Ann Inst Stat Math (2018) 70:969–982
https://doi.org/10.1007/s10463-017-0614-0

Smoothed nonparametric tests and approximations of p-values

Yoshihiko Maesono[1] · **Taku Moriyama**[2] · **Mengxin Lu**[2]

Received: 21 December 2015 / Revised: 11 July 2017 / Published online: 23 August 2017

Abstract We propose new smoothed sign and Wilcoxon's signed rank tests that are based on kernel estimators of the underlying distribution function of the data. We discuss the approximations of the p-values and asymptotic properties of these tests. The new smoothed tests are equivalent to the ordinary sign and Wilcoxon's tests in the sense of Pitman's asymptotic relative efficiency, and the differences between the ordinary and new tests converge to zero in probability. Under the null hypothesis, the main terms of the asymptotic expectations and variances of the tests do not depend on the underlying distribution. Although the smoothed tests are not distribution-free, making use of the specific kernel enables us to obtain the Edgeworth expansions, being free of the underlying distribution.

Keywords Edgeworth expansion · Kernel estimator · Sign test · Significance probability · Wilcoxon's signed rank test

✉ Yoshihiko Maesono
maesono@math.kyushu-u.ac.jp

Taku Moriyama
moritaku3542168@gmail4.com

Mengxin Lu
lumengxin1989@hotmail.co.jp

1 Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

2 Graduate School of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed (*i.i.d.*) random variables with a distribution function $F(x-\theta)$, where the associated density function satisfies $f(-x) = f(x)$ and θ is an unknown location parameter. Here, we consider a test and the confidence interval of the parameter θ. This setting is called a one-sample location problem. Numerous nonparametric test statistics have been proposed to test the null hypothesis $H_0 : \theta = 0$ vs. $H_1 : \theta > 0$, e.g., the sign test and Wilcoxon's signed rank test (see Hájek et al. 1999). These tests are distribution-free and have discrete distributions. As pointed out by Lehmann and D'abrera (2006) and Brown et al. (2001), because of the discreteness of the test statistics, the p-values jump in response to a small change in data values when the sample size n is small or moderate. Brown et al. (2001) discussed a smoothed version of the sign test and then obtained an Edgeworth expansion. In particular, they proposed a smoothed median estimator and a corresponding smoothed sign test. The test was, however, not distribution-free. Their smoothed sign test has good properties, but its Pitman's asymptotic relative efficiency ($A.R.E.$) did not coincide with that of the ordinary sign test. Furthermore, to use the Edgeworth expansion, they needed estimators of the unknown parameters.

In this paper, we first consider another smoothed sign test that is based on a kernel estimator of the distribution function and examine its asymptotic properties. We show that the difference between the two sign tests converges to zero in probability. In addition, we obtain an Edgeworth expansion, being free of the underlying distribution. Next, we propose a smoothed Wilcoxon's signed rank test. We show that the difference between the two Wilcoxon's tests converges to zero in probability as well and study an Edgeworth expansion.

Let us define the indicator function $I(A) = 1$ (if A occurs), $= 0$ (if A fails); the sign test is equivalent to

$$S = S(\boldsymbol{X}) = \sum_{i=1}^{n} I(X_i \geq 0),$$

where $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$. Wilcoxon's signed rank test is equivalent to the Mann–Whitney test:

$$W = W(\boldsymbol{X}) = \sum_{1 \leq i \leq j \leq n} I(X_i + X_j \geq 0).$$

Now, put $s = S(\boldsymbol{x})$ and $w = W(\boldsymbol{x})$ for observed values $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$. If the p-value $P_0(S \geq s)$ ($P_0(W \geq w)$), where $P_0(\cdot)$ denotes a probability under the null hypothesis H_0, is small enough, we reject the null hypothesis H_0.

Moreover, let us define

$$\Omega_{|x|} = \left\{ \boldsymbol{x} \in \boldsymbol{R}^n \parallel |x_1| < |x_2| < \cdots < |x_n| \right\}$$

and

$$\Omega_\alpha = \left\{ \boldsymbol{x} \in \Omega_{|x|} \;\middle\|\; \frac{s - E_0(S)}{\sqrt{V_0(S)}} \geq z_{1-\alpha}, \quad \text{or} \quad \frac{w - E_0(W)}{\sqrt{V_0(W)}} \geq z_{1-\alpha} \right\},$$

