Z. Wahrscheinlichkeitstheorie verw. Gebiete
39, 235–255 (1977)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete
© by Springer-Verlag 1977

Tests for Symmetry

André Antille and Götz Kersting

Institut für Mathematische Statistik der Universität Göttingen,
Lotzestr. 13, D-3400 Göttingen, Federal Republic of Germany

I. Introduction

Tests for symmetry of a distribution function about an unknown value μ are investigated. Two quite different methods are presented and compared. The first procedure consists of applying a non-parametric test for symmetry around 0, after μ has been estimated. The sign and Wilcoxon statistics are considered, the mean or the median being our estimates for μ. The second method was proposed by T. Gasser and seems to be new: non-parametric tests are applied to the differences of symmetrically located intervals of order statistics. A trimmed sign test only is presented. Analogous results have been obtained for trimmed statistics of Wilcoxon type but such tests do not seem to be good. Further, the asymptotic variance depends on the underlying density in a rather complicated way. The reason for throwing away the $[\varepsilon n]$ smallest, resp., greatest observations is that unsymmetry in the tails of a distribution is irrelevant to practical purposes. The asymptotic variance $\sigma^2_{4,\varepsilon}$ of the trimmed sign statistic $S_{4,\varepsilon}$ possesses a limit σ^2_4 for ε tending to 0. It would be interesting to know whether $S_{4,0}$ is asymptotically normal with variance σ^2_4. We think this is the case but we have not been able to prove it. Technical difficulties arise because of the behaviour of the derivatives of F^{-1} near 0 and 1, F being the underlying distribution function.

For both methods the asymptotic distribution under symmetry is given. The statistics are no longer distribution free but surprisingly, the asymptotic variance, in some cases, varies little with the density. The asymptotic distributions for alternatives where the distribution function is of the form $F(x+n^{-1/2}g(x))$ are computed. Comparisons of the power are made in the case where $g(x)=x$, $x\geqq 0$, and 0 elsewhere. This corresponds to a contraction of the positive axis. We had also considered as alternatives the contamination model and the convolution of a symmetric distribution with an asymptotically negligible unsymmetric distribution. The results are not satisfying, the alternatives differing too little from the hypothesis.

For the normal, double exponential and Cauchy distributions, a measure of relative efficiency to the Neyman-Pearson test is given. This shows that test based on $S_{4,\varepsilon}$ and on the Wilcoxon statistic with μ estimated by the median, are serious competitors among scale and translation invariant tests for symmetry. Further, if the underlying distribution does not have too heavy tails, tests based on the sign statistic with the mean as estimate for μ, also seem to be good. More details and comparisons with other tests will be given in a forthcoming paper.

II. Notation, Results

Let $(\Omega, \mathscr{A}, P)$ be a probability space and $X_1, \ldots, X_n$, i.i.d. real random variables with distribution function F and density f. Let $X_{(1)}, \ldots, X_{(n)}$ be the order statistics and $\bar{X}_n$, M_n the mean and the median of $X_1, \ldots, X_n$. Throughout the paper, Ψ is used for F^{-1} and $I(A)$ for the indicator function of a set A.

Tests for symmetry of F about an unknown value μ are derived from the following statistics:

$$S_1 = n^{-1/2} \sum_{1 \leqq i \leqq n} (I(X_i - \bar{X}_n \leqq 0) - 1/2),$$

$$S_2 = n^{-3/2} \sum_{1 \leqq i < j \leqq n} (I(X_i + X_j \leqq 2\bar{X}_n) - 1/2),$$

$$S_3 = n^{-3/2} \sum_{1 \leqq i < j \leqq n} (I(X_i + X_j \leqq 2M_n) - 1/2),$$

$$S_{4,\varepsilon} = n^{-1/2} \sum_{[\varepsilon n]+1 \leqq i \leqq [n/2]} (I(Y_i - Y_{n-i+2} \leqq 0) - 1/2),$$

where $Y_i = X_{(i)} - X_{(i-1)}$, $i = 2, \ldots, n$, and $0 < \varepsilon < 1/2$. $[a]$ means the greatest integer smaller or equal to a. S_i, $i = 1, 2, 3$ are modified sign and Wilcoxon statistics, where μ has been estimated by the mean or the median. $S_{4,\varepsilon}$ is a trimmed sign statistic based on $Y_i - Y_{n-i+2}$, the differences of symmetrically located intervals of order statistics.

