Bernoulli **11**(1), 2005, 131–189

Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances

EUSTASIO DEL BARRIO[1], EVARIST GINÉ[2] and FREDERIC UTZET[3]

[1]*Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, 47005 Valladolid, Spain*
[2]*Department of Mathematics, U-3009, University of Connecticut, Storrs CT 06269-3009, USA. E-mail: gine@math.uconn.edu*
[3]*Departament de Matemàtiques, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain*

Weighted L_2 functionals of the empirical quantile process appear as a component of many test statistics, in particular in tests of fit to location–scale families of distributions based on weighted Wasserstein distances. An essentially complete set of distributional limit theorems for the squared empirical quantile process integrated with respect to general weights is presented. The results rely on limit theorems for quadratic forms in exponential random variables, and the proofs use only simple asymptotic theory for probability distributions in $\mathbb{R}^n$. The limit theorems are then applied to determine the asymptotic distribution of the test statistics on which weighted Wasserstein tests are based. In particular, this paper contains an elementary derivation of the limit distribution of the Shapiro–Wilk test statistic under normality.

Keywords: distributional limit theorems; tests of fit to location–scale families; weighted L_2 norms of the quantile process; weighted Wasserstein distance

1. Introduction

Let $X, X_1, \ldots, X_n, \ldots$ be independent and identically distributed (i.i.d.) random variables with cumulative distribution function F, density f and quantile function $F^{-1}(t) := \inf\{y : F(y) \geqslant t\}$, $0 < t < 1$. For each $n \in \mathrm{N}$, let

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i), \qquad t \in \mathbb{R},$$

$$F_n^{-1}(t) = \inf\{y : F(y) \geqslant t\}, \qquad t \in (0, 1), \tag{1.1}$$

denote respectively the empirical distribution and quantile functions. The *empirical quantile process* (or quantile process for short), defined for each $n \in \mathrm{N}$ as

1350–7265 © 2005 ISI/BS

$$v_n(t) = \sqrt{n}\big(F_n^{-1}(t) - F^{-1}(t)\big), \qquad t \in (0, 1), \tag{1.2}$$

is the basic component of many interesting statistics and a large body of literature is devoted to it (see, for example, Shorack and Wellner 1986; Csörgő and Horváth 1993). A less general statistic that is also the main component of many statistics is the second moment with respect to Lebesgue measure, and in general with respect to any measure with density $w(t)$ on (0, 1), of the quantile process, $\|v_n\|_{2,w}^2 := \int_0^1 v_n^2(t)w(t)\mathrm{d}t$. This is the main component in statistics used, for example, in tests of fit based on the correlation coefficient (see Lockhart and Stephens 1998; and references therein), and in the related tests of fit based on Wasserstein distances (del Barrio *et al*. 1999a). These classes of tests contain some very important members, such as the Shapiro–Wilk test for normality.

One of the two goals of this paper is to give a complete description of convergence distribution of $\|v_n\|_{2,w}^2$ (and some variations). Although there are many important results in the literature on this subject (for example, Csörgő and Horváth 1988, 1993; de Wet and Venter 1972; Gregory 1977; Guttorp and Lockhart 1988; LaRiccia and Mason 1986; Mason 1984), these fall short of covering all the possibilities, and a comprehensive treatment is lacking and could be useful. The general picture given here is suggested by a heuristic description in Lockhart and Stephens (1998) together with an interesting example in del Barrio *et al*. (1999a). The second object is to apply the results obtained to testing the fit of empirical data to location–scale families of distributions by means of *weighted* Wasserstein distances, as suggested by the work of del Barrio *et al*. (1999a; 2000), de Wet (2000; 2002) and Csörgő (2002). Here we give an account of the content of this paper, together with a succint digression on each of these two topics.

1.1. The L_2 norm of the quantile process

It is well known that under certain standard conditions that allow to the general quantile process to be compared with the uniform, the quantile process $v_n(t)$ converges in law to the process $B_0(t) = B(t)/f(F^{-1}(t))$ in $\ell^\infty[a, b]$ for any $0 < a < b < 1$, where $B(t)$ is a Brownian bridge (Csörgő and Horváth 1993). The covariance of the limiting process is

$$\eta(s, t) = \frac{s \wedge t - st}{f(F^{-1}(t))f(F^{-1}(s))}. \tag{1.3}$$

We will see that, assuming additional conditions needed to handle the integrals on shrinking neighbourhoods of the end-points 0 and 1, we have the following four areas:

Case 1. If $\int_0^t \eta(t, t)w(t)\mathrm{d}t < \infty$ then $v_n(t)$ converges to $B_0(t)$ in law in $L_2(0, 1)$ and, in particular, $\|v_n\|_{2,w}^2 \to_d \|B_0\|_{2,w}^2$; note that the limit has a generalized chi-square distribution, namely, the distribution of $\sum \lambda_k Z_k^2$, where the variables Z_k are i.i.d. normal and λ_k are the eigenvalues of the covariance η.

Case 2. If $\int_0^1\int_0^1 \eta^2(s, t)w(s)w(t)\mathrm{d}s\,\mathrm{d}t < \infty$, but $\int_0^t \eta(t, t)w(t)\mathrm{d}t = \infty$, then

$$\|v_n\|_{2,w}^2 - \int_{1/n}^{1-1/n} \eta(t, t)w(t)\mathrm{d}t \to \int_0^1 \frac{B^2(t) - \mathrm{E}B^2(t)}{f^2(F^{-1}(t))} w(t)\mathrm{d}t,$$

where the last integral is defined in the natural limiting L_2 sense, and equals, in distribution, the probability law of $\sum \lambda_k(Z_k^2 - 1)$.

Case 3. If $\int_0^1 \int_0^1 \eta^2(s, t)w(s)w(t)\mathrm{d}s\,\mathrm{d}t = \infty$ but we are in the borderline case where $f(F^{-1}(t))/\sqrt{w(t)}$ is regularly varying with unit exponent at 0 and at 1, or is regularly varying at one of these two points and strictly of larger order at the other, then we have a normal limit,

$$\frac{1}{a_n}\left(\|v_n\|_{2,w}^2 - b_n\right) \to_d Z,$$

where Z is standard normal, $a_n^2 = \int_{1/n}^{1-1/n} \int_{1/n}^{1-1/n} \eta^2(s, t)w(s)w(t)\mathrm{d}s\,\mathrm{d}t$ and $b_n = \int_{1/\mathrm{n}}^{1-1\mathrm{n}} \eta(t, t)w(t)\mathrm{d}t$.

Case 4. If $f(F^{-1}(t))/\sqrt{w(t)}$ is regularly varying at 0 and at 1 with exponent larger than one (or is regularly varying at one end and of strictly larger order of magnitude at the other), then other limits may arise, which have to do with integrals of the centred squared partial sum process of independent exponential random variables. Here, in some subcases one must centre $\|v_n\|_{2,w}^2$ and in all of them one must divide by a sequence tending to infinity faster than in case 3.

These are all the possible limits for sufficiently regular densities. Case 1 was essentially known although perhaps not in its present generality (it can be deduced, for example, from Mason 1984); case 2 was known, at least implicitly, only for the normal distribution (de Wet and Venter 1972), whereas case 3 was known explicitly for the exponential, logistic and Gumbel distributions (McLaren and Lockhart 1987), and del Barrio *et al.* (2000) contains an example of case 4 (this example motivated us to consider this case).

The value of cases 3 and 4 for testing fit is limited since the limiting distribution only depends on the tails of the distribution (the middle part of the integral tends to zero when normalized by denominators that tend to infinity), and we will only consider them in the particular but important case $w(t) \equiv 1$.

All this is proved in a unified way and under very weak hypotheses. At least two methods of proof are available. A shorter, high-powered method would reduce the problem to a Gaussian one by means of special constructions – such as those of Csörgő *et al.* (1986) or Mason (1991) – and then work with the L_2 norms of integrals of the square of the Brownian bridge (for instance, by applying known results or using the method of moments). However, the results can easily be obtained from scratch using only elementary probability: one reduces the problem to the same problem for the uniform quantile process (at least in the first three cases) and then observes that the second moment of the empirical uniform quantile process is basically the square of the norm of a linear combination of independent exponential random variables with values in L_2, which can be treated using the central limit theorem in $\mathbb{R}^d$ plus Hilbert space approximation, at least in cases 1 and 2; case 4 is easy (but involved) whereas case 3 requires a martingale central limit theorem. This approach, which is the one we take, seems to have the advantage of requiring slightly less integrability than the approach by means of Gaussian approximations (see, for example, Csörgő 2002; Csörgő and Horváth 1988). Mason (1984) and Gregory (1977) pioneered different versions of the approach taken here.

In Section 2 we show how $\|v_n\|_{2,w}^2$ (almost) equals in distribution the norm of an L_2 linear combination of exponentials, in Section 3 we prove limit theorems for such linear combinations, and in Section 4 we prove the main results on convergence in law for the uniform and the general quantile process, as described, by just combining the results from the previous two sections.

1.2. Tests of fit based on weighted Wasserstein distances

Wasserstein tests of fit, introduced in del Barrio *et al.* (1999a; 2000), provide a powerful method for assessing fit of empirical data to a location–scale family of distributions. They can be described as follows. Let $\mathcal{H} = \{G_{\mu,\sigma} : G_{\mu,\sigma}(x) = G_0((x-\mu)/\sigma);\ \mu \in \mathbb{R},\ \sigma > 0\}$ be a location–scale family and assume that the first two moments of G_0 are 0 and 1, respectively. Given two probability measures on the real line with finite second moments and distribution functions F and G, respectively, $\mathcal{W}(F, G) = (\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 \mathrm{d}t)^{1/2}$ is the L_2 Wasserstein distance between them (see, for example, Bickel and Freedman 1981). Denote by $\mu(F)$ and $\sigma^2(F)$ the mean and the variance, respectively, of F, that is, $\mu(F) = \int_0^1 F^{-1}(t)\mathrm{d}t$ and $\sigma^2(F) = \int_0^1 (F^{-1}(t))^2 \mathrm{d}t - (\mu(F))^2$. It can be shown (see del Barrio *et al.* (1999a) that $\mathcal{W}^2(F, \mathcal{H}) := \inf_{H \in \mathcal{H}} \mathcal{W}^2(F, H) = \sigma^2(F) - (\int_0^1 F^{-1} G_0^{-1})^2$ and, from this, that

$$\frac{\mathcal{W}^2(F, \mathcal{H})}{\sigma^2(F)} = 1 - \frac{\left(\int_0^1 F^{-1}(t) G_0^{-1}(t)\mathrm{d}t\right)^2}{\sigma^2(F)}, \tag{1.4}$$

is a location- and scale-invariant measure of the discrepancy between F and $\mathcal{H}$. Its empirical version,

$$\mathcal{R}_n := \frac{\mathcal{W}^2(F_n, \mathcal{H})}{\sigma^2(F_n)} = 1 - \frac{\left(\int_0^1 F_n^{-1}(t) G_0^{-1}(t)\mathrm{d}t\right)^2}{\sigma^2(F_n)}, \qquad (1.5),$$

is the Wasserstein test statistic for $H_0 : F \in \mathcal{H}$, H_0 being rejected for large observed values of $\mathcal{R}_n$.

The Wasserstein test of normality turns out to be equivalent to the well-known Shapiro–Wilk test, sharing its good power properties. However, both are inefficient procedures for testing fit to location–scale families that, like the exponential family, for instance, have heavier tails (just as with tests of fit based on the correlation coefficient: see Lockhart and Stephens 1998, Section 5). The null asymptotics of $\mathcal{R}_n$ provides a good insight into the cause of this inefficiency. To study this asymptotic distribution under H_0 we can assume $F = G_0$ (by the location and scale invariance of $\mathcal{R}_n$) and denote the empirical quantile process as $v_n(t) = \sqrt{n}(F_n^{-1}(t) - F^{-1}(t))$, to obtain (see del Barrio *et al.* 1999a) that

$$n\mathcal{R}_n = \frac{1}{\sigma^2(F_n)}\left[\int_0^1 v_n^2(t)\mathrm{d}t - \left(\int_0^1 v_n(t)\mathrm{d}t\right)^2 - \left(\int_0^1 v_n(t)F^{-1}(t)\mathrm{d}t\right)^2\right]$$

$$= \frac{1}{\sigma^2(F_n)}\int_0^1 (v_n(t) - \langle v_n, 1\rangle 1 - \langle v_n, F^{-1}\rangle F^{-1}(t))^2 \mathrm{d}t$$

$$= \frac{1}{\sigma^2(F_n)}\int_0^1 \hat{v}_n^2(t)\mathrm{d}t, \tag{1.6}$$

where $\langle f, g\rangle = \int_0^1 f\cdot g$ and $\hat{v}_n = v_n - \langle v_n, 1\rangle 1 - \langle v_n, F^{-1}\rangle F^{-1}$. It is shown in del Barrio *et al*. (1999a) that under normality there exist constants a_n such that $n\mathcal{R}_n - a_n$ converges in law to a non-degenerate distribution. More precisely, if Φ (ϕ) denote the standard normal distribution (density) function and B is a Brownian bridge, then

$$n\mathcal{R}_n - a_n \underset{d}{\rightarrow} \int_0^1 \frac{B^2(t) - \mathrm{E}B^2(t)}{\phi^2(\Phi^{-1}(t))}\mathrm{d}t - \left(\int_0^1 \frac{B(t)}{\phi(\Phi^{-1}(t))}\mathrm{d}t\right)^2 - \left(\int_0^1 \frac{B(t)\Phi^{-1}(t)}{\phi(\Phi^{-1}(t))}\mathrm{d}t\right)^2. \tag{1.7}$$

In the exponential case $n\mathcal{R}_n$ is not shift tight, but it can be shown that, for some constants a_n, $(n/\sqrt{\log n})\mathcal{R}_n - a_n$ is asymptotically normal. However, if we fix $\delta \in (0, \frac{1}{2})$ then

$$\frac{1}{\sqrt{\log n}}\int_\delta^{1-\delta} (\hat{v}_n(t))^2 \mathrm{d}t \underset{\mathrm{Pr}}{\rightarrow} 0,$$

hence the asymptotic distribution of $\mathcal{R}_n$ depends only on the tails of F: the Wasserstein exponentiality test cannot detect alternatives that have approximately exponential tails.

As a possible remedy to this inefficiency, de Wet (2000; 2002) and Csörgő (2002) proposed replacing the Wasserstein distance $\mathcal{W}$ by a weighted version $\mathcal{W}_w(F, G) := (\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 w(t)\mathrm{d}t)^{1/2}$, for some positive measurable function w, and the test statistic $\mathcal{R}_n$ by

$$\mathcal{R}_n^w = \frac{\mathcal{W}_w^2(F_n, \mathcal{H})}{\sigma_w^2(F_n)},$$

where, here and in what follows, we set

$$\mu_w(F) = \int_0^1 F^{-1}(t)w(t)\mathrm{d}t, \qquad \sigma_w^2(F) = \int_0^1 (F^{-1}(t))^2 w(t)\mathrm{d}t - (\mu_w(F))^2.$$

$\mathcal{R}_n^w$ is location- and scale-invariant, hence its null distribution can be studied assuming, as above, that $F = G_0$. Under the assumptions

$$\int_0^1 w(t)\mathrm{d}t = 1, \tag{1.8}$$

$$\int_0^1 G_0^{-1}(t)w(t)\mathrm{d}t = 0 \tag{1.9}$$

and

$$\int_0^1 (G_0^{-1}(t))^2 w(t)\mathrm{d}t = 1, \tag{1.10}$$

we can mimic, step by step, the computations leading to (1.6) and obtain that

$$\begin{aligned} n\mathcal{R}_n^w &= \frac{1}{\sigma_w^2(F_n)}\left[\int_0^1 v_n^2(t)w(t)\mathrm{d}t - \left(\int_0^1 v_n(t)w(t)\mathrm{d}t\right)^2 - \left(\int_0^1 v_n(t)F^{-1}(t)w(t)\mathrm{d}t\right)^2\right] \\ &= \frac{1}{\sigma_w^2(F_n)}\int_0^1 (v_n(t) - \langle v_n, 1\rangle_w 1 - \langle v_n, F^{-1}\rangle_w F^{-1}(t))^2 w(t)\mathrm{d}t \\ &= \frac{1}{\sigma_w^2(F_n)}\int_0^1 \hat{v}_n^2(t)w(t)\mathrm{d}t, \end{aligned} \tag{1.11}$$

where now $\langle f, g\rangle_w = \int_0^1 (f\cdot g)w$ and $\hat{v}_n = v_n - \langle v_n, 1\rangle_w 1 - \langle v_n, F^{-1}\rangle_w F^{-1}$. Thus, the asymptotic distribution of $n\mathcal{R}_n^w$ under the null hypothesis can be obtained through the analysis of weighted L_2 functionals of the quantile process v_n. This is done in Section 5, where some examples are also presented. When specialized to the normal case, one obtains an elementary derivation of the asymptotic distribution of the Shapiro–Wilk test statistic under the null hypothesis, in the spirit of del Barrio (2001), though of course the present results apply to many more distributions.

2. Reduction of the L_2-norm of the quantile process to the norm of a linear combination of exonential variables with L_2 non-random coefficients

Let U_i, $i \in \mathbb{N}$, be i.i.d. uniform (0,1) random variables; for each n, let $G_n(t)$ be the empirical cdf associated with $U_1, \ldots, U_n$, let $G_n^{-1}(t)$ be the quantile function and let $u_n(t)$ be the associated uniform quantile process, that is,

$$u_n(t) = \sqrt{n}(G_n^{-1}(t) - t), \qquad 0 < t < 1. \tag{2.1}$$

If $\{\xi_n\}_{n=1}^{\infty}$ are i.i.d. random variables with common exponential distribution of mean 1 and $S_n = \xi_1 + \ldots + \xi_n$, then the well-known distributional identity

$$(U_{n:1}, \ldots, U_{n:n}) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}}\right)$$

allows us to rewrite $G_n^{-1}(t)$ as S_j/S_{n+1} if $(j-1)/n < t \leqslant j/n$ and, consequently,

$$u_n(t) \stackrel{d}{=} \frac{n}{S_{n+1}}\sum_{j=1}^{n+1} Z_{n,j}(t), \tag{2.2}$$

where

$$Z_{n,j}(t) = n^{-1/2} a_{n,j}(t)\xi_j, \qquad a_{n,j}(t) = (1-t)I_{\{j-1<nt\}} - tI_{\{j-1\geqslant nt\}}. \tag{2.3}$$

So if

$$L_n := \int_{1/n}^{1-1/n} \left(\frac{u_n(t)}{g(t)}\right)^2 \mathrm{d}t \tag{2.4}$$

for some weight function g non-vanishing on (0, 1), then, with $\|\cdot\|_2$ denoting the L_2 norm with respect to Lebesgue measure on the unit interval,

$$L_n \stackrel{d}{=} \left(\frac{n}{S_{n+1}}\right)^2 \left\| \sum_{i=1}^{n+1} c_{n,i}\xi_i \right\|_2^2 \tag{2.4$'$}$$

for certain functions $c_{n,i}(t)$ which we assume in $L_2(0, 1)$ (in the case of (2.4′), but not always below, $c_{n,i} = n^{-1/2}a_{n,i}(t)I_{[1/n,1-1/n]}(t)/g(t)$). By the law of large numbers, weak convergence of the statistic $a_nL_n - b_n$ then reduces to weak convergence of

$$a_n \left\| \sum_{i=1}^{n+1} c_{n,i}\xi_i \right\|_2^2 - b_n\left(\frac{S_{n+1}}{n}\right)^2,$$

and the second variable is almost a constant if b_n does not grow too fast.

Since the $a_{n,j}(t)$ have a relatively complicated expression, it is convenient to isolate here as a lemma some estimates for $a_{n,j}(t)$ to be used below. Let frac(·) denote the fractional part of a number. It is also convenient to introduce the notation $f \otimes g$ for the function of two variables $f(x)g(y)$. We recall that the map $(f, g) \mapsto f \otimes g$ is a continuous bilinear map between $L_2(0, 1) \times L_2(0, 1)$ and $L_2((0, 1) \times (0, 1))$ and that $\langle f_1 \otimes g_1, f_2 \otimes g_2\rangle = \langle f_1, g_1\rangle\langle f_2, g_2\rangle$.

Lemma 2.1. *Set* $\tilde{m}_n = \sum_{j=1}^{n+1} a_{n,j}$ *and* $\tilde{K}_n = \sum_{j=1}^{n+1} a_{n,j} \otimes a_{n,j}$. *Then, for every* $0 < s, t < 1$,

(i) $\tilde{m}_n(t) = \mathrm{frac}(n(1-t)) - t$; *also, if* $1/n \leqslant t \leqslant 1 - 1/n$ *then*

$$|\tilde{m}_n(t)| \leqslant 1 \leqslant \frac{n}{n-1} nt(1-t);$$

(ii) if $s \leqslant t$, *then* $\tilde{K}_n(s, t) = [n(1-t)]s + \mathrm{frac}(n(1-s))(1-t) + st$;

(iii) $(1/n)\tilde{K}_n(s, t) \to s \wedge t - st$; *further, if* $1/n \leqslant s, t \leqslant 1 - 1/n$, *then*

$$\frac{1}{2}(s\wedge t - st) \leqslant \frac{1}{n}\tilde{K}_n(s, t) \leqslant \frac{1}{n}\sum_{j=1}^{n+1} |a_{n,j}(s)a_{n,j}(t)| \leqslant 3(s \wedge t - st).$$

Proof *(sketch).* To prove (i), fix t and observe that each term in $\sum_{j=1}^{n+1} a_{n,j}(t)$ equals either $1-t$ (the first $n - [n(1-t)]$ terms) or $-t$ (the remaining $[n(1-t)] + 1$ terms). Hence

$$m_n(t) = (n - [n(1-t)])(1-t) - ([n(1-t)] + 1)t = \mathrm{frac}(n(1-t)) - t.$$

The identity in (ii) can be proved in a similar way. Fix $s \leqslant t$. In the corresponding sum for $\tilde{K}_n(s, t)$ there are three types of summands: the first $n - [n(1-s)]$, each equal to $(1-s)(1-t)$; the next $[n(1-s)] - [n(1-t)]$, each equal to $-s(1-t)$; and the remaining $[n(1-t)] + 1$, each equal to st. This gives (ii) and the right-hand inequality in (iii). The left-hand inequality in (iii) is a trivial consequence of (ii). □

Here is a simple but useful observation about linear combinations of exponential variables with coefficients in L_2:

Lemma 2.2. *Let* $Y(t) = \sum_{k=1}^{n} c_k(t)\xi_k$ *for some* $n \in \mathbb{N}$ *and* $c_k \in L_2(0, 1)$, *and where the variables* ξ_k *are independantly exponentially distributed with parameter 1. Then there exists an absolute constant* $C < \infty$ *such that*

$$\mathrm{E}\|Y\|_2^4 \leqslant C\Big(\mathrm{E}\|Y\|_2^2\Big)^2. \tag{2.5}$$

Proof. By convexity,

$$\mathrm{E}\|Y\|_2^4 \leqslant 8\mathrm{E}\Bigg\|\sum_k c_k(t)(\xi_k - 1)\Bigg\|_2^4 + 8\Bigg\|\sum_k c_k(t)\Bigg\|_2^4,$$

where, letting $c_{i,j} := \int_0^1 c_i(t)c_j(t)\mathrm{d}t$, the last summand is just

$$\Bigg\|\sum_k c_k(t)\Bigg\|_2^4 = \left(\int_0^1\left(\sum_k c_k(t)\right)^2 \mathrm{d}t\right)^2 = \left(\sum_{i,j} c_{i,j}\right)^2.$$

To estimate the first summand, we use symmetrization, followed by randomization by an independent Rademacher sequence $\{\varepsilon_n\}$ (these variables are independent, independent of $\{\xi_i\}$, symmetric and take only the values 1 and -1), and Khinchine's inequality (see, for example, de la Peña and Giné 1999, p. 16), to obtain, letting E_ε denote integration only with respect to the Rademacher variables,

$$\begin{aligned}
\mathrm{E}\left\|\sum_k c_k(t)(\xi_k-1)\right\|_2^4 &\leqslant 2^4\mathrm{E}\left\|\sum_k c_k(t)\varepsilon_k(\xi_k-1)\right\|_2^4 \\
&\leqslant 9\cdot 2^4\mathrm{E}\left(\mathrm{E}_\varepsilon\left\|\sum_k c_k(t)\varepsilon_k(\xi_k-1)\right\|_2^2\right)^2 \\
&= 9\cdot 2^4\mathrm{E}\left(\int_0^1\sum_k c_k^2(t)(\xi_k-1)^2\mathrm{d}t\right)^2 \\
&= 9\cdot 2^4\mathrm{E}\left(\sum_k c_{k,k}(\xi_k-1)^2\right)^2 \\
&= 9\cdot 2^4\left(9\sum_k c_{k,k}^2+\sum_{i\neq j}c_{i,i}c_{j,j}\right) \\
&\leqslant 6^4\left(\sum_k c_{k,k}\right)^2. \qquad (2.6)
\end{aligned}$$

Collecting the above estimates, we obtain inequality (2.5) with $C = 3^4\cdot 2^7$. □

Next we consider the general quantile process. Let F be a twice differentiable distribution function such that $f := F'$ is non-vanishing on $\operatorname{supp} F := \{F \neq 0, 1\}^\circ := (a_F, b_F)$ and

$$r := \sup_{0<t<1}\frac{t(1-t)\left|f'(F^{-1}(t))\right|}{f^2(F^{-1}(t))} < \infty, \tag{2.7}$$

where $F^{-1}(t)$ is the corresponding quantile function. Condition (2.7), from Csörgő and Révész (1978), is a natural condition to have if we wish to relate general and uniform quantile processes: see their Lemma 1.1, Chapter 6, and comments after its proof. Since we are considering only distributional results, there is no loss of generality in taking $X_i = F^{-1}(U_i)$, where U_i are i.i.d. uniform on [0, 1]. In this case,

$$F_n^{-1}(t) = F^{-1}\big(G_n^{-1}(t)\big), \tag{2.8}$$

where G_n^{-1} is the quantile function corrresponding to the uniform variables $U_1, \ldots, U_n$. We continue denoting by v_n the quantile processes associated with the sequence $X_i = F^{-1}(U_i)$,

$$v_n(t) := \sqrt{n}\big(F_n^{-1}(t) - F^{-1}(t)\big) = \sqrt{n}\big(F^{-1}(G_n^{-1}(t)) - F^{-1}(t)\big), \qquad n\in\mathbb{N}. \tag{2.9}$$

Our aim is to relate the (weighted) L_2 norms of v_n and $u_n/f(F^{-1})$. Let w be a non-negative measurable function on (0, 1) and denote by $\|\cdot\|_{2,w,n}$ and $\langle\cdot,\cdot\rangle_{w,n}$ respectively the norm and the inner product in the space $L_2((1/n, 1-1/n), w(t)\mathrm{d}t)$.

