South African Statistical Journal
Vol. 54, No. 1, 1–13
https://doi.org/10.37920/sasj.2020.54.1.1

OPTIMALITY IN WEIGHTED L_2-WASSERSTEIN GOODNESS-OF-FIT STATISTICS

Tertius de Wet[1]
Department of Statistics and Actuarial Science, University of Stellenbosch, South Africa
e-mail: *tdewet@sun.ac.za*

Veronica Humble
Legal & General Investment Management, London, UK
e-mail: *veronica.esaulova@gmail.com*

In Del Barrio, Cuesta-Albertos, Matran and Rodriguez-Rodriguez (1999) and Del Barrio, Cuesta-Albertos and Matran (2000), the authors introduced a new class of goodness-of-fit statistics based on the L_2-Wasserstein distance. It was shown that the desirable property of loss of degrees-of-freedom holds only under normality. Furthermore, these statistics have some limitations in their applicability to heavier-tailed distributions. To overcome these problems, the use of weight functions in the statistics was proposed and investigated by De Wet (2000), De Wet (2002) and Csörgő (2002). In the former the issue of loss of degrees-of-freedom was considered and in the latter the application to heavier-tailed distributions. In De Wet (2000) and De Wet (2002) it was shown how the weight functions could be chosen in order to retain the loss of degrees-of-freedom property separately for location and scale families. The weight functions that give this property, are the ones that give asymptotically optimal estimators for respectively the location and scale parameters – thus estimation optimality. In this paper we show that in the location case, this choice of "estimation optimal" weight function also gives "testing optimality", where the latter is measured in terms of approximate Bahadur efficiencies.

Key words: Bahadur approximate efficiency, Degrees-of-freedom, Goodness-of-fit, Location families, Optimal weight function, Weighted Wasserstein distance.

1. Introduction

In their paper Del Barrio et al. (1999), the authors introduced and studied a new class of statistics for testing for normality, based on the L_2-Wasserstein distance (see also Krauczi, 2009 for some simulation results of these statistics and Ramdas, Trillos and Cuturi, 2017). In Del Barrio et al. (2000) they extended the statistics to apply to general location-scale families of distributions and found the limiting distribution in terms of quadratic functionals of the Brownian bridge process. In this paper it was also shown that the desirable property of loss of degrees-of-freedom holds only under normality. Furthermore, these statistics have some limitations in terms of their applicability only to fairly light-tailed distributions – the latter was discussed quite extensively by Csörgő (2000). To overcome these problems, the use of weight functions in these statistics was proposed and investigated in De Wet (2000), De Wet (2002) and Csörgő (2002). In the former the issue of retaining loss of degrees-of-freedom was investigated and in the latter the issue of application to heavier-tailed distributions.

[1] Corresponding author.
MSC2010 subject classifications. 62F05, 62F03.

More specifically, in De Wet (2000) and De Wet (2002) it was shown how the weight functions could be chosen in order to retain the loss of degrees-of-freedom property separately for location and scale families. In the special case of the normal distribution, these two weight functions coincide, being both identically one, and thus for the normal distribution a loss of two degrees-of-freedom can be achieved for the location-scale case, using the identity weight function. The asymptotic distribution of the weighted test statistics was derived for general location-scale families independently by Csörgő (2003) and Del Barrio, Gine and Utzet (2005) (also see Del Barrio, 2007) using different methods and under different conditions.

In this paper we return to the issue of the loss of degrees-of-freedom. It was shown in De Wet (2002) that the weight functions that give the loss of degrees-of-freedom are the ones that give asymptotically optimal estimators for respectively the location and scale parameter. The question then arises whether this choice of weight functions also gives some form of optimality to the test statistic used. We will answer this question in the affirmative for the case of location alternatives. We now introduce these ideas in a more technical fashion.

For two distribution functions F_1 and F_2 on the real line, the L_2-Wasserstein distance between F_1 and F_2 is defined as

$$W(F_1, F_2) = \int_0^1 \left(F_1^{-1}(t) - F_2^{-1}(t)\right)^2 dt,$$

(see e.g. Bickel and Freedman, 1981 and Shorack and Wellner, 1986). Using this as point of departure, Del Barrio et al. (1999) and Del Barrio et al. (2000) proposed using $W(F_n, F_0)$ as test statistic for the hypothesis $H_0 : F = F_0$, with F_0 fully specified. Here F_n denotes the usual empirical distribution function of the sample. This they then extend to testing in a location-scale family by considering

$$\inf_{\mu,\sigma} W\left(F_n, F_0\left(\frac{\cdot - \mu}{\sigma}\right)\right).$$

In order to make the test statistic scale invariant, they divide the latter by the scale estimator obtained from the infimum.

