Tatra Mt. Math. Publ. **26** (2003), 337–356

TATRA MOUNTAINS
Mathematical Publications

WEIGHTED CORRELATION TESTS FOR GAMMA AND LOGNORMAL FAMILIES

Sándor Csörgő — Tamás Szabó

ABSTRACT. A recent general theory of weighted correlation tests is exposed and is illustrated on gamma and lognormal scale families. The emphasis here is on obtaining computable asymptotic distributions under the null hypotheses, which in the gamma case requires solving an associated eigenvalue problem.

1. Introduction

A new approach to testing goodness of fit to location and scale families, based on the L_2-Wasserstein distance, was recently proposed by del Barrio, Cuesta-Albertos, Matrán and Rodríguez-Rodríguez [5] and del Barrio, Cuesta-Albertos and Matrán [6]. The limitations of the arising correlation tests in their demand of very light underlying tails were delineated by Csörgő [3], and a remedial involvement of a measurable weight function $w\colon (0,1) \to [0,\infty)$ was proposed by Csörgő [4] and, with a different motivation, by de Wet [7, 8].

Letting $Q_F(t) = F^{-1}(t) := \inf\{x \in \mathbb{R}\colon F(x) \geq t\}$, $0 < t < 1$, stand for the quantile function of a distribution function F on the real line $\mathbb{R}$ and $G(\cdot)$ be a known distribution function on $\mathbb{R}$ for which $0 < \mu(G,w) := \int_0^1 Q_G^2(t)\,w(t)\,dt = \int_{\mathbb{R}} x^2 w\bigl(G(x)\bigr)\,dG(x) < \infty$ and $G_\sigma(x) = G(x/\sigma)$, $x \in \mathbb{R}$, for an unknown scale parameter $\sigma > 0$, we briefly expose the idea for the associated scale family $\mathcal{G} = \{G_\sigma\colon \sigma > 0\}$. Consider any distribution function F for which $0 < \mu(F,w) < \infty$. Then, since $Q_{G_\sigma}(\cdot) = \sigma Q_G(\cdot)$, the weighted L_2-Wasserstein distance $\mathcal{W}_w(F,\mathcal{G})$

2000 Mathematics Subject Classification: Primary 62F05, 62E20; Secondary 60F05.
Keywords: goodness of fit to scale families, weighted correlation tests, asymptotic distributions, gamma and lognormal scale families, loss of degree of freedom.

Research supported in part by the Hungarian National Foundation for Scientific Research, Grants T–032025 and T–034121.

between F and the family $\mathcal{G}$, scaled to F, is given by

$$d_w^2(F,\mathcal{G}) := \frac{\mathcal{W}_w^2(F,\mathcal{G})}{\mu(F,w)} := \frac{1}{\mu(F,w)} \inf_{\sigma>0} \int_0^1 \left[Q_F(t) - Q_{G_\sigma}(t)\right]^2 w(t)\,dt$$

$$= 1 - \frac{\left[\int_0^1 Q_F(t)Q_G(t)w(t)\,dt\right]^2}{\mu(F,w)\mu(G,w)},$$

where the infimum is taken for $\sigma = \sigma_*(w) := \int_0^1 Q_F(t)Q_G(t)w(t)\,dt/\mu(G,w)$, so that $1 - d_w^2(F,\mathcal{G})$ is the correlation between the functions Q_F and Q_G in the measure space $\big((0,1),\mathcal{B}(0,1),\lambda_w\big)$, where $\mathcal{B}(0,1) = \mathcal{B} \cap (0,1)$ for the σ-algebra $\mathcal{B}$ of Borel sets in $\mathbb{R}$ and the measure $\lambda_w(D) = \int_D w(t)\,dt$, $D \in \mathcal{B}(0,1)$, is not necessarily finite.

Given a sample $X_1,\dots,X_n$ from F, with the pertaining order statistics $X_{1,n} \le X_{2,n} \le \dots \le X_{n,n}$, sample distribution function $F_n(x) = \sum_{k=1}^n I\{X_k \le x\}/n$, $x \in \mathbb{R}$, and sample quantile function $Q_n(\cdot) := Q_{F_n}(\cdot)$, for which $Q_n(t) = X_{k,n}$, $(k-1)/n < t \le k/n$, $k = 1,2,\dots,n$, the natural scale-free test statistic for the null hypothesis $H_0\colon F \in \mathcal{G}$ is, as explained in Csörgő [4], a cut version of $d_w^2(F_n,\mathcal{G})$ given by

$$S_n(w) := 1 - \frac{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_n(t)Q_G(t)w(t)\,dt\right]^2}{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_G^2(t)w(t)\,dt\right]\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_n^2(t)w(t)\,dt\right]}.$$

Put $-\infty \le a_G := \sup\{x\colon G(x) = 0\} < \inf\{x\colon G(x) = 1\} =: b_G \le \infty$ for the endpoints of the support of G. Then, letting $\{B(t)\colon 0 \le t \le 1\}$ stand for a Brownian bridge, a sample-continuous Gaussian process with mean zero and covariance $\mathbb{E}\big(B(u)B(v)\big) = \min(u,v) - uv$, $0 \le u,v \le 1$, and $\xrightarrow{\mathcal{D}}$ denote convergence in distribution, with all asymptotic relations understood as $n \to \infty$ unless otherwise specified, for the asymptotic distribution of this test statistic we have the following

THEOREM 1. *Let $w(\cdot)$ be a nonnegative measurable function on the interval $(0,1)$. Suppose that $F \in \mathcal{G} = \{G_\sigma\colon \sigma > 0\}$, where G is twice differentiable on the open interval (a_G,b_G), such that $0 < \mu(G,w) = \int_0^1 Q_G^2(t)w(t)\,dt < \infty$,*

$g(x) = G'(x) > 0$ for all $x \in (a_G, b_G)$ and

$$A(G) := \sup_{0<t<1} \frac{t(1-t)\left|g'\big(Q_G(t)\big)\right|}{g^2\big(Q_G(t)\big)} < \infty\,. \tag{1.1}$$

If

$$J(G,w) := \int_0^1\!\!\int_0^1 \frac{\big[\min(s,t) - st\big]^2}{g^2\big(Q_G(s)\big)g^2\big(Q_G(t)\big)}\, w(s)w(t)\,ds\,dt < \infty\,, \tag{1.2}$$

and

$$a_n(G,w,\delta) := \frac{1}{n^{1-\delta}} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\big[t(1-t)\big]^\delta}{g^2\big(Q_G(t)\big)}\, w(t)\,dt \to 0 \quad \text{for some } \delta \in \left(\tfrac12, 1\right), \tag{1.3}$$

then

$$\begin{aligned} nS_n(w) - c_n(w) \xrightarrow{\mathcal{D}} V(G,w) := {} & \frac{1}{\mu(G,w)} \int_0^1 \frac{B^2(t) - t(1-t)}{g^2\big(Q_G(t)\big)}\, w(t)\,dt \\ & - \left[\frac{1}{\mu(G,w)} \int_0^1 \frac{B(t)Q_G(t)}{g\big(Q_G(t)\big)}\, w(t)\,dt\right]^2, \end{aligned}$$

where, setting

$$c_n^\diamond(G,w) := \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{t(1-t)}{g^2\big(Q_G(t)\big)}\, w(t)\,dt \qquad \text{and} \qquad \mu_n(G,w) := \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_G^2(t)w(t)\,dt\,,$$

$$c_n(w) = \frac{c_n^\diamond(G,w)}{\mu_n(G,w)}\,.$$

This result is a simplified weaker version, sufficient for the applications in the present paper, of the second part of Theorem 1 in Csörgő [4], in which condition (1.3) unites two finer sufficient conditions, each of which is the first possibility of two alternative conditions. The integrals in the limiting random variable $V(G,w)$ make sense in the space $L_2(\Omega, \mathcal{A}, \mathbb{P})$ over the probability space where $B(\cdot)$ is defined.