Lemma 2.3 *Let F be a distribution function which is twice differentiable on its open support (a_F, b_F), with $f(x) := F'(x) > 0$ for all $a_F < x < b_F$, and which satisfies condition* (2.7). *Assume further that w is a non-negative measurable function such that*

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \int_{1/n}^{1-1/n} \frac{t^{1/2}(1-t)^{1/2}}{f^2(F^{-1}(t))} w(t)\mathrm{d}t = 0. \tag{2.10}$$

Then, if u_n is the uniform quantile process and υ_n is the quantile process defined by (2.8) *and* (2.9),

$$\|\upsilon_n\|^2_{2,w,n} - \left\| \frac{u_n}{f(F^{-1})} \right\|^2_{2,w,n} \to 0 \quad \textit{and} \quad \left\| \upsilon_n - \frac{u_n}{f(F^{-1})} \right\|_{2,w,n} \to 0 \tag{2.11}$$

in probability.

Proof. υ_n and u_n are related by the limited Taylor expansion

$$\begin{aligned}\upsilon_n(t) &= \sqrt{n}\big(F^{-1}\big(G_n^{-1}(t)\big) - F^{-1}(t)\big) \\ &= \frac{\sqrt{n}\big(G_n^{-1}(t) - t\big)}{f(F^{-1}(t))} + \frac{1}{2\sqrt{n}} n\big(G_n^{-1}(t) - t\big)^2 \frac{f'(F^{-1}(\xi))}{f^3(F^{-1}(\xi))} \\ &= \frac{u_n(t)}{f(F^{-1}(t))} + \frac{1}{2\sqrt{n}} \frac{f'(F^{-1}(\xi))}{f^3(F^{-1}(\xi))} u_n^2(t)\end{aligned} \tag{2.12}$$

for some ξ between t and $G_n^{-1}(t)$. The object is to estimate, using the Taylor development (2.12), the terms in the difference

$$\|\upsilon_n\|^2_{2,w,n} - \left\| \frac{u_n}{f(F^{-1})} \right\|^2_{2,w,n} = \left\| \upsilon_n - \frac{u_n}{f(F^{-1})} \right\|^2_{2,w,n} + 2\left\langle \upsilon_n - \frac{u_n}{f(F^{-1})}, \frac{u_n}{f(F^{-1})} \right\rangle_{w,n}. \tag{2.13}$$

Since, by (2.12),

$$\begin{aligned}&\left(\upsilon_n(t) - \frac{u_n(t)}{f(F^{-1}(t))}\right)^2 \\ &\quad = \frac{1}{4n} \cdot \frac{u_n^4(t)}{f^2(F^{-1}(t)) t^2 (1-t)^2} \cdot \left(\frac{f'(F^{-1}(\xi))\xi(1-\xi)}{f^2(F^{-1}(\xi))} \cdot \frac{t(1-t)}{\xi(1-\xi)} \cdot \frac{f(F^{-1}(t))}{f(F^{-1}(\xi))} \right)^2\end{aligned} \tag{2.14}$$

and

$$\begin{aligned}&\left(\upsilon_n(t) - \frac{u_n(t)}{f(F^{-1}(t))}\right) \frac{u_n(t)}{f(F^{-1}(t))} \\ &\quad = \frac{1}{2\sqrt{n}} \cdot \frac{u_n^3(t)}{f^2(F^{-1}(t)) t(1-t)} \cdot \frac{t(1-t)}{\xi(1-\xi)} \cdot \frac{f(F^{-1}(t))}{f(F^{-1}(\xi))} \cdot \frac{\xi(1-\xi) f'(F^{-1}(\xi))}{f^2(F^{-1}(\xi))},\end{aligned} \tag{2.15}$$

the following bounds will be useful. First we observe that, as shown, for example, in Csörgő and Horváth (1993, Lemma 6.1.1, p. 369), condition (2.7) implies that, for all $t_1, t_2 \in (0, 1)$,

$$\frac{f(F^{-1}(t_1))}{f(F^{-1}(t_2))} \leqslant \left(\frac{t_1 \vee t_2}{t_1 \wedge t_2} \cdot \frac{1 - t_1 \wedge t_2}{1 - t_1 \vee t_2}\right)^r,$$

so that, for ξ between t and $G_n^{-1}(t)$,

$$\frac{f(F^{-1}(t))}{f(F^{-1}(\xi))} \leqslant \left(\frac{t}{G_n^{-1}(t)} \cdot \frac{1 - G_n^{-1}(t)}{1 - t}\right)^r \vee \left(\frac{G_n^{-1}(t)}{t} \cdot \frac{1 - t}{1 - G_n^{-1}(t)}\right)^r.$$

Likewise,

$$\frac{t(1-t)}{\xi(1-\xi)} \leqslant \frac{t}{G_n^{-1}(t)} \vee \frac{1-t}{1 - G_n^{-1}(t)}.$$

Moreover, using the representation of uniform quantiles as sums S_n of exponential random variables (as above), we have

$$\frac{1-t}{1-G_n^{-1}(t)} \stackrel{d}{=} \frac{t}{G_n^{-1}(t)} \sim_d \frac{S_{n+1}}{n} \cdot \frac{nt}{S_{[nt]}}.$$

So, by the law of large numbers,

$$\sup_{n^{-1} \leqslant t \leqslant 1-n^{-1}} \frac{t(1-t)}{\xi(1-\xi)} = O_P(1) \tag{2.16}$$

and

$$\sup_{n^{-1} \leqslant t \leqslant 1-n^{-1}} \frac{f(F^{-1}(t))}{f(F^{-1}(\xi))} = O_P(1). \tag{2.17}$$

Finally, using first Lemma 2.2 and then parts (i) and (iii) of Lemma 2.1, we obtain that there is a finite constant C independent of n such that

$$\begin{aligned}
\mathrm{E}\left(\frac{S_{n+1}}{n} u_n(t)\right)^4 &= \frac{1}{n^2} \mathrm{E}\left(\sum_{j=1}^{n+1} a_{n,j}(t)\xi_j\right)^4 \leqslant C \frac{1}{n^2}\left(\mathrm{E}\left(\sum_{j=1}^{n+1} a_{n,j}(t)\xi_j\right)^2\right)^2 \\
&= \frac{C}{n^2}\left(\sum_{j=1}^{n+1} a_{n,j}^2(t) + \left(\sum_{j=1}^{n+1} a_{n,j}(t)\right)^2\right)^2 \\
&\leqslant \frac{C}{n^2}\left(\tilde{K}_n(t, t) + |\tilde{m}_n(t)|\right)^2 \leqslant 25Ct^2(1-t)^2,
\end{aligned} \tag{2.18}$$

and therefore also

$$\mathrm{E}\left|\frac{S_{n+1}}{n} u_n(t)\right|^3 \leqslant \tilde{C}[t(1-t)]^{3/2}, \tag{2.19}$$

for some finite constant $\tilde{C}$.

Combining the estimates (2.16)–(2.19) and the bound in (2.7) with the identities in (2.14) and (2.15), we obtain

$$\left\| v_n - \frac{u_n}{f(F^{-1})} \right\|_{2,w,n}^2 = O_P(1) \times \frac{1}{n} \int_{1/n}^{1-1/n} \frac{w(t)}{f^2(F^{-1}(t))} \mathrm{d}t \tag{2.20}$$

and

$$\left\langle v_n - \frac{u_n}{f(F^{-1})}, \frac{u_n}{f(F^{-1})} \right\rangle_{w,n} = O_P(1) \times \frac{1}{\sqrt{n}} \int_{1/n}^{1-1/n} \frac{t^{1/2}(1-t)^{1/2}}{f^2(F^{-1}(t))} w(t) \mathrm{d}t. \tag{2.21}$$

Since, by dominated convergence, if the integral in (2.21) tends to zero (which it does by hypothesis (2.10)), then so does the integral in (2.20), combining these estimates with the identity (2.13), the lemma follows. □

The above lemma should be compared with Theorem 6.2.3 in Csörgő and Horváth (1993) for $p = 2$.

In fact, as mentioned in the Introduction, we are interested not in $\|v_n\|_{2,w,n}$ but rather in $\|v_n\|_{2,w}$, where $\|\cdot\|_{2,w}$ denotes the L_2 norm with respect to the measure $w(t)\mathrm{d}t$ over the whole interval (0, 1). So, we must deal next with $\int_0^{1/n} v_n^2(t)w(t)\mathrm{d}t$ and $\int_{1-1/n}^1 v_n^2(t)w(t)\mathrm{d}t$. To do so we impose conditions which are related to but weaker than the usual von Mises conditions on domains of attraction (see Parzen 1979; Schuster 1984).

Lemma 2.4. *Let F be a distribution function which is twice differentiable on its open support (a_F, b_F), with $f(x) := F'(x) > 0$ for all $a_F < x < b_F$. Assume that F satisfies condition* (2.7)*, that*

$$\text{either } a_F > -\infty \text{ or } \liminf_{x \to 0+} \frac{|f'(F^{-1}(x))|x}{f^2(F^{-1}(x))} > 0, \tag{2.22}$$

and that

$$\text{either } b_F < \infty \text{ or } \liminf_{x \to 0+} \frac{|f'(F^{-1}(1-x))|x}{f^2(F^{-1}(1-x))} > 0. \tag{2.23}$$

Assume further that w is a bounded non-negative measurable function such that

$$\lim_{x \to 0+} \frac{x\int_0^x w(t)\mathrm{d}t}{f^2(F^{-1}(x))} = 0, \qquad \lim_{x \to 0+} \frac{x\int_{1-x}^1 w(t)\mathrm{d}t}{f^2(F^{-1}(1-x))} = 0. \tag{2.24}$$

Then

$$\|v_n\|_{2,w}^2 - \|v_n\|_{2,w,n}^2 \to 0 \tag{2.25}$$

in probability.

Proof. We will only consider the upper end of the difference in (2.25) since the lower end can be dealt with in the same way. Set $U_{(n)} := \max_{i \leqslant n} U_i$, where the U_i are, as before, i.i.d. random variables uniform on (0, 1). We have

$$\int_{1-1/n}^{1} v_n^2(t)w(t)\mathrm{d}t = n\int_{1-1/n}^{1} \big(F^{-1}(U_{(n)}) - F^{-1}(t)\big)^2 w(t)\mathrm{d}t$$

$$\leqslant 2n\int_{1-1/n}^{1} \big(F^{-1}(U_{(n)}) - F^{-1}(1-1/n)\big)^2 w(t)\mathrm{d}t$$

$$+ 2n\int_{1-1/n}^{1} \big(F^{-1}(1-1/n) - F^{-1}(t)\big)^2 w(t)\mathrm{d}t. \tag{2.26}$$

For some ξ between $U_{(n)}$ and $1 - 1/n$ we have, by the mean value theorem,

$$n\big(F^{-1}(U_{(n)}) - F^{-1}(t)\big)^2 = \frac{1}{nf^2(F^{-1}(1-1/n))} n^2(U_{(n)} - 1 + 1/n)^2 \left(\frac{f(F^{-1}(1-1/n))}{f(F^{-1}(\xi))}\right)^2.$$

By the well-known limit theorem for $U_{(n)}$ (see Leadbetter *et al.* 1983, p. 23), $n^2(U_{(n)} - 1 + 1/n)^2 = O_P(1)$, and, by condition (2.7), as in the proof of Lemma 2.3,

$$\frac{f(F^{-1}(1-1/n))}{f(F^{-1}(\xi))} \leqslant \left(\frac{1-1/n}{U_{(n)}} \cdot n(1 - U_{(n)})\right)^r \vee \left(\frac{U_{(n)}}{1-1/n} \cdot \frac{1}{n(1-U_{(n)})}\right)^r,$$

which is $O_P(1)$ because $U_{(n)} \to 1$ in probablity and the limit in distribution of $n(U_{(n)} - 1)$ is a non-vanishing random variable on $(-\infty, 0]$. This and condition (2.24) imply that

$$\lim_{n\to\infty} 2n\int_{1-1/n}^{1} \big(F^{-1}(U_{(n)}) - F^{-1}(1-1/n)\big)^2 w(t)\mathrm{d}t = 0$$

in probability. Regarding the second summand on the right-hand side of (2.26), it is obvious that its limit is zero if $b_F < \infty$, since then F^{-1} is continuous at $t = 1$ (recall w is bounded). If $b_F = \infty$ and (2.23) holds, we have, applying L'Hôpital's rule twice and momentarily assuming that w is continuous near 0 and 1 for the second application of the rule,

$$\lim_{x\to 0+} \frac{1}{x}\int_{1-x}^{1} (F^{-1}(1-x) - F^{-1}(t))^2 w(t)\mathrm{d}t = \lim_{x\to 0+} -\frac{2\int_{1-x}^{1}(F^{-1}(1-x) - F^{-1}(t))w(t)\mathrm{d}t}{f(F^{-1}(1-x))}$$

$$= \lim_{x\to 0+} -\frac{2\int_{1-x}^{1} w(t)\mathrm{d}t}{f'(F^{-1}(1-x))}$$

$$= \frac{-2x\int_{1-x}^{1} w(t)\mathrm{d}t}{f^2(F^{-1}(1-x))}\left(\frac{xf'(F^{-1}(1-x))}{f^2(F^{-1}(1-x))}\right)^{-1} \to 0$$

by (2.23) and (2.24). If w is not continuous near 0 and 1, then this limit still holds by the argument at the end of the proof of Proposition 4.3 in del Barrio *et al.* (1999b). Combining these limits with inequality (2.26) proves the lemma. □

As a consequence of these two lemmas we have:

Proposition 2.5. *Let F be a distribution function which is twice differentiable on its open support (a_F, b_F), with $f(x) := F'(x) > 0$ for all $a_F < x < b_F$. Assume that F satisfies conditions* (2.7), (2.22) *and* (2.23). *Let w be a bounded non-negative measurable function for which the limits* (2.24) *hold. Assume further that* (2.10) *holds. Then*

$$\|v_n\|_{2,w}^2 - \left\| \frac{u_n}{f(F^{-1})} \right\|_{2,w,n}^2 \to 0 \tag{2.27}$$

in probability. If, moreover, $h \in L_2(w(t)\mathrm{d}t)$ and the sequence $\{\langle u_n/f(F^{-1}), h\rangle_{w,n}\}$ is stochastically bounded, then

$$\langle v_n, h\rangle_w^2 - \left\langle \frac{u_n}{f(F^{-1})}, h \right\rangle_{w,n}^2 \to 0 \tag{2.28}$$

in probability.

Proof. The conclusion (2.27) is a direct consequence of Lemmas 2.3 and 2.4. The limit (2.28) follows from the same two lemmas, stochastic boundedness of $\{\langle u_n/f(F^{-1}), h\rangle_{w,n}\}$, Hölder's inequality and the identities

$$\langle v_n, h\rangle_{w,n}^2 - \langle \frac{u_n}{f(F^{-1})}, h\rangle_{w,n}^2$$
$$= \left\langle v_n - \frac{u_n}{f(F^{-1})}, h \right\rangle_{w,n} \left(\left\langle v_n - \frac{u_n}{f(F^{-1})}, h \right\rangle_{w,n} + 2\left\langle \frac{u_n}{f(F^{-1})}, h \right\rangle_{w,n} \right)$$

and

$$\langle v_n, h\rangle_w^2 - \langle v_n, h\rangle_{w,n}^2 = \left(\int_{(0,1/n]\cup[1-1/n,1)} v_n h w \right) \left(\int_{(0,1/n]\cup[1-1/n,1)} v_n h w + 2\langle v_n, h\rangle_{w,n} \right).$$

(Note that, by the first identity, $\{\langle v_n, h\rangle_{w,n}\}$ is stochastically bounded.) □

Obviously, combining this proposition with (2.3) and (2.4′) for $g = f(F^{-1})$, reduces convergence in distribution of $\|v_n\|_{2,w}^2$ and of $\|\hat{v}_n\|_{2,w}^2$ to convergence in distribution of L_n, which is a function of exponential random variables that will be relatively easy to handle. The conditions under which this has been established here are weaker than those usually found in the literature.

3. Convergence in law of L_2 linear combinations of exponential random variables and shift convergence of their norms

We should point out that the results that follow do not require the variables ξ_i to be exponential, but only to be integrable enough; however, we stay with exponential variables, which is what we need. In this section, the functions $c_{n,i}$ in the expression (2.4′) for L_n are allowed to be *arbitrary* functions in $L_2(0, 1)$. Given a triangular array $c_{n,i}$, $i \leqslant n$, $n \in \mathbb{N}$, of functions in $L_2(0, 1)$, we set

$$Y_n(t) := \sum_{i=1}^{n} c_{n,i}\xi_i, \qquad Z_n(t) = \left(\frac{n-1}{S_n}\right) Y_n(t), \qquad t \in [0, 1]. \tag{3.1}$$

Define $c_{n,i,j} = c_{i,j}$ as

$$c_{n,i,j} = c_{i,j} = \int_0^1 c_{n,i}(t)c_{n,j}(t)\mathrm{d}t := \langle c_{n,i}, c_{n,j}\rangle, \qquad i, j = 1, \ldots, n, \ n \in \mathbb{N}. \tag{3.2}$$

It will also be convenient to introduce the functions

$$K_n(s, t) = \sum_{i=1}^{n} c_{n,i}(s)c_{n,i}(t), \quad m_n(t) = \sum_{i=1}^{n} c_{n,i}(t), \qquad t \in [0, 1], \ n \in \mathbb{N}, \tag{3.3}$$

which are the covariance and the mean functions, respectively, of the random processes $Y_n(t)$. With tensor notation $K_n = \sum_i c_{n,i} \otimes c_{n,i}$, obviously

$$\|K_n\|_2^2 = \sum_{i,j} \langle c_{n,i}, c_{n,j}\rangle^2 = \sum_{i,j} c_{n,i,j}^2, \qquad \|m_n\|_2^2 = \sum_{i,j} \langle c_{n,i}, c_{n,j}\rangle = \sum_{i,j} c_{n,i,j}. \tag{3.4}$$

Before turning to convergence, we examine some interesting integrability issues. The first result of this subsection is based on the Paley–Zygmund argument – see, for example, de la Peña and Giné (1999, pp. 119–124), in particular their Corollary 3.3.4 – which we restate for ease of reference:

Lemma 3.1. *Let V be a random variable such that*

$$\mathrm{E}V^4 \leqslant C(\mathrm{E}V^2)^2.$$

Then, for all $t > 0$,

$$I_{\mathrm{E}V^2 \geqslant 2t^2} \leqslant 4C \Pr\{|V| > t\},$$

that is,

$$\mathrm{E}V^2 < 2a^2 \text{ whenever } \Pr\{|V| > a\} < \frac{1}{4C}.$$

Proof. This follows immediately upon observing that

$$\mathrm{E}V^2 \leqslant t^2 + \mathrm{E}(V^2 I_{|V|>t}) \leqslant t^2 + (\mathrm{E}V^4)^{1/2}(\Pr\{|V| > t\})^{1/2}.$$

□

Proposition 3.2. *The sequence $\{\|Y_n\|_2\}$, with Y_n as defined in* (3.1)*, is stochastically bounded if and only if both conditions*

$$\sup_n \sum_{k=1}^{n} \|c_{n,k}\|_2^2 < \infty \tag{3.5}$$

and

$$\sup_n \|m_n\|_2^2 = \sup_n \sum_{1\leqslant i,j\leqslant n} \langle c_{n,i}, c_{n,j}\rangle < \infty \tag{3.6}$$

are satisfied, and the same is true for the sequence $\{\|Z_n\|_2\}$*. Moreover,* $\lim_n \|Y_n\|_2 = 0$ *in probability if and only if*

$$\lim_n \sum_{k=1}^{n} \|c_{n,k}\|_2^2 = \lim_n \sum_{1\leqslant i,j\leqslant n} \langle c_{n,i}, c_{n,j}\rangle = 0, \tag{3.7}$$

and the same is true for the sequence $\{\|Z_n\|_2\}$*.*

Proof. We will use the abbreviated notation $c_{i,j}$ for $c_{n,i,j}$. Since

$$\begin{aligned}
\mathrm{E}\|Y_n\|_2^2 &= \int_0^1 \mathrm{E}\left(\sum_k c_{n,k}(t)\xi_k\right)^2 \mathrm{d}t \\
&= \int_0^1 \left(\sum_k c_{n,k}^2(t) + \sum_{i,j} c_{n,i}(t)c_{n,j}(t)\right)\mathrm{d}t \\
&= \sum_k c_{k,k} + \sum_{i,j} c_{i,j},
\end{aligned} \tag{3.8}$$

it follows that conditions (3.5) and (3.6) are sufficient for tightness of the sequence $\{\|Y_n\|_2\}$ and that (3.7) is sufficient for its convergence to zero in probability. Sufficiency for tightness and convergence of $\{\|Z_n\|_2\}$ follows from this and the law of large numbers. Necessity in both cases follows immediately from Lemmas 3.1 and 2.2. □

One can say a little more about the way Y_n converges:

Proposition 3.3. *If the sequence* $\{\|Y_n\|_2\}$ *is stochastically bounded, then*

$$\sup_n \mathrm{E}\|Y_n\|_2^m < \infty$$

for all $m < \infty$*.*

The proof is based on Hoffmann-Jørgensen's inequality and the following lemma:

Lemma 3.4. *Let* $\delta \in (0, 1)$ *and let* $x_1, \ldots, x_n$ *be real numbers such that* $x_j \in [0, 1)$*,* $j = 1, \ldots, n$*, and* $1 - \prod_{j=1}^{n}(1 - x_j) \leqslant \delta$*. Then*

$$1-\prod_{j=1}^{n}(1-x_j^2)\leqslant\frac{2\delta}{1-\delta}\left(1-\prod_{j=1}^{n}(1-x_j)\right). \tag{3.9}$$

