The empirical process in Mallows distance, with application to goodness-of-fit tests

Richard Samworth* and Oliver Johnson
Statistical Laboratory, University of Cambridge,
Wilberforce Road, Cambridge, CB3 0WB, UK.

April 20, 2005

Abstract

Let F be a distribution function which is supported on an interval and possesses a smooth density, and let $\hat{F}_n$ denote the empirical distribution of a random sample of size n from F. We study the asymptotics of the Mallows distance between $\hat{F}_n$ and F. Four different types of limiting result are identified, depending on the tail behaviour of a two-sided hazard function, and examples of each type are given. We discuss applications to goodness-of-fit testing for location-scale and scale families.

Running Title: Empirical process in Mallows distance

Keywords: Bootstrap, empirical process, goodness-of-it, hazard function, Mallows distance, probability metric, Wasserstein distance.

Mathematics Subject Classification: Primary 62E20; Secondary 60F25; 62G30.

*Email:r.samworth@statslab.cam.ac.uk

1 Introduction

The theory of empirical processes has been a central theme in Probability and Statistics for over half a century. Often, a choice of metric is made to compare the empirical distribution of a sample and its underlying population. One such choice is the Mallows distance, also known as the L_2-Wasserstein or Kantorovich distance. This metric has found extensive applications to a wide variety of fields; see Rachev (1984) for a review.

Definition 1. *Let F and G be distribution functions with finite second moment. The Mallows metric $d_2(F, G)$ is defined by*

$$d_2(F, G) = \inf_{\mathcal{T}_{X,Y}} \left\{ \mathbb{E}(X - Y)^2 \right\}^{1/2},$$

where $\mathcal{T}_{X,Y}$ is the set of all pairs of random variables (X, Y) whose marginal distribution functions are F and G respectively.

Basic properties of the Mallows distance are given in Major (1978) and Bickel and Freedman (1981). We recall from these works that we may write

$$d_2(F, G) = \left(\int_0^1 \{F^{-1}(p) - G^{-1}(p)\}^2 \, dp \right)^{1/2},$$

where, for example, $F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}$, and that convergence in the Mallows distance is equivalent to convergence in distribution together with convergence of the second moments.

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with distribution function F, and let $\hat{F}_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}$ denote the empirical distribution of the sample. In this paper, we are concerned with conditions under which an appropriately normalised version of $d_2(\hat{F}_n, F)$ converges in distribution to a non-degenerate limit. This problem and others related to it have been previously studied by several authors, notably Shorack and Wellner (1986), Csörgő and Horváth (1990), Csörgő and Horváth (1993), del Barrio et al. (1999) and del Barrio et al. (2000). Nevertheless, to the best of our knowledge, limiting results have only been given when F belongs to the normal or Weibull families, or is supported on a finite interval.

The Hungarian construction of Kolmós, Major and Tusnády (1975, 1976) and Csörgő and Révész (1978) (see also Csörgő and Horváth (1993)) is a crucial tool in the analysis. In a certain sense, it allows the approximation of $n^{1/2} f(F^{-1}(p))\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}$ by $B_n(p)$, where $\{B_n(p) : 0 \le p \le 1\}$ is a sequence of Brownian bridges, and it may therefore be expected that

$$n \int_0^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2 \, dp \xrightarrow{d} \int_0^1 \frac{B^2(p)}{f(F^{-1}(p))^2} \, dp, \tag{1.1}$$

where $\{B(p) : 0 \le p \le 1\}$ is another Brownian bridge. The results in Section 2 may be summarised as follows:

1. In Section 2.1, we give simple sufficient conditions in terms of a two-sided hazard function for (1.1) to hold. In fact, (1.1) requires surprisingly restrictive conditions, which fail even when F is the normal distribution function.

2. This problem can be overcome to a limited extent by permitting a centering sequence of constants to be subtracted from the left-hand side of (1.1) and modifying the right-hand side accordingly. In Section 2.2, simple sufficient conditions are expressed in terms of the hazard function, and these are satisfied by the normal distribution and a certain range of Weibull distributions.

3. A third limiting regime is found in Section 2.3 by scaling as well as centering the both sides of (1.1). This allows the restrictions on the hazard function to be relaxed substantially, and distributions which fall into this regime include a large subclass of those whose extremes belong to the domains of attraction of the Weibull or Gumbel laws.

4. An interesting feature of the scaling sequence employed in the third regime is that it diverges to infinity slower than any positive power of n. In the fourth class, however, the asymptotic behaviour of the Mallows distance is seen in Section 2.4 to be dominated by the extremes, and we can have $d_2(\hat{F}_n, F) = O_p\{n^{-1+\frac{2}{\alpha}} \ell(n)\}$ for any $\alpha > 2$, where ℓ is a slowly varying function. This class includes the family of Pareto distributions (for which $\ell \equiv 1$), and others whose extremes are in the domain of attraction of a Fréchet distribution. Although this result appears to have fewer statistical applications than those mentioned earlier, it is interesting to see the rate of convergence, and the fact that the limiting distribution is of a rather different form.

In Section 3 we discuss the application of these results to goodness-of-fit tests for location-scale and scale families. We derive the asymptotic null distributions of test statistics based on minimising the Mallows distance between the empirical distribution of the sample and the family in question. These tests may therefore be applied to a very wide range of location-scale and scale families. Some concluding remarks are presented in Section 4.

2 Asymptotic results for the empirical process

Suppose that the random variable X has distribution function F and continuous density f. Define $F^{-1}(p) = \inf\{x \in \mathbb{R} : F(x) \geq p\}$ for $p \in (0, 1]$, and let $a = \sup\{x : F(x) = 0\}, b = F^{-1}(1)$. Note that we allow $a = -\infty$, $b = \infty$, and assume:

(A1): f is positive and continuously[1] differentiable on (a, b).

Thus, in particular, we suppose that the support of the underlying distribution is an interval.

Definition 2. *The two-sided hazard function $h : (a, b) \to \mathbb{R}$ of X is given by*

$$h(x) = \begin{cases} f(x)/\bar{F}(x) & \text{if } x \geq F^{-1}(1/2), \\ f(x)/F(x) & \text{if } x < F^{-1}(1/2), \end{cases}$$

where $\bar{F}(x) = \mathbb{P}(X > x)$.

The distribution function may be recovered from the hazard function and the median by means of the equations

$$F(x) = 1 - \frac{1}{2}\exp\left(-\int_{F^{-1}(1/2)}^{x} h(t)\,dt\right), \quad x \geq F^{-1}(1/2), \tag{2.1}$$

$$F(x) = \frac{1}{2}\exp\left(-\int_{x}^{F^{-1}(1/2)} h(t)\,dt\right), \quad x < F^{-1}(1/2). \tag{2.2}$$

As was mentioned in the introduction, the behaviour of the empirical process in the Mallows distance will depend crucially on the tail behaviour of the hazard function. We now describe each of the different regimes in turn.

[1]While it is not always strictly necessary to assume f is continuously differentiable, rather than merely differentiable, this is assumed throughout for convenience.

2.1 Hazard function rapidly diverging: no centering required

In this subsection, our assumptions on the hazard function are as follows:

(A2): $\sup_{x\in(a,b)}\left|\frac{d}{dx}\left(\frac{1}{h(x)}\right)\right| < \infty$;

(A3): $h(x) \to \infty$ as $x \to a$ and as $x \to b$;

(A4): $\int_a^b \frac{1}{h(x)}\,dx < \infty$.

Before describing the asymptotic behaviour of the empirical process, we pause briefly to consider the strength of these conditions. Table 1 presents the hazard functions of several common distributions, together with ticks or crosses according to whether they satisfy each of the conditions above. We see that (**A2**) is very weak, and is satisfied by almost all distributions encountered in applications. Assumptions (**A3**) and (**A4**) are quite restrictive tail conditions, with (**A4**) failing even for the normal distribution, although it holds for all Weibull distributions with parameter $\alpha > 2$.

Distribution	$h(x)$	(**A2**) $\sup\left\|\frac{d}{dx}(1/h(x))\right\| < \infty$	(**A3**) $h(x) \to \infty$	(**A4**) $\int 1/h(x)\,dx < \infty$
Uniform	$1/(b-x)$	$\surd$	$\surd$	$\surd$
Weibull	$\alpha x^{\alpha-1}$	$\surd$	$\surd$ iff $\alpha > 1$	$\surd$ iff $\alpha > 2$
Normal	$\phi(x)/\bar{\Phi}(x)$	$\surd$	$\surd$	$\times$
Lognormal	$\frac{\phi(\log x)}{x\bar{\Phi}(\log x)}$	$\surd$	$\times$	$\times$
Pareto	α/x	$\surd$	$\times$	$\times$

Table 1: Hazard functions of common distributions, and whether they satisfy conditions (**A2**)–(**A4**). For simplicity, each hazard function is only presented for $x \geq F^{-1}(1/2)$, although no overall conclusions are altered by considering the left-hand tail as well.

