TESTS FOR NORMALITY BASED ON DENSITY ESTIMATORS OF CONVOLUTIONS

ANTON SCHICK, YISHI WANG AND WOLFGANG WEFELMEYER

Abstract. Recent results show that densities of convolutions can be estimated by local U-statistics at the root-n rate in various norms. Motivated by this and the fact that convolutions of normal densities are normal, we introduce new tests for normality which use as test statistics weighted L_1-distances between the standard normal density and local U-statistics based on standardized observations. We show that such test statistics converge at the root-n rate and determine their limit distributions as functionals of Gaussian processes. We also address a choice of bandwidth. Simulations show that our tests are competitive with other tests of normality.

1. Introduction

Suppose we observe independent and identically distributed random variables $X_1, \ldots, X_n$ with mean μ and finite variance σ^2. We want to test the hypothesis that the observations are normally distributed. The literature contains a wealth of goodness-of-fit tests for this purpose, in particular the tests of Lilliefors (1967), Shapiro–Wilk (1965), Csörgő (1986), and the BHEP test introduced by Epps and Pulley (1983) and Baringhaus and Henze (1988).

Let Φ denote the standard normal distribution function, and let F_2 denote the distribution function of

$$\frac{X_1 + X_2 - 2\mu}{\sigma\sqrt{2}}.$$

The Lévy characterization says that X_1 has a normal distribution if and only if $F_2 = \Phi$. This can be used to test for normality. Arcones and Wang (2006) estimate $F_2(x)$ by the U-statistic

$$\mathbb{F}_2(x) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathbf{1}[\hat{Z}_i + \hat{Z}_j \le x]$$

based on estimated observations $\hat{Z}_j = (X_j - \bar{X})/(\hat{\sigma}\sqrt{2})$ of $Z_j = (X_j - \mu)/(\sigma\sqrt{2})$, where $\bar{X}$ is the sample mean and $\hat{\sigma}$ is the sample standard deviation. They propose the Kolmogorov–Smirnov-type test statistic

$$\sup_{x \in \mathbb{R}} |\mathbb{F}_2(x) - \Phi(x)|. \tag{1}$$

Key words and phrases. Convolution-type kernel density estimator, goodness-of-fit test.

Anton Schick was supported by NSF Grant DMS0906551.

Instead of comparing distribution functions, we can also compare densities. Densities of convolutions can be estimated at the root-n rate by local U-statistics; see Frees (1994), Saavedra and Cao (2000) for pointwise rates, and Schick and Wefelmeyer (2004, 2007) and Giné and Mason (2007) for rates in various norms and functional central limit theorems in the corresponding function spaces. Du and Schick (2007) consider estimating derivatives of convolution densities. The results of Giné and Mason (2007) and Du and Schick (2007) cover random bandwidths.

Motivated by these results, we estimate the density f_2 of F_2 at x by the local U-statistic

$$\hat{f}_2(x) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} k_b(x - \hat{Z}_i - \hat{Z}_j)$$

with $k_b(x) = k(x/b)/b$ for a kernel k and a bandwidth b. This suggests test statistics of the form $\|\hat{f}_2 - \varphi\|$ with $\|\ \|$ a norm and φ the standard normal density. Here we use the V-norm

$$\|\hat{f}_2 - \varphi\|_V = \int |\hat{f}_2(x) - \varphi(x)| V(x)\, dx, \tag{2}$$

which is a weighted L_1-norm, with positive weight function V. Let L_V denote the space of functions with finite V-norm. Schick and Wefelmeyer (2004) have shown root-n consistency and a functional central limit theorem in L_V for the version $\tilde{f}_2$ of $\hat{f}_2$ which is obtained by replacing the estimated $\hat{Z}_j$ by the true Z_j,

$$\tilde{f}_2(x) = \frac{2}{n(n-1)} \sum_{1\le i<j\le n} k_b(x - Z_i - Z_j).$$

More precisely, their result shows that, under mild assumptions, the expansion

$$\|\tilde{f}_2 - f_2 - \mathbb{H}\|_V = o_p(n^{-1/2})$$

holds with

$$\mathbb{H}(x) = \frac{1}{n}\sum_{j=1}^{n} \big(2g(x - Z_j) - 2f_2(x)\big)$$

and g the density of Z_1, and that $n^{1/2}\mathbb{H}$ converges in distribution in L_V to a centered Gaussian process. Under the null hypothesis of normality, g is the normal density with mean 0 and variance 1/2, and $f_2 = \varphi$.