In Del Barrio et al. (1999) the limiting distribution was found in the case of testing for normality and shown that it has the loss of degrees-of-freedom property in the sense that the limiting distribution loses two terms in the Karhunen-Loève expression for the limiting random variable, compared to the case where the parameters are known. In Del Barrio et al. (2000) (see also Del Barrio et al., 2005 and Del Barrio, 2007) it was shown that the normal distribution uniquely has this property. To retain it more generally, De Wet (2000) and De Wet (2002) proposed using a weighted Wasserstein distance measure, viz

$$W^{(w)}(F_1, F_2) = \int_0^1 \left(F_1^{-1}(t) - F_2^{-1}(t)\right)^2 w(t)\, dt,$$

with w an appropriate (positive) weight function on $(0, 1)$. He showed that the property could be realised separately for unknown location and scale parameters, by choosing w as below.

Let $f_0 = F_0'$ and $Q_0 = F_0^{-1}$. Then, in the location case, take

$$w(t) = I_1^{-1} L_1'(Q_0(t)) = J_1(t) \text{ (say)}, \tag{1}$$

with

$$L_1(y) = -f_0'(y)/f_0(y) \tag{2}$$

and

$$I_1 = \int_{-\infty}^{\infty} L_1'(y) f_0(y)\, dy \equiv \int_0^1 L_1'(Q_0(u))\, du.$$

The case of a normal distribution will be an example of special interest to us. In that case it follows easily that

$$L_1(y) = y,\ L_1'(y) = 1,\ I_1 = 1,\ J_1(t) \equiv 1.$$

In the scale case, take:

$$w(t) = I_2^{-1} L_2'(Q_0(t)) / Q_0(t) = J_2(t) \text{ (say)}, \tag{3}$$

with

$$L_2(y) = -1 - y f_0'(y) / f_0(y)$$

and

$$I_2 = \int_{-\infty}^{\infty} L_2^2(y) f_0(y)\, dy \equiv \int_0^1 L_2'(Q_0(u)) Q_0(u)\, du.$$

Considering again the normal case as an example, we obtain the following

$$L_2(y) = y^2 - 1,\ L_2'(y) = 2y,\ I_2 = 2,\ J_2(t) \equiv 1.$$

We note that one degree-of-freedom is lost in each case. The rationale for these weight functions is that they lead to asymptotically optimal estimators obtained from the minimized weighted Wasserstein distance. This naturally leads to the question whether these choices of weight functions also give the corresponding test statistics some optimality property. We show below that in the location case, this is indeed the case in terms of Bahadur efficiencies based on approximate Bahadur slopes (see e.g. Bahadur, 1967 or Nikitin, 1995). To be more precise, consider the null hypothesis $H_0 : F = F_0$, with F_0 fully specified, and denote the alternative hypothesis by $H_A : F = F_\theta$, where we write F_0 for F_{θ_0}. Let $X_1, X_2, \ldots, X_n$ be a sample on F and F_n the empirical distribution function of the sample. To test H_0 we use the weighted Wasserstein statistic

$$T_n \equiv n \int_0^1 \left(F_n^{-1}(t) - Q_0(t)\right)^2 w(t)\, dt, \tag{4}$$

for a given choice of w. We will show that if w is chosen as in (1) and θ is a location parameter, then this choice w of gives the best approximate Bahadur slope amongst all weight functions satisfying the specified conditions.

Before proceeding to the next section, we mention some further aspects related to these statistics. In Csörgő (2002) weighted Wasserstein statistics were also considered, albeit for a different reason. It is known that correlation-type statistics have certain limitations when used for heavier tailed distributions (see e.g. Lockhart 1991 and McLaren and Lockhart 1987). In order to overcome this limitation, Csörgő (2002) proposed using a weight function in the Wasserstein statistic and showed precisely to what extent this alleviates the problem. Furthermore, in Csörgő (2003) and independently in Del Barrio et al. (2005), the asymptotic distribution of the general weighted Wasserstein statistic was derived under different sets of conditions. The proofs of our results depend heavily on these.

The layout of the paper is as follows: In the next section we discuss approximate Bahadur slopes as they apply to the Wasserstein statistics. In Section 3 we show optimality of the weight function J_1, and also state and prove our main results. In Section 4 the efficiencies are compared for a number of different weight functions. Some concluding remarks are made in the final section.

2. Bahadur slopes for Wasserstein statistics

The Bahadur efficiency is a well-known measure for comparing the efficiencies of different test statistics. It is based on so-called Bahadur slopes (either exact or approximate) and indicates the rate at which the attained level of a test statistic converges to zero under the alternative (typically at an exponential rate). See e.g. Nikitin (1995) for a very clear exposition of this and other measures of efficiency of test statistics. We will follow Gregory (1980) in using approximate Bahadur efficiencies. We briefly summarise Bahadur's results for this (see e.g. Bahadur, 1967 or Nikitin, 1995 for more details and Grané and Fortiana, 2008, for a recent related, but different, application of Bahadur efficiency).