Theorem 1. *Let $X_1, \ldots, X_n$ be independent and identically distributed with distribution function F and density f, and let $\hat{F}_n$ denote the empirical distribution function. Assume that (**A1**)–(**A4**) hold. Then*

$$nd_2^2(\hat{F}_n, F) \xrightarrow{d} \int_0^1 \frac{B^2(p)}{f(F^{-1}(p))^2}\,dp \tag{2.3}$$

as $n \to \infty$, where $\{B(p) : 0 \le p \le 1\}$ is a Brownian bridge.

Proof. We apply a particular form of the Hungarian construction as follows. Assumption (**A1**) covers Conditions (i) and (ii) of Lemma 6.1.1 of Csörgő and Horváth (1993), while (**A2**) is equivalent to Condition (iii) of the same lemma. These are used in their Theorem 6.2.1(ii) to prove that for any $\nu \in (0, 1/2]$, there exists a probability space on which may be defined a sequence of Brownian bridges $\{B_n(p), 0 \le p \le 1\}$ such that

$$n^{(1/2)-\nu} \sup_{\frac{1}{n+1} \le p \le \frac{n}{n+1}} \frac{|\rho_n(p) - B_n(p)|}{\{p(1-p)\}^\nu} = O_p(1), \tag{2.4}$$

where $\rho_n(p) = n^{1/2} f\big(F^{-1}(p)\big)\big\{\hat{F}_n^{-1}(p) - F^{-1}(p)\big\}$. Now write

$$\begin{aligned}\left| n d_2^2(\hat{F}_n, F) - \int_0^1 \frac{B_n^2(p)}{f(F^{-1}(p))^2} \right| dp &\le \left| \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\rho_n^2(p) - B_n^2(p)}{f(F^{-1}(p))^2}\, dp \right| \\ &\quad + \left| \int_A \frac{B_n^2(p)}{f(F^{-1}(p))^2}\, dp \right| + \left| \int_A \frac{\rho_n^2(p)}{f(F^{-1}(p))^2}\, dp \right| \\ &\equiv |L_n^{(1)}| + |L_n^{(2)}| + |L_n^{(3)}|,\end{aligned}$$

say, where $A = (0, \frac{1}{n+1}) \cup (\frac{n}{n+1}, 1)$. The result will follow if we can show that each of these three terms converges in probability to zero.

1. *Bound for the contribution of $L_n^{(1)}$, using (**A1**),(**A2**) and (**A3**):*

Proceeding as in the proof of Proposition 2 of del Barrio et al. (1999) using the inequality $|x^2 - y^2| \le (x-y)^2 + 2|x-y||y|$, we find $L_n^{(1)} \xrightarrow{p} 0$ as $n \to \infty$ provided that $\lim_{n\to\infty} A_n^{(1)} = 0$ and $\lim_{n\to\infty} A_n^{(2)} = 0$, where

$$A_n^{(1)} = n^{2\nu-1} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\{p(1-p)\}^{2\nu}}{f(F^{-1}(p))^2}\, dp \quad \text{and} \quad A_n^{(2)} = n^{\nu-\frac{1}{2}} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\{p(1-p)\}^{\nu+\frac{1}{2}}}{f(F^{-1}(p))^2}\, dp. \tag{2.5}$$

Now, for $\nu \in (0, 1/2)$, we see that $A_n^{(1)}$ clearly converges to 0 if the integral over $(0, 1)$ converges, and otherwise l'Hôpital's rule gives that the limit is zero provided that $h(x) \to \infty$ as $x \to a$ and as $x \to b$, which is (**A3**). The same argument applies to $A_n^{(2)}$.

2. *Bound for the contribution of $L_n^{(2)}$, using (**A1**) and (**A4**):*

The case $p = 2$, $q(\cdot) = f(F^{-1}(\cdot))$ of Lemma 5.3.2 of Csörgő and Horváth (1993) may be restated as

$$\mathbb{P}\left(\int_0^1 \frac{B^2(p)}{f(F^{-1}(p))^2}\,dp < \infty\right) = \begin{cases} 1 & \text{if } \int_a^b 1/h(x)\,dx < \infty \\ 0 & \text{if } \int_a^b 1/h(x)\,dx = \infty. \end{cases}$$

Thus, under condition (**A4**), we see that $L_n^{(2)} \xrightarrow{a.s.} 0$ as $n \to \infty$.

3. *Bound for the contribution of $L_n^{(3)}$, using (**A1**) and (**A3**)*:

It suffices to show that

$$\int_0^{\frac{1}{n}} n\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\,dp \xrightarrow{p} 0 \quad \text{and} \quad \int_{\frac{n-1}{n}}^1 n\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\,dp \xrightarrow{p} 0,$$

which is Condition (3.16) of del Barrio et al. (2000). As the arguments used to prove the two statements are very similar, here we prove only the second. Let $X_{(n)} = \max_i X_i$, and observe that

$$n\int_{\frac{n-1}{n}}^1 \left(\hat{F}_n^{-1}(p) - F^{-1}(p)\right)^2 dp$$
$$\leq 2\left\{X_{(n)} - F^{-1}\left(\tfrac{n-1}{n}\right)\right\}^2 + 2n\int_{F^{-1}\left(\frac{n-1}{n}\right)}^b \left\{x - F^{-1}\left(\tfrac{n-1}{n}\right)\right\}^2 f(x)\,dx. \qquad (2.6)$$

If $b < \infty$, then the first term of (2.6) clearly converges in probability to zero. On the other hand, if $b = \infty$, then Theorem 4.1.2 of Galambos (1978) gives that $X_{(n)} - F^{-1}\left(\frac{n-1}{n}\right)$ converges in probability to zero if and only if

$$\lim_{x\to\infty} \frac{\bar{F}(t+x)}{\bar{F}(x)} = 0. \qquad (2.7)$$

for each $t > 0$. But using (2.1) we see that (**A3**) is sufficient to ensure that (2.7) holds.

To deal with the second term in (2.6), we use arguments based on those which establish Poincaré inequalities in Sysoeva (1965). The simplest possible case of the main theorem of Sysoeva (1965) gives a Hardy inequality; that is, if $t \geq F^{-1}(\frac{n-1}{n})$ and $G : [t, b) \to [0, \infty]$ is a differentiable function with $G(t) = 0$ and $\mathbb{E}\{G(X)^2\} < \infty$, then

$$\int_t^b f(x)G(x)^2\,dx \leq 4\int_t^b \frac{\bar{F}(x)^2}{f(x)} g(x)^2\,dx = 4\int_t^b \frac{\bar{F}(x)}{h(x)} g(x)^2\,dx, \qquad (2.8)$$

where $g(x) = G'(x)$.

[For the sake of completeness, note that (2.8) can be proved by integration by parts:

$$\int_t^b f(x)G(x)^2\,dx = 2\int_t^b \bar{F}(x)G(x)g(x)\,dx$$
$$\leq 2\Big(\int_t^b f(x)G(x)^2 dx\Big)^{1/2}\Big(\int_t^b \frac{\bar{F}(x)^2}{f(x)} g(x)^2 dx\Big)^{1/2},$$

by Cauchy-Schwarz.] Applying (2.8) with $G(x) = x - t$ and $t = F^{-1}(\frac{n-1}{n})$, we find that

$$n\int_{F^{-1}(\frac{n-1}{n})}^b \big\{x - F^{-1}\big(\tfrac{n-1}{n}\big)\big\}^2 f(x)\,dx \leq 4n\int_{F^{-1}(\frac{n-1}{n})}^b \frac{\bar{F}(x)}{h(x)}\,dx$$
$$= 4n\int_n^\infty \frac{1}{x^2 h(F^{-1}(\frac{x-1}{x}))^2}\,dx. \tag{2.9}$$

This right-hand side converges to zero using l'Hôpital's rule and (**A3**), and this completes the proof. □

2.2 Hazard function diverging at moderate rate: centering required

The integral in (2.3) is infinite almost surely when (**A4**) fails. Nevertheless, it is still meaningful in an L_2-sense under weaker conditions. This idea was first exploited by del Barrio et al. (1999) in their study of a normal underlying population, and was used to cover the case of a Weibull distribution with parameter $\alpha \in (\frac{4}{3}, 2]$ by Csörgő in his discussion of the paper by del Barrio et al. (2000). In order to extend these results to a more general context, we require a further condition on the hazard function:

(A5): $\liminf_{x\to a} \frac{d}{dx}\big(\frac{1}{h(x)}\big) > -1/2$ and $\limsup_{x\to b} \frac{d}{dx}\big(\frac{1}{h(x)}\big) < 1/2$.