Table 1 Number of samples in which S and W have comparatively smaller p-values

Sample size	$n = 10$	$n = 20$	$n = 30$
$z_{0.90}$			
S	25	69,080	59,092,679
W	82	94,442	87,288,529
W/S	3.28	1.367	1.477
$z_{0.95}$			
S	25	32,705	30,857,108
W	48	47,387	43,957,510
W/S	1.92	1.449	1.425
$z_{0.975}$			
S	5	12,704	14,028,374
W	21	21,267	22,049,240
W/S	4.2	1.674	1.572

where $z_{1-\alpha}$ is the $(1-\alpha)th$ quantile of the standard normal distribution $N(0,1)$, and $E_0(\cdot)$ and $V_0(\cdot)$ are, respectively, the expectation and variance under H_0. The observed values $S(\boldsymbol{x})$ and $W(\boldsymbol{x})$ are invariant under a permutation of $x_1, \ldots, x_n$, so it is sufficient to consider the case that $|x_1| < |x_2| < \cdots < |x_n|$; there are 2^n times combinations of $\text{sign}(x_i) = \pm 1 (i = 1, \ldots, n)$. We count samples in which the exact p-value of one test is smaller than the p-value of the other test in the tail area Ω_α. Table 1 shows the number of samples in which the p-value of S (W) is smaller than that of W (S) in the tail area. In Table 1, row S indicates a number of samples in which the p-value of S is smaller than that of W, and row W means the number of samples in which the p-value of W is smaller than that of S. W/S is the ratio of W and S. For each sample, there is one tie of p-values.

Remark 1 Table 1 shows that W is preferable if one wants a small p-value and that S is preferable if one does not want to reject the null hypothesis H_0. Thus, a practitioner could make an arbitrary choice of the test statistics. This problem comes from the discreteness of the distributions of the test statistics.

On the other hand, it is possible to use an estimator of $F(0)$ as a test statistic. Define the empirical distribution function by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x).$$

Then, $F_n(0)$ is equivalent to the sign test S, that is,

$$S = n - nF_n(0-).$$

As usual, a kernel estimator $\widetilde{F}_n$ can be used to get a smooth estimator of the distribution function. It is natural to use $\widetilde{F}_n(0)$ as a smoothed sign test. Let k be a kernel function that satisfies

$$\int_{-\infty}^{\infty} k(u)\mathrm{d}u = 1,$$

and write K be the integral of k,

$$K(t) = \int_{-\infty}^{t} k(u)\mathrm{d}u.$$

The kernel estimator of $F(x)$ is defined by

$$\widetilde{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where h_n is a bandwidth that satisfies $h_n \to 0$ and $nh_n \to \infty$ $(n \to \infty)$. We can use

$$\widetilde{S} = n - n\widetilde{F}_n(0) = n - \sum_{i=1}^{n} K\left(-\frac{X_i}{h_n}\right)$$

for testing H_0 and can regard $\widetilde{S}$ as the smoothed sign test. Under H_0, the main terms of the asymptotic expectation and variance of $\widetilde{S}$ do not depend on F; i.e., they are asymptotically distribution-free. Furthermore, we can obtain an Edgeworth expansion, being free of F.

We can construct a smoothed Wilcoxon's signed rank test in a similar fashion. Since the main term of the Mann–Whitney statistic can be regarded as an estimator of the probability $P(\frac{X_1+X_2}{2} > 0)$, we propose the following smoothed test statistic:

$$\widetilde{W} = \frac{n(n+1)}{2} - \sum_{1\leq i\leq j\leq n} K\left(-\frac{X_i + X_j}{2h_n}\right).$$

The smoothed test $\widetilde{W}$ is not distribution-free. However, under H_0, the asymptotic expectation and variance do not depend on F, and we can obtain the Edgeworth expansion of $\widetilde{W}$. The resulting Edgeworth expansion does not depend on F if we use a symmetric fourth-order kernel and bandwidth of $h_n = o(n^{-1/4})$.

Also, we will show that the difference between the standardized S and $\widetilde{S}$ and the difference between the standardized W and $\widetilde{W}$ go to zero in probability. Accordingly, the smoothed test statistics are equivalent in the sense of the first-order asymptotic.

The rest of this paper is organized as follows. In Sect. 2, we discuss the asymptotic properties of $\widetilde{S}$ and $\widetilde{W}$ and obtain the Edgeworth expansions with $o(n^{-1})$ residual terms. The confidence intervals of θ based on $\widetilde{S}$ and $\widetilde{W}$ are also discussed. In Sect. 3, we discuss the selection of the bandwidth and kernel function. In Sect. 4, we study higher order approximations. Some proofs are given in Appendix (the complete proofs are found in Maesono et al. (2016)).

2 Asymptotic properties of smoothed tests

We assume that the kernel k is symmetric, i.e., $k(-u) = k(u)$. Using the properties of the kernel estimator, we can obtain expectations $E_\theta(\widetilde{S})$, $E_\theta(\widetilde{W})$ and variances $V_\theta(\widetilde{S})$, $V_\theta(\widetilde{W})$. Because of the symmetry of the underlying distribution f and the kernel k, we get

$$F(-x) = 1 - F(x) \quad \text{and} \quad \int_{-\infty}^{\infty} uk(u)\mathrm{d}u = 0.$$