Let $X_1, \ldots, X_n$ be i.i.d. each with distribution P_θ, for an unknown parameter $\theta \in \Theta$. As a test for $H_0 : \theta \in \Theta_0 \subset \Theta$ reject the hypothesis for large values of a statistic $V_n \equiv V_n(X_1, \ldots, X_n)$. Denote the limiting distribution function of V_n by G and suppose that for each $\theta \in \Theta_0$

$$P_\theta\left(V_n \leq t\right) \rightarrow G\left(t\right), \text{ as } n \rightarrow \infty.$$

Define

$$L_n = 1 - G\left(V_n\right),$$

and suppose $\{V_n\}$ satisfies, for all $\theta \in \Theta - \Theta_0$,

$$n^{-\frac{1}{2}} V_n \rightarrow b\left(\theta\right), \text{ in } P_\theta \text{ probability.}$$

Also, suppose that for some constant $0 < a < \infty$,

$$\log\left(1 - G\left(t\right)\right) = -\frac{1}{2} a t^2 \left(1 + o\left(1\right)\right), \text{as } t \rightarrow \infty.$$

Then, Bahadur showed that

$$\lim_{n\rightarrow\infty} n^{-1} \log L_n - \frac{1}{2} a b^2\left(\theta\right) \equiv -\frac{1}{2} s\left(\theta\right) \text{ in } P_\theta \text{ probability.}$$

Since $-2 \log L_n \approx n s(\theta)$, Bahadur termed $s(\theta) = a b^2(\theta)$ the approximate slope of the sequence $\{V_n\}$. The ratio of the approximate slopes of two sequences of test statistics is called their approximate Bahadur efficiency.

We now apply these results to Wasserstein statistics. With T_n as defined above we have, using the results of e.g. Del Barrio et al. (2005), under the null hypothesis and under conditions specified there, as $n \rightarrow \infty$, that

$$T_n - a_n \xrightarrow{D} \int_0^1 \left(B^2\left(t\right) - E B^2\left(t\right)\right) Q_0'\left(t\right)^2 w\left(t\right) dt \overset{D}{\equiv} \sum_m \gamma_m \left(Z_m^2 - 1\right), \tag{5}$$

with $B(t)$ a Brownian bridge process and $\{Z_m\}$ i.i.d. $N(0,1)$ and $\{\gamma_m\}$ the eigenvalues of the covariance kernel

$$K_w\left(s,t\right) = \left(s \wedge t - st\right) Q_0'\left(s\right) Q_0'\left(t\right) \left(w\left(s\right) w\left(t\right)\right)^{\frac{1}{2}}.$$

See Del Barrio et al. (2005) Theorem 4.6 (ii). Here $\{a_n\}$ is a sequence of constants specified in the theorem. From Lemma 2.4 of Gregory (1980) it follows that

$$\lim_{x\rightarrow\infty} \left[\frac{-\log\left(1 - G\left(x\right)\right)}{x}\right] = 2\gamma_1,$$

with $\gamma_1 = \max_m \{\gamma_m\}$. Now, following the approach of Gregory (1980), define

$$T_n^* = \max(T_n - a_n, 0)$$

and let G^* denote its distribution function.

For our goodness-of-fit testing problem we have $\Theta_0 = \{\theta_0\}$, and under appropriate conditions (see the next section), we expect,

$$n^{-1}(T_n - a_n) \xrightarrow{P_\theta} \int_0^1 \left(F_\theta^{-1}(t) - Q_0(t)\right)^2 w(t)\, dt \tag{6}$$

and thus

$$n^{-\frac{1}{2}} T_n^* \xrightarrow{P_\theta} \left(\int_0^1 \left(F_\theta^{-1}(t) - Q_0(t)\right)^2 w(t)\, dt\right)^{\frac{1}{2}}.$$

Also, for $x > 0$

$$G^*(x) = G\left(x^2\right),$$

and thus from Gregory's result

$$\lim_{x\to\infty}\left[\frac{-\log(1 - G^*(x))}{x^2}\right] = 2\gamma_1.$$

Using these in the result of Bahadur, we find that the approximate slope is given by

$$s^{(w)}(\theta) = \int_0^1 \left(F_\theta^{-1}(t) - Q_0(t)\right)^2 w(t)\, dt/\gamma_1. \tag{7}$$

Writing $\gamma^{(w)}$ for γ to indicate the dependence on the weight function, we have that for two weight functions w_1 and w_2, $s^{(w_1)} \geq s^{(w_2)}$ (i.e. w_1 is preferable to w_2) if

$$\gamma_1^{(w_2)} \int_0^1 \left(F_\theta^{-1}(t) - Q_0(t)\right)^2 w_1(t)\, dt \geq \gamma_1^{(w_1)} \int_0^1 \left(F_\theta^{-1}(t) - Q_0(t)\right)^2 w_2(t)\, dt.$$

Remark. This result simplifies considerably in the case of a location alternative, i.e. where $F_\theta(x) = F_0(x - \theta)$. Clearly, in this case, w_1 is preferable to w_2 if

$$\gamma_1^{(w_2)} \int_0^1 w_1(t)\, dt \geq \gamma_1^{(w_1)} \int_0^1 w_2(t)\, dt. \tag{8}$$

Note in this case the parameter θ cancels out.

