GOODNESS-OF-FIT TESTS VIA PHI-DIVERGENCES

By Leah Jager* and Jon A. Wellner†

University of Washington and University of Washington

A unified family of goodness-of-fit tests is introduced and studied. The new family of test statistics $S_n(s)$ includes both the supremum version of the Anderson-Darling statistic and the test statistic of Berk and Jones (1979) as special cases. The new family is based on phi-divergences somewhat analogously to the phi-divergence tests for multinomial families introduced by Cressie and Read (1984), and is indexed by a real parameter $s \in \mathbb{R}$: $s = 2$ gives the Anderson - Darling test statistic, $s = 1$ gives the Berk-Jones test statistic, $s = 1/2$ gives a new (Hellinger - distance type) statistic, $s = 0$ corresponds to the "reversed Berk-Jones" statistic studied by Jager and Wellner (2004), and $s = -1$ gives a "studentized" (or empirically weighted) version of the Anderson - Darling statistic. We also introduce corresponding integral versions of the new statistics.

We show that the asymptotic null distribution theory of Jaeschke (1979) and Eicker (1979) for the Anderson-Darling statistic, and of Berk and Jones (1979) and Wellner and Koltchinskii (2003) for the Berk-Jones statistic, applies to the whole family of statistics $S_n(s)$ with $s \in [-1, 2]$. We also provide new finite-sample approximations to the null distributions and show how the new approximations can be used to obtain accurate computation of quantiles.

On the side of power behavior, we show that for $0 < s < 1$ and fixed alternatives the test statistics always converge almost surely to their corresponding natural parameter. For $1 < s < \infty$ we provide necessary and sufficient conditions on the alternative d.f. F for convergence to the corresponding natural parameter to hold, and show that the "Poisson boundary" phenomena noted by Berk and Jones for their statistic continues to hold for $s \geq 1$ and $s < 0$ by identifying the Poisson boundary distributions explicitly. We also briefly discuss further large deviation results and connections between the work of Berk and Jones (1979) and Groeneboom and Shorack (1981).

We extend the results of Donoho and Jin (2004) by showing that all our new tests for $s \in [-1, 2]$ have the same "optimal detection boundary" for normal shift mixture alternatives as Tukey's "higher-criticism" statistic and the Berk-Jones statistic.

*Supported in part by NSF grants DMS-0503822 and NSF (VIGRE Grant) DMS - 0354131

†Supported in part by NSF grants DMS-0203320 and DMS-0503822 and by NI-AID grant 2R01 AI291968-04

AMS 2000 subject classifications: Primary 62G10, 62G20; secondary 62G30

Keywords and phrases: alternatives, combining p-values, confidence bands, goodness-of-fit, Hellinger, large deviations, multiple comparisons, normalized empirical process, phi-divergence, Poisson boundaries

1. Introduction. In this paper we introduce and study a new family of goodness-of-fit tests which includes both the supremum version of the Anderson-Darling statistic (or equivalently, Tukey's "higher criticism" statistics as discussed by Donoho and Jin (2004)) and the test statistic of Berk and Jones (1979) as special cases. The new family is based on phi-divergences somewhat analogously to the phi-divergence tests for multinomial families introduced by Cressie and Read (1984), and is indexed by a real parameter $s \in \mathbb{R}$: $s = 2$ gives the Anderson - Darling test statistic, $s = 1$ gives the Berk-Jones test statistic, $s = 1/2$ gives a new (Hellinger - distance type) statistic, $s = 0$ corresponds to the "reversed Berk-Jones" statistic studied by Jager and Wellner (2004a), and $s = -1$ gives a "studentized" (or empirically weighted) version of the Anderson - Darling statistic. We introduce the corresponding integral versions of the new statistics (but will study them in detail elsewhere).

We show that the null distribution theory of the entire family of statistics can be handled exactly for sample sizes up to $n = 3000$ via Noe's recursion formulas along the lines explored for the Berk-Jones statistic by Owen (1995). We generalize the asymptotic distribution theory of Jaeschke (1979) and Eicker (1979) for the Anderson-Darling statistic and Berk and Jones (1979) and Wellner and Koltchinskii (2003) for the Berk-Jones statistic by showing that the existing null distribution theory for $s = 1$ and $s = 2$ applies to (an appropriate version of) the whole family of statistics. We also provide new finite-sample approximations to the null distributions by an argument that "backs off" from the asymptotic theory, and show how the new approximations can be used to obtain accurate computation of quantiles for small sample sizes.

We generalize the results of Owen (1995) by showing that our family of test statistics provides a corresponding family of confidence bands.

On the side of power behavior, we show that for $0 < s < 1$ and fixed alternatives the test statistics always converge almost surely to their corresponding natural parameter. For $1 < s < \infty$ we provide necessary and sufficient conditions on the alternative d.f. F for convergence to the corresponding natural parameter to hold, and show that the "Poisson boundary" phenomena noted by Berk and Jones for their statistic continues to hold for $1 \le s < \infty$ and for $s < 0$ by identifying the Poisson boundary distributions explicitly. We also briefly discuss further large deviation results and connections between the work of Berk and Jones (1979) and Groeneboom and Shorack (1981).

We extend the results of Donoho and Jin (2004) by showing that all our new tests for $s \in [-1, 2]$ have the same "optimal detection boundary" for normal shift mixture alternatives as Tukey's "higher-criticism" statistic and the Berk-Jones statistic.

2. The test statistics. Consider the classical goodness-of-fit problem: suppose that $X_1, \ldots, X_n$ are i.i.d. F, and let $\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^{n} 1\{X_i \le x\}$ be the empirical distribution function of the sample. We want to test

$$H: \ F = F_0 \qquad \text{versus} \qquad K: \ F \neq F_0,$$

where F_0 is continuous. By the probability integral transformation, we can, without loss of generality, suppose that F_0 is the uniform distribution on $[0,1]$, $F_0(x) = (x \wedge 1) \vee 0$, and that all the distribution functions F in the alternative K are defined on $[0,1]$. The basic idea behind our new family of tests is simple. For fixed $x \in (0,1)$, the interval is divided into two sub-intervals $[0,x]$ and $(x,1]$, and we can test the (pointwise) null hypothesis $H_x: \ F(x) = x$ versus the (pointwise) alternative $K_x: \ F(x) \neq x$ using any of the general phi-divergence test statistics $K_\phi(\mathbb{F}_n(x), x)$ proposed by Csiszár (1963) (see also Csiszár (1967) and Ali and Silvey (1966)) and studied further in a multinomial context by Cressie and Read (1984), where ϕ is a convex function mapping $[0,\infty)$ to the extended reals $\mathbb{R} \cup \{\infty\}$ (cf. Liese and Vajda (1987), pages 10 and 212, Vajda (1989)). Then our proposed test statistics are of the form

$$S_n(\phi) \equiv \sup_x K_\phi(\mathbb{F}_n(x), x) \qquad \text{or} \qquad T_n(\phi) \equiv \int_0^1 K_\phi(\mathbb{F}_n(x), x)dx$$

where the supremum and/or integral over x may require some restriction depending on the choice of ϕ.

In our particular case , we define $\phi = \phi_s$ for $s \in \mathbb{R}$ by

$$\phi_s(x) \equiv \begin{cases} [1 - s + sx - x^s]/[s(1-s)], & s \neq 0, 1 \\ x(\log x - 1) + 1 \equiv h(x), & s = 1 \\ \log(1/x) + x - 1 \equiv \tilde{h}(x), & s = 0 \end{cases}$$

(cf. Liese and Vajda (1987), page 34), so that

$$\begin{aligned} K_s(u,v) &= v\phi_s(u/v) + (1-v)\phi_s((1-u)/(1-v)) \\ &= \frac{1}{s(1-s)}\{1 - u^s v^{1-s} - (1-u)^s(1-v)^{1-s}\}, \qquad s \neq 0, 1. \end{aligned}$$

Note that this definition makes ϕ_s continuous in s for all x in $(0,1)$, and hence K_s is continuous in s for all $(u,v) \in (0,1)^2$. Also note that $K_{\lambda+1}(p,q) = I_2^\lambda(\underline{p} : \underline{q})$ where $\underline{p} = (p, 1-p)$, $\underline{q} = (q, 1-q)$, and $I_2^\lambda(\underline{p} : \underline{q})$ is as defined in (5.1), Cressie and Read (1984), page 456. Then our proposed test statistics $S_n(s)$ and $T_n(s)$ for $s \in \mathbb{R}$, are defined by

$$S_n(s) \equiv \begin{cases} \sup_{0<x<1} K_s(\mathbb{F}_n(x), x), & \text{if } \ s \geq 1, \\ \sup_{X_{(1)} \leq x < X_{(n)}} K_s(\mathbb{F}_n(x), x), & \text{if } \ s < 1, \end{cases}$$

and

$$T_n(s) \equiv \begin{cases} \int_0^1 K_s(\mathbb{F}_n(x), x)dx, & \text{if } \ s > 0, \\ \\ \int_{X_{(1)}}^{X_{(n)}} K_s(\mathbb{F}_n(x), x)dx, & \text{if } \ s \leq 0. \end{cases}$$

The reasons for changing the definitions of the statistics by restricting the supremum or integral for different values of s will be explained in the remarks below and in section 3; basically the restrictions must be imposed for some appropriate value of s in order to maintain the same null distribution theory for all values of s in $[-1,2]$.

The most notable special cases of these statistics are $s \in \{-1,0,1/2,1,2\}$: it is easily checked that

$$\begin{aligned}
K_2(u,v) &= \frac{1}{2}\frac{(u-v)^2}{v(1-v)},\\
K_1(u,v) &= u\log\left(\frac{u}{v}\right)+(1-u)\log\left(\frac{1-u}{1-v}\right),\\
K_{1/2}(u,v) &= 4\{1-\sqrt{uv}-\sqrt{(1-u)(1-v)}\}\\
&= 2\{(\sqrt{u}-\sqrt{v})^2+(\sqrt{1-u}-\sqrt{1-v})^2\},\\
K_0(u,v) &= K_1(v,u) = v\log\left(\frac{v}{u}\right)+(1-v)\log\left(\frac{1-v}{1-u}\right),\\
K_{-1}(u,v) &= K_2(v,u) = \frac{1}{2}\frac{(u-v)^2}{u(1-u)}.
\end{aligned}$$

It follows that:

(a) $S_n(2)$ is (1/2 times) the square of the supremum form of the Anderson-Darling statistic (or, in its one-sided form, Tukey's "higher criticism statistic", see Donoho and Jin (2004) and section 5);
(b) $S_n(1)$ is the statistic studied by Berk and Jones (1979);
(c) $S_n(1/2)$ is (4 times) the supremum of the pointwise Hellinger divergences between Bernoulli($\mathbb{F}_n(x)$) and Bernoulli($F_0(x)$); as far as we know, this is a new goodness-of-fit statistic (as are all the statistics $S_n(s)$ for $s \notin \{-1,0,1,2\}$).
(d) $S_n(0)$ is the "reversed Berk - Jones" statistic introduced by Jager and Wellner (2004a);
(e) $S_n(-1)$ is (1/2 times) a "studentized" version of the supremum form of the Anderson-Darling statistic; see e.g. Eicker (1979), page 116.
(f) $T_n(1)$ is the integral form of the Berk-Jones statistic introduced by Einmahl and McKeague (2003).
(g) $T_n(2)$ is the classical (integral form of) the Anderson-Darling statistic introduced by Anderson and Darling (1952).

Remark 2.1. For each $r \in \mathbb{R}$, define $D_r : (0,1)^2 \to \mathbb{R}^+$ by $D_r(u,v) = K_{r+1/2}(u,v)$. Then $D_{-r}(u,v) = D_r(v,u)$, so the families of statistics $S_n(s)$ and $T_n(s)$ have a natural symmetry about $s = 1/2$. We will continue to use the "$s-$parametrization" of these families for reasons of notational simplicity.
Remark 2.2. The symmetry noted in the previous remark provides one rationale for

changing the region to which the supremum is restricted at $s = 1/2$ rather than at $s = 1$, but this seems to result in a "bigger discontinuity" in the null distribution theory; for further discussion see section 3.

Remark 2.3. Note that in the case of item (a) in the above list, Anderson and Darling (1952) only allow the supremum to be taken over a fixed subset $[a, b]$ of $(0, 1)$ with $0 < a \leq b < 1$; see the paragraph above their example 1 on page 207 and their example 2, pages 210-211. This also seems to be the case in Kendall and Kendall (1980) who arrived at this statistic in connection with some problems in geometrical probability.

3. Distributions under the null hypothesis.

3.1. *Finite sample critical points via Noé's recursion.* Owen (1995) showed how to use the recursions of Noé (1972) to obtain finite sample critical points of the Berk-Jones statistic $R_n = S_n(1)$ for values of n up to 1000. (Also see Shorack and Wellner (1986), pages 362-366 for an exposition of Noé's methods.) Jager and Wellner (2004a) pointed out a minor error in the derivations of Owen (1995) and extended his results to the reversed Berk-Jones statistic $S_n(0)$. Jager (2006) gives exact finite sample computations for the whole family of statistics via Noé's recursions for values of n up to 3000. (The C- and R- programs are available at the second author's web-site.) We will not give details of the finite-sample computations here but refer the interested reader to Jager and Wellner (2004a) and Jager (2006). See subsection 3.3 for plots of finite sample critical points together with several asymptotic approximations.

3.2. *Asymptotic distribution theory for $S_n(s)$ under the null hypothesis.* Limit distribution theory for $S_n(2)$ and $S_n(-1)$ under the null hypothesis follows from the work of Jaeschke (1979) and Eicker (1979); see Shorack and Wellner (1986), chapter 16, pages 597 - 615 for an exposition. These results are closely related to the classical results of Darling and Erdös (1956). Berk and Jones (1979) stated the asymptotic distribution of their statistic $R_n = S_n(1)$; for details of the proof see Wellner and Koltchinskii (2003). Here we show that the limit distribution of $nS_n(s) - r_n$ is the same double-exponential extreme value distribution for all $-1 \leq s \leq 2$ where

$$r_n = \log_2 n + \frac{1}{2} \log_3 n - \frac{1}{2} \log(4\pi)$$

with $\log_2 n \equiv \log(\log n)$ and $\log_3 n \equiv \log(\log_2 n)$.

Theorem 3.1 *(Limit distribution under null hypothesis.) Suppose that the null hypothesis H holds so that F is the uniform distribution on $[0, 1]$. Then for $-1 \leq s \leq 2$ it follows that*

$$nS_n(s) - r_n \to_d Y_4 \sim E_v^4$$

where $E_v^4(x) = \exp(-4\exp(-x)) = P(Y_4 \leq x)$.

Define

$$b_n = \sqrt{2\log_2 n}, \qquad c_n = 2\log_2 n + \frac{1}{2}\log_3 n - \frac{1}{2}\log(4\pi),$$
$$d_n = n^{-1}(\log n)^5, \qquad Z_n \equiv \sup_{d_n \le x \le 1-d_n} \frac{\sqrt{n}|\mathbb{F}_n(x) - x|}{\sqrt{x(1-x)}},$$

As will be seen, the proof involves the following four facts: *Fact 1.* $Z_n/b_n \to_p 1$. *Fact 2.* $b_n Z_n - c_n \to_d Y_4 \sim E_v^4$. *Fact 3.* $(1/2)c_n^2/b_n^2 = r_n + o(1)$. *Fact 4.* $nS_n(s) \doteq (1/2)Z_n^2$ (where $\doteq$ means that that the quantity on the right side determines the limiting distribution of the random variables on the left side).

In the ranges $s > 2$ and $s < -1$ we do not know a theorem describing the behavior of the statistics $S_n(s)$ under the null hypothesis.