We show that under the null hypothesis, the stochastic expansion

$$\|\hat{f}_2 - \varphi - \mathbb{G}\|_V = o_p(n^{-1/2}) \tag{3}$$

holds with

$$\mathbb{G}(x) = \frac{1}{n}\sum_{j=1}^{n} \big(2g(x - Z_j) - 2\varphi(x) - 2Z_j x\varphi(x) - (Z_j^2 - 1/2)(x^2-1)\varphi(x)\big),$$

and that $n^{1/2}\mathbb{G}$ converges in distribution in L_V to a centered Gaussian process. This implies that the random variable $\|n^{1/2}\mathbb{G}\|_V$ converges in distribution to some random variable Γ. Thus, under the null hypothesis, the test statistic $T = n^{1/2}\|\hat{f}_2 - \varphi\|_V$ also converges in distribution to Γ. A test of asymptotic level α is obtained by rejecting the null hypothesis if T exceeds the $(1-\alpha)$-quantile of Γ. Since the distribution of Γ is intractable, the quantile must be determined by simulation. We compare the power of our test for various choices of weight function V with other tests available in the literature. Our simulations show that our tests are quite competitive.

2. Result

We now give conditions under which the stochastic expansion (3) holds. As in Schick and Wefelmeyer (2004) we require the weight function V to satisfy the following conditions:

The function V is continuous at 0 with $V(0) = 1$ and satisfies

$$V(x+y) \le V(x)V(y), \quad x, y \in \mathbb{R},$$
$$V(sx) \le V(x), \quad |s| \le 1, \quad x \in \mathbb{R}.$$

In addition, we require the integrability condition

$$\int (1+|x|)^2 V^2(x)\varphi(x)\,dx < \infty.$$

Possible choices of V are $V(x) = 1$, which yields the usual L_1-norm, and $V(x) = (1+|x|)^r$ and $V(x) = \exp(r|x|)$ with $r \ge 0$.

We assume that the kernel k is an absolutely continuous and symmetric density and satisfies the integrability conditions

$$\int (1+|x|)^2 V^2(\sqrt{2}x)(k^2(x) + x^2 k'^2(x))\,dx < \infty,$$
$$\int k(x)(1+|x|)^2 V(\sqrt{2}x)\,dx < \infty.$$

Possible kernels are the normal kernel or absolutely continuous densities with compact support such as the Epanechnikov kernel.

To allow for some data-driven bandwidth selection we assume that the bandwidth b is of the form $b = \hat{\lambda}c$ for nonstochastic $c = c_n$ satisfying $nc_n \to \infty$ and $nc_n^4 \to 0$, and positive random variables $\hat{\lambda} = \hat{\lambda}_n$ such that $\hat{\lambda}_n + 1/\hat{\lambda}_n$ is bounded in probability.

Theorem 1. *Suppose the weight function V, the kernel k and the bandwidth b are as above. Then, under the null hypothesis of normality, we have the stochastic expansion (3), and the test statistic $T = n^{1/2}\|\hat{f}_2 - \varphi\|_V$ converges in distribution to Γ.*