In the next section we will show that the weight function J_1 discussed above, satisfies (8), i.e. we will show that for any w

$$\gamma_1^{(w)} \int_0^1 J_1(t)\, dt \geq \gamma_1^{(J_1)} \int_0^1 w(t)\, dt, \tag{9}$$

with J_1 given by (1).

3. Optimality results

In this section we show that the weight function J_1 has the highest approximate Bahadur efficiency among all other weight functions, in the case of location alternatives. The results are stated with the proofs deferred to the Appendix.

Our results will make essential use of Theorem 4.6 (ii), page 173, of Del Barrio et al. (2005). This result holds under their conditions (GH), (2.10) and (4.18). We call this set of conditions (DGU). Under this condition, (5) holds. In Theorem 1 below, we will show that for location alternatives, with $\theta \neq \theta_0$, (6) also holds under this set of conditions. In De Wet (2002) it was shown that in the location case, the term that is lost in the limiting random variable $\sum_m \gamma_m(Z_m^2 - 1)$, is the one corresponding to the eigenvalue $\gamma_k = I_1^{-1}$ (see Remark 2.1 there).

Also, in De Wet (2002), Theorem 2.4, it was shown that the eigenfunction corresponding to this is

$$g_k(t) = (J_1(t))^{\frac{1}{2}}.$$

We now show that $\gamma_k = I_1^{-1}$ actually corresponds to the largest eigenvalue, i.e. $k = 1$.

Lemma 1. *Assuming* $\log(1/f_0(x))$ *is convex,* $(J_1(t))^{\frac{1}{2}}$ *is the eigenfunction of* K_{J_1} *corresponding to the largest eigenvalue.*

Remark. From Lemma 1 and (9), for optimality we thus need to show that

$$\gamma_1^{(w)} \int_0^1 J_1(t)\,dt \geq I_1^{-1} \int_0^1 w(t)\,dt.$$

From Remark 2.1 of De Wet (2002) we also know that

$$\int_0^1 J_1(t)\,dt = 1,$$

and for optimality we thus need to prove that

$$\gamma_1^{(w)} \geq I_1^{-1} \int_0^1 w(t)\,dt. \tag{10}$$

This result and its proof is given in Theorem 1 below. The approach to the proof is to not work with K_{J_1} but with its so-called related kernel, denoted by c_{J_1}. The latter is now briefly discussed. Define

$$\psi(s,u) = I(s \leq u) - u = \begin{cases} 1-u & \text{if } s \leq u \\ -u & \text{if } s > u \end{cases}$$

and

$$q(s,t) = \psi(s,t)\, Q_0'(t)\, (w(t))^{\frac{1}{2}}.$$

Let Γ_A be the integral operator corresponding to a function A, i.e.

$$(\Gamma_A g)(s) = \int_0^1 A(s,t)\, g(t)\,dt.$$

It then follows that

$$\Gamma_{K_w} = \Gamma_q^* \Gamma_q, \tag{11}$$

with Γ^* denoting the adjoint of Γ.

Now define

$$\Gamma_{c_w} = \Gamma_q \Gamma_q^*. \tag{12}$$

Then c_w is called the related kernel of K_w, and vice versa (see e.g. De Wet, 1980 in this regard). It follows quite easily that c_w is given explicitly as

$$c_w(s,t) = \int_0^1 \psi(s,u)\,\psi(t,u)\,Q_0'(u)^2\,w(u)\,du. \tag{13}$$

From (11) and (12) it follows directly that K_w and c_w have the same eigenvalues. This implies in particular that $\gamma_1^{(w)}$ is also the largest eigenvalue of c_w. Furthermore, the relationship between the (orthonormal) eigenfunctions is as follows. Let $\|\cdot\|$ denote the L_2-norm (see De Wet, 1980). If g denotes an eigenfunction of K_w and h the corresponding eigenfunction of c_w, then we have immediately from (11) and (12) that $h = \Gamma_q g / \left\|\Gamma_q g\right\|$, i.e.

$$h(s) = \int_0^1 q(s,v)\,g(v)\,dv \Big/ \left\| \int_0^1 q(s,v)\,g(v)\,dv \right\|, \tag{14}$$

and similarly that $g = \Gamma_q^* h / \left\|\Gamma_q^* h\right\|$.

Write $g_1(t) = (J_1(t))^{\frac{1}{2}}$ for the eigenfunction corresponding to the largest eigenvalue $\gamma_1^{(J_1)} = I^{-1}$. The next lemma gives the corresponding eigenfunction of the related kernel c_{J_1}.