3.3. *Using the asymptotic distribution theory to approximate critical points,* $s \in [-1, 2]$. Wellner and Koltchinskii (2003) noted (for the case $s = 1$) that in view of fact 3 a slightly better centering of $nS_n(s)$ might involve $c_n^2/(2b_n^2)$. Thus to approximate the upper $1-\alpha$ quantile $q_n(s,\alpha)$ of $nS_n(s)$ defined by

$$P_{unif}(nS_n(s) \le q_n(s,\alpha)) = 1 - \alpha,$$

the limiting distribution of theorem 3.1 suggests two possible approximations, namely

$$q_n^{(1)}(s,\alpha) = y_{4,\alpha} + r_n, \qquad q_n^{(2)}(s,\alpha) = y_{4,\alpha} + c_n^2/(2b_n^2)$$

where $y_{4,\alpha} = -\log[(1/4)\log(1/(1-\alpha))]$ satisfies $P(Y_4 \le y_{4,\alpha}) = 1 - \alpha$.

But because the convergence in fact 1 of the previous subsection is rather slow (note that the interval $[d_n, 1-d_n]$ is non-empty only for $n > 1010388$), we consider the following "backing-off the asymptotics" argument to obtain a third approximation: from facts 2 and 4 above, we deduce that the approximate distribution of Z_n is that of $(Y_4 + c_n)/b_n$ and hence the approximate distribution of $nS_n(s)$ is that of

$$\frac{1}{2}\left(\frac{Y_4 + c_n}{b_n}\right)^2 = \frac{1}{2b_n^2}Y_4^2 + \frac{c_n}{b_n^2}Y_4 + \frac{1}{2}\frac{c_n^2}{b_n^2}.$$

Equivalently, the approximate distribution of $nS_n(s) - (1/2)c_n^2/b_n^2$ is that of

$$\frac{1}{2}\left(\frac{Y_4 + c_n}{b_n}\right)^2 - \frac{1}{2}\frac{c_n^2}{b_n^2} = \frac{1}{2b_n^2}Y_4^2 + \frac{c_n}{b_n^2}Y_4.$$

Thus we propose approximating

$$P_{unif}(nS_n(s) - (1/2)c_n^2/b_n^2 \le x) \qquad \text{by} \qquad P\left(\frac{1}{2b_n^2}Y_4^2 + \frac{c_n}{b_n^2}Y_4 \le x\right).$$

Note that since $b_n^2 \to \infty$ (slowly!) and $c_n/b_n^2 \to 1$, the probability on the right side in the last display converges to $P(Y_4 \leq x)$. Furthermore, by inverting the quadratic form,

$$P\left(\frac{1}{2b_n^2}Y_4^2 + \frac{c_n}{b_n^2}Y_4 \leq x\right) = P(l_n(x) \leq Y_4 \leq u_n(x)) = e^{-4\exp(-u_n(x))} - e^{-4\exp(-l_n(x))}$$

where $u_n(x)$ and $l_n(x)$ are given by

$$u_n(x) = -c_n + c_n\sqrt{1 + 2b_n^2 x/c_n^2}, \qquad l_n(x) = -c_n - c_n\sqrt{1 + 2b_n^2 x/c_n^2}.$$

Thus our third approximation to the quantiles $q_n(s, \alpha)$ of $S_n(s)$ is

$$q_n^{(3)}(s, \alpha) = x_{\alpha,n} + c_n^2/(2b_n^2),$$

where $x_{\alpha,n}$ satisfies

$$1 - \alpha = P(l_n(x_{\alpha,n}) \leq Y_4 \leq u_n(x_{\alpha,n})) = e^{-4\exp(-u_n(x_{\alpha,n}))} - e^{-4\exp(-l_n(x_{\alpha,n}))}.$$

Note that none of the three approximations depend on s. Interestingly, it happens that

$$q_n^{(1)}(s, \alpha) < q_n^{(2)}(s, \alpha) < q_n(s, \alpha) < q_n^{(3)}(s, \alpha)$$

for $s \in [0, 1)$ and wide ranges of n and α, so we can develop empirical formulas as a function of s, n, and α to give accurate approximations to $q_n(s, \alpha)$ for sample sizes $10 \leq n \leq 3000$. See Figure 1 for $\alpha = .05$ and $s \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. The resulting approximation formula for $q_n(s, \alpha)$ is

$$q_n^{(4)}(s, \alpha) = (1 - \lambda_n(s, \alpha))q_n^{(2)}(s, \alpha) + \lambda_n(s, \alpha)q_n^{(3)}(s, \alpha)$$

where

$$\begin{aligned} \lambda_n(s, \alpha) &= 1 - \exp(-f_n(s, \alpha)), \\ f_n(s, \alpha) &= a_0(s, \alpha) + a_1(s, \alpha)\log_2 n + a_2(s, \alpha)(\log_2 n)^2. \end{aligned}$$

Values of the coefficients $a_j(s, .05)$ are given for three different ranges of n in Table 3.3. For example,

$$f_n(.5, .05) = \begin{cases} -0.8004235 + 0.8371563\log_2 n - 0.1358954(\log_2 n)^2, & 10 < n \leq 100 \\ -0.9739494 + 1.098979\log_2 n - 0.2332922(\log_2 n)^2, & 100 < n \leq 1000 \\ -0.6745409 + 0.7843495\log_2 n - 0.15006064(\log_2 n)^2, & 1000 < n \leq 3000. \end{cases}$$

TABLE 1
Coefficients for $\alpha = .05$, $10 < n \leq 3000$ $s = 0.0(0.1)0.9$

s	$a_j(s, .05)$	$10 < n \leq 10^2$	$10^2 < n \leq 10^3$	$10^3 < n \leq 3 \times 10^3$
0.0	$a_0(.0, .05)$	-0.7480087	-6.316381	-6.310909
	$a_1(.0, .05)$	0.3328438	7.573198	7.648025
	$a_2(.0, .05)$	0.6203775	-1.739940	-1.780574
0.1	$a_0(.1, .05)$	-1.051773	-3.549094	-2.457953
	$a_1(.1, .05)$	1.060974	4.454303	3.326228
	$a_2(.1, .05)$	0.1030335	-1.051539	-0.75991
0.2	$a_0(.2, .05)$	-1.056712	-2.307136	-1.446272
	$a_1(.2, .05)$	1.165262	2.914192	2.015244
	$a_2(.2, .05)$	-0.07139836	-0.6824301	-0.4476649
0.3	$a_0(.3, .05)$	-0.9847481	-1.644847	-1.024914
	$a_1(.3, .05)$	1.092687	2.038614	1.388977
	$a_2(.3, .05)$	-0.1299329	-0.4674763	-0.2972222
0.4	$a_0(.4, .05)$	-0.8930882	-1.242862	-0.8071487
	$a_1(.4, .05)$	0.9681655	1.481478	1.024084
	$a_2(.4, .05)$	-0.1421381	-0.328973	-0.2088906
0.5	$a_0(.5, .05)$	-0.8004235	-0.9739494	-0.6745409
	$a_1(.5, .05)$	0.8371563	1.098979	0.7843495
	$a_2(.5, .05)$	-0.1358954	-0.2332922	-0.15006064
0.6	$a_0(.6, .05)$	-0.7107262	-0.7766191	-0.5794551
	$a_1(.6, .05)$	0.7149534	0.8204617	0.6131262
	$a_2(.6, .05)$	-0.1226809	-0.1636567	-0.1091296
0.7	$a_0(.7, .05)$	-0.618766	-0.6186249	-0.4990619
	$a_1(.7, .05)$	0.6029367	0.6096428	0.4838249
	$a_2(.7, .05)$	-0.1071244	-0.1115457	-0.07843385
0.8	$a_0(.8, .05)$	-0.5353572	-0.4758341	-0.4084025
	$a_1(.8, .05)$	0.5241284	0.442001	0.3732231
	$a_2(.8, .05)$	-0.09916116	-0.07218545	-0.05350028
0.9	$a_0(.9, .05)$	-0.4107134	-0.3459936	0.3291721
	$a_1(.9, .05)$	0.4398488	0.3393939	-0.3915376
	$a_2(.9, .05)$	-0.09080127	-0.05263163	0.1449166

For further tables with more values of s and α, see Jager (2006).

Figure 1 gives a plot of our approximate quantiles and the exact finite sample quantiles for $10 \leq n \leq 3000$ and $s \in \{j/10 : \ j = 0, 1, \ldots, 9\}$. Figure 2 gives the resulting residual plots for $s = 0, .2, .5,$ and $.8$.

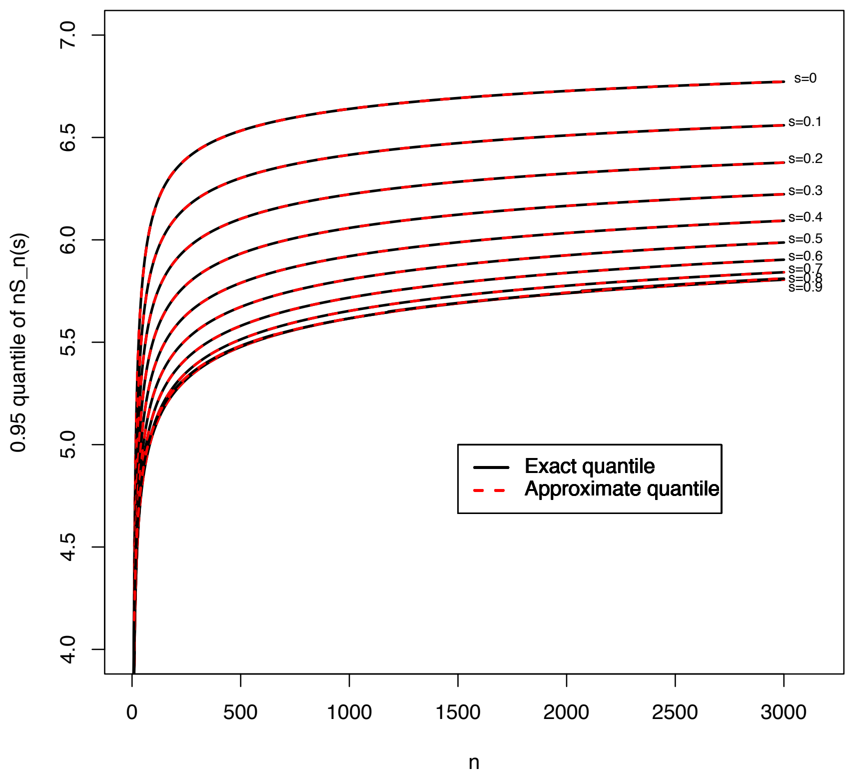

Fig 1. *Exact and approximate quantiles of* $nS_n(s)$, $10 \leq n \leq 3000$

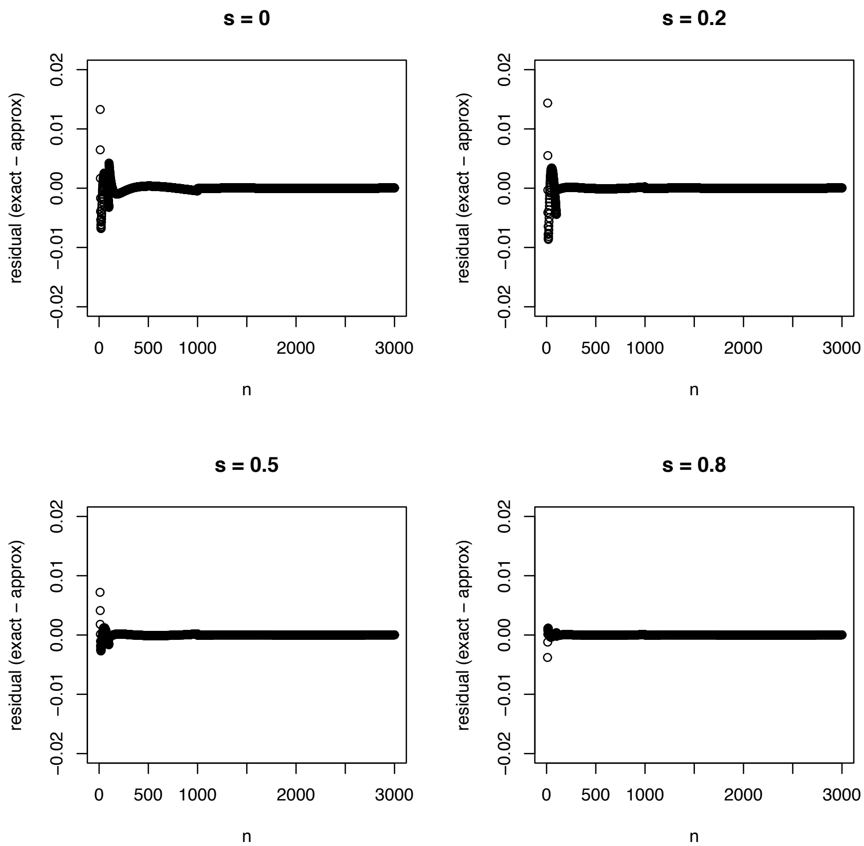

FIG 2. *Exact quantiles minus approximate quantiles of* $nS_n(s)$, $10 \le n \le 3000$

3.4. *Confidence bands.* Owen (1995) showed how the Berk-Jones statist $R_n = S_n(1)$ can be inverted to obtain confidence bands for an unknown distribution function F. Similarly, the family of statistics $S_n(s)$ yields a new family of confidence bands for F as follows: given a continuous d.f. F on $\mathbb{R}$, define

$$S_n(s,F) \equiv \begin{cases} \sup_{-\infty<x<\infty} K_s(\mathbb{F}_n(x), F(x)), & \text{if } \ s \ge 1, \\ \sup_{X_{(1)} \le x < X_{(n)}} K_s(\mathbb{F}_n(x), F(x), & \text{if } \ s < 1. \end{cases}$$

Then for each fixed $\alpha \in (0,1)$ and n,

$$P_F(S_n(s,F) \le q_n(s,\alpha)) = P_{F_0}(S_n(s) \le q_n(s,\alpha)) = 1-\alpha.$$

Hence

$$\{F: \ S_n(s,F) \le q_n(s,\alpha)\} = \{F: \ L_n(x;s,\alpha) \le F(x) \le U_n(s;s,\alpha) \ \text{ for all } \ x \in \mathbb{R}\}$$

yields a family of $1-\alpha$ confidence bands for F. Here $L_n(x;s,\alpha)$ and $U_n(x;s,\alpha)$ are random functions determined by s, α, n and the data in a straightforward way; see Owen (1995), Jager and Wellner (2004a), and Jager (2006) for details.

3.5. *Asymptotic distribution theory for $T_n(s)$ under the null hypothesis.* Limit distribution theory for $T_n(2)$ was established by Anderson and Darling (1952). Einmahl and McKeague (2003) noted that this carries over to $T_n(1)$ (for a proof, see Wellner and Koltchinskii (2003)) and extended $T_n(1)$ to other testing problems. Here we show that the limit distribution of $nT_n(s)$ is ((1/2) times) the Anderson-Darling limit distribution for all $s \in [-1, 2]$, namely the distribution of

$$A^2 \equiv \int_0^1 \frac{[\mathbb{U}(t)]^2}{t(1-t)} dt \stackrel{d}{=} \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)} \tag{1}$$

where $\mathbb{U}$ is a standard Brownian bridge process on $[0, 1]$ and $Z_1, Z_2, \ldots$ are i.i.d. $N(0, 1)$; see e.g. Shorack and Wellner (1986), pages 224-227.

Theorem 3.2 *(Limit distribution of $T_n(s)$ under the null hypothesis.) Suppose that the null hypothesis H holds so that F is the uniform distribution on $[0, 1]$. Then for $-\infty < s \leq 2$ it follows that $nT_n(s) \rightarrow_d A^2/2$ where A^2 is the Anderson-Darling limit defined in (1).*

We will not study the statistics $T_n(s)$ further in this paper, but intend to continue their study elsewhere.

