limiting distribution is concentrated on $(0, \infty)$ or on $(-\infty, 0)$.

The local asymptotic results in Theorems 1.5 and 1.6 can be strengthened to produce the following distributional limit theorem for $\gamma(F_n, G_m)$.

Theorem 1.7. Assume $\Gamma^* = \{t_1, \ldots, t_k\}$ where $t_i$ is a regular contact point with intensities $r_L(t_i), r_R(t_i)$ and constants $C_L(t_i), C_R(t_i), i = 1, \ldots, k$. Set $r_i = \max(r_L(t_i), r_R(t_i))$ if $t_i \in (0,1)$, and $r_i = \max(r_L(t_i), r_R(t_i)) - \frac{1}{2}$ if $t_i \in \{0,1\}$. Then, if $r_0 = \max_{1 \leq i \leq k} r_i$,

$$(n + m)^{\frac{1}{r_0}} (\gamma(F_n, G_m) - \gamma(F, G)) \xrightarrow{D} \sum_{i=1}^{k} I(r_i = r_0) T_{rL(t_i), rR(t_i)}(t; C_L(t_i), C_R(t_i)).$$

Theorem 1.7 shows that the rate of convergence of $\gamma(F_n, G_m)$ is determined by the maximal intensity of contact, and that only points with maximal intensity contribute to the limiting distribution, with adjustments to take into account the different role of inner and extremal contact points. If there are extremal contact points then the rate of convergence can be $(n + m)^c$ for any $c \in (0,1]$. When there are only inner contact points the rate is $(n + m)^c$ with $c \in (0, \frac{1}{2}]$. The only case in which $\gamma(F_n, G_m)$ is asymptotically normal is when the inner contact points have intensity one, all of them with constants $C_L = -C_R$ and there is no extremal contact point or its influence vanishes faster.

1.2. Organization of the paper

The remaining sections of this work are organized as follows. Section 2 includes some key results on quantile functions and analyzes the structure of the contact sets. We will explicitly formulate several results on quantile functions. Some are classical, but, in fact it is not an easy task to find a comprehensive reference on quantile functions, with the notable exception of Appendix A in [8] on ‘Inverse
Distribution Functions’. We observe that [8] is devoted to the analysis of convergence rates of Kantorovich transport distances between probability measures on the real line, which can be expressed in terms of quantile functions as \( \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt \), thus our problem corresponds to the limiting case \( p = 0 \). Remarkably, this problem also encompasses a wide range of convergence rates.

In Section 3 we provide the proofs of Theorems 1.1, 1.2 and 1.3. Most of the limit theorems that we give for Galton’s rank statistic are based on convenient representations of empirical quantile functions, combined with some type of strong approximation. Using representation (14) below, we can derive limit theorems for Galton’s rank statistic relying on strong approximations for uniform quantile processes, rather than using strong approximations for general quantile processes (as, for instance, in Chapter 6 in [11]). This results in a significant gain in generality, since approximations for general quantile processes typically require strong smoothness assumptions (existence of densities plus additional conditions on them) that we can circumvent with this approach.

Section 4 gives the proofs of Theorems 1.5, 1.6 and 1.7. The key ingredients for this will be, as in Section 3, a convenient representation of the quantile processes and some application of strong approximations. With some simple localization results (Lemma 4.1 and Corollary 4.2) we see that the asymptotic behaviour of \( \gamma(F_n, G_n) \) can be studied through that of the localized terms \( \ell_{m,n}^{\omega_0} \) with \( \omega_0 \) in the contact set. Some results on the asymptotic independence between lower, central and upper order statistics allow then to complete the proof of Theorem 1.7. Subsection 4.1 in that section provides some examples of contact points with different positions and contact intensities. This subsection also includes a simplified version of Theorem 1.7 under conditions that guarantee that \( F_G \) is smooth (see Theorem 4.9); and a further limit theorem (Theorem 4.10) for the case when \( F \) and \( G \) have finite supports. This is an interesting example which can be handled with our approach even though the contact points here are not regular contact points.

A separate Appendix, [4], presents additional material on the \( F_G \) transform and the strong approximation result that we will use through the paper.

1.3. Notation

We end this Introduction with some words on notation. Through the paper \( \mathcal{L}(X) \) will denote the law of the random vector or r.v. \( X \). We will consider a generic probability space \((\Omega, \sigma, P)\), where the involved random objects are defined. Given the (measurable) sets \( A, B \), by \( IA \) we will denote the indicator function of \( A \) and \( A \setminus B \) will denote the set \( \{ x \in A : x \notin B \} \). As before, \( t \) will denote the Lebesgue measure on the unit interval \((0,1)\). Convergence (almost sure, in probability, weak or in law) will be denoted by \( a_p \), \( p \rightarrow \), and \( w \rightarrow \), respectively. Given \( x \in \mathbb{R} \), we will use \( \lfloor x \rfloor \) to denote the smaller integer greater or equal than \( x \), and \( x^+ := \sup\{ x, 0 \} \) and \( x^- = -\inf\{ x, 0 \} \). Also we use the notation \( f(x-) := \lim_{y \rightarrow x-} f(y) \) and \( f(x+) := \lim_{y \rightarrow x+} f(y) \) for the lateral limits of a real function, \( f \), whenever these limits exist, and \( \text{sgn}(x) \) (defined as 0 if \( x = 0 \) and \( x/|x| \) otherwise). Also recall that for real numbers \( a, b, a^{w(b)} \) will denote either \( a^+ \) or \( a^- \) depending on whether \( b > 0 \) or \( b < 0 \).

Throughout, \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_m \) will be independent samples of i.i.d. r.v.’s such that \( \mathcal{L}(X_i) \) and \( \mathcal{L}(Y_i) \) have respective d.f.’s \( F \) and \( G \). As above, \( F_n \) and \( G_m \) will denote the respective empirical d.f.’s based on the \( X_i \)'s and \( Y_j \)'s samples. Occasionally, we will use the superscript \( \omega \) in functions computed from the sample values \( X_i(\omega), i = 1, \ldots, n \) or \( Y_j(\omega), j = 1, \ldots, m \), (for instance, the empirical d.f. \( F_n^\omega \) or the empirical quantile function \( (F_n^\omega)^{-1} \)). Without loss of generality we can (and often do) assume that the samples have been obtained from independent \( U(0, 1) \) samples \( U_1, \ldots, U_n \) and \( V_1, \ldots, V_m \) through the transformations \( X_i = F^{-1}(U_i), Y_j = G^{-1}(V_j) \). From now on, we will denote the empirical quantile functions of these uniform samples by \( U_n \) and \( \mathcal{V}_m \). We have the obvious
relations $F_n^{-1} = F^{-1}(U_n), G_m^{-1} = G^{-1}(V_m)$. Writing $u_n$ and $v_m$ for the quantile processes based on the $U_i$'s and the $V_j$'s, respectively, ($u_n(t) = \sqrt{n}(U_n(t) - t)$ and similarly for $v_m$) we note that

$$F_n^{-1}(t) = F^{-1}\left(t + \frac{u_n(t)}{\sqrt{n}}\right) \quad \text{and} \quad G_m^{-1}(t) = G^{-1}\left(t + \frac{v_m(t)}{\sqrt{m}}\right).$$

(14)

We recall the main role in the paper of $\tilde{\Gamma}$, the contact set between the identity and the function

$$F_G(t) := F(G^{-1}(t)).$$

(15)

We also note that, while the role of $F$ and $G$ is symmetric in the definitions of $\Gamma$ and $\Gamma^*$, this is not true in the case of $\tilde{\Gamma}$. For a more clear description of the relations among these sets we sometimes write $\tilde{\Gamma}_F = \Gamma$, $\tilde{\Gamma}_G = \{t \in (0,1) : G(t) = t\}$ and $\tilde{\Gamma}^*_F$ (resp. $\tilde{\Gamma}^*_G$) for the set of generalized contact points between $F_G$ (resp. $G_F$) and the identity (see (19)). Obviously, $\tilde{\Gamma}_F \subset \tilde{\Gamma}^*_F$.

2. Quantile transforms and contact sets

Quantile functions defined as in (1) provide a useful description of probabilities on the real line in terms of nondecreasing, left-continuous functions on $(0,1)$. In fact, every nondecreasing left-continuous real function, $H$, defined on $(0,1)$ is the unique quantile function associated to just a unique d.f.: as a dual relation to (1), such a function $H$ is the quantile function associated to the d.f.

$$F(x) = \sup\{t \in (0,1) : H(t) \leq x\}.\quad (16)$$

As already noted, it will be convenient at some points to extend $F^{-1}$ to 0 and 1 in the obvious way: $F^{-1}(0) := F^{-1}(0^+)$ and $F^{-1}(1) := F^{-1}(1^-)$.

In this section we present some relevant facts on the relation between quantile functions and the composite functions $F_G$ defined in (15) without any smoothness assumption on the d.f.'s. We must begin by stressing the fact that, in general, we cannot guarantee even lateral continuity of $F_G$ (that would we only guaranteed for $F(G^{-1}(t-))$ on the left and for $F(G^{-1}(t+))$ on the right). Also, the well known fact that, for $t \in (0,1)$: $t \leq F(x) \iff F^{-1}(t) \leq x$, leads to the relations

$$F_G(t) = \max\{s \in [0,1] : F^{-1}(s) \leq G^{-1}(t)\}, \quad \text{for } t \in (0,1)\quad (17)$$

$$F^{-1}(t) > G^{-1}(s) \iff t > F_G(s) \text{ for } t, s \in (0,1).\quad (18)$$

We note that $F_G(t)$ and $G_F(t)$ could be different even when $t \in \Gamma$. This possibility is naturally related to the behaviour of the composition $F(F^{-1}(t))$. Clearly, $F(F^{-1}(t)-) \leq t \leq F(t)$, thus $F_F(t) = t$ when $F$ is continuous at $F^{-1}(t)$, but this could fail otherwise. More precisely, for $t \in (0,1)$, we have that $F_F(t) = t$ is equivalent to $t \in \text{Im}(F)$, where $\text{Im}(F) := \{F(x), x \in \mathbb{R}\}$ (see Lemma A.3 in [8]). Now let $t \in (0,1) \cap \Gamma$: if $t \in \text{Im}(F)$, then $t = F_F(t) = F_G(t)$, while if $t \notin \text{Im}(F)$, then $t < F_F(t) = F_G(t)$. We collect these facts and some easy consequences in the following lemma.

Lemma 2.1. Let $F, G$ be arbitrary d.f.'s. For $t \in (0,1)$, with the above notation, we have:

a) If $t \in \Gamma \cap \text{Im}(F)$, then $F_G(t) = t$.

b) If $t \in \Gamma \setminus \text{Im}(F)$, then $F_G(t) > t$.

c) If $F_G(t) = t$, then either $t \in \Gamma$ or $F^{-1}(t) < G^{-1}(t)$.

d) If $F_G(t) = t$ and $G_F(t) = t$, then $t \in \Gamma$. 

The conclusions in Lemma 2.1 can be rewritten with the notation of (4) and (5). Item a), for instance, becomes $\Gamma \cap \text{Im}(F) \subset \hat{\Gamma}$. Generalized contact points in the sense of Definition 1.4 (that is, points in $\Gamma^*$) can also be characterized in terms of the composite functions $F_G$ and $G_F$.

As already noted, the consideration of virtual contact points associated to left-jump discontinuities of $F_G$ would be superfluous: they are in fact contact points in the strict sense or they are associated to right-jump discontinuities of $G_F$ (see Proposition 2.5). In consequence, we consider a point $t_0 \in (0, 1)$ as a contact point of $F_G$ and the identity whenever

$$F_G(t_0) = t_0, \text{ or } F_G(t_0) < t_0 \leq F_G(t_0+) \quad (19)$$

Note that the second condition is equivalent to $F_G(t_0) < t_0 \leq F_G(s)$ for all $s > t_0$, hence also to $G^{-1}(t_0) < F^{-1}(t_0) \leq G^{-1}(t_0+)$, which is condition (iii) in Definition 1.4. Therefore, we have that

**Proposition 2.2.** The virtual contact points of $F^{-1}$ and $G^{-1}$ are exactly the virtual contact points of $F_G$ or $G_F$ with the identity.

Since $\hat{\Gamma}^*_F \setminus \hat{\Gamma}_F$ is contained in the (at most countable) set of discontinuity points of the nondecreasing function $F_G$, we have $\ell(\hat{\Gamma}^*_F \setminus \hat{\Gamma}_F) = 0$ and $\ell(\hat{\Gamma}_F^*) = \ell(\hat{\Gamma}_F)$. Proposition 2.2 means that $\Gamma^* \setminus \Gamma = (\hat{\Gamma}_F^* \setminus \hat{\Gamma}_F) \cup (\hat{\Gamma}_G^* \setminus \hat{\Gamma}_G)$. We explore next the situation for contact points in the strict sense.