Lemma 2. *Assume $f_0'(y) \to 0$, as $y \to \pm\infty$. Then the eigenfunction of c_{J_1} corresponding to its largest eigenvalue $\gamma_1^{(J_1)} = I_1^{-1}$, is given by*

$$h_1(s) = I_1^{-\frac{1}{2}} \left(-L_1\left(Q_0(s)\right)\right). \tag{15}$$

We now give our main result. Note that we take $\theta_0 = 0$ without loss of generality.

Theorem 1. *Under the set of conditions DGU and those of Lemmas 1 and 2, and in the case of location alternatives, we have that (10) holds, i.e.*

$$\gamma_1^{(w)} \geq I_1^{-1} \int_0^1 w(t)\,dt.$$

Remark. (i) This theorem says that the weight function J_1 gives the highest Bahadur efficiency amongst all weight functions satisfying the DGU conditions. In the next section we compare its Bahadur efficiency relative to two well-known weight functions, viz. the Cramér-von Mises and Anderson-Darling weight functions, in the case of the normal and logistic distributions.

(ii) The other interesting case is that of scale alternatives. Here, however, we do not necessarily work with the largest eigenvalue (e.g. in the case of normal distributions, we have the scale case corresponding to the second largest eigenvalue). We thus need to approach this situation using a different proof.

4. Relative Bahadur efficiencies

In this section we calculate the relative Bahadur efficiencies of the optimal weight function compared to two well-known ones, viz. the Cramér-von Mises and Anderson-Darling weights, applying it to the normal and logistic distributions. Consider first the Cramér-von Mises weight function, denoted by $w_{(CvM)}$. Here we need to have $Q_0'(t)^2 w_{(CvM)} \equiv 1$, giving

$$w_{(CvM)}(t) = f_0(Q_0(t))^2.$$

For this weight function the corresponding eigenvalues are $\gamma_m^{(CvM)} = (\pi m)^{-2}, m = 1, 2, \ldots$ Thus $\gamma_1^{(CvM)} = (\pi)^{-2}$ (see e.g. Shorack and Wellner, 1986).

For the Anderson-Darling weight function, denoted by $w_{(AD)}$, we need to have $Q_0'(t)^2 w_{(AD)} = [t(1-t)]^{-1}$, giving

$$w_{(AD)}(t) = f_0(Q_0(t))^2 [t(1-t)]^{-1}. \tag{16}$$

The corresponding eigenvalues are $\gamma_m^{(AD)} = [m(m+1)]^{-1}, m = 1, 2, \ldots$ Thus $\gamma_1^{(AD)} = \frac{1}{2}$ (see e.g. Shorack and Wellner, 1986).

Now from (7) and (8) we find that the relative efficiency using a weight function w_1 compared to using weight function w_2, is given by

$$e_{w_1,w_2} = \frac{\gamma_1^{(w_2)}}{\gamma_1^{(w_1)}} \cdot \frac{\int_0^1 w_1(t)\,dt}{\int_0^1 w_2(t)\,dt}.$$

Thus, with the special choice using the optimal weight function J_1, remembering that $\gamma_1^{(J_1)} = I_1^{-1}$ and noting that $\int_0^1 J_1(t)\,dt = 1$ (see De Wet, 2002, equation (2.8)), we have that for a generic w the relative efficiency is equal to

$$e_{J_1,w} = I_1 \gamma_1^{(w)} \Big/ \int_0^1 w(t)\,dt. \tag{17}$$

4.1 Cramér-von Mises weights

Note that in this case we have

$$\int_0^1 w_{(CvM)}(t)\,dt = \int_0^1 f_0(Q_0(t))^2\,dt = \int_{-\infty}^{\infty} f_0(x)^3\,dx.$$

This latter integral is easily found to equal $(2\pi\sqrt{3})^{-1}$ and $(30)^{-1}$ for the standard normal and logistic distributions, respectively. Furthermore, from De Wet (2002) we have $I_1 = 1$ and $I_1 = \frac{1}{3}$ for these two distributions, respectively. Using (17), this then gives the following relative efficiencies: For the normal: $e_{J_1,CvM} = 1.1026$. For the logistic: $e_{J_1,CvM} = 1.1032$.

4.2 Anderson-Darling weights

In this case we have

$$\int_0^1 w_{(AD)}(t)\,dt = \int_{-\infty}^{\infty} f_0(x)^3 [F(x)(1-F(x))]^{-1}\,dx.$$

This latter integral is easily found to equal 0.4805 and $\frac{1}{6}$ for the standard normal and logistic distributions, respectively. Again $I_1 = 1$ and $I_1 = \frac{1}{3}$ for the two distributions, respectively. With $\gamma_1^{(AD)} = \frac{1}{2}$, this then gives the relative efficiency for the normal distribution as $e_{J_1,AD} = 1.0406$.