Let us define

$$e_1(\theta) = E_\theta\left[1 - K\left(-\frac{X_1}{h_n}\right)\right].$$

Using the transformation $u = -x/h_n$, integration by parts, and a Taylor-series expansion, we get

$$\begin{aligned} e_1(\theta) &= 1 - \int_{-\infty}^{\infty} K(u) f(-\theta - h_n u)\frac{1}{h_n}\mathrm{d}u \\ &= 1 - \int_{-\infty}^{\infty} k(u) F(-\theta - h_n u)\mathrm{d}u \\ &= F(\theta) + O(h_n^2), \end{aligned}$$

which yields

$$E_\theta(\widetilde{S}) = n\left\{F(\theta) + O\left(h_n^2\right)\right\}.$$

Similarly, we have

$$E_\theta\left[K^2\left(-\frac{X_1}{h_n}\right)\right] = F(-\theta) + O(h_n),$$

hence,

$$V_\theta(\widetilde{S}) = n\left[\{1 - F(\theta)\}F(\theta) + O(h_n)\right].$$

On the other hand, since $\widetilde{W}$ takes the form of the U-statistic, we can use asymptotic properties of the U-statistic (see Lee 1990). The expectation and variance of $\widetilde{W}$ are given by

$$\begin{aligned} E_\theta(\widetilde{W}) &= \tfrac{n(n+1)}{2}\left\{G(\theta) + O\left(h_n^2\right)\right\}, \\ V_\theta(\widetilde{W}) &= n(n+1)^2\left\{\textstyle\int_{-\infty}^{\infty} F^2(u + 2\theta) f(u)\mathrm{d}u - G^2(\theta) + O\left(h_n^2\right)\right\}, \end{aligned}$$

where

$$G(\theta) = \int_{-\infty}^{\infty} F(2\theta + u)f(u)\mathrm{d}u$$

is the distribution function of $(X_1 + X_2)/2$.

Direct computations yield the following theorem.

Theorem 1 *Let us assume that f' exists and is continuous in a neighborhood of $-\theta$, and $h_n = cn^{-d}(c > 0, \frac{1}{4} < d < \frac{1}{2})$. If*

$$0 < \lim_{n\to\infty} V_\theta \left[1 - K\left(-\frac{X_1}{h_n}\right)\right] < \infty,$$
$$0 < \lim_{n\to\infty} Cov_\theta \left[1 - K\left(-\frac{X_1+X_2}{2h_n}\right),\quad 1 - K\left(-\frac{X_1+X_3}{2h_n}\right)\right] < \infty,$$

and the kernel k is symmetric around zero, then,

$$\lim_{n\to\infty} E_\theta \left\{\frac{S - E_\theta(S)}{\sqrt{V_\theta(S)}} - \frac{\widetilde{S} - E_\theta(\widetilde{S})}{\sqrt{V_\theta(\widetilde{S})}}\right\}^2 = 0,$$
$$\lim_{n\to\infty} E_\theta \left\{\frac{W - E_\theta(W)}{\sqrt{V_\theta(W)}} - \frac{\widetilde{W} - E_\theta(\widetilde{W})}{\sqrt{V_\theta(\widetilde{W})}}\right\}^2 = 0.$$

Since S and W are asymptotically normal, $\widetilde{S}$ and $\widetilde{W}$ are also asymptotically normal. Pitman's $A.R.E$.s of $\widetilde{S}$ and $\widetilde{W}$ coincide with S and W, respectively.

For the sign test S, it is difficult to improve the normal approximation because of the discreteness of the distribution function of S. The standardized sign test S takes values with jump order $n^{-1/2}$, so we cannot prove the validity of the formal Edgeworth expansion. On the other hand, since $\widetilde{S}$ is a smoothed statistic and has a continuous type distribution, we can obtain an Edgeworth expansion and prove its validity. García-Soidán et al. (1997) discussed the Edgeworth expansion and proved its validity for the kernel estimators. Huang and Maesono (2014) obtained an explicit formula when $h_n = cn^{-d}$ $(c > 0,\ \frac{1}{4} < d < \frac{1}{2})$. Bickel et al. (1986) proved the validity of the Edgeworth expansion of the U-statistic with an $o(n^{-1})$ residual term. Since the standardized W and $\widetilde{W}$ are asymptotically equivalent, we can obtain the Edgeworth expansion of $\widetilde{W}$. The resulting Edgeworth approximations do not depend on the underlying distribution F, if we use the fourth-order kernel, i.e.,

$$\int u^\ell k(u)\mathrm{d}u = 0\ (\ell = 1, 2, 3) \quad \text{and} \quad \int u^4 k(u)\mathrm{d}u \neq 0.$$

Using the results of García-Soidán et al. (1997) and Huang and Maesono (2014), we can prove the following theorem.