4. Limit theory under alternatives. The main reason for interest in our new class of statistics is the excellent power behavior of individual members of the class which have previously been studied separately and somewhat in isolation from each other: see e.g. Berk and Jones (1979) (for the Berk-Jones statistic), Durbin, Knott and Taylor (1975) and D'Agostino and Stephens (1986) for $T_n(2)$ compared to other integral goodness-of-fit statistics, and Nikitin (1995) for treatment of Bahadur efficiencies for many goodness-of-fit statistics. This interest has received new impetus via the use of appropriate one-sided versions of the test statistics $S_n(2)$ in the context of multiple testing problems; see e.g. Donoho and Jin (2004), Jin (2004), and Meinshausen and Rice (2006). See Cayón, Jin and Treaster (2005) for an interesting application to detection of non-Gaussianity in the cosmic microwave background data gathered by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite, and see Cai, Jin and Low (2005) for further work on estimation aspects of the problem in connection with the developments in Meinshausen and Rice (2006).

Here we study convergence of the family of statistics to their "natural parameters" under fixed alternatives, comment briefly on the Bahadur efficiency results of Berk and Jones (1979) in light of the results of Groeneboom and Shorack (1981), and show that the optimal detection boundary results of Donoho and Jin (2004) extend to the whole family of statistics $S_n(s)$ for $s \in [-1, 2]$. In spite of the negative results of Janssen (2000) for goodness-of-fit statistics in general, much remains to be learned about the power behavior of the family $\{S_n(s)\}$.

4.1. *Almost sure convergence to natural parameter.* Let F_0 be the Uniform$(0,1)$ distribution function as in section 2. The Kolmogorov statistic $D_n \equiv \|\mathbb{F}_n - F_0\|_\infty$ has the property that for *any* distribution function F in K, if $X_1, \ldots, X_n$ are i.i.d. F, then

$$D_n \rightarrow_{a.s.} \|F - F_0\|_\infty \equiv d(F).$$

We call $d(F) = \|F - F_0\|_\infty$ the *natural parameter* for the Kolmogorov statistic D_n. As Berk and Jones (1979) pointed out for their statistic $R_n = S_n(1)$, under alternatives $F \in K$ the convergence

$$S_n(1) = R_n = \sup_{0<x<1} K_1(\mathbb{F}_n(x), x) \rightarrow_{a.s.} \sup_{0<x<1} K_1(F(x), x) \equiv r(F)$$

holds only under some condition on F (the exact condition will be given below), and for a slightly more extreme F, namely what we call the "Poisson boundary distribution function", the behavior changes to convergence in distribution to a functional of a Poisson process rather than convergence to a natural parameter. Thus Berk and Jones (1979) showed that if $F(x) = 1/(1 + \log(1/x))$, then

$$S_n(1) = R_n \rightarrow_d \sup_{t>0} \frac{\mathbb{N}(t)}{t} \stackrel{d}{=} \frac{1}{U} \tag{2}$$

where $\mathbb{N}$ is a standard Poisson process and $U \sim$ Uniform$(0,1)$.

It turns out that in the range $0 < s < 1$ the statistics $S_n(s)$ behave analogously to the Kolmogorov statistic D_n under fixed alternatives. Namely we show that in this range the statistics converge almost surely to their "natural parameter" for all d.f.'s $F \in K$.

Proposition 4.1 *Suppose that $X_1, \ldots, X_n$ are i.i.d. $F \in K$ and that $0 < s < 1$. Then $S_n(s) \rightarrow_{a.s.} \sup_{0<x<1} K_s(F(x), x) \equiv S_\infty(s, F)$.*

On the other hand, in the range $s > 1$ we have the following criterion for almost sure convergence of the statistics $S_n(s)$ to their natural parameters:

Proposition 4.2 *Suppose that $X_1, \ldots, X_n$ are i.i.d. $F \in K$ and that $s > 1$. Then $S_n(s) \rightarrow_{a.s.} \sup_{0<x<1} K_s(F(x), x) \equiv S_\infty(s, F)$ if and only if F satisfies*

$$\int_0^1 \frac{1}{(F^{-1}(u)(1 - F^{-1}(u)))^{(s-1)/s}} du < \infty,$$

By the (inverse) probability integral transformation, the convergence in the last display is equivalent to $E_F[X(1-X)]^{-(1-1/s)} < \infty$.

As Berk and Jones (1979) show, if for some $\gamma > 0$, the distribution function F satisfies

$$\begin{aligned} F(x) &\le \left\{\log(1/x)(\log_2(1/x))^{1+\gamma}\right\}^{-1}, && x \le \gamma, \quad \text{and} \\ 1 - F(x) &\le \left\{\log(1/(1-x))(\log_2(1/(1-x)))^{1+\gamma}\right\}^{-1}, && x \ge 1-\gamma, \end{aligned}$$

then $R_n \equiv S_n(1) \to_{a.s.} \sup_{0<x<1} K_1(F(x), x) \equiv S_\infty(1, F) \equiv r(F) < \infty$. It can be shown that this convergence holds if and only if $\int_0^1 [x(1-x)]^{-1} F(x)(1-F(x))dx < \infty$; see Jager (2006) for details. We do not yet know sharp conditions for $S_n(s) \to_{a.s.} S_\infty(s, F)$ under $F \in K$ when $s \le 0$.

4.2. *Poisson boundaries for $s \ge 1$ and $s < 0$.* As noted in the previous subsection, the statistic $R_n = S_n(1)$ has a "Poisson boundary" d.f. $F_1 \in K$ for which $R_n = S_n(1) \to_d 1/U$ rather than $R_n = S_n(1) \to_{a.s.} r(F) \equiv S_\infty(1, F)$. Here we note that this behavior persists for the entire range $s \ge 1$ and for $s < 0$.

For each fixed $s \in [0, 1)^c$, define the distribution function F_s on $[0, 1]$ by $F_s(0) = 0$ and, for $0 < x \le 1$, by

$$F_s(x) = \begin{cases} \left(1 + \frac{x^{1-s}-1}{s-1}\right)^{-1/s}, & 1 < s < \infty, \\ (1 + \log(1/x))^{-1}, & s = 1, \\ \left(1 - s(x^{s-1} - 1)\right)^{1/s}, & s < 0. \end{cases} \tag{3}$$

Note that $F_s(x) \to F_1(x)$ as $s \searrow 1$ for $0 \le x \le 1$.

The following proposition includes the result (2) of Berk and Jones (1979) when $s = 1$, and it agrees with the case $b = 1/2$ of Theorem 2 of Jager and Wellner (2004b) when $s = 2$.

Proposition 4.3 *(Poisson boundaries for $s \ge 1$ and $s < 0$).*
(i) Fix $s \ge 1$ and suppose that $X_1, \ldots, X_n$ are i.i.d. F_s given in (3). Then

$$S_n(s) \to_d \frac{1}{s}\left(\sup_{t>0} \frac{\mathbb{N}(t)}{t}\right)^s \stackrel{d}{=} \frac{1}{sU^s}.$$

(ii) Fix $s < 0$ and suppose that $X_1, \ldots, X_n$ are i.i.d. F_s given in (3). Then

$$S_n(s) \to_d \frac{1}{(1-s)}\left(\sup_{t \ge S_1} \frac{t}{\mathbb{N}(t)}\right)^{-s} \tag{4}$$

where $S_1 = E_1$ is the first jump point of $\mathbb{N}$*.*

Remark 4.1. The distribution of $\sup_{t\ge S_1}(t/\mathbb{N}(t))$, which is also the limiting distribution of $\sup\{(t/\mathbb{G}_n(t)) : t \ge \xi_{(1)}\}$ where $\mathbb{G}_n$ is the empirical distribution function of n i.i.d. Uniform$(0,1)$ random variables $\xi_1, \ldots, \xi_n$, is given by

$$P(\sup_{t\ge S_1}(t/\mathbb{N}(t)) > x) = \exp(-x) + \sum_{k=1}^{\infty} \frac{(k-1)^{k-1}}{k!} x^k \exp(-kx), \qquad x > 1;$$

see Wellner (1978), pages 1008-1009 and Shorack and Wellner (1986), page 412. This yields an explicit formula for the distribution of the random variable on the right side of (4).

Remark 4.2. Although the family of distributions F_s satisfies $F_s(x) \to \exp(-(1/x - 1)) \equiv F_0(x)$ as $s \nearrow 0$, it appears that the natural limit in distribution under F_0 is $S_n(0) \to_d 1 = \sup_{0<x<1} K_0(F_0(x), x)$ in this case, so apparently convergence to the natural parameter continues to hold under F_0. We do not know if there is a (more extreme) d.f. $\tilde{F}_0$ for which $S_n(0) \to_d g(\mathbb{N})$ for a non-degenerate functional g of a standard Poisson process $\mathbb{N}$. On the other hand it is clear that the natural parameter $\sup_{0<x<1} K_0(F(x), x)$ is infinite for many alternative d.f.'s "more extreme" than $\widetilde{F}_0(x) = \exp(-(1/x - 1))$; for example, take $F(x) = \exp(-(x^{-r} - 1))$ with $r > 1$.

Remark 4.3. If $F \in K$ has Poisson boundary behavior at both 0 and 1, then natural generalizations of Proposition 4.3 involving two independent Poisson processes can easily be proved. For example, if F is the standard arcsin law with density $\pi^{-1}u^{-1/2}(1-u)^{-1/2}1_{(0,1)}(u)$, then

$$S_n(2) \to_d \frac{2}{\pi^2} \max\left\{ \left(\sup_{t>0} \frac{\mathbb{N}(t)}{t}\right)^2, \left(\sup_{t>0} \frac{\widetilde{\mathbb{N}}(t)}{t}\right)^2 \right\}$$

where $\mathbb{N}$, $\widetilde{\mathbb{N}}$ are independent standard Poisson processes.

4.3. *Bahadur efficiency comparisons.* Berk and Jones (1979) studied the Bahadur efficiency of their statistic $S_n(1) = R_n$ relative to weighted Kolmogorov statistics based on the work of Abrahamson (1967). As pointed out by Groeneboom and Shorack (1981) however, the Bahadur efficacies of the weighted Kolmogorov statistics are 0 for weights heavier than the (quite light) logarithmic weight function $\psi(x) = -\log(x(1-x))$ because the null distribution large-deviation result is degenerate for heavier weights. A second difficulty for Bahadur efficiency comparisons is that both the weighted Kolmogorov statistics and the Berk-Jones statistic fail to converge almost surely to their natural parameters for sufficiently extreme alternative d.f.'s F, and as noted by Berk and Jones (1979) for the Berk-Jones statistic and by Jager and Wellner (2004b) for the weighted Kolmogorov statistics, there is a certain "Poisson boundary" d.f. F for which the statistics converge in distribution to a functional of a Poisson process. Thus comparisons of goodness-of-fit statistics of the supremum type via Bahadur efficiency is rendered difficult by breakdowns in both the large-deviation theory under the null hypothesis and by failure of the statistics to converge almost surely under fixed alternatives. Nevertheless, it would be interesting to be able to make comparisons where possible.

To this end we consider variants of our statistics in the range $0 < s < 1$ with the supremum unrestricted as follows:

$$S_n^{ur}(s) = \sup_{0<x<1} K_s(\mathbb{F}_n(x), x),$$
$$S_n^{ur,+}(s) = \sup_{0<x<1} K_s^+(\mathbb{F}_n(x), x), \qquad S_n^{ur,-}(s) = \sup_{0<x<1} K_s^-(\mathbb{F}_n(x), x),$$

where

$$K_s^+(u,v) = \begin{cases} K_s(u,v) & \text{if } 0 < v < u < 1, \\ 0 & \text{if } 0 \le u \le v \le 1, \\ \infty & \text{otherwise.} \end{cases}$$

Also set $K(u,v) = K_1(u,v) = u\log(u/v) + (1-u\log((1-u)/(1-v))$ for $(u,v) \in (0,1)^2$, and $K^+(u,v) = K_1^+(u,v)$. Although we do not yet have large deviation results for the statistics $S_n(s)$ or $S_n^+(s) \equiv \sup_{X_{(1)} \le x < X_{(n)}} K_s^+(\mathbb{F}_n(x), x)$, we can establish the following large deviation results for $S_n^{ur,+}(s)$ and $S_n^{ur,-}(s)$.

Theorem 4.4 *Suppose that $X_1, \ldots, X_n$ are i.i.d. with continuous d.f. F_0, the Uniform distribution on $(0,1)$. Fix $s \in (0,1)$. Then:*

$$\begin{aligned} n^{-1}\log P_0(S_n^{ur,+}(s) \ge a) &\to -\inf_{0<x<1} K^+(\tau_s^+(x,a), x) \\ &= -\frac{\log[1 - s(1-s)a]}{1-s} \equiv -g_s^+(a) \end{aligned} \tag{5}$$

for each $0 \le a < 1/[s(1-s)]$ where $\tau_s^+(x,a) = \inf\{t : \ K_s^+(t,x) \ge a\}$. Furthermore

$$\begin{aligned} n^{-1}\log P_0(S_n^{ur,-}(s) \ge a) &\to -\inf_{0<x<1} K^-(\tau_s^-(x,a), x) \\ &= -\frac{\log[1 - s(1-s)a]}{1-s} \equiv -g_s^-(a) \end{aligned}$$

for each $a \ge 0$ where $\tau_s^-(x,a) = \sup\{t : \ K_s^-(t,x) \ge a\}$.

Combining Theorem 4.4 with proposition 4.1, we have the following corollary for the Bahadur efficacies of the statistics $S_n^{ur,+}(s)$ and $S_n^{ur,-}(s)$ with $0 < s < 1$:

Corollary 4.5 *Let F be a continuous distribution function on $[0,1]$. Then the Bahadur efficacy of $S_n^{ur,+}(s)$ at the alternative F is*

$$\epsilon_s^{\pm}(F) = g_s^{\pm}(S_\infty^{\pm}(s,F)) = g_s^{+}(S_\infty^{\pm}(s,F))$$

where g_s^+ is defined in (5) and $S_\infty^{\pm}(s,F) \equiv \sup_{0<x<1} K_s^{\pm}(F(x),x)$.

Remark. Note that $\lim_{s \nearrow 1} g_s^+(a) = a$, in agreement with theorem 2.2, page 50, of Berk and Jones (1979).

Remark. Since $g_s^+(a) = g_s^-(a) \sim sa$ as $s \searrow 0$, the Bahadur efficacies of the statistics $S_n^{ur,\pm}(s)$ tend to be smaller than the efficacies of the Berk-Jones statistic $S_\infty^+(1,F) = r(F)$ (when the latter exists), and especially so for small s. This, together with extensive numerical computations of Jager (2006), strengthens the case in favor of the statistics $S_n^+(s) = \sup_{X_{(1)} \le x < X_{(n)}} K_s^+(\mathbb{F}_n(x), x)$ with restricted supremum. Unfortunately we do not yet know the large deviation behavior of these statistics with restricted supremum.