Proof. Set $F(x_1, \ldots, x_n)=1-\prod_{j=1}^{n}(1-x_j^2)$, $G(x_1, \ldots, x_n)=1-\prod_{j=1}^{n}(1-x_j)$ and $A=\{(x_1, \ldots, x_n)\in\mathbb{R}^n : x_i\in[0, 1),\ i=1, \ldots, n,\ 1-\prod_{j=1}^{n}(1-x_j)\leqslant\delta\}$. Note that $(x_1, \ldots, x_n)\in A$ implies $x_i\leqslant\delta$, $i=1, \ldots, n$. Since

$$\frac{\partial F}{\partial x_i}=\frac{2x_i}{1-x_i^2}(1-F)\ \text{and}\ \frac{\partial G}{\partial x_i}=\frac{1}{1-x_i}(1-G),$$

it follows that

$$\frac{\partial F/\partial x_i}{\partial G/\partial x_i}=\frac{2x_i}{1+x_i}\prod_{i=1}^{n}(1+x_i). \tag{3.10}$$

It is easy to see, using Lagrange multipliers for example, that

$$\sup_{(x_1,\ldots,x_n)\in A}\prod_{i=1}^{n}(1+x_i)=(2-(1-\delta)^{1/n})^n\leqslant\frac{1}{1-\delta}.$$

This, combined with (3.10) shows that

$$\frac{\partial F/\partial x_i}{\partial G/\partial x_i}\leqslant\frac{2\delta}{1-\delta}\ \text{ for every } (x_1, \ldots, x_n)\in A.$$

Therefore

$$F(x_1, \ldots, x_n)=\int_0^1\sum_{i=1}^{n}x_i\frac{\partial F}{\partial x_i}\bigg|_{t(x_1,\ldots,x_n)}\mathrm{d}t\leqslant\frac{2\delta}{1-\delta}\int_0^1\sum_{i=1}^{n}x_i\frac{\partial G}{\partial x_i}\bigg|_{t(x_1,\ldots,x_n)}\mathrm{d}t$$
$$=\frac{2\delta}{1-\delta}G(x_1, \ldots, x_n).$$

□

Proof of Proposition 3.3. Let us assume $\{\|Y_n\|_2\}$ is stochastically bounded, and let $m\geqslant 1$. By convexity, as in the proof of Proposition 3.2,

$$\mathrm{E}\|Y_n\|_2^{2m}\leqslant 2^{2m-1}\mathrm{E}\left\|\sum_k c_{n,k}(t)(\xi_k-1)\right\|_2^{2m}+2^{2m-1}\left\|\sum_k c_{n,k}(t)\right\|_2^{2m},$$

where the power of the norm in the last summand is just

$$\left\|\sum_k c_{n,k}(t)\right\|_2^{2m}=\left(\int_0^1\left(\sum_k c_{n,k}(t)\right)^2\mathrm{d}t\right)^m=\left(\sum_{i,j}c_{i,j}\right)^m,$$

which is uniformly bounded by Proposition 3.2. To bound the first summand we can proceed as in (2.6) and obtain

$$\mathrm{E}\left\| \sum_k c_{n,k}(t)(\xi_k - 1) \right\|_2^{2m} \leqslant (2m-1)^m \cdot 2^{2m} \mathrm{E}\left(\sum_k c_{k,k}(\xi_k - 1)^2 \right)^m .$$

By Hoffmann-Jørgensen's inequality for sums of non-negative independent random variables (see, for example Theorem 1.2.5 and the comment below it in de la Peña and Giné 1999, p. 12), there is a universal constant $K < \infty$ such that

$$\mathrm{E}\left(\sum_k c_{k,k}(\xi_k - 1)^2 \right)^m \leqslant K \left[\left(\mathrm{E}\left(\sum_k c_{k,k}(\xi_k - 1)^2 \right) \right)^m + \mathrm{E}\left(\max_k c_{k,k}(\xi_k - 1)^2 \right)^m \right].$$

The first summand on the right-hand side of this inequality does not exceed $(\sum c_{k,k})^m$, which is uniformly bounded by Proposition 3.2. So the proposition will be proved if we show

$$\sup_n \mathrm{E}\left(\max_k c_{k,k}(\xi_k - 1)^2 \right)^m < \infty. \tag{3.11}$$

Since $\mathrm{E}\max_k c_{k,k}(\xi_k - 1)^2 \leqslant \sum_k c_{k,k}$, it follows from Proposition 3.2 that this sequence of maxima is tight. Thus, for every $\delta > 0$ there exists $M(\delta) < \infty$ such that

$$\sup_n \Pr\left\{ \max_{k \leqslant n} c_{n,k,k}(\xi_k - 1)^2 > M \right\} < \delta$$

for all $M \geqslant M(\delta)$, and we can assume that $M(\delta) > 2$ for all $\delta > 0$. Also, since $\sup_n \sum_k c_{n,k,k} < \infty$, we can assume without loss of generality that $c_{n,k,k} = c_{k,k} \leqslant 1$ for all k and n. Then, since

$$\Pr\{|\xi - 1| > t\} = \mathrm{e}^{1-t} \qquad \text{for } t \geqslant 1,$$

these probabilities are

$$\Pr\left\{ \max_{k \leqslant n} c_{n,k,k}(\xi_k - 1)^2 > M \right\} = 1 - \prod_{k=1}^n \left(1 - \mathrm{e}^{1-\sqrt{M/c_{n,k,k}}} \right) := h_n(M).$$

So, we have

$$\sup_n \left[1 - \prod_{k=1}^n (1 - \mathrm{e}^{1-\sqrt{M/c_{n,k,k})}} \right] := \sup_n h_n(M) < \delta$$

for all $M \geqslant M(\delta)$, and we can apply Lemma 2.4 to $h_n(M)$. If we take $\delta = 1/(3 \cdot 4^m)$, then this lemma gives $h_n(4R) < (3/4^{m+1})h(R)$ for all $R \geqslant M(1/(3 \cdot 4^m))$. This inequality, together with the fact that $h_n(M)$ is decreasing, shows that, for all $n \in \mathbb{N}$,

$$
\begin{aligned}
\mathrm{E}\left(\max_{k\leqslant n} c_{n,k,k}(\xi_k-1)^2\right)^m &= m\int_0^\infty t^{m-1}h_n(t)\mathrm{d}t \\
&\leqslant M^m + \sum_{k=0}^\infty m\int_{4^k M}^{4^{k+1}M} t^{m-1}h_n(t)\mathrm{d}t \\
&\leqslant M^m + M^m\sum_{k=0}^\infty 4^{(k+1)m}h_n(4^k M) \\
&\leqslant M^m\left[1+4^m h_n(M)\sum_{k=0}^\infty\left(\frac{3}{4}\right)^k\right] < 3M^m,
\end{aligned}
$$

proving (3.11) and the proposition. □

With these preliminaries on integrability out of the way, we now consider convergence in law of the sequence $\{\|Y_n\|_2\}$. We consider several cases, corresponding to the different cases for convergence of the square integral of the quantile process described in the Introduction.

3.1. Convergence of the processes Y_n

Here we obtain necessary and sufficient conditions for weak convergence of Y_n as L_2-valued random vectors; then convergence of $\|Y_n\|_2^2$ will be an immediate consequence of the continuous mapping theorem for weak convergence. Note that

$$
\mathrm{P_r}(\|c_{n,i}\xi_i\|_2 > \epsilon) = \exp(-\epsilon/\|c_{n,i}\|_2)
$$

and therefore, the triangular array $\{c_{n,i}\xi : i = 1, \ldots, n;\ n \in \mathbb{N}\}$ is infinitesimal if and only if

$$
\max_i \|c_{n,i}\|_2 \to 0 \tag{3.12}
$$

as $n\to\infty$. The next theorem gives necesary and sufficient conditions for the convergence in law in $L_2(0, 1)$ of $\{Y_n\}$ under (3.12). Under infinitesimality, the only possible limits of $\{Y_n\}$ are Gaussian, with a trace-class covariance operator. K_n and m_n are defined as in (3.3).

Theorem 3.5. *Assuming condition* (3.12) *holds, the sequence* $\{Y_n\}$ *converges in law in* $L_2(0, 1)$ *if and only if the following conditions hold:*

(i) *There exists a symmetric, positive semi-definite, trace-class kernel* $K(s, t) \in L_2((0, 1)\times(0, 1))$ *such that*

$$
K_n \underset{L_2}{\longrightarrow} K. \tag{3.13}
$$

(ii) *If* $\lambda_i \geqslant 0$ *are the eigenvalues of* K *then*

$$\sum_{i=1}^{n} \|c_{n,i}\|_2^2 \to \sum_{i=1}^{\infty} \lambda_i. \tag{3.14}$$

(iii) *There exists* $m \in L_2(0, 1)$ *such that*

$$m_n \underset{L_2}{\longrightarrow} m. \tag{3.15}$$

If (i), (ii) and (iii) hold, then Y_n *converges in law in* $L_2(0, 1)$ *to an* $L_2(0, 1)$*-valued Gaussian random variable* Y *with mean function* m *and covariance operator* Φ_K *given by*

$$\Phi_K(f, g) = \int_0^1 \int_0^1 K(s, t) f(s) g(t) \mathrm{d}s \, \mathrm{d}t$$

for $f, g \in L_2(0, 1)$.

Proof. *Necessity.* Let us assume first that the L_2-valued random vectors Y_n converge in law. Then $\{\|Y_n\|_2\}$ also converges in law and, moreover, by Proposition 3.3, its moments converge as well (to the moments of the limit). This implies, in particular, that the sequence

$$\mathrm{E}\|Y_n\|_2^2 = \sum_{i=1}^{n} \|c_{n,i}\|_2^2 + \left\| \sum_{i=1}^{n} c_{n,i} \right\|_2^2, \quad n \in \mathbb{N}, \tag{3.16}$$

converges. Note also that convergence in law of Y_n to Y plus uniform integrability of $\{\|Y_n\|_2\}$, which is a consequence of moment convergence, ensure that $\mathrm{E}Y_n \to_{L_2} \mathrm{E}Y$ and, therefore, that

$$\sum_{i=1}^{n} c_{n,i} \underset{L_2}{\longrightarrow} m := \mathrm{E}Y. \tag{3.17}$$

Now, (3.16) and (3.17) imply (3.15) and also that the left-hand side of (3.14) converges to a finite limit. We have also proved that $\{Y_n - \mathrm{E}Y_n\}$ converges in law:

$$Y_n - \mathrm{E}Y_n = \sum_{i=1}^{n} c_{n,i}(\xi_i - 1) \underset{d}{\to} Y - \mathrm{E}Y. \tag{3.18}$$

Also, (3.18) and uniform integrability imply

$$\sum_{i=1}^{n} \|c_{n,i}\|_2^2 = \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2 \to \mathrm{E}\|Y - \mathrm{E}Y\|_2^2. \tag{3.19}$$

Another consequence of (3.18) is that

$$(Y_n - \mathrm{E}Y_n) \otimes (Y_n - \mathrm{E}Y_n) \underset{d}{\to} (Y - \mathrm{E}Y) \otimes (Y - \mathrm{E}Y) \tag{3.20}$$

in $L_2((0, 1) \times (0, 1))$ ((3.20) follows from (3.18) and continuity of the map $(f, g) \mapsto f \otimes g$). Now, since $\|f \otimes g\|_2 = \|f\|_2 \|g\|_2$, convergence of moments of $\|Y_n\|_2$ also ensures uniform integrability of $(Y_n - \mathrm{E}Y_n) \otimes (Y_n - \mathrm{E}Y_n)$ and, just as above, we obtain that

$$K_n = \mathrm{E}((Y_n - \mathrm{E}Y_n) \otimes (Y_n - \mathrm{E}Y_n)) \underset{L_2}{\longrightarrow} K := \mathrm{E}((Y - \mathrm{E}Y) \otimes (Y - \mathrm{E}Y)).$$

Now let λ_i, ϕ_i be, respectively, the eigenvalues and the corresponding eigenfunctions associated with the kernel K. Then

$$\lambda_i = \int_0^1\int_0^1 K(s,\, t)\phi_i(s)\phi_i(t)\mathrm{d}s\,\mathrm{d}t = \mathrm{E}\langle Y - \mathrm{E}Y,\, \phi_i\rangle^2.$$

Therefore,

$$\mathrm{E}\|Y - \mathrm{E}Y\|_2^2 = \mathrm{E}\left(\sum_{i=1}^{\infty}\langle Y - \mathrm{E}Y,\, \phi_i\rangle^2\right) = \sum_{i=1}^{\infty}\lambda_i,$$

which, combined with (3.19), yields (3.13) and (3.14).

Sufficiency. Assume now that (3.12)–(3.15) hold. Let us denote $\xi_{i,\epsilon} = \xi_i I_{\{\|c_{n,i}\xi_i\|_2 \leqslant \epsilon\}}$ and $\xi_i^\epsilon = \xi_i I_{\{\|c_{n,i}\xi_i\|_2 > \epsilon\}}$ for $\epsilon > 0$ and $i \in \mathbb{N}$. Since $\mathrm{E}\xi_i I_{\{\xi_i > t\}} = (t+1)\mathrm{e}^{-t} \leqslant 1/t^2$ for large enough t, condition (3.14) implies that, for large enough n,

$$\begin{aligned}\left\|\mathrm{E}\sum_{i=1}^n c_{n,i}\xi_i^\epsilon\right\|_2 &= \left\|\sum_{i=1}^n c_{n,i}\mathrm{E}\xi_i^\epsilon\right\|_2 \\ &\leqslant \sum_{i=1}^n \|c_{n,i}\|_2 \mathrm{E}\xi_i^\epsilon \leqslant \frac{1}{\epsilon^2}\left(\sum_{i=1}^n \|c_{n,i}\|_2^2\right)\max_i \|c_{n,i}\|_2 \to 0.\end{aligned}$$

Hence, by the central limit theorem on Hilbert spaces (see Araujo and Giné 1980, Corollary 3.7.8), if $\{Y_n\}$ is shift convergent in law, we can take the shifts to be the expected values $\mathrm{E}Y_n$ and therefore, by the same central limit theorem, the proof reduces to showing that:

(a) $\sum_{j=1}^n \Pr(\|c_{n,j}\xi_j\|_2 > \epsilon) \to 0$ as $n \to \infty$, for every $\epsilon > 0$;
(b) for every $\epsilon > 0$ and every $f \in L_2(0,\, 1)$,

$$\sum_{j=1}^n \mathrm{var}(\langle c_{n,j}\xi_{j,\epsilon},\, f\rangle) \to \Phi_K(f,\, f);$$

(c) there exists a complete orthonormal system of functions $\{\phi_i\}_{i \geqslant 1}$ in $L_2(0,\, 1)$ such that

$$\lim_{k\to\infty}\limsup_{n\to\infty}\sum_{j=1}^n\left(\mathrm{E}\|c_{n,j}\xi_{j,\epsilon} - \mathrm{E}c_{n,j}\xi_{j,\epsilon}\|^2 - \sum_{i=1}^k \mathrm{E}\langle c_{n,j}\xi_{j,\epsilon} - \mathrm{E}c_{n,j}\xi_{j,\epsilon},\, \phi_i\rangle^2\right) = 0.$$

To check (a), we see that, as a consequence of (3.12) and (3.13),

$$\begin{aligned}\sum_{j=1}^n \Pr(\|c_{n,j}\xi_j\| > \epsilon) &\leqslant \frac{1}{\epsilon^2}\sum_{j=1}^n \mathrm{E}\|c_{n,j}\xi_j^\epsilon\|_2^2 = \frac{1}{\epsilon^2}\sum_{j=1}^n \|c_{n,j}\|_2^2\mathrm{E}(\xi_j^\epsilon)^2 \\ &\leqslant \frac{1}{\epsilon^2}\left(\sum_{j=1}^n \|c_{n,j}\|_2^2\right)\mathrm{E}\xi_1^2 I_{\{\xi_1 > \epsilon/\max_i\|c_{n,i}\|_2\}} \to 0\end{aligned}$$

as $n \to \infty$. This calculation shows also that, for every $f \in L_2(0,\, 1)$,

$$\sum_{j=1}^{n} \mathrm{var}(\langle c_{n,j}\xi_j^\epsilon, f\rangle) \leqslant \sum_{j=1}^{n} \mathrm{E}\langle c_{n,j}\xi_j^\epsilon, f\rangle^2$$

$$\leqslant \|f\|_2^2 \sum_{j=1}^{n} \mathrm{E}\|c_{n,j}\xi_j^\epsilon\|_2^2 \to 0,$$

which implies that (b) is equivalent to

$$\sum_{j=1}^{n} \mathrm{var}(\langle c_{n,j}\xi_j, f\rangle) \to \Phi_K(f, f). \tag{3.21}$$

In order to prove (3.21), we recall that $\mathrm{cov}(\sum_{j=1}^n c_{n,j}(s)\xi_j, \sum_{j=1}^n c_{n,j}(t)\xi_j) = K_n(s, t)$, which, combined with (3.13), implies that

$$\sum_{j=1}^{n} \mathrm{var}(\langle c_{n,j}\xi_j, f\rangle) = \mathrm{var}\left(\int_0^1 \left(\sum_{j=1}^{n} c_{n,j}(t)\xi_j\right) f(t)\mathrm{d}t\right)$$

$$= \int_0^1\int_0^1 \mathrm{cov}\left(\sum_{j=1}^{n} c_{n,j}(s)\xi_j, \sum_{j=1}^{n} c_{n,j}(t)\xi_j\right) f(s)f(t)\mathrm{d}s\,\mathrm{d}t$$

$$= \int_0^1\int_0^1 K_n(s, t)f(s)f(t)\mathrm{d}s\,\mathrm{d}t \to \int_0^1\int_0^1 K(s, t)f(s)f(t)\mathrm{d}s\,\mathrm{d}t$$

$$= \Phi_K(f, f),$$

proving (3.21). Finally, to show that (c) holds, let $\{\phi_i\}$ be a complete orthonormal system of eigenfunctions of the covariance operator Φ_K and let $\{\lambda_i\}$ be the associated eigenvalues. Then, using (3.14) and the fact that, as shown above, $\sum_{j=1}^n \mathrm{E}\|c_{n,j}\xi_j^\epsilon\|_2^2 \to 0$, we have

$$\lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{E}\|c_{n,j}\xi_{j,\epsilon} - \mathrm{E}c_{n,j}\xi_{j,\epsilon}\|^2 = \lim_{n\to\infty} \sum_{j=1}^{n} \mathrm{E}\|c_{n,j}\xi_j - \mathrm{E}c_{n,j}\xi_j\|^2$$

$$= \lim_{n\to\infty} \sum_{j=1}^{n} \|c_{n,j}\|^2 = \sum_{i=1}^{\infty} \lambda_i$$

and, by (3.21),

$$\lim_{n\to\infty} \sum_{j=1}^{n}\sum_{i=1}^{k} \mathrm{E}\langle c_{n,j}\xi_{j,\epsilon} - \mathrm{E}c_{n,j}\xi_{j,\epsilon}, \phi_i\rangle^2 = \sum_{i=1}^{k} \Phi_K(\phi_i, \phi_i) = \sum_{i=1}^{k} \lambda_i,$$

which completes the proof. □

The above proof shows that the sufficiency part of Theorem 3.5 also holds if we only assume that the random variables ξ_i are i.i.d. and square integrable (with trivial adjustments to account for mean and variance possibly different from 1).

As an inmediate consequence of Theorem 3.5, we obtain sufficient conditions for convergence in law of the L_2 norms of linear combinations of independent exponential random variables:

Corollary 3.6. *Suppose that* (3.12)–(3.15) *hold. Let S be a metric space and let $H: L_2(0, 1) \mapsto S$ be a continuous function. Then*

$$H(Y_n) \underset{d}{\to} H(Y),$$

where Y is an $L_2(0, 1)$-valued Gaussian random variable with mean function m and covariance operator Φ_K given by

$$\Phi_K(f, g) = \int_0^1 \int_0^1 K(s, t) f(s) g(t) \mathrm{d}s \, \mathrm{d}t$$

for $f, g \in L_2(0, 1)$.

Below, we will apply this corollary to $H(f) = \|f\|_2^2 - \sum_{k=1}^2 \langle f, h_k \rangle^2$ with $h_k \in L_2$.

Remark 3.1. The limiting random process Y in Theorem 3.5 and Corollary 3.6 is centred if and only if $\|m_n\|_2^2 = \sum_{i,j} c_{n,i,j} \to 0$. The type of argument employed in the proof of Theorem 3.5 shows that, under the infinitesimality condition (3.12), conditions (3.13) and (3.14) are necessary and sufficient for convergence in law of the processes $Y_n - \mathrm{E} Y_n$ and that the limiting random process has then a centred Gaussian distribution with covariance operator Φ_K.

3.2. Shift convergence of $\|Y_n\|_2^2$: *I*

It can be proved that shift tightness of $\{\|Y_n\|_2^2\}$ implies tightness of the sequence centred at expectations, and even tightness of the sequence $\{\|Z_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2\}$, but this is marginal to our analysis here and will therefore be omitted (it would only add a comment on the sharpness of the results that follow). In Section 3.1, we examined the case when the kernels K_n associated with Y_n converge in L_2 to a trace-class kernel. In that case, $Y_n - \mathrm{E} Y_n \to_d Y - m = \sum_i \sqrt{\lambda_i} \phi_i Z_i$ and $\|Y - m\|_2^2 = \sum_i \lambda_i Z_i^2$, where $\{Z_i\}$ is an ortho-Gaussian sequence (a sequence of i.i.d. standard normal random variables). Of course, convergence of this series requires $\sum_i \lambda_i < \infty$. However, if we allow centring, then

$$\|Y_n - \mathrm{E} Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E} Y_n\|_2^2 \underset{d}{\to} \sum_i \lambda_i (Z_i^2 - 1)$$

and, clearly, in order to make sense of this limit it suffices (and is also necessary) that $\sum_i \lambda_i^2 < \infty$, a weaker condition. We deal here with this situation, that is, we relax the assumptions on K in Theorem 3.5 by only assuming that $K \in L_2((0, 1) \times (0, 1))$. In this case, the operator induced by K on $L_2(0, 1)$ is Hilbert–Schmidt, that is, its eigenvalues $\{\lambda_k\}$ satisfy $\sum_k \lambda_k^2 < \infty$ (Dunford and Schwartz 1963, XI.6 and XI.8.44). Then, with considerable abuse of notation, we define

$$\|Y - \mathrm{E}Y\|_2^2 - \mathrm{E}\|Y - \mathrm{E}Y\|_2^2 := \sum_k \lambda_k (Z_k^2 - 1), \tag{3.22}$$

where the variables Z_k are independent standard normal (Y may not exist but the series does converge almost surely). (Omission of the dependence of Y on K will not result in confusion.)

We start with a useful lemma on the asymptotic normality of sums of independent exponential random variables.

Lemma 3.7. *If $\{a_{n,i} : i = 1, \ldots, n;\ n \in \mathbb{N}\}$ is a triangular array of real numbers then $\sum_{i=1}^n a_{n,i}(\xi_i - 1) \to_d \sigma Z$, where Z is a standard normal random variable, if and only if*

$$\max_i |a_{n,i}| \to 0 \quad \text{and} \quad \sum_{i=1}^n a_{n,i}^2 \to \sigma^2. \tag{3.23}$$

Morevover, if $\max_i |a_{n,i}| \to 0$ then the only possible limit laws of $\sum_{i=1}^n a_{n,i}(\xi_i - 1)$ are normal, and convergence in law is equivalent to convergence of $\sum_{i=1}^n a_{n,i}^2$.

Proof (sketch). Arguing as in Proposition 3.2, we see that convergence in law implies convergence of the second moments, showing that the second part of (3.23) is necessary. If the first condition is not satisfied, then (after reordering the indices if necessary) we can find a subsequence n' such that $a_{n',1} \to a > 0$. In that subsequence the possible limits in distribution would be laws of type $\nu * \mu$, ν being the law of $a(\xi_1 - 1)$. But the Gaussian family is factor closed, which implies that the first part of (3.23) is also necessary. Conversely, if (3.23) holds then convergence follows in a straightforward way from the classical central limit theorem for triangular arrays. □

The main argument of this section is contained in the proof of the following proposition.