The appealing idea of de Wet's [7, 8] for the choice of w is to make the underlying linear estimator $\int_0^1 Q_n(t)Q_G(t)w(t)\,dt/\mu(G,w)$, the empirical counterpart of $\sigma_*(w)$, asymptotically efficient and thereby achieve an asymptotic loss of degree of freedom, due to the estimation of the parameter σ; the latter property will be elucidated in the examples below. For this de Wet [7, 8] substitutes the classical theory due to Chernoff, Gastwirth and Johns [2] for the asymptotic attainability of the Cramér–Rao bound. Thus, assuming that G is three times differentiable in (a_G, b_G) and $x^2 g'(x) \to 0$ both as $x \downarrow a_G$ and $x \uparrow b_G$, whenever possible, de Wet's choice of $w(\cdot) = w_G(\cdot)$ is

$$w_G(t) = \frac{1}{I_G} \frac{L'_G\big(Q_G(t)\big)}{Q_G(t)}, \quad 0 < t < 1, \quad \text{where} \quad L_G(x) = -1 - \frac{x g'(x)}{g(x)} \tag{1.4}$$

for all $x \in (a_G, b_G)$, and where $I_G = \int_{-\infty}^{\infty} L_G^2(x) g(x)\, dx = \int_0^1 L'_G\big(Q_G(t)\big) Q_G(t)\, dt$, so that $\mu(G, w_G) = \int_0^1 Q_G^2(t) w_G(t)\, dt = 1$.

Csörgő [4] brings two important examples for the application of Theorem 1. One of these is the class of Weibull$_\alpha$ scale families, where the choice of a de Wet weight function w_α is possible and works for each shape parameter $\alpha > 0$. The other is the class of Pareto$_\alpha$ scale families, where reasonable weight functions still produce asymptotic tests despite the fact that no de Wet choice is possible for any shape parameter $\alpha > 0$. The aim of the present paper is to further expose weighted correlation tests by working out two more examples that are of interest on their own right: for gamma and lognormal scale families. The initial challenge in both is the lack of a closed analytic formula for the quantile function. Furthermore, the emphasis here is on obtaining computable asymptotic distributions, which turns out rather surprisingly for the gamma and well for both families. An extra theoretical feature of a lognormal family is that, its tail being between all Weibull and all Pareto tails, it roughly represents a lower boundary for the feasibility of the choice in (1.4). The implementation of the test in the gamma case is described before the proof of Theorem 2, and the lognormal case is reduced to the use of the same table.

2. Gamma families

Consider the Gamma$_\alpha$ scale family $\mathcal{G}_\alpha = \big\{G_{\alpha,\sigma}(x) = G_\alpha(x/\sigma),\ x \in \mathbb{R} : \sigma > 0\big\}$ of distribution functions, where $G_\alpha(x) = \int_{-\infty}^{x} g_\alpha(y)\, dy$, $x \in \mathbb{R}$, given by the gamma density function

$$g_\alpha(x) = \begin{cases} \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} & \text{if } x > 0, \\ 0 & \text{if } x < 0, \end{cases}$$

of order $\alpha > 0$, where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$ is the usual gamma function. Let $Q_\alpha(t) = G_\alpha^{-1}(t)$ be the associated quantile function, uniquely given by the equation

$$G_\alpha\big(Q_\alpha(t)\big) = \frac{1}{\Gamma(\alpha)} \int_0^{Q_\alpha(t)} y^{\alpha-1} e^{-y}\, dy = t\,, \qquad 0 \le t < 1\,,$$

so that $Q_\alpha(0) = 0$ and $Q_\alpha(1) = \lim_{t \uparrow 1} Q_\alpha(t) = \infty$. Of course, the distribution function $G_\alpha(\cdot)$ is infinitely many times differentiable on the open support $(0, \infty)$ and since

$$g'_\alpha(x) = \frac{(\alpha-1)x^{\alpha-2}e^{-x} - x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}\,, \qquad x > 0\,,$$

we see that $x^2 g'_\alpha(x) \to 0$ both as $x \downarrow 0$ and $x \to \infty$. So, the conditions for the 'optimal' weight function $w_\alpha(\cdot) := w_{G_\alpha}(\cdot)$ in (1.4) are satisfied, resulting in a loss of degree of freedom in the asymptotic distribution of the weighted correlation test statistic $S_n(\alpha) := S_n(w_\alpha)$ under the null hypothesis $H_0\colon F \in \mathcal{G}_\alpha$ for any known shape parameter $\alpha > 0$, provided the conditions of Theorem 1 are also satisfied to allow us to get to the limiting world. We have

$$L_\alpha(x) = -1 - \frac{x g'_\alpha(x)}{g_\alpha(x)} = x - \alpha\,, \qquad \text{so that } L'_\alpha(x) = 1\,,\ x > 0\,,$$

and the Fisher information per observation from G_α is (cf. Chernoff *et al.* [2])

$$I_\alpha = \int_0^1 L'_\alpha\big(Q_\alpha(t)\big) Q_\alpha(t)\, dt = \int_0^\infty x g_\alpha(x)\, dx = \frac{\int_0^\infty x^\alpha e^{-x}\, dx}{\Gamma(\alpha)} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\,.$$

Hence de Wet's weight function in (1.4) for the present situation is

$$w_\alpha(t) = \frac{1}{\alpha Q_\alpha(t)}\,, \qquad 0 < t < 1\,.$$

Then the corresponding weighted correlation statistic is

$$S_n(\alpha) = S_n(w_\alpha) = 1 - \frac{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_n(t) Q_\alpha(t) w_\alpha(t)\, dt\right]^2}{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_\alpha^2(t) w_\alpha(t)\, dt\right] \left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_n^2(t) w_\alpha(t)\, dt\right]}$$

$$= 1 - \frac{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_n(t)\,dt\right]^2}{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_\alpha(t)\,dt\right]\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{Q_n^2(t)}{Q_\alpha(t)}dt\right]}$$

$$= 1 - \frac{\left[\frac{1}{n(n+1)}X_{1,n} + \frac{1}{n}\sum_{k=2}^{n-1} X_{k,n} + \frac{1}{n(n+1)}X_{n,n}\right]^2}{\mu_n(\alpha)\left[d_{1,n}^*(\alpha)X_{1,n}^2 + \sum_{k=2}^{n-1} d_{k,n}(\alpha)X_{k,n}^2 + d_{n,n}^*(\alpha)X_{n,n}^2\right]},$$

where

$$\mu_n(\alpha) = \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_\alpha(t)\,dt = \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} x g_\alpha(x)\,dx = \alpha \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} g_{\alpha+1}(x)\,dx$$

$$= \alpha\left[G_{\alpha+1}\left(Q_\alpha\left(\frac{n}{n+1}\right)\right) - G_{\alpha+1}\left(Q_\alpha\left(\frac{1}{n+1}\right)\right)\right],$$

$$d_{1,n}^*(\alpha) = \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{1}{n})} \frac{g_\alpha(x)}{x}\,dx\,, \qquad d_{n,n}^*(\alpha) = \int_{Q_\alpha(\frac{n-1}{n})}^{Q_\alpha(\frac{n}{n+1})} \frac{g_\alpha(x)}{x}\,dx\,,$$

$$d_{k,n}(\alpha) = \int_{\frac{k-1}{n}}^{\frac{k}{n}} \frac{1}{Q_\alpha(t)}\,dt = \int_{Q_\alpha(\frac{k-1}{n})}^{Q_\alpha(\frac{k}{n})} \frac{g_\alpha(x)}{x}\,dx\,, \qquad k = 2, \dots, n-1\,.$$