5. Attainment of the Ingster - Donoho - Jin optimal detection boundary. Jin (2004) and Donoho and Jin (2004) consider testing in a "sparse heterogenous mixture" problem defined as follows: Suppose that $Y_1, \ldots, Y_n$ are i.i.d. G on $\mathbb{R}$ and consider testing

$$\begin{aligned} &H_0: \quad G = \Phi, \qquad \text{the standard } N(0,1) \text{ distribution function} \\ &\text{versus} \\ &H_1: \quad G = (1-\epsilon)\Phi + \epsilon\Phi(\cdot - \mu) \quad \text{for some } \ \epsilon \in (0,1), \ \mu > 0 \end{aligned}$$

In particular they consider the $n-$dependent alternatives $H_1^{(n)}$ given by

$$H_1^{(n)}: \quad G_n = (1-\epsilon_n)\Phi + \epsilon_n\Phi(\cdot - \mu_n) \quad \text{for} \ \ \epsilon_n = n^{-\beta}, \ \ \mu_n = \sqrt{2r\log n} \tag{6}$$

where $1/2 < \beta < 1$ and $0 < r < 1$. By transforming to $X_i \equiv 1 - \Phi(Y_i)$ i.i.d. $F = 1 - G(\Phi^{-1}(1-\cdot))$ (with the X_i's taking values in $[0,1]$), the testing problem becomes: test

$$\begin{aligned} &H_0: \quad F = F_0, \qquad \text{the Uniform}(0,1) \text{ distribution function} \\ &\text{versus} \\ &H_1: \quad F = F_0(u) + \epsilon\{(1-u) - \Phi(\Phi^{-1}(1-u) - \mu)\} > F_0(u). \end{aligned}$$

(The corresponding $n-$ dependent sequence is $F_n(u) = u + \epsilon_n\{(1-u) - \Phi(\Phi^{-1}(1-u) - \mu_n)\}$ with the same choice of ϵ_n and μ_n as in (6).) Donoho and Jin (2004) consider several different test statistics, among which the principal contenders were Tukey's "higher criticism" statistic HC_n^* defined by

$$HC_n^* \equiv \sup_{X_{(1)} \le x < X_{([\alpha_0 n])}} \frac{\sqrt{n}(\mathbb{F}_n(x) - x)}{\sqrt{x(1-x)}}$$

for some $\alpha_0 > 0$ (they seem to usually take $\alpha_0 = 1/2$), and a one-sided version of the Berk-Jones statistic

$$BJ_n^+ \equiv n \sup_{X_{(1)} \le x < 1/2} K_1^+(\mathbb{F}_n(x), x)$$

where $K_s^+(u, v) \equiv K_s(u, v) 1\{0 < v < u < 1\}$.

Jin (2004) (see also Ingster (1998), Ingster (1997)) showed that the likelihood ratio test of H_0 versus $H_1^{(n)}$ has a "detection boundary" described as follows: set

$$\rho^*(\beta) = \begin{cases} \beta - 1/2, & 1/2 < \beta \le 3/4, \\ (1 - \sqrt{1-\beta})^2, & 3/4 < \beta < 1 \end{cases}$$

Then for $r > \rho^*(\beta)$ the likelihood ratio test (which makes use of knowledge of β and r) is size and power consistent against $H_1^{(n)}$ as $n \to \infty$. Donoho and Jin (2004) show that the tests of H_0 versus $H_1^{(n)}$ based on HC_n^* and BJ_n^+ are also size and power consistent as $n \to \infty$ and that both of these tests dominate several other tests based on multiple comparison procedures such as the sample range, sample maximum, FDR (False Discovery Rate) and Fisher's method; see e.g. Figure 1 of Donoho and Jin (2004) and their theorems 1.4 and 1.5.

We show here that the tests based on appropriate one-sided versions of the statistics $S_n(s)$, namely

$$nS_n^+(s) \equiv n \sup_{X_{(1)} \le x \le 1/2} K_s^+(\mathbb{F}_n(x), x)$$

have the same detection boundary for testing H_0 versus $H_1^{(n)}$ as the statistics HC_n^* and BJ_n^+.

Theorem 5.1 *For each $s \in [-1, 2]$ $\rho_s(\beta) = \rho^*(\beta)$ for $1/2 < \beta < 1$.*

While this may not be too suprising for $1 \le s \le 2$ in view of the Donoho-Jin results for $nS_n^+(1) = BJ_n^+$ and $nS_n^+(2) = (1/2)(HC_n^*)^2$, it seems new and interesting for $s \in [-1, 1)$. Figure 3 gives smoothed histograms of the values of the statistics $nS_n(s) - r_n$ under the null hypothesis H_0 (red) and under the alternative hypothesis $H_1^{(n)}$ (blue) for $n = .5 \times 10^6$, $r = .15$, and $\beta = 1/2$. This should be compared with figure 2 on page 978 of Donoho and Jin (2004) showing values of HC_n^* and HC_n^+, corresponding to our $s = 2$; their $HC_n^+ \equiv \sup_{1/n \le x \le 1/2} \sqrt{n}(\mathbb{F}_n(x) - x)^+ / \sqrt{x(1-x)}$.

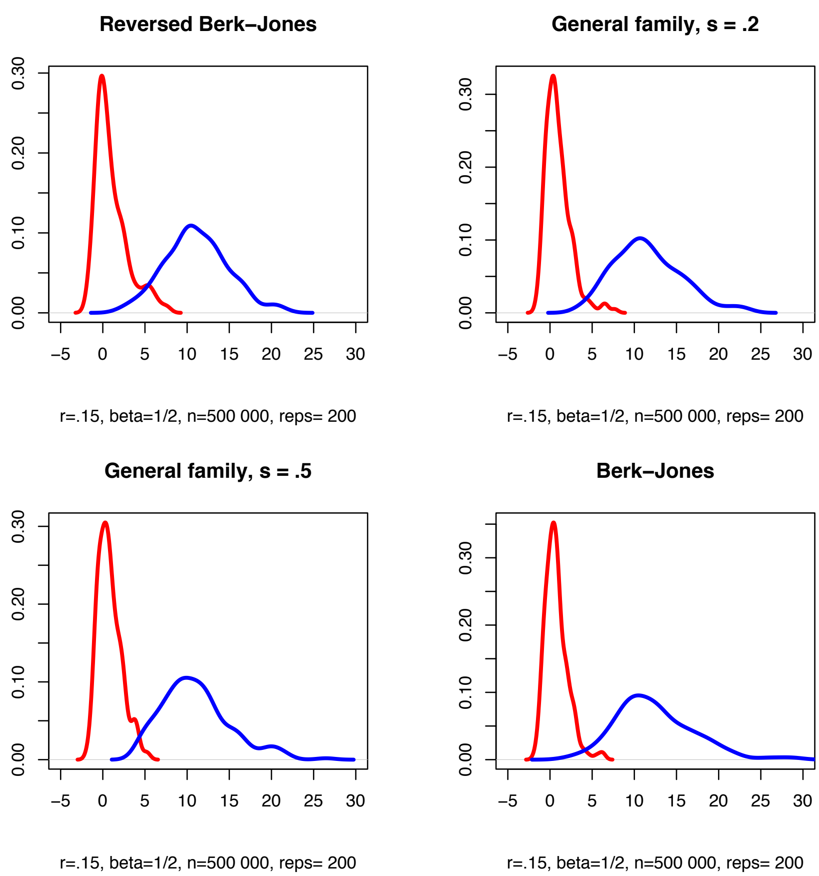

FIG 3. *Smoothed histograms of (reps) values of the statistics* $S_n^+(s)$ *under the null hypothesis* H_0 *(red) and alternative hypothesis* $H_1^{(n)}$ *(blue) with* $r = .15$, $\beta = 1/2$ *for* $s = 0, .2, .5, 1$

6. Discussion and some further problems. Here is a brief listing of some of the remaining open problems.

- Problem 1: What is the limit distribution of $S_n(s)$ under the null hypothesis when $s < -1$ or $s > 2$?
- Problem 2: What are necessary and sufficient conditions for $S_n(s)$ to converge to its natural parameter under fixed alternatives F for $s \le 0$?
- Problem 3: What is the large deviation behavior of $S_n(s)$ under the null hypothesis for $0 < s < 1$?
- Problem 4: Is there an appropriate contiguity theory for the statistics $S_n(s)$? [The only example involving something similar of which we are aware is Theorem A1 of Bickel and Rosenblatt (1973), but their results do not seem to apply to the statistics $S_n(s)$.]
- It is fairly easy to construct versions of our statistics $S_n(s)$ in more general settings by replacing the intervals $[0, x]$ and $(x, 1]$ with sets C and C^c for C in some class of sets $\mathcal{C}$. Then for testing $H_0 : \; P = P_0$ versus $H_1 : \; P_1 \ne P_0$ a natural generalization of the statistics $S_n(s)$ is
$$S_n(s, \mathcal{C}) = \sup_{C \in \mathcal{C}} K_s(\mathbb{P}_n(C), P_0(C))$$
where $\mathbb{P}_n$ is the empirical measure of $X_1, \ldots, X_n$ i.i.d. P. Problem 5: Do the statistics $S_n(s, \mathcal{C})$ have reasonable power behavior for some of the "chimeric alternatives" of Khmaladze (1998) for some choice of $\mathcal{C}$?

7. Proofs.

7.1. *Proofs for section 3.*

Proof of Theorem 3.1. We first carry out the proof for $-1 \le s < 1$, and then indicate the changes that are necessary for $1 \le s \le 2$. Fix $s \in [-1, 1)$. Note that
$$\frac{\partial}{\partial u} K_s(u, v)\Big|_{u=v} = \phi_s\left(\frac{u}{v}\right) - \phi_s\left(\frac{1-u}{1-v}\right)\Big|_{u=v} = \phi_s(1) - \phi_s(1) = 0$$
and
$$\frac{\partial^2}{\partial u^2} K(u, v) = \left(\frac{u}{v}\right)^{s-2} \frac{1}{v} + \left(\frac{1-u}{1-v}\right)^{s-2} \frac{1}{1-v} \equiv D_s(u, v)\, .$$
Hence it follows by Taylor expansion of $u \mapsto K_s(u, v)$ about $u = v$ that
$$\begin{aligned} K_s(u, v) &= K_s(v, v) + \frac{\partial}{\partial u} K_s(u, v)\Big|_{u=v} (u - v) + \frac{1}{2} \frac{\partial^2}{\partial u^2} K_s(u, v)\Big|_{u=u^*} (u - v)^2 \\ &= 0 + 0 + \frac{1}{2}(u - v)^2 D_s(u^*, v) \\ &= \frac{1}{2}(u - v)^2 D_s(u^*, v) \end{aligned}$$

for some u^* satisfying $|u^* - v| \le |u - v|$. This yields

$$K_s(\mathbb{F}_n(x), x) = \frac{1}{2}(\mathbb{F}_n(x) - x)^2 D_s(\mathbb{F}_n^*(x), x) \tag{7}$$

for $0 < x < 1$ where $|\mathbb{F}_n^*(x) - x| \le |\mathbb{F}_n(x) - x|$; i.e. $x \le \mathbb{F}_n^*(x) \le \mathbb{F}_n(x)$ on the event $x \le \mathbb{F}_n(x)$ and $\mathbb{F}_n(x) \le \mathbb{F}_n^*(x) \le x$ on the event $\mathbb{F}_n(x) \le x$.

We can write (7) as

$$K_s(\mathbb{F}_n(x), x) = \frac{1}{2}\frac{(\mathbb{F}_n(x) - x)^2}{x(1-x)}\left\{1 + x(1-x)D_s(\mathbb{F}_n^*(x), x) - 1\right\} \tag{8}$$

where

$$\begin{aligned}
|\mathrm{Rem}_n(x)| &\equiv \left|x(1-x)D_s(\mathbb{F}_n^*(x), x) - 1\right| \\
&= \left|x(1-x)\left\{\left(\frac{\mathbb{F}_n^*(x)}{x}\right)^{s-2}\frac{1}{x} + \left(\frac{1-\mathbb{F}_n^*(x)}{1-x}\right)^{s-2}\frac{1}{1-x}\right\} - 1\right| \\
&= \left|(1-x)\left(\frac{x}{\mathbb{F}_n^*(x)}\right)^{2-s} + x\left(\frac{1-x}{1-\mathbb{F}_n^*(x)}\right)^{2-s} - (1-x) - x\right| \\
&\le (1-x)\left|\left(\frac{x}{\mathbb{F}_n^*(x)}\right)^{2-s} - 1\right| + x\left|\left(\frac{1-x}{1-\mathbb{F}_n^*(x)}\right)^{2-s} - 1\right|.
\end{aligned}$$

Fix $\delta \in (0, 1/2)$. Now for $x \in [\delta, 1-\delta]$, $\mathbb{F}_n(x) \in [\delta/2, 1-\delta/2]$ a.s. for $n \ge N_\omega$, so, much as in Wellner and Koltchinskii (2003)

$$\sup_{\delta \le x \le 1-\delta} |\mathrm{Rem}_n(x)| = O_p(n^{-1/2})\,.$$

For $0 < v \le 1/2$ and $1 \le s \le 2$ the function $u \mapsto D_s(u, v)$ is monotone for $u \in (0, 1/2]$, while for $0 < v \le 1/2$ and $-1 \le s \le 1$ the function $u \mapsto D_s(u, v)$ is monotone for $u \in (0, b(v, s)]$ where $b(v, s) \equiv 1/(1 + c(v, s))$ with $c(v, s) \equiv ((1-v)/v)^{(1-s)/(3-s)}$, so

$$x(1-x)D_s(\mathbb{F}_n^*(x), x) \le x(1-x)\left\{D_s(x, x) \vee D_s(\mathbb{F}_n(x), x)\right\}$$

on the set $\{\mathbb{F}_n(x) < 1/2 \wedge b(x, s)\}$. Since $P(\mathbb{F}_n(\delta) \ge 1/2 \wedge b(\delta, s)) \to 0$ for $\delta < 1/2$ and $s \in [-1, 2]$, we get

$$\begin{aligned}
\sup_{X_{(1)} \le x \le \delta} |\mathrm{Rem}_n(x)| &\le \sup_{X_{(1)} \le x \le \delta}\left|\left(\frac{x}{\mathbb{F}_n(x)}\right)^{2-s} - 1\right| \\
&\quad + \sup_{X_{(1)} \le x \le \delta}\left|\left(\frac{1-x}{1-\mathbb{F}_n(x)}\right)^{2-s} - 1\right| \\
&= O_p(1) + o_p(1) = O_p(1) \qquad (9)
\end{aligned}$$

by Shorack and Wellner (1986), inequality 1, page 415, and inequality 2, (10.3.6), page 416. Alternatively, see Wellner (1978), lemma 2, page 75, and remark 1 (ii). Similarly

$$\begin{aligned} \sup_{1-\delta \le x < X_{(n)}} |\mathrm{Rem}_n(x)| &\le \sup_{1-\delta \le x < X_{(n)}} \Big| \left(\frac{x}{\mathbb{F}_n(x)} \right)^{2-s} - 1 \Big| \\ &\quad + \sup_{1-\delta \le x < X_{(n)}} \Big| \left(\frac{1-x}{1-\mathbb{F}_n(x)} \right)^{2-s} - 1 \Big| \\ &= o_p(1) + O_p(1) = O_p(1) \,. \end{aligned} \tag{10}$$

Now we write

$$S_n(s) = S_n(s,I) \vee S_n(s,II) \vee S_n(s,III)$$

where

$$\begin{aligned} S_n(s,I) &\equiv \sup_{\delta \le x \le 1-\delta} K_s(\mathbb{F}_n(x), x) = \sup_{\delta \le x \le 1-\delta} \frac{1}{2} (\mathbb{F}_n(x) - x)^2 D_s(\mathbb{F}_n^*(x), x) \,, \\ S_n(s,II) &\equiv \sup_{X_{(1)} \le x \le \delta} K_s(\mathbb{F}_n(x), x) = \frac{1}{2} \sup_{X_{(1)} \le x \le \delta} (\mathbb{F}_n(x) - x)^2 D_s(\mathbb{F}_n^*(x), x) \,, \\ S_n(s,III) &\equiv \sup_{1-\delta \le x < X_{(n)}} K_s(\mathbb{F}_n(x), x) = \frac{1}{2} \sup_{1-\delta \le x < X_{(n)}} (\mathbb{F}_n(x) - x)^2 D_s(\mathbb{F}_n^*(x), x) \,. \end{aligned}$$