**Proposition 2.3.** If $t_0 \in \Gamma \cap (0, 1)$ then $F_G(t_0) = t_0$, or $G_F(t_0) = t_0$ (that is $t_0 \in \hat{\Gamma}_F \cup \hat{\Gamma}_G$), or the set $\{ t \in (0, 1) : F^{-1}(t) = G^{-1}(t) = F^{-1}(t_0) \}$ is a non-degenerate interval (hence, in the latter case, the point $x_0 = F^{-1}(t_0)$ is a common discontinuity point of $F$ and $G$ and $t_0$ cannot be an isolated element of $\Gamma^*$).

**Proof.** It is easy to see that if $t_0 \in (0, 1) \setminus (\text{Im}(F) \cup \text{Im}(G))$ satisfies $F^{-1}(t_0) = G^{-1}(t_0)$, then the point $x_0 = F^{-1}(t_0)$ would have positive mass under both distributions, hence the set $\{ t \in (0, 1) : F^{-1}(t) = G^{-1}(t) = F^{-1}(t_0) \}$ is a non-degenerate interval. Any other point in $\Gamma \cap (0, 1)$ must belong to $\text{Im}(F) \cup \text{Im}(G)$ and, by Lemma 2.1, must satisfy either $F_G(t_0) = t_0$ or $G_F(t_0) = t_0$. \[\square\]

**Proposition 2.4.** Let $t_0 \in (0, 1)$ be such that $t_0 \in \hat{\Gamma}_F \cup \hat{\Gamma}_G$, that is $F_G(t_0) = t_0$ or $G_F(t_0) = t_0$. Then $t_0 \in \Gamma^*$ (to is a contact point between $F^{-1}$ and $G^{-1}$).

**Proof.** For any $t_0 \in (0, 1)$ such that $F_G(t_0) = t_0$ (the case $G_F(t_0) = t_0$ is identical), we must have one of the following exclusive possibilities:

i) $G_F(t_0) < t_0$, and then $F^{-1}(t_0) < G^{-1}(t_0)$, and

ii) $F_G(t_0) = t_0 = G_F(t_0)$, or $F_G(t_0) = t_0 < G_F(t_0)$, which lead to $F^{-1}(t_0) = G^{-1}(t_0)$.

If i) holds, then we would have $G^{-1}(t_0) \leq F^{-1}(t_0)$ (this follows easily from the fact that the strict inequality $G^{-1}(t_0) > F^{-1}(t_0)$ would imply $F_G(t_0) = F(G^{-1}(t_0)) > t_0$). Hence, i) implies $F^{-1}(t_0) < G^{-1}(t_0) \leq F^{-1}(t_0+)$. \[\square\]

The next proposition allows to obviate the analysis of contact points associated to left-discontinuities.

**Proposition 2.5.** Let $t_0 \in (0, 1)$. If $F_G(t_0- \leq t_0 < F_G(t_0)$, then $G_F(t_0) \leq t_0 \leq G_F(t_0+)$, or the point $x_0 = G^{-1}(t_0)$ is a common discontinuity point of $F$ and $G$ and $t_0$ cannot be an isolated element of $\Gamma^*$. 


Proof. If we suppose that \( G^{-1}(t) = G^{-1}(t_0) \) for some \( t < t_0 \), then for every sequence \( \{t_n\} \) such that \( t_n \to t_0 \), \( F(G^{-1}(t_n)) = F(G^{-1}(t_0)) \) will hold eventually, thus leading to the absurd \( F_G(t_0) = F_G(t_0) \). Therefore it must be \( x_0 := G^{-1}(t_0) > G^{-1}(t) \) for every \( t < t_0 \), and \( F_G(t_0) = F(G^{-1}(t_0)) \). Moreover, the discontinuity of \( F \) and its link with \( F^{-1} \) easily show that \( F^{-1}(t) = x_0 \) if \( t \in (F(x_0) - F(x_0)) \). Now, on the first hand, from the hypothesis we obtain \( F^{-1}(t_0) \leq x_0 \) and \( F^{-1}(s) = x_0 \) for every \( s \in (t_0, F(x_0)) \), hence
\[
G_F(t_0+) = G(x_0) = G(G^{-1}(t_0)) \geq t_0. \tag{20}
\]
On the other hand, from the relation \( t_0 < F_G(t_0) \) we obtain \( F^{-1}(t_0) \leq G^{-1}(t_0) \), hence \( F^{-1}(t_0) = G^{-1}(t_0) \) or, alternatively, \( F^{-1}(t_0) < G^{-1}(t_0) \) which gives \( G_F(t_0) < t_0 \). This relation and (20) imply that \( G_F(t_0) < t_0 \leq G_F(t_0+) \). Finally, if \( F^{-1}(t_0) = G^{-1}(t_0) \) and \( x_0 = F^{-1}(t_0) \) is a continuity point of \( G \), from (20) we obtain that \( G_F(t_0) = t_0 \) what proves the result. \( \Box \)

We conclude this section with some easy consequences of the last results.

Corollary 2.6. Let \( t_0 \in (0,1) \) such that \( F_G(t_0) = t_0 \) (resp. \( F_G(t_0) = t_0 \)) then \( t_0 \) is a contact point (possibly virtual) between \( G_F \) (resp. \( F_G \)) and the identity.

Corollary 2.6 states that \( \hat{\Gamma}_F \subset \hat{\Gamma}_G^* \). From the comments after Proposition 2.2 we see that \( \ell(\hat{\Gamma}_F) \leq \ell(\hat{\Gamma}_G) \). The same argument shows that \( \ell(\hat{\Gamma}_G) \leq \ell(\hat{\Gamma}_F) \), hence \( \ell(\hat{\Gamma}_F) = \ell(\hat{\Gamma}_G) \). This guarantees the symmetric roles of \( F \) and \( G \) in the condition \( \ell(\hat{\Gamma}) = 0 \) in Theorems 1.1, 1.2 and 1.3.

Proposition 2.7. If \( \Gamma^* \) is finite then \( \Gamma^* = \hat{\Gamma}_F^* \cup \hat{\Gamma}_G^* \). In particular, \( \hat{\Gamma}_F^* \) and \( \hat{\Gamma}_G^* \) are finite.

We remark that, while \( \hat{\Gamma}_F^* \cup \hat{\Gamma}_G^* \subset \Gamma^* \) always holds (this follows from Proposition 2.4), the set \( \Gamma^* \) can be much bigger that \( \hat{\Gamma}_F^* \cup \hat{\Gamma}_G^* \) (recall the comments in the Introduction; the case \( G = F \), with \( F \) the d.f. of the Bernoulli law with mean \( p \) gives a simple example of this).

In Section 4 we prove distributional limit theorems for \( \gamma(F_n,G_m) \) under the assumption that \( \Gamma^* \) is finite, say, \( \Gamma^* = \{t_1 < \cdots < t_r\} \). The differences \( F^{-1}(t) - G^{-1}(t) \) must have constant sign in the open intervals \( (t_i,t_{i+1}) \) (the same happens in \( (0,t_1) \) or \( (t_r,1) \) if 0 or 1 are not contact points). The next result will enable us to focus on neighbourhoods of isolated contact points to study \( \gamma(F_n,G_m) \).

Lemma 2.8. Assume \( 0 < a \leq b < 1 \) are such that \( [a,b] \cap \Gamma^* = \emptyset \), and also that \( \text{sgn}(F^{-1}(t) - G^{-1}(t)) > 0 \) (resp. \( \text{sgn}(F^{-1}(t) - G^{-1}(t)) < 0 \)) for every \( t \in [a,b] \). Then there exists \( \delta > 0 \) such that \( F^{-1}(t) - G^{-1}(t + \delta) < -\delta \) (resp. \( F^{-1}(t) - G^{-1}(t + \delta) > \delta \)) for every \( t \in [a,b] \).

Proof. Let us consider the case \( \text{sgn}(F^{-1}(t) - G^{-1}(t)) > 0 \). Assume, on the contrary, that there exist a sequence \( \{t_k\} \subset [a,b] \) such that \( F^{-1}(t_k) - G^{-1}(t_k) \to 0 \). Then we can choose \( \{t_k^+\} \subset (0,1) \) such that, \( t_k < t_k^+ < t_k - t_k^+ \to 0 \), \( F^{-1}(t_k) - G^{-1}(t_k)^+ \to 0 \) and \( G^{-1}(t_k) - G^{-1}(t_k^+) \to 0 \). Since \( [a,b] \) is compact, we can assume that \( \{t_k\} \), thus also \( \{t_k^+\} \), converges. We write \( t_0 \in [a,b] \) for the common limit. By taking subsequences, if necessary, we can also assume that both sequences are monotone.

Now, we only need to consider four possible cases. If, for instance, \( \{t_k\} \) and \( \{t_k^+\} \) are increasing, then we would obtain that \( F^{-1}(t_0) = G^{-1}(t_0) \) which is impossible by assumption. If the sequence \( \{t_k\} \) is increasing and \( \{t_k^+\} \) is decreasing, then \( F^{-1}(t_k) \to F^{-1}(t_0) \) and \( G^{-1}(t_k) \to G^{-1}(t_0^+) \), and we would have that \( F^{-1}(t_0) = G^{-1}(t_0) \) and, consequently, \( t_0 \) would be a contact point what is not possible either, because \( [a,b] \cap \Gamma^* = \emptyset \). The two remaining cases lead to similar contradictions. \( \Box \)
We conclude this section with two observations that will be exploited in later sections. First, we note that \( \text{sgn}(t - F_G(t)) = \text{sgn}(F^{-1}(t) - G^{-1}(t)) \) for every \( t \notin \Gamma^* \). To check this recall relation (18), giving that \( F^{-1}(t) > G^{-1}(t) \) if and only if \( t > F_G(t) \). This also implies that \( F^{-1}(t) \leq G^{-1}(t) \) if and only if \( t \leq F_G(t) \), but \( F_G(t) \) cannot happen if \( t \notin \Gamma^* \) (Proposition 2.4). On the other hand, if \( t < F_G(t) \) then \( F^{-1}(t) \leq G^{-1}(t) \) but, again, \( F^{-1}(t) = G^{-1}(t) \) is not possible if \( x \notin \Gamma^* \). This means that \( \text{sgn}(t - F_G(t)) \) is constant in the intervals \( (t_i, t_{i+1}) \) as above.

Our second observation arises from the fact that every nondecreasing left-continuous real function, \( H \), defined on \((0, 1)\) is the quantile function associated to the d.f. given by (16). We can apply Lemma 2.8 to the quantile function \( H(t) = F_G(t-) \) and the identity and conclude, for instance, that in a compact interval where \( t - F_G(t) > 0 \) there exists some \( \delta > 0 \) such that \( t - F_G(t) \geq \delta \).

3. Consistency of Galton’s rank order statistic

In this section we provide proofs of Theorems 1.1, 1.2 and 1.3. These results show that the statistic is consistent if and only if the contact set \( \Gamma = \{ t : t = F_G(t) \} \) has zero Lebesgue measure. The key to the proof of Theorem 1.1 is the following lemma. Here we denote \( \Gamma_0 := \text{Im}(F) \cap \Gamma \cap (0, 1) \).

**Lemma 3.1.** Let \( F, G \) be arbitrary d.f.’s. With the notation above, we have:

\[
\gamma(F_n, G_m) - \gamma(F, G) - \ell(\{ F_n^{-1} > G_m^{-1} \} \cap \Gamma) \xrightarrow{a.s.} 0 \quad \text{as } n, m \to \infty, \tag{21}
\]

\[
\gamma(F_n, G_m) - \gamma(F, G) - \ell(\{ F_n^{-1} > G_m^{-1} \} \cap \Gamma_0) \xrightarrow{a.s.} 0 \quad \text{as } n, m \to \infty, \tag{22}
\]

\[
\gamma(F_n, G_m) - \gamma(F, G) - \ell(\{ F_n^{-1} > G_m^{-1} \} \cap \hat{\Gamma}) \xrightarrow{a.s.} 0 \quad \text{as } n, m \to \infty. \tag{23}
\]

**Proof.** By right continuity, if \( t \notin \text{Im}(F) \), then there exists \( \delta_t > 0 \) such that \( \{ t, t + \delta_t \} \cap \text{Im}(F) = \emptyset \) and \( F_F(s) = t + \delta_t \), for every \( s \in [t, t + \delta_t] \). From this, it easy to see that there exists an at most countable family of disjoint intervals \( I_k = [a_k, b_k] \), with \( a_k < b_k \) which is a partition of the complement of \( \text{Im}(F) \) and \( F_F(s) = b_k \), for every \( s \in [a_k, b_k] \).