In general, the need to turn to the computer for the computation of these weights and $\mu_n(\alpha)$ is unavoidable. This is true even for the important case $\alpha = 1$, testing for exponentiality, when $Q_1(t) = \log 1/(1-t)$, $0 \le t < 1$. The result concerning the asymptotic null distribution for testing for a gamma scale family is

Theorem 2. *If $F \in \mathcal{G}_\alpha$ for some $\alpha > 0$, then*

$$n\alpha S_n(\alpha) - c_n^*(\alpha) \xrightarrow{\mathcal{D}} V\,,$$

where

$$c_n^*(\alpha) = \alpha \frac{\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{t(1-t)}{g_\alpha^2\left(Q_\alpha(t)\right)Q_\alpha(t)}\,dt}{\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} Q_\alpha(t)\,dt} = \alpha \frac{\int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} \frac{G_\alpha(x)[1-G_\alpha(x)]}{g_\alpha(x)x}\,dx}{\int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} x g_\alpha(x)\,dx}$$

$$= \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} \frac{G_\alpha(x)\big[1-G_\alpha(x)\big]}{g_\alpha(x)x}\,dx + o(1) = \log\log(n+1) + \gamma + o(1)$$

with $\gamma = -\Gamma'(1) = 0.577215664\ldots$ *standing for Euler's constant and*

$$V = \int_0^1 \frac{B^2(t)-t(1-t)}{g_\alpha^2(Q_\alpha(t))Q_\alpha(t)}\,dt - \left[\frac{1}{\sqrt{\alpha}}\int_0^1 \frac{B(t)}{g_\alpha(Q_\alpha(t))}\,dt\right]^2$$

$$\stackrel{\mathcal{D}}{=} \sum_{k=0}^{\infty} \frac{Z_{k+1}^2-1}{k+1} - Z_1^2 = -1 + \sum_{m=2}^{\infty} \frac{Z_m^2-1}{m}\,,$$

where $Z_1, Z_2, \ldots$ *are independent standard normal random variables.*

Although it is more precise to compute the centering constants $c_n^*(\alpha)$ also on a computer for moderate sample sizes n, it is remarkable that their asymptotic behavior depends on α only in the negligible third-order term $o(1)$. The distributional convergence to V is obtained by an application of Theorem 1, and we mentioned the inherent difficulty in this application. However, the main result here is the distributional equality $V \stackrel{\mathcal{D}}{=} -1 + \sum_{m=2}^{\infty}(Z_m^2-1)/m$. While the cancellation of the leading random term in the expansion of the first term of V by the second term was to be expected, this is what is called loss of degree of freedom, due to the (asymptotically efficient) estimation of the unknown scale parameter, the happy fact that the weights $1/m$ and hence the distribution of V do not depend on α is hardly believable at first sight. In this regard the only promising feature at this point is that the mean-zero normal random variable

$$Z(\alpha) = \frac{1}{\sqrt{\alpha}}\int_0^1 \frac{B(t)}{g_\alpha(Q_\alpha(t))}\,dt\,,$$

the square of which is subtracted, has variance

$$E\big(Z^2(\alpha)\big) = \frac{1}{\alpha}\int_0^1\int_0^1 \frac{\min(s,t)-st}{g_\alpha(Q_\alpha(s))g_\alpha(Q_\alpha(t))}\,ds\,dt$$

$$= \frac{1}{\alpha}\int_0^\infty\int_0^\infty \big[\min\big(G_\alpha(x),G_\alpha(y)\big) - G_\alpha(x)G_\alpha(y)\big]\,dx\,dy$$

$$= \frac{1}{\alpha}\left\{\int_0^\infty y^2 g_\alpha(y)\,dy - \left[\int_0^\infty y g_\alpha(y)\,dy\right]^2\right\} = \frac{1}{\alpha}\left\{\alpha^2+\alpha-\alpha^2\right\} = 1\,,$$

so that $Z(\alpha)$ does not depend on α, it is standard normal.

It is not the first time that a phenomenon like this occurs. Thirty years ago de Wet and Venter [9] developed a test statistic for general scale families, which is a ratio of two asymptotically efficient estimators of the scale parameter. In particular, for all gamma families the asymptotic distribution of their test statistic under H_0 is that of $V+1$, so they are asymptotically equivalent to $n\alpha S_n(\alpha) - c_n^*(\alpha) + 1$. Since the two techniques are different, it appears to be a hard problem to try and prove this directly. One of our aims being the exposition of correlation statistics based on a minimized weighted Wasserstein distance, this would be pointless anyway.

Fortunately for us, de Wet and Venter [9] tabulated the distribution function $\Psi(y) = H(y-1)$, where $H(x) = \mathbb{P}\{V \le x\} = \Psi(x+1)$, $y, x \in \mathbb{R}$. We only have to shift the argument to obtain the values of H between $H(-4) = \Psi(-3) \approx 0$ and $H(5) = \Psi(6) \approx 1$ in their Table 1, for example to find approximate p-values of our test based on $n\alpha S_n(\alpha) - c_n^*(\alpha)$. Specifically, shifting their Table 2, for the most frequently used percentiles of H we obtain $H^{-1}(0.8) = -0.8186$, $H^{-1}(0.85) = 0.0943$, $H^{-1}(0.9) = 0.4725$, $H^{-1}(0.95) = 1.0950$, and $H^{-1}(0.99) = 2.4960$. For example, for large n, we reject the null hypothesis $F \in \mathcal{G}_\alpha$ at approximate level 0.05 if $n\alpha S_n(\alpha) - c_n^*(\alpha) > 1.095$. Simulation studies would be needed to determine how large n should be for reliable use of these asymptotic critical points.

The special case $\alpha = 1$ of Theorem 3 in Csörgő [4] for testing a Weibull$_\alpha$ scale family and the special case $\alpha = 1$ of Theorem 2 above are identical, as they should be, for testing exponentiality.

Proof. Multiplying over by α, convergence in distribution to V follows from Theorem 1 if we show the three conditions that

$$A(\alpha) = \sup_{x>0} A_\alpha(x) < \infty\,, \tag{2.1}$$

$$J(\alpha) := J(G_\alpha, w_\alpha) = \frac{1}{\alpha^2} \int_0^1 \int_0^1 \frac{\big[\min(s,t) - st\big]^2}{g_\alpha^2(Q_\alpha(s)) g_\alpha^2(Q_\alpha(t)) Q_\alpha(s) Q_\alpha(t)}\, ds\, dt < \infty \tag{2.2}$$

and

$$a_n(\alpha, \delta) := a_n(G_\alpha, w_\alpha, \delta) = \frac{1}{\alpha\, n^{1-\delta}} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\big[t(1-t)\big]^\delta}{g_\alpha^2(Q_\alpha(t)) Q_\alpha(t)}\, dt \to 0 \tag{2.3}$$

for some $\delta \in (1/2, 1)$, where in (2.1), using that $g'_\alpha(x)/g^2_\alpha(x) = \Gamma(\alpha)e^x[\alpha - 1 - x]/x^\alpha$,

$$A_\alpha(x) = \frac{1}{\Gamma(\alpha)}\left[\int_0^x y^{\alpha-1}e^{-y}\,dy\right]\left[\int_x^\infty y^{\alpha-1}e^{-y}\,dy\right]\frac{e^x}{x^\alpha}\,|\alpha - 1 - x|\,, \qquad x > 0\,.$$