Theorem 2 *Let us assume that the conditions of Theorem 1 hold and the kernel k is symmetric. If* $|f'(x)| \le M (M > 0)$, $\int |u^4 k(u)| \mathrm{d}u < \infty$ *and the bandwidth satisfies* $h_n = cn^{-d}$ $(c > 0,\ \frac{1}{4} < d < \frac{1}{2})$, *then,*

$$P_0\left(\frac{\widetilde{S} - E_0(\widetilde{S})}{\sqrt{V_0(\widetilde{S})}} \le y\right) = \Phi(y) - \frac{1}{24n}(y^3 - 3y)\phi(y) + o(n^{-1}),$$

$$P_0\left(\frac{\widetilde{W} - E_0(\widetilde{W})}{\sqrt{V_0(\widetilde{W})}} \le y\right) = \Phi(y) - \left(\frac{7}{20}y^3 - \frac{21}{20}y\right)\phi(y) + o(n^{-1}).$$

The Edgeworth expansions in Theorem 2 do not depend on the underlying distribution F. However, in order to use the normal approximations or the Edgeworth expansions, we have to obtain approximations of $E_0(\widetilde{S})$, $V_0(\widetilde{S})$, $E_0(\widetilde{W})$ and $V_0(\widetilde{W})$. Let us define

$$A_{i,j} = \int_{-\infty}^{\infty} K^i(u)k(u)u^j \mathrm{d}u.$$

We have the following higher order approximations of the expectations and variances under the null hypothesis H_0.

Theorem 3 *Let us assume that the kernel is symmetric, and let* M_1, M_2 *and* M_3 *be positive constants. If exactly one of the following conditions holds:* (a) $|f^{(5)}(x)| \le M_1$ *and* $h_n = o(n^{-1/4})$, (b) $|f^{(4)}(x)| \le M_2$ *and* $h_n = o(n^{-3/10})$, (c) $|f^{(3)}(x)| \le M_3$ *and* $h_n = o(n^{-1/3})$, *then,*

$$E_0(\widetilde{S}) = \tfrac{n}{2} + o(n^{-1/2}),$$

$$V_0(\widetilde{S}) = \tfrac{n}{4} - 2nh_n f(0)A_{1,1} - \tfrac{nh_n^3}{3} f''(0)A_{1,3} + o(1),$$

$$E_0(\widetilde{W}) = \tfrac{n(n+1)}{4} + o(n^{1/2}), \tag{1}$$

$$V_0(\widetilde{W}) = \tfrac{n^2(2n+3)}{24} - 4n^3h_n^2 A_{0,2} \textstyle\int_{-\infty}^{\infty} \{f(x)\}^3 \mathrm{d}x + o(n^2). \tag{2}$$

Remark 2 In order to get the above approximations, we used a Taylor-series expansion of the integral. We can divide up the integral at discrete points, so we do not need to worry about the differentiability of the density function at finite number of points.

3 Selection of bandwidth and kernel function

We discuss the selection of the bandwidth and the kernel function. Azzalini (1981) recommended a bandwidth of $cn^{-1/3}$ for the estimation of the distribution function. Actually, we compared several bandwidths in simulation studies and found that the approximations were not good when the convergence rate of the bandwidth was slower than $n^{-1/3}$. When the convergence rate of the bandwidth was faster than $n^{-1/3}$, the

Table 2 p-value approximations of smoothed sign $\widetilde{S}$ with $k_{e,2}$ and $k_{e,4}$ ($h_n = n^{-1/3}$)

$\alpha = 0.05$		$n = 10$	$n = 20$	$n = 30$
$k_{e,2}$	N(0,1)	0.03437	0.03803	0.03834
$k_{e,4}$	N(0,1)	0.05224	0.05397	0.05374
$k_{e,2}$	Logis.	0.04038	0.04256	0.04293
$k_{e,4}$	Logis.	0.05382	0.05395	0.05347
$k_{e,2}$	D.exp	0.02603	0.03062	0.03193
$k_{e,4}$	D.exp	0.04272	0.04468	0.04797

approximations were similar to the case of $n^{-1/3}$. Thus, hereafter, we will use the bandwidth $h_n = n^{-1/3}$. Epanechnikov (1969) showed the optimality of

$$k_{e,2}(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$$

as a kernel for the estimation of the density function ($k_{e,2}$ is a second-order kernel). We observed from several simulation studies that the second-order kernels ($A_{0,1} = 0$, $A_{0,2} \neq 0$) do not give good approximations of the p-values. We compare normal approximations of $\widetilde{S}$ based on $k_{e,2}(u)$ and a modified fourth-order kernel

$$k_{e,4}(u) = \frac{15}{8}\left(1 - \frac{7}{3}u^2\right)k_{e,2}(u)$$

with the bandwidth $h_n = n^{-1/3}$. We simulated the following probabilities from 100,000 random samples from normal, logistic, and double exponential distributions:

$$P_0\left(\frac{\widetilde{S} - n/2}{\sqrt{n/4}} \geq z_{1-\alpha}\right).$$

Table 2 shows that the fourth-order kernel gives good approximations. Note that in the table, the normal distribution is denoted as $N(0, 1)$, the logistic distribution as Logis., and the double exponential distribution as D.Exp. We performed similar simulations using a Gaussian kernel and obtained similar results. The choice of kernel function does not affect the approximations of the p-values, but the order degree of the kernel is important, as expected. Although the fourth-order kernels loose the monotonicity of the distribution functions of $\widetilde{S}$ and $\widetilde{W}$, they loose monotonicity at most 3% points of x when $n = 10$.