II. 1. Asymptotic Distribution under Symmetry

The following theorems give the asymptotic distribution of the four statistics under some regularity conditions. The asymptotic variances under the normal, logistic, double exponential and Cauchy distributions are given in Table 1 at the end of this subsection. For $S_{4,\varepsilon}$, the limit for ε tending to 0 only has been computed.

Theorem 1. *Assume that F has a continuous derivative f at μ. If $f(\mu) \neq 0$ and $\int_{-\infty}^{+\infty} x^2 f(x)\,dx < \infty$, then S_1 is asymptotically normal $\mathcal{N}(0, \sigma_1^2)$, with $\sigma_1^2 = 1/4 + f(\mu)[\sigma^2 f(\mu) - \mu_0]$, where $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx$ and $\mu_0 = 2\int_{\mu}^{\infty} (x-\mu) f(x)\,dx$.*

Theorem 2. *Suppose that f is absolutely continuous and $f' \in L_2(-\infty, +\infty)$.*

If $\int_{-\infty}^{+\infty} x^2 f(x)\,dx < \infty$, then S_2 is asymptotically normal $\mathcal{N}(0, \sigma_2^2)$, with

Table 1

	Normal	Logistic	Double exponential	Cauchy
σ_1^2	0,09085	0,10904	0,25	
σ_2^2	0,00376	0,00805	0,02083	
σ_3^2	0,03155	0,02778	0,02083	0,02083
σ_4^2	0,10938	0,11621	0,12500	0,27744

$$\sigma_2^2 = 1/12 + \sigma^2 \left(\int_{-\infty}^{+\infty} f^2(x)\, dx \right)^2 + 2 \left(\int_{-\infty}^{+\infty} f^2(x)\, dx \right) \left(\int_{-\infty}^{+\infty} x(1/2 - F(x))\, f(x)\, dx \right),$$

where $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\, dx.$

If $f(\mu) \neq 0$, *then* S_3 *is asymptotically normal* $\mathcal{N}(0, \sigma_3^3)$, *where*

$$\sigma_3^2 = 1/48 + 1/4 \left(f(\mu)^{-1} \int_{-\infty}^{+\infty} f^2(x)\, dx - 1/2 \right)^2,$$

Theorem 3. *Assume that the support of F is a (possibly infinite) interval. If f is strictly positive on this interval and twice continuously differentiable, then the statistic* $S_{4,\varepsilon}$ *is asymptotically normal* $\mathcal{N}(0, \sigma_{4,\varepsilon}^2)$, *where*

$$\sigma_{4,\varepsilon}^2 = 1/8 + 1/4 \int_{\varepsilon}^{1/2} (\ln \Psi'(1/2) - \ln \Psi'(t))\, dt + 1/8 \int_{\varepsilon}^{1/2} (\ln \Psi'(1/2) - \ln \Psi'(t))^2\, dt + \varepsilon/16 (\ln \Psi'(1/2) - \ln \Psi'(\varepsilon))^2 - \varepsilon/4.$$

The limit σ_4^2 *of* $\sigma_{4,\varepsilon}^2$ *for* ε *tending to 0, exists and is equal to*

$$\sigma_{4,0}^2 = 1/16 + 1/16 \left(\int_{-\infty}^{+\infty} (1 + \ln f(t)/f(\mu))^2 f(t)\, dt \right),$$

if the integral exists.

II. 2. Asymptotic Distribution under the Alternative

Let F be symmetric about 0. Consider the distribution function $H_n(x) = F(x + n^{-1/2} g(x))$. We assume that $x + n^{-1/2} g(x)$ is, for large n, a monotonous increasing function. Otherwise our definition does not make sense. If the X_i are distributed according to H_n it is shown that the statistics are asymptotically normal with the same standard deviation as under F and a mean different from 0. The quotient of

Table 2

	Normal	Logistic	Double exponential	Cauchy
μ_1	$-0{,}15915$	$-0{,}17328$	$-0{,}25$	
μ_2	$-0{,}03296$	$-0{,}04166$	$-0{,}06250$	
μ_3	0,07957	0,07385	0,06250	0,051
μ_4	$-0{,}125$	$-0{,}125$	$-0{,}125$	$-0{,}125$
$\lvert\mu_1\rvert/\sigma_1$	0,528	0,524	0,5	
$\lvert\mu_2\rvert/\sigma_2$	0,538	0,464	0,433	
$\lvert\mu_3\rvert/\sigma_3$	0,447	0,443	0,433	0,35
$\lvert\mu_4\rvert/\sigma_4$	0,377	0,366	0,353	0,237
$\lvert\mu_5\rvert/\sigma_5$	1	0,845	1	0,5