Since the asymptotic equalities (with the ratio of the two sides going to 1)

$$\int_0^x y^{\alpha-1}e^{-y}\,dy = \Gamma(\alpha)G_\alpha(x) \sim \frac{x^\alpha}{\alpha} \qquad \text{as } x \downarrow 0 \tag{2.4}$$

and

$$\int_x^\infty y^{\alpha-1}e^{-y}\,dy = \Gamma(\alpha)\big[1 - G_\alpha(x)\big] \sim x^{\alpha-1}e^{-x} \qquad \text{as } x \to \infty \tag{2.5}$$

hold, we have $\lim_{x\downarrow 0} A_\alpha(x) = |\alpha - 1|/\alpha$ and $\lim_{x\to\infty} A_\alpha(x) = 1$, which imply (2.1). To prove (2.2) we use the substitutions $Q_\alpha(t) = x$ and $Q_\alpha(s) = y$ to get

$$\begin{aligned}
\alpha^2 J(\alpha) &= \int_0^\infty\int_0^x \frac{G^2_\alpha(y)\big[1 - G_\alpha(x)\big]^2}{g_\alpha(y)g_\alpha(x)yx}\,dy\,dx + \int_0^\infty\int_x^\infty \frac{G^2_\alpha(x)\big[1 - G_\alpha(y)\big]^2}{g_\alpha(y)g_\alpha(x)yx}\,dy\,dx \\
&= 2\int_0^\infty\int_0^x \frac{G^2_\alpha(y)\big[1 - G_\alpha(x)\big]^2}{g_\alpha(y)g_\alpha(x)yx}\,dy\,dx \\
&= \frac{2}{\Gamma^2(\alpha)}\int_0^\infty \frac{\left[\int_x^\infty t^{\alpha-1}e^{-t}\,dt\right]^2}{x^\alpha e^{-x}}\left\{\int_0^x \frac{\left[\int_0^y t^{\alpha-1}e^{-t}\,dt\right]^2}{y^\alpha e^{-y}}\,dy\right\}dx\,.
\end{aligned}$$

Clearly the interesting cases are when x is small and when x and y are both large. If $x < 1$, then, as when obtaining the upper bound for (2.4),

$$\frac{2}{\Gamma^2(\alpha)}\int_0^1 \frac{\left[\int_x^\infty t^{\alpha-1}e^{-t}\,dt\right]^2}{x^\alpha e^{-x}}\left\{\int_0^x \frac{\left[\int_0^y t^{\alpha-1}e^{-t}\,dt\right]^2}{y^\alpha e^{-y}}\,dy\right\}dx$$

$$\leq \frac{2}{\Gamma^2(\alpha)}\int_0^1 \frac{\Gamma^2(\alpha)}{x^\alpha e^{-x}}\left[\int_0^x \frac{y^{2\alpha}}{\alpha^2 y^\alpha e^{-y}}\,dy\right]dx \leq \int_0^1 \frac{2e}{x^\alpha}\,\frac{e\,x^{\alpha+1}}{\alpha^2(\alpha+1)}\,dx \leq \frac{e^2}{\alpha^2(\alpha+1)}\,.$$

If $x, y > K$ for some $K \geq 1$ large enough, then by (2.5) and the analogous asymptotic equality $\int_K^x y^{-\alpha} e^y \, dy \sim x^{-\alpha} e^x$ as $x \to \infty$,

$$\frac{2}{\Gamma^2(\alpha)} \int_K^\infty \frac{\left[\int_x^\infty t^{\alpha-1} e^{-t}\, dt\right]^2}{x^\alpha e^{-x}} \left\{ \int_K^x \frac{\left[\int_0^y t^{\alpha-1} e^{-t}\, dt\right]^2}{y^\alpha e^{-y}}\, dy \right\} dx$$

$$\leq \frac{2}{\Gamma^2(\alpha)} \int_K^\infty 4x^{\alpha-2} e^{-x} \left[\int_K^x \frac{\Gamma^2(\alpha)}{y^\alpha e^{-y}}\, dy \right] dx \leq 8 \int_K^\infty x^{\alpha-2} e^{-x} \frac{2e^x}{x^\alpha}\, dx = \frac{16}{K}.$$

The remaining easier cases can be handled similarly, and hence (2.2) is also satisfied.

Next we show (2.3). We only have to look at the integral near 0 and 1. Let $\varepsilon \in (0,1)$ be as small as needed for the inequalities below. Then, on the one hand, using (2.4) we see that for all n large enough,

$$a_{n,\alpha}^{\delta,\varepsilon} := \frac{1}{n^{1-\delta}} \int_{\frac{1}{n+1}}^{\varepsilon} \frac{\big[t(1-t)\big]^\delta}{g^2\big(Q_\alpha(t)\big) Q_\alpha(t)}\, dt = \frac{\Gamma(\alpha)}{n^{1-\delta}} \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\varepsilon)} \frac{G_\alpha^\delta(x)\big[1-G_\alpha(x)\big]^\delta}{x^{\alpha-1} e^{-x} x}\, dx$$

$$\leq \frac{2e^{Q_\alpha(\varepsilon)}\Gamma(\alpha)}{\alpha^\delta \big[\Gamma(\alpha)\big]^\delta n^{1-\delta}} \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\varepsilon)} \frac{1}{x^{\alpha-\alpha\delta}}\, dx$$

and, since $Q_\alpha(t) \sim \big(\alpha\Gamma(\alpha)t\big)^{1/\alpha}$ as $t \downarrow 0$, also that the bound is $O(1/n^{1-\delta})$, $O\big((\log n)/n^{1-\delta}\big)$, or $O(1/n^{1/\alpha})$ according as $\alpha(1-\delta)$ is less than, equal to, or greater than 1, so $a_{n,\alpha}^{\delta,\varepsilon} \to 0$ for any $\delta \in (0,1)$. On the other hand, using (2.5) and then a suitable corresponding statement for the arising integral, we have

$$b_{n,\alpha}^{\delta,\varepsilon} := \frac{1}{n^{1-\delta}} \int_{1-\varepsilon}^{\frac{n}{n+1}} \frac{\big[t(1-t)\big]^\delta}{g^2\big(Q_\alpha(t)\big) Q_\alpha(t)}\, dt = \frac{\Gamma(\alpha)}{n^{1-\delta}} \int_{Q_\alpha(1-\varepsilon)}^{Q_\alpha(\frac{n}{n+1})} \frac{G_\alpha^\delta(x)\big[1-G_\alpha(x)\big]^\delta}{x^{\alpha-1} e^{-x} x}\, dx$$

$$\leq \frac{2^\delta \Gamma(\alpha)}{\big[\Gamma(\alpha)\big]^\delta n^{1-\delta}} \int_{Q_\alpha(1-\varepsilon)}^{Q_\alpha(\frac{n}{n+1})} x^{(\delta-1)(\alpha-1)-1} e^{-(\delta-1)x}\, dx$$

$$\leq \frac{2^{1+\delta}\big[\Gamma(\alpha)\big]^{1-\delta}}{(1-\delta)n^{1-\delta}} \frac{\Big[Q_\alpha^{\alpha-1}\big(\frac{n}{n+1}\big) e^{-Q_\alpha(\frac{n}{n+1})}\Big]^{\delta-1}}{Q_\alpha\big(\frac{n}{n+1}\big)} \leq \frac{2^{2+\delta}}{(1-\delta)n^{1-\delta}} \frac{\big[1-\frac{n}{n+1}\big]^{\delta-1}}{Q_\alpha\big(\frac{n}{n+1}\big)}$$

for all n large enough since $\Gamma(\alpha)(1-t) \sim Q_\alpha^{\alpha-1}(t)e^{-Q_\alpha(t)}$ as $t \uparrow 1$, and this bound is asymptotically equal to $2^{2+\delta}/\big[(1-\delta)Q_\alpha\big(\frac{n}{n+1}\big)\big]$ and hence goes to 0. Thus $b_{n,\alpha}^{\delta,\varepsilon} \to 0$ for any $\delta \in (0,1)$, as well, which proves (2.3).