The Glivenko-Cantelli Theorem gives that for some \( \Omega_0 \in \sigma \), with \( P(\Omega_0) = 1 \), if \( \omega \in \Omega_0 \), then

\[
\sup_t |F_n^\omega(t) - F(t)| \to 0 \quad \text{and} \quad \sup_t |G_m^\omega(t) - G(t)| \to 0.
\]

Now, using the elementary Skorohod theorem (see e.g. Lemma A.5 in \([8]\)), for every \( \omega \in \Omega_0 \), the set

\[
T^\omega := \left\{ t \in (0, 1) : (F_n^\omega)^{-1}(t) \to F^{-1}(t) \text{ and } (G_m^\omega)^{-1}(t) \to G^{-1}(t) \right\} \setminus \{a_1, a_2, \ldots\}
\]

has Lebesgue measure one. Therefore, if \( \omega \in \Omega_0 \),

\[
\gamma(F_n^\omega, G_m^\omega) - \gamma(F, G) - \ell((F_n^\omega)^{-1} > (G_m^\omega)^{-1}) \cap \Gamma
\]

\[
= \ell((F_n^\omega)^{-1} > (G_m^\omega)^{-1}, F_n^\omega < G_m^\omega) \cap T^\omega
\]

\[
- \ell((F_n^\omega)^{-1} > (G_m^\omega)^{-1}, F_n^\omega > G_m^\omega) \cap T^\omega
\]

which converges to 0 because both sets within brackets converge to the empty set. This proves (21). To prove (22) we show that if \( \omega \in \Omega_0 \), then

\[
d_n := \ell(\left\{ (F_n^\omega)^{-1} > (G_m^\omega)^{-1} \right\} \cap \Gamma) - \ell(\left\{ (F_n^\omega)^{-1} > (G_m^\omega)^{-1} \right\} \cap \Gamma_0) \to 0. \tag{24}
\]
To check this, notice that
\[ d_n = \ell \left( \left\{ (F_n^\omega)^{-1} > (G_m^\omega)^{-1} \right\} \cap T^\omega \cap \Gamma \cap \left( \cup_k (a_k, b_k) \right) \right) \]
\[ = \ell \left( \left\{ t > (F_n^\omega)_{G_m}(t) \right\} \cap T^\omega \cap \Gamma \cap \left( \cup_k (a_k, b_k) \right) \right). \]

Now, Glivenko-Cantelli again, and the construction of \( T^\omega \) yield that if \( \omega \in \Omega_0 \) and \( t \in T^\omega \cap \Gamma \), then
\[ 0 = \lim_n |F_n^\omega| (G_m^\omega)^{-1} (t) - F [ (G_m^\omega)^{-1} (t) ] \text{ and } \lim_n (G_m^\omega)^{-1} (t) = G^{-1} (t) = F^{-1} (t). \] (25)

From here, if \( t \in T^\omega \cap \Gamma \cap \left( (a_k, b_k) \right) \) for some \( k \), then, eventually, \( F [(G_m^\omega)^{-1} (t)] = b_k > t \) which, combined with the first statement in (25), is eventually impossible if \( t > (F_n^\omega)_{G_m}(t) \); thus (24) follows.

The proof of (23) is now obvious taking into account that, from Lemma 2.1
\[ \Gamma_0 \subset \hat{\Gamma} \subset \Gamma \cup \{ F^{-1} < G^{-1} \}. \] (26)

\[\square\]

**Proof of Theorem 1.1.** Sufficiency is a trivial consequence of Lemma 3.1. To prove necessity, if \( \gamma (F_n, G_m) \xrightarrow{a.s.} \gamma (F, G) \), as \( n, m \to \infty \), according to Lemma 3.1, we have that
\[ D_n := \ell \left( \left\{ F_n^{-1} > G_m^{-1} \right\} \cap \hat{\Gamma} \right) \xrightarrow{a.s.} 0, \] (27)

From (3) in [4], we have
\[ D_n = \ell \left( \left\{ \cup U_n > F_G (V_m) \right\} \cap \hat{\Gamma} \right) \geq \ell \left( \left\{ t : \cup U_n (t) > t \right\} \cap \{ t : t \geq F_G (V_m (t)) \} \cap \hat{\Gamma} \right). \]

Now Fubini’s theorem and independence between samples yield
\[ E[D_n] \geq \int_{\Gamma} P [\cup U_n (t) > t] P [t \geq F_G (V_m (t))] dt \geq \int_{\Gamma} P [\cup U_n (t) > t] P [t > V_m (t)] dt, \] (28)

where the last inequality follows from the fact that, since \( F_G \) is nondecreasing, \( F_G (t) = t \) and \( V_m (t) < t \) imply that \( F_G (V_m (t)) \leq t \). On the other hand, for every \( t \in (0, 1) \), the factors inside the integral converge to 1/2. But, since \( |D_n| \leq 1 \), (27) implies \( E[D_n] \to 0 \). This and convergence to 1/4 of the last integrand in (28) imply that \( \ell (\hat{\Gamma}) = 0 \). \[\square\]

Next, we give a proof of Theorem 1.2 (recall that it does not assume any smoothness on \( F \) or \( G \)).

**Proof of Theorem 1.2.** Assuming, w.l.o.g., the construction in Theorem A.6 in [4] we have
\[ \hat{\gamma}_n := \gamma (F_n, G) = \ell \left\{ t : F^{-1} \left( t + \frac{u_n (t)}{\sqrt{n}} \right) > G^{-1} (t) \right\} \]
\[ = \ell \left\{ t : t + \frac{u_n (t)}{\sqrt{n}} > F_G (t) \right\} = \ell \left\{ t : u_n (t) > \sqrt{n} (F_G (t) - t) \right\}, \]

and similarly \( \gamma := \gamma (F, G) = \ell \{ t : F_G (t) - t < 0 \} \). Therefore, we see that
\[ \hat{\gamma}_n - \gamma = \ell \left\{ t : u_n (t) > \sqrt{n} (F_G (t) - t) \geq 0 \right\} - \ell \left\{ t : 0 > \sqrt{n} (F_G (t) - t) \geq u_n (t) \right\}. \]
Obviously, for the Brownian bridges $B_n^F(t)$,
\[
\ell \{ t : u_n(t) > \sqrt{n}(F_G(t) - t) > 0 \} \leq \ell \{ t : B_n^F(t) + K \frac{\log n}{\sqrt{n}} \geq \sqrt{n}(F_G(t) - t) > 0 \} + \ell \{ t : |B_n^F(t) - u_n(t)| > K \frac{\log n}{\sqrt{n}} \}.
\]

By Theorem A.6 in [4], the last summand eventually vanishes. For a fixed Brownian bridge $B(t)$ and $t \in (0, 1)$ such that $F_G(t) - t > 0$, we have $B(t) + K \frac{\log n}{\sqrt{n}} < \sqrt{n}(F_G(t) - t)$ eventually. This and the dominated convergence Theorem imply that
\[
\ell \{ t : B(t) + K \frac{\log n}{\sqrt{n}} \geq \sqrt{n}(F_G(t) - t) > 0 \} \xrightarrow{a.s.} 0.
\]

As a result we obtain that
\[
\ell \{ t : u_n(t) > \sqrt{n}(F_G(t) - t) > 0 \} \xrightarrow{P} 0.
\]

Similarly we see that
\[
\ell \{ t : 0 > \sqrt{n}(F_G(t) - t) \geq u_n(t) \} \xrightarrow{P} 0
\]
and conclude that
\[
\gamma_n - \gamma = \ell \{ t : u_n(t) \geq 0, F_G(t) = t \} + o_P(1).
\]

Next, we observe that, eventually,
\[
\ell \{ t \in \tilde{\Gamma} : B_n^F(t) - K \frac{\log n}{\sqrt{n}} \geq 0 \} \leq \ell \{ t \in \tilde{\Gamma} : u_n(t) \geq 0 \} \leq \ell \{ t \in \tilde{\Gamma} : B_n^F(t) + K \frac{\log n}{\sqrt{n}} \geq 0 \}.
\]

Now,
\[
\ell \{ t \in \tilde{\Gamma} : B(t) + K \frac{\log n}{\sqrt{n}} \geq 0 \} \rightarrow \ell \{ t \in \tilde{\Gamma} : B(t) \geq 0 \},
\]
and
\[
\ell \{ t \in \tilde{\Gamma} : B(t) - K \frac{\log n}{\sqrt{n}} \geq 0 \} \rightarrow \ell \{ t \in \tilde{\Gamma} : B(t) \geq 0 \}.
\]

This and (29) show the announced result. \qed

We recall that the set involved in the limit law in the last result is $\tilde{\Gamma}$, which generally does not coincide with $\Gamma$ (see Lemma 2.1 and (26) for more details). For a better understanding of the links between Theorem 1.1 and 1.2, we note that degeneracy in the limit law is equivalent to $\ell(\Gamma) = 0$. This is an obvious consequence of the next, simple result.

**Lemma 3.2.** If $B(t)$ is a standard Brownian bridge on $[0, 1]$, for any Borel set $A$ in $[0, 1]$, the r.v. $\ell(\{ B > 0 \} \cap A)$ is a.s. constant if and only if $\ell(A) = 0$.

**Proof.** If $\ell(A) = 0$ then, obviously, $\ell(\{ B > 0 \} \cap A) = 0$. Assume now that $\ell(A) > 0$. It is well known that $\ell(\{ t \in [0, 1] : B(t) = 0 \}) = 0$ (this follows easily from Fubini’s Theorem). Moreover, if $B$ is a Brownian bridge then $B = _d - B$. Hence, $\ell(\{ B < 0 \} \cap A) = _d \ell(\{ B > 0 \} \cap A)$, while $\ell(\{ B < 0 \} \cap A) +$
\(\ell(\{B > 0\} \cap A) = \ell(A)\). This implies that \(E(\ell(\{B > 0\} \cap A)) = \ell(A)/2\). Thus, if \(\ell(\{B > 0\} \cap A)\) were a.s. constant, that constant should equal \(\ell(A)/2\). However, \(\ell(B > 0)\) stochastically dominates \(\ell(\{B > 0\} \cap A)\), and degeneracy on the value \(\ell(A)/2\) would lead to the conclusion that the \(U(0,1)\) law stochastically dominates Dirac’s measure on \(\ell(A)/2\), which cannot hold if \(\ell(A) > 0\).

To deal with Galton’s rank statistic in the two-sample case we must adapt the argument in the proof of Theorem 1.2. This is done with Lemma 3.3, which will play an important role in our development. It relies on the strong approximation given in Theorem A.6 in [4]. Given two real functions \(f, g\) and versions of independent sequences of Brownian bridges \(\{B_n^F\}, \{B_m^G\}\) and of uniform quantile processes, \(u_n\) and \(v_m\), as in Theorem A.6 in [4], we set

\[
\begin{align*}
f_n(t) &:= f(t + \frac{u_n(t)}{\sqrt{n}}) \quad \text{and} \quad g_n := g(t + \frac{v_n(t)}{\sqrt{m}}), \\
f_n^*(t) &:= f(t + \frac{B_n^F(t)}{\sqrt{n}}) \quad \text{and} \quad g^*_n := g(t + \frac{B_m^G(t)}{\sqrt{m}}).
\end{align*}
\]

**Lemma 3.3.** Consider \(A \subset (0,1)\) such that \(\ell(A) > 0\). With the notation and construction of Theorem A.6 in [4], if we assume that \(f, g\) are two real Lipschitz functions, then there exists \(L > 0\) such that, if \(C_{n,m} := L(\frac{\log n}{n} + \frac{\log m}{m})\), then whenever \(n, m \to \infty\), eventually,

\[
\begin{align*}
\ell\{t \in A : f_n(t) > g_n(t) + C_{n,m}\} &\leq \ell\{t \in A : f_n(t) > g_n(t)\} \\
&\leq \ell\{t \in A : f_n(t) > g_n(t) - C_{n,m}\}. 
\end{align*}
\]

**Proof:** Since \(f\) is Lipschitz, for \(t \in A\) we have that

\[
|f_n(t) - f_n^*(t)| = |f(t + \frac{u_n(t)}{\sqrt{n}}) - f(t + \frac{B_n^F(t)}{\sqrt{n}})| \leq \|f\|_{\text{Lip}} \|u_n - B_n^F\|_{\infty},
\]

with a similar bound for \(|g_n(t) - g_n^*(t)|\). These bounds and (8) in [4] imply that on a probability one set, eventually,

\[
\sup_{t \in A} \left| (f_n - g_n) - (f_n^* - g_n^*) \right| \leq L\left( \frac{\log n}{n} + \frac{\log m}{m} \right) = C_{n,m}
\]

for some positive constant \(L\) (depending only on \(f\) and \(g\)). Observe that

\[
\begin{align*}
\ell\{t \in A : f_n(t) > g_n(t)\} &\leq \ell\{t \in A : f_n(t) > g_n(t) - C_{n,m}\} \\
&\leq \ell\{t \in A : |(f_n(t) - f_n^*(t)) - (g_n(t) - g_n^*(t))| > C_{n,m}\},
\end{align*}
\]

\[
\begin{align*}
\ell\{t \in A : f_n(t) > g_n(t) + C_{n,m}\} &\leq \ell\{t \in A : f_n(t) > g_n(t)\} \\
&\leq \ell\{t \in A : |(f_n(t) - f_n^*(t)) - (g_n(t) - g_n^*(t))| > C_{n,m}\}.
\end{align*}
\]

On a probability one set the second summands on the last two upper bounds eventually vanish. Hence, on that probability one set, (31) eventually holds.

We will apply Lemma 3.3 to the cases in which \(f = F^{-1}\) and \(g = G^{-1}\) and when \(f\) is the identity and \(g = F_G\) (see Section A in [4] for the analysis of the Lipschitz condition on \(F_G\)).