Here $\mu_4=\mu_{4,0}$ and $\sigma_4^2=\sigma_{4,0}^2$. μ_5 and σ_5 are the asymptotic mean and standard deviation of the Neyman-Pearson statistic, when the X_i are distributed according to H_n.

these values will precisely be our measure of efficiency. The asymptotic means μ_i and the quotients $|\mu_i|/\sigma_i$, under the normal, logistic, double exponential and Cauchy distributions are given in Table 2 at the end of this subsection. The same quantities were also computed for the Neyman-Pearson statistics and are presented in the last line of Table 2.

Theorem 4. *Suppose that*

a) *F is twice continuously differentiable with bounded derivatives;*

b) *f' is monotonous on (a, ∞) for some $a\geqq 0$;*

c) *g admits two bounded continuous derivatives;*

d) $(f')^2\,|x|^3\in L_1(-\infty, +\infty)$, $f^2x^2\in L_1(-\infty, +\infty)$.

Then under the assumption of Section II.1, *the statistics S_i, $i=1,2,3$ are under H_n asymptotically normal $\mathcal{N}(\mu_i, \sigma_i^2)$, $i=1,2,3$, where*

$$\mu_1=f(0)\int_{-\infty}^{\infty}(g(0)-g(x))\,f(x)\,dx$$

$$\mu_2=-\left(\int_{-\infty}^{+\infty}f^2(x)\,dx\right)\left(\int_{-\infty}^{+\infty}g(x)\,f(x)\,dx\right)+\int_{-\infty}^{\infty}f^2(x)\,g(x)\,dx,$$

$$\mu_3=\int_{-\infty}^{\infty}f^2(x)(g(x)-g(0))\,dx.$$

Theorem 5. *Under the assumptions of Theorem* 3, *if g is twice continuously differentiable with bounded derivatives, then $S_{4,\varepsilon}$ is asymptotically normal $\mathcal{N}(\mu_{4,\varepsilon}, \sigma_{4,\varepsilon}^2)$, where*

$$\mu_{4,\varepsilon}=1/4\int_{\varepsilon}^{1/2}(g'(\Psi(t))-g'(-\Psi(t)))\,dt.$$

The limiting value μ_4 for ε tending to zero is given by

$$\mu_{4,0}=1/4\int_{-\infty}^{0}(g'(s)-g'(-s))\,f(s)\,ds.$$

III. Proofs

III.1. Proof of Theorem 1

Although the result has already been proved by Gastwirth [4], we give here a lemma which will be used later for approximating the median of the sample.

Lemma 1. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with a density f. Suppose that f is continuous in a neighbourhood of 0.*

Let

$$Z_n(t) = n^{-1/2} \sum_{1 \leqq i \leqq n} (I(X_i \leqq t\, n^{-1/2}) - I(X_i \leqq 0)) - t\, f(0).$$

Then for any fixed positive number B,

$$\sup\{|Z_n(t)| : t \in [-B, B]\} \xrightarrow{P} 0,$$

as n tends to infinity.

Proof. For simplicity consider t as varying in the interval $[0, 1]$. An easy computation shows that

$$\sup\{|E(Z_n(t))| : t \in [0, 1]\} \to 0,$$

as n tends to infinity. So it is sufficient to prove Lemma 1 for the process $\bar{Z}_n(t)$ which is obtained from $Z_n(t)$ by centering it at expectation. For this let m be a fixed positive integer and $\delta > 0$.

Define $s_k = k/m$, $k = 0, 1, \ldots, m$.