It is well known that the quality of kernel estimators depends crucially on the choice of bandwidth, see e.g. Wand and Jones (1995). This is to some extent also true for the local U-statistic $\tilde{f}_2$. If μ and σ are known, we use a bandwidth which minimizes the mean integrated squared error $M(b) = E[(\int(\tilde{f}_2(x) - \phi(x))^2\,dx]$. Using the Hoeffding decomposition, we obtain

$$M(b) = \frac{4\tau}{n} + \frac{2}{n(n-1)b}\|K^2\|_1 + \frac{b^4}{4}\|(\phi'')^2\|_1 \sigma_K^4 + O\Big(\frac{b^2}{n} + \frac{1}{n^2}\Big) + o(b^4),$$

where $\tau = \int (g^2 * g(x) - \phi^2(x))dx$, $\|K^2\|_1 = \int K^2(u)du$, $\|(\phi'')^2\|_1 = \int(\phi''(u))^2du$, $\sigma_K^2 = \int u^2K(u)du$, and $g = \phi_{1/2}$ is the normal density with mean zero and variance $1/2$. The minimizer of $M(b)$ is asymptotically equal to the minimizer of

$$\frac{2}{n^2 b}\|K^2\|_1 + \frac{b^4}{4}\|(\phi_2'')^2\|_1\sigma_K^4,$$

which is

$$b_* = \Big(\frac{2||K^2||_1}{n^2||(\phi'')^2||_1\sigma_K^4}\Big)^{1/5}. \tag{4}$$

Estimating μ and σ is not expected to have a large effect on the optimal bandwidth. For this reason we work with b_* in the simulations below.

Brownrigg and Khmaladze (2010) have observed that it is difficult to distinguish between a normal density and the convolution of this density with another (centered) density. For such alternatives, our estimator will also not work well. It is not essential here that the first density is normal. One explanation of this phenomenon is the following. Suppose we observe independent copies $X_1, \dots, X_n$ of a random variable X. Consider, for simplicity, the problem of testing the one-point hypothesis that X has density f. Introduce a one-dimensional family of alternatives under which X is distributed as $Y + aZ$, where Y has density f and Z is independent of Y and has a mean zero density g. Here a is a (non-negative) parameter. Under such an alternative, X has density $f_a(x) = \int f(x - az)g(z)\,dz$, and, under appropriate conditions, its derivative with respect to a is $\dot f_a(x) = -\int zf'(x-az)g(z)\,dz$. Hence $\dot f_0 \equiv 0$. In particular, the score function vanishes at $a = 0$. This will typically imply that alternatives are contiguous even for parameters a of order $n^{-1/4}$, not just for the usual order $n^{-1/2}$.

Specifically, let f be the standard normal density, and let g be the density of the uniform distribution on the interval $(-1/2, 1/2)$. Then $f_a(x) = \int_{-1/2}^{1/2} f(x - az)\,dz$. The square-root $s_a(x) = \sqrt{f_a(x)}$ is twice continuously differentiable in a with derivatives

$$\dot s_a(x) = -\frac{\int_{-1/2}^{1/2} f'(x - az)z\,dz}{2s_a(x)}$$

and

$$\ddot s_a(x) = \frac{\int_{-1/2}^{1/2} f''(x - az)z^2\,dz}{2s_a(x)} - \frac{\big(\int_{-1/2}^{1/2} f'(x - az)z\,dz\big)^2}{4s_a^3(x)}.$$

Since $\dot s_0(x) = 0$ for all x, we have

$$s_a(x) - s_0(x) = a^2\int_0^1 (1 - t)\ddot s_{ta}(x)\,dt.$$

For the standard normal density f we have $-f'(x) = xf(x) = xs(x)s(x)$ with $s(x) = \sqrt{f(x)} = s_0(x)$. Two applications of the Cauchy–Schwarz inequality yield

$$\begin{aligned}\Big(\int_{-1/2}^{1/2} f'(x-az)z\,dz\Big)^2 &\le s_a^2(x)\int_{-1/2}^{1/2}(x-az)^2 f(x-az)z^2\,dz\\ &\le s_a^3(x)\Big(\int_{-1/2}^{1/2}(x-az)^4 f(x-az)z^4\,dz\Big)^{1/2}.\end{aligned}$$