We end the section with the proof of the two-sample analogue of Theorem 1.2.
Proof of Theorem 1.3. By taking subsequences we can assume \( \frac{n}{n+m} \to \lambda \in (0,1) \). Also, after Lemma 3.1, it suffices to prove that

\[
\ell \{ t \in \hat{\Gamma} : F_n^{-1}(t) > G_m^{-1}(t) \} \overset{w}{\to} \ell \{ t \in \hat{\Gamma} : B(t) > 0 \},
\]

which, using Theorem A.6 in [4] and Lemma 3.3, will hold if we have the convergences

\[
\ell \{ t \in \hat{\Gamma}, t + \frac{B_F(t)}{\sqrt{n}} > F_G(t + \frac{B_G(t)}{\sqrt{m}} + C_{n,m}) \} \overset{w}{\to} \ell \{ t \in \hat{\Gamma} : B(t) > 0 \} \quad (32)
\]

\[
\ell \{ t \in \hat{\Gamma}, t + \frac{B_F(t)}{\sqrt{n}} > F_G(t + \frac{B_G(t)}{\sqrt{m}} - C_{n,m}) \} \overset{w}{\to} \ell \{ t \in \hat{\Gamma} : B(t) > 0 \}. \quad (33)
\]

Both terms can be handled similarly, hence we will address here only (32). First, we note that \( \ell \{ t \in \hat{\Gamma} : F_G(t) \leq x \} = \ell ((0,x] \cap \hat{\Gamma}) \), thus it defines a measure with density function \( I_F(t) \), and, by the Lebesgue differentiation theorem,

\[
\lim_{h \to 0} \frac{F_G(t+h) - t}{h} = 1 \quad \text{for almost every } t \in \hat{\Gamma}. \quad (34)
\]

Now, from

\[
\ell \{ t \in \hat{\Gamma}, t + \frac{B_F(t)}{\sqrt{n}} > F_G(t + \frac{B_G(t)}{\sqrt{m}} + C_{n,m}) \} \quad (35)
\]

\[
= \ell \{ t \in \hat{\Gamma}, \sqrt{\frac{m+n}{n} B_F(t)} > \sqrt{\frac{m+n}{m} B_G(t)}(F_G(t + \frac{B_G(t)}{\sqrt{m}} + C_{n,m}) \}
\]

\[
= \ell \{ t \in \hat{\Gamma}, \sqrt{\frac{m+n}{n} B_F(t)} > \sqrt{\frac{m+n}{m} B_G(t)}(F_G(t + \frac{B_G(t)}{\sqrt{m}}) - t) + \frac{B_G(t)}{\sqrt{m}} + C_{n,m} \},
\]

where \( B_F \) and \( B_G \) are independent standard Brownian bridges, (34), the expression of \( C_{n,m} \) and dominated convergence imply convergence to

\[
\ell \{ t \in \hat{\Gamma}, \lambda^{-1/2} B_F(t) - (1-\lambda)^{-1/2} B_G(t) > 0 \}.
\]

Finally, independence between \( B_F \) and \( B_G \) gives that \( \lambda^{-1/2} B_F(t) - (1-\lambda)^{-1/2} B_G(t) \) is a scaled Brownian bridge (it can be written as \( (\lambda^{-1} + (1-\lambda)^{-1})^{1/2} B(t) \), where \( B(t) \) is a standard Brownian bridge). Therefore the limit law in (35) is that \( \ell \{ t \in \hat{\Gamma} : B(t) > 0 \}. \quad \square
\]

Remark 3.4. Obviously, for any Borel set \( A \subseteq [0,1] \), and a Brownian bridge \( B \), the support of the distribution of \( \ell \{ B > 0 \} \cap A \) is contained in \( [0,\ell(A)] \). One could conjecture that it should be even uniform on \( (0,\ell(A)) \). However, a second thought shows that, in fact, depends on the set \( A \) and that it could even be non-continuous. It is well known (see e.g. pag. 42 in [20]) that \( P(B(t) \neq 0 \text{ for } a < t < b) = 0 \) if \( 0 < a < b < 1 \), thus if \( A \) is contained in \( [a,b] \), then the probability of the event \( \ell \{ B > 0 \} \cap A = \ell(A) \} \) is strictly positive. The distribution has two atoms: at \( \ell(A) \) and at 0. \quad \square

4. Rates of convergence

When the set \( \hat{\Gamma} \) is negligible, Theorems 1.2 and 1.3 yield convergence of Galton’s rank statistic to the index \( \gamma(F,G) \). We investigate in this section the rate of convergence in this result when the contact set \( \Gamma^* \) (recall Definition 1.4) is finite. The following simple result will be crucial in our analysis.
Lemma 4.1. Assume that \([a, b] \subset [0, 1]\) and \(\Gamma^*\) is such that \(t - F_G(t) > \delta > 0\) for every \(t \in [a, b]\). If \(\frac{n}{n + m} \to \lambda \in (0, 1)\) then, for every \(\varepsilon > 0\) such that \(a + \varepsilon < b - \varepsilon\), we have a.s. eventually

\[\{t \in [a + \varepsilon, b - \varepsilon]: F_n^{-1}(t) > G_m^{-1}(t)\} = \{t \in [a + \varepsilon, b - \varepsilon]: F^{-1}(t) > G^{-1}(t)\}.\]

The same conclusion holds if \(t - F(t) < -\delta\) for every \(t \in [a, b]\).

Proof: We have \(F^{-1}(t) > G^{-1}(t)\) for every \(t \in [a, b]\). Using the representation (14),

\[\{t \in [a + \varepsilon, b - \varepsilon]: F_n^{-1}(t) > G_m^{-1}(t)\} = \{t \in [a + \varepsilon, b - \varepsilon]: t + \frac{u_n(t)}{\sqrt{n}} > F_G(t + \frac{v_m(t)}{\sqrt{m}})\}.\]

Without loss of generality we can assume that the chosen version of \(u_n\) satisfies \(\sup_{0 \leq t \leq 1} |u_n(t)|\) is a.s. bounded, and the same for \(v_m\). Then, a.s., we have that for all \(t \in [a + \varepsilon, b - \varepsilon]\), eventually \(t + \frac{v_m(t)}{\sqrt{m}} \in [a, b]\) and therefore \(F_G(t + \frac{v_m(t)}{\sqrt{m}}) < t + \frac{v_m(t)}{\sqrt{m}} - \delta < t + \frac{u_n(t)}{\sqrt{n}}\) for large enough \(n\) and \(m\) and the result follows. The same argument fixes the case \(t - F_G(t) < -\delta\). \(\square\)

Now, using the notation (6), we obtain from Lemmas 4.1 and 2.8 and the subsequent comments:

Corollary 4.2. If \(\Gamma^* = \{t_1, \ldots, t_k\}\), \(k > 0\), \(\frac{n}{n + m} \to \lambda \in (0, 1)\) and \(\eta > 0\) is such that \(\{t_i\} = \Gamma^* \cap (t_i - \eta, t_i + \eta), i = 1, \ldots, k\), then for \(s > 0\)

\[n^s(\gamma(F_n, G_m) - \gamma(F, G)) = n^s \sum_{i=1}^k \ell_{n,m}^{t_i} + o_P(1).\]

The main consequence of Lemma 4.1 and Corollary 4.2 is that when \(\Gamma^*\) is finite the key to the asymptotic behaviour of \(\gamma(F_n, G_m)\) is the (joint) asymptotic behaviour of \(\ell_{n,m}^{t_i}\). We address this problem in this section when \(\Gamma^*\) consists of regular contact points. We note that these regular contact points (recall (8)) are elements of \(\Gamma^*_{F,G}\). This, apparently, excludes contact points in \(\Gamma^*_F\) but not in \(\Gamma^*_G\) or points which would be regular if we exchange the roles of \(F\) and \(G\) but are not with the present definition. However, these cases can often be handled with the same approach. To see this, observe that when \(\Gamma^*\) is finite (recall the concluding remarks in Section 2) we have that \(\ell(t \in A: F_n^{-1}(t) \leq G_m^{-1}(t)) = \ell(t \in A: F_n^{-1}(t) < G_m^{-1}(t))\) for every measurable \(A\).

If we assume further that \(F\) and \(G\) have no common discontinuity point (see Proposition A.2 and the more general Proposition A.3, involving just local conditions, in [4]), then \(\ell(t \in A: F_n^{-1}(t) \leq G_m^{-1}(t)) = \ell(t \in A: F_n^{-1}(t) < G_m^{-1}(t))\) a.s. and we have the a.s. equality

\[\ell_{n,m}^{t_0} = -\ell\left(\{F_n^{-1} < G_m^{-1}\} \cap (t_0 - \eta, t_0 + \eta)\right) - \ell\left(\{F^{-1} < G^{-1}\} \cap (t_0 - \eta, t_0 + \eta)\right) =: -\ell_{m,n}^{t_0}.\]

Observe that \(\ell_{m,n}^{t_0}\) is the same statistic as \(\ell_{n,m}^{t_0}\) after exchanging the roles of the \(X\) and the \(Y\) samples. Hence, we restrict our analysis to points in \(\Gamma^*_F\). Our results hold for points in \(\Gamma^*_G\) with obvious changes.

We note that for every regular contact point, \(t_0\), there exists \(\eta^* > 0\) such that

\[\text{sgn}(F_G(t) - t)\text{ is non-null and constant on each of } (t_0 - \eta^*, t_0) \text{ and } (t_0, t_0 + \eta^*).\] (37)

Taking \(\eta^*\) small enough (to exclude other contact points from the interval), from the final comments in Section 2: \(\text{sgn}(F_G(t) - t) = \text{sgn}(G^{-1}(t) - F^{-1}(t))\) for every \(t \in (t_0 - \eta^*, t_0) \cup (t_0, t_0 + \eta^*)\). Now,
if (37) holds, the study of $\ell_{n,m}^t$ can be carried out through the study, for $\eta \in (0, \eta^*)$, of the pieces

$$L_{n,m}^\geq := \int_{t_0-\eta}^{t_0} I_{\{F_n^{-1}(s) > G_m^{-1}(s)\}} \, ds \quad \text{and} \quad R_{n,m}^\geq := \int_{t_0}^{t_0+\eta} I_{\{F_n^{-1}(s) > G_m^{-1}(s)\}} \, ds,$$

$$L_{n,m}^\leq := \int_{t_0-\eta}^{t_0} I_{\{F_n^{-1}(s) \leq G_m^{-1}(s)\}} \, ds \quad \text{and} \quad R_{n,m}^\leq := \int_{t_0}^{t_0+\eta} I_{\{F_n^{-1}(s) \leq G_m^{-1}(s)\}} \, ds,$$

corresponding to the interval(s) (if any) where $F^{-1} > G^{-1}$. For example, for a crossing point $t_0$ such that $F^{-1} < G^{-1}$ on $(t_0-\eta, t_0)$ and $F^{-1} > G^{-1}$ on $(t_0, t_0+\eta)$, $\ell_{n,m}^t = L_{n,m}^\geq - R_{n,m}^\leq$ (for small enough $\eta$ this happens when $C_L(t_0) > 0, C_R(t_0) < 0$). We use this notation in the following proofs.

**Proof of Theorem 1.5.** We assume, for instance, that $C_L > 0$, and $r_L \geq r_R$, thus $r = r_L$. The other cases can be handled similarly. We note that $\ell_{n,m}^t = L_{n,m}^\geq + R_{n,m}^\leq$ if $C_R > 0$, while $\ell_{n,m}^t = L_{n,m}^\geq - R_{n,m}^\leq$ if $C_R < 0$. We consider first the case $r_L > 1$. We set $d_n = (n + m)^{1/2r}$ and prove next that

$$d_n L_{n,m}^\geq \overset{w}{\to} \ell \{ y < 0 : B_\lambda(t_0) > C_L |y|^r \}. \tag{38}$$

To check this we note that, using (8) in [4], (30), Lemma 3.3 and (15), it is enough to prove that

$$d_n \ell \{ t \in \mathcal{I} : t + \frac{B_1(t)}{\sqrt{n}} > F_G(t + \frac{B_2(t)}{\sqrt{m}}) - C_{n,m} \} \overset{w}{\to} \ell \{ y < 0 : B_\lambda(t_0) > C_L |y|^r \} \tag{39}$$

and similarly with $d_n \ell \{ t \in \mathcal{I} : t + \frac{B_1(t)}{\sqrt{n}} > F_G(t + \frac{B_2(t)}{\sqrt{m}}) + C_{n,m} \} \rightarrow \ell \{ y < 0 : B_\lambda(t_0) > C_L |y|^r \}$ whenever $t < 0$ or $t > 0$. Then

$$I \{ t \in \mathcal{I} : t + \frac{B_1(t)}{\sqrt{n}} - F_G(t + \frac{B_2(t)}{\sqrt{m}}) - C_{n,m} \} = I \{ t \in \mathcal{I} : t + \frac{B_1(t)}{\sqrt{n}} > t_0 + \xi_m + C(\xi_m^*) |\xi_m^*| + o(\xi_m^*) - C_{n,m} \}$$

$$= I \{ t \in \mathcal{I} : B_{n,m}^{1/2}(t) > \sqrt{n+m}(C(\xi_m^*) |\xi_m^*| + o(\xi_m^*) - C_{n,m}) \}, \tag{40}$$

where $\alpha_n = \sqrt{(n+m)/n}$, $\beta_m = \sqrt{(n+m)/m}$, $B_{n,m}^{1/2}(t) = \alpha_n B_1(t) - \beta_m B_2(t)$, $\xi_m = (t + \frac{B_2(t)}{\sqrt{m}} - t_0)$ and we have used that $t_0 = F_G(t_0)$. Denoting $\xi_m^*(y) = \frac{y}{d_n} + \frac{B_G(t_0) + \frac{y}{d_n}}{\sqrt{m}}$, the change of variable $t = t_0 + \frac{y}{d_n}$, and (40) lead to

$$d_n \ell \{ t \in \mathcal{I} : t + \frac{B_1(t)}{\sqrt{n}} > F_G(t + \frac{B_2(t)}{\sqrt{m}}) - C_{n,m} \} \overset{w}{\to} \ell \{ y < 0 : B_\lambda(t_0) > C_L |y|^r \} \tag{41}$$

$$= \int_{-\eta d_n}^{0} \{ B_{n,m}^{1/2}(t_0 + \frac{y}{d_n}) > C(\xi_m^*(y))(\frac{(n+m)^{1/2r}}{d_n} y + \frac{(n+m)^{1/2r}}{d_n} B_2(t_0 + \frac{y}{d_n}) |\xi_m^*(y)| - C_{n,m}) \} dy.$$