Proposition 3.8. *If $\max_i \|c_{n,i}\|_2 \to 0$ and $K_n \to_{L_2} K$, so that K is necessarily in $L_2((0, 1) \times (0, 1))$, then*

$$\|Y_n - \mathrm{E}Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2 \underset{d}{\to} \|Y - \mathrm{E}Y\|_2^2 - \mathrm{E}\|Y - \mathrm{E}Y\|_2^2,$$

where $\|Y - \mathrm{E}Y\|_2^2 - \mathrm{E}\|Y - \mathrm{E}Y\|_2^2$ is as defined in (3.22).

Proof. Let $\{\phi_k\}$ be a complete orthonormal system of eigenfunctions of K, of eigenvalue λ_k for each k. Then

$$\|Y_n - \mathrm{E}Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2 = \sum_{k=1}^{\infty} \big(\langle Y_n - \mathrm{E}Y_n, \phi_k \rangle^2 - \mathrm{E}\langle Y_n - \mathrm{E}Y_n, \phi_k \rangle^2 \big),$$

where $\langle Y_n - \mathrm{E}Y_n, \phi_k \rangle = \sum_{i=1}^n \langle c_{n,i}, \phi_k \rangle (\xi_i - 1)$. By Lemma 3.7, if (3.23) holds for $a_{n,i} = \langle c_{n,i}, \phi_k \rangle$ then the only possible limit laws of $\langle Y_n - \mathrm{E}Y_n, \phi_k \rangle$ are normal, and there is convergence if and only if $\{\sum_{i=1}^n \langle c_{n,i}, \phi_k \rangle^2\}$ converges. Since $K_n \to_{L_2} K$, we have in particular,

$$\begin{aligned}
\mathrm{E}(\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle\langle Y_n - \mathrm{E}Y_n, \phi_l\rangle) &= \int_0^1\int_0^1 K_n(s, t)\phi_k(s)\phi_l(t)\mathrm{d}s\,\mathrm{d}t \\
&\to \int_0^1\int_0^1 K(s, t)\phi_k(s)\phi_l(t)\mathrm{d}s\,\mathrm{d}t \\
&= \begin{cases} \lambda_k & \text{if } k = l \\ 0 & \text{if } k \neq l, \end{cases}
\end{aligned}$$

and therefore $(\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle)_{k=1}^M \to_d (\lambda_k Z_k)_{k=1}^M$. Then, by the continuous mapping theorem for weak convergence,

$$\sum_{k=1}^{M}\left(\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle^2 - \mathrm{E}\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle^2\right) \underset{d}{\to} \gamma_M := \sum_{k=1}^{M}\lambda_k(Z_j^2 - 1). \tag{3.24}$$

Since $\sum_{k=1}^\infty \lambda_k^2 = \|K\|_2^2 < \infty$, γ_M converges almost surely and in L_2 and, with some abuse of notation, as explained above, we denote this limit as $\|Y - \mathrm{E}Y\|_2^2 - \mathrm{E}\|Y - \mathrm{E}Y\|_2^2$, that is, we have,

$$\gamma_M \underset{L_2}{\longrightarrow} \|Y - \mathrm{E}Y\|_2^2 - \mathrm{E}\|Y - \mathrm{E}Y\|_2^2. \tag{3.25}$$

Set $c_{n,i}^M = c_{n,i} - \sum_{k=1}^M \langle c_{n,i}, \phi_k\rangle\phi_k$, $Y_n^M = \sum_{i=1}^n c_{n,i}^M \xi_i$ and $K_n^M = \sum_{i=1}^n c_{n,i}^M \otimes c_{n,i}^M$. Observe that

$$\sum_{k=M+1}^{\infty}\left(\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle^2 - \mathrm{E}\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle^2\right) = \|Y_n^M - \mathrm{E}Y_n^M\|_2^2 - \mathrm{E}\|Y_n^M - \mathrm{E}Y_n^M\|_2^2. \tag{3.26}$$

We claim that, under (3.13),

$$\lim_{M\to\infty}\limsup_{n\to\infty}\mathrm{var}(\|Y_n^M - \mathrm{E}Y_n^M\|_2^2) = 0. \tag{3.27}$$

We recall that $K = \sum_{k=1}^\infty \lambda_k\phi_k \otimes \phi_k$ and define $K^M = \sum_{k=M+1}^\infty \lambda_k\phi_k \otimes \phi_k$. We can easily see that $\|K^M\|_2^2 = \sum_{k=M+1}^\infty \lambda_k^2$. Now, since

$$\{\phi_k \otimes \phi_k\}_k \cup \{\tfrac{1}{\sqrt{2}}(\phi_k \otimes \phi_l + \phi_k \otimes \phi_l)\}_{k\neq l}$$

is an orthonormal basis for $L_2((0, 1) \times (0, 1))$, we have that

$$\|K_n^M - K^M\|_2^2 = \sum_k \langle K_n^M - K^M, \phi_k \otimes \phi_k\rangle^2 + 2\sum_{k\neq l}\langle K_n^M - K^M, \phi_k \otimes \phi_l\rangle^2 \tag{3.28}$$

(here we have used the fact that $\langle f \otimes f, g \otimes h\rangle = \langle f \otimes f, h \otimes g\rangle$). Observe that $\langle K^M, \phi_k \otimes \phi_l\rangle = \langle K, \phi_k \otimes \phi_k\rangle$ if $k = l > M$ and $\langle K^M, \phi_k \otimes \phi_l\rangle = 0$ otherwise, and also that $\langle K_n^M, \phi_k \otimes \phi_l\rangle = \langle K_n, \phi_k \otimes \phi_l\rangle$ if $k, l > M$ and $\langle K_n^M, \phi_k \otimes \phi_l\rangle = 0$ otherwise. Combining this with (3.28), we obtain that

$$\begin{aligned}\|K_n^M - K^M\|_2^2 &= \sum_{k>M} \langle K_n - K, \phi_k \otimes \phi_k\rangle^2 + 2 \sum_{k\neq l; k,l>M} \langle K_n - K, \phi_k \otimes \phi_l\rangle^2 \\ &\leqslant \sum_k \langle K_n - K, \phi_k \otimes \phi_k\rangle^2 + 2\sum_{k\neq l} \langle K_n - K, \phi_k \otimes \phi_l\rangle^2 \\ &= \|K_n - K\|_2^2.\end{aligned}$$

The last inequality implies that $K_n^M \to_{L_2} K^M$ and also that $\|K_n^M\|_2 \to \|K^M\|_2$. Now this convergence, combined with the fact that $\mathrm{var}(\|Y_n^M - \mathrm{E}Y_n^M\|_2^2) \leqslant 8\|K_n^M\|_2^2$, proves claim (3.27). The proposition now follows from (3.24), (3.25), (3.26) and (3.27) through a standard 3ϵ argument. □

We should remark that this proposition also holds if we replace the sequence of exponential random variables by an i.i.d. sequence of square-integrable random variables, with only formal changes in the proof.

Both Theorem 3.5 and Proposition 3.8 are exercises on the central limit theorem in Hilbert space; however, Proposition 3.8 can be seen as a limit theorem for quadratic forms, and this subject has a long history, reviewed, for example, in Guttorp and Lockhart (1988). Theorem 1 in de Wet and Venter (1973) and Theorem 5 in Guttorp and Lockhart (1988) could seemingly apply to give Proposition 3.8; however, the conditions in either theorem are quite difficult to verify and we have been unable to check them in the case of interest to us, whereas the conditions in Proposition 3.8 are very easy to decide in general.

Our next result gives sufficient conditions for convergence in law of $\{\|Y_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2\}$. Actually, in order to have a result directly applicable to Wasserstein distances, we must sacrifice simplicity and consider a slightly more complicated functional. A warning on notation: we write $\langle Y - \mathrm{E}Y, h\rangle = \sum\sqrt{\lambda_k}\langle\phi_k, h\rangle Z_k$ for ϕ_k orthonormal, $h \in L_2(0, 1)$ and $\sum\lambda^2 < \infty$ even though $Y - \mathrm{E}Y$ may not make sense.

Theorem 3.9. *Let h_ℓ, $\ell = 1, \ldots, r$, be functions in $L_2(0, 1)$. If $\max_i\|c_{n,i}\|_2 \to 0$, $K_n \to_{L_2} K$ and $m_n \to_{L_2} m$, then*

$$\begin{aligned}&\|Y_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2 - \sum_{\ell=1}^r \langle Y_n - \mathrm{E}Y_n, h_\ell\rangle^2 \\ &\quad \underset{d}{\to} \|Y - \mathrm{E}Y\|_2^2 - \mathrm{E}\|Y - \mathrm{E}Y\|_2^2 + 2\langle Y - \mathrm{E}Y, m\rangle - \sum_{\ell=1}^r \langle Y - \mathrm{E}Y, h_\ell\rangle^2 \\ &\quad := \sum_{k=1}^\infty \lambda_k (Z_k^2 - 1) + 2\sum_{k=1}^\infty \sqrt{\lambda_k}\langle m, \phi_k\rangle Z_k - \sum_{\ell=1}^r \left(\sum_{k=1}^\infty \sqrt{\lambda_k}\langle h_\ell, \phi_k\rangle Z_k\right)^2\end{aligned}$$

where $\|Y - \mathrm{E}\|_2^2 - \mathrm{E}\|Y - \mathrm{E}Y\|_2^2$ is defined as in (3.22) and $\{Z_k\}$ is an ortho-Gaussian sequence.

Proof. We require the proof of Proposition 3.8 rather than its statement. First, we note that

$$\|Y_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2 = (\|Y_n - \mathrm{E}Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2) + 2\langle Y_n - \mathrm{E}Y_n, \mathrm{E}Y_n\rangle. \tag{3.29}$$

As in the previous proof,

$$\begin{aligned}\mathrm{E}\langle Y_n - \mathrm{E}Y_n, \mathrm{E}Y_n\rangle\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle &= \langle K_n, m_n \otimes \phi_k\rangle \\ &\to \langle K, m \otimes \phi_k\rangle = \lambda_k\langle m, \phi_k\rangle\end{aligned}$$

and also, similarly,

$$\begin{aligned}\mathrm{E}\langle Y_n - \mathrm{E}Y_n, \mathrm{E}Y_n\rangle\langle Y_n - \mathrm{E}Y_n, h_\ell\rangle &= \langle K_n, m_n \otimes h_\ell\rangle \to \langle K, m \otimes h_\ell\rangle, \\ \mathrm{E}\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle\langle Y_n - \mathrm{E}Y_n, h_\ell\rangle &= \langle K_n, \phi_k \otimes h_\ell\rangle \to \lambda_k\langle \phi_k, h_\ell\rangle.\end{aligned}$$

This implies that for each M we have convergence in law of the vector

$$(\langle Y_n - \mathrm{E}Y_n, \phi_1\rangle, \ldots, \langle Y_n - \mathrm{E}Y_n, \phi_M\rangle, \langle Y_n - \mathrm{E}Y_n, \mathrm{E}Y_n\rangle, \langle Y_n - \mathrm{E}Y_n, h_1\rangle, \ldots \langle Y_n - \mathrm{E}Y_n, h_\ell\rangle)$$

to the Gaussian vector

$$\left(\lambda_1 Z_1, \ldots, \lambda_M Z_M, \sum_{k=1}^{\infty}\sqrt{\lambda_k}\langle m, \phi_k\rangle Z_k, \sum_{k=1}^{\infty}\sqrt{\lambda_k}\langle h_1, \phi_k\rangle Z_k, \ldots, \sum_{k=1}^{\infty}\sqrt{\lambda_k}\langle h_\ell, \phi_k\rangle Z_k\right).$$

This gives weak convergence, for every $M < \infty$, of the random variables

$$\sum_{k=1}^{M}\big(\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle^2 - \mathrm{E}\langle Y_n - \mathrm{E}Y_n, \phi_k\rangle^2\big) + 2\langle Y_n - \mathrm{E}Y_n, \mathrm{E}Y_n\rangle - \sum_{\ell=1}^{r}\langle Y - \mathrm{E}Y, h_\ell\rangle^2,$$

by analogy with (3.24). By (3.29) these random varibles are 'finite-dimensional' approximations of the sequence of interest, and the result now follows by the approximation argument in the previous proof, as a consequence of the limit (3.27). □

The hypotheses in this theorem are quite natural. We will not deal with the question of whether they are necessary (given infinitesimality); however, note that the existence of K and m are necessary in order to define the limit.

3.3. Shift convergence of $\|Y_n\|_2^2$: *II*

There are some situations in which K_n is not convergent in L_2 but, nevertheless, $\{\|Y_n - \mathrm{E}Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2\}$ is weakly convergent. From the definitions we see that

$$\begin{aligned}\|Y_n - \mathrm{E}Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2 &= \sum_{1\leqslant i\neq j\leqslant n} c_{n,i,j}(\xi_i - 1)(\xi_j - 1) + \sum_{i=1}^n c_{n,i,i}\left[(\xi_i - 1)^2 - 1\right] \\ &= \sum_{i=1}^n \left[2\left(\sum_{j=1}^{i-1} c_{n,i,j}(\xi_j - 1)\right)(\xi_i - 1) + c_{n,i,i}\big((\xi_i - 1)^2 - 1\big)\right] \\ &= \sum_{i=1}^n x_{n,i},\end{aligned}$$

where

$$x_{n,i} = 2\left(\sum_{j=1}^{i-1} c_{n,i,j}(\xi_j - 1)\right)(\xi_i - 1) + c_{n,i,i}\big((\xi_i - 1)^2 - 1\big)$$

(and we use the convention that $\sum_{j=1}^{0} a_j = 0$). If $\{\xi_i'\}$ denotes an independent copy of the sequence $\{\xi_i\}$ and we set

$$\tilde{x}_{n,i} = 2\left(\sum_{j=1}^{i-1} c_{n,i,j}(\xi_j - 1)\right)(\xi_i' - 1) + c_{n,i,i}\big((\xi_i' - 1)^2 - 1\big)$$

for $i = 1, \ldots, n$ and $\mathcal{F}_{n,i} = \sigma(\xi_1, \ldots, \xi_i)$, then, for each $n \in \mathbb{N}$, $\{x_{n,i}\}$ and $\{\tilde{x}_{n,i}\}$ are tangent sequences with respect to $\{\mathcal{F}_i\}$, that is, $\mathcal{L}(x_{n,i}|\mathcal{F}_{n,i-1}) = \mathcal{L}(\tilde{x}_{n,i}|\mathcal{F}_{n,i-1})$ and the random variables $\tilde{x}_{n,i}$ are conditionally independent given the sequence $\{\xi_i\}$. Hence, $\{\tilde{x}_{n,i}\}$ is a decoupled tangent sequence to $\{x_{n,i}\}$ (see, for example, de la Peña and Giné 1999, Chapter 6). Decoupling introduces enough independence among the summands in $\sum_{i=1}^n \tilde{x}_{n,i}$ to enable us to use the central limit theorem in order to obtain their asymptotic distribution. The principle of conditioning – Theorem 1.1 in Jakubowski (1986), reproduced in de la Peña and Giné (1999, Theorem 7.1.4) – can then be used to conclude convergence in law of $\sum_{i=1}^n x_{n,i}$ itself. The proof of our next result follows this approach.

Theorem 3.10. *Let Z be a standard normal random variable. If*

$$\max_i \|c_{n,i}\| \to 0, \tag{3.30}$$

$$2\|K_n\|_2^2 + 6\sum_{i=1}^n \|c_{n,i}\|_2^4 \to \sigma^2 \tag{3.31}$$

and

$$\sum_{j\neq k}\left(\sum_{i:i>j\vee k}\langle c_{n,i}, c_{n,j}\rangle\langle c_{n,i}, c_{n,k}\rangle\right)^2 + \sum_j\left(\sum_{i:i>j}\langle c_{n,i}, c_{n,j}\rangle\langle c_{n,i}, c_{n,i} + c_{n,j}\rangle\right)^2 \to 0, \tag{3.32}$$

then

$$\|Y_n - \mathrm{E}Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2 \underset{d}{\to} \sigma Z. \tag{3.33}$$

If, instead of conditions (3.30)*,* (3.31) *and* (3.32)*, we have*

$$\max_i\Big(\|c_{n,i}\|_2^2+|\langle c_{n,i},\,m_n\rangle|\Big)\to 0, \tag{3.34}$$

$$2\|K_n\|_2^2+2\sum_{i=1}^n\|c_{n,i}\|_2^4+4\sum_i\langle c_{n,i},\,c_{n,i}+m_n\rangle^2\to\sigma^2 \tag{3.35}$$

and

$$\sum_{j,k}\left(\sum_{i:i>j\vee k}\langle c_{n,i},\,c_{n,j}\rangle\langle c_{n,i},\,c_{n,k}\rangle\right)^2+\sum_j\left(\sum_{i:i>j}\langle c_{n,i},\,c_{n,j}\rangle\langle c_{n,i},\,c_{n,i}+m_n+c_{n,j}\rangle\right)^2\to 0, \tag{3.36}$$

then

$$\|Y_n\|_2^2-\mathrm{E}\|Y_n\|_2^2\underset{d}{\to}\sigma Z. \tag{3.37}$$

Proof. We first prove the limit in (3.33). If we set $\tilde{U}_n=\sum_{i=1}^n\tilde{x}_{n,i}$, with $\tilde{x}_{n,i}$ defined as above, the principle of conditioning (Jakubowski 1986) reduces the proof to showing that

$$\mathcal{L}(\tilde{U}_n|\{\xi_i\})\underset{w}{\to}N(0,\,\sigma^2)$$

in probability. Arguing as in the proof of Lemma 3.7, we can see that this is equivalent to proving that

$$A_n:=\max_i\mathrm{E}(\tilde{x}_{n,i}^2|\{\xi_j\})=4\max_i\left(c_{i,i}^2+\left(c_{i,i}+\sum_{j=1}^{i-1}c_{i,j}(\xi_j-1)\right)^2\right)\underset{\Pr}{\to}0 \tag{3.38}$$

and

$$B_n:=\sum_i\mathrm{E}(\tilde{x}_{n,i}^2|\{\xi_j\})=4\sum_i c_{i,i}^2+4\sum_i\left(c_{i,i}+\sum_{j=1}^{i-1}c_{i,j}(\xi_j-1)\right)^2\underset{\Pr}{\to}\sigma^2. \tag{3.39}$$

After a straightforward but cumbersome computation that we omit, it is evident that

$$\mathrm{E}B_n=2\|K_n\|_2^2+6\sum_{i=1}^n\|c_{n,i}\|_2^4$$

and

$$\mathrm{var}(B_n)=16\left(\sum_{j\neq k}\left(\sum_{i:i>j\vee k}c_{i,j}c_{i,k}\right)^2+4\sum_j\left(\sum_{i:i>j}c_{i,j}(c_{i,i}+c_{i,j})\right)^2\right),$$

which, by (3.31) and (3.32), inmediately give (3.39). We now check (3.38), which is equivalent to $\max_i c_{i,i}\to 0$ and $\max_i|\sum_{j=1}^{i-1}c_{i,j}(\xi_j-1)|\to_{\Pr}0$. This last convergence follows from (3.32) and the use of a refined Octaviani maximal inequality (see, Proposition 1.1.2 in de la Peña and Giné 1999):

$$
\begin{aligned}
P\left(\max_i \left|\sum_{j=1}^{i-1} c_{i,j}(\xi_j - 1)\right| > t\right) &\leqslant 3 \max_i P\left(\left|\sum_{j=1}^{i-1} c_{i,j}(\xi_j - 1)\right| > t/3\right) \\
&\leqslant \frac{27}{t^2} \max_i \sum_{j=1}^{i-1} c_{i,j}^2 \leqslant \frac{27}{t^2} \max_i \sum_{j=1}^{n} c_{i,j}^2 \\
&= \frac{27}{t^2} \max_i \langle c_{n,i} \otimes c_{n,i}, K_n \rangle \\
&\leqslant \frac{27}{t^2} \|K_n\|_2 \max_i \|c_{n,i}\|_2^2 \to 0.
\end{aligned}
$$

This concludes the proof of the limit (3.33).

We now turn to the limit (3.37). The fact that

$$
\begin{aligned}
\|Y_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2 &= \sum_{1 \leqslant i \neq j \leqslant n} c_{i,j}(\xi_i - 1)(\xi_j - 1) + \sum_{i=1}^{n} \left(c_{i,i}\left[(\xi_i - 1)^2 - 1\right] + 2\langle c_{n,i}, m_n \rangle (\xi_i - 1)\right) \\
&= \sum_{i=1}^{n} y_{n,i},
\end{aligned}
$$

where $y_{n,i} = 2(\sum_{j=1}^{i-1} c_{n,i,j}(\xi_j - 1))(\xi_i - 1) + c_{n,i,i}((\xi_i - 1)^2 - 1) + 2\langle c_{n,i}, m_n \rangle (\xi_i - 1)$, can be used to conclude (3.37) by reproducing the proof of (3.33) almost verbatim. □

The principle of conditioning used in Theorem 3.10, which could be easily replaced by the Brown–Eagleson central limit theorem for martingales, has been used before in analogous situations. We will just mention Hall (1984), who uses it in density estimation, in order to prove a limit theorem for degenerate U-statistics with varying kernels. His result is different from ours and does not apply here, but there are similarities in the proofs.

The assumptions in the above theorem are quite tight (for instance, it can be shown that they are necessary for the limits (3.38) and (3.39)), but we will give as a corollary a slicker set of (stronger) sufficient conditions, more adapted to the quantile process case. We introduce a convenient definition by setting

$$
(K_n \circ K_n)(s, t) := \int_0^1 K_n(s, u) K_n(t, u) \mathrm{d}u = \sum_{i,j} \langle c_{n,i}, c_{n,j} \rangle c_{n,i} \otimes c_{n,j}.
$$

It can be easily checked that $\|K_n \circ K_n\|_2^2 = \sum_{j,k} \left(\sum_i c_{n,i,j} c_{n,i,k}\right)^2$ and also that

$$
\|K_n \circ K_n\|_2^2 = \int_0^1 \int_0^1 \int_0^1 \int_0^1 K_n(s, t) K_n(u, v) K_n(s, u) K_n(t, v) \mathrm{d}s\, \mathrm{d}t\, \mathrm{d}u\, \mathrm{d}v.
$$

Corollary 3.11. *If*

$$\sum_{i=1}^{n}\langle c_{n,i},\, m_n\rangle^2 \to 0, \quad \sum_{i=1}^{n}\|c_{n,i}\|_2^4 \to 0 \quad \text{and} \quad \|K_n \circ K_n\|_2 \to 0, \tag{3.40}$$

$$K_n^*(s,\, t) := \sum_i |c_{n,i}(s)c_{n,i}(t)| \leqslant CK_n(s,\, t) \tag{3.41}$$

for some absolute constant $C < \infty$, and

$$\|K_n\|_2^2 \to \sigma^2/2, \tag{3.42}$$

then (3.37) *holds.*

Proof. Conditions (3.34) and (3.35) obviously hold and only (3.36) requires verification. First, we note that

$$\begin{aligned}
&\sum_j\Bigg(\sum_{i:i>j}\langle c_{n,i},\, c_{n,j}\rangle\langle c_{n,i},\, c_{n,i}+m_n+c_{n,j}\rangle\Bigg)^2\\
&\quad\leqslant 2\sum_j\Bigg(\sum_{i:i>j}\langle c_{n,i},\, c_{n,j}\rangle^2\Bigg)^2+2\sum_j\Bigg(\sum_{i:i>j}\langle c_{n,i},\, c_{n,j}\rangle\langle c_{n,i},\, c_{n,i}+m_n\rangle\Bigg)^2\\
&\quad\leqslant 2\sum_{j,k}\Bigg(\sum_{i:i>j\vee k}\langle c_{n,i},\, c_{n,j}\rangle\langle c_{n,i},\, c_{n,k}\rangle\Bigg)^2+2\sum_j\Bigg(\sum_{i:i>j}\langle c_{n,i},\, c_{n,j}\rangle^2\Bigg)\Bigg(\sum_{i:i>j}\langle c_{n,i},\, c_{n,i}+m_n\rangle^2\Bigg)\\
&\quad\leqslant 2\sum_{j,k}\Bigg(\sum_{i:i>j\vee k}\langle c_{n,i},\, c_{n,j}\rangle\langle c_{n,i},\, c_{n,k}\rangle\Bigg)^2+2\|K_n\|_2^2\sum_i\langle c_{n,i},\, c_{n,i}+m_n\rangle^2,
\end{aligned}$$

and observe that the second term on the right of the last inequality tends to zero by (3.42) and the first two limits in (3.40), whereas the first term tends to zero as a consequence of (3.41) and the third limit in (3.40):

$$\begin{aligned}
&\sum_{j,k}\Bigg(\sum_{i:i>j\vee k}\langle c_{n,i},\, c_{n,j}\rangle\langle c_{n,i},\, c_{n,k}\rangle\Bigg)^2\leqslant\sum_{j,k}\Bigg(\sum_i|\langle c_{n,i},\, c_{n,j}\rangle\langle c_{n,i},\, c_{n,k}\rangle|\Bigg)^2\\
&\quad\leqslant\sum_{j,k}\langle K_n^*,\, |c_{n,j}|\otimes|c_{n,k}|\rangle^2\leqslant\langle K_n^*\otimes K_n^*,\, \sum_{j,k}\big(|c_{n,j}|\otimes|c_{n,k}|\big)\otimes\big(|c_{n,j}|\otimes|c_{n,k}|\big)\rangle\\
&\quad=\iiiint K_n^*(s,\, t)K_n^*(u,\, v)K_n^*(s,\, u)K_n^*(t,\, v)\mathrm{d}s\,\mathrm{d}t\mathrm{d}u\mathrm{d}v\\
&\quad\leqslant C^4\|K_n\circ K_n\|_2^2\to 0.
\end{aligned}$$

So, (3.36) holds and the corollary follows from Theorem 3.11. □

Note that if $\|K_n \circ K_n\|_2 \to 0$ we cannot have $K_n \to K$ in L_2 unless $K = 0$.