For the standard normal density f we also have $f''(x) = (x^2-1)f(x)$. The Cauchy–Schwarz inequality therefore yields

$$\Big|\int_{-1/2}^{1/2} f''(x-az)z^2\,dz\Big| \le s_a(x)\Big(\int_{-1/2}^{1/2}\big((x-az)^2-1\big)^2 f(x-az)z^4\,dz\Big)^{1/2}.$$

From this we derive that

$$|\ddot{s}_a(x)| \le \Big(\int_{-1/2}^{1/2}\big(1+(x-az)^4\big)f(x-az)z^4\,dz\Big)^{1/2}$$

and can conclude

$$\begin{aligned}\int |\ddot{s}_a(x)|^2\,dx \ &\le \int\int_{-1/2}^{1/2}\big(1+(x-az)^4\big)f(x-az)z^4\,dz\,dx\\ &= \int(1+x^4)f(x)\,dx\int_{-1/2}^{1/2} z^4\,dz \le 1.\end{aligned}$$

Since $\ddot{s}_a(x) \to \ddot{s}_0(x)$ as is easily checked, and

$$\ddot{s}_0(x) = \frac{f''(x)}{2\sqrt{f(x)}}\int_{-1/2}^{1/2} z^2\,dz = \frac{1}{24}(x^2-1)\sqrt{f(x)},$$

we conclude from the above that

$$\int\Big(s_a(x) - s_0(x) - \frac{1}{48}a^2(x^2-1)\sqrt{f(x)}\Big)^2\,dx = o(a^4).$$

This means that s_a is twice Hellinger differentiable at $a = 0$, with vanishing first derivative. It implies that the likelihood ratio of the observations $(X_1,\dots,X_n)$ under $a = tn^{-1/4}$ and under $a = 0$ is asymptotically normal for each $t > 0$. Hence the distributions of the observations $(X_1,\dots,X_n)$ under $a = 0$ and $a = tn^{-1/4}$ are mutually contiguous. We refer to Le Cam and Yang (2000, Section 6.2) for these results.

3. Simulations

For practical purposes, our test statistics need to be evaluated by numerical integration, and for this it is convenient to replace the domain of integration by a symmetric interval. This is done in the simulations reported below. We evaluate the integral from -3 to 3 and divide the interval into 24 equally spaced subintervals of length $1/4$. The integrals over these subintervals are evaluated using the 7-point closed Newton–Cotes formula: for a subinterval $[a,b]$, we pick $h=(b-a)/6=1/24$, and approximate the integral $\int_a^b g(x)dx$ by

$$\frac{h}{140}\Big(41g(a){+}216g(a+h)+27g(a+2h)+272g(a+3h) \\ +27g(a+4h)+216g(a+5h)+41g(a+6h)\Big).$$

We will compare our tests based on (2) with several other tests of normality. The test statistic AW of Arcones and Wang (2006) is described in (1). We also consider the QH test proposed by Chen and Shapiro (1995). It has power comparable or superior to the original Shapiro–Wilk test and is based on the test statistic

$$1-\frac{1}{(n-1)\hat\sigma}\sum_{j=1}^{n-1}\frac{X_{i+1}-X_i}{H_{i+1}-H_i},$$

where H_i is the $(i-3/8)/(n+1/4)$ quantile of the standard normal distribution. The CS test of Csörgő (1986) is based on the test statistic

$$\sup_{|t|\le T}\big||\hat\chi(t)|^2-\exp(-t^2)\big|$$

with

$$\hat\chi(t)=\frac{1}{n}\sum_{j=1}^{n}\exp(it\sqrt{n}\hat Z_j).$$

As recommended by Csörgő (1986), in the simulations we use instead the test statistic

$$\sup_{-10^2\le k\le 10^2}\Big||\hat\chi((1.47)10^{-2}k)\Big|^2-\exp\big(-((1.47)10^{-2}k)^2\big)\Big|.$$

The BHEP test proposed by Epps and Pulley (1983) and Baringhaus and Henze (1988) is based on the test statistic

$$\int\big(\hat\chi(t)-\exp(-t^2/2)\big)^2\exp(-t^2/2)\,dt.$$

The following table shows the critical values for size .05 of our test statistic (2) with three different norms,

$$V_1(x)=1,\qquad V_2(x)=(1+|x|)^4,\qquad V_3(x)=\exp(|x|).$$

In all three cases we use the bandwidth b_* as given in (4). As kernel k we take the standard normal density ϕ. The critical values were generated for sample sizes 30, 50, 100 and 200 using 10,000 simulations.