Since the Brownian bridges have continuous trajectories, they are bounded and:

$$\sup_{y \in [-\eta d_n, 0]} \xi_m^*(y) \leq \sup_{x \in [0, 1]} \frac{|B_2(t)|}{\sqrt{m}} \to 0,$$

$$\inf_{y \in [-\eta d_n, 0]} \xi_m^*(y) \geq -\eta - \frac{\sup_{x \in [0, 1]} |B_2(t)|}{\sqrt{m}} \to -\eta.$$
Thus, eventually, for every $y < 0$, $\xi_m^*(y) \in [-n^*, n^*]$ and $C(\xi_m^*(y)) \geq \min(|C_L|, |C_R|) > 0$. This, the fact that $B_1$ and $B_2$ are a.s. bounded, and also that $(n + m)^{1/2r}$ is either equal to one or, else, goes to infinity, yield that, a.s., the order of $\sqrt{n + m} \xi_m^*(y)^r$ is $|y|^r$ or higher. Finally, by definition of $C_{n,m}$, there exists $M > 0$ (depending on the trajectories of the Brownian bridges) such that

$$I \left\{ y \in [-n^*, 0); B_n^{1/2}(t + \frac{y}{dn}) > C(\xi_m^*(y)) \right\} \leq I(-M \leq y \leq 0).$$

Now, if we fix $y < 0$ such that $B_\lambda(t_0) \neq C_L |y|^r$, then, a.s.,

$$I \left\{ B_n^{1/2}(t_0 + \frac{y}{dn}) > C(\xi_m^*(y)) \right\} \leq I \left\{ B_\lambda(t_0) > C_L |y|^r \right\}.$$

From here, dominated convergence yields (39), hence (38). We note that the limit in (38) equals

$$\text{sgn}(C_L) \left( \frac{(B_\lambda(t_0))^{\text{sgn}(C_L)}}{|C_L|} \right)^{1/r_0}.$$ (42)

A completely similar analysis shows that

$$d_n^r R_{n,m}^{\gamma_0} \xrightarrow{w} \text{sgn}(C_R) \left( \frac{(B_\lambda(t_0))^{\text{sgn}(C_R)}}{|C_R|} \right)^{1/r_0} I(r_R = r_0)$$

when $C_R > 0$ ($d_n^r R_{n,m}^{\gamma_0}$ vanishes in probability if $r_R < r_L = r_0$). Furthermore, we are using the same strong approximation to handle $d_n^r L_{n,m}^{\gamma_0}$ and $d_n^r R_{n,m}^{\gamma_0}$, which implies that there is weak convergence of $(d_n^r L_{n,m}^{\gamma_0}, d_n^r R_{n,m}^{\gamma_0})$ and, consequently, of $d_n^r L_{n,m}^{\gamma_0} = d_n^r (L_{n,m}^{\gamma_0} + R_{n,m}^{\gamma_0})$. This completes the proof in the case $r_L \geq r_R$, $C_L > 0, C_R > 0$. The other cases with $r > 1$ follow similarly.

The case $r_L = r_R = 1$ goes along the same lines, only note that, a.s., if $y < 0$ satisfies that $B_\lambda(t_0) \neq C_\ast |y + B_2(t_0)/\sqrt{1-\lambda}|$, where $\ast = L$ or $R$ whenever $\text{sgn} \left( y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right) = -1$ or $+1$, we would have

$$I \left\{ B_n^{1/2}(t_0 + \frac{y}{dn}) > C(\xi_m^*(y)) \right\} \leq I \left\{ B_\lambda(t_0) > C_\ast |y + B_2(t_0)/\sqrt{1-\lambda}| \right\}$$

and, by dominated convergence,

$$d_n^r L_{n,m}^{\gamma_0} \xrightarrow{w} \ell \left\{ y < 0; B_\lambda(t_0) > C_\ast |y + B_2(t_0)/\sqrt{1-\lambda}| \right\}$$

$$= \ell \left\{ y < 0; B_\lambda(t_0) > -C_L \left( y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}, y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right) < 0 \right\}$$

$$+ \ell \left\{ y < 0; B_\lambda(t_0) > C_R \left( y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}, y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} > 0 \right) \right\}.$$

The right side of the interval is dealt with in a similar way. In the case $C_L > 0, C_R > 0$, we obtain

$$d_n^r R_{n,m}^{\gamma_0} \xrightarrow{w} \ell \left\{ y > 0; B_\lambda(t_0) > -C_L \left( y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}, y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} < 0 \right) \right\}$$

$$+ \ell \left\{ y > 0; B_\lambda(t_0) > C_R \left( y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}, y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} > 0 \right) \right\}.$$

Hence,

$$d_n^r \ell_0 = d_n^r (L_{n,m}^{\gamma_0} + R_{n,m}^{\gamma_0}) \xrightarrow{w} \ell \left\{ y; B_\lambda(t_0) > -C_L (y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}, y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} < 0 \right\}.$$
Galton rank-order statistic

\[ + \ell(y : B_\lambda(t_0) > C_R(y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}), y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}) > 0) \]

\[ = \frac{(B_\lambda(t_0))^+}{C_L} + \frac{(B_\lambda(t_0))^+}{C_R} = T_{1,1}(t_0; C_L, C_R). \]

If \( C_L > 0, C_R < 0 \) we get

\[ d_n R_{n,m}^c \xrightarrow{w} \ell\{y > 0 : B_\lambda(t_0) < -C_L\left(y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}\right), y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} < 0\} \]

\[ + \ell\{y > 0 : B_\lambda(t_0) < C_R\left(y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}\right), y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} > 0\}. \]

Therefore,

\[ d_n \ell_{n,m}^{t_0} = d_n (L_{n,m}^c - R_{n,m}^c) \xrightarrow{w} \ell\{y > 0 : B_\lambda(t_0) > -C_L\left(y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}\right), y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} < 0\} \]

\[ - \ell\{y > 0 : y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} < 0\} \]

\[ - \ell\{y : B_\lambda(t_0) < C_R\left(y + \frac{B_2(t_0)}{\sqrt{1-\lambda}}\right), y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} > 0\} \]

\[ + \ell\{y > 0 : y + \frac{B_2(t_0)}{\sqrt{1-\lambda}} > 0\} \]

\[ = \frac{(B_\lambda(t_0))^+}{C_L} - \frac{(B_2(t_0))^+}{\sqrt{1-\lambda}} + \frac{(B_\lambda(t_0))^+}{C_R} + \frac{(B_2(t_0))^+}{\sqrt{1-\lambda}} \]

\[ = \frac{(B_\lambda(t_0))^+}{C_L} + \frac{(B_\lambda(t_0))^+}{C_R} + \frac{B_2(t_0)}{\sqrt{1-\lambda}} = T_{1,1}(t_0; C_L, C_R). \]

The remaining cases are completely similar. We omit further details. \( \square \)

Some comments are in order here. First note that, by focusing on the transform \( F_G \), Theorem 1.5 is able to handle virtual contact points for \( F^{-1} \) and \( G^{-1} \). As an illustration of this claim, assume \( F^{-1}(t_0) < G^{-1}(t_0) \leq G^{-1}(t_0^+) < F^{-1}(t_0^+) \) \( (t_0^+ \) is then a virtual crossing point). As noted above, \( F_G(t) = t_0 \) in an interval \((t_0 - \eta, t_0 + \eta)\) for \( \eta \) small enough, and (7) holds with \( r_R = r_L = 1, C_R(t_0) = -1 \) and \( C_L(t_0) = +1 \). Thus, Theorem 1.5 applies and gives (43) below.

The case \( F^{-1}(t_0) < G^{-1}(t_0) \leq F^{-1}(t_0^+) < G^{-1}(t_0^+) \) (a virtual tangency point) can be handled similarly, although it does not fit exactly in the setup of Theorem 1.5. In this case we have that, for some small enough \( \eta, \delta > 0 \), \( F_G(t) = t_0 \), \( t \in (t_0 - \eta, t_0) \), \( F_G(t) > t + \delta \), \( t \in (t_0, t_0 + \eta) \). It is easy to see that, eventually, \( t_{n,m}^{t_0} = \ell\{t \in (t_0 - \eta, t_0 + \eta) : t + \frac{u_{\alpha}(t)}{\sqrt{m}} > t_0, t + \frac{v_{\alpha}(t)}{\sqrt{m}} < t_0\} \). From this point one can argue as in the proof of Theorem 1.5 to obtain (44) below.

We include in the following proposition these results for virtual contact points. Notice that this proposition includes the possibility of non-continuous d.f.’s \( F \) or \( G \).

**Proposition 4.3.** Let \( t_0 \in \Gamma^* \cap (0,1) \), such that for some \( \eta_0 > 0 \), \((t_0 - \eta_0, t_0 + \eta_0) \cap \Gamma^* = \{t_0\}\). Then, for every small enough \( \eta > 0 \), if \( n/(n + m) \to \lambda \in (0,1) \) as \( n, m \to \infty \), we have that:

(i) **(virtual crossing points)** If \( F^{-1}(t_0) < G^{-1}(t_0) \leq G^{-1}(t_0^+) < F^{-1}(t_0^+) \), then:

\[ \frac{\sqrt{n + m t_0}}{\sqrt{X}} \xrightarrow{w} B_1(t_0). \]  

(43)
(ii) (virtual tangency points) If $F^{-1}(t_0) < G^{-1}(t_0) \leq F^{-1}(t_0+) < G^{-1}(t_0+)$, then,
\[
\sqrt{n + m\ell_{n,m}^0} \overset{w}{\to} \ell \{ y : -\frac{B_2(t_0)}{\sqrt{1-\lambda}} > y, -\frac{B_1(t_0)}{\sqrt{\lambda}} < y \} = \left(\frac{B_1(t_0)}{\sqrt{\lambda}} - \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right)^+. \tag{44}
\]

We can easily adapt Proposition 4.3 to the case $G^{-1}(t_0) < F^{-1}(t_0) \leq F^{-1}(t_0+) < G^{-1}(t_0+)$. In this case $F_G(t_0) < t_0 < F_G(t_0+)$ (but $G_F(t) = t_0$ for $t$ close to $t_0$). We call this kind of virtual crossing a vertical crossing, while we will refer to case (i) as a horizontal crossing. For vertical crossing points the argument above yields $\sqrt{n + m\ell_{n,m}^0} \overset{w}{\to} \left(\frac{B_1(t_0)}{\sqrt{\lambda}} - \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right)^- \text{.}$ Also, (ii) corresponds to an upper tangency point, in the sense that $F_G$ touches the identity at $t_0$ but remains above it in $(t_0 - \eta, t_0 + \eta) \setminus \{t_0\}$. With obvious changes we can deal with lower tangency points, obtaining then $\sqrt{n + m\ell_{n,m}^0} \overset{w}{\to} \left(\frac{B_1(t_0)}{\sqrt{\lambda}} - \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right)^- \text{.}$

A further observation is that, since any contact order $r \geq 1$ is possible (see Example 4.5), we can obtain any rate of convergence $(m + n)^{-s}$, $s \leq 1/2$ for $\ell_{n,m}^0 \text{.}$ As previously mentioned, only the case $r_L(t_0) = r_R(t_0) = 1$ and $C_L(t_0) = -C_R(t_0)$ leads to asymptotic normality.