Note that for the logistic distribution direct calculations give $J_1(t) = 6t(1-t)$, while from (16) it follows that $w_{(AD)}(t) = t(1-t)$. Thus J_1 and $w_{(AD)}$ are essentially the same and the Anderson-Darling weight function should give Bahadur efficiency of 1. This is indeed the case as a straightforward calculation shows, viz

$$e_{J_1,AD} = I_1\gamma_1^{(AD)} \Big/ \int_0^1 w_{(AD)}(t)\,dt = \frac{1}{3}\cdot\frac{1}{2}\Big/\frac{1}{6} = 1.$$

Remark. Although we have not proved optimality in the scale case, we can still calculate the relative efficiencies of the weight function J_2 in (3) (i.e., the one leading to a loss of a degree-of-freedom). We compare this also to the Cramér-von Mises and Anderson-Darling weight functions, in the case of the normal and exponential distributions. Straightforward calculations lead to the following relative efficiencies: For the normal: $e_{J_2,CvM} = 6.6160$. For the exponential: $e_{J_2,CvM} = 1.3678$. For the Anderson-Darling weight function, we get: For the normal: $e_{J_2,AD} = 3.7027$. For the exponential: $e_{J_2,AD} = 1.2373$. Here the gain in efficiency using the weight function J_2 is substantially more than in the location case.

5. Further aspects

In Gregory (1980) quadratic tests are considered and he finds some interesting results for such statistics, also in terms of Bahadur efficiency. Two particularly interesting results are given. One is a result on the connection between Bahadur and Pitman efficiencies as studied by Wieand (1976). Gregory (1980) in his Theorem 3.1 gives conditions under which Wieand's condition III* holds implying equality of limiting Bahadur and limiting Pitman efficiencies. A second result of Gregory (1980) for quadratic statistics considers contiguous alternatives and gives a condition for an optimal quadratic statistic with respect to Bahadur efficiency. Although the Wasserstein goodness-of-fit statistics considered in this paper are not quadratic statistics as defined by Gregory (1980), they are asymptotically equivalent to quadratic statistics. Extension of Gregory's result to statistics that are asymptotically quadratic will lead to an alternative proof of the result in this paper as well as leading to extensions thereof.

Acknowledgements. The authors would like to acknowledge their appreciation to the late Prof Sandor Csörgő for the opportunity to discuss early ideas of this paper with him during a visit by one of the authors (TdW) to the University of Szeged. The first author's research was partially supported by research grant 108874 from the National Research Foundation.

Appendix

Proof of lemma 1. By the convexity assumption on $\log(1/f_0(x))$, it follows that J_1 is positive on $(0,1)$ and this directly implies that K_{J_1} is positive on $(0,1)^2$. Now, let g_1 denote the normalized eigenfunction of K_{J_1} corresponding to its largest eigenvalue. It is well known that for this eigenfunction (see

e.g. Tricomi, 1957)

$$\int_0^1 \int_0^1 K_{J_1}(s,t)\, g_1(s)\, g_1(t)\, dsdt = \sup_{\{g:\int g^2=1\}} \int_0^1 \int_0^1 K_{J_1}(s,t)\, g(s)\, g(t)\, dsdt. \tag{18}$$

It follows quite readily from this that g_1 is nonnegative on $(0,1)$ (but of course not always zero).

Suppose this does not hold, i.e. there are subsets of $(0,1)$ where it is strictly positive and subsets where it is strictly negative. In such a case we can write

$$g_1(t) = g_{1+}(t) - g_{1-}(t),$$

where, on the respective subsets $g_{1+}(t) > 0$, $g_{1-}(t) > 0$, while $g_{1+}(t)\, g_{1-}(t) = 0$, $t \in (0,1)$. Also, $|g_1(t)| = g_{1+}(t) + g_{1-}(t)$ and $1 = \|g_1\|^2 = \||g_1|\|^2 = \|g_{1+}\|^2 + \|g_{1-}\|^2$.

Now, because of strict positiveness on certain subsets

$$\begin{aligned}
&\int_0^1 \int_0^1 K_{J_1}(s,t)\, g_1(s)\, g_1(t)\, dsdt \\
&= \int_0^1 \int_0^1 K_{J_1}(s,t)\, g_{1+}(s)\, g_{1+}(t)\, dsdt + \int_0^1 \int_0^1 K_{J_1}(s,t)\, g_{1-}(s)\, g_{1-}(t)\, dsdt \\
&\quad - \int_0^1 \int_0^1 K_{J_1}(s,t)\, g_{1+}(s)\, g_{1-}(t)\, dsdt - \int_0^1 \int_0^1 K_{J_1}(s,t)\, g_{1-}(s)\, g_{1+}(t)\, dsdt \\
&< \int_0^1 \int_0^1 K_{J_1}(s,t)\, |g_1(s)|\, |g_1(t)|\, dsdt.
\end{aligned}$$