As we will also need to assume that the hazard function diverges in the tails, we must have $\limsup_{x\to a} \frac{d}{dx}\big(\frac{1}{h(x)}\big) \geq 0$ and $\liminf_{x\to b} \frac{d}{dx}\big(\frac{1}{h(x)}\big) \leq 0$. Thus, condition (**A5**) rules out large oscillations of the reciprocal of the hazard function in the tails, and holds in particular if the hazard function is eventually monotone at both ends.

In (2.10) below, the integral on the right-hand side is to be understood as the L_2-limit of

$$\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{B^2(p) - \mathbb{E}\{B^2(p)\}}{f(F^{-1}(p))^2}\,dp.$$

Theorem 2. *Let $X_1, \ldots, X_n$ be independent and identically distributed with distribution function F and density f. Assume that (**A1**), (**A2**), (**A3**) and (**A5**) hold. Then*

$$nd_2^2(\hat{F}_n, F) - \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{p(1-p)}{f(F^{-1}(p))^2}\,dp \xrightarrow{d} \int_0^1 \frac{B^2(p) - \mathbb{E}\{B^2(p)\}}{f(F^{-1}(p))^2}\,dp \tag{2.10}$$

if and only if

$$\int_a^b \frac{1}{h(x)^3}\,dx < \infty.$$

Before proving this theorem, we state and prove a lemma:

Lemma 3. *Assume (**A1**).*

- *(i) If $\int_0^1 \int_0^1 \frac{\{s \wedge t - st\}^2}{f(F^{-1}(s))^2 f(F^{-1}(t))^2}\,ds\,dt < \infty$ and (**A2**) holds, then $\int_a^b \frac{1}{h(x)^3}\,dx < \infty$.*
- *(ii) If $\int_a^b \frac{1}{h(x)^3}\,dx < \infty$ and (**A3**) and (**A5**) hold, then $\int_0^1 \int_0^1 \frac{\{s \wedge t - st\}^2}{f(F^{-1}(s))^2 f(F^{-1}(t))^2}\,ds\,dt < \infty$.*

Proof. To prove (i), we only show that $\int_{F^{-1}(1/2)}^b \frac{1}{h(x)^3}\,dx < \infty$ because the arguments for the two tails are very similar. Observe that

$$\begin{aligned}\int_{1/2}^1 \int_t^1 \frac{(1-s)^2 t^2}{f(F^{-1}(s))^2 f(F^{-1}(t))^2}\,ds\,dt &= \int_{F^{-1}(1/2)}^b \int_y^b \frac{\bar{F}(x)^2 F(y)^2}{f(x)f(y)}\,dx\,dy \\ &\geq \frac{1}{4} \int_{F^{-1}(1/2)}^b \frac{1}{h(y)^3} \left\{ \frac{h(y)^2}{\bar{F}(y)} \int_y^b \frac{f(x)}{h(x)^2}\,dx \right\} dy,\end{aligned}$$

since $F(y)^2 \geq 1/4$ for $y \geq F^{-1}(1/2)$. Thus it suffices to show that the term in braces in the final term above is bounded away from zero as $y \to b$. Now integration by parts yields

$$\frac{\bar{F}(y)}{h(y)^2} = \int_y^b \frac{f(x)}{h(x)^2}\,dx + 2\int_y^b \bar{F}(x) \frac{1}{h(x)} \Big(\frac{1}{h(x)}\Big)'\,dx = \int_y^b \frac{f(x)}{h(x)^2} \Big\{1 + 2\Big(\frac{1}{h(x)}\Big)'\Big\}\,dx. \tag{2.11}$$

Hence, by rearranging,

$$\liminf_{y\to b} \frac{h(y)^2}{\bar{F}(y)} \int_y^b \frac{f(x)}{h(x)^2}\,dx \geq \frac{1}{1+2\limsup_{y\to b}(\frac{1}{h(y)})'},$$

and under (**A2**), this right-hand side is strictly positive. This proves (i).

To prove (ii), by symmetry it is enough to prove the double integral is finite when the inner integral is taken over values of s lying between t and 1. Now, by Fubini's theorem,

$$\begin{aligned}
&\int_0^1 \int_t^1 \frac{(1-s)^2t^2}{f(F^{-1}(s))^2 f(F^{-1}(t))^2}\,ds\,dt \\
&= \int_0^{1/2} \int_{1/2}^1 \frac{(1-s)^2t^2}{f(F^{-1}(s))^2 f(F^{-1}(t))^2}\,ds\,dt + 2\int_{1/2}^1 \int_t^1 \frac{(1-s)^2t^2}{f(F^{-1}(s))^2 f(F^{-1}(t))^2}\,ds\,dt \\
&\leq \int_2^\infty \frac{dy}{y^2 h(F^{-1}(\frac{1}{y}))^2} \int_2^\infty \frac{dx}{x^2 h(F^{-1}(\frac{x-1}{x}))^2} + 2\int_{F^{-1}(\frac{1}{2})}^b \frac{1}{h(y)^3}\left\{\frac{h(y)^2}{\bar{F}(y)} \int_y^b \frac{f(x)}{h(x)^2}\,dx\right\} dy.
\end{aligned}$$

The first term on the right-hand side is finite under (**A3**), while the second is finite provided that the term in braces is bounded as $y \to b$. But from (2.11),

$$\limsup_{y\to b} \frac{h(y)^2}{\bar{F}(y)} \int_y^b \frac{f(x)}{h(x)^2}\,dx \leq \frac{1}{1+2\liminf_{y\to b}(\frac{1}{h(y)})'},$$

so the result follows from (**A5**). □

Proof of Theorem 2.
By a virtually identical argument to that employed in the proof of Theorem 1, we see that the stated convergence in distribution holds if $\tilde{L}_n^{(2)} \xrightarrow{p} 0$, where

$$\tilde{L}_n^{(2)} = \int_A \frac{B^2(p) - \mathbb{E}\{B^2(p)\}}{f(F^{-1}(p))^2}\,dp.$$

In turn, then, it is enough to show that the sequence of random variables (M_n) converges in $L_2(0,1)$, where

$$M_n = \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{B^2(p) - \mathbb{E}\{B^2(p)\}}{f(F^{-1}(p))^2}\,dp.$$

But

$$\mathbb{E}(M_n^2) \to \int_0^1 \int_0^1 \frac{2(s \wedge t - st)^2}{f(F^{-1}(s))^2 f(F^{-1}(t))^2} \, ds \, dt$$

and by Lemma 3, this is finite (and so M_n converges in L_2) provided that $1/h(x)^3$ is integrable.

Conversely, if $\int_a^b \frac{1}{h(x)^3}\, dx = \infty$, then Lemma 3 shows that A_n does not converge in L_2, so the right-hand side of (2.10) does not make sense as an L_2-limit.

□

Theorem 2 applies in particular when the underlying distribution is normal or a Weibull distribution with parameter $\alpha \in \left(\frac{4}{3}, 2\right]$, which were the cases studied in del Barrio et al. (2000). In each case it is straightforward using Table 1 to verify the integrability of $1/h(x)^3$. Nevertheless, this remains quite a restrictive condition on the hazard function – it fails, for example, with the exponential distribution – and further work is required in the next section to broaden the scope of our results.

2.3 Intermediate hazard function behaviour: centering and scaling required

We follow the idea of Csörgő in his stimulating discussion of the Weibull scale family in del Barrio et al. (2000), and seek conditions under which a centered and scaled version of $nd_2^2(\hat{F}_n, F)$ may be expressed asymptotically as a sequence of random variables of precise order 1 in probability, defined in terms of the sequence of Brownian bridges from the Hungarian construction (2.4). We will require the following strengthening of (**A5**):

(A6): $\lim_{x\to a} \frac{d}{dx}\left(\frac{1}{h(x)}\right) = 0$ and $\lim_{x\to b} \frac{d}{dx}\left(\frac{1}{h(x)}\right) = 0$.