By the monotonicity of $u \mapsto D_s(u,v)$ for $u \le 1/2$ again, with probability tending to 1,

$$\begin{aligned} S_n(s,II) &\le \frac{1}{2} \sup_{X_{(1)} \le x \le \delta} (\mathbb{F}_n(x) - x)^2 \{ D_s(\mathbb{F}_n(x), x) \vee D_s(x,x) \} \\ &\ge \frac{1}{2} \sup_{X_{(1)} \le x \le \delta} (\mathbb{F}_n(x) - x)^2 \{ D_s(\mathbb{F}_n(x), x) \wedge D_s(x,x) \} \,, \end{aligned}$$

and similarly, by the monotonicity of $u \mapsto D_s(u,v)$ for $1/2 \le u < 1$,

$$\begin{aligned} S_n(s,III) &\le \frac{1}{2} \sup_{1-\delta \le x < X_{(n)}} (\mathbb{F}_n(x) - x)^2 \{ D_s(\mathbb{F}_n(x), x) \vee D_s(x,x) \} \\ &\ge \frac{1}{2} \sup_{1-\delta \le x < X_{(n)}} (\mathbb{F}_n(x) - x)^2 \{ D_s(\mathbb{F}_n(x), x) \wedge D_s(x,x) \} \,. \end{aligned}$$

For $S_n(s,I)$,

$$\begin{aligned} S_n(s,I) &= \frac{1}{2} \sup_{\delta \le x \le 1-\delta} \frac{(\mathbb{F}_n(x) - x)^2}{x(1-x)} \left\{ 1 + O_p(n^{-1/2}) \right\} \\ &\le \frac{1}{2} \sup_{\delta \le x \le 1-\delta} (\mathbb{F}_n(x) - x)^2 \{ D_s(\mathbb{F}_n(x), x) \vee D_s(x,x) \} \left\{ 1 + O_p(n^{-1/2}) \right\} \,. \end{aligned}$$

In the second region the argument above leading to (9) yields

$$\begin{aligned} S_n(s,II) &\le \frac{1}{2}\sup_{X_{(1)}\le x\le\delta}(\mathbb{F}_n(x)-x)^2\{D_s(\mathbb{F}_n(x),x)\vee D_s(x,x)\} \\ &\ge \frac{1}{2}\sup_{X_{(1)}\le x\le\delta}(\mathbb{F}_n(x)-x)^2\{D_s(\mathbb{F}_n(x),x)\wedge D_s(x,x)\}\,, \end{aligned}$$

and similarly for $S_n(s,III)$. It follows that

$$\begin{aligned} S_n(s) &\le \frac{1}{2}\sup_{X_{(1)}\le x< X_{(n)}}(\mathbb{F}_n(x)-x)^2\,\{D_s(\mathbb{F}_n(x),x)\vee D_s(x,x)\}\left\{1+O_p(n^{-1/2})\right\} \\ &= \frac{1}{2}\sup_{X_{(1)}\le x< X_{(n)}}\left\{\frac{(\mathbb{F}_n(x)-x)^2}{x(1-x)}\,\{1\vee x(1-x)D_s(\mathbb{F}_n(x),x)\}\right\} \\ &\qquad\cdot\left\{1+O_p(n^{-1/2})\right\} \end{aligned} \tag{11}$$

and, on the other hand,

$$\begin{aligned} S_n(s) &\ge \frac{1}{2}\sup_{X_{(1)}\le x< X_{(n)}}(\mathbb{F}_n(x)-x)^2\,\{D_s(\mathbb{F}_n(x),x)\wedge D_s(x,x)\}\left\{1+O_p(n^{-1/2})\right\} \\ &= \frac{1}{2}\sup_{X_{(1)}\le x< X_{(n)}}\left\{\frac{(\mathbb{F}_n(x)-x)^2}{x(1-x)}\,\{1\wedge x(1-x)D_s(\mathbb{F}_n(x),x)\}\right\} \\ &\qquad\cdot\left\{1+O_p(n^{-1/2})\right\}\,. \end{aligned} \tag{12}$$

Now we break the supremum into the regions $[X_{(1)},d_n]$, $[d_n,1-d_n]$, and $[1-d_n,X_{(n)})$ with $d_n=(\log n)^k/n$ for any $k\ge 1$. Note that with $k=5$ as chosen below the middle interval is non-empty, i.e. $d_n\le 1/2$, if $n\ge 1010388$, so we are really relying on *large* n here! In our Monte-Carlo experiments to understand this limiting behavior we have replaced $d_n=(\log n)^5/n$ by $\tilde{d}_n=(\log n)^5/(50n)$ since $\tilde{d}_n\le .43$ for all n. Then we have

$$n\sup_{X_{(1)}\le x\le d_n}\frac{(\mathbb{F}_n(x)-x)^2}{x(1-x)}=o_p(b_n^2)$$

where $b_n=\sqrt{2\log_2 n}$; see Shorack and Wellner (1986), (26), page 602. Moreover,

$$\sup_{X_{(1)}\le x\le d_n}\left|x(1-x)D_s(\mathbb{F}_n(x),x)\right|=O_p(1)\,,$$

so

$$n\sup_{X_{(1)}\le x\le d_n}\frac{(\mathbb{F}_n(x)-x)^2}{x(1-x)}\,(1\,\#\,x(1-x)D_s(\mathbb{F}_n(x),x))=o_p(b_n^2) \tag{13}$$

for $\# = \wedge$ or $\# = \vee$, and similarly for the region $[1 - d_n, X_{(n)}]$ by a symmetric argument. On the other hand if we define

$$Z_n \equiv \sup_{d_n \le x \le 1-d_n} \frac{\sqrt{n}|\mathbb{F}_n(x) - x|}{\sqrt{x(1-x)}},$$

then, for $k \ge 5$

$$\frac{Z_n}{b_n} \to_p 1, \tag{14}$$

and

$$b_n Z_n - c_n \to_d Y_4 \sim E_v^4 \tag{15}$$

where $c_n = 2\log_2 n + (1/2)\log_3 n - (1/2)\log(4\pi)$ (see e.g. Shorack and Wellner (1986), page 600, (16.1.20)) and (16.1.17)). [Note that for the middle bracket in (13) we have

$$x(1-x)D_s(\mathbb{F}_n(x), x) \quad = \quad (1-x)\left(\frac{x}{\mathbb{F}_n(x)}\right)^{2-s} + x\left(\frac{1-x}{1-\mathbb{F}_n(x)}\right)^{2-s},$$

so

$$\sup_{X_{(1)} \le x \le X_{(1)}} x(1-x)D_s(\mathbb{F}_n(x), x) \quad \le \quad \left(\sup_{X_{(1)} \le x \le d_n} \frac{x}{\mathbb{F}_n(x)}\right)^{2-s} + o_p(1),$$

so by Wellner (1978), remark 1, the probability of large values of the main term can be bounded by

$$\begin{aligned} P\left(\left(\sup_{X_{(1)} \le x \le d_n} \frac{x}{\mathbb{F}_n(x)}\right)^{2-s} > \lambda\right) &= P\left(\sup_{X_{(1)} \le x \le d_n} \frac{x}{\mathbb{F}_n(x)} > \lambda^{1/(2-s)}\right) \\ &\le P\left(\sup_{X_{(1)} \le x \le 1} \frac{x}{\mathbb{F}_n(x)} > \lambda^{1/(2-s)}\right) \\ &\le e \cdot \lambda^{1/(2-s)} \exp(-\lambda^{1/(2-s)}).] \end{aligned}$$

Furthermore,

$$\left\|\frac{\mathbb{F}_n(x) - x}{x}\right\|_{d_n}^1 = O(a_n) \tag{16}$$

almost surely where

$$a_n^2 \equiv \frac{\log_2 n}{n d_n} = \frac{\log_2 n}{(\log n)^k} \to 0;$$

see Shorack and Wellner (1986), page 424, (4.5.10) and (4.5.11). It follows from (11), (12), and (13) - (16) that

$$\begin{aligned} nS_n(s) &= \frac{1}{2}\left\{ \sup_{d_n \le x \le 1-d_n} \frac{n(\mathbb{F}_n(x)-x)^2}{x(1-x)}\,(1+O_p(a_n)) \bigvee o_p(b_n^2) \right\} \\ &\qquad \times \left\{1+O_p(n^{-1/2})\right\} \\ &= \frac{1}{2}\left\{Z_n^2 \bigvee o_p(b_n^2)\right\} + o_p(1)\,. \qquad (17) \end{aligned}$$

Hence we can write

$$\begin{aligned} \frac{1}{2}Z_n^2 &= \frac{1}{2}(Z_n - c_n/b_n)(Z_n + c_n/b_n) + \frac{1}{2}\frac{c_n^2}{b_n^2} \\ &= \frac{1}{2}b_n(Z_n - c_n/b_n)\frac{Z_n + c_n/b_n}{b_n} + \frac{1}{2}\frac{c_n^2}{b_n^2} \end{aligned}$$

It follows that

$$\begin{aligned} nS_n(s) - \frac{1}{2}\frac{c_n^2}{b_n^2} &= b_n(Z_n - c_n/b_n)\frac{Z_n + c_n/b_n}{2b_n} \bigvee \left(o_p(b_n^2) - \frac{1}{2}\frac{c_n^2}{b_n^2}\right) + o_p(1) \\ &= b_n(Z_n - c_n/b_n)\frac{Z_n/b_n + c_n/b_n^2}{2} \bigvee \left(o_p(1) - 1/2\right) b_n^2 + o_p(1) \\ &\to_d Y_4\frac{1+1}{2} \bigvee \{-\infty\} = Y_4\,; \qquad (18) \end{aligned}$$

here we used $c_n^2/b_n^2 \sim b_n^2$ in the second equality. Since

$$\frac{1}{2}\frac{c_n^2}{b_n^2} = \log_2 n + (1/2)\log_3 n - (1/2)\log(4\pi) + o(1) = r_n + o(1)$$

this yields

$$P(nS_n(s) - r_n \le x) \quad \to \quad \exp(-4\exp(-x))\,, \qquad (19)$$

and completes the proof of Theorem 1. Note that the centering $c_n^2/(2b_n^2)$ emerges naturally in the course of this proof. This completes the proof for the case $s \in [-1,1)$. For $1 \le s \le 2$, there are two additional terms that enter, and both of these are $o_p(b_n^2)$ from the arguments in the previous section. The case $s = 2$ is easy since in this case $v(1-v)D_s(u,v) = 1$ for all u, while the result was stated for the case $s = 1$ by Berk and Jones (1978) and proved in Wellner and Koltchinskii (2003). □

Proof of Theorem 3.2. The fact that $nT_n(2) = A_n^2/2 \to_d A^2/2$ is classical; see Shorack and Wellner (1986), page 148. That $nT_n(1) \to_d A^2$ was noted by Einmahl and McKeague (2003) and proved by Wellner and Koltchinskii (2003). To show that the claim holds for $s \neq 1, 2$, we first consider $s \in (-\infty, 0) \cup (0, 1) \cup (1, 2]$. First note that since

$$K_s(u,v) = \{1 - u^s v^{1-s} - (1-u)^s (1-v)^{1-s}\}/(s(1-s))$$

we have $K_s(0,v) = \{1-(1-v)^{1-s}\}/s(1-s)$ and $K_s(1,v) = \{1 - v^{1-s}\}/s(1-s)$. It follows that

$$\begin{aligned} 2nT_n(s) &= \int_{[X_{(1)},X_{(n)})} 2nK_s(\mathbb{F}_n(x),x)dx \\ &\quad + \left(2n\int_0^{X_{(1)}} \{1-(1-x)^{1-s}\}dx + 2n\int_{X_{(n)}}^1 \{1-x^{1-s}\}dx \right) \frac{1_{(0,2]}(s)}{s(1-s)} \\ &\equiv I_n + II_n + III_n \,. \end{aligned}$$

Since

$$\int_0^y \{1-(1-x)^{1-s}\}dx \sim \int_0^y (1-s)xdx = \frac{1}{2}(1-s)y^2$$

as $y \to 0$, it follows that

$$II_n \sim n(1-s)X_{(1)}^2/s(1-s) = o_p(1)\,,$$

and similarly $III_n = o_p(1)$ as well. To handle I_n, fix $\alpha \in (1/2, 1)$ and set $a_n = n^{-\alpha}$. We know that

$$I_n^0 \equiv \int_{[X_{(1)},X_{(n)})} \frac{n(\mathbb{F}_n(x)-x)^2}{x(1-x)}dx \to_d \int_0^1 \frac{\mathbb{U}^2(t)}{t(1-t)}dt\,,$$

so it suffices to show that $I_n - I_n^0 = o_p(1)$. But since

$$K_s(u,v) = \frac{(u-v)^2}{2}\left\{ \left(\frac{u^*}{v}\right)^{s-2}\frac{1}{v} + \left(\frac{1-u^*}{1-v}\right)^{s-2}\frac{1}{1-v} \right\}$$

where $|u^* - v| \leq |u - v|$, we can write

$$\begin{aligned} I_n - I_n^0 &= \int_{[X_{(1)},X_{(n)})} \frac{n(\mathbb{F}_n(x)-x)^2}{x(1-x)} \\ &\qquad \left\{ \left[\left(\frac{x}{\mathbb{F}_n^*(x)}\right)^{2-s} - 1 \right](1-x) + \left[\left(\frac{1-x}{1-\mathbb{F}_n^*(x)}\right)^{2-s} - 1 \right] x \right\} dx \\ &= \int_{[X_{(1)},a_n)} \text{same}\, dx + \int_{a_n}^{1-a_n} \text{same}\, dx + \int_{(1-a_n,X_{(n)})} \text{same}\, dx \\ &\equiv C_n + D_n + E_n \end{aligned}$$

where $|\mathbb{F}_n^*(x) - x| \le |\mathbb{F}_n(x) - x|$ for each x. Now $D_n \to_p 0$ since

$$\left\| \frac{x}{\mathbb{F}_n(x)} - 1 \right\|_{a_n}^{1-a_n} \to_p 0 \qquad \text{and} \qquad \left\| \frac{1-x}{1-\mathbb{F}_n(x)} - 1 \right\|_{a_n}^{1-a_n} \to_p 0$$

by Wellner (1978), Theorem 0, page 77, and hence we also have

$$\left\| \left(\frac{x}{\mathbb{F}_n(x)} \right)^{2-s} - 1 \right\|_{a_n}^{1-a_n} \to_p 0 \qquad \text{and} \qquad \left\| \left(\frac{1-x}{1-\mathbb{F}_n(x)} \right)^{2-s} - 1 \right\|_{a_n}^{1-a_n} \to_p 0 \,.$$

Since $|\mathbb{F}_n^*(x) - x| \le |\mathbb{F}_n(x) - x|$, these last two displays also continue to hold with $\mathbb{F}_n$ replaced by $\mathbb{F}_n^*$. It remains only to show that $C_n \to_p 0$ and $E_n \to_p 0$. To handle C_n note that

$$\begin{aligned} C_n &\le \left\| \left(\frac{x}{\mathbb{F}_n^*(x)} \right)^{2-s} - 1 \right\|_{X_{(1)}}^{a_n} \int_{X_{(1)}}^{a_n} \frac{n(\mathbb{F}_n(x) - x)^2}{x(1-x)} dx \\ &\le \left\| \left(\frac{x}{\mathbb{F}_n(x)} \right)^{2-s} - 1 \right\|_{X_{(1)}}^{1} \int_{0}^{a_n} \frac{n(\mathbb{F}_n(x) - x)^2}{x(1-x)} dx \\ &\equiv \left\| \left(\frac{x}{\mathbb{F}_n(x)} \right)^{2-s} - 1 \right\|_{X_{(1)}}^{1} Z_n = O_p(1) o_p(1) \end{aligned}$$

where the $O_p(1)$ follows from the fact that $\|x/\mathbb{F}_n(x)\|_{X_{(1)}}^1 = O_p(1)$ (see e.g. Wellner (1978), remark 1, page 75), and the $o_p(1)$ follows since, for $\epsilon > 0$ we have, by Markov's inequality,

$$P(Z_n > \epsilon) \le \frac{E(Z_n)}{\epsilon} = \frac{1}{\epsilon} \int_0^{a_n} 1 \, dx = \frac{a_n}{\epsilon} \to 0 \,.$$

Finally $E_n \to_p 0$ by symmetry. This completes the proof for $s \le 2$, $s \ne 0, 1$.