Remark 3.2. The results on convergence or shift convergence in law of $\|Y_n\|_2^2$ derived so far in this paper assume infinitesimality on the coefficients $c_{n,i}$ ($\max_i \|c_{n,i}\| \to 0$ in probability). Of course, if conditions of this type are removed, other asymptotic distributions can be obtained. It is straightforward to see, for instance, that if

$$\sum_{1 \leqslant i,j \leqslant n} (c_{n,i,j} - \gamma_{i,j})^2 \to 0, \tag{3.43}$$

for some real numbers $\{\gamma_{i,j}\}$ satisfying $\sum_{i,j} \gamma_{i,j}^2 < \infty$, then

$$\begin{aligned}\|Y_n - \mathrm{E}Y_n\|_2^2 - \mathrm{E}\|Y_n - \mathrm{E}Y_n\|_2^2 &= \sum_{i,j=1}^{n} c_{n,i,j}\left[(\xi_i - 1)(\xi_j - 1) - \delta_{i,j}\right] \\ &\underset{L_2}{\to} \sum_{i,j=1}^{\infty} \gamma_{i,j}\left[(\xi_i - 1)(\xi_j - 1) - \delta_{i,j}\right].\end{aligned}$$

Note that the limiting random variable is well defined because the condition $\sum_{i,j} \gamma_{i,j}^2 < \infty$ implies that the associated partial sums are L_2 convergent. If, further,

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{n} c_{n,i,j} - \beta_i \right)^2 \to 0, \tag{3.44}$$

for some real numbers β_i such that $\sum_{i=1}^{\infty} \beta_i^2 < \infty$, then we also have that

$$\|Y_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2 \underset{L_2}{\longrightarrow} \sum_{i,j=1}^{\infty} \gamma_{i,j}\left[(\xi_i - 1)(\xi_j - 1) - \delta_{i,j}\right] + 2\sum_{i=1}^{\infty} \beta_i(\xi_i - 1). \tag{3.45}$$

3.4 Shift convergence of $\|Z_n\|_2^2$

Y_n can be replaced by Z_n in Theorem 3.5 as an immediate consequence of the law of large numbers, while it can be replaced in Theorem 3.9 and Corollary 3.11 because of the following proposition.

Proposition 3.12. *Suppose* $\|Y_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2$ *converges in law. Then* $\|Z_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2$ *converges in law to the same limit if and only if*

$$\frac{\mathrm{E}\|Y_n\|_2^2}{\sqrt{n}} \to 0. \tag{3.46}$$

In particular, this condition is satisfied if

$$\frac{\sum_i c_{n,i,i}}{\sqrt{n}} \to 0 \quad \text{and} \quad \sum_i \langle c_{n,i}, m_n \rangle^2 \to 0. \tag{3.47}$$

If (3.46) *holds, we also have* $\langle Z_n - \mathrm{E}Y_n, h\rangle - \langle Y_n - \mathrm{E}Y_n, h\rangle \to 0$ *in probability for any* $h \in L_2(0, 1)$.

Proof. Since

$$\|Z_n\|_2^2 - \|Y_n\|_2^2 = \left(\frac{n-1}{S_n}\right)^2 \left(1 + \frac{S_n}{n-1}\right)\left(1 - \frac{S_n}{n-1}\right)\|Y_n\|_2^2 = O_P\left(n^{-1/2}\right)\|Y_n\|_2^2,$$

by the central limit theorem and the law of large numbers, the necessity and sufficiency of condition (3.46) follow from Lemmas 2.2 and 3.1. Now, by (3.8) and Cauchy–Schwarz,

$$\frac{1}{\sqrt{n}}\mathrm{E}\|Y_n\|_2^2 = \frac{1}{\sqrt{n}}\left|\sum_i\left(c_{i,i} + \sum_j c_{i,j}\right)\right| \leqslant \frac{1}{\sqrt{n}}\sum_i c_{i,i} + \sum_i\left\langle c_{n,i}, \sum_j c_{n,j}\right\rangle^2,$$

which gives the sufficiency of (3.47). The last statement follows because, by (3.46) and the law of large numbers, $\|Z_n - Y_n\|_2^2 \to 0$ in probability. □

4. Convergence in law of weighted L_2 functionals of the quantile process

At the risk of overburdening the reader, we distinguish between the uniform and the general quantile processes.

4.1. The uniform quantile process

Recall from the beginning of Section 2 that if u_n is the uniform quantile process, then

$$L_n := \int_{1/n}^{1-1/n}\left(\frac{u_n(t)}{g(t)}\right)^2 \mathrm{d}t \stackrel{d}{=} \left(\frac{n}{S_{n+1}}\right)^2 \left\|\sum_{i=1}^{n+1} c_{n,i}\xi_i\right\|_2^2,$$

where $c_{n,i} = n^{-1/2}a_{n,i}(t)I_{[1/n,1-1/n]}(t)/g(t)$ and the $a_{n,i}$ are as defined in (2.3). With the help of Lemma 2.1, the results of Section 3 can easily be specialized to this situation.

(a) *The infinitesimality condition* (3.12). It follows from the definitions that

$$\frac{1}{2n}\int_{1/n}^{1-1/n}\frac{t^2 + (1-t)^2}{g^2(t)}\mathrm{d}t \leqslant \max_i\|c_{n,i}\|_2^2 \leqslant \frac{1}{n}\int_{1/n}^{1-1/n}\frac{1}{g^2(t)}\mathrm{d}t,$$

and from this we conclude that condition (3.12) is equivalent to

$$\frac{1}{n}\int_{1/n}^{1-1/n}\frac{1}{g^2(t)}\mathrm{d}t \to 0. \tag{4.1}$$

(b) *Convergence of K_n and definition of K.* Also from the definitions (in (3.3) and Lemma 2.1), we have

$$K_n(s,\ t) = \frac{1}{n}\frac{\tilde{K}_n(s,\ t)}{g(s)g(t)} I\{1/n \leq s,\ t \leq 1 - 1/n\},$$

so that, by Lemma 2.1 (iii),

$$K_n(s,\ t) \to K(s,\ t) := \frac{(s \wedge t - st)}{g(s)g(t)} \tag{4.2}$$

pointwise, hence, by Lemma 2.1 (iii) and dominated convergence, $K_n \to_{L_2} K$ if and only if $K \in L_2((0,\ 1) \times (0,\ 1))$, if and only if

$$\int_0^1 \int_0^1 \frac{(s \wedge t - st)^2}{g^2(s)g^2(t)} \mathrm{d}s\,\mathrm{d}t < \infty. \tag{4.3}$$

Next we see that the limiting kernel K is trace-class and the limit (3.14) holds if and only if

$$\int_0^1 \frac{t(1-t)}{g^2(t)} \mathrm{d}t < \infty. \tag{4.4}$$

In fact, by Lemma 2.1 (iii),

$$\sum_{i=1}^{n+1} \|c_{n,i}\|_2^2 = \frac{1}{n}\int_{1/n}^{1-1/n} \frac{\tilde{K}_n(t,\ t)}{g^2(t)} \mathrm{d}t \to \int_0^1 \frac{t(1-t)}{g^2(t)} \mathrm{d}t$$

regardless of whether the limiting integral is finite or not. Thus (4.4) is necessary in order to obtain a finite limit in (3.14). On the other hand, if (4.4) holds and $B_g(t) = B(t)/g(t)$, where $B(t)$, $0 < t < 1$, is a Brownian bridge, then B_g is a centred, $L_2(0,\ 1)$-valued Gaussian process with covariance function $K(s,\ t)$. Thus, if λ_i and ϕ_i denote the eigenvalues and eigenfunctions, respectively, of the kernel K, then

$$\begin{aligned}\int_0^1 \frac{t(1-t)}{g^2(t)} \mathrm{d}t &= \int_0^1 \mathrm{E}B_g(t)^2\,\mathrm{d}t = \mathrm{E}\int_0^1 \frac{B^2(t)}{g^2(t)} \mathrm{d}t = \mathrm{E}\sum_{i=1}^{\infty}\left(\int_0^1 \frac{B(t)}{g(t)}\phi_i(t)\mathrm{d}t\right)^2 \\ &= \sum_{i=1}^{\infty} \mathrm{E}\left(\int_0^1 \frac{B(t)}{g(t)}\phi_i(t)\mathrm{d}t\right)^2 = \sum_{i=1}^{\infty}\int_0^1\int_0^1 \frac{s \wedge t - st}{g(s)g(t)}\phi_i(s)\phi_i(t)\mathrm{d}s\,\mathrm{d}t \\ &= \sum_{i=1}^{\infty} \lambda_i < \infty.\end{aligned}$$

Hence, if (4.4) holds then K is trace-class and (3.14) holds.

(c) *Convergence of m_n to $m = 0$ assuming infinitesimality.* If the infinitesimality condition (4.1) holds, then

$$\left\| \sum_{i=1}^{n+1} c_{n,i} \right\|_2^2 = \frac{1}{n}\int_0^1 \frac{\tilde{m}_n(t)^2}{g^2(t)} \mathrm{d}t \leq \frac{1}{n}\int_0^1 \frac{1}{g^2(t)} \mathrm{d}t \to 0,$$

showing that $m_n \to 0$ in L_2.

Finally, note that condition (4.4) implies condition (4.1) by dominated convergence, and condition (4.3) because, since $(s \wedge t - st)^2 \leqslant s(1-s)t(1-t)$, we have

$$\int_0^1\int_0^1 \frac{(s \wedge t - st)^2}{g^2(s)g^2(t)} \mathrm{d}s\,\mathrm{d}t \leqslant \int_0^1\int_0^1 \frac{s(1-s)t(1-t)}{g^2(s)g^2(t)} \mathrm{d}s\,\mathrm{d}t = \left(\int_0^1 \frac{t(1-t)}{g^2(t)}\mathrm{d}t\right)^2.$$

Summarizing, Theorem 3.5 and the law of large numbers for S_n/n give the following:

Theorem 4.1. *Let $u_n(g)$ denote the weighted uniform quantile process, that is, $u_n(g)(t) = (u_n(t)/g(t))I\{1/n \leqslant t \leqslant 1-1/n\}$, $0 < t < 1$, where g is a non-zero measurable function. Assume condition* (4.1)*. Then the sequence of processes $\{u_n(g)\}$ is weakly convergent in $L_2(0, 1)$ to a non-degenerate limit if and only if*

$$\int_0^1 \frac{t(1-t)}{g^2(t)}\mathrm{d}t < \infty. \tag{4.4}$$

In this case,

$$u_n(g) \underset{d}{\to} B_g$$

in $L_2(0, 1)$, where $B_g(t) = B(t)/g(t)$ and B is a Brownian bridge. In particular, if h_ℓ, $\ell = 1, \ldots, r$, are functions in $L_2(0, 1)$, then

$$\int_{1/n}^{1-1/n} \frac{u_n^2(t)}{g^2(t)}\mathrm{d}t - \sum_{\ell=1}^r \left(\int_{1/n}^{1-1/n} \frac{u_n(t)h_\ell(t)}{g(t)}\mathrm{d}t\right)^2 \underset{d}{\to} \int_0^1 \frac{B^2(t)}{g^2(t)}\mathrm{d}t - \sum_{\ell=1}^r \left(\int_{1/n}^{1-1/n} \frac{B(t)h_\ell(t)}{g(t)}\mathrm{d}t\right)^2.$$

Only the necessity part of this theorem may be considered new; the sufficiency is well known (see, Mason 1984; Csörgő and Horváth 1988; 1993, p. 354).

Since, under infinitesimality, $m_n \to 0$ in L_2, we have that the analogue of Theorem 3.9 for $Y_n = (S_{n+1}/n)u_n$ holds under conditions (4.1) and (4.3). In order to get rid of the factor S_{n+1}/n, according to Proposition 3.12 we must have $\mathrm{E}\|Y_n\|_2^2/\sqrt{n} \to 0$, which follows from conditions (3.47). Now, if conditions (4.1) and (4.3) hold, then so do conditions (3.47): the second condition in (3.47) is obvious because $m_n \to 0$ in L_2 and $\sup_n\|K_n\|_2 < \infty$ (as K_n converges in L_2), so

$$\sum_i \langle c_{n,i}, m_n\rangle^2 = \langle K_n, m_n \otimes m_n\rangle \leqslant \|K_n\|_2\|m_n\|_2^2 \to 0,$$

and the first folows because, by Lemma 2.1 (iii),

$$\frac{1}{\sqrt{n}}\sum_i \langle c_{n,k}, c_{n,k}\rangle = \frac{1}{\sqrt{n}}\int_{1/n}^{1-1/n} \frac{\tilde{K}_n(t, t)}{ng^2(t)}\mathrm{d}t \leqslant \frac{3}{\sqrt{n}}\int_{1/n}^{1-1/n} \frac{t(1-t)}{g^2(t)}\mathrm{d}t,$$

and it is easy to see that this last expression tends to zero as a consequence of (4.1) (divide the domain of integration at the points $1/\sqrt{n}$, $1/2$ and $1 - 1/\sqrt{n}$). Hence, Theorem 3.9 and Proposition 3.12 together give:

Theorem 4.2. *Let* h_ℓ, $\ell = 1, \ldots, r$, *be square-integrable functions on* $(0, 1)$. *If conditions* (4.1) *and* (4.3) *hold, then*

$$\int_{1/n}^{1-1/n} \frac{u_n^2(t)}{g(t)^2}\,\mathrm{d}t - \int_{1/n}^{1-1/n} \frac{t(1-t)}{g^2(t)}\,\mathrm{d}t - \sum_{\ell=1}^{r}\left(\int_{1/n}^{1-1/n} \frac{u_n(t)h_\ell(t)}{g(t)}\,\mathrm{d}t\right)^2$$

$$\underset{d}{\rightarrow} \int_0^1 \frac{B^2(t) - \mathrm{E}B^2(t)}{g^2(t)}\,\mathrm{d}t - \sum_{\ell=1}^{r}\left(\int_{1/n}^{1-1/n} \frac{B(t)h_\ell(t)}{g(t)}\,\mathrm{d}t\right)^2, \tag{4.5}$$

where the integral of $(B^2 - \mathrm{E}B^2)/g^2$ *is defined in a limiting* L_2 *sense.*

Proof. By Theorem 3.9 and the above observations, it only remains to show that we can actually replace $\mathrm{E}\|Y_n\|_2^2$ by the centring constants in (4.5), and that $\int_{1/n}^{1-1/n} \mathrm{E}Y_n(t)h_\ell(t)\mathrm{d}t \to 0$ for all ℓ. By (4.1) and Lemma 2.1, we have

$$\left|\mathrm{E}\|Y_n\|_2^2 - \int_{1/n}^{1-1/n} \frac{t(1-t)}{g^2(t)}\,\mathrm{d}t\right| \leqslant \frac{1}{n}\int_{1/n}^{1-1/n} \frac{|nt(1-t) - \tilde{K}_n(t, t) - \tilde{m}_n^2(t, t)|}{g^2(t)}\,\mathrm{d}t$$

$$\leqslant \frac{4}{n}\int_{1/n}^{1-1/n} \frac{1}{g^2(t)}\,\mathrm{d}t \to 0.$$

And, since as shown above, $\|\mathrm{E}Y_n\|_2^2/\sqrt{n} \to 0$, and since $h_\ell \in L_2(0, 1)$, we obviously have that $\int_{1/n}^{1-1/n} \mathrm{E}Y_n(t)h_\ell(t)\mathrm{d}t \to 0$. □

For $g(t) = \phi(\Phi^{-1}(t))$, where ϕ and Φ denote the standard normal density and distribution function respectively, this result goes back, in one form or other, to de Wet and Venter (1972), but it seems to be new in the generality in which it is given here. See also Gregory (1977) and del Barrio *et al.* (1999a).

Next we examine the normal convergence case (as a consequence of Corollary 3.11). We will further relax integrability (so condition (4.3) will not hold), but, for convenience, will impose regular variation of g at at least one of the end-points 0 and at 1. Standard use of the basic properties of regular variation shows that, if g is regularly varying at 0 and at 1 with exponent α, then the hypotheses of Theorem 4.2, (4.1) and (4.3), both hold for $\alpha < 1$ and fail if $\alpha > 1$. We now study the borderline case in which $\alpha = 1$ and (4.3) fails, that is,

$$L(x) := 2\int_x^{1-x}\int_x^{1-x} \frac{(s \wedge t - st)^2}{g^2(s)g^2(t)}\,\mathrm{d}s\,\mathrm{d}t \to \infty \qquad \text{as } x \to 0. \tag{4.6}$$

This case will fall within the scope of Corollary 3.11 and we will obtain normal convergence. Besides the function $L(x)$ just defined, it is convenient to introduce two more functions,

$$M(x) = 2\int_x^{1-x}\int_x^{1-x} \frac{(s \wedge t - st)}{g^2(s)g^2(t)}\,\mathrm{d}s\,\mathrm{d}t$$

and

$$R(x) = \int_x^{1-x}\int_x^{1-x}\int_x^{1-x}\int_x^{1-x} \frac{(s\wedge t - st)(s\wedge u - su)(t\wedge v - tv)(u\wedge v - uv)}{g^2(s)g^2(t)g^2(u)g^2(v)}\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v,$$

and establish their relationship with L. The next lemma is an exercise in regular variation and L'Hôpital's rule.

Lemma 4.3. *Assume that $g > 0$ is regularly varying at 0 and $\lim_{x\to 0} g(x)/g(1-x) = c \in [0, \infty)$, or that $g > 0$ is regularly varying at 1 and $\lim_{x\to 0} g(1-x)/g(x) = c \in [0, \infty)$. Assume also that $L(x) \to \infty$ as $x \to 0$. Then*

$$\lim_{x\to 0} \frac{xM(x)}{L(x)} = 0 \tag{4.7}$$

and

$$\lim_{x\to 0} \frac{R(x)}{L^2(x)} = 0. \tag{4.8}$$

Proof. For notational convenience we prove this lemma only under the assumption that g is symmetric about $\frac{1}{2}$ (the general case can be proved in the same way). To prove (4.7), it suffices to realize (using symmetry of g) that

$$L'(x) = -\frac{8x^2}{g^2(x)}\int_x^{1-x} \frac{s^2}{g^2(s)}\mathrm{d}s \quad \text{and} \quad M'(x) = -\frac{8x}{g^2(x)}\int_x^{1-x} \frac{s}{g^2(s)}\mathrm{d}s,$$

that, since $L(x) \to \infty$ as $x \to 0$, we necessarily have that $\int_0^1 t(1-t)/g^2(t)\mathrm{d}t = \infty$ and that this fact and regular variation imply the following asymptotic equivalences:

$$\int_x^{1-x} \frac{s^2}{g^2(s)}\mathrm{d}s \sim \int_x^{1-x} \frac{s}{g^2(s)}\mathrm{d}s \sim \int_x^{\delta} \frac{1}{g^2(s)}\mathrm{d}s \sim \frac{x}{g^2(x)}. \tag{4.9}$$

From (4.9) we obtain that

$$L'(x) \sim -\frac{8x^3}{g^4(x)} \quad \text{and} \quad M'(x) \sim -\frac{8x^2}{g^4(x)} \tag{4.10}$$

as $x \to 0$. Regular variation and (4.9) also imply that

$$\begin{aligned} M(x) &= 8\int_x^{1/2} \frac{t}{g^2(t)}\int_t^{1-t} \frac{s}{g^2(s)}\mathrm{d}s\,\mathrm{d}t \sim 8\int_x^{\delta} \frac{t}{g^2(t)}\int_t^{\delta} \frac{1}{g^2(s)}\mathrm{d}s\,\mathrm{d}t \\ &\sim 8\int_x^{\delta} \frac{t^2}{g^4(t)}\mathrm{d}t \sim \frac{8x^3}{g^4(x)}. \end{aligned} \tag{4.11}$$

Now, (4.10), (4.11) and L'Hôpital's rule show that

$$\lim_{x\to 0} \frac{xM(x)}{L(x)} = \lim_{x\to 0} \frac{xM'(x)}{L'(x)} + \lim_{x\to 0} \frac{M(x)}{L'(x)} = 1 - 1 = 0,$$

which proves (4.7).