Now we sketch the proof of the last asymptotic equation for $c_n^*(\alpha)$ in the theorem. Since $\int_x^\infty y^{\alpha-1}e^{-y}\,dy = x^{\alpha-1}e^{-x} + (\alpha-1)\int_x^\infty y^{\alpha-2}e^{-y}\,dy$, it must be done for

$$\begin{aligned}
c_n^\diamond(\alpha) &:= \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} \frac{G_\alpha(x)\big[1-G_\alpha(x)\big]}{g_\alpha(x)x}\,dx = \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} G_\alpha(x)\,\frac{\int_x^\infty y^{\alpha-1}e^{-y}\,dy}{x^\alpha e^{-x}}\,dx \\
&= \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} \frac{G_\alpha(x)}{x}\,dx + \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} G_\alpha(x)\,\frac{(\alpha-1)\int_x^\infty y^{\alpha-2}e^{-y}\,dy}{x^\alpha e^{-x}}\,dx \\
&=: c_n^{(1)}(\alpha) + c_n^{(2)}(\alpha)\,.
\end{aligned}$$

Integrating by parts again, using the asymptotics for $Q_\alpha(\cdot)$ and a well-known expansion (see [10], for example) of $\Gamma'(\alpha)/\Gamma(\alpha)$, the first term is

$$\begin{aligned}
c_n^{(1)}(\alpha) &= \big[G_\alpha(x)\log x\big]_{x=Q_\alpha(\frac{1}{n+1})}^{x=Q_\alpha(\frac{n}{n+1})} - \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} g_\alpha(x)\log x\,dx \\
&= \frac{n}{n+1}\log Q_\alpha\left(\frac{n}{n+1}\right) - \frac{1}{n+1}\log Q_\alpha\left(\frac{1}{n+1}\right) - \int_{Q_\alpha(\frac{1}{n+1})}^{Q_\alpha(\frac{n}{n+1})} \frac{x^{\alpha-1}\log x}{\Gamma(\alpha)e^x}\,dx \\
&= \log\log(n+1) - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + o(1) \\
&= \log\log(n+1) + \left[\gamma + \frac{1}{\alpha} - \sum_{k=1}^\infty\left(\frac{1}{k} - \frac{1}{\alpha+k}\right)\right] + o(1)
\end{aligned}$$

and, by the obvious expansion of the incomplete gamma function ([10], p. 135)

$$G_\alpha(x) = \frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)} + \frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)}\sum_{k=1}^\infty \frac{x^k}{(\alpha+1)(\alpha+2)\cdots(\alpha+k)}\,, \qquad x \ge 0\,,$$

integrating term-wise and using Fubini's theorem, the second term becomes

$$c_n^{(2)}(\alpha) = \frac{\alpha-1}{\Gamma(\alpha+1)}\int_0^\infty \left[1 + \sum_{k=1}^\infty \frac{x^k}{(\alpha+1)\cdots(\alpha+k)}\right]\left[\int_x^\infty y^{\alpha-2}e^{-y}\,dy\right]dx + o(1)$$

$$= \frac{\alpha-1}{\Gamma(\alpha+1)} \int_0^\infty y^{\alpha-2} e^{-y} \left[\int_0^y dx \right] dy$$

$$+ \frac{\alpha-1}{\Gamma(\alpha+1)} \sum_{k=1}^\infty \int_0^\infty \frac{y^{\alpha-2} e^{-y}}{(\alpha+1)\cdots(\alpha+k)} \left[\int_0^y x^k \, dx \right] dy + o(1)$$

$$= \frac{\alpha-1}{\Gamma(\alpha+1)} \Gamma(\alpha) + \frac{\alpha-1}{\Gamma(\alpha+1)} \sum_{k=1}^\infty \frac{\Gamma(\alpha+k)}{(k+1)(\alpha+1)\cdots(\alpha+k)} + o(1)$$

$$= -\frac{1}{\alpha} + 1 + \sum_{k=1}^\infty \frac{\alpha-1}{(k+1)(\alpha+k)} + o(1)$$

$$= -\frac{1}{\alpha} + \sum_{k=1}^\infty \left[\frac{1}{k} - \frac{1}{k+1} \right] + \sum_{k=1}^\infty \left[\frac{1}{k+1} - \frac{1}{\alpha+k} \right] + o(1),$$

so that $c_n^{(2)}(\alpha) = -\frac{1}{\alpha} + \sum_{k=1}^\infty \left[\frac{1}{k} - \frac{1}{\alpha+k} \right] + o(1)$. Therefore, $c_n^\diamond(\alpha) = c_n^{(1)}(\alpha) + c_n^{(2)}(\alpha) = \log\log(n+1) + \gamma + o(1)$, indeed, as claimed.

Finally we turn to the distributional equation for V. Condition (2.2) so far ensured that the random integrals in V exist as limits in $L_2(\Omega, \mathcal{A}, \mathbb{P})$. It is also the necessary and sufficient condition ([11], Sections 66, 97, 98) that the Hilbert-Schmidt covariance operator $T\colon L_2(0,1) \to L_2(0,1)$ associated with the stochastic process

$$Z_\alpha(t) = \frac{B(t)}{g_\alpha\big(Q_\alpha(t)\big)\sqrt{Q_\alpha(t)}}, \qquad 0 < t < 1,$$

given by $Th(t) = \int_0^1 \operatorname{Cov}\big(Z_\alpha(s), Z_\alpha(t)\big) h(s)\, ds$, that is

$$Th(t) = \int_0^1 \frac{\min(s,t) - st}{g_\alpha\big(Q_\alpha(s)\big) g_\alpha\big(Q_\alpha(t)\big) \sqrt{Q_\alpha(s)}\sqrt{Q_\alpha(t)}} h(s)\, ds, \qquad 0 < t < 1,$$

has a countable number of different eigenvalues λ_k with pertaining eigenfunctions h_k: $Th_k(t) = \lambda_k h_k(t)$, $0 < t < 1$; here k runs through a countable set to be determined later on. Our first goal is the joint determination of λ_k and $h_k(\cdot)$.

Since we know from de Wet's [8] Theorem 3.1 that $\sqrt{Q_\alpha(\cdot)}$ is an eigenfunction (this is what motivated the choice of $w_\alpha(\cdot)$ in the first place, eventually leading to the loss of degree of freedom property), it is reasonable to search for the eigenfunctions in the form of $h_k(\cdot) = f_k\big(Q_\alpha(\cdot)\big)\sqrt{Q_\alpha(\cdot)}$, where, for each k, $f_k(\cdot)$

is twice continuously differentiable on $(0,\infty)$ such that beyond $h_k \in L_2(0,1)$,

$$\lim_{x\downarrow 0} G_\alpha(x) f_k(x) = 0 \qquad \text{and} \qquad \lim_{x\to\infty} \big[1 - G_\alpha(x)\big] f_k(x) = 0\,. \tag{2.6}$$