By the triangle inequality:

$$\begin{aligned} &\{\sup\{|\bar{Z}_n(t)| : t \in [0, 1]\} \geqq \delta\} \\ &\quad \subseteq \{\max\{|\bar{Z}_n(s_k)| : k = 0, \ldots, m\} \geqq \delta/2\} \\ &\quad \cup \{\max_{0 \leqq k \leqq m-1} \sup\{|\bar{Z}_n(t) - \bar{Z}_n(s_k)| : t \in [s_k, s_{k+1}]\} \geqq \delta/2\} \\ &\quad = A_1 \cup A_2 \qquad \text{(say).} \end{aligned}$$

By Markov inequality,

$$P\{A_1\} \leqq \sum_{1 \leqq k \leqq m} P\{|\bar{Z}_n(s_k)| \geqq \delta/2\} \leqq (2/\delta)^2 \sum_{1 \leqq k \leqq m} \operatorname{Var}(\bar{Z}_n(s_k)).$$

But $\operatorname{Var}(\bar{Z}_n(s_k)) \leqq (F(s_k/\sqrt{n}) - F(0))$, and by assumption there exists a K such that $\left| F\left(\frac{s_k}{\sqrt{n}}\right) - F(0) \right| \leqq K\, n^{-1/2}$, for n large enough.

Therefore,

$$P\{A_1\} \leqq \left(\frac{2}{\delta}\right)^2 m \cdot K\, n^{-1/2}, \text{ for } n \text{ large enough.}$$

We can now easily find a bound for $P\{A_2\}$ by using the monotony of $I\left(0 \leqq X_i \leqq \frac{t}{\sqrt{n}}\right)$ as a function of t.

Let t be fixed in $[s_k, s_{k+1}]$ and suppose for example that

$$\bar{Z}_n(t) - \bar{Z}_n(s_k) \geqq 0.$$

Then the absolute value of the left-hand side is smaller than

$$|\bar{Z}_n(s_{k+1}) - \bar{Z}_n(s_k)| + n^{-1/2} \sum_{1 \leqq i \leqq n} P\left\{\frac{s_k}{\sqrt{n}} \leqq X_i \leqq \frac{s_{k+1}}{\sqrt{n}}\right\}.$$

A similar argument in the case where $\bar{Z}_n(t) - \bar{Z}_n(s_k) \leqq 0$, leads to

$$\begin{aligned} A_2 \subset & \left\{\max\{|\bar{Z}_n(s_{k+1}) - \bar{Z}_n(s_k)|: k=0, \ldots, m-1\} \geqq \frac{\delta}{8}\right\} \\ & \cup \left\{\max\left\{n^{-1/2} \sum_{1 \leqq i \leqq n} P\left\{\frac{s_k}{\sqrt{n}} \leqq X_i \leqq \frac{s_{k+1}}{\sqrt{n}}\right\}: k=0, \ldots, m-1\right\} \geqq \frac{\delta}{8}\right\} \\ & = B_1 \cup B_2 \qquad \text{(say).} \end{aligned}$$

By assumption there exists a $K > 0$, such that

$$\operatorname{Var}(\bar{Z}_n(s_{k+1}) - \bar{Z}_n(s_k)) \leqq \frac{K}{\sqrt{n}} (s_{k+1} - s_k) = \frac{K}{\sqrt{n}} \cdot m^{-1},$$

and

$$P\left\{\frac{s_k}{\sqrt{n}} \leqq X_i \leqq \frac{s_{k+1}}{\sqrt{n}}\right\} \leqq \frac{K}{\sqrt{n}} m^{-1}, \qquad \text{for } n \text{ large enough.}$$

Hence

$$P\{B_1\} \leqq \left(\frac{8}{d}\right)^2 \frac{K}{\sqrt{n}},$$

and

$$P\{B_2\} = 0, \qquad \text{for } m, n \text{ large enough.}$$

Lemma 1 follows by letting, for a fixed δ, n tend to infinity and then m.

Theorem 1 follows directly from the lemma.

Proof. Suppose without loss of generality that $\mu = 0$. Then since

$$\int_{-\infty}^{+\infty} x^2 f(x)\, dx < \infty, \qquad \sqrt{n}\, \bar{X}_n = o_p(1),$$

S_1 can be written as

$$Z_n(n^{1/2} \bar{X}_n) + n^{1/2} \bar{X}_n f(0) + n^{-1/2} \sum_{1 \leqq i \leqq n} \{I(X_i \leqq 0) - 1/2\}.$$

By Lemma 1, the first term tends to 0 in probability and Theorem 1 follows.

For later purposes we now use Lemma 1 to approximate the median M_n by a statistic which is easier to handle.