Remark 3 If we use the normal approximations of the standardized $\widetilde{S}$ and $\widetilde{W}$, $A_{1,1}$ affects the approximations. We recommend the fourth-order kernel because the value of $A_{1,1}$ with the fourth-order kernel is much smaller than that of the second-order one.

Since the distributions of $\widetilde{S}$ and $\widetilde{W}$ depend on F, we compared the significance probabilities of $\widetilde{S}$ and $\widetilde{W}$ in simulations. We used the kernel $k_{e,4}$ and bandwidth $h_n =$

Table 3 Number of samples in which S and W have comparatively smaller p-values ($k_{e,4},\ h_n = n^{-1/3}$)

Sample size	$n = 10$	$n = 20$	$n = 30$
$z_{0.90}$			
$\widetilde{S}$	5658	5978	6142
$\widetilde{W}$	7263	7066	7050
$\widetilde{W}/\widetilde{S}$	1.284	1.182	1.148
$z_{0.95}$			
$\widetilde{S}$	2921	3017	3164
$\widetilde{W}$	3407	3616	3515
$\widetilde{W}/\widetilde{S}$	1.166	1.199	1.111
$z_{0.975}$			
$\widetilde{S}$	1133	1440	1599
$\widetilde{W}$	1628	1785	1780
$\widetilde{W}/\widetilde{S}$	1.437	1.240	1.113

Table 4 Closeness of p-values of S & $\widetilde{S}$, and W & $\widetilde{W}$ ($k_{e,4},\ h_n = n^{-1/3}$)

$\alpha = 0.05$		$n = 10$	$n = 20$	$n = 30$		$n = 10$	$n = 20$	$n = 30$
S	N(0,1)	0.05474	0.05867	0.04950	D.exp	0.05446	0.05617	0.04947
$\widetilde{S}$	N(0,1)	0.04981	0.05358	0.04548	D.exp	0.04434	0.04915	0.04187
W	N(0,1)	0.05271	0.04797	0.05092	D.exp	0.05271	0.04942	0.04998
$\widetilde{W}$	N(0,1)	0.05226	0.04864	0.05046	D.exp	0.05117	0.04879	0.04962

$n^{-1/3}$. We estimated the significance probabilities in the tail area $\widetilde{\Omega}_\alpha$ from 100,000 random samples from a normal distribution:

$$\widetilde{\Omega}_\alpha = \left\{ \boldsymbol{x} \in \boldsymbol{R}^n \;\middle\|\; \frac{\widetilde{s}(\boldsymbol{x}) - E_0(\widetilde{S})}{\sqrt{V_0(\widetilde{S})}} \geq z_{1-\alpha}, \quad \text{or} \quad \frac{\widetilde{w}(\boldsymbol{x}) - E_0(\widetilde{W})}{\sqrt{V_0(\widetilde{W})}} \geq z_{1-\alpha} \right\}.$$

For the simulated sample $\boldsymbol{x} \in \boldsymbol{R}^n$, we calculated the p-values based on the normal approximation. In Table 3, $\widetilde{S}$ means that the p-values of $\widetilde{S}$ are smaller than that of $\widetilde{W}$, etc. Comparing Tables 1 and 3, we can see that the differences between the p-values of $\widetilde{S}$ and $\widetilde{W}$ are smaller than those of S and W.

Next we checked how close the ordinary and smoothed tests are to each other when the sample size n is small. Since S has a discrete distribution, we chose a nearest value α' to 0.05, i.e., $P(S \geq s_{\alpha'}) = \alpha' \approx 0.05$. After that, we simulated the p-values $P(S \geq s_{\alpha'})$ and $P(\widetilde{S} \geq \frac{n}{2} + \frac{\sqrt{n}}{2} z_{1-\alpha'})$ from 100,000 repetitions for underlying normal $(N(0, 1))$ and double exponential (D.exp) distributions.

Table 4 shows that the difference between the smoothed and ordinary sign tests are small, so we can regard $\widetilde{S}$ as a smoothing statistic of S. We got similar results for $\widetilde{W}$ and W.