Let us now consider (4.8). For any ordering of the variables $s < t < u < v$ in the multiple integral, the smallest, s, appears as s^2, the largest, v, as $(1-v)^2$ and the other two as $t(1-t)$ and $u(1-u)$. Thus we have

$$R(x) = 24\int_x^{1-x} \frac{(1-v)^2}{g^2(v)} \int_x^v \frac{u(1-u)}{g^2(u)} \int_x^u \frac{t(1-t)}{g^2(t)} \int_x^t \frac{s^2}{g^2(s)} \mathrm{d}s\,\mathrm{d}t\,\mathrm{d}u\,\mathrm{d}v$$

and

$$R'(x) = -48\frac{x^2}{g^2(x)}\int_x^{1-x} \frac{u(1-u)}{g^2(u)} \int_x^u \frac{t(1-t)}{g^2(t)} \int_x^t \frac{s^2}{g^2(s)} \mathrm{d}s\,\mathrm{d}t\,\mathrm{d}u =: -48\frac{x^2}{g^2(x)} R_1(x).$$

We now see that

$$R_1'(x) = -\frac{x(1-x)}{g^2(x)}\int_x^{1-x} \frac{t(1-t)}{g^2(t)} \int_x^t \frac{s^2}{g^2(s)} \mathrm{d}s\,\mathrm{d}t - \frac{x^2}{g^2(x)}\int_x^{1-x} \frac{t(1-t)}{g^2(t)} \int_x^t \frac{s(1-s)}{g^2(s)} \mathrm{d}s\,\mathrm{d}t$$

$$=: -\frac{x(1-x)}{g^2(x)} R_{1,1}(x) - \frac{x^2}{g^2(x)} R_{1,2}(x).$$

By regular variation,

$$R_{1,1}'(x) = -\frac{x(1-x)}{g^2(x)}\int_x^{1-x} \frac{s^2}{g^2(s)}\mathrm{d}s - \frac{x^2}{g^2(x)}\int_x^{1-x} \frac{s(1-s)}{g^2(s)}\mathrm{d}s \sim \frac{-x}{g^2(x)}\int_x^{\delta} \frac{1}{g^2(s)}\mathrm{d}s \sim \frac{-x^2}{g^4(x)}.$$

$R_{1,2}'$ can be estimated in a similar way, yielding

$$\lim_{x\to 0} R_{1,2}(x)/R_{1,1}(x) = \lim_{x\to 0} R_{1,2}'(x)/R_{1,1}'(x) = 0.$$

Thus,

$$R_1'(x) \sim \frac{-x}{g^2(x)} R_{1,1}(x) \sim \frac{-x}{g^2(x)}\int_x^{\delta} R_{1,1}'(t)\mathrm{d}t \sim \frac{x}{g^2(x)}\int_x^{\delta} \frac{t^2}{g^4(t)}\mathrm{d}t \sim \frac{x^4}{g^6(x)} \tag{4.12}$$

(here the last equivalence is, again, a consequence of regular variation). Observe now that, by (4.9) and (4.10),

$$\frac{R'(x)}{(L^2(x))'} = \frac{R'(x)}{2L(x)L'(x)} \sim 3\frac{R_1(x)}{\int_x^{\delta}(1/g^2(s))\mathrm{d}s\,L(x)}$$

and, therefore, L'Hôpital's rule reduces the proof to showing that

$$\frac{L(x)(1/g^2(x)) - \int_x^{\delta}(1/g^2(s))\mathrm{d}s\,L'(x)}{R_1'(x)} \to \infty.$$

Now (4.9), (4.10) and (4.12) show that $(\int_x^{\delta}(1/g^2(s))\mathrm{d}s\,L'(x))/R_1'(x) \to 8$ and further reduce the proof to checking that $(x^2(\int_x^{\delta} g(u)^{-2}\mathrm{d}u)^2)/L(x) \to 0$. But this follows from L'Hôpital's rule and the following chain of equivalences (where (4.9) is used twice):

$$\lim_{x\to 0}\frac{x^2\left(\int_x^\delta (1/g(u)^2)\mathrm{d}u\right)^2}{L(x)} = \lim_{x\to 0}\frac{\left(2x\int_x^\delta (1/g^2(u))\mathrm{d}u - 2x^2/g^2(x)\right)\int_x^\delta (1/g^2(u))\mathrm{d}u}{L'(x)}$$

$$= \frac{1}{4}\lim_{x\to 0}\frac{x/g^2(x) - \int_x^\delta (1/g^2(u))\mathrm{d}u}{x/g^2(x)}$$

$$= \frac{1}{4}\left(1 - \lim_{x\to 0}\frac{\int_x^\delta (1/g^2(u))\mathrm{d}u}{x/g^2(x)}\right) = 0. \tag{4.13}$$

□

Theorem 4.4. *Assume that $g > 0$ is regularly varying at 0 and $\lim_{x\to 0} g(x)/g(1-x) = c \in [0, \infty)$, or that $g > 0$ is regularly varying at 1 and $\lim_{x\to 0} g(1-x)/g(x) = c \in [0, \infty)$. Assume also that $L(x) \to \infty$ as $x \to 0$. Let Z denote a standard normal random variable. Then*

$$\frac{1}{\sqrt{L(1/n)}}\left(\int_{1/n}^{1-1/n}\frac{u_n^2(t)}{g^2(t)}\mathrm{d}t - \int_{1/n}^{1-1/n}\frac{t(1-t)}{g^2(t)}\mathrm{d}t\right)\underset{d}{\to} Z.$$

Proof. As in the previous lemma, we only consider the case where g is symmetric about $\frac{1}{2}$. We can apply Corollary 3.11 with

$$c_{n,i}(t) = \frac{1}{\sqrt{nL^{1/2}(1/n)}}\frac{a_{n,i}(t)}{g(t)} I\{1/n \leqslant t \leqslant 1 - 1/n\}.$$

Now we have that

$$2\|K_n\|_2^2 = \frac{2}{n^2 L(1/n)}\int_{1/n}^{1-1/n}\int_{1/n}^{1-1/n}\frac{\tilde{K}_n^2(s,\ t)}{g^2(s)g^2(t)}\mathrm{d}s\,\mathrm{d}t$$

and

$$\sum_{i=1}^{n+1}\|c_{n,i}\|_2^4 = \frac{2}{n^2 L(1/n)}\int_{1/n}^{1-1/n}\int_{1/n}^{1-1/n}\frac{\sum_{i=1}^{n+1} a_{n,i}^2(s)a_{n,i}^2(t)}{g^2(s)g^2(t)}\mathrm{d}s\,\mathrm{d}t.$$

We claim that

$$2\|K_n\|_2^2 \to 1, \quad \sum_{i=1}^{n+1}\|c_{n,i}\|_2^4 \to 0 \quad \text{and} \quad \sum_{i=1}^{n+1}\langle c_{n,i},\ c_{n,i} + m_n\rangle^2 \to 0. \tag{4.14}$$

In fact, from Lemma 2.1 we obtain that $\sum_{i=1}^{n+1} a_{n,i}^2(s)a_{n,i}^2(t) \leqslant 3n(s \wedge t - st)$, which, by Lemma 4.3, implies that

$$\sum_{i=1}^{n+1}\|c_{n,i}\|_2^4 \leqslant 6\frac{(1/n)M(1/n)}{L(1/n)} \to 0$$

and proves the second part of (4.14). The first part can be obtained using parts (ii) and (iii) of Lemma 2.1 to see that

$$|\tilde{K}_n(s, t)^2 - n^2(s \wedge t - st)^2| = |\tilde{K}_n(s, t) + n(s \wedge t - st)||\tilde{K}_n(s, t) - n(s \wedge t - st)|$$
$$\leqslant 8n(s \wedge t - st)$$

and, consequently, that

$$|2\|K_n\|_2^2 - 1| = \frac{2}{n^2 L(1/n)} \left| \int_{1/n}^{1-1/n} \int_{1/n}^{1-1/n} \frac{\tilde{K}_n^2(s, t) - n^2(s \wedge t - st)^2}{g^2(s) g^2(t)} \mathrm{d}s\, \mathrm{d}t \right|$$
$$\leqslant 16 \frac{(1/n)M(1/n)}{L(1/n)} \to 0.$$

Finally, the third part of claim (4.14) is a consequence of

$$\sum_{i=1}^{n+1} \langle c_{n,i}, m_n \rangle^2 = \langle K_n, m_n \otimes m_n \rangle$$
$$= \frac{1}{n^2 L(1/n)} \int_{1/n}^{1-1/n} \int_{1/n}^{1-1/n} \frac{\tilde{K}_n(s, t) \tilde{m}_n(s) \tilde{m}_n(t)}{g^2(s) g^2(t)} \mathrm{d}s\, \mathrm{d}t \leqslant \frac{(1/n)M(1/n)}{L(1/n)} \to 0,$$

since $(a+b)^2 \leqslant 2a^2 + 2b^2$. The limits (4.14) prove the first two limits in (3.40) and the limit in (3.42) (with $\sigma^2 = 1$). Lemma 2.1(iii) gives that (3.41) is also satisfied (with $C = 6$). Finally, the third limit in (3.40) follows from Lemma 4.3 since

$$\|K_n \circ K_n\|_2^2 \leqslant \frac{81 R(1/n)}{n^4 L^2(1/n)} \to 0.$$

Corollary 3.11 now implies that $\|Y_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2 \to_w N(0, 1)$. Conditions (3.47) from Proposition 3.12 are also satisfied because of the last two limits in (4.14) (see the argument immediately before Theorem 4.2) and therefore we also have $\|Z_n\|_2^2 - \mathrm{E}\|Y_n\|_2^2 \to_w N(0, 1)$. Now all that is left to show is that we can replace $\mathrm{E}\|Y_n\|_2^2$ by $L^{-1/2}(1/n) \int_{1/n}^{1-1/n} t(1-t) g^{-2}(t) \mathrm{d}t$ as centring constants. Arguing as in the proof of Theorem 4.2, we see that

$$\left| \mathrm{E}\|Y_n\|_2^2 - \frac{1}{L^{1/2}(1/n)} \int_{1/n}^{1-1/n} \frac{t(1-t)}{g^2(t)} \mathrm{d}t \right| \leqslant \frac{4}{n L^{1/2}(1/n)} \int_{1/n}^{1-1/n} \frac{1}{g^2(t)} \mathrm{d}t \to 0,$$

where the last limit is a consequence of (3.44). □

This result is new in the present generality. For weights related to the exponential distribution and to the Weibull distribution with exponent $0 < \alpha \leqslant \frac{4}{3}$, see respectively Csörgő (2000) and Csörgő (2002). For results similar to those in the last two theorems, but for the L_1 norm of the empirical process instead, see del Barrio (1999b).

Finally, we consider smaller functions g, corresponding in some way to Remark 3.2. The following result is only given for completeness and only symmetric weights are considered in the proof since the one-sided analogue is already contained in Csörgő and Horváth (1988).

Proposition 4.5. *Let g be a positive function on $(0, 1)$, regularly varying at 0 and at 1 with exponent $\alpha > 1$ and with equal or smaller exponent at the other extreme, and such that $\lim_{x\to 0} g(x)/g(1-x) := c \in [0, \infty]$. Set*

$$\mathrm{E}(x) := \int_x^{1-x} \frac{t(1-t)}{g^2(t)} \mathrm{d}t.$$

Then

$$\frac{1}{\mathrm{E}(1/n)} \int_{1/n}^{1-1/n} \frac{u_n^2(t)}{g^2(t)} \mathrm{d}t \underset{d}{\rightarrow} \frac{1}{\alpha - 1} \left(\frac{c^2}{1+c^2} \int_1^\infty \frac{(S_{[y]}^{(1)} - y)^2}{y^{2\alpha}} \mathrm{d}y + \frac{1}{1+c^2} \int_1^\infty \frac{(S_{[y]}^{(2)} - y)^2}{y^{2\alpha}} \mathrm{d}y \right), \tag{4.15}$$

where $\{S_{[y]}^{(1)} : y \geqslant 1\}$ is the partial sum process associated with the sequence $\{\xi_j\}$ of independent exponential random variables, that is, $S_{[y]}^{(1)} = \sum_{j=1}^{[y]} \xi_j$, and $\{S_{[y]}^{(2)}\}$ is an independent copy of $\{S_{[y]}^{(1)}\}$.

Proof. As above, we only consider the case where g is symmetric. Symmetry of g and the fact that $a_{n,j}(1-t) = -a_{n,n+2-j}(t)$ show that

$$\begin{aligned}\frac{1}{\mathrm{E}(1/n)} \int_{1/n}^{1-1/n} \frac{u_n^2(t)}{g^2(t)} \mathrm{d}t &= \left(\frac{n}{S_{n+1}}\right)^2 \frac{1}{n\mathrm{E}(1/n)} \int_{1/n}^{1-1/n} \frac{\left(\sum_{j=1}^{n+1} a_{n,j}(t)\xi_j\right)^2}{g^2(t)} \mathrm{d}t \\ &= \left(\frac{n}{S_{n+1}}\right)^2 (V_n^{(1)} + V_n^{(2)}),\end{aligned}$$

where

$$V_n^{(1)} = \frac{1}{nE(1/n)} \int_{1/n}^{1/2} \frac{\left(\sum_{j=1}^{n+1} a_{n,j}(t)\xi_j\right)^2}{g^2(t)} \mathrm{d}t \quad \text{and} \quad V_n^{(2)} = \frac{1}{n\mathrm{E}(1/n)} \int_{1/n}^{1/2} \frac{\left(\sum_{j=1}^{n+1} a_{n,j}(t)\xi_{n+2-j}\right)^2}{g^2(t)} \mathrm{d}t.$$

We set $b_{n,j}(t) = I\{j - 1 < nt\}$ and define

$$W_n^{(1)} = \frac{1}{n\mathrm{E}(1/n)} \int_{1/n}^{1/2} \frac{\left(\sum_{j=1}^{n+1} b_{n,j}(t)\xi_j - nt\right)^2}{g^2(t)} \mathrm{d}t$$

and, similarly, $W_n^{(2)}$, replacing ξ_j with ξ_{n+2-j}. Now, since

$$|(V_n^{(1)})^{1/2} - (W_n^{(1)})^{1/2}|^2 \leqslant \left(\sqrt{n}\left(1 - \frac{S_{n+1}}{n}\right)\right)^2 \frac{1}{E(1/n)} \int_{1/n}^{1/2} \frac{t^2}{g^2(t)} \mathrm{d}t = O_P(1) \underset{\mathrm{Pr}}{\rightarrow} 0$$

and $V_n^{(1)} = O_P(1)$, we see that $V_n^{(1)} - W_n^{(1)} = o_P(1)$. Analogously, $V_n^{(2)} - W_n^{(2)} = o_P(1)$, showing that (4.15) is equivalent to convergence in law of $W_n^{(1)} + W_n^{(2)}$ to the right-hand side of (4.15). Obviously, $W_n^{(1)} \stackrel{d}{=} W_n^{(2)}$ and, moreover, $W_n^{(1)}$ and $W_n^{(2)}$ are asymptotically independent: they are indeed independent if n is odd since $b_{n,j}(t) = 0$ if $t \leqslant 1/2$ and $j > (n+1)/2$ and therefore $W_n^{(1)}$ depends only on $\xi_1, \ldots, \xi_{(n+1)/2}$ while $W_n^{(2)}$ depends only on $\xi_{(n+3)/2}, \ldots, \xi_{n+1}$; the overlapping that arises if n is even is negligible. Hence, in order to prove (4.15) it suffices to show that

$$W_n^{(1)} \underset{d}{\to} \frac{1}{\alpha - 1} \int_1^\infty \frac{(S_{[y]}^{(1)} - y)^2}{y^{2\alpha}} \, \mathrm{d}y. \tag{4.16}$$

To see this, we note that

$$W_n^{(1)} = \sum_{i,j} c_{n,i,j} (\xi_i - 1)(\xi_j - 1) + 2 \sum_i d_{n,i} (\xi_i - 1) + e_n$$

with

$$c_{n,i,j} = \frac{1}{n^2 \mathrm{E}(1/n)} \int_{i \vee j - 1}^{n/2} \frac{1}{g^2(y/n)} \mathrm{d}y \quad \text{if } i \vee j > 1, \qquad c_{n,1,1} = c_{n,2,2},$$

$$d_{n,i} = \frac{1}{n^2 \mathrm{E}(1/n)} \int_{i-1}^{n/2} \frac{([y] - y)}{g^2(y/n)} \mathrm{d}y \quad \text{if } i > 1, \qquad d_{n,1} = d_{n,2}$$

$$e_n = \frac{1}{n^2 \mathrm{E}(1/n)} \int_1^{n/2} \frac{([y] - y)^2}{g^2(y/n)} \mathrm{d}y.$$

Similarly,

$$\frac{1}{\alpha - 1} \int_1^\infty \frac{(S_{[y]}^{(1)} - y)^2}{y^{2\alpha}} \mathrm{d}y = \sum_{i,j} \gamma_{i,j} (\xi_i - 1)(\xi_j - 1) + 2 \sum_i \delta_i (\xi_i - 1) + \epsilon,$$

where

$$\gamma_{i,j} = \frac{1}{\alpha - 1} \int_{i \vee j - 1}^\infty \frac{1}{y^{2\alpha}} \mathrm{d}y \quad \text{if } i \vee j > 1, \qquad \gamma_{1,1} = \gamma_{2,2},$$

$$\delta_i = \frac{1}{\alpha - 1} \int_{i-1}^\infty \frac{([y] - y)}{y^{2\alpha}} \mathrm{d}y \quad \text{if } i > 1, \qquad \delta_1 = \delta_2,$$

$$\epsilon = \frac{1}{\alpha - 1} \int_1^\infty \frac{([y] - y)^2}{y^{2\alpha}} \mathrm{d}y.$$

Standard regular variation techniques show that $\sum_{i,j} (c_{n,i,j} - \gamma_{i,j})^2 \to 0$, $\sum_i (d_{n,i} - \delta_i)^2 \to 0$ and $e_n \to \epsilon$, yielding as in Remark 3.2 that

$$W_n^{(1)} \underset{L_2}{\to} \frac{1}{\alpha - 1} \int_1^\infty \frac{(S_{[y]}^{(1)} - y)^2}{y^{2\alpha}} \mathrm{d}y$$

and proving (4.16). □

4.2. The general quantile process

By Proposition 2.5, we can transfer the results in the previous section corresponding to Theorems 4.1, 4.2 and 4.4 to the general quantile process just by taking $g(t) = f(F^{-1}(t))/\sqrt{w(t)}$. We will need the following conditions on the cdf F and the weight w:

F is twice differentiable on its open support (a_F, b_F), with $f(x) = F'(x) > 0$, and satisfies conditions (2.7), (2.22) and (2.23).w is a bounded non-negative measurable function on (0, 1) and satisfies conditions (2.24). (GH)

These, together with (2.10), are the conditions under which we can transfer results on u_n to v_n by Proposition 2.5. We exclude (2.10) because it will be subsumed by other conditions – in fact because, by dominated convergence,

$$\int_0^1 \frac{t(1-t)}{f^2(F^{-1}(t))} w(t)\mathrm{d}t < \infty \Rightarrow (2.10) \Rightarrow \frac{1}{n}\int_{1/n}^{1-1/n} \frac{w(t)\mathrm{d}t}{f^2(F^{-1}(t))} \to 0.$$

We then have:

Theorem 4.6. *Let B be a Brownian bridge on* (0, 1) *and let Z be a standard normal random variable.*

(i) *If F and w satisfy* (GH) *and*

$$\int_0^1 \frac{t(1-t)}{f^2(F^{-1}(t))} w(t)\mathrm{d}t < \infty, \tag{4.17}$$

then

$$v_n(t) \to \frac{B(t)\sqrt{w(t)}}{\sqrt{f^2(F^{-1}(t))}} \qquad \text{in law in } L_2(0, 1);$$

in particular,

$$\int_0^1 v_n^2(t) w(t)\mathrm{d}t \to \int_0^1 \frac{B^2(t)}{f^2(F^{-1}(t))} w(t)\mathrm{d}t \qquad \text{in distribution}.$$

(ii) *If F and w satisfy* (GH) *and*

$$\frac{1}{\sqrt{n}}\int_{1/n}^{1-1/n} \frac{t^{1/2}(1-t)^{1/2}}{f^2(F^{-1}(t))} w(t)\mathrm{d}t \to 0 \tag{2.10}$$

and

$$\int_0^1\int_0^1 \frac{(s \wedge t - st)^2}{f^2(F^{-1}(s))f^2(F^{-1}(t))} w(s)w(t)\mathrm{d}s\,\mathrm{d}t < \infty, \tag{4.18}$$

then

$$\int_0^1 v_n^2(t)w(t)\mathrm{d}t - \int_{1/\mathrm{n}}^{1-1/\mathrm{n}} \frac{t(1-t)}{f^2(F^{-1}(t))} w(t)\mathrm{d}t \to \int_0^1 \frac{B^2(t) - EB^2(t)}{f^2(F^{-1}(t))} w(t)\mathrm{d}t \qquad \text{in distribution}.$$

(iii) *Assume F is twice differentiable on its open support (a_F, b_F), with $f(x) = F'(x) > 0$, that F satisfies condition (2.7) and that the function $g := f(F^{-1})$ is either regularly varying with exponent one at 0 and $\lim_{x\to 0} g(x)/g(1-x) = c \in [0, \infty)$, or g is regularly varying with exponent one at 1 and $\lim_{x\to 0} g(1-x)/g(x) = c \in [0, \infty)$. Assume also that*

$$L(x) := 2\int_x^{1-x}\int_x^{1-x} \frac{(s\wedge t - st)^2}{f^2(F^{-1}(s)f^2(F^{-1}(t))} \mathrm{d}s\,\mathrm{d}t \to \infty \tag{4.19}$$

as $x \to 0$. Then

$$\frac{1}{\sqrt{L(1/n)}}\left(\int_0^1 v_n^2(t)\mathrm{d}t - \int_{1/n}^{1-1/n} \frac{t(1-t)}{f^2(F^{-1}(t))}\mathrm{d}t\right) \to Z \quad \textit{in distribution.} \tag{4.20}$$

As in the case of the uniform quantile process, we could also have added terms of the form $\sum_{\ell=1}^{r}(\int_0^1 v_n(t)h_\ell(t)w(t)\mathrm{d}t)^2$ to the limiting results above, with corresponding changes in the limit. However, although we will need this in the next section, the statement is cleaner as it is now.

Proof of Theorem 4.6. By Proposition 2.5 and the remark above on condition (2.10), the statements (i) and (ii) do not require proof. But part (iii) does (Proposition 2.5) does not apply in this case). As usual, we assume $f(F^{-1})$ symmetric about $\frac{1}{2}$. If we can replace $\|v_n\|_2^2$ in (4.20) by $\|u_n/f(F^{-1})\|_{2,n}^2 = \int_{1/n}^{1-1/n} u_n^2(t)/f(F^{-1}(t))\mathrm{d}t$, the result will follow from Theorem 4.4. By the proof of Lemma 2.4, we can replace $\|v_n\|_2^2$ by $\|v_n\|_{2,n}^2$ if we show that

$$\lim_{x\to 0}\frac{x^2}{f^2(F^{-1}(x))\sqrt{L(x)}} = 0 \text{ and } \lim_{x\to 0}\frac{1}{x\sqrt{L(x)}}\int_0^x (F^{-1}(x) - F^{-1}(t))^2\mathrm{d}t = 0,$$

and, by the proof of Lemma 2.3 (see (2.20) and (2.21)), we can replace $\|v_n\|_{2,n}^2$ by $\|u_n/f(F^{-1})\|_{2,n}^2$ if

$$\lim_{n\to\infty}\frac{1}{\sqrt{nL(1/n)}}\int_{1/n}^{1-1/n}\frac{t^{1/2}(1-t)^{1/2}}{f^2(F^{-1}(t))}\mathrm{d}t = 0.$$

The first and third of these limits follows just like the limit (4.13) in the proof of Lemma 4.3, using L'Hôpital and the equivalence (4.9). To show that the second limit also holds, let us set $h(x) = \int_0^x (F^{-1}(x) - F^{-1}(t))^2\mathrm{d}t$ and observe that

$$\begin{aligned} h'(x) &= \frac{2}{f(F^{-1}(x))}\int_0^x (F^{-1}(x) - F^{-1}(t))\mathrm{d}t = \frac{2}{f(F^{-1}(x))}\int_0^x\int_t^x \frac{1}{f(F^{-1}(u))}\mathrm{d}u\,\mathrm{d}t \\ &= \frac{2}{f(F^{-1}(x))}\int_0^x \frac{u}{f(F^{-1}(u))}\mathrm{d}u \simeq \frac{2x^2}{f^2(F^{-1}(x))}, \end{aligned}$$

the last equivalence being a consequence of regular variation. This, (4.9) and regular variation imply, in turn,

$$h(x) = \int_0^x h'(y)\mathrm{d}y \simeq 2\int_0^x y^2/f^2(F^{-1}(y))\mathrm{d}y \simeq 2x^3/f^2(F^{-1}(x)) \simeq 2x^2\int_x^\delta 1/f^2(F^{-1}(y))\mathrm{d}y$$

and, consequently, that

$$\lim_{x\to 0}\frac{1}{x\sqrt{L(x)}}\int_0^x (F^{-1}(x) - F^{-1}(t))^2\mathrm{d}t = 2\lim_{x\to 0}\frac{x\int_0^\delta 1/f^2(F^{-1}(y))\mathrm{d}y}{\sqrt{L(x)}} = 0$$

by (4.13). □

Example 4.1. Consider, for $\alpha > 0$, the distribution functions

$$F_\alpha(x) = \begin{cases} \frac{1}{2}\mathrm{e}^{-|x|^\alpha} & \text{if } x \leqslant 0, \\ 1 - \frac{1}{2}\mathrm{e}^{-x^\alpha} & \text{if } x \geqslant 0, \end{cases}$$

and take $w \equiv 1$. Let f_α be the corresponding densities, which are symmetric about zero. Then it is easy but somewhat cumbersome to show that

$$f_\alpha\big(F_\alpha^{-1}(x)\big) \approx x(1-x)\log^{(\alpha-1)/\alpha}\frac{1}{x(1-x)}, \quad f'_\alpha\big(F_\alpha^{-1}(x)\big) \approx x(1-x)\log^{(\alpha-2)/\alpha}\frac{1}{x(1-x)},$$

where $a(x) \approx b(x)$ means that $0 < \lim_{x\to 0} a(x)/b(x) < \infty$ and likewise for $x \to 1$, whereas $0 < \inf_{t\in I} a(t)/b(t) \leqslant \sup_{x\in I} a(t)/b(t) < \infty$ for any closed interval I contained in (0, 1) (for instance $f_\alpha\big(F_\alpha^{-1}(x)\big) = \alpha x|\log 2x|^{(\alpha-1)/\alpha} + \alpha(1-x)|\log 2(1-x)|^{(\alpha-1)/\alpha}$). So, $f_\alpha\big(F_\alpha^{-1}\big)$ is symmetric about $\frac{1}{2}$ and of regular variation with unit exponent at 0 (and at 1). It then follows easily that F_α falls under part (i) of Theorem 4.6 if and only if $\alpha > 2$, part (ii) if and only if $\frac{4}{3} < \alpha \leqslant 2$, hence for the normal distribution, and part (iii) if and only if $0 < \alpha \leqslant \frac{4}{3}$, in particular for the symmetric exponential distribution. As mentioned above, if the tail probabilities are of different order, the largest dominates and these theorems still hold, so that the same conclusions apply to the one-sided families.