Table 1: Critical values for size .05.

	V_1	V_2	V_3
$n=30$	0.15647500	5.642335	0.6261157
$n=50$	0.12657350	4.661203	0.5101362
$n=100$	0.09080627	3.353047	0.3691862
$n=200$	0.06573999	2.428222	0.2669440

The following tables present the power of the above tests for some selected alternatives.

Table 2: Power for Gamma(18,1)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.1375	0.1201	0.1044	0.1275	0.1418	0.1480	0.1476
$n = 50$	0.2018	0.1660	0.1262	0.1901	0.2181	0.2309	0.2326
$n = 100$	0.3849	0.3045	0.1805	0.3877	0.4159	0.4503	0.4432
$n = 200$	0.6817	0.5502	0.2362	0.6757	0.6937	0.7355	0.7307

Table 3: Power for Weibull(10,2)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.2144	0.1966	0.1529	0.1989	0.2289	0.2382	0.2385
$n = 50$	0.3386	0.2879	0.1988	0.3225	0.3680	0.3813	0.3851
$n = 100$	0.6409	0.5469	0.3155	0.6318	0.6717	0.7072	0.7001
$n = 200$	0.9176	0.8417	0.4532	0.9089	0.9196	0.9425	0.9389

Table 4: Power for Beta(23,3)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.4238	0.3645	0.2360	0.3331	0.4033	0.4329	0.4270
$n = 50$	0.6717	0.5543	0.3273	0.5423	0.6457	0.7076	0.6986
$n = 100$	0.9634	0.8808	0.4985	0.8976	0.9423	0.9709	0.9642
$n = 200$	1.0000	0.9972	0.6961	0.9987	0.9997	1.0000	1.0000

Table 5: Power for rlnorm(Nn,0,.2)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.1962	0.1740	0.1471	0.1825	0.2065	0.2103	0.2094
$n = 50$	0.2968	0.2458	0.1959	0.2915	0.3239	0.3402	0.3441
$n = 100$	0.5619	0.4567	0.2976	0.5616	0.5924	0.6280	0.6215
$n = 200$	0.8601	0.7464	0.4187	0.8559	0.8704	0.8953	0.8907

Table 6: Power for F(30,20)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.6477	0.5950	0.4899	0.5880	0.6536	0.6661	0.6643
$n = 50$	0.8627	0.7960	0.6508	0.8103	0.8650	0.8880	0.8854
$n = 100$	0.9939	0.9810	0.8816	0.9872	0.9926	0.9969	0.9960
$n = 200$	1.0000	0.9998	0.9852	1.0000	1.0000	1.0000	1.0000

Table 7: Power for Gumbel(0,1)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.4644	0.4194	0.3512	0.4248	0.4727	0.4848	0.4849
$n = 50$	0.6860	0.6084	0.4817	0.6473	0.7017	0.7268	0.7304
$n = 100$	0.9473	0.8948	0.7163	0.9310	0.9474	0.9630	0.9606
$n = 200$	0.9991	0.9954	0.9086	0.9985	0.9990	0.9995	0.9994

Table 8: Power for rt(Nn,df=30,ncp=5)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.1269	0.1151	0.1146	0.1273	0.1333	0.1345	0.1355
$n = 50$	0.1798	0.1474	0.1447	0.1892	0.2095	0.2131	0.2170
$n = 100$	0.3234	0.2591	0.2147	0.3485	0.3500	0.3633	0.3633
$n = 200$	0.5475	0.4429	0.2930	0.5883	0.5886	0.6038	0.6046