With this key observation our integral equation to solve is

$$\begin{aligned}(1-t)\int_0^t \frac{s f_k(Q_\alpha(s))}{g_\alpha(Q_\alpha(s))}\,ds + t\int_t^1 \frac{(1-s) f_k(Q_\alpha(s))}{g_\alpha(Q_\alpha(s))}\,ds \\ = \lambda_k f_k(Q_\alpha(t))\sqrt{Q_\alpha(t)}\, g_\alpha(Q_\alpha(t))\sqrt{Q_\alpha(t)} \\ = \lambda_k f_k(Q_\alpha(t))\, g_\alpha(Q_\alpha(t))\, Q_\alpha(t)\,, \qquad 0 < t < 1\,,\end{aligned}$$

which is nothing but

$$\big[1 - G_\alpha(x)\big]\int_0^x G_\alpha(y) f_k(y)\,dy + G_\alpha(x)\int_x^\infty \big[1 - G_\alpha(y)\big] f_k(y)\,dy = \lambda_k f_k(x) g_\alpha(x) x$$

for all $x > 0$. Differentiating and then dividing by $g_\alpha(\cdot)$, we obtain

$$\begin{aligned}\int_x^\infty \big[1 - G_\alpha(y)\big] f_k(y)\,dy - \int_0^x G_\alpha(y) f_k(y)\,dy \\ = \lambda_k \left[f_k'(x) x + f_k(x) \left\{ \frac{g_\alpha'(x)}{g_\alpha(x)}\, x + 1 \right\} \right] \\ = \lambda_k \big[f_k'(x) x + f_k(x)(\alpha - x)\big]\,, \qquad x > 0\,.\end{aligned}$$

Differentiating once more,

$$\begin{aligned}-f_k(x) &= -\big[1 - G_\alpha(x)\big] f_k(x) - G_\alpha(x) f_k(x) \\ &= \lambda_k \big[f_k''(x) x + f_k'(x) + f_k'(x)(\alpha - x) - f_k(x)\big] \\ &= \lambda_k \big[f_k''(x) x + f_k'(x)(\alpha + 1 - x) - f_k(x)\big]\,,\end{aligned}$$

from which

$$f_k''(x) x + f_k'(x)(\alpha + 1 - x) + \left(\frac{1}{\lambda_k} - 1\right) f_k(x) = 0\,, \qquad x > 0\,.$$

One can also work backward from here and see that under (2.6) this differential equation is equivalent to $T f_k(Q_\alpha(\cdot))\sqrt{Q_\alpha(\cdot)} = \lambda_k f_k(Q_\alpha(\cdot))\sqrt{Q_\alpha(\cdot)}$, the eigenvalue problem. In this differential equation we recognize the one (Szegő [12], p. 100) which characterizes the classical Laguerre polynomials

$$L_k^{(\alpha)}(x) = \sum_{j=0}^{k} \binom{k+\alpha}{k-j} \frac{(-x)^j}{j}\,, \qquad x > 0\,,$$

defined by

$$\int_0^\infty L_n^{(\alpha)}(x) L_m^{(\alpha)}(x) g_{\alpha+1}(x)\,dx = \binom{n+\alpha}{n}\delta_{nm}\,, \qquad n,m=0,1,2,\dots\,, \tag{2.7}$$

where δ_{nm} is Kronecker's symbol: the equation has $L_k^{(\alpha)}(\cdot)$ as the unique polynomial solution if and only if

$$\frac{1}{\lambda_k} - 1 = k\,, \quad \text{that is}\,, \quad \lambda_k = \frac{1}{k+1} \qquad \text{for all } k=0,1,2,\dots\,.$$

Note that (2.6) is trivially satisfied with $f_k(\cdot) = L_k^{(\alpha)}(\cdot)$ for all $k=0,1,2,\dots$.

Introducing now the normalized eigenfunctions

$$h_k^{(\alpha)}(t) = \frac{L_k^{(\alpha)}(Q_\alpha(t))\sqrt{Q_\alpha(t)}}{\sqrt{\alpha\binom{k+\alpha}{k}}}\,, \qquad 0 \le t < 1\,,\ \ k=0,1,2,\dots\,,$$

notice that (2.7) is equivalent to the ordinary orthogonality relation

$$\int_0^1 h_n^{(\alpha)}(t)\, h_m^{(\alpha)}(t)\,dt = \delta_{nm}\,, \qquad n,m=0,1,2,\dots\,.$$

Then the zero-mean jointly normal random variables

$$Z_{k+1} = \sqrt{k+1}\int_0^1 Z_\alpha(t)\, h_k^{(\alpha)}(t)\,dt\,, \qquad k=0,1,2,\dots\,,$$

have covariance

$$\begin{aligned} E(Z_{k+1}Z_{m+1}) &= \int_0^1\int_0^1 \frac{\big[\min(s,t)-st\big]\, h_k^{(\alpha)}(s)\, h_m^{(\alpha)}(t)}{g_\alpha(Q_\alpha(s))\, g_\alpha(Q_\alpha(t))\sqrt{Q_\alpha(s)}\sqrt{Q_\alpha(t)}}\,ds\,dt\,, \\ &= \frac{\sqrt{k+1}\sqrt{m+1}}{k+1}\int_0^1 h_k^{(\alpha)}(t)\, h_m^{(\alpha)}(t)\,dt = \delta_{km} \end{aligned}$$

for all $m,k=0,1,2,\dots$, and hence are independent standard normal variables. In terms of this sequence, the Karhunen-Loève expansion of $Z_\alpha(t)$, holding in $L_2(\Omega,\mathcal{A},\mathbb{P})$ pointwise, is

$$Z_\alpha(t) = \sum_{k=0}^\infty \frac{h_k^{(\alpha)}(t)}{\sqrt{k+1}}\, Z_{k+1}\,, \qquad 0 < t < 1\,,$$

and we have

$$
\begin{aligned}
V &= \int_0^1 \big[Z_\alpha^2(t) - E\big(Z_\alpha^2(t)\big)\big]\,dt - \left[\frac{1}{\sqrt{\alpha}}\int_0^1 Z_\alpha(t)\sqrt{Q_\alpha(t)}\,dt\right]^2 \\
&= \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\frac{Z_{k+1}Z_{m+1} - E(Z_{k+1}Z_{m+1})}{\sqrt{k+1}\sqrt{m+1}}\int_0^1 h_k^{(\alpha)}(t)\,h_m^{(\alpha)}(t)\,dt \\
&\quad - \left[\sum_{k=0}^{\infty}\frac{Z_{k+1}}{\sqrt{k+1}}\int_0^1 h_k^{(\alpha)}(t)\frac{\sqrt{Q_\alpha(t)}}{\sqrt{\alpha\binom{0+\alpha}{0}}}\,dt\right]^2 \\
&= \sum_{k=0}^{\infty}\frac{Z_{k+1}^2 - 1}{k+1} - \left[\sum_{k=0}^{\infty}\frac{Z_{k+1}}{\sqrt{k+1}}\int_0^1 h_k^{(\alpha)}(t)\,h_0^{(\alpha)}(t)\,dt\right]^2 \\
&= \sum_{k=0}^{\infty}\frac{Z_{k+1}^2 - 1}{k+1} - Z_1^2 = -1 + \sum_{m=2}^{\infty}\frac{Z_m^2 - 1}{m}\,.
\end{aligned}
$$

The formal manipulations may be made rigorous as in the proof of Theorem 3.6 in del Barrio *et al.* [6], all random integrals and infinite series exist as $L_2(\Omega, \mathcal{A}, \mathbb{P})$ limits, proving the last statement of the theorem. □