Theorem 4. *Assume (**A1**) and (**A2**). Suppose that at least one of the following statements holds:*

*(a) $h(x) \to \infty$ as $x \to b$ and the part of (**A5**) concerning the right-hand tail holds;*

*(b) the part of (**A6**) concerning the right-hand tail holds.*

Suppose further that at least one of the corresponding versions of (a) and (b) holds for the left-hand tail. Then

$$\frac{1}{c_n}\left\{nd_2^2(\hat{F}_n, F) - \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{p(1-p)}{f(F^{-1}(p))^2}\,dp\right\} = \frac{1}{c_n}\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{B_n^2(p) - \mathbb{E}\{B_n^2(p)\}}{f(F^{-1}(p))^2}\,dp + o_p(1), \tag{2.12}$$

with

$$c_n^2 = \int_{F^{-1}(\frac{1}{n+1})}^{F^{-1}(\frac{n}{n+1})} \frac{1}{h(x)^3}\,dx.$$

Proof. As the argument involves treating the two tails separately, here we prove only that if (a) or (b) holds for the right-hand tail, then

$$\frac{1}{c_n}\left(\int_{1/2}^{1} \rho_n^2(p)\,dp - \int_{1/2}^{\frac{n}{n+1}} \frac{p(1-p)}{f(F^{-1}(p))^2}\,dp\right) = \frac{1}{c_n}\int_{1/2}^{\frac{n}{n+1}} \frac{B_n^2(p) - \mathbb{E}\{B_n^2(p)\}}{f(F^{-1}(p))^2}\,dp + o_p(1), \tag{2.13}$$

with

$$c_n^2 = \int_{F^{-1}(1/2)}^{F^{-1}(\frac{n}{n+1})} \frac{1}{h(x)^3}\,dx.$$

Notice that the first terms on the right-hand sides of (2.13) and (2.12) are of precise order 1 in probability, by Lemma 3.

First, suppose that (a) holds for the right-hand tail. Since $\mathbb{E}\{B^2(p)\} = p(1-p)$, we have as in the proof of Theorem 1 that

$$\int_{1/2}^{1} \rho_n^2(p)\,dp - \int_{1/2}^{\frac{n}{n+1}} \frac{p(1-p)}{f(F^{-1}(p))^2}\,dp = \int_{1/2}^{\frac{n}{n+1}} \frac{B_n^2(p) - \mathbb{E}\{B_n^2(p)\}}{f(F^{-1}(p))^2}\,dp + o_p(1),$$

so the result follows from the fact that (c_n) is bounded away from zero for large n.

Now suppose that (b) holds for the right-hand tail. Combining the error (2.5) from the Hungarian construction for $\nu \in (0, 1/2)$ with the two terms in (2.6) from the tail of the empirical process, it suffices to show that the three terms

$$n^{2\nu-1}\int_{1/2}^{\frac{n}{n+1}} \frac{\{p(1-p)\}^{2\nu}}{f(F^{-1}(p))^2}\,dp, \quad \left\{X_{(n)} - F^{-1}\left(\tfrac{n-1}{n}\right)\right\}^2 \quad \text{and} \quad n\int_n^{\infty} \frac{1}{x^2 h(F^{-1}(\frac{x-1}{x}))^2}\,dx \tag{2.14}$$

are $o(c_n)$, $o_p(c_n)$ and $o(c_n)$ respectively as $n \to \infty$. As a preliminary, we observe that the function $\ell(y) = 1/h(F^{-1}(\frac{y-1}{y}))$ is slowly varying. This follows by the argument on p. 15 of Bingham, Goldie and Teugels (1987), because

$$\frac{y\ell'(y)}{\ell(y)} = \frac{d}{dx}\Big(\frac{1}{h(x)}\Big)\Big|_{x=F^{-1}\left(\frac{y-1}{y}\right)} \to 0 \tag{2.15}$$

as $y \to \infty$ using (**A6**), and because the function in (2.15) is continuous. It follows that

$$\begin{aligned} n^{2\nu-1}\int_{1/2}^{\frac{n}{n+1}} \frac{\{p(1-p)\}^{2\nu}}{f(F^{-1}(p))^2}\,dp &\le n^{2\nu-1}\int_2^{n+1} \frac{1}{y^{2\nu}h(F^{-1}(\frac{y-1}{y}))^2}\,dy \\ &\sim \frac{1}{(1-2\nu)h(F^{-1}(\frac{n}{n+1}))^2}, \end{aligned}$$

by Karamata's theorem (Theorem 1.5.11(i) of Bingham, Goldie and Teugels (1987)). Similarly, for any constant C,

$$c_n = \left(\int_2^{n+1} \frac{1}{yh(F^{-1}(\frac{y-1}{y}))^4}\,dy\right)^{1/2} \ge \frac{C}{h(F^{-1}(\frac{n}{n+1}))^2},$$

for sufficiently large n. Thus the first term in (2.14) is $o(c_n)$ as $n \to \infty$.

For the second term, Theorem 2.7.2 of Galambos (1978) shows that the part of (**A6**) concerning the right-hand tail is sufficient to ensure that

$$\mathbb{P}\left(\frac{X_{(n)} - F^{-1}(\frac{n-1}{n})}{b_n} \le x\right) \to \exp(-e^{-x}), \quad x \in \mathbb{R},$$

with $b_n = n\int_{F^{-1}(\frac{n-1}{n})}^{b} \bar{F}(x)\,dx$. But

$$b_n = n\int_n^{\infty} \frac{1}{x^2 h(F^{-1}(\frac{x-1}{x}))}\,dx \sim \frac{1}{h(F^{-1}(\frac{n-1}{n}))}, \tag{2.16}$$

by the second part of Karamata's theorem (Theorem 1.5.11(ii) of Bingham, Goldie and Teugels (1987)). Therefore $\{X_{(n)} - F^{-1}(\frac{n-1}{n})\}^2 = O_p(b_n^2) = o_p(c_n)$. A very similar argument applies shows that the third term in (2.14) is $o(c_n)$ as $n \to \infty$. □

While it is disappointing not to be able to give a more explicit expression for the limiting distribution in Theorem 4, there are two important points to note. Firstly, the contributions of all of the order statistics to $\int_0^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2 \, dp$ are of the same order (it is not the case, for instance, that the contributions from the extreme order statistics dominate those from the central ones). Secondly, our asymptotically equivalent sequence on the right-hand side of (2.12) is constructed from a Gaussian process. Neither of these results will remain true when the hazard function decays at a faster rate, as in the next section.

Theorem 4 is applicable to many distributions of interest.

1. Recall that the Weibull distributions with parameter $\alpha \in (0, \frac{4}{3}]$ were not covered by Theorems 1 and 2. However, for the right-hand tail, statement (a) holds for $\alpha \in (1, \frac{4}{3}]$ and (b) holds for all $\alpha \in \big(0, \frac{4}{3}\big]$, while the version of (a) for the left-hand tail holds for all α. It is straightforward to verify (as Csörgő found by a different method) that we may take $c_n^2 = \log\log n$ for $\alpha = 4/3$ and $c_n^2 = \log^{\frac{4}{\alpha}-3} n$ for $\alpha < 4/3$.

2. For the inverse Gaussian distribution, with $f(x) = (2\pi x^3)^{-1/2} \exp\big\{-\frac{(x-1)^2}{2x}\big\}$ for $x > 0$ and $h(x) \to 1/2$ as $x \to \infty$, we may take $c_n^2 = \log n$.

3. The lognormal is another interesting case, as its hazard function decays to zero in the right-hand tail faster than any of the Weibull distributions. Nevertheless, the right-hand tail still satisfies condition (b) of Theorem 4, while the left-hand tail satisfies (a). We may take $c_n^2 = \exp\big\{4\sqrt{2\log n}\big\}/\log^{\frac{3}{2}} n$.

2.4 Hazard function decaying rapidly: non-Gaussian limit

Of the distributions in Table 1, the only family not covered so far is the class of Pareto distributions. This class satisfies the following condition:

(A7): $b = \infty$ and $xh(x) \to \alpha$ as $x \to \infty$, for some $\alpha \in (0, \infty)$.

By Theorems 8.13.2 and 1.5.12 of Bingham, Goldie and Teugels (1987), a distribution function F belongs to the domain of attraction of the Fréchet distribution with parameter α, i.e.

$$\mathbb{P}\left(\frac{X_{(n)}}{F^{-1}(\frac{n-1}{n})} \leq x\right) \to \begin{cases} \exp(-x^{-\alpha}) & \text{if } x > 0, \\ 0 & \text{otherwise,} \end{cases}$$

if and only if we may write $F^{-1}(p) = (1-p)^{-1/\alpha}\ell(\frac{1}{1-p})$, where ℓ is slowly varying. If (**A1**) is satisfied, then (**A7**) is equivalent to the stronger condition that ℓ is normalised slowly varying, as defined in (1.3.4) of Bingham, Goldie and Teugels (1987).