Table 9: Power for rinvgauss(Nn,10,1/120)

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.3374	0.2944	0.2351	0.3003	0.3491	0.3561	0.3577
$n = 50$	0.5335	0.4477	0.3246	0.4831	0.5415	0.5760	0.5735
$n = 100$	0.8482	0.7432	0.4931	0.8161	0.8609	0.8933	0.8856
$n = 200$	0.9928	0.9679	0.6974	0.9874	0.9912	0.9956	0.9946

Table 10: Power for Double Exponential

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.3407	0.4882	0.3763	0.4377	0.3927	0.3373	0.3594
$n = 50$	0.4806	0.6809	0.5313	0.5867	0.5447	0.4298	0.4745
$n = 100$	0.7571	0.9073	0.8179	0.8174	0.8109	0.6557	0.7171
$n = 200$	0.9647	0.9949	0.9829	0.9721	0.9745	0.8845	0.9367

Table 11: Power for t-distribution with df=5

	QH	CS	$BHEP$	AW	V_1	V_2	V_3
$n = 30$	0.2459	0.3431	0.2354	0.2902	0.2689	0.2643	0.2656
$n = 50$	0.3396	0.4818	0.2943	0.3811	0.3663	0.3440	0.3568
$n = 100$	0.5232	0.7016	0.4887	0.5547	0.5574	0.5007	0.5240
$n = 200$	0.7750	0.9069	0.7444	0.7840	0.7836	0.7019	0.7398

The above simulations show that our tests are competitive with other normality tests. The first eight alternative distributions are skewed, while the last two are symmetric. Our tests dominate for the skewed distributions, while the CS test dominates for the symmetric ones.

4. Proof of Theorem 1

To stress the dependence of $\tilde{f}_2$ on the bandwidth b, we write now $\tilde{f}_{2,b}$ instead of $\tilde{f}_2$. Note that

$$\hat{Z}_j = \frac{X_j - \bar{X}}{\hat{\sigma}\sqrt{2}} = \hat{s}(Z_j - \bar{Z})$$

with $\hat{s} = \sigma/\hat{\sigma}$ and $\bar{Z} = (1/n)\sum_{j=1}^n Z_j$. Thus

$$x - \hat{Z}_i - \hat{Z}_j = \hat{s}\Big(\frac{x}{\hat{s}} + 2\bar{Z} - Z_i - Z_j\Big).$$

This shows that

$$\hat{f}_2(x) = \frac{1}{\hat{s}}\tilde{f}_{2,b/\hat{s}}\Big(\frac{x}{\hat{s}} + 2\bar{Z}\Big).$$

Under our assumptions on V we have, with $V_*(x) = V(\sqrt{2}x)$,

$$R_n = \|\tilde{f}_{2,b/\hat{s}} - \varphi - \mathbb{H}\|_{V_*} = o_p(n^{-1/2}),$$

and $n^{1/2}\mathbb{H}$ converges in distribution in L_{V_*}. This follows from Theorem 4.2 in Du and Schick (2007), applied with $p = 1$, $r = 0$, $m = 1$ and $\xi = 1$. Thus, on the event $\hat{s} \le \sqrt{2}$, whose probability tends to one, we have

$$\begin{aligned}\int \Big|\frac{1}{\hat{s}}\tilde{f}_{2,b/\hat{s}}\Big(\frac{x}{\hat{s}} + 2\bar{Z}\Big) - \frac{1}{\hat{s}}\varphi\Big(\frac{x}{\hat{s}} + 2\bar{Z}\Big) - \frac{1}{\hat{s}}\mathbb{H}\Big(\frac{x}{\hat{s}} + 2\bar{Z}\Big)\Big| V(x)\,dx \\ \le \int |\tilde{f}_{2,b/\hat{s}}(u) - \varphi(u) - \mathbb{H}(u)| V(\hat{s}(u - 2\bar{Z}))\,du \\ \le V(2\hat{s}\bar{Z})R_n = o_p(n^{-1/2}).\end{aligned}$$