We claim that the result holds for $s = 0$ by continuity, and it holds for $s = 1$ via Einmahl and McKeague (2003) and Wellner and Koltchinskii (2003). Thus it holds for all $s \le 2$. ■

7.2. *Proofs for section 4.*

Proof of Proposition 4.1. We first prove the claim for the "unrestricted version" of the statistics $S_n^{ur}(s)$ defined by

$$S_n^{ur}(s) \equiv \sup_{0<x<1} K_s(\mathbb{F}_n(x), x),$$

and then show that the difference between $S_n(s)$ and $S_n^{ur}(s)$ is negligible. Now for for

$s \in (0,1)$ and $C_s \equiv 1/(s(1-s))$, we have

$$\begin{aligned}
&|S_n^{ur}(s) - S_0(s)| \\
&\quad \le \quad C_s \sup_{0<x<1} |\{1 - \mathbb{F}_n(x)^s x^{1-s} - (1-\mathbb{F}_n(x))^s (1-x)^{1-s}\} \\
&\qquad\qquad - \{1 - F(x)^s x^{1-s} - (1-F(x))^s (1-x)^{1-s}\}| \\
&\quad \le \quad C_s \left\{ \sup_x |(\mathbb{F}_n(x)^s - F(x)^s) x^{1-s}| + \sup_x |\{(1-\mathbb{F}_n(x))^s - (1-F(x))^s\} x^{1-s}| \right\} \\
&\quad \le \quad C_s \left\{ \sup_x |\mathbb{F}_n(x)^s - F(x)^s| + \sup_x |(1-\mathbb{F}_n(x))^s - (1-F(x))^s| \right\} \\
&\quad \to_{a.s.} \quad 0\,.
\end{aligned}$$

Thus the proposition will be proved if we show that

$$S_n^{ur}(s) - S_n(s) \to_{a.s.} 0\,. \tag{20}$$

Now write $S_n^0(s) = \max\{R_n, M_n, L_n\}$ where

$$\begin{aligned}
M_n &\equiv \sup_{X_{(1)} \le x < X_{(n)}} K_s(\mathbb{F}_n(x), x) = S_n(s)\,, \\
L_n &\equiv \sup_{x < X_{(1)}} K_s(\mathbb{F}_n(x), x)\,, \quad \text{and} \\
R_n &\equiv \sup_{x \ge X_{(n)}} K_s(\mathbb{F}_n(x), x)\,.
\end{aligned}$$

Note that

$$S_n^0(s) - S_n(s) = \max\{L_n, M_n, R_n\} - M_n = \begin{cases} 0 & \text{if } M_n \ge L_n \vee R_n, \\ L_n - M_n & \text{if } L_n > M_n \vee R_n, \\ R_n - M_n & \text{if } R_n > M_n \vee L_n\,. \end{cases}$$

Now set

$$\alpha_0 \equiv \alpha_0(F) = \sup\{x : F(x) = 0\} \ge 0, \qquad \alpha_1 \equiv \alpha_1(F) = \inf\{x : F(x) = 1\} \le 1.$$

Note that

$$\begin{aligned}
L_n \quad &= \quad \sup_{x < X_{(1)}} K_s(\mathbb{F}_n(x), x) = \frac{1}{s(1-s)} \{1 - (1 - X_{(1)})^{1-s}\} \\
&\to_{a.s.} \quad \frac{1}{s(1-s)} \{1 - (1-\alpha_0)^{1-s}\} \equiv l_0(s, F)\,,
\end{aligned}$$

and, on the other hand

$$\begin{aligned}
M_n &= \sup_{X_{(1)} \le x < X_{(n)}} K_s(\mathbb{F}_n(x), x) \ge K_s(\mathbb{F}_n(X_{(1)}), X_{(1)}) = K_s(1/n, X_{(1)}) \\
&= \frac{1}{s(1-s)}\{1 - (1/n)^s X_{(1)}^s - (1-1/n)^s (1 - X_{(1)})^{1-s}\} \equiv L_n^0 \\
&\to_{a.s.} \frac{1}{s(1-s)}\{1 - (1-\alpha_0)^{1-s}\} = l_0(s, F)\,.
\end{aligned}$$

Similarly,

$$R_n = \sup_{x \ge X_{(n)}} K_s(\mathbb{F}_n(x), x) = \frac{1}{s(1-s)}\{1 - X_{(n)}^{1-s}\} \to_{a.s.} \frac{1}{s(1-s)}\{1 - \alpha_1^{1-s}\} \equiv r_0(s, F)\,,$$

while

$$\begin{aligned}
M_n &= \sup_{X_{(1)} \le x < X_{(n)}} K_s(\mathbb{F}_n(x), x) \ge K_s(\mathbb{F}_n(X_{(n)}-), X_{(n)}) = K_s(1 - 1/n, X_{(n)}) \\
&= \frac{1}{s(1-s)}\{1 - (1-1/n)^s X_{(n)} - (1/n)^s (1 - X_{(n)})^{1-s}\} \equiv R_n^0 \\
&\to_{a.s.} \frac{1}{s(1-s)}\{1 - \alpha_1^{1-s}\} = r_0(s, F)\,.
\end{aligned}$$

By combining these pieces It follows that

$$\begin{aligned}
0 \le S_n^0(s) - S_n(s) &= \max\{L_n, M_n, R_n\} - M_n \\
&= \left\{ \begin{array}{ll} 0 & \text{if } \ M_n \ge L_n \vee R_n, \\ L_n - M_n & \text{if } \ L_n > M_n \vee R_n, \\ R_n - M_n & \text{if } \ R_n > M_n \vee L_n \end{array} \right\} \\
&\le \left\{ \begin{array}{ll} 0 & \text{if } \ M_n \ge L_n \vee R_n, \\ L_n - L_n^0 & \text{if } \ L_n > L_n^0 \vee R_n, \\ R_n - R_n^0 & \text{if } \ R_n > R_n^0 \vee L_n \end{array} \right\} \\
&\to_{a.s.} 0\,.
\end{aligned}$$

This shows that (20) holds and completes the proof. ■

Proof of Proposition 4.2. Recall that when $s = 2$ such a condition follows from Theorem 3 in Jager and Wellner (2004): taking $b = 1/2$ and applying continuous mapping we conclude that

$$S_n(2) \to_{a.s.} \sup_{0<x<1} K_2(F(x), x) \qquad \text{if and only if} \qquad E\{[X(1-X)]^{-1/2}\} < \infty.$$

Similarly, for $1 < s < \infty$,

$$\begin{aligned} S_n(s) &= \sup_{0<x<1} \left\{ \mathbb{F}_n(x)^s x^{1-s} + (1-\mathbb{F}_n(x))^s (1-x)^{1-s} - 1 \right\} \frac{1}{s(s-1)} \\ &\to_{a.s.} \sup_{0<x<1} \left\{ F(x)^s x^{1-s} + (1-F(x))^s (1-x)^{1-s} - 1 \right\} \frac{1}{s(s-1)} \end{aligned}$$

if and only if

$$\|\mathbb{F}_n(x)^s x^{1-s} - F(x)^s x^{1-s}\| \to_{a.s.} 0$$

and

$$\|(1-\mathbb{F}_n(x))^s (1-x)^{1-s} - (1-F(x))^s (1-x)^{1-s}\| \to_{a.s.} 0,$$

if and only if

$$\| \left(\frac{\mathbb{F}_n(x)}{x^{(s-1)/s}} \right)^s - \left(\frac{F(x)}{x^{(s-1)/s}} \right)^s \| \to_{a.s.} 0$$

and

$$\| \left(\frac{1-\mathbb{F}_n(x)}{(1-x)^{(s-1)/s}} \right)^s - \left(\frac{1-F(x)}{(1-x)^{(s-1)/s}} \right)^s \| \to_{a.s.} 0,$$

if and only if

$$\| \left(\frac{\mathbb{G}_n(F(x))}{x^{(s-1)/s}} \right)^s - \left(\frac{F(x)}{x^{(s-1)/s}} \right)^s \| \to_{a.s.} 0$$

and

$$\| \left(\frac{1-\mathbb{G}_n(F(x))}{(1-x)^{(s-1)/s}} \right)^s - \left(\frac{1-F(x)}{(1-x)^{(s-1)/s}} \right)^s \| \to_{a.s.} 0.$$

Since $g(u) = u^s$ is uniformly continuous on bounded sets, these last two convergences occur if and only

$$\| \frac{\mathbb{G}_n(F(x))}{x^{(s-1)/s}} - \frac{F(x)}{x^{(s-1)/s}} \| \to_{a.s.} 0$$

and

$$\| \frac{1-\mathbb{G}_n(F(x))}{(1-x)^{(s-1)/s}} - \frac{1-F(x)}{(1-x)^{(s-1)/s}} \| \to_{a.s.} 0.$$

These in turn hold if and only if

$$\left\|\frac{\mathbb{G}_n(u) - u}{F^{-1}(u)^{(s-1)/s}}\right\| \to_{a.s.} 0, \qquad \text{and} \qquad \left\|\frac{1 - \mathbb{G}_n(u) - (1-u)}{(1 - F^{-1}(u))^{(s-1)/s}}\right\| \to_{a.s.} 0.$$

But in view of Wellner (1977), these convergences hold if and only if F satisfies

$$\int_0^1 \frac{1}{(F^{-1}(u)(1 - F^{-1}(u)))^{(s-1)/s}} du < \infty.$$

By the (inverse) probability integral transformation, the convergence in the last display is equivalent to

$$E[X(1-X)]^{(1-s)/s} < \infty.$$

This completes the proof of the claimed equivalences. □.

Proof of Proposition 4.3. For $s = 1$, this follows from Berk and Jones (1979) pages 55 - 56. Thus it suffices to prove the claimed convergences for $s > 1$ and $s < 0$.

For $s > 1$, fix $\alpha \in (0, 1)$. We begin by breaking the supremum over $(0, 1)$ into the regions $0 < F_s(x) < n^{-\alpha}$, $n^{-\alpha} \le F_s(x) \le 1 - n^{-\alpha}$, and $1 - n^{-\alpha} < F_s(x) < 1$:

$$\begin{aligned} S_n(s) &= \sup_{0<x<1} \{\mathbb{F}_n(x)^s x^{1-s} + (1 - \mathbb{F}_n(x))^x (1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &= \sup_{x:0<F_s(x)<n^{-\alpha}} \{\mathbb{F}_n(x)^s x^{1-s} + (1 - \mathbb{F}_n(x))^x (1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &\quad \vee \sup_{x:n^{-\alpha}<F_s(x)<1-n^{-\alpha}} \{\mathbb{F}_n(x)^s x^{1-s} + (1 - \mathbb{F}_n(x))^x (1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &\quad \vee \sup_{x:1-n^{-\alpha}<F_s(x)<1} \{\mathbb{F}_n(x)^s x^{1-s} + (1 - \mathbb{F}_n(x))^x (1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &\equiv I_n(s) \vee II_n(s) \vee III_n(s). \end{aligned}$$

For the main term, $I_n(s)$, let $\mathbb{G}_n$ be the empirical d.f. of n i.i.d. Uniform$(0, 1)$ random variables and use $\mathbb{F}_n \stackrel{d}{=} \mathbb{G}_n(F_s)$ to write

$$\begin{aligned} s(s-1)I_n(s) &\stackrel{d}{=} \sup_{0<F_s(x)<n^{-\alpha}} \left\{ \left(\frac{\mathbb{G}_n(F_s(x))}{F_s(x)}\right)^s F_s(x)^s x^{1-s} \right. \\ &\qquad \left. + \left(\frac{1 - \mathbb{G}_n(F_s(x))}{1 - F_s(x)}\right)^s (1 - F_s(x))^s (1-x)^{1-s} - 1 \right\} \end{aligned}$$

where

$$\sup_{x:F_s(x)<n^{-\alpha}} \left(\frac{\mathbb{G}_n(F_s(x))}{F_s(x)}\right)^s = \sup_{0<t<n^{-\alpha}} \left(\frac{\mathbb{G}_n(t)}{t}\right)^s \to_d \sup_{t>0} \left(\frac{\mathbb{N}(t)}{t}\right)^s,$$

$$F_s(x)^s x^{1-s} = \frac{x^{1-s}}{1+(x^{1-s}-1)/(s-1)} \to s-1$$

uniformly in $x \in [0, n^{-\alpha}]$, while

$$\sup_{x:F_s(x)<n^{-\alpha}} \left| \frac{1-\mathbb{G}_n(F_s(x))}{1-F_s(x)} - 1 \right| \to_{a.s.} 0,$$

and

$$\sup_{x:F_s(x)<n^{-\alpha}} |(1-F_s(x))^s(1-x)^{1-s} - 1| \to 0.$$

Combining these last five displays shows that $I_n(s) \to_d s^{-1} \sup_{t>0}(\mathbb{N}(t)/t)^s$; note that the limit variable is $\geq 1/s$ almost surely.

To handle the term $II_n(s)$, write

$$\begin{aligned} II_n(s) \quad &\stackrel{d}{=} \sup_{n^{-\alpha}<F_s(x)<1-n^{-\alpha}} \left\{ \left(\frac{\mathbb{G}_n(F_s(x))}{F_s(x)} \right)^s F_s(x)^s x^{1-s} \right. \\ &\quad + \left. \left(\frac{1-\mathbb{G}_n(F_s(x))}{1-F_s(x)} \right)^s (1-F_s(x))^s(1-x)^{1-s} - 1 \right\} \frac{1}{s(s-1)} \end{aligned}$$

where now the two terms involving the ratio of the empirical d.f. to the true d.f. F_s converge almost surely to 1. Hence we conclude that

$$II_n(s) \to_{a.s.} \sup_{0<x<1} K_s(F_s(x), x) = \frac{1}{s}$$

where the equality follows after some calculation. Finally, it is easily shown that $III_n(s) \to_{a.s.} 0$.

For $s < 0$, fix $\alpha \in (0, 1)$. We begin by breaking the supremum over $(0, 1)$ into the regions $X_{(1)} \leq x < F_s^{-1}(n^{-\alpha})$, $F_s^{-1}(n^{-\alpha}) \leq x \leq F_s^{-1}(1 - n^{-\alpha})$, and $F_s^{-1}(1 - n^{-\alpha}) < x < X_n)$:

$$\begin{aligned} S_n(s) &= \sup_{X_{(1)} \leq x < X_{n)}} \{\mathbb{F}_n(x)^s x^{1-s} + (1-\mathbb{F}_n(x))^x(1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &= \sup_{x:X_{(1)} \leq x < F_s^{-1}(n^{-\alpha})} \{\mathbb{F}_n(x)^s x^{1-s} + (1-\mathbb{F}_n(x))^x(1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &\quad \vee \sup_{x:\ F_s^{-1}(n^{-\alpha}) \leq x \leq F_s^{-1}(1-n^{-\alpha})} \{\mathbb{F}_n(x)^s x^{1-s} + (1-\mathbb{F}_n(x))^x(1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &\quad \vee \sup_{x:\ F_s^{-1}(1-n^{-\alpha}) < x < X_{(n)}} \{\mathbb{F}_n(x)^s x^{1-s} + (1-\mathbb{F}_n(x))^x(1-x)^{1-s} - 1\} \frac{1}{s(s-1)} \\ &\equiv I_n(s) \vee II_n(s) \vee III_n(s). \end{aligned}$$

For the main term, $I_n(s)$, let $\mathbb{G}_n$ be the empirical d.f. of n i.i.d. Uniform$(0,1)$ random variables and use $\mathbb{F}_n \stackrel{d}{=} \mathbb{G}_n(F_s)$ to write

$$\begin{aligned} s(s-1)I_n(s) \quad &\stackrel{d}{=} \sup_{x: X_{(1)} \le x < F_s^{-1}(n^{-\alpha})} \left\{ \left(\frac{F_s(x)}{\mathbb{G}_n(F_s(x))} \right)^{-s} F_s(x)^s x^{1-s} \right. \\ &\qquad \left. + \left(\frac{1-F_s(x)}{1-\mathbb{G}_n(F_s(x))} \right)^{-s} (1-F_s(x))^s (1-x)^{1-s} - 1 \right\} \end{aligned}$$

where

$$\sup_{x: X_{(1)} \le x < F_s^{-1}(n^{-\alpha})} \left(\frac{F_s(x)}{\mathbb{G}_n(F_s(x))} \right)^{-s} = \sup_{\xi_{(1)} \le t < n^{-\alpha}} \left(\frac{t}{\mathbb{G}_n(t)} \right)^{-s} \rightarrow_d \sup_{t \ge S_1} \left(\frac{t}{\mathbb{N}(t)} \right)^{-s},$$

$$F_s(x)^s x^{1-s} = x^{1-s}(1 - s(x^{-(1-s)} - 1)) \to -s$$

uniformly in $x \in [0, F^{-1}(n^{-\alpha})]$, while

$$\sup_{x: F_s(x) < n^{-\alpha}} \left| \frac{1-F_s(x)}{1-\mathbb{G}_n(F_s(x))} - 1 \right| \rightarrow_{a.s.} 0,$$

and

$$\sup_{x: F_s(x) < n^{-\alpha}} |(1-F_s(x))^s (1-x)^{1-s} - 1| \to 0.$$

Combining these last five displays shows that $I_n(s) \rightarrow_d (1-s)^{-1} \sup_{t \ge S_1} (t/\mathbb{N}(t))^{-s}$; note that the limit variable is $\ge 1/(1-s)$ almost surely.