Theorem 5. *Assume that (**A1**) and (**A2**) hold, and that (**A7**) holds for some $\alpha > 2$. Let (Y_n) denote a sequence of independent exponential random variables with mean 1, and write $\tilde{S}(u) = \sum_{1 \leq i < u+1} Y_i$, for $u \geq 1$. Then*

$$\frac{n}{F^{-1}(\frac{n-1}{n})^2} \int_{1/2}^{1} \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\, dp$$
$$\overset{d}{\to} \frac{1}{Y_1^{2/\alpha}} - \frac{2\alpha}{(\alpha-1)Y_1^{1/\alpha}} + \frac{\alpha}{\alpha-2} + \int_1^\infty \frac{1}{u^{2/\alpha}}\left\{\left(\frac{\tilde{S}(u)}{u}\right)^{-1/\alpha} - 1\right\}^2 du \quad (2.17)$$

as $n \to \infty$.

The proof of Theorem 5 is deferred to the Appendix. For clarity of exposition, the result is stated for the right-hand tail. Under corresponding conditions, an analogous theorem may be proved for the left-hand tail. Moreover, if the right-hand tail satisfies the hypotheses of Theorem 5, and the left-hand tail also satisfies the hypotheses of any of the earlier theorems, then the results for the two tails may be combined to yield a limit theorem for $d_2(\hat{F}_n, F)$, as we now demonstrate by example.

For the Pareto distribution with parameter $\alpha > 2$, Theorem 5 implies in particular that $\int_{1/2}^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\, dp$ is of precise order $n^{-1+\frac{2}{\alpha}}$ in probability. On the other hand, the left-hand tail satisfies the hypotheses of Theorem 1. It follows that $\int_0^{1/2} \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\, dp = O_p(n^{-1})$, and hence that

$$n^{1-\frac{2}{\alpha}} d_2^2(\hat{F}_n, F) \overset{d}{\to} \frac{1}{Y_1^{2/\alpha}} - \frac{2\alpha}{(\alpha-1)Y_1^{1/\alpha}} + \frac{\alpha}{\alpha-2} + \int_1^\infty \frac{1}{u^{2/\alpha}}\left\{\left(\frac{\tilde{S}(u)}{u}\right)^{-1/\alpha} - 1\right\}^2 du.$$

Another family of distributions covered by Theorem 5 is the log-logistic, with hazard function $h(x) = \frac{\alpha x^{\alpha-1}}{1+x^\alpha}$ in the right-hand tail, provided that $\alpha > 2$. We also mention here the situation where (**A1**) and (**A2**) hold, $a = -\infty$, $b = \infty$ and $|x|h(x) \to \alpha$ as

$x \to \pm\infty$, for some $\alpha > 2$. In this case, the contributions from both tails are of the same stochastic order, and

$$\frac{nd_2^2(\hat{F}_n, F)}{F^{-1}(\frac{n-1}{n})^2} \xrightarrow{d} \frac{1}{Y_1^{2/\alpha}} - \frac{2\alpha}{(\alpha-1)Y_1^{1/\alpha}} + \frac{\alpha}{\alpha-2} + \int_1^\infty \frac{1}{u^{2/\alpha}}\left\{\left(\frac{\tilde{S}(u)}{u}\right)^{-1/\alpha} - 1\right\}^2 du$$
$$+ \frac{1}{Z_1^{2/\alpha}} - \frac{2\alpha}{(\alpha-1)Z_1^{1/\alpha}} + \frac{\alpha}{\alpha-2} + \int_1^\infty \frac{1}{u^{2/\alpha}}\left\{\left(\frac{\tilde{T}(u)}{u}\right)^{-1/\alpha} - 1\right\}^2 du,$$

where (Z_n) is a sequence of independent exponential random variables with mean 1, independent of the (Y_n), and $\tilde{T}(u) = \sum_{1\leq i<u+1} Z_i$, for $u \geq 1$. This follows from Theorem 5 and Rossberg's lemma on the asymptotic independence of functions of order statistics (cf. Lemma 5.1.4 of Csörgő and Horváth (1993)).

Finally, Theorem 5 may be adapted to cover the situation where $\alpha = 2$, but F still has finite variance, as in the example in Proposition 3.4 of del Barrio et al. (2000). The centering sequence $\int_{\frac{n}{n+1}}^1 F^{-1}(p)^2\,dp$ is required, and the $\alpha/(\alpha-2)$ term should be removed from the right-hand side of (2.17).

3 Application to goodness-of-fit tests

In this section, we consider goodness-of-fit tests of location-scale and scale families based on minimising the Mallows distance between the empirical distribution of the sample and the set of distributions specified by the null hypothesis. Let G denote a distribution function with mean m and finite variance $\tau^2 > 0$. Further, let $G_{\mu,\sigma}(\cdot) = G\left(\frac{\cdot-\mu}{\sigma}\right)$ and $G_\sigma(\cdot) = G\left(\frac{\cdot}{\sigma}\right)$, so that $\mathcal{G}_{\mathrm{ls}} = \{G_{\mu,\sigma} : \mu \in \mathbb{R}, \sigma > 0\}$ and $\mathcal{G}_{\mathrm{s}} = \{G_\sigma : \sigma > 0\}$ are respectively the location-scale and scale families associated with G. As in the previous section, we assume $X_1, \ldots, X_n$ are independent and identically distributed with distribution function F, and let $\hat{F}_n$ denote the empirical distribution function. We wish to test the null hypothesis $H_0 : F \in \mathcal{G}_{\mathrm{ls}}$ or $H_0 : F \in \mathcal{G}_{\mathrm{s}}$.

In the location-scale case, we seek to compute $d_2(\hat{F}_n, \mathcal{G}_{\mathrm{ls}}) = \inf_{\mu,\sigma} d_2(\hat{F}_n, G_{\mu,\sigma})$, and a calculation based on that in del Barrio et al. (2000) shows that this infimum is attained at

$$(\hat{\mu}, \hat{\sigma}) = \left(\bar{X}_n - \frac{m}{\tau^2}\int_0^1 \{\hat{F}_n^{-1}(p) - \bar{X}_n\}G^{-1}(p)\,dp\,,\; \frac{1}{\tau^2}\int_0^1 \{\hat{F}_n^{-1}(p) - \bar{X}_n\}G^{-1}(p)\,dp\right),$$

where $\bar{X}_n = n^{-1}\sum X_i$. It follows that

$$d_2(\hat{F}_n, \mathcal{G}_{\text{ls}}) = S_n^2 - \frac{1}{\tau^2}\left(\int_0^1 \{\hat{F}_n^{-1}(p) - \bar{X}_n\} G^{-1}(p)\, dp\right)^2,$$

where $S_n^2 = n^{-1}\sum(X_i - \bar{X}_n)^2$ is the sample variance. Considerations of location and scale invariance then suggest the test statistic

$$R_n = 1 - \frac{1}{\tau^2 S_n^2}\left(\int_0^1 \{\hat{F}_n^{-1}(p) - \bar{X}_n\} G^{-1}(p)\, dp\right)^2, \tag{3.1}$$

with large values leading to the rejection of the null hypothesis.

For scale families, we write $d_2(\hat{F}_n, \mathcal{G}_{\text{s}}) = \inf_\sigma d_2(\hat{F}_n, G_\sigma)$, and the infimum is attained at $\hat{\sigma} = \mu_2(G)^{-1}\int_0^1 \hat{F}_n^{-1} G^{-1}(p)\, dp$, and as in Csörgő's discussion of del Barrio et al. (2000), scale invariance leads to the test statistic

$$T_n = 1 - \frac{1}{\mu_2(G)\mu_2(\hat{F}_n)}\left(\int_0^1 \hat{F}_n^{-1}(p) G^{-1}(p)\, dp\right)^2, \tag{3.2}$$

where $\mu_2(F) = \int x^2\, dF(x)$, so that $\mu_2(\hat{F}_n) = n^{-1}\sum X_i^2$.