Finally, we note that the limiting expressions become simpler under regularity. In fact, if $h(t) = F_G(t) - t$ is $r$ times differentiable with continuity at a point $t_0 \in (0, 1)$, such that $h(t_0) = 0$ and with derivatives $h(k)(t_0) = 0$, $k = 1, \ldots, r - 1$ and $h^{(r)}(t_0) \neq 0$, then $t_0$ is an isolated contact point in the sense of (8) with $r_L(t_0) = r_R(t_0) = r$. For odd $r \geq 3$ we have $C_R(t_0) = -C_L(t_0) = \frac{h^{(r)}(t_0)}{r!}$ and the conclusion in Theorem 1.5 reads
\[
(n + m)^{\frac{1}{2}} \ell_{n,m}^0 \overset{w}{\to} \left(\frac{r!}{|h^{(r)}(t_0)|}\right)^{1/r} \left(\left(\frac{\lambda}{\lambda - 1}\right) \left(\frac{B_1(t_0)}{\sqrt{\lambda}} - \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right)^+ - \left(\frac{\lambda}{\lambda - 1}\right) \left(\frac{\lambda}{\lambda - 1}\right) \left(\frac{B_1(t_0)}{\sqrt{\lambda}} - \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right)^- \right), \tag{45}
\]
while for $r = 1$ it becomes
\[
(n + m)^{\frac{1}{2}} \ell_{n,m}^0 \overset{w}{\to} \text{sgn}(h^{(1)}(t_0)) \left(\frac{r!}{|h^{(1)}(t_0)|} B_1(t_0) + \frac{1}{\sqrt{1-\lambda}} (1 + \frac{1}{\lambda^{(1)}}) B_2(t_0) \right). \tag{46}
\]

For even $r \geq 2$ we have $C_R(t_0) = C_L(t_0) = \frac{h^{(r)}(t_0)}{r!}$ and Theorem 1.5 yields
\[
(n + m)^{\frac{1}{2}} \ell_{n,m}^0 \overset{w}{\to} \text{sgn}(h^{(r)}(t_0)) 2^{1/r} \left(\frac{r!}{|h^{(r)}(t_0)|}\right)^{1/r} \left(\left(\frac{\lambda}{\lambda - 1}\right) \left(\frac{B_1(t_0)}{\sqrt{\lambda}} - \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right)^+ - \left(\frac{\lambda}{\lambda - 1}\right) \left(\frac{\lambda}{\lambda - 1}\right) \left(\frac{B_1(t_0)}{\sqrt{\lambda}} - \frac{B_2(t_0)}{\sqrt{1-\lambda}} \right)^- \right)^{1/r}. \tag{47}
\]

When the contact point is extremal, that is, when $t_0 \in \{0, 1\}$, the limiting r.v.'s in Theorem 1.5 vanish. We prove now Theorem 1.6, showing that in this case there is weak convergence, at a faster rate, to a nondegenerate limiting distribution.

**Proof of Theorem 1.6.** Let us take $t_0 = 0$ and $C > 0$. The cases with $C < 0$ and/or $t_0 = 1$ are similar. We handle first the case $r = 1$. Then for small enough $\eta$ we have $\ell \left(\{ F^{-1} > G^{-1} \} \cap (0, \eta) \right) = 0$. We recall that
\[
\rho_{n,m}^0 = \ell \left(\{ F^{-1} > G^{-1} \} \cap (0, \eta) \right). \tag{48}
\]

We use the well-known fact that the joint law of $\frac{1}{\sqrt{n+1}} (S_1^1, \ldots, S_n^1)$ is the same as that the ordered sample of size $n$ of i.i.d. $U(0, 1)$ r.v.'s. Thus,
\[
(X_1, \ldots, X_n) \overset{d}{=} \left( F^{-1} \left( \frac{S_1^1}{\sqrt{n+1}} \right), \ldots, F^{-1} \left( \frac{S_n^1}{\sqrt{n+1}} \right) \right). \tag{49}
\]
with a similar expression for the Y-sample. From (48) and (49) we see that

$$
\ell_{n,m}^0 = \int_0^\eta \left\{ F^{-1}\left( \frac{S_1^1}{n+m} \right) > G^{-1}\left( \frac{S_2^2}{n+m} \right) \right\} dt = \frac{1}{n+m} \int_0^{(n+m)\eta} I\{ F^{-1}(\xi_n^1(y)) > G^{-1}(\xi_m^2(y)) \} dy,
$$

where \( \xi_n^1(y) := \frac{S_1^1}{n+m} y / S_{n+1}^1 \) and \( \xi_m^2(y) := \frac{S_2^2}{m+n} y / S_{m+1}^2 \). Now (8) yields that, if \( y \in (0, (n+m)\eta) \), then

$$
I\{ F^{-1}(\xi_n^1(y)) > G^{-1}(\xi_m^2(y)) \} = I\{ \xi_n^1(y) > F_C(\xi_m^2(y)) \}
= I\{ (n+m)\xi_n^1(y) > (1+C)(n+m)\xi_m^2(y) + (n+m)o(\xi_m^2(y)) \}.
$$

(50)

The SLLN implies that there exists \( \Omega_0 \), with \( P(\Omega_0) = 1 \), such that for every \( \omega \in \Omega_0 \), \( \frac{S_1^1}{n} \to 1 \) and \( \frac{S_2^2}{m} \to 1 \). Therefore, for any \( \delta^* > 0 \), if \( \omega \in \Omega_0 \), eventually

$$
0 < \xi_m^2(y) \leq \frac{S_2^2[1-(\lambda+\delta^*)]\lambda y+1}{S_{m+1}^2} \to 0.
$$

(51)

This and (50) show that if \( \omega \in \Omega_0 \),

$$
I\{ F^{-1}(\xi_n^1(y)) > G^{-1}(\xi_m^2(y)) \} \to I\{ (1-\lambda)S_1^1[\lambda y] > \lambda(1+C)S_2^2[1-(\lambda+\delta^*)]y \},
$$

for every \( y \) not belonging to the countable set \( \left\{ \frac{j}{1-\lambda} : j = 0, 1, \ldots \right\} \cup \left\{ \frac{j}{\lambda} : j = 0, 1, \ldots \right\} \). Clearly, in \( \Omega_0 \) we have

$$
\lim_{y \to \infty} \frac{(1-\lambda)S_1^1}{\lambda S_2^2[1-(1-\lambda)y]} = 1.
$$

Hence, the fact that \( C > 0 \) gives that, in \( \Omega_0 \),

$$
I\{ (1-\lambda)S_1^1[\lambda y] > \lambda(1+C)S_2^2[1-(\lambda+\delta^*)]y \}
= 0 \text{ for large enough } y.
$$

This shows that

$$
\int_0^\infty I\{ (1-\lambda)S_1^1[\lambda y] > \lambda(1+C)S_2^2[1-(\lambda+\delta^*)]y \} dy \text{ is an a.s. finite r.v..}
$$

We will conclude (13) as soon as we prove that for every \( \omega \in \Omega_0 \) we can apply dominated convergence. To check this, notice that (51) gives that, for \( m \) large enough,

$$
I\{ (m+n)\xi_n^1(y) > (m+n)(1+C)\xi_m^2(y) + (m+n)o(\xi_m^2(y)) \}
\leq I\{ (m+n)\xi_n^1(y) > (1+C/2)(m+n)\xi_m^2(y) \}.
$$

(52)

Now, for every \( \omega \in \Omega_0 \), there exist a natural number and a positive real number depending on \( \omega \), \( N(\omega) \) and \( Y(\omega) \), such that, if \( n \geq N(\omega) \) then both \( S_{n+1}^1/n \) and \( S_{m+1}^2/m \) are close to one, and, if we take \( y \geq Y(\omega) \), then, both \( \frac{S_1^1}{n+m} y \) and \( \frac{S_2^2}{n+m} y \) are close to \( y \). This completes the proof for the case \( r = 1 \), since (52) gives that, for all \( n \geq N(\omega) \),

$$
I\{ \xi_n^1(y) > (1+C(\xi_m^2(y)))(m+n)\xi_m^2(y) + (m+n)o(\xi_m^2(y)) \} \leq I_{0,Y(\omega)}.
$$

For the case \( r > 1 \) we assume, again, \( C > 0 \) and observe that the r.v. \( T_{r+1}(0; C, C) \) is a.s. finite (this follows, for instance, from the fact that, a.s., \( W_0(y)/y \to 0 \) as \( y \to \infty \)). Now \( \ell_{n,m}^0 \) has the same
expression as in (48). We will use the same notation as in Lemma 3.3. First, we have that
\[
e^{\theta_{n,m}} = \ell\left\{ t \in (0, \eta) : F^{-1}(t + \frac{u_n(t)}{\sqrt{n}}) > G^{-1}(t + \frac{v_m(t)}{\sqrt{m}}) \right\}
\]
\[
e^{\phi_{n,m}} = \ell\left\{ t \in (0, \eta) : t + \frac{u_n(t)}{\sqrt{n}} > G(t + \frac{v_m(t)}{\sqrt{m}}) \right\}.
\]

Therefore, if we take \( f \) equal to the identity and \( g = F_G \) in Lemma 3.3, we only need to show that
\[
d_n e^{\theta_{n,m}} = d_n \int_0^\eta I\left(\alpha_n d_n W_{F}(t) > \beta_m d_n W_{G}(t) + \sqrt{n+m}C(t + \frac{G(t)}{\sqrt{m}})\right) dt
\]
\[
= \int_0^\eta I\left(\alpha_n d_n W_{F}(y) - y W_{F}(d_n) > \beta_m d_n W_{G}(y) - y W_{G}(d_n) + C(|\xi_n(y)| d_n |\xi_n(y)|^r + d_m^2 C(|\xi_n(y)|^r) - L_{n,m})\right) dy,
\]
where \( \alpha_n = ((m+n)/n)^{1/2}, \beta_m = ((m+n)/m)^{1/2} \) and \( \xi_n(y) = y d_n + \frac{1}{\sqrt{m}} G_{m}(\frac{y}{d_n}) \).

As it is well known, there exists \( \Omega_0 \in \sigma \), with \( P(\Omega_0) = 1 \) such that, if \( \omega \in \Omega_0 \), then, \( W_i \) is continuous, \( W_i(x)/x \to 0 \), as \( x \to \infty, i = F, G \) and the set
\[
\left\{ y : (1-\theta)^{-1/2} W_F(y) + C(0) y^r \right\}
\]
has Lebesgue measure zero. If we fix \( \omega \in \Omega_0 \), then, we have that
\[
\sup_{y \in [0,d_n]} |\xi_n(y)| \leq \eta + \frac{1}{\sqrt{m}d_n} \sup_{y \in [0,d_n]} \left| W_F(y) - y W_F(d_n)/d_n \right| \to \eta,
\]
and we can conclude that, eventually, \( \{\xi_n(y) : y \in [0,d_n]\} \subset [0, \eta^r] \), and, consequently, from an index onward, \( \inf_{y \in [0,d_n]} C(\xi(y)) \geq \inf_{y \in [0,\eta^r]} C(h) > 0 \). On the other hand, we have \( d_n^2 L_{n,m} \to 0 \) and
\[
d_n^r |\xi_n(y)|^r = \left| y + \beta_m W_G(y) - y W_G(d_n)/d_n \right|^r = y(1 + o(1)) \to \infty, \text{ as } y \to \infty.
\]

Therefore, there exists a constant \( M \) (which possibly depends on the chosen \( \omega \)) such that
\[
I\left(\alpha_n d_n W_{F}(y) - y W_{F}(d_n) > \beta_m d_n W_{G}(y) - y W_{G}(d_n) + C(|\xi_n(y)| d_n |\xi_n(y)|^r + d_m^2 C(|\xi_n(y)|^r) - L_{n,m})\right) \leq I\{0 \leq y \leq M\}.
\]
for every large enough $n$. Moreover, 

\[
I\left\{ \alpha_n \left( W_F(y) - y \frac{W_G(d_n)}{d_n} \right) > \beta_n \left( W_G(y) - y \frac{W_G(d_n)}{d_n} + C(\xi_n(y))d_n(\xi_n(y))' + d_n(\alpha(\xi_n(y))') - L_{n,m} \right) \right\} 
\to I\left\{ \lambda^{-1/2} W_F(y) - (1-\lambda)^{-1/2} W_G(y) > C(0)y^r \right\}.
\]

These observations allow to apply dominated convergence to conclude that, for this $\omega$,

\[
d_{\omega} I\left\{ (\tilde{F}_n^{-1} > \tilde{G}_n^{-1} - L_{n,m}) \cap (0,\eta) \right\} 
\to I\left\{ y \in (0,\infty) : \lambda^{-1/2} W_F(y) - (1-\lambda)^{-1/2} W_G(y) > C(0)y^r \right\}.
\]

The fact that $(\lambda(1-\lambda))^{1/2}(\lambda^{-1/2} W_F(y) - (1-\lambda)^{-1/2} W_G(y))$ is a Brownian motion yields (53). □

Now, the main issue to prove Theorem 1.7 is to prove asymptotic independence between the localized statistics around central and extremal contact points.