Example 4.2. Likewise, if

$$f_\alpha(x) = \alpha x^{\alpha-1}\mathrm{e}^{-x^\alpha}, \quad x > 0,\ c > 0,$$

is the Weibull family of densities, then

$$f(F^{-1}(u)) = \alpha(1-u)\log^{(\alpha-1)/\alpha}\frac{1}{1-u},$$

and, as in Example 4.1, f_α falls under part (i), (ii) or (iii) of Theorem 4.6 according as to whether $\alpha > 3$, $\frac{4}{3} < \alpha \leqslant 2$ or $0 < \alpha \leqslant \frac{4}{3}$.

Example 4.3. The following example is due to McLaren and Lockhart (1987). For the logistic distribution $F(x) = (1 + \mathrm{e}^{-x})^{-1}$, $x \in \mathbb{R}$, and the exponential with parameter 1, both falling under part (iii) of Theorem 4.6, computation of L and the centring gives

$$\frac{1}{2^{3/2}\sqrt{\log n}}\left(\int_0^1 v_n^2(t)\mathrm{d}t - 2\log n\right) \to Z \quad \text{in distribution,}$$

and for the extreme value distribution $H(x) = \exp(-e^{-x})$, also falling under part (iii),

$$\frac{1}{2\sqrt{\log n}}\left(\int_0^1 v_n^2(t)\mathrm{d}t - \log n\right) \to Z \quad \text{in distribution.}$$

We cannot apply Proposition 2.5 when $f(F^{-1}(t))$ is too small at 0 and 1 (regularly varying of exponent 1 or more), as we saw in part (iii) of Theorem 4.6 (exponent 1). Here is a situation where the exponent is larger than one.

Theorem 4.7. *Assume F satisfies conditions* (2.7), $f(F^{-1})$ *varies regularly at* 0 *or at* 1 *with exponent* $\gamma \in \left(1, \frac{3}{2}\right)$ *and with equal or smaller exponent at the other extreme, and* $\lim_{x\to 0}|F^{-1}(x)|/F^{-1}(1-x) = c \in [0, \infty]$. *Then, denoting* $\alpha = 1 - \gamma$,

$$\frac{1}{\mathrm{E}(1/n)}\int_0^1 v_n^2(t)\mathrm{d}t \to_w \frac{2}{|\alpha|}\left[\frac{c^2}{1+c^2}\int_0^\infty \left((S_{[y]+1}^{(1)})^\alpha - y^\alpha\right)^2 \mathrm{d}y + \frac{1}{1+c^2}\int_0^\infty \left((S_{[y]+1}^{(2)})^\alpha - y^\alpha\right)^2 \mathrm{d}y\right].$$

If F satisfies conditions (2.7), $f(F^{-1})$ *varies regularly at* 0 *or at* 1 *with exponent* $\gamma = \frac{3}{2}$ *(i.e.,* $\alpha = -\frac{1}{2}$) *and with equal or smaller exponent at the other extreme,* $\lim_{x\to 0}|F^{-1}(x)|/F^{-1}(1-x) = c \in [0, \infty]$ *and* $\int_0^1 (F^{-1}(t))^2\mathrm{d}t < \infty$, *then*

$$\frac{1}{\mathrm{E}(1/n)}\left[\int_0^1 v_n^2(t)\mathrm{d}t - c_n\right] \to_w 4\left[\frac{c^2}{1+c^2}\left(\frac{1}{S_1^{(1)}} - \frac{4}{\sqrt{S_1^{(1)}}} + \int_1^\infty \left((S_{[y]+1}^{(1)})^{-1/2} - y^{-1/2}\right)^2 \mathrm{d}y\right)\right.$$

$$\left. + \frac{1}{1+c^2}\left(\frac{1}{S_1^{(2)}} - \frac{4}{\sqrt{S_1^{(2)}}} + \int_1^\infty \left((S_{[y]+1}^{(2)})^{-1/2} - y^{-1/2}\right)^2 \mathrm{d}y\right)\right],$$

where $c_n = \int_0^{1/n}(F^{-1}(t))^2\mathrm{d}t + \int_{1-1/n}^1 (F^{-1}(t))^2\mathrm{d}t$. *In both cases* $\{S_{[y]+1}^{(1)} : y \geqslant 0\}$ *is the partial sum process associated the sequence* $\{\xi_j\}$ *of independent exponential random variables, that is,* $S_{[y]+1}^{(1)} = \sum_{j=1}^{[y]+1}\xi_j$, *and* $\{S_{[y]+1}^{(2)} : y \geqslant 0\}$ *is an independent copy of* $\{S_{[y]+1}^{(1)}\}$.

Remark 4.1. Regular variation of $f(F^{-1})$ with exponent γ at 0, written $f(F^{-1}) \in RV_\gamma(0)$, is essentially equivalent to regular variation of F^{-1} with exponent $\alpha \in (-\frac{1}{2}, 0)$. In fact, if $f(F^{-1}) \in RV_\gamma(0)$, then $F^{-1} \in RV_\alpha(0)$ and, provided f is monotone in a neighbourhood of $-\infty$, if $F^{-1} \in RV_\alpha(0)$ then $f(F^{-1}) \in RV_\gamma(0)$ (see, for example, Resnick 1987, Propositions 0.6 and 0.7). With the assumption of regular variation, finiteness of $\int_0^1 v_n^2(t)\mathrm{d}t$ requires $\alpha \geqslant -\frac{1}{2}$. Thus, Theorem 4.7 completes the picture of all the possible limiting distributions of $\int_0^1 v_n^2(t)\mathrm{d}t$ for distributions with regularly varying tails.

Remark 4.2. It follows easily from the law of the iterated logarithm that

$$\limsup_{y\to\infty}\frac{\left|(S_{[y]+1}^{(1)})^\alpha - y^\alpha\right|}{y^{\alpha-1/2}\sqrt{2\log\log y}} = \frac{1}{|\alpha|} \quad \text{almost surely}$$

for all $\alpha < 0$ (see, Samorodnitsky and Taqqu 1994, p. 31). This implies that the limiting integrals in Theorem 4.7 are almost surely finite. Integrability of $((S^{(1)}_{[y]+1})^\alpha - y^\alpha)^2$ at 0 requires $\alpha > \frac{1}{2}$. When $\alpha = \frac{1}{2}$ the effect of the centring constants, c_n, is to remove this lack of integrability, still leading to a limiting distribution.

Next we collect some elementary properties of regularly varying functions that will be useful in our proof of Theorem 4.7.

Lemma 4.8. (i)$l \in RV_0(0)$ *is positive and* $\epsilon > 0$ *then*

$$\lim_{n\to\infty}(\log n)^{-\epsilon}\frac{l(\log n/n)}{l(1/n)} = 0$$

(ii) *If* $l \in RV_\alpha(0)$ *and* $\alpha < 0$ *then*

$$\lim_{n\to\infty}\frac{l(\log n/n)}{l(1/n)} = 0.$$

(iii) *If* $l \in RV_\alpha(0)$ *and* $\beta > -\alpha$ *then* $x^\beta l(x) \to 0$ *as* $x \to 0$.

Proof. (ii) is a trivial consequence of (i), and the proof of (iii) can be found in Resnick (1987), so we only prove (i). By Karamata's theorem (Resnick 1987, p. 17) l can be written as $l(x) = c(x)\exp(\int_x^1(\epsilon(t)/t)\mathrm{d}t)$ with $c(x) \to c \in (0, \infty)$ and $\epsilon(x) \to 0$ as $x \to 0$. Therefore, taking n_0 large enough to ensure that $|\epsilon(t)| < \epsilon/2$ for $t \leqslant \log n_0/n_0$ and $n \geqslant n_0$ we have that

$$(\log n)^{-\epsilon}\frac{l(\log n/n)}{l(1/n)} \leqslant 2(\log n)^{-\epsilon}\exp\left(\frac{\epsilon}{2}\int_{1/n}^{\log n/n}\frac{1}{t}\mathrm{d}t\right) = 2(\log n)^{-\epsilon/2} \to 0.$$

□

Proof of Theorem 4.7. We will assume in this proof that $0 > \alpha > -\frac{1}{2}$. The case $\alpha = -\frac{1}{2}$ can be handled with straightforward changes. We set, as in the proof of Proposition 4.5, $b_{n,j} = I\{j - 1 < nt\}$ and $H^{-1}(x) = -F^{-1}(1 - x)$ and observe, using the fact that $b_{n,j}(1 - t) = 1 - b_{n,n+2-j}(t)$ except in a null set, that

$$\begin{aligned}\frac{1}{\mathrm{E}(1/n)}\int_0^1 v_n^2(t)\mathrm{d}t &= \frac{n}{\mathrm{E}(1/n)}\int_0^{\log n/n}\left(F^{-1}\left(\frac{1}{S_{n+1}}\sum_{j=1}^{n+1} b_{n,j}(t)\xi_j\right) - F^{-1}(t)\right)^2 \mathrm{d}t \\ &\quad + \frac{n}{\mathrm{E}(1/n)}\int_{1-\log n/n}^{1}\left(F^{-1}\left(\frac{1}{S_{n+1}}\sum_{j=1}^{n+1} b_{n,j}(t)\xi_j\right) - F^{-1}(t)\right)^2 \mathrm{d}t \\ &\quad + \frac{1}{\mathrm{E}(1/n)}\int_{\log n/n}^{1-\log n/n} v_n^2(t)\mathrm{d}t \\ &= \frac{n}{\mathrm{E}(1/n)}\int_0^{\log n/n}\left(F^{-1}\left(\frac{1}{S_{n+1}}\sum_{j=1}^{n+1} b_{n,j}(t)\xi_j\right) - F^{-1}(t)\right)^2 \mathrm{d}t \\ &\quad + \frac{n}{\mathrm{E}(1/n)}\int_0^{\log n/n}\left(H^{-1}\left(\frac{1}{S_{n+1}}\sum_{j=1}^{n+1} b_{n,j}(t)\xi_{n+2-j}\right) - H^{-1}(t)\right)^2 \mathrm{d}t \\ &\quad + \frac{1}{\mathrm{E}(1/n)}\int_{\log n/n}^{1-\log n/n} v_n^2(t)\mathrm{d}t \\ &=: V_n^{(1)} + V_n^{(2)} + V_n^{(3)}.\end{aligned}$$

We also set

$$\begin{aligned}W_n^{(1)} &:= \frac{n}{\mathrm{E}(1/n)}\int_0^{\log n/n}\left(F^{-1}\left(\frac{1}{n}\sum_{j=1}^{n+1} b_{n,j}(t)\xi_j\right) - F^{-1}(t)\right)^2 \mathrm{d}t, \\ W_n^{(2)} &:= \frac{n}{\mathrm{E}(1/n)}\int_0^{\log n/n}\left(H^{-1}\left(\frac{1}{n}\sum_{j=1}^{n+1} b_{n,j}(t)\xi_{n+2-j}\right) - H^{-1}(t)\right)^2 \mathrm{d}t.\end{aligned}$$

Observe that $W_n^{(1)}$ and $W_n^{(2)}$ are independent since they are functions of disjoint sets of independent exponential random variables ξ_j. We now proceed by showing that the central part, $V_n^{(3)}$, is negligible and that the upper and lower integrals are asymptotically independent and weakly convergent to the above stated limits. This will be achieved by proving the following three claims:

Claim 1.

$$V_n^{(3)} \underset{\mathrm{Pr}}{\to} 0.$$

Claim 2.

$$(V_n^{(i)})^{1/2} - (W_n^{(i)})^{1/2} \underset{\mathrm{Pr}}{\to} 0, \quad i = 1, 2.$$

Claim 3.

$$W_n^{(1)} \to_w \frac{2c^2}{|\alpha|(1+c^2)} \int_0^\infty \left((S_{[y]+1}^{(1)})^\alpha - y^\alpha \right)^2 \mathrm{d}y,$$

$$W_n^{(2)} \to_w \frac{2}{|\alpha|(1+c^2)} \int_0^\infty \left((S_{[y]+1}^{(2)})^\alpha - y^\alpha \right)^2 \mathrm{d}y.$$

Proof of Claim 1. We first show that

$$V_n^{(3)} - \frac{1}{\mathrm{E}(1/n)} \int_{\log n/n}^{1-\log n/n} \frac{u_n^2(t)}{f^2(F^{-1}(t))} \mathrm{d}t \underset{\mathrm{Pr}}{\to} 0.$$

As in the proof of Proposition 2.3, this reduces to showing that

$$\frac{1}{n\mathrm{E}(1/n)} \int_{\log n/n}^{1-\log n/n} \frac{1}{f^2(F^{-1}(t))} \mathrm{d}t \to 0 \quad \text{and} \quad \frac{1}{\sqrt{n}\mathrm{E}(1/n)} \int_{\log n/n}^{1-\log n/n} \frac{t^{1/2}(1-t)^{1/2}}{f^2(F^{-1}(t))} \mathrm{d}t \to 0.$$

For ease of computation we will assume in the remainder of the proof of this claim that $c = 1$ and replace $\mathrm{E}(1/n)$ by $F^{-1}(1/n)^2$ in the last two denominators (the ratio of the two sequences converges, by regular variation, to a positive constant). Extension to general c is straightforward. Regular variation implies that

$$\lim_{x\to 0} \frac{x/f(F^{-1}(x))}{F^{-1}(x)} = \alpha \quad \text{and} \quad \lim_{x\to 0} \frac{x/f^2(F^{-1}(x))}{\int_x^{1-x} 1/(f^2(F^{-1}(t)))\mathrm{d}t} = \frac{1}{2} - \alpha,$$

which implies, in turn, that

$$\lim_{n\to\infty} \frac{1}{nF^{-1}(1/n)^2} \int_{\log n/n}^{1-\log n/n} \frac{1}{f^2(F^{-1}(t))} \mathrm{d}t = \frac{\alpha^2}{1/2-\alpha} \lim_{n\to\infty} \frac{l^{(1)}(\log n/n)}{l^{(1)}(1/n)} = 0,$$

where $l^{(1)}(x) = x/f^2(F^{-1}(x)) \in RV_{2\alpha-1}(0)$ and $2\alpha - 1 \in (-2, -1)$ and the last limit follows from Lemma 4.8. Similarly,

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}F^{-1}(1/n)^2} \int_{\log n/n}^{1-\log n/n} \frac{t^{1/2}(1-t)^{1/2}}{f^2(F^{-1}(t))} \mathrm{d}t = \frac{\alpha^2}{1/4-\alpha} \lim_{n\to\infty} \frac{l^{(2)}(\log n/n)}{l^{(2)}(1/n)} = 0,$$

since $l^{(2)}(x) = x^{3/2}/f^2(F^{-1}(x)) \in RV_{2\alpha-1/2}(0)$ and $2\alpha - \frac{1}{2} \in (-\frac{3}{2}, -\frac{1}{2})$. Now we can prove Claim 1 by showing that

$$\frac{1}{F^{-1}(1/n)^2}\int_{\log n/n}^{1-\log n/n}\frac{u_n^2(t)}{f^2(F^{-1}(t))}\mathrm{d}t \underset{\mathrm{Pr}}{\to} 0.$$

But taking expectations, we can see that it suffices to show that

$$\lim_{n\to\infty}\frac{1}{F^{-1}(1/n)^2\int_{\log n/n}^{1-\log n/n}\frac{t(1-t)}{f^2(F^{-1}(t))}\mathrm{d}t = 0.}$$

Using again regular variation properties and Lemma 4.8, we have that

$$\lim_{n\to\infty}\frac{1}{F^{-1}(1/n)^2}\int_{\log n/n}^{1-\log n/n}\frac{t(1-t)}{f^2(F^{-1}(t))}\mathrm{d}t = |\alpha|\lim_{n\to\infty}\frac{l^{(3)}(\log n/n)}{l^{(3)}(1/n)} = 0,$$

now with $l^{(3)}(x) = x^2/f^2(F^{-1}(x)) \in RV_{-2\alpha}(0)$ and $-2\alpha \in (-1, 0)$. This completes the proof of Claim 1.

Proof of Claim 2. We will show that $(V_n^{(1)})^{1/2} - (W_n^{(1)})^{1/2} \to_{\mathrm{Pr}} 0$. It suffices to show that

$$\frac{1}{F^{-1}(1/n)^2}\int_0^{\log n}(F^{-1}(S_{[y]+1}^{(1)}/S_{n+1}) - F^{-1}(S_{[y]+1}^{(1)}/n))^2\mathrm{d}y \to_{\mathrm{Pr}} 0.$$

For ease of notation we will omit the superscript from $S_{[y]+1}^{(1)}$. Similarly as in (2.12), we consider a Taylor expansion

$$F^{-1}\left(\frac{S_j}{S_{n+1}}\right) - F^{-1}\left(\frac{S_j}{n}\right) = \left(1 - \frac{S_{n+1}}{n}\right)\frac{S_j}{S_{n+1}}\frac{1}{f\left(F^{-1}\left(\frac{S_j}{n}\right)\right)}$$
$$+\frac{1}{2}\left(1 - \frac{S_{n+1}}{n}\right)^2\left(\frac{S_j}{S_{n+1}}\right)^2\frac{f'(F^{-1}(\xi))}{f^3(F^{-1}(\xi))},$$

for some ξ between S_j/S_{n+1} and S_j/n, which enables us to write, using the obvious analogues of (2.16) and (2.17), and the fact that $\sup_{n\geqslant 1} n\mathrm{E}(1 - S_{n+1}/n)^2 < \infty$, that

$$\left|F^{-1}\left(\frac{S_j}{S_{n+1}}\right) - F^{-1}\left(\frac{S_j}{n}\right)\right| \leqslant O_P(1)\left(\frac{1}{\sqrt{n}}\frac{S_j/n}{f(F^{-1}(S_j/n))} + \frac{1}{n}\frac{S_j/n}{f(F^{-1}(S_j/n))}\right)$$
$$\leqslant O_P(1)\frac{1}{\sqrt{n}}\frac{S_j/n}{f(F^{-1}(S_j/n))},$$

where $O_P(1)$ stands for a stochastically bounded sequence which does not depend on $j \in [1, \log n]$. We now take $\epsilon > 0$ such that $2\gamma - 3 + \epsilon < 0$. From this bound and regular variation (Lemma 4.8(iii)) we obtain that

$$\int_0^{\log n}\left(F^{-1}\left(\frac{S_{[y]+1}}{S_{n+1}}\right)-F^{-1}\left(\frac{S_{[y]+1}}{n}\right)\right)^2 \mathrm{d}y \leqslant O_P(1)\frac{1}{n}\int_0^{\log n}\frac{\left(S_{[y]+1}/n\right)^2}{f^2(F^{-1}(S_{[y]+1}/n))}\mathrm{d}y$$
$$\leqslant O_P(1)\frac{1}{n}\sum_{j=1}^{[\log n+1]}\frac{\left(S_j/n\right)^2}{f^2(F^{-1}(S_j/n))}$$
$$\leqslant O_P(1)\frac{1}{n}\sum_{j=1}^{[\log n+1]}\left(\frac{S_j}{n}\right)^{2-2\gamma-\epsilon}$$
$$\leqslant O_P(1)n^{2\gamma-3+\epsilon}\sum_{j=1}^{[\log n+1]} j^{2-2\gamma-\epsilon}\to 0.$$

This completes the proof of Claim 2 (note that we need not divide by $F^{-1}(1/n)^2$ to obtain the equivalence of the two sequences if $\gamma<\frac{3}{2}$; if $\gamma=\frac{3}{2}$ that division still gives the result).

Proof of Claim 3. We set

$$W_{n,k}^{(1)}=\frac{n}{\mathrm{E}(1/n)}\int_0^{k/n}\left(F^{-1}\left(\frac{1}{n}\sum_{j=1}^{n+1}b_{n,j}(t)\xi_j\right)-F^{-1}(t)\right)^2\mathrm{d}t$$

and

$$W_k=\frac{2c^2}{|\alpha|(1+c^2)}\int_0^k\left((S_{[y]+1}^{(1)})^\alpha-y^\alpha\right)^2\mathrm{d}y.$$

With the change of variable $t=y/n$ we can rewrite $W_{n,k}^{(1)}$ as

$$W_{n,k}^{(1)}=b_n\int_0^k\left(\frac{F^{-1}(S_{[y]+1}^{(1)}/n)-F^{-1}(y/n)}{F^{-1}(1/n)}\right)^2\mathrm{d}y,$$

where $b_n=F^{-1}(1/n)^2/\mathrm{E}(1/n)\to 2c^2/|\alpha|(1+c^2)$, and we conclude that, by regular variation, $W_{n,k}^{(1)}\to_{\mathrm{Pr}}W_k$. To prove that

$$W_n^{(1)}\to_w\frac{2c^2}{|\alpha|(1+c^2)}\int_0^\infty((S_{[y]+1}^{(1)})^\alpha-y^\alpha)^2\mathrm{d}y$$

it suffices, using a 3ϵ argument, to show that

$$\lim_{k\to\infty}\limsup_{n\to\infty}P(|W_{n,k}^{(1)}-W_n^{(1)}|>\epsilon)=0 \tag{4.21}$$

for all $\epsilon>0$. As in the proof of Claim 2, we consider a Taylor expansion

$$F^{-1}\left(\frac{S_{[y]+1}}{n}\right)-F^{-1}\left(\frac{y}{n}\right)=\frac{S_{[y]+1}-y}{nf(F^{-1}(y/n))}+\frac{1}{2}\frac{(S_{[y]+1}-y)^2}{n^2}\frac{f'(F^{-1}(\xi))}{f^3(F^{-1}(\xi))},$$

for some ξ between $S_{[y]+1}/n$ and y/n, which enables us to write, using the obvious equivalents of (2.16) and (2.17), and the fact that $\sup_{y\geqslant 1}\mathrm{E}\big((S_{[y]+1}-y)/y^{1/2}\big)^2<\infty$, that

$$\left|F^{-1}\left(\frac{S_{[y]+1}}{n}\right)-F^{-1}\left(\frac{y}{n}\right)\right|\leqslant O_P(1)\left(\frac{1}{\sqrt{n}}\frac{\sqrt{y}/\sqrt{n}}{f(F^{-1}(y/n))}+\frac{1}{n}\frac{1}{f(F^{-1}(y/n))}\right),$$

where $O_P(1)$ stands for a stochastically bounded sequence which does not depend on $y\in[k,\log n]$. From this bound we obtain that

$$\begin{aligned}|W_{n,k}^{(1)}-W_{n}^{(1)}|&=\frac{b_n}{F^{-1}(1/n)^2}\int_k^{\log n}\Big(F^{-1}(S_{[y]+1}^{(1)}/n)-F^{-1}(y/n)\Big)^2\mathrm{d}y\\&\leqslant O_P(1)\frac{1}{F^{-1}(1/n)^2}\left(\frac{1}{n}\int_k^{\log n}\frac{y/n}{f^2(F^{-1}(y/n))}\mathrm{d}y+\frac{1}{n^2}\int_k^{\log n}\frac{1}{f^2(F^{-1}(y/n))}\mathrm{d}y\right)\\&=O_P(1)\frac{1}{F^{-1}(1/n)^2}\left(\int_{k/n}^{\log n/n}\frac{t}{f^2(F^{-1}(t))}\mathrm{d}t+\frac{1}{n}\int_{k/n}^{\log n/n}\frac{1}{f^2(F^{-1}(t))}\mathrm{d}t\right).\end{aligned}$$

From regular variation we obtain that

$$\frac{1}{F^{-1}(1/n)^2}\left(\int_{k/n}^{\log n/n}\frac{t}{f^2(F^{-1}(t))}\mathrm{d}t+\frac{1}{n}\int_{k/n}^{\log n/n}\frac{1}{f^2(F^{-1}(t))}\mathrm{d}t\right)\to C_1k^{2-2\gamma}+C_2k^{1-2\gamma}$$

and this, combined with the last estimate and the fact that $2-2\gamma<0$, completes the proof of (4.21) and, consequently, of Claim 3. □

5. Weighted Wasserstein tests of fit to location–scale families of distributions

Finally, we apply the foregoing to weighted Wasserstein tests. Recall that $\mathcal{R}_n^w=\mathcal{W}_w^2(F_n,\mathcal{H})/\sigma_w^2(F_n)$ relates to the quantile process v_n via (1.11) (assuming conditions (1.8)–(1.10)).