4 Higher order approximation

We discuss higher order approximations based on Edgeworth expansions. If the conditions of $A_{1,1}=0$ and $A_{1,3}=0$ hold, we can use the Edgeworth expansion of $\widetilde{S}$. If the kernel is fourth-order symmetric, $A_{0,2}=0$ and we can use the Edgeworth expansion of $\widetilde{W}$. The conditions of $A_{1,1}=0$ and $A_{1,3}=0$ seem restrictive, but we can still construct the desired kernel. Let us define

$$k^*(u)=\left(\frac{1}{4}(\sqrt{105}-3)+\frac{1}{2}(5-\sqrt{105})|u|\right)I(|u|\leq 1),$$

which is fourth-order symmetric with $A_{1,1}=0$. This kernel k^* may take a negative value, and hence, $\widetilde{F}_n(x)$ is not monotone as a function of x. However, our main purpose is to test the null hypothesis H_0 and to construct the confidence interval; that means we do not need to worry about it. As mentioned above, $\widetilde{F}_n(x)$ looses monotonicity at most 3% of its points around the origin when $n=10$. For the smoothed Wilcoxon's rank test, we need only assume that the kernel k is fourth-order symmetric. While it is theoretically possible to construct a polynomial-type kernel that satisfies $A_{1,1}=A_{1,3}=0$, it is rather complicated to do so, and it takes a couple of pages to write out the full form. Thus, we will only consider the kernel k^* here. It may be possible to construct another type of kernel that satisfies $A_{1,1}=0$ and $A_{1,3}=0$. We postpone this endeavor to a future work.

If the equation

$$V_0(\widetilde{S})=\frac{n}{4}+o(1)$$

holds, we can use the Edgeworth expansion of $\widetilde{S}$ for testing H_0 and constructing a confidence interval without making any estimators. We can get an approximation of the α-quantile ($P_0(\widetilde{S}\leq\widetilde{s}_\alpha)=\alpha+o(n^{-1})$), i.e.,

$$\widetilde{s}_\alpha=\frac{n}{2}+\frac{\sqrt{n}}{2}z_\alpha+\frac{1}{48\sqrt{n}}\left(z_\alpha^3-3z_\alpha\right). \tag{3}$$

For the significance level $0<\alpha<1$, if the observed value $\widetilde{s}$ satisfies $\widetilde{s}\geq\widetilde{s}_{1-\alpha}$, we reject the null hypothesis H_0. Since the distribution function of $X_i-\theta$ is $F(x)$, we can construct the confidence interval of θ by using Eq. (3). For the observed value $\boldsymbol{x}=(x_1,\ldots,x_n)$, let us define

$$\widetilde{s}(\theta|\boldsymbol{x})=n-\sum_{i=1}^{n}K\left(\frac{\theta-x_i}{h_n}\right),$$
$$\widehat{\theta}_U=\arg\min_\theta\left\{\widetilde{s}(\theta|\boldsymbol{x})\leq\widetilde{s}_{\alpha/2}\right\},$$

and

$$\widehat{\theta}_L=\arg\max_\theta\left\{\widetilde{s}_{1-\alpha/2}\leq\widetilde{s}(\theta|\boldsymbol{x})\right\},$$

Table 5 Comparison of normal approximation and Edgeworth expansion with the kernel $A_{1,1} = 0$ ($k^*,\ h_n = n^{-1/3}(\log n)^{-1}$)

$\widetilde{s}$	$n = 30$	$A_{1,1} = 0$		$\widetilde{s}$	$n = 30$	$A_{1,1} = 0$	
$z_{0.99}$	True	Edge.	Nor.	$z_{0.95}$	True	Edge.	Nor.
N(0,1)	0.00842	0.01021	$\underline{0.01}$	N(0,1)	0.05013	0.04993	$\underline{0.05}$
Logis.	0.00937	0.01021	$\underline{0.01}$	Logis.	0.0491	$\underline{0.04993}$	0.05
D. exp.	0.00908	0.01021	$\underline{0.01}$	D.Exp.	0.04903	$\underline{0.04993}$	0.05
$\widetilde{s}$	n=100	$A_{1,1} = 0$		$\widetilde{s}$	n=100	$A_{1,1} = 0$	
$z_{0.99}$	True	Edge.	normal	$z_{0.95}$	True	Edge.	Nor.
N(0,1)	0.00962	0.01006	$\underline{0.01}$	N(0,1)	0.04903	$\underline{0.04998}$	0.05
Logis.	0.00954	0.01006	$\underline{0.01}$	Logis.	0.04892	$\underline{0.04998}$	0.05
D. exp.	0.0099	$\underline{0.01006}$	0.01	D.Exp.	0.04937	$\underline{0.04998}$	0.05

where $0 < \alpha < 1$. The $1 - \alpha$ confidence interval is given by $\widehat{\theta}_L \le \theta \le \widehat{\theta}_U$.