3. Lognormal families

Consider the Lognormal$_\alpha$ scale family $\mathcal{G}_\alpha = \{G_{\alpha,\sigma}(x) = G_\alpha(x/\sigma),\ x \in \mathbb{R} : \sigma > 0\}$ of distribution functions, where, writing $\log^2 x = (\log x)^2$, the generating distribution function $G_\alpha(x) = \int_{-\infty}^{x} g_\alpha(y)\,dy$, $x \in \mathbb{R}$, is given by the lognormal density function

$$
g_\alpha(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\alpha x}\exp\left\{-\frac{\log^2 x}{2\alpha^2}\right\} = \frac{1}{\sqrt{2\pi}\alpha x^{1+2^{-1}\alpha^{-2}\log x}} & \text{if } x > 0\,, \\ 0 & \text{if } x < 0\,, \end{cases}
$$

where α^2 is a known initial variance: $G_{\alpha,\sigma}(x) = \mathbb{P}\{\sigma e^{Z_{(\alpha)}} \le x\}$, $x \ge 0$, where $Z_{(\alpha)}$ is a $\mathrm{N}(0, \alpha^2)$ random variable. This is often used to model relatively long-tail phenomena. Let $Q_\alpha(t) = G_\alpha^{-1}(t)$ be the corresponding quantile function, given by $G_\alpha\big(Q_\alpha(t)\big) = t \in [0, 1)$, for which $Q_\alpha(0) = 0$ and $Q_\alpha(1) = \infty$. Since

$$
g'_\alpha(x) = -\left[\frac{1}{\alpha x^2} + \frac{\log x}{\alpha^3 x^2}\right]\frac{1}{\sqrt{2\pi}}\,e^{-\frac{\log^2 x}{2\alpha^2}}\,,
$$

we see that $x^2 g'_\alpha(x) \to 0$ both as $x \downarrow 0$ and $x \to \infty$, so that the choice of the weight function $w_\alpha(\cdot) := w_{G_\alpha}(\cdot)$ according to (1.4) is still possible. We have

$$L_\alpha(x) = -1 - \frac{x g'_\alpha(x)}{g_\alpha(x)} = \frac{\log x}{\alpha^2}, \qquad \text{so that} \quad L'_\alpha(x) = \frac{1}{\alpha^2 x}, \quad x > 0,$$

and $I_\alpha := I_{G_\alpha} = \int_0^1 L'_\alpha\big(Q_\alpha(t)\big) Q_\alpha(t)\, dt = 1/\alpha^2$, and so

$$w_\alpha(t) = \frac{1}{I_\alpha} \frac{L'_\alpha\big(Q_\alpha(t)\big)}{Q_\alpha(t)} = \frac{1}{Q_\alpha^2(t)}, \qquad 0 < t < 1. \tag{3.1}$$

We have $\mu(G_\alpha, w_\alpha) = \int_0^1 Q_\alpha^2(t) w_\alpha(t)\, dt = 1$, as it should be. Let $\varphi(x)$ and $\Phi(x) = \int_{-\infty}^x \varphi(y)\, dy$, $x \in \mathbb{R}$, be the standard normal density and distribution functions and let $\Phi^{-1}(t)$ be the standard normal quantile function: $\Phi\big(\Phi^{-1}(t)\big) = t \in (0,1)$. Since

$$G_\alpha(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \frac{1}{\alpha y} e^{-\frac{1}{2}\left(\frac{\log y}{\alpha}\right)^2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\log x}{\alpha}} e^{\frac{t^2}{2}} dt = \Phi\left(\frac{\log x}{\alpha}\right), \qquad x > 0,$$

we have $t = G_\alpha\big(Q_\alpha(t)\big) = \Phi\big(\alpha^{-1} \log Q_\alpha(t)\big)$, thus $\Phi^{-1}(t) = \alpha^{-1} \log Q_\alpha(t)$, and so

$$Q_\alpha(t) = e^{\alpha \Phi^{-1}(t)}, \qquad 0 \le t < 1. \tag{3.2}$$

Given a sample $X_1, X_2, \ldots, X_n$ from F, the weight function $w_\alpha(\cdot)$ in (3.1) now gives the test statistic

$$\begin{aligned} S_n(\alpha) := S_n(w_\alpha) &= 1 - \frac{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{Q_n(t)}{Q_\alpha(t)}\, dt\right]^2}{\left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} dt\right] \left[\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{Q_n^2(t)}{Q_\alpha^2(t)}\, dt\right]} \\ &= 1 - \frac{(n+1)\left[b^*_{1,n}(\alpha) X_{1,n} + \sum_{k=2}^{n-1} b_{k,n}(\alpha) X_{k,n} + b^*_{n,n}(\alpha) X_{n,n}\right]^2}{(n-1)\left[d^*_{1,n}(\alpha) X_{1,n}^2 + \sum_{k=2}^{n-1} d_{k,n}(\alpha) X_{k,n}^2 + d^*_{n,n}(\alpha) X_{n,n}^2\right]}, \end{aligned}$$

where, using (3.2) and seeing that machine computations are again unavoidable,

$$b^*_{1,n}(\alpha) = \int_{\Phi^{-1}\left(\frac{1}{n+1}\right)}^{\Phi^{-1}\left(\frac{1}{n}\right)} e^{-\alpha x} \varphi(x)\, dx, \qquad b^*_{n,n}(\alpha) = \int_{\Phi^{-1}\left(\frac{n-1}{n}\right)}^{\Phi^{-1}\left(\frac{n}{n+1}\right)} e^{-\alpha x} \varphi(x)\, dx,$$

$$b_{k,n}(\alpha) = \int_{\frac{k-1}{n}}^{\frac{k}{n}} e^{-\alpha\Phi^{-1}(t)}\,dt = \int_{\Phi^{-1}(\frac{k-1}{n})}^{\Phi^{-1}(\frac{k}{n})} e^{-\alpha x}\varphi(x)\,dx\,, \qquad k=2,\dots,n-1\,,$$

$$d^*_{1,n}(\alpha) = \int_{\Phi^{-1}(\frac{1}{n+1})}^{\Phi^{-1}(\frac{1}{n})} e^{-2\alpha x}\varphi(x)\,dx\,, \qquad d^*_{n,n}(\alpha) = \int_{\Phi^{-1}(\frac{n-1}{n})}^{\Phi^{-1}(\frac{n}{n+1})} e^{-2\alpha x}\varphi(x)\,dx\,,$$

$$d_{k,n}(\alpha) = \int_{\frac{k-1}{n}}^{\frac{k}{n}} e^{-2\alpha\Phi^{-1}(t)}\,dt = \int_{\Phi^{-1}(\frac{k-1}{n})}^{\Phi^{-1}(\frac{k}{n})} e^{-2\alpha x}\varphi(x)\,dx\,, \qquad k=2,\dots,n-1\,.$$

Setting $c_n(\alpha) = c_n(w_\alpha)$, the result for testing lognormal scale families is

Theorem 3. *If $F \in \mathcal{G}_\alpha$ for some $\alpha > 0$, then*

$$\frac{nS_n(\alpha)}{\alpha^2} - c_n^\star \xrightarrow{\mathcal{D}} V,$$

where, with $\gamma = -\Gamma'(1) = 0.577215\dots$ still standing for Euler's constant,

$$\begin{aligned} c_n^\star &= \frac{n+1}{n-1}\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{t(1-t)}{\varphi^2\big(\Phi^{-1}(t)\big)}\,dt = \frac{n+1}{n-1}\int_{\Phi^{-1}(\frac{1}{n+1})}^{\Phi^{-1}(\frac{n}{n+1})} \frac{\Phi(x)\big[1-\Phi(x)\big]}{\varphi(x)}\,dx \\ &= \log\log(n+1) + \gamma + \log 2 + o(1) \end{aligned}$$

and

$$\begin{aligned} V &= \frac{1}{\alpha^2}\int_0^1 \frac{B^2(t)-t(1-t)}{g_\alpha^2(Q_\alpha(t))Q_\alpha^2(t)}\,dt - \left[\frac{1}{\alpha}\int_0^1 \frac{B(t)}{g_\alpha(Q_\alpha(t))Q_\alpha(t)}\,dt\right]^2 \\ &= \int_0^1 \frac{B^2(t)-t(1-t)}{\varphi^2\big(\Phi^{-1}(t)\big)}\,dt - \left[\int_0^1 \frac{B(t)}{\varphi\big(\Phi^{-1}(t)\big)}\,dt\right]^2 \\ &\stackrel{\mathcal{D}}{=} \sum_{m=1}^\infty \frac{Z_m^2-1}{m} - Z_1^2 = -1 + \sum_{m=2}^\infty \frac{Z_m^2-1}{m}\,, \end{aligned}$$

where, again, $Z_1, Z_2, \dots$ are independent standard normal random variables.