For $0 < \delta < \sqrt{2}$, define maps L_δ on L_{V_*} by

$$L_\delta(h) = \sup_{|s-1|\le\delta}\sup_{|t|\le\delta}\int V(x)|sh(sx+t) - h(x)|\,dx.$$

These maps are continuous, and $L_\delta(h)$ decreases to zero for each h in L_{V_*} as δ decreases to 0. Therefore Dini's theorem yields

$$\sup_{|s-1|\le\delta}\sup_{|t|\le\delta}\sup_{h\in K}\int V(x)|sh(sx+t) - h(x)|\,dx \to 0$$

as $\delta \to 0$ for every compact subset K of L_{V_*}. Hence we have

$$\int \Big|\frac{1}{\hat{s}}\mathbb{H}\Big(\frac{x}{\hat{s}} + 2\bar{Z}\Big) - \mathbb{H}(x)\Big| V(x)\,dx = o_p(n^{-1/2})$$

in view of the tightness of $n^{1/2}\mathbb{H}$ in L_{V_*}. Finally, it is easy to verify that

$$\int \Big|\frac{1}{\hat{s}}\varphi\Big(\frac{x}{\hat{s}} + 2\bar{Z}\Big) - \varphi(x) - (\hat{s} - 1)(x^2 - 1)\varphi(x) + 2\bar{Z}x\varphi(x)\Big| V(x)\,dx = o_p(n^{-1/2})$$

in view of $\hat{s} - 1 = O_p(n^{-1/2})$ and $\bar{Z} = O_p(n^{-1/2})$. It is easy to check that

$$\hat{s} - 1 = \frac{1}{n}\sum_{j=1}^n Z_j^2 - \frac{1}{2} + o_p(n^{-1/2}).$$

The result follows.

References

[1] Arcones, M. A. and Wang, Y. (2006). Some new tests for normality based on U-processes. *Statist. Probab. Lett.* **76**, 69–82.

[2] Baringhaus, L. and Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* **35**, 339–348.

[3] Brownrigg, R. and Khmaladze, E. (2010). Strange facts about the marginal distributions of processes based on the Ornstein–Uhlenbeck process. To appear in: *J. Comput. Finance.*

[4] Chen, L. and Shapiro, S. (1995). An alternative test for normality based on normalized spacings. *J. Statist. Comput. Simulation* **53**, 269–288.

[5] Csörgő, S. (1986). Testing for normality in arbitrary dimension. *Ann. Statist.* **14**, 708–723.

[6] Du, J. and Schick, A. (2007). Root-n consistency and functional central limit theorems for estimators of derivatives of convolutions of densities. *Internat. J. Statist. Management Systems* **2**, 67–87.

[7] Epps, T. W. and Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika* **70**, 723–726.

[8] Frees, E.W. (1994) . Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89**, 517–525.

[9] Giné, E. and Mason, D. M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.* **35**, 1105–1145.

[10] Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics. Some Basic Concepts.* 2nd ed. Springer, New York.

[11] Lilliefors, H. W. (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Amer. Statist. Assoc.* **62**, 399–402.

[12] Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average processes. *Canad. J. Statist.* **28**, 799-815.

[13] Schick, A. and Wefelmeyer, W. (2004). Root-n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.* **16**, 925-935.

[14] Schick, A. and Wefelmeyer, W. (2007). Root-n consistent density estimators of convolutions in weighted L_1-norms. *J. Statist. Plann. Inference* **137**, 1765–1774.

[15] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality: Complete samples. *Biometrika* **52**, 591–611.

[16] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman & Hall, London.

Anton Schick, Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902

Yishi Wang, Department of Mathematics and Statistics, University of North Carolina, Wilmington, NC 28403

Wolfgang Wefelmeyer, Mathematisches Institut, Universität zu Köln, 50931 Köln, Germany.