**Proof of Theorem 1.7:** From Corollary 4.2 it suffices to prove that $(n + m)\frac{1}{\sqrt{n}} (\ell_{n,m}^i)_{1 \leq i \leq k}$ converges weakly. This follows trivially if $\Gamma^* \subset (0,1)$ after checking that the strong approximation used in the proof of Theorem 1.5 allows to deal with all the $\ell_{n,m}^i$ simultaneously. Hence, it suffices to prove asymptotic independence among $\ell_{n,m}^0$, $(\ell_{n,m}^i)_{i: t_i \in (0,1)}$ and $\ell_{n,m}^1$ when 0 or 1 (or both) are contact points. Let us assume, for instance, that $\Gamma^* = \{0 < t_1 \cdots < t_s < 1\}$ and set $A_n = (n + m)\frac{1}{\sqrt{n}} \ell_{n,m}^0$, $B_n = (n + m)\frac{1}{\sqrt{n}} (\ell_{n,m}^i)_{1 \leq i \leq s}$ and $C_n = (n + m)\frac{1}{\sqrt{n}} \ell_{n,m}^1$. We have that there exist $A, B, C$ such that $A_n \xrightarrow{w} A$, $B_n \xrightarrow{w} B$, $C_n \xrightarrow{w} C$. Assume $(\tilde{A}, \tilde{B}, \tilde{C})$ is a random vector with $\tilde{A}, \tilde{B}, \tilde{C}$ independent, $\tilde{A} \overset{d}{=} A$, $\tilde{B} \overset{d}{=} B$ and $\tilde{C} \overset{d}{=} C$ and consider $(\tilde{A}_n, \tilde{B}_n, \tilde{C}_n)$, with the same properties with respect $(A_n, B_n, C_n)$. $A_n$ is a function of the smallest $\lceil \eta n \rceil$ elements in the $X$ sample and the smallest $\lceil \eta m \rceil$ elements in the $Y$ sample. Similarly, $B_n$ and $C_n$ are functions of the central and upper order statistics. If $d_{TV}$ denotes the distance in total variation, then there exists a universal constant $H > 0$ such that

\[
d_{TV}(\mathcal{L}(A_n, B_n, C_n), \mathcal{L}(\tilde{A}_n, \tilde{B}_n, \tilde{C}_n)) \leq H \left[ \frac{\eta(1 - t_1 - \eta)}{t_1 - 2\eta} + \frac{\eta(t_s + \eta)}{1 - t_s - 2\eta} \right]^{1/2}
\]

for small enough $\eta$ (this follows from Theorem 4.2.9 and Lemma 3.3.7 in [19]). If $\rho$ denotes the Prokhorov metric, then the fact that $\rho(\mu_1, \mu_2) \leq d_{TV}(\mu_1, \mu_2)$ implies

\[
\rho(\mathcal{L}(A_n, B_n, C_n), \mathcal{L}(\tilde{A}_n, \tilde{B}_n, \tilde{C}_n)) \leq H \left[ \frac{\eta(1 - t_1 - \eta)}{t_1 - 2\eta} + \frac{\eta(t_s + \eta)}{1 - t_s - 2\eta} \right]^{1/2},
\]

We prove now that $(A_n, B_n, C_n) \xrightarrow{w} (\tilde{A}, \tilde{B}, \tilde{C})$. Obviously $(\tilde{A}_n, \tilde{B}_n, \tilde{C}_n) \xrightarrow{w} (\tilde{A}, \tilde{B}, \tilde{C})$. Having weakly convergent components, $(A_n, B_n, C_n)$ is tight. To complete the proof it suffices to show that for any weakly convergent subsequence $(A_n', B_n', C_n') \xrightarrow{w} \gamma$, necessarily $\gamma = \mathcal{L}(\tilde{A}, \tilde{B}, \tilde{C})$. To check this, we observe that, since $\rho$ metrizes the weak convergence, we have

\[
\rho(\gamma, \mathcal{L}(\tilde{A}, \tilde{B}, \tilde{C})) \leq H \left[ \frac{\eta(1 - t_1 - \eta)}{t_1 - 2\eta} + \frac{\eta(t_s + \eta)}{1 - t_s - 2\eta} \right]^{1/2}.
\]

Now, using Corollary 4.2 we see that we can repeat the argument leading to (54) for every small enough $\eta$. Hence, $\rho(\gamma, \mathcal{L}(\tilde{A}, \tilde{B}, \tilde{C})) = 0$. This completes the proof. □
4.1. Some examples and extensions

We provide here some simple examples showing different limiting distributions for $\ell_{n,m}^{0}$. Later we give simple sufficient conditions under which extremes have no influence on the asymptotic behaviour of $\gamma(F_n, G_m)$ and give a simplified version of Theorem 1.7 under the assumption that $F$ and $G$ have regular densities. Finally, we consider the case of finitely supported distributions.

**Example 4.4.** In this example $G(t) = t$ (the uniform law on $(0, 1)$). For $r > 0$ we consider the quantile function $F^{-1}(t) = \frac{1}{2} + \text{sgn}(t - \frac{1}{2})|t - 1/2|^r$, $0 \leq t \leq 1$. Then, $F(x) = \frac{1}{2} + \text{sgn}(x - \frac{1}{2})|x - \frac{1}{2}|^{1/r}$, $\frac{1}{2} - \frac{1}{2r} \leq x \leq \frac{1}{2} + \frac{1}{2r}$, $F_G = F$ and $F_G(\frac{1}{2}) = \frac{1}{2}$. Thus, $\frac{1}{2}$ is a contact point. If $r < 1$ then $F_G(t) = \frac{1}{r}|t - \frac{1}{2}|^{r-1}$. In particular, $F_G$ is Lipschitz in a neighbourhood of $\frac{1}{2}$. We easily check that $\Delta(h) = -h + \text{sgn}(h)|h|^{1/r} = -h + o(h)$, that is, $\frac{1}{2}$ is an isolated regular contact point (a crossing point) with intensities $r_L = r_R = 1$ and constants $C_R = -C_L = -1$. We can apply Theorem 1.5 to conclude that

$$(n + m)^{1/2} \ell_{n,m}^{0} \frac{w}{\sqrt{1-\lambda}} \to B_{\frac{2}{1-\lambda}}(\frac{1}{2}).$$

If $r > 1$ then $F_G(\frac{1}{2}) = +\infty$ and $F_G$ is not Lipschitz around the contact point. However, following the reasoning after Corollary 4.2, we have that $\ell_{n,m}^{0} = -\ell_{n,m}^{0}$ and we can handle this case exchanging the roles of the $F$ and $G$ samples and studying $G_F(t) = F^{-1}(t)$. Now $G_F(t) = r|t - \frac{1}{2}|^{r-1}$ and $G_F$ is Lipschitz in a neighbourhood of $\frac{1}{2}$. Furthermore, $\Delta(h) = -h + \text{sgn}(h)|h|^{r-1}$, hence, for $r > 1$ we have that

$$(n + m)^{1/2} \ell_{n,m}^{0} \frac{w}{\sqrt{1-\lambda}} \to B_{\frac{2}{1-\lambda}}(\frac{1}{2}).$$

**Example 4.5.** Now $F$ denotes the d.f. of the uniform law on $(0, 1)$ and $G^{-1}(t) = t + \text{sgn}(t - \frac{1}{2})|t - 1/2|^r$, $0 \leq t \leq 1$. As before, $\frac{1}{2}$ is a contact point. For $r \geq 1$ $F_G = G^{-1}$ is differentiable, with $F_G'(t) = 1 + r|t - 1/2|^{r-1}$. We have $\Delta(h) = \text{sgn}(h)|h|^r$, that is, Theorem 1.5 can be applied here with $r_L = r_R = r$, $C_R = -C_L = 1$. Thus, for $r = 1$ we get

$$(n + m)^{1/2} \ell_{n,m}^{0} \frac{w}{\sqrt{1-\lambda}} \to \frac{B_{\frac{2}{1-\lambda}}(\frac{1}{2})}{\sqrt{1-\lambda}} + \frac{B_{\frac{2}{1-\lambda}}(\frac{1}{2})}{\sqrt{1-\lambda}}.$$

while for $r > 1$ (the case $0 < r < 1$ can be handled exchanging the roles of the two samples) we obtain

$$(n + m)^{1/r} \ell_{n,m}^{0} \frac{w}{\sqrt{1-\lambda}} \to ((B_{\frac{2}{1-\lambda}}(\frac{1}{2}))^{1/r} - ((B_{\frac{2}{1-\lambda}}(\frac{1}{2}))^{-1/r}.$$

**Example 4.6.** Let $F = F_\nu$, with density $f_\nu$, be a Student $t$-distribution with $\nu > 0$ degrees of freedom and $G(x) = G_\nu(x) = F_\nu(x - \mu)$, for some $\mu > 0$. In this case $F^{-1}(0) = G^{-1}(0) = -\infty$ and $F_G(0) = 0$. To ease notation, we set $s = (\nu + 1)/2$, write $K$ for a non-null generic constant which can change from line to line (in particular, $f_\nu(t) = K(\nu + t^2)^{-\nu} - \nu$ and $f_\nu(x)$ when $\int_0^x f_\nu(t) dt = 1$ as $x \to x_0$.

Using l'Hôpital’s rule we see that $F_G(t) \approx K\nu^{-\nu+1}$ as $t \to \infty$ and, as a consequence, $F_G^{-1}(h) \approx K\nu^{1/(1-2s)} h^{-1}$ as $h \to 0+$. Furthermore,

$$F_G(h) = \frac{F_\nu(F_\nu^{-1}(h) + \mu)}{F_\nu(F_\nu^{-1}(h))} = \left(\nu + (F_\nu^{-1}(h) + \mu)^2\right)^{-s} \nu + (F_\nu^{-1}(h))^{-s} \to 1, \quad \text{as } h \to 0+. \quad (55)$$
Some simple but tedious computations give that

\[
F''_G(h) \approx K \left( \frac{\mu (F'_w^{-1}(h))^{2s} + O \left( \left( F'_w^{-1}(h) \right)^{2s-1} \right)}{\nu + (F'_w^{-1}(h))^2} \right);
\]

therefore, \( F''_G(h) \approx K \left( F'_w^{-1}(h) \right)^{2s-2} \). Consequently, \( F''_G(h) \approx Kh^{(2s-2)/(1-2s)} \). Now, applying l’Hôpital’s rule twice we get that \( \Delta(h) \approx Kh^{\frac{2s-2}{1-2s}} = Kh^{\frac{\nu+1}{\nu}} \), that is,

\[
\Delta(h) = Kh^{(\nu+1)/\nu} + o(h^{(\nu+1)/\nu}) \quad \text{for some } K \neq 0 \text{ as } h \to 0 + .
\]

We see from (55) that \( F'_G \) is bounded. Hence \( F_G \) is Lipschitz and Theorem 1.6 can be applied here with \( r_R = \frac{\nu+1}{\nu} \) to obtain that, for any \( \nu > 0 \),

\[
(n + m)^{\frac{\nu}{\nu+2}} \int_{t_{n,m}^0}^{
\text{Example 4.7.} \quad \text{Let } F \text{ (resp. } G) \text{ be centered (resp. with mean } \mu > 0) \text{ normal distributions with common variance } \sigma^2. \text{ Let } f \text{ denote the density function of } F. \text{ Now, } F_G(t) = F(\nu^{-1}(t) + \mu), t \in [0, 1] \text{ and }

\[
F'_G(t) = \frac{f(F^{-1}(t) + \mu)}{f(F^{-1}(t))} = e^{-(2\mu F^{-1}(t) + \mu^2)/2\sigma^2} \to \infty, \quad \text{as } t \to 0 + .
\]

This implies that \( F_G \) is not Lipschitz in a neighbourhood of 0. However, we can use the fact that \( \ell^{t_0}_{n,m} = - \left( \int_0^n I(F_n^{-1}(t) < G^{-1}_m(t))dt - \int_0^n I(F_n^{-1}(t) > G^{-1}_m(t))dt \right) = - \left( \int_0^n I(F_n^{-1}(t) < G^{-1}_m(t))dt - \int_0^n I(F_n^{-1}(t) > G^{-1}_m(t))dt \right) \). Hence, \( \ell^{t_0}_{n,m} = - \ell^{t_0}_{m,n} \), where \( \ell^{t_0}_{m,n} \), as before, denotes the same statistic as \( \ell^{t_0}_{n,m} \), but exchanging the roles of \( F \) and \( G \). Now \( G_F(0) = 0 \) and 0 is an isolated regular contact point as in (8), with \( r_R = 1 \) and \( C_R = -1 \). Using Theorem 1.6 we conclude that

\[
(n + m)\ell^{t_0}_{n,m} \to \lambda(1 - \lambda) \int_0^\infty I_{\{1-(\lambda(S^2_i)_{\lambda>1})>0\}} dy = \lambda(1 - \lambda) \int_0^\infty I_{\{1-(\lambda(S^2_i)_{\lambda<1})<0\}} dy = 0,
\]

since, a.s., \( S^2_i > 0 \) for every \( i \geq 1 \). Thus the rate of convergence here is faster than \( (n + m)^{-1} \). \( \square \)

We obtain now some consequences of Theorem 1.7. If the extremal contact points have a non-null contribution to the limiting distribution, then it cannot be normal. We pay now attention to finding conditions under which \( \sqrt{n + m} \ell^{t_1}_{n,m} \), vanishes for \( t = 0, 1 \). We pay special attention to the rate \( \sqrt{n + m} \) because only this rate can result in a normal limit. Theorem 1.6 provides some answer to this problem, but we will give here simpler sufficient conditions. If the supports of \( F \) and \( G \) are bounded and

\[
\liminf |F^{-1}(t) - G^{-1}(t)| > 0 \text{ when } t \to 0 + \text{ or } t \to 1 -, \tag{56}
\]

then the convergence to zero of \( \ell^{0}_{n,m} \) and \( \ell^{1}_{n,m} \) can be dealt with as in Lemma 4.1.