Similarly, if the observed value $\widetilde{w}$ satisfies $\widetilde{w} \ge \widetilde{w}_{1-\alpha}$, we reject the null hypothesis H_0, where

$$\widetilde{w}_\alpha = \frac{n(n+1)}{4} + \frac{n\sqrt{2n+3}}{2\sqrt{6}} \left\{ z_\alpha + \frac{1}{n}\left(\frac{7}{20} z_\alpha^3 - \frac{21}{20} z_\alpha \right) \right\}. \tag{4}$$

Using $\widetilde{w}_\alpha$ in (4), we can construct the confidence interval of θ. For the observed value $\boldsymbol{x} = (x_1, \ldots, x_n)$, let us define

$$\widetilde{w}(\theta|\boldsymbol{x}) = \frac{n(n+1)}{2} - \sum_{1 \le i \le j \le n} K\left(\frac{2\theta - x_i - x_j}{2h_n} \right),$$
$$\widehat{\theta}_U^* = \arg\min_\theta \left\{ \widetilde{w}(\theta|\boldsymbol{x}) \le \widetilde{w}_{\alpha/2} \right\}$$

and

$$\widehat{\theta}_L^* = \arg\max_\theta \left\{ \widetilde{w}_{1-\alpha/2} \le \widetilde{w}(\theta|\boldsymbol{x}) \right\}.$$

Thus, we have the $1 - \alpha$ confidence interval $\widehat{\theta}_L^* \le \theta \le \widehat{\theta}_U^*$.

Table 5 compares the simple normal approximation and Edgeworth expansion using the kernel k^* and the bandwidth $h_n = n^{-1/3}(\log n)^{-1}$. Since we do not know exact distributions of the smoothed sign test $\widetilde{S}$, we estimated the values $P(\frac{\widetilde{S}-E_0(\widetilde{S})}{\sqrt{V_0(\widetilde{S})}} \ge z_{1-\alpha})$ from 100,000 replications of the data and denote them as "True" in the table. "Edge." and "Nor." denote the Edgeworth and simple normal approximations, respectively. The underlying distributions are normal, logistic, and double exponential ones. The double exponential distribution is not differentiable at the origin (zero), but as mentioned in Remark 2, we don't have to worry about that.

Table 6 Coverage probabilities of S, $\widetilde{S}$, W and $\widetilde{W}$ $(k^*,\ h_n = n^{-1/3}(\log n)^{-1})$

$n = 10$, 90%	$s(\theta\|\boldsymbol{x})$	$\widetilde{s}(\theta\|\boldsymbol{x})$	$w(\theta\|\boldsymbol{x})$	$\widetilde{w}(\theta\|\boldsymbol{x})$
N(0,1)	0.9760	0.8910	0.9145	0.8910
Logis.	0.9788	0.8990	0.9158	0.8970
D.exp	0.9806	0.8938	0.9189	0.9024
$n = 20$, 90%	$s(\theta\|\boldsymbol{x})$	$\widetilde{s}(\theta\|\boldsymbol{x})$	$w(\theta\|\boldsymbol{x})$	$\widetilde{w}(\theta\|\boldsymbol{x})$
N(0,1)	0.9610	0.8921	0.9035	0.8982
Logis.	0.9610	0.8922	0.9026	0.8971
D.exp	0.9589	0.8966	0.9055	0.8984

Remark 4 If we use a symmetric fourth-order kernel, which satisfies $A_{1,1} = 0$, the $n^{-1/2}$ term of the Edgeworth expansion is zero, and hence, the simple normal approximation means that the residual term is already $o(n^{-1/2})$. Comparing the $n^{-1/2}$ terms, we can see that the effect of the n^{-1} term is small; thus, the Edgeworth expansion with the $o(n^{-1})$ residual term is comparable to the simple normal approximation when the sample size n is small. When the sample size is large, the Edgeworth approximation is better.

Finally, we simulated the coverage probabilities based on S, $\widetilde{S}$, W, and $\widetilde{W}$, by making 100,000 repetitions. We used the intervals $\widehat{\theta}_L \le \theta \le \widehat{\theta}_U$ and $\widehat{\theta}^*_L \le \theta \le \widehat{\theta}^*_U$, where the kernel k^* and $h_n = n^{-1/3}(\log n)^{-1}$. For S and W, we constructed conservative confidence intervals whose coverage probabilities are equal to or greater than $1 - \alpha$ when the sample size is small. Table 6 shows that the coverage probabilities of the smoothed statistics are less conservative.

Remark 5 If the sample size n is large enough, the higher order approximation works well. In that case, we recommend the Edgeworth expansion with the pair $(k^*,\ h_n = n^{-1/3}(\log n)^{-1})$. If the sample size is moderate, the normal approximation based on the pair $(k_{e,4},\ h_n = n^{-1/3})$ works well. In that case, we recommend the fourth-order kernel and bandwidth $n^{-1/3}$.

Acknowledgements The authors would like to thank the editor and two anonymous referees for their careful reading and valuable comments, which helped us to improve the manuscript significantly. The authors gratefully acknowledge JSPS KAKENHI Grant Nos. JP15K11995 and JP16H02790.

5 Appendix

Appendix gives brief proofs of Theorems 1 and 3. The complete proofs are found in Maesono et al. (2016).