Here we only have to prove distributional convergence to the first expression given for V, as a consequence of Theorem 1. The second equation for V is a simple algebraic fact resulting from (3.2), while the final distributional equality follows from the proof of Theorem 3.6 in del Barrio *et al.* [6], which is for testing for normality. In fact, the asymptotic distribution here is nothing but the one obtained when testing for normality with a known variance and an unknown mean; this is to be expected in view of the equation $G_\alpha(x/\sigma) = \Phi\big(\alpha^{-1}[\log x - \log\sigma]\big)$, $x > 0$, which for $\sigma = 1$ led to (3.2). The secret of all this is that the weight function in (3.1) transforms the lognormal scale testing problem, along with the whole proof below, to the normal location testing problem in an optimal fashion. The identity of the limiting distributions in Theorems 2 and 3 here, and that of Theorem 3 in Csörgő [4] for all Weibull cases, is remarkable. In particular, the numerical results for $H(x) = \mathbb{P}\{V \le x\}$, $x \in \mathbb{R}$, and its quantiles $H^{-1}(\cdot)$, mentioned in the previous section, still apply. The asymptotic evaluation of $c_n^\star$ in the theorem above was done by Bickel and van Zwet [1], formula (5.56) on p. 973.

Proof. Noting that

$$c_n^\star = \frac{c_n(\alpha)}{\alpha^2} = \frac{\frac{1}{\alpha^2}\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{t(1-t)}{g_\alpha^2\big(Q_\alpha(t)\big)Q_\alpha^2(t)}\,dt}{\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{Q_\alpha^2(t)}{Q_\alpha^2(t)}\,dt} = \frac{n+1}{n-1}\int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{t(1-t)}{\varphi^2\big(\Phi^{-1}(t)\big)}\,dt\,,$$

convergence in distribution to V will follow from Theorem 1 once we show that conditions (1.1)–(1.3) are satisfied in the present situation. For the first of these,

$$A(\alpha) := A(G_\alpha) = \sup_{0<t<1} \sqrt{2\pi}\,t(1-t)\left|\alpha + \frac{\Phi^{-1}(t)}{\alpha}\right| e^{[\Phi^{-1}(t)]^2/2}\,,$$

so that $A(\alpha) = \sup_{x\in\mathbb{R}} A_\alpha(x)$, where $A_\alpha(x) = \Phi(x)\big[1-\Phi(x)\big]\big|\alpha + \alpha^{-1}x\big|/\varphi(x)$. Since $\Phi(x) \sim \varphi(x)/|x|$ as $x \to -\infty$ and $1-\Phi(x) \sim \varphi(x)/x$ as $x \to \infty$, we have $A_\alpha(x) \to 1/\alpha$ both as $x \to -\infty$ and $x \to \infty$, which imply that $A(\alpha) < \infty$.

Next, concerning the finiteness of $J(\alpha) := J(G_\alpha, w_\alpha)$ in (1.2), we have

$$\begin{aligned}\frac{J(\alpha)}{\alpha^4} &= \int_0^1\!\!\int_0^1 \frac{\big[\min(s,t)-st\big]^2}{\alpha^4 g_\alpha^2\big(Q_\alpha(s)\big)g_\alpha^2\big(Q_\alpha(t)\big)Q_\alpha^2(s)Q_\alpha^2(t)}\,ds\,dt\\ &= \int_0^1\!\!\int_0^1 \frac{\big[\min(s,t)-st\big]^2}{\varphi^2\big(\Phi^{-1}(s)\big)\varphi^2\big(\Phi^{-1}(t)\big)}\,ds\,dt < \infty\end{aligned}$$

as established by del Barrio *et al.* [5, 6] for the normal case.

Finally, for $a_n(\alpha,\delta) := a_n(G_\alpha, w_\alpha, \delta)$, for some $\delta \in (1/2,1)$, in condition (1.3),

$$\frac{a_n(\alpha,\delta)}{\alpha^2} = \frac{1}{n^{1-\delta}} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\big[t(1-t)\big]^\delta}{\alpha^2 g_\alpha^2\big(Q_\alpha(t)\big)Q_\alpha^2(t)}\,dt = \frac{1}{n^{1-\delta}} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\big[t(1-t)\big]^\delta}{\varphi^2\big(\Phi^{-1}(t)\big)}\,dt \to 0\,,$$

in fact for all $\delta \in (0,1)$ by (2.7) in [5], proving the theorem. □

In conclusion we point out, just as in Csörgő [4] for the Weibull and Pareto cases, that it would be equally interesting in the present two cases to see what happens with the statistics in Theorems 2 and 3 if α is viewed as a nuisance parameter and is, therefore, estimated, aiming a test for the null hypotheses $F \in \bigcup_{\{\alpha>0\}} \mathcal{G}_\alpha$.

REFERENCES

[1] BICKEL, P. J.—van ZWET, W. R.: *Asymptotic expansions for the power of distribution-free tests in the two-sample problem*, Ann. Statist. **6** (1978), 937–1004.

[2] CHERNOFF, H.—GASTWIRTH, J. L.—JOHNS, M. V.: *Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation*, Ann. Math. Statist. **38** (1967), 52–72.

[3] CSÖRGŐ, S.: *Testing for Weibull scale families as a test case for Wasserstein correlation tests*, Discussion of del Barrio, Cuesta-Albertos and Matrán [6], Test **9** (2000), 54–70.

[4] CSÖRGŐ, S.: *Weighted correlation tests for scale families*, Test **11** (2002), 219–248.

[5] DEL BARRIO, E.—CUESTA-ALBERTOS, J. A.—MATRÁN, C.—RODRÍGUEZ-RODRÍGUEZ, J. M.: *Tests of goodness of fit based on the L_2-Wasserstein distance*, Ann. Statist. **27** (1999), 1230–1239.

[6] DEL BARRIO, E.—CUESTA-ALBERTOS, J. A.—MATRÁN, C.: *Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests*, with discussion, Test **9** (2000), 1–96.

[7] DE WET, T.: *Discussion of del Barrio, Cuesta-Albertos and Matrán* [6], Test **9** (2000), 74–79.

[8] DE WET, T.: *Goodness-of-fit tests for location and scale families based on a weighted L_2-Wasserstein distance measure*, Test **11** (2002), 89–107.

[9] DE WET, T.—VENTER, J. H.: *A goodness of fit test for a scale parameter family of distributions*, South African Statist. J. **7** (1973), 35–46.

[10] ERDÉLYI, A.—MAGNUS, W.—OBERHETTINGER, F.—TRICOMI, F. G.: *Higher Transcendental Functions*, Vol. II, McGraw-Hill, New York, 1953.

[11] RIESZ, F.—SZ.-NAGY, B.: *Functional Analysis*, Frederick Ungar Publishing Co., New York, 1955.

[12] SZEGŐ, G.: *Orthogonal Polynomials*, Fourth Edition, Colloquium Publications XXIII, American Mathematical Society, Providence, Rhode Island, 1975.

Received March 26, 2002

Bolyai Institute
University of Szeged
Aradi vértanúk tere 1
H–6720 Szeged
HUNGARY

E-mail: csorgo@math.u-szeged.hu
szbtmsz@math.u-szeged.hu