However, this contradicts the fact that g_1 is the eigenfunction corresponding to the largest eigenvalue and thus satisfies (18). Therefore g_1 is nonnegative on $(0,1)$. Since $(J)^{\frac{1}{2}}$ is an eigenfunction, it is either orthogonal to g_1 or equal to g_1. However, since $K_{J_1} > 0$, $(J)^{\frac{1}{2}} > 0$, it follows that

$$\int_0^1 \int_0^1 K_{J_1}(s,t)\, (J_1(s))^{\frac{1}{2}}\, g_1(t)\, dsdt > 0,$$

and thus orthogonality cannot hold. Therefore $(J)^{\frac{1}{2}} = g_1$, the eigenfunction corresponding to the largest eigenvalue.

Proof of Lemma 2. We have from (14) that

$$h_1(s) = \int_0^1 q(s,v)\, g_1(v)\, dv \Big/ \left\| \int_0^1 q(s,v)\, g_1(v)\, dv \right\|.$$

Here

$$\begin{aligned}
\int_0^1 q(s,v)\, g_1(v)\, dv &= \int_0^1 \psi(s,v)\, Q_0'(v)\, J_1(v)^{\frac{1}{2}}\, J_1(v)^{\frac{1}{2}}\, dv \\
&= -\int_0^s v Q_0'(v)\, J_1(v)\, dv + \int_s^1 (1-v)\, Q_0'(v)\, J_1(v)\, dv \\
&= I_1^{-1} \left[\int_s^1 (1-v)\, Q_0'(v)\, L_1'(Q_0(v))\, dv - \int_0^s v Q_0'(v)\, L_1'(Q_0(v))\, dv \right]
\end{aligned}$$

$$= I_1^{-1}\left[-L_1\left(Q_0\left(s\right)\right) - \int_{-\infty}^{\infty} f_0'\left(x\right) dx\right]$$
$$= I_1^{-1}\left[-L_1\left(Q_0\left(s\right)\right)\right].$$

To obtain h_1, we need the norm of the latter expression:

$$\left\|I_1^{-1}\left[-L_1\left(Q_0\left(s\right)\right)\right]\right\| = I_1^{-1}\left(\int_0^1 L_1^2\left(Q_0\left(s\right)\right) ds\right)^{\frac{1}{2}} = I_1^{-1}\left(\int_{-\infty}^{\infty} L_1^2\left(y\right) f_0\left(y\right) dy\right)^{\frac{1}{2}} = I_1^{-\frac{1}{2}},$$

since, from (2),

$$I_1 = \int_{-\infty}^{\infty} L_1'\left(y\right) f_0\left(y\right) dy = L_1\left(y\right) f_0\left(y\right)\big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} L_1\left(y\right) f_0'\left(y\right) dy = 0 + \int_{-\infty}^{\infty} L_1^2\left(y\right) f_0\left(y\right) dy,$$

using the assumption $f_0'\left(y\right) \to 0$ as $y \to \pm\infty$. This completes the proof.

Proof of Theorem 1. We first show that under the assumed conditions, (6) holds. Note that by (2.10) in DGU, clearly $n^{-1}a_n \to 0$. Denote by $T_n(\theta)$ the statistic (4) when $F_\theta(x) = F_0(x-\theta)$ is the underlying distribution function, and thus $T_n(0)$ the statistic under H_0. Let G_n^{-1} denote the empirical quantile function of a sample from the uniform distribution on $(0,1)$. Then it is well known that we can write

$$F_n^{-1}\left(t\right) \overset{D}{=} F_\theta^{-1}\left(G_n^{-1}\left(t\right)\right) = \theta + Q_0\left(G_n^{-1}\left(t\right)\right).$$

Using this, gives

$$n^{-1}\left(T_n\left(\theta\right) - a_n\right) \overset{D}{=} \int_0^1 \left[Q_0\left(G_n^{-1}\left(t\right)\right) - Q_0\left(t\right) + \theta\right]^2 w\left(t\right) dt - n^{-1}a_n$$
$$\overset{D}{=} \theta^2 \int_0^1 w\left(t\right) dt + n^{-1}T_n\left(0\right) - n^{-1}a_n + 2\theta \int_0^1 \left[Q_0\left(G_n^{-1}\left(t\right)\right) - Q_0\left(t\right)\right] w\left(t\right) dt$$
$$= \theta^2 \int_0^1 w\left(t\right) dt + n^{-1}\left(T_n\left(0\right) - a_n\right) + 2\theta R_n \text{ (say)}.$$

Clearly from (5) $n^{-1}(T_n(0) - a_n) \xrightarrow{P} 0$. Furthermore, since $|R_n|^2 \leq n^{-1}T_n(0) \cdot \int_0^1 w\left(t\right) dt$, and $n^{-1}a_n \to 0$, it follows that $R_n \xrightarrow{P} 0$, and thus

$$n^{-1}(T_n(\theta) - a_n) \xrightarrow{P} \theta^2 \int_0^1 w\left(t\right) dt.$$

(6) therefore holds.