Proof of Theorem 1 For the ordinary sign test S, we have

$$V_\theta(S) = nF(\theta)[1 - F(\theta)].$$

Then, it is sufficient to show that

$$E_\theta\left[\{S - F(\theta)\}\left\{\widetilde{S} - E_\theta(\widetilde{S})\right\}\right] = n\left\{F(\theta)[1 - F(\theta)] + O(h_n)\right\}.$$

Since S and $\widetilde{S}$ are sums of $i.i.d.$ random variables, we have

$$\begin{aligned} &E_\theta\left[\{S - E_\theta(S)\}\left\{\widetilde{S} - E_\theta(\widetilde{S})\right\}\right] \\ &\quad = nE_\theta\left[\{I(X_1 \geq 0) - E_\theta(I(X_1 \geq 0))\}\left\{1 - K\left(-\frac{X_1}{h_n}\right) - e_1(\theta)\right\}\right]. \end{aligned}$$

Using the transformation $u = x/h_n$, integration by parts, and a Taylor expansion, we get

$$\int_{-\infty}^{\infty} I(x \geq 0)\left[1 - K\left(-\frac{x}{h_n}\right)\right] f(x - \theta)\mathrm{d}x = F(\theta) + O(h_n).$$

Since $E_\theta(I(x \geq 0)) = F(\theta)$ and $E_\theta(1 - K) = F(\theta) + O(h_n^2)$, we have

$$E_\theta\left[\{S - E_\theta(S)\}\left\{\widetilde{S} - E_\theta(\widetilde{S})\right\}\right] = n\{F(\theta) - [F(\theta)]^2 + O(h_n)\}.$$

Thus, we get the desired result. □

Similarly, we can show that the difference between W and $\widetilde{W}$ goes to zero.

Proof of Theorem 3 Assuming that the density f is differentiable, we have

$$\begin{aligned} \frac{1}{n}E_0(\widetilde{S}) &= 1 - \int_{-\infty}^{\infty} K\left(-\frac{x}{h_n}\right) f(x)\mathrm{d}x = 1 - \int_{-\infty}^{\infty} k(u)F(-h_n u)\mathrm{d}u \\ &= 1 - F(0) + h_n f(0)A_{0,1} - \frac{h_n^2}{2} f'(0)A_{0,2} + \frac{h_n^3}{6} f''(0)A_{0,3} \\ &\quad - \frac{h_n^4}{24} f^{(3)}(0)A_{0,4} + \frac{h_n^5}{120} f^{(4)}(0)A_{0,5} + O(h_n^6). \end{aligned}$$

Similarly, we can show that

$$E_0\left\{K^2\left(-\frac{X_1}{h_n}\right)\right\} = F(0) - 2h_n f(0)A_{1,1} + h_n^2 f'(0)A_{1,2} - \frac{h_n^3}{3} f''(0)A_{1,3} + O(h_n^4)$$

and

$$\begin{aligned} \frac{1}{n}V_0(\widetilde{S}) = F(0)\{1 - F(0)\} - 2h_n f(0)A_{1,1} + h_n^2 f'(0)\{A_{1,2} - F(0)A_{0,2}\} \\ - \frac{h_n^3}{3} f''(0)\{A_{1,3} - F(0)A_{0,3}\} + O(h_n^4). \end{aligned}$$

Note that $k(-u) = k(u)$ yields $A_{0,1} = A_{0,3} = A_{0,5} = 0$. Furthermore, since $f(-x) = f(x)$, we get

$$f'(0) = 0, \quad f''(-x) = f''(x), \quad f^{(3)}(-x) = -f^{(3)}(x), \quad \text{and} \quad f^{(3)}(0) = 0.$$

We can derive equations (1) and (2) in a similar manner (see Maesono et al. 2016). □

References

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, *68*(1), 326–328.

Bickel, P., Götze, F., Van Zwet, W. (1986). The edgeworth expansion for u-statistics of degree two. *The Annals of Statistics*, *14*(4), 1463–1484.

Brown, B., Hall, P., Young, G. (2001). The smoothed median and the bootstrap. *Biometrika*, *88*(2), 519–534.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications*, *14*(1), 153–158.

García-Soidán, P. H., González-Manteiga, W., Prada-Sánchez, J. (1997). Edgeworth expansions for non-parametric distribution estimation with applications. *Journal of Statistical Planning and Inference*, *65*(2), 213–231.

Hájek, J., Šidák, Z., Sen, P. K. (1999). *Theory of rank tests*. San Diego: Academic press.

Huang, Z., Maesono, Y. (2014). Edgeworth expansion for kernel estimators of a distribution function. *Bulletin of Informatics and Cybernetics*, *46*, 1–10.

Lee, A. (1990). *U-statistics. Theory and practice*. New York: Marcel Dekker.

Lehmann, E. L., D'abrera, H. (2006). *Nonparametrics*: *Statistical methods based on ranks*. New York: Springer.

Maesono, Y., Moriyama, T., Lu, M. (2016). Smoothed nonparametric tests and their properties. *arXiv preprint* ArXiv:1610.02145