Note that, in the case of non-bounded support, (56) does not exclude that 0 or 1 could be contact points (recall Example 4.7). For this case the following criterion on the tails can be useful to guarantee asymptotic negligibility of \( \ell^{0}_{n,m} \) and \( \ell^{1}_{n,m} \) in presence of inner contact points:

\[
\int_{(0,\varepsilon)\cup(1-\varepsilon, 1)} \left( \frac{\sqrt{f(1-t)}}{f(F^{-1}(t))} \right)^p dt < \infty \quad \text{and} \quad \int_{(0,\varepsilon)\cup(1-\varepsilon, 1)} \left( \frac{\sqrt{f(1-t)}}{g(G^{-1}(t))} \right)^p dt < \infty, \tag{57}
\]
for some $p > 1$ and $\varepsilon > 0$. In fact, let us assume that (57) and (56) hold and that, for instance, 0 is a contact point and that $\inf(F^{-1}(t) - G^{-1}(t)) > \delta > 0$ on $(0, \eta) \subset (0, \varepsilon)$. Noting that

$$(F^{-1} - G^{-1} > \delta) \cap (F^{-1}_n - G^{-1}_m \leq 0) \subset (|F^{-1}_n - F^{-1}| > \delta/2) \cup (|G^{-1}_m - G^{-1}| > \delta/2):$$

$$n^{1/2} \int_0^n I_{\{|F^{-1}_n(t) - G^{-1}_m(t)| \leq 0\}} dt \leq n^{1/2} \left( \int_0^n I_{\{|F^{-1}_n(t) - F^{-1}(t)| \geq \delta/2\}} dt + \int_0^n I_{\{|G^{-1}_m(t) - G^{-1}(t)| \geq \delta/2\}} dt \right) \leq \frac{n^{1/2} - \varepsilon^{1/2}}{2(\delta/2)^p} \left( \int_0^n |\sqrt{n}(F^{-1}_n(t) - F^{-1}(t))|^p dt + \int_0^n |\sqrt{n}(G^{-1}_m(t) - G^{-1}(t))|^p dt \right) \overset{p}{\to} 0,$$

where the last convergence follows from the fact that by (57) and Theorem 5.3, p. 46 in [8], the integrals in parentheses are stochastically bounded.

Now, assume that $F$ and $G$ have density functions, $f$ and $g$, and enough differentiability, and consider the function $h(t) = F_G(t) - t$ (used to obtain (45)). Then the set of contact points with intensity $k$ is

$$\Gamma_k := \left\{ t \in \Gamma : h^j(t) = 0, j = 0, \ldots, k-1 \text{ and } h^k(t) \neq 0 \right\}, \quad k \in \mathbb{N},$$

and the derivatives of $h$ and those of $f$ and $g$ are related as follows.

**Lemma 4.8.** If $t_0 \in \Gamma_k$ for some $k \geq 1$, and we denote $x_0 = F^{-1}(t_0)$, then,

$$h^k(t_0) = \begin{cases} 
\frac{f(x_0)}{g(x_0)} - 1, & \text{if } k = 1 \\
\frac{f^{k-1}(x_0) - g^{k-1}(x_0)}{f^k(x_0)} & \text{if } k > 1.
\end{cases}$$

The above considerations, combined to (45), (46) and (47) give the next version of Theorem 1.7.

**Theorem 4.9.** Assume that $F$ and $G$ have positive densities $f$ and $g$ on possibly unbounded intervals which are $k_0$ times continuously differentiable. Assume further that the set of contact points is finite with maximal intensity $k_0$ and that condition (56) holds. Suppose in addition that either the supports are bounded or that condition (57) is satisfied. Then, if $B_1$ and $B_2$ are independent Brownian bridges, and $n, m \to \infty$ with $\frac{n}{n+m} \to \lambda \in (0, 1)$,

(i) if $k_0 = 1$ and $x_i = F^{-1}(t_i)$,

$$(n + m)^{1/2}(\gamma(F_n, G_m) - \gamma(F, G)) \overset{w}{\to} \sum_{t_i \in \Gamma_1} \left( \frac{g(x_i)}{|f(x_i) - g(x_i)|} \frac{B_1(t_i)}{\sqrt{\lambda}} + \frac{f(x_i)}{|f(x_i) - g(x_i)|} \frac{B_2(t_i)}{\sqrt{1-\lambda}} \right),$$

(ii) if $k_0 \geq 3$ is odd

$$(n + m)^{1/2} (\gamma(F_n, G_m) - \gamma(F, G)) \overset{w}{\to} \sum_{t_i \in \Gamma_{k_0}} \left( \frac{k_0!}{h^{k_0}(t_i)} \right)^{1/k_0} \left( ((B_\lambda(t_i))^{1/k_0})^+ - ((B_\lambda(t_i))^{1/k_0})^- \right),$$
(iii) if $k_0$ is even

$$(n+m)^{1/2k_0} (\gamma(F_n,G_m) - \gamma(F,G)) \xrightarrow{w} \sum_{t_i \in \Gamma_{k_0}} \text{sgn}(h^{k_0})(t_i) 2\left(\frac{k_0!}{h^{k_0}(t_i)}\right)^{1/k_0} ((B_{\lambda}(t_i))^{\text{sgn}(h^{k_0})(t_i)})^{1/k_0}.$$

We see from Theorem 4.9 that asymptotic normality (arguably, the most useful case for statistical applications) holds, with the standard $\sqrt{n+m}$ rate, only when $F$ and $G$ have a finite number of ‘simple’ crossings. In all the other cases we get a slower rate and a nonnormal limit.

While Theorem 1.7 (hence, also Theorem 4.9) involves only the case when $\Gamma^*=\Gamma_F^*$ consists of regular contact points, the comments about virtual contact points between $F_G$ and the identity that led to (44) apply to the global analysis of $\gamma(F_n,G_m)$. As an important example, we consider the case when $F$ and $G$ are finitely supported. More precisely, let us assume $F$ and $G$ have a finite support $x_1 < x_2 < \cdots < x_k$, with probabilities $p_1, p_2, \ldots, p_k$ and $q_1, q_2, \ldots, q_k$, respectively, with $p_i + q_i > 0$ (although $p_i$ or $q_i$ could be null), $i = 1, \ldots, k$. We set $P_i := \sum_{j=1}^i p_j$ and $Q_i := \sum_{j=1}^i q_j$, $i = 1, \ldots, k - 1$. Then, $F_G(t) = P_i$ for $i \in (Q_{i-1}, Q_i]$. Hence, the only possible inner contact points are $P_i$, $Q_i$, $i = 1, \ldots, k - 1$ and all the possible contact points are either horizontal crossings ($P_i$ if $Q_{i-1} < P_i < Q_i$), vertical crossings ($Q_i$ if $Q_i < P_i < Q_{i+1}$), upper tangency points ($Q_i$ if $Q_{i-1} < Q_i < P_i$), and lower tangency points ($P_i$ if $P_{i-1} < P_i = Q_{i-1} < Q_i$), using the same terms as in the discussion following Proposition 4.3. Combining that discussion with Corollary 4.2 we obtain the following consequence.

**Theorem 4.10.** With the above notation, if $F$ and $G$ are finitely supported and $\mathcal{H}$, $\mathcal{V}$, $\mathcal{U}$ and $\mathcal{L}$ denote, respectively, the sets of horizontal and vertical crossings, upper and lower tangency points for $F$ and $G$. If, additionally, $B_1$ and $B_2$ are independent Brownian bridges then, assuming that $\frac{n}{n+m} \rightarrow \lambda \in (0,1)$,

$$\sqrt{n+m}(\gamma(F_n,G_m) - \gamma(F,G)) \xrightarrow{w} \sum_{t \in \mathcal{H}} \frac{B_1(t)}{\sqrt{\lambda}} - \sum_{t \in \mathcal{V}} \frac{B_2(t)}{\sqrt{1-\lambda}} + \sum_{t \in \mathcal{U}} \left( \frac{B_1(t)}{\sqrt{\lambda}} - \frac{B_2(t)}{\sqrt{1-\lambda}} \right)^+ - \sum_{t \in \mathcal{L}} \left( \frac{B_1(t)}{\sqrt{\lambda}} - \frac{B_2(t)}{\sqrt{1-\lambda}} \right)^-.$$

Similar to Theorem 1.5, we get a Gaussian limiting distribution only when all the contact points are crossing points (which, necessarily, have orders $r_L = r_R = 1$). In the case $F = G$ we have $Q_{i-1} < Q_i = P_i < P_{i+1}$ for all $i$, that is, every $P_i$ is an upper tangency point and Theorem 4.10 yields

$$\sqrt{n+m}(\gamma(F_n,G_m) \xrightarrow{w} \sum_{i=1}^{k-1} \left( \frac{B_1(P_i)}{\sqrt{\lambda}} - \frac{B_2(P_i)}{\sqrt{1-\lambda}} \right)^+. \quad (58)$$

Using the fact that $\sqrt{1-\lambda}B_1 - \sqrt{\lambda}B_2$ is a Brownian bridge, we can, equivalently, write (58) as

$$\sqrt{\frac{n+m}{n+m}}(\gamma(F_n,G_m) \xrightarrow{w} \sum_{i=1}^{k-1} (B_1(P_i))^+. \quad (58)$$

**Acknowledgements**

Research partially supported by FEDER, Spanish Ministerio de Economía y Competitividad, grants MTM2017-86061-C2-1, 2 and Junta de Castilla y León, grants VA005P17 and VA002G18.
Supplementary Material

Supplement to “The complex behaviour of Galton rank-order statistic"
This Appendix contains some properties of the $F_G$ transform, including a technical discussion on conditions which guarantee that $F_G$ is Lipschitz or locally Lipschitz. Moreover, it includes the strong approximation result that we have used in several proofs through the paper.

References

[1] Álvarez-Esteban, P.C.; del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2017). Models for the assessment of treatment improvement: the ideal and the feasible. *Statist. Sci.*, **32**, 469–485.
[2] Aly, Emad-Eldin A.A. (1986). Strong Approximations of the Q-Q Process. *J. Multiv. Analysis* **20**, 114–128.
[3] Aly, Emad-Eldin A.A.; Csörgo, M. and Horvath, L. (1987). P-P Plots, Rank Processes and Chernoff-Savage Theorems. In *New Perspectives in Theoretical and Applied Statistics* (M.L. Puri, J.P. Vilaplana and W. Wertz Eds.), 135–156. Wiley, New York
[4] del Barrio, E.; Cuesta-Albertos, J.A. and Matrán, C. (2021). Supplement to “The complex behaviour of Galton rank-order statistic”. The DOI will be added by the typesetter.
[5] Behnen, K. and Neuhaus, G. (1983). Galton tests as linear rank tests with estimated scores and its local asymptotic efficiency. *Ann. Statist.* **11**(2), 588–599.
[6] Bickel, P.J. and Hodges, J.L., Jr. (1967). The asymptotic theory of Galton’s test and a related simple estimate of location. *Ann. Math. Statist.* **38**, 73–89.
[7] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley.
[8] Bobkov, S. and Ledoux, M. (2016). One-dimensional empirical measures, order statistics and Kantorovich transport distances. *Memoirs Am. Math. Soc.* Vol.: 261, Number 1259.
[9] Chung, K. L., and Feller, W. (1949). On fluctuations in coin-tossing. *Proc. Nat. Acad. Sci. of USA*, **35**, 605–608.
[10] Csáki, E. and Vincze, I. (1961). On some problems connected with the Galton-test. *Publ. Math. Inst. Hungar. Acad. Sci.* **6**, 97–109
[11] Csörgo, M. and Horvath, L. (1993). *Weighted approximations in probability and statistics*. Wiley.
[12] Darwin, C. (1876). *The effect of Cross- and Self-fertilization in the Vegetable Kingdom*. John Murray.
[13] Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* **2**(2), 267–277.
[14] Feller, W. (1968). *An Introduction to Probability Theory and its Applications* Vol. I. Wiley.
[15] Gross, S., and Holland, P. W. (1968). The Distribution of Galton Statistic. *Ann. Math. Statist.*, **39**(6), 2114–2117.
[16] Hodges, J.L. (1955). Galton rank-order test. *Biometrika* **42**, 261–262.
[17] Lehmann, E.L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26**, 399–419.
[18] Lévy, P. (1939). Sur certains processus stochastiques homogènes. *Compositio Math.* **7**, 283–339.
[19] Reiss, R.D. (1989). *Approximate Distributions of Order Statistics With Applications to Nonparametric Statistics*. Springer.
[20] Shorack, J.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley and Sons, New York.
[21] Sparre-Andersen, E. (1953). On the fluctuations of sums of random variables. *Math. Scand.* **1**, 263–285.
[22] Zhuang, W.W., Hu, B.Y., and Chen, J. (2019). Semiparametric inference for the dominance index under the density ratio model. *Biometrika*, **106**, 1, 229–241