That in each case we have explicit expressions for the values of the parameter(s) which minimise the Mallows distance between $\hat{F}_n$ and the class of null hypothesis distributions is an attractive feature of the tests. Location and scale invariance means that the distribution of R_n under any member of $\mathcal{G}_{\text{ls}}$ is the same, so we may choose $G = F$ and moreover assume F has mean zero and unit variance when deriving the asymptotic null distribution. The importance of the results of the previous section in this context stems from the fact that we may study the asymptotic null distributions of R_n and T_n through the modified statistics $R_n^* = \tau^2 S_n^2 R_n$ and $T_n^* = \mu_2(G)\mu_2(\hat{F}_n)T_n$. Recalling that now $m = 0$, $\tau^2 = 1$, $\mu_2(G) = 1$, we may express these as

$$R_n^* = d_2^2(\hat{F}_n, F) - \left[\int_0^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}\, dp\right]^2 - \left[\int_0^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\} F^{-1}(p)\, dp\right]^2, \tag{3.3}$$

$$T_n^* = d_2^2(\hat{F}_n, F) - \left[\int_0^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\} F^{-1}(p)\, dp\right]^2.$$

The notation in the corollary below is as in Section 2. We sketch the details of the proof in the Appendix.

Corollary 6. *(i) Assume the conditions of Theorem 1. Then, under H_0,*

$$nR_n \xrightarrow{d} \int_0^1 \frac{B^2(p)}{f(F^{-1}(p))^2}\,dp - \left\{\int_0^1 \frac{B(p)}{f(F^{-1}(p))}\,dp\right\}^2 - \left\{\int_0^1 \frac{B(p)F^{-1}(p)}{f(F^{-1}(p))}\,dp\right\}^2. \tag{3.4}$$

(ii) Assume the conditions of Theorem 2. Then, under H_0,

$$nR_n - \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{p(1-p)}{f(F^{-1}(p))^2}\,dp \xrightarrow{d} \int_0^1 \frac{B^2(p) - \mathbb{E}\{B^2(p)\}}{f(F^{-1}(p))^2}\,dp - \left\{\int_0^1 \frac{B(p)}{f(F^{-1}(p))}\,dp\right\}^2 - \left\{\int_0^1 \frac{B(p)F^{-1}(p)}{f(F^{-1}(p))}\,dp\right\}^2, \tag{3.5}$$

where the first integral on the right-hand side is understood in the L_2-sense.

(iii) Assume the conditions of Theorem 4, and that $\int_a^b 1/h(x)^3\,dx = \infty$. Then, under H_0, there exists a probability space such that

$$\frac{1}{c_n}\left\{nR_n - \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{p(1-p)}{f(F^{-1}(p))^2}\,dp\right\} = \frac{1}{c_n}\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{B_n^2(p) - \mathbb{E}\{B_n^2(p)\}}{f(F^{-1}(p))^2}\,dp + o_p(1). \tag{3.6}$$

(iv) Assume the conditions of Theorem 5 for the right-hand tail, and that the left-hand tail either satisfies the conditions of Theorem 4, or $a = -\infty$ and $-xh(x) \to \alpha'$ as $x \to -\infty$, for some $\alpha' > \alpha$. Then, under H_0,

$$\frac{nR_n}{F^{-1}(\frac{n-1}{n})^2} \xrightarrow{d} \frac{1}{Y_1^{2/\alpha}} - \frac{2\alpha}{(\alpha-1)Y_1^{1/\alpha}} + \frac{\alpha}{\alpha-2} + \int_1^\infty \frac{1}{u^{2/\alpha}}\left\{\left(\frac{\tilde{S}(u)}{u}\right)^{-1/\alpha} - 1\right\}^2 du. \tag{3.7}$$

In each part of the Corollary 6, we may replace R_n with T_n, provided that we remove the second integral from the right-hand sides of (3.4) and (3.5). A direct consequence of the version of Corollary 6(iv) when adapted to T_n is that for the Pareto distribution function $F(x) = 1 - x^{-\alpha}$ for $x > 1$ and $\alpha > 2$, we have

$$\mu_2(F)n^{1-\frac{2}{\alpha}}T_n \xrightarrow{d} \frac{1}{Y_1^{2/\alpha}} - \frac{2\alpha}{(\alpha-1)Y_1^{1/\alpha}} + \frac{\alpha}{\alpha-2} + \int_1^\infty \frac{1}{u^{2/\alpha}}\left\{\left(\frac{\tilde{S}(u)}{u}\right)^{-1/\alpha} - 1\right\}^2 du,$$

where $\mu_2(F) = \alpha/(\alpha-2)$, which proves the conjecture made by Csörgő on p. 69 of del Barrio et al. (2000), as well as giving the form of the limiting distribution. We should, however, point out that tests based on part (iv) of Corollary 6 do not distinguish distributions which are different from F but have the same tails. Such tests would also suffer from a poor rate of convergence to the asymptotic null distribution by comparison with that experienced for tests based on parts (i), (ii) and (iii) of Corollary 6. Finally, we mention that it is equally straightforward to obtain critical values for tests based on Corollary 6(iii) as for the other cases, despite the limiting distribution having a less explicit form.

4 Concluding remarks

In certain circumstances, the tests in Section 3 may be extended to situations where other parameters are unknown. For example, let $G_\alpha(x) = 1 - \exp(-x^\alpha)$ for $x > 0$ denote the distribution function of the Weibull distribution with parameter α, and let $\mathcal{G}_\alpha(\cdot) = \{G_\alpha(\frac{\cdot}{\sigma}) : \sigma > 0\}$ denote the corresponding scale family. Suppose we are interested in testing the null hypothesis $H_0 : F \in \cup_{\{\epsilon\leq\alpha\leq 4/3\}}\mathcal{G}_\alpha$ for some $\epsilon > 0$. Suppose further that we are given an estimator $\hat{\alpha}$, such as the maximum likelihood estimator, which under H_0 satisfies $\hat{\alpha} = \alpha_0 + O_p(n^{-1/2})$ uniformly for $\alpha_0 \in [\epsilon, 4/3]$, where α_0 is the true value of α. If we estimate the scale by $\hat{\sigma}_{\hat{\alpha}}$, where

$$\hat{\sigma}_\alpha = \frac{1}{\mu_2(G_\alpha)}\int_0^1 \hat{F}_n^{-1}(p)G_\alpha^{-1}(p)\,dp,$$

then under H_0 it may be shown that $\hat{\sigma}_{\hat{\alpha}} = \hat{\sigma}_{\alpha_0} + O_p(n^{-1/2})$, uniformly for $\alpha_0 \in [\epsilon, 4/3]$. It follows that under H_0,

$$\begin{aligned} d_2^2(\mathcal{G}_{\hat{\alpha}}, \mathcal{G}_{\alpha_0}) &= \int_0^1 \left(\hat{\sigma}_{\hat{\alpha}} \log^{\frac{1}{\hat{\alpha}}} \frac{1}{1-p} - \hat{\sigma}_{\alpha_0} \log^{\frac{1}{\alpha_0}} \frac{1}{1-p}\right)^2 dp \\ &= \hat{\sigma}_{\hat{\alpha}}^2 \Gamma\Big(1 + \frac{2}{\hat{\alpha}}\Big) - 2\hat{\sigma}_{\hat{\alpha}}\hat{\sigma}_{\alpha_0}\Gamma\Big(1 + \frac{1}{\hat{\alpha}} + \frac{1}{\alpha_0}\Big) + \hat{\sigma}_{\alpha_0}^2\Gamma\Big(1 + \frac{2}{\alpha_0}\Big) = O_p(n^{-1}), \end{aligned}$$

uniformly for $\alpha_0 \in [\epsilon, 4/3]$. Thus $d_2^2(\hat{F}_n, \mathcal{G}_{\hat{\alpha}}) = d_2^2(\hat{F}_n, \mathcal{G}_{\alpha_0}) + O_p(n^{-1})$, uniformly in α_0. We conclude that the modified test statistic $\tilde{T}_n$ obtained by first estimating α_0 by $\hat{\alpha}$ and then substituting this expression into (3.2) has the same asymptotic null distribution as the test statistic T_n which would be used if α_0 were known.

Finally, we briefly mention a further application of the results of Section 2, to bootstrap performance. If $X_1, \ldots, X_n$ are independent and identically distributed with mean μ and finite variance σ^2 (which for simplicity we assume to be known), then for the purposes of constructing confidence intervals for μ it is natural to consider the root $n^{1/2}(\bar{X}_n - \mu)$. We denote the sampling distribution of this root by $H_n(F)$. Conditional on $X_1, \ldots, X_n$, let $H_n(\hat{F}_n)$ denote distribution of the bootstrap estimator of $H_n(F)$, so that $H_n(\hat{F}_n)$ is the distribution of $n^{1/2}(\bar{X}_n^* - \bar{X}_n)$, where $X_1^*, \ldots, X_n^*$ are conditionally independent from $\hat{F}_n$, and $\bar{X}_n^* = n^{-1} \sum X_i^*$. Then a calculation on p. 74 of Shao and Tu (1995) shows that

$$d_2\big(H_n(\hat{F}_n), H_n(F)\big) \leq d_2(\hat{F}_n, F), \tag{4.1}$$

so that the results of Section 2 give bounds on the rate of convergence of the bootstrap distribution. It would be interesting to know when both sides of (4.1) are of the same stochastic order.