Now, use again the well-known fact that the largest eigenvalue of a kernel is given by

$$\gamma_1^{(w)} = \sup_{\left\{g:\int g^2=1\right\}} \int_0^1 \int_0^1 K_w\left(s,t\right) g\left(s\right) g\left(t\right) dsdt$$

(see e.g. Tricomi, 1957). Thus, also

$$\gamma_1^{(w)} = \sup_{\left\{g:\int g^2=1\right\}} \int_0^1 \int_0^1 c_w\left(s,t\right) g\left(s\right) g\left(t\right) dsdt.$$

Choosing in particular the eigenfunction h_1 of c_{J_1} corresponding to $\gamma_1^{(J_1)}$, we have

$$\gamma_1^{(w)} \geq \int_0^1 \int_0^1 c_w(s,t)\, h_1(s)\, h_1(t)\, ds dt \equiv D_w \text{ (say)}.$$

Substituting for c_w from (13) and for h_1 from (15), we have

$$\begin{aligned}
D_w &= \int_0^1 \int_0^1 \left[\int_0^1 \psi(s,u)\, \psi(t,u)\, Q_0'(u)^2\, w(u)\, du \right] \cdot I_1^{-1} L_1(Q_0(s))\, L_1(Q_0(t))\, ds dt \\
&= I_1^{-1} \int_0^1 Q_0'(u)^2\, w(u) \left[\int_0^1 \psi(s,u)\, L_1(Q_0(s))\, ds \right]^2 du \\
&= I_1^{-1} \int_0^1 Q_0'(u)^2\, w(u) \left[(1-u) \int_0^u L_1(Q_0(s))\, ds - u \int_u^1 L_1(Q_0(s))\, ds \right]^2 du \\
&= I_1^{-1} \int_0^1 Q_0'(u)^2\, w(u) \left[\int_0^u L_1(Q_0(s))\, ds - u \int_0^1 L_1(Q_0(s))\, ds \right]^2 du \\
&= I_1^{-1} \int_0^1 Q_0'(u)^2\, w(u) \left[-f_0(Q_0(u)) \right]^2 du \\
&= I_1^{-1} \int_0^1 w(u)\, du.
\end{aligned}$$

From this $\gamma_1^{(w)} \geq I_1^{-1} \int_0^1 w(t) dt$, and the theorem follows.

References

Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *The Annals of Mathematical Statistics*, **38**, 303–324.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, **9**, 1196–1217.

Csörgő, S. (2000). Discussion of Del Barrio, Cuesta-Albertos and Matran. *TEST*, **9**, 54–70.

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

Csörgő, S. (2003). Weighted correlation tests for location-scale families. *Mathematical and Computer Modelling*, **38**, 753–762.

De Wet, T. (1980). Cramér-von Mises tests for independence. *Journal of Multivariate Analysis*, **10**, 38–50.

De Wet, T. (2000). Discussion of Del Barrio, Cuesta-Albertos and Matran. *TEST*, **9**, 74–79.

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

Del Barrio, E. (2007). Empirical and quantile processes in the asymptotic theory of goodness-of-fit tests. *In* Ranicki, A. (Editor) *Lectures on Empirical Processes*. European Mathematical Society, Zürich, 1–92.

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

Del Barrio, E., Cuesta-Albertos, J. A., Matran, C., and Rodriguez-Rodriguez, J. M. (1999). Tests of goodness-of-fit based on the L_2-Wasserstein distance. *The Annals of Statistics*, **27**, 1230–1239.

Del Barrio, E., Gine, E., and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, **11**, 131–189.

Grané, A. and Fortiana, J. (2008). Karhunen-Loève basis in goodness-of-fit tests decomposition: An Evaluation. *Communications in Statistics – Theory and Methods*, **37**, 3144–3163.

Gregory, G. G. (1980). On efficiency and optimality of quadratic tests. *The Annals of Statistics*, **8**, 116–131.

Krauczi, E. (2009). A study of the quantile correlation test for normality. *TEST*, **18**, 156–165.

Lockhart, R. A. (1991). Overweight tails are inefficient. *The Annals of Statistics*, **19**, 2254–2258.

McLaren, C. G. and Lockhart, R. A. (1987). On the asymptotic efficiency of certain correlation tests of fit. *The Canadian Journal of Statistics*, **15**, 159–167.

Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press, London.

Ramdas, A., Trillos, N. G., and Cuturi, M. (2017). On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, **19**, 1–15.

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

Tricomi, F. G. (1957). *Integral Equations*. Interscience Publishers, New York.

Wieand, H. S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *The Annals of Statistics*, **4**, 1003–1011.

Manuscript received 2019-12-03, revised 2020-03-04, accepted 2020-03-05.