To handle the term $II_n(s)$, write

$$\begin{aligned} II_n(s) \quad &\stackrel{d}{=} \sup_{n^{-\alpha} < F_s(x) < 1-n^{-\alpha}} \left\{ \left(\frac{F_s(x)}{\mathbb{G}_n(F_s(x))} \right)^{-s} F_s(x)^s x^{1-s} \right. \\ &\qquad \left. + \left(\frac{1-F_s(x)}{1-\mathbb{G}_n(F_s(x))} \right)^{-s} (1-F_s(x))^s (1-x)^{1-s} - 1 \right\} \frac{1}{s(s-1)} \end{aligned}$$

where now the two terms involving the ratio of the empirical d.f. to the true d.f. F_s converge almost surely to 1. Hence we conclude that

$$II_n(s) \rightarrow_{a.s.} \sup_{0<x<1} K_s(F_s(x), x) = \frac{1}{1-s}$$

where the equality follows after some calculation. Finally, it is easily shown that $III_n(s) \rightarrow_{a.s.} 0$. □

To prove Theorem 4.4 and its corollary, we will use the following lemma from Chernoff.

Lemma 7.1 *Let $X_1, X_2, \ldots$ be i.i.d. with continuous distribution F_0.*
(a) If $t < F_0(x)$, then $n^{-1} \log P(\mathbb{F}_n(x) \leq t) \to -K^-(t, F_0(x))$.
(b) If $t > F_0(x)$, then $n^{-1} \log P(\mathbb{F}_n(x) \geq t) \to -K^+(t, F_0(x))$.
In both cases, the convergence is from below.

Proof of Lemma 7.1. This follows from Theorem 1 of Chernoff (1952). □
Proof of Theorem 4.4. We first prove the theorem for $S_n^{ur+}(s)$. Since $K_s^+(t,x)$ is continuous in t and strictly increasing on $(x,1)$, then for $0 < a < (1-x^{1-s})/[s(1-s)]$, there is a unique $\tau = \tau(x)$ in $(x,1)$ for which $K_s^+(\tau,x) = a$ and $\{t : K_s^+(t,x) \geq a\} = [\tau, \infty)$. If $a \geq (1-x^{1-s})/[s(1-s)]$, then $\tau = 1$ necessarily.

For any fixed $x \in (0,1)$ we have

$$\begin{aligned} \frac{1}{n} \log P(S_n^{ur+}(s) \geq a) &\geq \frac{1}{n} \log P(K_s^+(\mathbb{F}_n(x), x) \geq a) \\ &= \frac{1}{n} \log P(\mathbb{F}_n(x) \geq \tau) \to -K^+(\tau, x) \end{aligned}$$

by Lemma 7.1. So

$$\liminf_{n\to\infty} \frac{1}{n} \log P(S_n^{ur+}(s) \geq a) \geq - \inf_{0<x<1} K^+(\tau(x), x). \tag{21}$$

Now for the reverse inequality. Let $\sup_{x|i}$ denote the supremum for $X_{(i)} \leq x \leq X_{(i+1)}$. Since $\mathbb{F}_n(x) = i/n$ on this range, we have

$$\sup_{x|i} K_s^+(\mathbb{F}_n(x), x) = K_s^+(i/n, X_{(i)}) \vee K_s^+(i/n, X_{(i+1)}) = K_s^+(i/n, X_{(i)}).$$

Note that for $x < X_{(1)}$ we have $\mathbb{F}_n(x) = 0$, and so $K_s^+(\mathbb{F}_n(x), x) = 0$ also. So we can write $S_n^{ur+}(s)$ as

$$S_n^{ur+}(s) = \max_{1\leq i\leq n} \{K_s^+(i/n, X_{(i)})\} = \max_{1\leq i\leq n} \{K_s^+(\mathbb{F}_n(X_{(i)}), X_{(i)})\}.$$

Now, using monotonicity of τ

$$\begin{aligned}
&\frac{1}{n}\log P(S_n^{ur+}(s)\geq a) = \frac{1}{n}\log P(\max_{1\leq i\leq n}\{K_s^+(\mathbb{F}_n(X_{(i)}),X_{(i)})\}\geq a)\\
\leq\ & \frac{1}{n}\log\sum_{i=1}^n P(K_s^+(\mathbb{F}_n(X_{(i)}),X_{(i)})\geq a) = \frac{1}{n}\log\sum_{i=1}^n P(\mathbb{F}_n(X_{(i)})\geq \tau(X_{(i)}))\\
=\ & \frac{1}{n}\log\sum_{i=1}^n P(i/n\geq \tau(X_{(i)})) = \frac{1}{n}\log\sum_{i=1}^n P(\tau^{-1}(i/n)\geq X_{(i)})\\
\leq\ & \frac{1}{n}\log\sum_{i=1}^n P(\mathbb{F}_n(\tau^{-1}(i/n))\geq \mathbb{F}_n(X_{(i)})) = \frac{1}{n}\log\sum_{i=1}^n P(\mathbb{F}_n(\tau^{-1}(i/n))\geq i/n)\\
\leq\ & \frac{1}{n}\log\sum_{i=1}^n e^{-nK^+(i/n,\tau^{-1}(i/n))}, \qquad \text{by lemma 7.1(b)},\\
\leq\ & \frac{1}{n}\log\sum_{i=1}^n e^{-n\min_{1\leq i\leq n}K^+(i/n,\tau^{-1}(i/n))} \leq \frac{1}{n}\log\sum_{i=1}^n e^{-n\inf_{0<x<1}K^+(x,\tau^{-1}(x))}\\
=\ & \frac{1}{n}\log\sum_{i=1}^n e^{-n\inf_{0<x<1}K^+(\tau(x),x)} = \frac{1}{n}\log\left[ne^{-n\inf_{0<x<1}K^+(\tau(x),x)}\right]\\
=\ & -\inf_{0<x<1}K^+(\tau(x),x)+\frac{\log n}{n}
\end{aligned}$$

So we conclude that

$$\limsup_{n\to\infty}\frac{1}{n}\log P(S_n^{ur+}(s)\geq a)\leq -\inf_{0<x<1}K^+(\tau(x),x). \tag{22}$$

Combining this last display with (21) yields the convergence part of (5). To prove the explicit formula for g_s^+, note that $K^+(\tau^+(x),x)$ is a decreasing function of x until $x=[1=s(1-s)a]^{1/(1-s)}$ where $K^+(\tau^+(x),x)=\infty$. Thus

$$\begin{aligned}
&\inf_{0<x<1}K^+(\tau^+(x),x)\\
&\quad= \lim_{\epsilon\searrow 0}K^+(\tau^+([1=s(1-s)a]^{1/(1-s)}-\epsilon),[1=s(1-s)a]^{1/(1-s)}-\epsilon)\\
&\quad= -\log[1-s(1-s)a]/(1-s),
\end{aligned}$$

so the given formula for the infimum in (5) holds. This completes the proof for $S_n^{ur+}(s)$. The proof for $S_n^{ur-}(s)$ is analogous using lemma 7.1(a). □

7.3. *Proofs for section 5.*

The following lemma extends Lemma A.4, page 988, Donoho and Jin (2004).

Lemma 7.2 *(i) For $0 < v \le u \le 1/2$, and $-1 \le s \le 2$*

$$K_s^+(u,v) \le \frac{1}{2}\frac{(u-v)^2}{v(1-v)} \equiv K_2(u,v). \tag{23}$$

(ii) Let $1 < s \le 2$ and $v = v(u)$ satisfy $0 < v \le u < 1$. Then, as $u \to 0$,

$$K_s(u,v) = \begin{cases} K_2(u,v)\left[1 + O(1-(v/u)^{2-s}) \vee O(((1-v)/(1-u))^{2-s} - 1)\right], & \textit{if } u/v \to 1 \\ v\phi_s(u/v)(1+o(1)) = u\{((u/v)^{s-1} - s\}(1+o(1))/(s(s-1)), & \textit{if } u/v \to \infty. \end{cases}$$

(iii) Let $s = 1$ and $v = v(u)$ satisfy $0 < v \le u < 1$. Then, as $u \to 0$,

$$K_1(u,v) = \begin{cases} K_2(u,v)\left[1 + O(u + (u/v) - 1)\right], & \textit{if } u/v \to 1 \\ u\log(u/v)(1+o(1)), & \textit{if } u/v \to \infty. \end{cases}$$

(iv) Let $s \in [-1,1) \setminus \{0\}$ and $v = v(u)$ satisfy $0 < v \le u < 1$. Then, as $u \to 0$,

$$K_s(u,v) = \begin{cases} K_2(u,v)\left[1 + O(1-(v/u)^{2-s}) \vee O(((1-v)/(1-u))^{2-s} - 1)\right], & \textit{if } u/v \to 1 \\ \frac{1}{1-s}u(1+o(1)), & \textit{if } u/v \to \infty. \end{cases}$$

(v) Let $s = 0$ and $v = v(u)$ satisfy $0 < v \le u < 1$. Then, as $u \to 0$,

$$K_0(u,v) = \begin{cases} K_2(u,v)\left[1 + O(1-(v/u)^{2}) \vee O(((1-v)/(1-u))^{2} - 1)\right], & \textit{if } u/v \to 1 \\ u(1+o(1)), & \textit{if } u/v \to \infty. \end{cases}$$

Remark. Note that for $1 < s \le 2$, as $u \to 0$ and $u/v \to \infty$,

$$\begin{aligned} v\phi_s(u/v) &= v\left\{(1-s) + s\frac{u}{v} - \left(\frac{u}{v}\right)^s\right\}\frac{1}{s(1-s)} \\ &= \frac{u}{s(s-1)}\left\{\left(\frac{u}{v}\right)^{s-1} + (s-1)\frac{v}{u} - s\right\} \\ &\sim \frac{u}{s(s-1)}\left\{\left(\frac{u}{v}\right)^{s-1} - s\right\}(1+o(1)). \end{aligned}$$

where the right side converges to $u\log(u/v)(1+o(1))$ as $s \searrow 1$.

Proof. (i) Letting $u = tv$, it suffices to show that for $0 < v \le 1/2$ and $1 \le t \le 1/(2v)$

$$K_s(tv,v) \le \frac{1}{2}\frac{(t-1)^2}{1-v}v$$

or, equivalently, since $K_s(tv,v) = v\phi_s(t) + (1-v)\phi_s((1-tv)/(1-t))$,

$$\phi_s(t) + \left(\frac{1}{v} - 1\right)\phi_s\left(\frac{1-tv}{1-v}\right) \le \frac{1}{2}\frac{(t-1)^2}{1-v}.$$

Let

$$f_s(t) \equiv \phi_s(t) + \left(\frac{1}{v} - 1\right)\phi_s\left(\frac{1-tv}{1-v}\right) - \frac{1}{2}\frac{(t-1)^2}{1-v}.$$

Now by direct calculation, $f_s(1) = 0$, and

$$\begin{aligned} f_s'(t) &= \phi_s'(t) + \left(\frac{1}{v} - 1\right)\phi_s'\left(\frac{1-tv}{1-v}\right)\left(\frac{-v}{1-v}\right) - \frac{t-1}{1-v} \\ &= \phi_s'(t) - \phi_s'\left(\frac{1-tv}{1-v}\right) - \frac{t-1}{1-v} \end{aligned}$$

so that $f_s'(1) = 0$. Furthermore

$$\begin{aligned} f_s''(t) &= \phi_s''(t) + \frac{v}{1-v}\phi_s''\left(\frac{1-tv}{1-v}\right) - \frac{1}{1-v} \\ &= \frac{1}{1-v}\left\{(1-v)\phi_s''(t) + v\phi_s''\left(\frac{1-tv}{1-v}\right) - 1\right\} \\ &= \frac{1}{1-v}\left\{(1-v)\left(\frac{1}{t}\right)^{2-s} + v\left(\frac{1-v}{1-tv}\right)^{2-s} - 1\right\} \\ &\le \frac{1}{1-v}\left\{\left((1-v)\frac{1}{t} + v\left(\frac{1-v}{1-tv}\right)\right)^{2-s} - 1\right\} \\ &\qquad \text{since } \ x^{2-s} \ \text{ is concave for } \ 1 \le s \le 2 \\ &= \frac{1}{1-v}\left\{\left(\frac{1-v}{t(1-tv)}\right)^{2-s} - 1\right\} \le 0 \end{aligned}$$

using $v(1-v) \le vt(1-vt)$ for $0 \le v \le vt \le 1/2$ implies $((1-v)/(t(1-tv)) \le 1$. Here we used

$$\phi_s'(x) = \frac{s - sx^{s-1}}{s(1-s)} = \frac{1-x^{s-1}}{1-s}, \qquad \phi_s''(x) = x^{s-2}.$$

Since $1 \le t \le 1/(2v)$ it follows that $v \le vt \le 1/2 < 1$, $1 - v \ge 1 - vt \ge 1/2 > 0$, and $1 \ge (1-vt)/(1-v) \ge 1/(2(1-v))$. When $s < 1$ we calculate

$$f_s'''(t) = \phi_s'''(t) - \left(\frac{v}{1-v}\right)^2 \phi_s'''\left(\frac{1-tv}{1-v}\right)$$

and note that $f_s'''(1) < 0$ while $f_s'''(t) = 0$ has a unique root, so to show $f_s''(t) \le 0$ it suffices to show $f_s''(1/(2v) \le 0$ for $0 \le v \le 1/2$. By a straightforward calculation we get

$$f_s''(1/(2v)) = (2v)^{2-s} + \frac{v}{1-v}\left(\frac{1-v}{1/2}\right)^{2-s} - \frac{1}{1-v}$$

which is ≤ 0 for $0 \leq v \leq 1/2$ if $s \geq -1$. This shows that $f_s''(t) \leq 0$ in the range $-1 \leq s < 1$, and completes the proof of (i).
(ii) By expanding $K_s(u,v)$ as a function of u as in (8),

$$K_s(u,v) = K_2(u,v)\{1 + v(1-v)D_s(u^*,v) - 1\}$$

with $|u^* - v| \leq |u - v|$; since $0 < v \leq u$, we necessarily have $0 < v \leq u^* \leq u$. Here

$$\begin{aligned} v(1-v)D_s(u^*,v) - 1 &= (1-v)\left\{\left(\frac{v}{u^*}\right)^{2-s} - 1\right\} + v\left\{\left(\frac{1-v}{1-u^*}\right)^{2-s} - 1\right\} \\ &\equiv I + II. \end{aligned}$$

Now $0 < v \leq u^* \leq u$ implies $1 \leq u^*/v \leq u/v$ so $v/u \leq v/u^* \leq 1$ and $1 - v \geq 1 - u^* \geq 1 - u$ implies $(1-v)/(1-u) \geq (1-v)/(1-u^*) \geq 1$. Thus $I \leq 0$ and $II \geq 0$. It follows that

$$v(1-v)D_s(u^*,v) - 1 \leq v\left\{\left(\frac{1-v}{1-u}\right)^{2-s} - 1\right\}.$$

Similarly,

$$v(1-v)D_s(u^*,v) - 1 \geq (1-v)\left\{\left(\frac{v}{u}\right)^{2-s} - 1\right\}.$$

in this range, and the claimed bound in the first part of (ii) follows.