A Appendix

Proof of Theorem 5.
In this proof, we assume $X_1 = F^{-1}(U_1), \ldots, X_n = F^{-1}(U_n)$, where $U_1, \ldots, U_n$ are independent uniform random variables on $(0, 1)$. We write $\hat{G}_n(x)$ for the empirical distribution function of $U_1, \ldots, U_n$ and observe that $\hat{F}_n^{-1}(p) = F^{-1}(\hat{G}_n^{-1}(p))$. In turn, by Lemma 3.0.1 of Csörgő and Horváth (1993), we may assume that the order statistics $U_{(1)}, \ldots, U_{(n)}$ are generated as $U_{(1)} = 1 - \frac{\tilde{S}(n)}{\tilde{S}(n+1)}, \ldots, U_{(n)} = 1 - \frac{\tilde{S}(1)}{\tilde{S}(n+1)}$, so that $\hat{G}_n^{-1}(p) = 1 - \frac{\tilde{S}(n(1-p))}{\tilde{S}(n+1)}$ almost everywhere. Let (a_n) be a sequence satisfying $a_n \leq n^{-1} \log n$ and $n a_n \to \infty$ as $n \to \infty$. We divide the integral in the left-hand side of (2.17) into three pieces, and treat each separately.

1. *The range* $1 - a_n \leq p < n/(n+1)$.

We write $F^{-1}(p) = (1-p)^{-1/\alpha} \ell\big(\frac{1}{1-p}\big)$ where ℓ is a slowly varying function. Then, following the proof of Theorem 6.4.5 of Csörgő and Horváth (1993), we have

$$\frac{n}{F^{-1}(\frac{n-1}{n})^2}\int_{1-a_n}^{\frac{n}{n+1}}\{\hat{F}_n^{-1}(p)-F^{-1}(p)\}^2\,dp$$
$$=\frac{n^{1-\frac{2}{\alpha}}}{\ell(n)^2}\int_{1-a_n}^{\frac{n}{n+1}}\left\{\left(\frac{1-p}{1-\hat{G}_n^{-1}(p)}\right)^{1/\alpha}\frac{\ell\left(\frac{1}{1-\hat{G}_n^{-1}(p)}\right)}{\ell(\frac{1}{1-p})}-1\right\}^2\frac{\ell(\frac{1}{1-p})^2}{(1-p)^{2/\alpha}}\,dp\equiv A_n,$$

say. We write $A_n=(A_n-B_n)+(B_n-C_n)+C_n$, with

$$B_n=\frac{n^{1-\frac{2}{\alpha}}}{\ell(n)^2}\int_{1-a_n}^{\frac{n}{n+1}}\left\{\left(\frac{1-p}{1-\hat{G}_n^{-1}(p)}\right)^{1/\alpha}-1\right\}^2\frac{\ell(\frac{1}{1-p})^2}{(1-p)^{2/\alpha}}\,dp$$

and

$$C_n=n^{1-\frac{2}{\alpha}}\int_{1-a_n}^{\frac{n}{n+1}}\left\{\left(\frac{1-p}{1-\hat{G}_n^{-1}(p)}\right)^{1/\alpha}-1\right\}^2\frac{1}{(1-p)^{2/\alpha}}\,dp.$$

Now $\frac{1-p}{1-\hat{G}_n^{-1}(p)}=O_p(1)$, so by Theorem 1.2.1 of Bingham, Goldie and Teugels (1987), we may write

$$\sup_{1-\frac{\log n}{n}\le p\le\frac{n}{n+1}}\left|\frac{\ell\left(\frac{1}{1-\hat{G}_n^{-1}(p)}\right)}{\ell(\frac{1}{1-p})}-1\right|=\epsilon_n,$$

where $\epsilon_n=o_p(1)$. Thus from the inequality $|x^2-y^2|\le(x-y)^2+2|x-y||y|$, we see that

$$|A_n-B_n|\le\epsilon_n^2\frac{n^{1-\frac{2}{\alpha}}}{\ell(n)^2}\int_{1-a_n}^{\frac{n}{n+1}}\{1-\hat{G}_n^{-1}(p)\}^{-2/\alpha}\ell(\tfrac{1}{1-p})^2\,dp$$
$$+2\epsilon_n\frac{n^{1-\frac{2}{\alpha}}}{\ell(n)^2}\int_{1-a_n}^{\frac{n}{n+1}}\{1-\hat{G}_n^{-1}(p)\}^{-1/\alpha}\left|\left(\frac{1-p}{1-\hat{G}_n^{-1}(p)}\right)^{1/\alpha}-1\right|\frac{\ell(\frac{1}{1-p})^2}{(1-p)^{1/\alpha}}\,dp \tag{A.1}$$

Now a weak form of Corollary A.3.1 of Csörgő and Horváth (1993) gives that for p in the range of the integral, $\ell(\frac{1}{1-p})^2/\ell(n)^2\le 1+\{n(1-p)\}^{1/\alpha}$ for sufficiently large n, so the first term in (A.1) is bounded above by

$$\epsilon_n^2 n^{1-\frac{2}{\alpha}}\int_{1-a_n}^{\frac{n}{n+1}}\left\{\frac{\tilde{S}(n(1-p))}{\tilde{S}(n+1)}\right\}^{-2/\alpha}\left[1+\{n(1-p)\}^{1/\alpha}\right]dp.$$

But $\tilde{S}(n+1)/n \stackrel{a.s.}{\to} 1$ by the strong law of large numbers, and a change of variable to $u = n(1-p)$ and the law of the iterated logarithm gives that the first term in (A.1) is $O_p\{\epsilon_n^2(na_n)^{1-\frac{1}{\alpha}}\}$.

For the second term in (A.1), we may apply the Cauchy-Schwarz inequality to see that it may be written as $O_p\{\epsilon_n(na_n)^{\frac{1}{2}-\frac{1}{2\alpha}}C_n^{1/2}\}$. But

$$\begin{aligned} C_n &= n^{1-\frac{2}{\alpha}} \int_{1-a_n}^{\frac{n}{n+1}} \left\{ \left(\frac{\tilde{S}(n(1-p))}{(1-p)\tilde{S}(n+1)} \right)^{-1/\alpha} - 1 \right\}^2 \frac{1}{(1-p)^{2/\alpha}}\, dp \\ &= \int_1^\infty u^{-2/\alpha} \left\{ \left(\frac{\tilde{S}(u)}{u} \right)^{-1/\alpha} - 1 \right\}^2 du + o_p(1) \end{aligned} \tag{A.2}$$

by a very similar argument to that used above (cf. also (6.4.18) and (6.4.19) in Csörgő and Horváth (1993)). Since the right-hand side of (A.2) is almost surely finite by the law of the iterated logarithm, it follows that $|A_n - B_n| = O_p\{\epsilon_n(na_n)^{\frac{1}{2}-\frac{1}{2\alpha}} + \epsilon_n^2(na_n)^{1-\frac{1}{\alpha}}\}$. Therefore, if we choose

$$a_n = \frac{1}{n} \min\{\log n, \epsilon_n^{-\alpha/(\alpha-1)}\},$$

then $A_n - B_n = o_p(1)$.

To show that $B_n - C_n = o_p(1)$, we use a stronger form of Corollary A.3.1 in Csörgő and Horváth (1993), namely that for p in the range of the integral and for any $\delta > 0$, we have

$$\left| \frac{\ell(\frac{1}{1-p})^2}{\ell(n)^2} - 1 \right| \le \delta\{n(1-p)\}^{1/\alpha}$$

for sufficiently large n. This fact, together with a virtually identical argument to that just given for the distributional limit of C_n, gives that $B_n - C_n = o_p(1)$. It therefore follows that

$$A_n = \int_1^\infty u^{-2/\alpha} \left\{ \left(\frac{\tilde{S}(u)}{u} \right)^{-1/\alpha} - 1 \right\}^2 du + o_p(1). \tag{A.3}$$

2. *The range* $n/(n+1) \le p < 1$.