Theorem 5.1. *Let w be a bounded non-negative measurable function satisfying condition* (1.8). *Let $\mathcal{H}$ be a location–scale family of distributions as defined in the Introduction, such that $\int_0^1(F^{-1}(t))^2w(t)\mathrm{d}t<\infty$ for any (hence for all) $F\in\mathcal{H}$, let $G_0\in\mathcal{H}$ be chosen so as to satisfy conditions* (1.9) *and* (1.10) *and let $g_0=G_0'$. Assume the distribution functions $F\in\mathcal{H}$ and the weight w satisfies conditions* (GH) *and that, moreover,*

$$\int_0^1\frac{t(1-t)}{f^2(F^{-1}(t))}w(t)\mathrm{d}t<\infty.\qquad(5.1)$$

Then, under the null hypothesis $F\in\mathcal{H}$, we have

$$n\mathcal{R}_n^w \underset{d}{\rightarrow} \int_0^1 \frac{B^2(t)}{g_0^2(G_0^{-1}(t))} w(t)\mathrm{d}t - \left(\int_0^1 \frac{B(t)}{g_0(G_0^{-1}(t))} w(t)\mathrm{d}t\right)^2 - \left(\int_0^1 \frac{B(t)G_0^{-1}(t)}{g_0(G_0^{-1}(t))} w(t)\mathrm{d}t\right)^2. \quad (5.2)$$

Note that the hypotheses on $F \in \mathcal{H}$ are either satisfied by all or by none of the functions in $\mathcal{H}$.

Proof. By equivariance we can assume $F = G_0$. The result follows directly from Theorem 4.6(i) once we show that $\sigma_w^2(F_n) \to 1$ in probability. By Theorem 4.6(i) $\|v_n\|_{2,w} = O_P(1)$, and therefore (recall $F = G_0$ and (1.10))

$$|\sigma_w(F_n) - 1| = |\|F_n^{-1}\|_{2,w} - \|F^{-1}\|_{2,w}| \leqslant \|F_n^{-1} - F^{-1}\|_{2,w} = \frac{1}{\sqrt{n}}\|v_n\|_{2,w} \underset{\Pr}{\rightarrow} 0.$$

□

If $\mathcal{H}$ in Theorem 5.1 were only a location family or only a scale family then the limit would exhibit the loss of only one degree of freedom, that is, one of the last two integrals would be absent from the limit in (5.2): see Csörgő (2002), where a theorem of this sort for scale families is proved.

Theorem 5.2. *Under the hypotheses of Theorem 5.1, except that condition* (5.1) *is now replaced by the weaker conditions* (2.10) *and*

$$\int_0^1\int_0^1 \frac{(s \wedge t - st)^2}{g_0^2\big(G_0^{-1}(t)\big)\, g_0^2\big(G_0^{-1}(s)\big)} w(s)w(t)\mathrm{d}s\,\mathrm{d}t < \infty, \quad (5.3)$$

we have

$$n\mathcal{R}_n^w - \int_{1/n}^{1-1/n} \frac{t(1-t)}{g_0^2(G_0^{-1}(t))} w(t)\mathrm{d}t \quad (5.4)$$

$$\underset{d}{\rightarrow} \int_0^1 \frac{B^2(t) - \mathrm{E}B^2(t)}{g_0^2(G_0^{-1}(t))} w(t)\mathrm{d}t - \left(\int_0^1 \frac{B(t)}{g_0(G_0^{-1}(t))} w(t)\mathrm{d}t\right)^2 - \left(\int_0^1 \frac{B(t)G_0^{-1}(t)}{g_0(G_0^{-1}(t))} w(t)\mathrm{d}t\right)^2.$$

Proof. As above, we can take $F = G_0$. By Theorem 4.6(ii) properly modified to account for the weighted integrals of the Brownian bridge (as done in Theorem 4.2), it suffices to prove that

$$\sigma_w(F_n) \underset{\Pr}{\rightarrow} 1 \text{ and } \left(\frac{1}{\sigma_w^2(F_n)} - 1\right)\int_{1/n}^{1-1/n} \frac{t(1-t)}{f^2(F^{-1}(t))} w(t)\mathrm{d}t \underset{\Pr}{\rightarrow} 0,$$

which, by condition (2.10), reduces to proving $\sqrt{n}(\sigma_w^2(F_n) - 1) = O_P(1)$. We have

$$\sqrt{n}(\sigma_w^2(F_n) - 1) = \frac{1}{\sqrt{n}}\|v_n\|_{2,w} + 2\langle v_n, F^{-1}\rangle_w.$$

By checking the proof of Theorem 4.2 (by way of Theorem 3.9), it is easy to see that

$\langle u_n/f(F^{-1}), F^{-1}\rangle_{w,n} \to_d \langle B/f(F^{-1}), F^{-1}\rangle_{w,n}$ (note that (1.10) implies $F^{-1} \in L_2(w(t)\mathrm{d}t)$). Hence Proposition 2.5 gives $\langle v_n, F^{-1}\rangle_w = O_P(1)$. Likewise, by Theorem 4.6(ii), $\|v_n\|_{2,w}$ is shift convergent in law with shifts $\int_{1/n}^{1-1/n} t(1-t)/f^2(F^{-1}(t))w(t)\mathrm{d}t$ which, by (2.10), are $o(\sqrt{n})$, so that $\|v_n\|_{2,w}/\sqrt{n} \to_{\Pr} 0$. □

A version of this theorem for scale families is proved in Csörgő (2002), however, the hypotheses there are stronger by factors of order $\log n$ or $(\log n)^2$, the integrals at the end-points are not treated analytically and the proof is different (it relies on strong approximations, which account for the stronger assumptions).

Next we consider convergence to a normal distribution. This case is less interesting in connection with testing since, as indicated in the Introduction, $\int_\delta^{1-\delta} v_n^2(t)w(t)\mathrm{d}t \to_d \int_\delta^{1-\delta} B^2(t)/f^2(F^{-1}(t))w(t)\mathrm{d}t$ if f does not vanish on $\operatorname{supp} F$, and therefore, if we divide by $\sqrt{L(1/n)} \to \infty$, as we must by Theorem 4.6(iii), this part of the statistic has no influence on the limit. So, when a distribution satisfies the hypotheses of Theorem 4.6(iii) (meaning that $g = f(F^{-1})$ is regularly varying with exponent 1, and $L(x)$ with this g tends to infinity), if one wishes to have a sensible test of fit, it is probably best to find a weight w so that one can apply Theorem 5.1 or 5.2. Hence, we will only consider the normal convergence case with weight $w \equiv 1$.

Theorem 5.3. *Let $\mathcal{H}$ be a location–scale family of distributions and assume for simplicity that the distribution $G_0 \in \mathcal{H}$ with mean 0 and variance 1 is the distribution function of a symmetric random variable. Assume that the following conditions hold for some (hence for all) $F \in \mathcal{H}$: F is twice differentiable on its open support (a_F, b_F), with $f(x) = F'(x) > 0$; F satisfies condition (2.7); and the function $g := f(F^{-1})$ is either regularly varying with unit exponent at 0 and $\lim_{x\to 0} g(x)/g(1-x) = c \in [0, \infty)$, or is regularly varying with unit exponent at 1 and $\lim_{x\to 0} g(1-x)/g(x) = c \in [0, \infty)$; and $L(x) \to \infty$ as $x \to 0$. Then, under the null hypothesis $F \in \mathcal{H}$,*

$$\frac{1}{\sqrt{L(1/n)}}\left(n\mathcal{R}_n - \int_{1/n}^{1-1/n} \frac{t(1-t)}{g_0^2\left(G_0^{-1}(t)\right)}\mathrm{d}t\right) \underset{d}{\to} Z, \tag{5.5}$$

where Z is standard normal and $n\mathcal{R}_n := n\mathcal{R}_n^w$ is as defined in (1.11) for $w \equiv 1$.

Proof. As in the previous two theorems, we can assume $F = G_0$. Since $f(F^{-1})$ is regularly varying of unit exponent at 0 and at 1, it follows that $F^{-1}(x) = \int_{1/2}^x \mathrm{d}t/f(F^{-1}(t))$ is slowly varying at 0 and at 1. Hence, $\int_0^1 |F^{-1}(t)|^r \mathrm{d}t < \infty$ for all $r \in \mathbb{R}$ and therefore, all the moments $\int_0^\infty |x|^r \mathrm{d}F(x)$, $r > 0$, are finite. In particular, if X_i are i.i.d. with distribution F, then

$$\int_0^1 v_n(t)\mathrm{d}t = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mathrm{E}X_i) = O_P(1)$$

by the central limit theorem, and $\sigma^2(F_n) \to 1$ almost surely by the law of large numbers. So it suffices to show that

$$\frac{1}{\sqrt{L(1/n)}}\left[\|v_n\|_2^2 - \langle v_n, F^{-1}\rangle^2 - \int_{1/n}^{1-1/n}\frac{t(1-t)}{f^2(F^{-1}(t))}\mathrm{d}t\right] \underset{d}{\to} Z.$$

The arguments in the proof of Theorem 4.6(iii) not only show that here we can replace $\|v_n\|_2^2$ by $\|u_n/f(F^{-1})\|_{2,n}$, but also that $\langle v_n, F^{-1}\rangle^2$ can be replaced by $\langle u_n/f(F^{-1}), F^{-1}\rangle_n$; therefore, the theorem will follow from Theorem 4.6(iii) (hence from Theorem 4.4) if we show that the sequence

$$\left\langle\frac{u_n}{f(F^{-1})}, F^{-1}\right\rangle_n := \int_{1/n}^{1-1/n}\frac{u_n(t)F^{-1}(t)}{f(F^{-1}(t))}\mathrm{d}t, \qquad n\in\mathbb{N}, \tag{5.6}$$

is stochastically bounded (as it will then tend to zero upon dividing by $\sqrt{L(1/n)}$). For this purpose, we show that the product of the nth variable in (5.6) by S_{n+1}/n has expected value tending to zero and variance dominated by a constant independent of n. By (2.2), Lemma 2.1(ii) and slow variation of F^{-1} at 0 and 1, we have

$$\begin{aligned}\left|\mathrm{E}\left\langle\frac{S_{n+1}}{n}\frac{u_n}{f(F^{-1})}, F^{-1}\right\rangle_n\right| &= \left|\mathrm{E}\int_{1/n}^{1-1/n}\frac{F^{-1}(t)\sum_{i=1}^{n+1}a_{n,i}\xi_i}{\sqrt{n}f(F^{-1}(t))}\mathrm{d}t\right| \\ &\leqslant \frac{1}{\sqrt{n}}\int_{1/n}^{1-1/n}\frac{|F^{-1}(t)|}{f(F^{-1}(t))}\mathrm{d}t \\ &= \frac{1}{\sqrt{n}}\int_{F^{-1}(1/n)}^{F^{-1}(1-1/n)}|u|\mathrm{d}u \\ &= \frac{(F^{-1}(1/n))^2 + (F^{-1}(1-1/n))^2}{2\sqrt{n}} \to 0.\end{aligned}$$

Let X be a random variable with distribution F. By Lemma 2.1(iii) and finiteness of the absolute moments of X, we have

$$\begin{aligned}
\operatorname{var}\left(\left\langle \frac{S_{n+1}}{n}\frac{u_n}{f(F^{-1})},\, F^{-1}\right\rangle_n\right) &= \mathrm{E}\left(\frac{1}{\sqrt{n}}\int_{1/n}^{1-1/n}\frac{F^{-1}(t)\sum_{i=1}^{n+1}a_{n,i}(\xi_i-1)}{f(F^{-1}(t))}\mathrm{d}t\right)^2\\
&= \frac{1}{n}\int_{1/n}^{1-1/n}\int_{1/n}^{1-1/n}\frac{\tilde{K}_n(s,\,t)F^{-1}(s)\|F^{-1}(t)}{f(F^{-1}(s))f(F^{-1}(t))}\mathrm{d}s\,\mathrm{d}t\\
&\leqslant 6\int_{1/n}^{1-1/n}\int_{1/n}^{t}\frac{s(1-t)|F^{-1}(s)\|F^{-1}(t)|}{f(F^{-1}(s))f(F^{-1}(t))}\mathrm{d}s\,\mathrm{d}t\\
&= 6\int_{F^{-1}(1/n)}^{F^{-1}(1-1/n)}\int_{F^{-1}(1/n)}^{v}F(u)(1-F(v))|u\|v|\mathrm{d}u\,\mathrm{d}v\\
&\leqslant 6\int_{F^{-1}(1/n)}^{0}\int_{F^{-1}(1/n)}^{v}F(u)|u\|v|\mathrm{d}u\,\mathrm{d}v\\
&\quad+\int_{0}^{F^{-1}(1-1/n)}\int_{F^{-1}(1/n)}^{0}F(u)(1-F(v))|u\|v|\mathrm{d}u\,\mathrm{d}v\\
&\quad+\int_{0}^{F^{-1}(1-1/n)}\int_{0}^{v}(1-F(v))|u\|v|\mathrm{d}u\,\mathrm{d}v\\
&\leqslant \frac{3}{4}\left(\mathrm{E}X^4+(\mathrm{E}X^2)^2\right)<\infty,
\end{aligned}$$

where at the last step we use Fubini and integration by parts. □

As in Theorem 4.6(iii), symmetry of G_0 is not necessary. Csörgő (2002) also proves a result for correlation tests where the limit is normal, but only for the special case of Weibull scale families.

Likewise, Theorem 4.7 can be used to obtain the limiting distribution of $n\mathcal{R}_n$ when $f(F^{-1})$ is regularly varying at the end-points with exponent $\gamma > 1$, but we refrain from doing so, to avoid too much repetition.

Example 5.1 *Gauss–Laplace location–scale families.* This is a modification of a result in Csörgő (2002). Consider the distribution functions $F_\alpha(x)$ from Example 4.1. In that example, Theorem 5.1 with $w \equiv 1$ holds for the location–scale family based on F_α if and only if $\alpha > 2$, Theorem 5.2 with $w \equiv 1$ holds if and only if $4/3 < \alpha \leqslant 2$, hence for the normal distribution (which gives Shapiro–Wilk), and Theorem 5.3 with $w \equiv 1$ holds for $0 < \alpha \leqslant \frac{4}{3}$, in particular for the symmetric exponential distribution. As mentioned above, if the tail probabilities are of different order, the largest dominates and the same conclusions apply to the one-sided families.

Example 5.2 *Testing fit to the Laplace location–scale family.* It follows from Example 5.1 and the comments immediately before Theorem 5.3 that a weighted Wasserstein test would be convenient for the Gauss–Laplace location–scale family when the index α is between 0 and $\frac{4}{3}$. For any given $\alpha > 0$, these families are (in terms of the densities):

$$\mathcal{H}_\alpha = \left\{ F_{\beta,\gamma} : f_{\beta,\gamma}(x) := \frac{\alpha}{2\gamma\Gamma(1/\alpha)} \mathrm{e}^{-|(x-\beta)/\gamma|^\alpha}, \, x \in \mathbb{R}, \, \beta \in \mathbb{R}, \, \gamma > 0 \right\}.$$

The weight should approach zero near 0 and 1. For simplicity we will only present a test for the Laplace family $\mathcal{H}_1$. Simple but tedious computations using the approximations in the previous example show that a weight of order $w(t) \sim 1/|\log t(1-t)|^\tau$ will allow us to apply Theorem 5.1 if $\tau > 1$ and Theorem 5.2 if $\frac{1}{2} < \tau \leqslant 1$ (the determining conditions are (5.1), which holds for all $\tau > 1$, and (5.5), which holds for $\frac{1}{2} < \tau \leqslant 1$). If w is too small near 0 and 1, we make the extreme part of the distribution count less, whereas possibly the limit has more variability as the integral of $B^2 - \mathrm{E}B^2$ is closer to being divergent. De Wet (2000) convincingly suggests taking $\tau = 1$ (see also Csörgő 2002). Specifically, we define

$$w(t) := \left(\frac{1}{\log \mathrm{e}/2t} I_{0<t\leqslant 1/2} + \frac{1}{\log \mathrm{e}/2(1-t)} I_{1/2<t\leqslant 1} \right) / W$$

where

$$W := \mathrm{e} \int_1^\infty u^{-1} \mathrm{e}^{-u} \mathrm{d}u,$$

and set also

$$V = \int_0^\infty \frac{u^2}{1+u} \mathrm{e}^{-u} \mathrm{d}u.$$

Take $G_0 := F_{0,\sqrt{W/V}}$. Then w and G_0 satisfy conditions (1.8)–(1.10), and the conditions (GH) and (5.3) hold as well (but not (5.1)). Then, Theorem 5.2 gives that, under the null hypothesis $F \in \mathcal{H}_1$,

$$\begin{aligned} n\mathcal{R}_n^w - \frac{2}{V}\left(\log\log \frac{n\mathrm{e}}{2} - \frac{\mathrm{W}}{2} \right) \underset{d}{\to} \frac{1}{V} &\left(\int_0^{1/2} \frac{B^2(t) - \mathrm{E}B^2(t)}{t^2 \log(\mathrm{e}/2t)} \mathrm{d}t + \int_{1/2}^1 \frac{B^2(t) - \mathrm{E}B^2(t)}{(1-t)^2 \log(\mathrm{e}/2(1-t))} \mathrm{d}t \right) \\ &- \frac{1}{VW} \left(\int_0^{1/2} \frac{B(t)}{t \log(\mathrm{e}/2t)} \mathrm{d}t + \int_{1/2}^1 \frac{B(t)}{(1-t)\log(\mathrm{e}/2(1-t))} \mathrm{d}t \right)^2 \\ &- \frac{1}{V^2} \left(\int_0^{1/2} \frac{B(t)\log 2t}{t \log(\mathrm{e}/2t)} \mathrm{d}t + \int_{1/2}^1 \frac{B(t)\log 2(1-t)}{(1-t)\log(\mathrm{e}/2(1-t))} \mathrm{d}t \right)^2. \end{aligned}$$

Note added after submission

Written independently and submitted at about the same time, Csörgő (2003) also considers weighted Wasserstein tests of fit to location–scale families. The methodology used is different, and the assumptions are not exactly the same.

Acknowledgements

This research has been partially supported by MCYT and FEDER grant no. BFM2002-04430-C02-00,01 and PAPIJCYL grant no. VA74/03 (del Barrio), NSF Grant no. DMS–0070382 (Giné) and CIRIT 2001SGR00174 and DGI MCYT BFM2003-00261 (Utzet).

References

Araujo, A. and Giné, E. (1980) *The Central Limit Theorem for Real and Banach Valued Random Variables*. New York: Wiley

Bickel, P. and Freedman, D. (1981) Some asymptotic theory for the bootstrap. *Ann Statist*., **9**, 1196–1216.

Csörgő, M. and Horváth, L. (1988) On the distribution of L_p norms of weighted uniform empirical and quantile processes. *Ann. Probab.*, **16**, 142–161.

Csörgő, M. and Horváth, L. (1993) *Weighted Approximations in Probability and Statistics*. New York: Wiley

Csörgő, M. and Révész, P. (1978) Strong approximations of the quantile process. *Ann. Statist*., **6**, 882–894.

Csörgő, M., Csörgő, S., Horváth, L. and Mason, D.M. (1986) Weighted empirical and quantile processes. *Ann. Probab.*, **14**, 31–85.

Csörgő, S. (2000) Testing for Weibull scale families as a test case for Wasserstein correlation tests. *Test*, **9**, 54–70. [Discussion to del Barrio *et al*. (2000).]

Csörgő, S. (2002) Weighted correlation tests for scale families. *Test*, **11**, 219–248.

Csörgő, S. (2003) Weighted correlation tests for location–scale families. *Math. Comput. Modelling*, **38**, 753–762.

de la Peña, V. and Giné, E. (1999) *Decoupling: From Dependence to Independence*. New York: Springer-Verlag.

de Wet, T. (2000) Discussion to del Barrio, Cuesta-Albertos and Matrán (2000). *Test*, **9**, 74–79.

de Wet, T. (2002) Goodness-of-fit tests for location and scale families based on a weighted L_2-Wasserstein distance measure. *Test*, **11**, 89–107.

de Wet, T. and Venter, J.H. (1972) Asymptotic distributions of certain test criteria of normality. *South African Statist. J.*, **6**, 135–149.

de Wet, T. and Venter, J.H. (1973) Asymptotic distributions for quadratic forms with applications to tests of fit. *Ann. Statist.*, **1**, 380–387.

del Barrio, E. (2001) An elementary derivation of the asymptotic distribution of statistics of Shapiro–Wilk type. Preprint.

del Barrio, E., Cuesta-Albertos, J., Matrán, C. and Rodríguez-Rodríguez, J. (1999a) Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.*, **27**, 1230–1239.

del Barrio, E., Giné, E and Matrán, C. (1999b) Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.*, **27**, 1009–1071.

del Barrio, E., Cuesta-Albertos, J. and Matrán, C. (2000) Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *Test*, **9**, 1–96.

Dunford, N. and Schwartz, J.T. (1963) *Linear Operators, Part II*. New York: Wiley.

Gregory, G. (1997) Functions of order statistics and tests of fit. *South African Statist. J.*, **11**, 99–118.

Guttorp, P. and Lockhart, R.A. (1988) On the asymptotic distribution of quadratic forms in uniform order statistics. *Ann. Statist.*, **16**, 433–449.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, **14**, 1–16.

Jacubowski, A. (1986) Principle of conditioning in limit theorems for sums of random variables. *Ann. Probab.*, **14**, 902–915.

LaRiccia, V. and Mason, D.M. (1986) Cramér–von Mises statistics based on the sample quantile function and estimated parameters. *J. Multivariate Anal.*, **18**, 93–106.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes*. New York: Spinger-Verlag.

Lockhart, R.A. and Stephens, M.A. (1998) The probability plot: test of fit based on the correlation coefficient. In N. Balakrishnan and C.R. Rao (eds), *Order Statistics: Applications*, Handbook of Statistics 17, pp. 453–473. Amsterdam: North-Holland.

Mason, D.M. (1984) Weak convergence of the weighted empirical quantile process in $L_2(0, 1)$. *Ann. Probab.*, **12**, 243–255.

Mason, D.M. (1991) A note on weighted approximations to the uniform empirical and quantile processes. In M.G. Hahn, D.M. Mason and D.C. Weiner (eds), *Sums, Trimmed Sums and Extremes*, Progr. Probab. 23, pp. 269–284. Boston: Birkhäuser.

McLaren, C.G. and Lockhart, R.A. (1987) On the asymptotic efficiency of certain correlation tests of fit. *Canad. J. Statist.*, **15**, 159–167.

Parzen, E. (1979) Nonparametric statistical data modeling. *J. Amer Statist. Assoc.*, **74**, 105–121.

Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* New York: Springer-Verlag.

Samorodnitski, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*. New York: Chapman & Hall.

Schuster, E.F. (1984) Classification of probability laws by tail behavior. *J. Amer. Statist. Assoc.*, **79**, 936–939.

Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* New York: Wiley.

Received January 2004 and revised April 2004