To prove the second part of (ii), note that when $u/v \to \infty$ we can write

$$\begin{aligned} \frac{K_s(u,v)}{v\phi_s(u/v)} &= \frac{1 - (\frac{u}{v})^s v - (\frac{1-u}{1-v})^s(1-v)}{v\{1 - s + s(u/v) - (u/v)^s\}} \\ &= \frac{\left(\frac{u}{v}\right)^s v + \left(\frac{1-u}{1-v}\right)^s (1-v) - 1}{\left(\frac{u}{v}\right)^s v - v(1-s) - su} \\ &= \frac{1 + \left[\left(\frac{1-u}{1-v}\right)^s (1-v) - 1\right]/[(u/v)^s v]}{1 - [v(1-s) + su]/[(u/v)^s v]} \\ &\equiv \frac{1 + A(u,v)}{1 - B(u,v)} \end{aligned}$$

where, for $1 < s \leq 2$,

$$B(u,v) = \frac{v(1-s) + su}{(u/v)^s v} = \frac{1-s}{(u/v)^s} + s(v/u)^{s-1} = o(1)$$

and

$$\begin{aligned}
A(u,v) &= \frac{\left(\frac{1-u}{1-v}\right)^s (1-v) - 1}{(u/v)^s v} \\
&= \frac{\left(\frac{1-u}{1-v}\right)^s [(1-v)-1] + \left(\frac{1-u}{1-v}\right)^s - 1}{(u/v)^s v} \\
&= -\frac{\left(\frac{1-u}{1-v}\right)^s}{(u/v)^s} + \frac{\left(\frac{1-u}{1-v}\right)^s - 1}{(u/v)^s v} \\
&= o(1) + \frac{1 - su + sv - 1}{(u/v)^s v} + o(1) \\
&= o(1) + s\frac{1-(u/v)}{(u/v)^s} = o(1) - s(v/u)^{s-1} = o(1).
\end{aligned}$$

Thus the second part of (ii) holds.

The first part of (iv) is proved exactly as in (ii). To prove the second part of (iv) we write

$$\begin{aligned}
K_s(u,v) &= \frac{1}{s(1-s)}\{1 - (u/v)^s v - ((1-u)/(1-v))^s(1-v)\} \\
&= \frac{1}{s(1-s)}\left\{v\left(1 - \left(\frac{u}{v}\right)^s\right) + (1-v)\left(1 - \left(\frac{1-u}{1-v}\right)^s\right)\right\} \\
&= \frac{1}{s(1-s)}\left\{u\left(\frac{u}{v}\right)^{s-1}\left(-1 + \left(\frac{v}{u}\right)^s\right) + \frac{(1-v)(s(u-v)+o(u)+o(v))}{1-sv+o(v)}\right\} \\
&= \frac{1}{s(1-s)}\left\{su(1-(v/u))(1+o(1)) - u(v/u)^{1-s}\{1-(v/u)^s\}\right\} \\
&= \frac{1}{1-s}u(1+o(1)).
\end{aligned}$$

(v) The proof of (v) is similar to the proof of (iv). □

Lemma 7.3 *A. Suppose that* $X_1, \ldots, X_n$ *are i.i.d.* F_n *with* $0 < \rho^*(\beta) < r < \beta/3$. *Then* $r < 1/4$ *and for any* $0 < r_0 < r$,

$$\sup_{n^{-4r} < x < n^{-4r_0}} \left| \frac{\mathbb{F}_n(x)}{x} - 1 \right| \to_p 0, \tag{24}$$

Proof. Note that $\mathbb{F}_n(\cdot) \stackrel{d}{=} \mathbb{G}_n(F_n(\cdot))$ where $\mathbb{G}_n$ is the empirical d.f. of n i.i.d. $U(0,1)$ random variables $\xi_1, \ldots, \xi_n$ and

$$F_n(x) = x + \epsilon_n\{(1-x) - \Phi(\Phi^{-1}(1-x) - \mu_n)\} \geq x.$$

Thus with $\|\cdot\|_a^b \equiv \sup_{a\le t\le b}|f(t)|$,

$$\sup_{n^{-4r}<x<n^{-4r_0}}\Big|\frac{\mathbb{F}_n(x)}{x}-1\Big| \tag{25}$$

$$= \Big\|\Big(\frac{\mathbb{F}_n(x)}{x}-1\Big)^+\Big\|_{n^{-4r}}^{n^{-4r_0}} \vee \Big\|\Big(1-\frac{\mathbb{F}_n(x)}{x}\Big)^+\Big\|_{n^{-4r}}^{n^{-4r_0}}$$

$$\stackrel{d}{=} \Big\|\Big(\frac{\mathbb{G}_n(F_n(x))}{x}-1\Big)^+\Big\|_{n^{-4r}}^{n^{-4r_0}} \vee \Big\|\Big(1-\frac{\mathbb{G}_n(F_n(x))}{x}\Big)^+\Big\|_{n^{-4r}}^{n^{-4r_0}}. \tag{26}$$

The second term in this last display converges to 0 in probability easily since $F_n(x)/x \ge 1$ implies that it is bounded by

$$\Big\|\Big(1-\frac{\mathbb{G}_n(F_n(x))}{F_n(x)}\Big)^+\Big\|_{n^{-4r}}^{1} \le \Big\|\Big(1-\frac{\mathbb{G}_n(t)}{t}\Big)^+\Big\|_{n^{-4r}}^{1} \to_p 0$$

by Theorem 0 of Wellner (1978). On the other hand,

$$\begin{aligned}\frac{\mathbb{G}_n(F_n(x))}{x}-1 &= \frac{\mathbb{G}_n(F_n(x))}{F_n(x)}\frac{F_n(x)}{x}-1\\ &= \Big(\frac{\mathbb{G}_n(F_n(x))}{F_n(x)}-1\Big)\frac{F_n(x)}{x}+\Big(\frac{F_n(x)}{x}-1\Big),\end{aligned}$$

so again by Theorem 0 of Wellner (1978) the first term of (26) converges to 0 in probability if

$$\limsup_n \|F_n(x)/x\|_{n^{-4r}}^1 < \infty, \quad \text{and} \quad \sup_{n^{-4r}<x<n^{-4r_0}}\Big(\frac{F_n(x)}{x}-1\Big) \to 0.$$

But this holds by a straightforward analysis using the asymptotics of Φ^{-1} when $r < \beta/3$. □

Now have the tools in place to prove our extension of the results of Donoho and Jin (2004).

Proof of Theorem 5.1. First consider $1 < s < 2$. As in Donoho and Jin (2004) we first consider the case $r < \beta/3$. Then $r < 1/4$ and we can choose $0 < r_0 < r < 1/4$. From Lemma 7.3 the convergence (24) holds. Thus by part (ii) of Lemma 7.2, it follows that for $n^{-4r} < x < n^{-4r_0}$ we have

$$nK_s^+(\mathbb{F}_n(x),x) = \frac{1}{2}\left(\frac{(\mathbb{F}_n(x)-x)^+}{\sqrt{x(1-x)}}\right)^2(1+o_p(1)),$$

and hence

$$nS_n^+(s) \ge \sup_{n^{-4r}<x<n^{-4r_0}} nK_s^+(\mathbb{F}_n(x),x) \ge \frac{1}{2}HC_{n,r,r_0}^{*2}(1+o_p(1)).$$

Thus $nS_n^+(s)$ separates H_0 and $H_1^{(n)}$ for $s \in (1, s)$ and $r < \beta/3$.

Now suppose that $r > (1 - \sqrt{1-\beta})^2$ (and still $1 < s < 2$). Since $(r+\beta)/(2\sqrt{r}) < 1$ we can pick a constant $q < 1$ such that

$$\frac{(r+\beta)}{2\sqrt{r}} \vee \sqrt{r} < \sqrt{q} < 1.$$

As argued by Donoho and Jin, under $H_1^{(n)}$, $n\mathbb{F}_n(n^{-q}) \sim \text{Binomial}(n, L_n n^{-[\beta+(\sqrt{q}-\sqrt{r})^2]})$ where $L_n n^{-[\beta+(\sqrt{q}-\sqrt{r})^2]} >> n^{-q}$; here L_n is a logarithmic term that does not contributed significantly to the argument. Hence we have $\mathbb{F}_n(n^{-q})/n^{-q} >> 1$, and thus from part (ii) of Lemma 7.2 again,

$$nK_s^+(\mathbb{F}_n(n^{-q}), n^{-q}) = \frac{n\mathbb{F}_n(n^{-q})}{s(s-1)} \left\{ \left(\frac{\mathbb{F}_n(n^{-q})}{n^{-q}} \right)^{s-1} - s \right\} (1 + o_p(1))$$

Hence we conclude that

$$nS_n^+(s) \geq nK_s^+(\mathbb{F}_n(n^{-q}), n^{-q}) = \frac{n\mathbb{F}_n(n^{-q})}{s(s-1)} \left\{ \left(\frac{\mathbb{F}_n(n^{-q})}{n^{-q}} \right)^{s-1} - s \right\} (1 + o_p(1))$$

so using $\beta + (\sqrt{q} - \sqrt{r})^2 < q < 1$ we conclude that $nS_n^+(s)$ separates H_0 and $H_1^{(n)}$ in this range.

Now consider $-1 \leq s < 1$. For this range of s the argument is exactly the same as above, but now using parts (iv) and (v) of Lemma 7.2. [Note that the conclusion of Lemma A.1 of Donoho and Jin (2004) can be strengthened considerably as follows: if $Z_n \sim \text{Bin}(n, \pi_n)$ with $\pi_n \to 0$ and $n\pi_n \to \infty$, then $Z_n \to_p \infty$; i.e. for any number $M > 0$ we have $P(Z_n \geq M) \to 1$. This follows easily from Theorem 0 of Wellner (1978) since $|Z_n/(n\pi_n) - 1| \to_p 0$ so $Z_n = (Z_n/n\pi_n)n\pi_n \to_p 1 \cdot \infty = \infty$. This also follows easily from the Paley-Zygmund inequality (see e.g. Kallenberg (1997), page 40: $P(Z_n > rE(Z_n)) \geq (1-r)_+^2 (EZ_n)^2/[EZ_n^2]$.] □

8. Acknowledgements. We owe thanks to Art Owen for sharing his computer programs, Robert Berk for helpful suggestions, and David Donoho and Jiashun Jin for pointing out several additional references.

REFERENCES

Abrahamson, I. G. (1967). Exact Bahadur efficiencies for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics. *Ann. Math. Statist.* **38** 1475–1490.

Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser. B* **28** 131–142.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statistics* **23** 193–212.

Berk, R. H. and Jones, D. H. (1978). Relatively optimal combinations of test statistics. *Scand. J. Statist.* **5** 158–162.

Berk, R. H. and Jones, D. H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Gebiete* **47** 47–59.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

Cai, T. T., Jin, J. and Low, M. G. (2005). Estimation and confidence sets for sparse normal mixtures. Tech. rep., Purdue University.

Cayón, L., Jin, J. and Treaster, A. (2005). Higher criticism statistic: detecting and identifying non-gaussianity in the WMAP first year data. *Mon. Not. R. Astron. Soc.* **362** 826–832.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics* **23** 493–507.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46** 440–464.

Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8** 85–108.

Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318.

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-fit techniques*, vol. 68 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York.

Darling, D. A. and Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994.

Durbin, J., Knott, M. and Taylor, C. C. (1975). Components of Cramér-von Mises statistics. II. *J. Roy. Statist. Soc. Ser. B* **37** 216–237.

Eicker, F. (1979). The asymptotic distribution of the suprema of the standardized empirical processes. *Ann. Statist.* **7** 116–138.

Einmahl, J. H. J. and McKeague, I. W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli* **9** 267–290.

Groeneboom, P. and Shorack, G. R. (1981). Large deviations of goodness of fit statistics and linear combinations of order statistics. *Ann. Probab.* **9** 971–987.

Ingster, Y. I. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69.

Ingster, Y. I. (1998). Minimax detection of a signal for l^n-balls. *Math. Methods Statist.* **7** 401–428 (1999).

Jaeschke, D. (1979). The asymptotic distribution of the supremum of the standardized empirical distribution function on subintervals. *Ann. Statist.* **7** 108–115.

Jager, L. (2006). Goodness-of-fit statistics based on phi-divergences. Ph.D. dissertation in progress, University of Washington Department of Statistics.

Jager, L. and Wellner, J. A. (2004a). A new goodness of fit test: the reversed berk-jones statistic. Tech. rep., University of Washington Department of Statistics.

Jager, L. and Wellner, J. A. (2004b). On the "Poisson boundaries" of the family of weighted Kolmogorov statistics. In *A festschrift for Herman Rubin*, vol. 45 of *IMS Lecture Notes Monogr. Ser.* Inst. Math. Statist., Beachwood, OH, 319–331.

Janssen, A. (2000). Global power functions of goodness of fit tests. *Ann. Statist.* **28** 239–253.

Jin, J. (2004). Detecting a target in very noisy data from multiple looks. In *A festschrift for Herman Rubin*, vol. 45 of *IMS Lecture Notes Monogr. Ser.* Inst. Math. Statist., Beachwood, OH, 255–286.

Kallenberg, O. (1997). *Foundations of modern probability.* Probability and its Applications (New York), Springer-Verlag, New York.

Kendall, D. G. and Kendall, W. S. (1980). Alignments in two-dimensional random sets of points. *Adv. in Appl. Probab.* **12** 380–424.

Khmaladze, E. V. (1998). Goodness of fit tests for "chimeric" alternatives. *Statist. Neerlandica* **52** 90–111.

Liese, F. and Vajda, I. (1987). *Convex statistical distances*, vol. 95 of *Teubner-Texte zur Mathematik [Teubner Texts in Mathematics].* BSB B. G. Teubner Verlagsgesellschaft, Leipzig.

Meinshausen, N. and Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34** xxx–yyy.

Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests.* Cambridge University Press, Cambridge.

Noé, M. (1972). The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics. *Ann. Math. Statist.* **43** 58–64.

Owen, A. B. (1995). Nonparametric likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.* **90** 516–521.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Vajda, I. (1989). *Theory of Statistical Inference and Information.* Theory and Decision Libarary, Series B: Mathematical and Statistical Methods, Kluwer, Dordrecht.

Wellner, J. A. (1977). A Glivenko-Cantelli theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* **5** 473–480.

Wellner, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. Verw. Gebiete* **45** 73–88.

Wellner, J. A. and Koltchinskii, V. (2003). A note on the asymptotic distribution of Berk-Jones type statistics under the null hypothesis. In *High dimensional probability, III (Sandjberg, 2002)*, vol. 55 of *Progr. Probab.* Birkhäuser, Basel, 321–332.

Leah Jager
University of Washington
Department of Statistics
Box 354322
Seattle, WA 98195-4322
e-mail: leah@stat.washington.edu

Jon A. Wellner
University of Washington
Department of Statistics
Box 354322
Seattle, WA 98195-4322
e-mail: jaw@stat.washington.edu