We have

$$\frac{n}{F^{-1}(\frac{n-1}{n})^2}\int_{\frac{n}{n+1}}^{1}\{\hat{F}_n^{-1}(p)-F^{-1}(p)\}^2\,dp$$

$$=\frac{nF^{-1}\big(1-\frac{\tilde{S}(1)}{\tilde{S}(n+1)}\big)^2}{(n+1)F^{-1}(\frac{n-1}{n})^2}-\frac{2nF^{-1}\big(1-\frac{\tilde{S}(1)}{\tilde{S}(n+1)}\big)}{F^{-1}(\frac{n-1}{n})^2}\int_{\frac{n}{n+1}}^{1}F^{-1}(p)\,dp+\frac{n\int_{\frac{n}{n+1}}^{1}F^{-1}(p)^2\,dp}{F^{-1}(\frac{n-1}{n})^2}$$

$$=\left\{\frac{(\frac{\tilde{S}(1)}{\tilde{S}(n+1)})^{-2/\alpha}\ell(\frac{\tilde{S}(n+1)}{\tilde{S}(1)})^2}{n^{2/\alpha}\ell(n)^2}-\frac{2\alpha(\frac{\tilde{S}(1)}{\tilde{S}(n+1)})^{-1/\alpha}}{(\alpha-1)n^{1/\alpha}}+\frac{\alpha}{\alpha-2}\right\}\{1+o_p(1)\}$$

$$=\tilde{S}(1)^{-2/\alpha}-\frac{2\alpha}{\alpha-1}\tilde{S}(1)^{-1/\alpha}+\frac{\alpha}{\alpha-2}+o_p(1). \tag{A.4}$$

3. *The range* $1/2\le p<1-a_n$.

Finally, we show that

$$\frac{n}{F^{-1}(\frac{n-1}{n})^2}\int_{1/2}^{1-a_n}\{\hat{F}_n^{-1}(p)-F^{-1}(p)\}^2\,dp\xrightarrow{p}0. \tag{A.5}$$

By the Hungarian construction (2.4) with $\nu=1/2$,

$$\frac{1}{F^{-1}(\frac{n-1}{n})^2}\left|\int_{1/2}^{1-a_n}n\{\hat{F}_n^{-1}(p)-F^{-1}(p)\}^2\,dp-\int_{1/2}^{1-a_n}\frac{B_n^2(p)}{f(F^{-1}(p))^2}\,dp\right|$$

$$\le\frac{O_p(1)}{F^{-1}(\frac{n-1}{n})^2}\int_{1/2}^{1-a_n}\frac{1-p}{f(F^{-1}(p))^2}\,dp+\frac{O_p(1)}{F^{-1}(\frac{n-1}{n})^2}\int_{1/2}^{1-a_n}\frac{|B_n(p)|(1-p)^{1/2}}{f(F^{-1}(p))^2}\,dp. \tag{A.6}$$

To show that the left-hand side of (A.6) is $O_p(1)$, it is enough to prove that the non-random part of the first term on the right-hand side of (A.6) converges to zero, since $\mathbb{E}|B_n(p)|\le(1-p)^{1/2}$. But, recalling that $F^{-1}(p)=(1-p)^{1/\alpha}\ell(\frac{1}{1-p})$, by (**A7**) we may write $f(F^{-1}(p))=(1-p)^{1+\frac{1}{\alpha}}\tilde{\ell}(\frac{1}{1-p})$, where $\tilde{\ell}$ is another slowly varying function which satisfies $\ell(x)\tilde{\ell}(x) \to \alpha$ as $x \to \infty$. Thus

$$\frac{1}{F^{-1}(\frac{n-1}{n})^2}\int_{1/2}^{1-a_n}\frac{1-p}{f(F^{-1}(p))^2}\,dp = \frac{n^{-2/\alpha}}{\ell(n)^2}\int_{1/2}^{1-a_n}\frac{1}{(1-p)^{1+\frac{2}{\alpha}}\tilde{\ell}(\frac{1}{1-p})^2}\,dp$$

$$\sim \frac{\alpha}{2\ell(n)\tilde{\ell}(1/a_n)^2}\frac{(1/a_n)^{2/\alpha}}{n^{2/\alpha}}$$

$$\sim \frac{\ell(1/a_n)^2}{2\ell(n)^2}\frac{(1/a_n)^{2/\alpha}}{n^{2/\alpha}} \to 0$$

as $n \to \infty$. We deduce (A.5) on noticing that

$$\mathbb{E}\left(\int_{1/2}^{1-a_n}\frac{B_n^2(p)}{f(F^{-1}(p))^2}\,dp\right) \to 0$$

by the same argument as above.

4.*Conclusion of the proof.*

Adding together (A.3), (A.4) and (A.5), we find that

$$\frac{n}{F^{-1}(\frac{n-1}{n})^2}\int_{1/2}^{1}\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\,dp \xrightarrow{d} \frac{1}{Y_1^{2/\alpha}} - \frac{2\alpha}{(\alpha-1)Y_1^{1/\alpha}} + \frac{\alpha}{\alpha-2}$$

$$+\int_1^\infty \frac{1}{u^{2/\alpha}}\left\{\left(\frac{\tilde{S}(u)}{u}\right)^{-1/\alpha} - 1\right\}^2 du,$$

as required.

□

Proof of Corollary 6.
We begin by studying the limiting behaviour of R_n^*. For (i) and (ii), the Hungarian construction (2.4) allows a simultaneous approximation of the three terms in (3.3), on applying the elementary inequalities

$$\left[\int_B\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}\,dp\right]^2 \le \int_B\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\,dp, \tag{A.7}$$

$$\left[\int_B\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}F^{-1}(p)\,dp\right]^2 \le \int_B\{\hat{F}_n^{-1}(p) - F^{-1}(p)\}^2\,dp \tag{A.8}$$

to the set $B = \left[\frac{1}{n+1}, \frac{n}{n+1}\right]$. The argument from the proof of Theorem 1 deals with the regions $p \in (0, \frac{1}{n+1})$ and $p \in (\frac{n}{n+1}, 1)$, and the fact that the second and third integrals in (3.4) and (3.5) are finite almost surely gives that (3.4) and (3.5) hold when R_n is replaced with R_n^*. Similarly, for (iii), we have that (3.6) holds for R_n^*, with the condition that $\int_a^b 1/h(x)^3\,dx = \infty$ ensuring that the second and third terms on the right-hand side of (3.5) are asymptotically negligible when divided by c_n. Finally, for (iv),

$$\int_0^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}\,dp = \bar{X}_n = O_p\Big(\frac{1}{n^{1/2}}\Big) = o_p\Big(\frac{F^{-1}(\frac{n-1}{n})}{n^{1/2}}\Big),$$

and

$$2\int_0^1 \{\hat{F}_n^{-1}(p) - F^{-1}(p)\}F^{-1}(p)\,dp = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 1) - d_2^2(\hat{F}_n, F) = o_p\Big(\frac{F^{-1}(\frac{n-1}{n})}{n^{1/2}}\Big)$$

using Rosenthal's inequality and Theorem 5. Thus again, (3.7) holds when R_n is replaced with R_n^*.

To transfer the limit theorems to R_n, in the case of (i) we write $nR_n - nR_n^* = nR_n^*\big(\frac{1}{S_n^2} - 1\big) = o_p(1)$. The calculations for (ii), (iii) and (iv) are very similar, as are the computations which allow the results for T_n^* to be transfered to T_n.

□

Acknowledgement: The first author is very grateful for helpful conversations with Michail Loulakis.

References

Bickel, P. J. and Freedman, D. A. (1981), *Some asymptotic theory for the bootstrap*, Ann. Statist., **9**, 1196–1217.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987), *Regular Variation*, Cambridge Univ. Press.

Csörgő, M. and Horváth, L. (1990), *On the distributions of L_p norms of weighted quantile processes*, Ann. Inst. Henri Poincaré, **26**, 65–90.

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

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

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

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

Galambos, J. (1978), *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.

Kolmós, J., Major, P. and Tusnády, G. (1975), *An approximation of partial sums of independent RV's and the sample DF, Part I*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, **32**, 111–131.

Kolmós, J., Major, P. and Tusnády, G. (1976), *An approximation of partial sums of independent RV's and the sample DF. Part II*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, **34**, 33–58.

Major, P. (1978), *On the invariance principle for sums of independent, identically distributed random variables*, J. Mult. Anal., **8**, 487–501.

Rachev, S. T. (1984), *The Monge-Kantorovich problem on mass transfer and its applications in stochastics*, Theor. Probab. Appl., **29**, 647–676.

Shao, J. and Tu, D. (1995), *The Jackknife and Bootstrap*, Springer-Verlag, New York.

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

Sysoeva, F. A. (1965), *Generalization of a certain Hardy inequality*, Izv. Vysš. Učebn. Zaved. Matematika, **49**, 140–143.