Let $X_1, \ldots, X_n$ be i.i.d. with density f, symmetric around 0. Suppose that $f(0) \neq 0$. Then under the assumptions of Lemma 1,

$$n^{1/2} M_n = (f(0))^{-1} n^{-1/2} \sum_{1 \leqq i \leqq n} \{1/2 - I(X_i \leqq 0)\} + o_P(1). \qquad (*)$$

Proof. It is known that $n^{1/2} M_n = O_p(1)$. By definition of M_n,

$$n^{-1/2} \sum_{1 \leqq i \leqq n} \{I(X_i \leqq M_n) - 1/2\} = 0,$$

which is equivalent to

$$Z_n(n^{1/2} M_n) + n^{1/2} M_n f(0) + n^{-1/2} \sum_{1 \leqq i \leqq n} \{I(X_i \leqq 0) - 1/2\} = 0.$$

Since by Lemma 1, $Z_n(n^{1/2} M_n) \xrightarrow{P} 0$, as n tends to infinity, relation $(*)$ is proved.

III.2. Proof of Theorem 2

Assume without loss of generality that $\mu = 0$. Under the assumptions of Theorem 2, the mean and the median are, whenever used, $O_P(n^{-1/2})$. It then follows from Antille [1], Theorem II.2 and Corollary, that

$$S_2 = \sqrt{n}\, \bar{X}_n \int_{-\infty}^{+\infty} f^2(x)\, dx + n^{-3/2} \sum_{1 \leqq i < j \leqq n} \{I(X_i + X_j \leqq 0) - \tfrac{1}{2}\} + o_P(1),$$

and

$$S_3 = \sqrt{n}\, M_n \int_{-\infty}^{+\infty} f^2(x)\, dx + n^{-3/2} \sum_{1 \leqq i < j \leqq n} \{I(X_i + X_j \leqq 0) - \tfrac{1}{2}\} + o_P(1).$$

The second term of the right-hand side can be approximated as follows:

Lemma 2. *Assume that f is symmetric about* 0. *Let*

$$T = n^{-3/2} \sum_{1 \leqq i < j \leqq n} \{I(X_i + X_j \leqq 0) - \tfrac{1}{2}\}.$$

Then,

$$T = n^{-1/2} \sum_{1 \leqq i \leqq n} \{\tfrac{1}{2} - F(X_i)\} + o_P(1).$$

Proof. We use the projection method of Hájek [5]: T is approximated by its projection $\bar{T} = \sum_{1 \leqq i \leqq n} E(T \mid X_i)$. $E(T \mid X_i)$ means here the conditional expectation of T given X_i. An easy computation shows that

$$\bar{T} = (n-1)\, n^{-3/2} \sum_{1 \leqq i \leqq n} \{\tfrac{1}{2} - F(X_i)\},$$

$\operatorname{Var}(\bar{T}) = (n-1)^2 n^{-2} (12)^{-1}$, and

$$\begin{aligned}\operatorname{Var}(T) &= (n-1)\,n^{-2}(2)^{-1}\operatorname{Var}(I(X_1+X_2\leqq 0))\\ &\quad +(n-1)(n-2)\,n^{-2}\operatorname{Cov}(I(X_1+X_2\leqq 0), I(X_1+X_3\leqq 0))\\ &= (n-1)\,n^{-2}(8)^{-1}+n^{-2}(n-2)(n-1)(12)^{-1}.\end{aligned}$$

Since $E(T-\bar{T})^2=\operatorname{Var}(T-\bar{T})=\operatorname{Var}(T)-\operatorname{Var}(\bar{T})$, Lemma 2 follows.

The statistic S_2 is then asymptotically equivalent to

$$n^{-1/2}\sum_{1\leqq i\leqq n}\left(X_i\int_{-\infty}^{+\infty} f^2(x)\,dy+\tfrac{1}{2}-F(X_i)\right).$$

By using the approximation we gave before for the median M_n, S_3 is easily shown to be asymptotically equivalent to

$$\left(\int_{-\infty}^{+\infty} f^2(x)\,dx\right)(f(0))^{-1}\,n^{-1/2}\sum_{1\leqq i\leqq n}\{\tfrac{1}{2}-I(X_i\leqq 0)\}$$
$$+n^{-1/2}\sum_{1\leqq i\leqq n}\{\tfrac{1}{2}-F(X_i)\},$$

and Theorem 2 is proved.

III.3. Proof of Theorem 4

Let $H_n=F\left(x+\frac{g(x)}{\sqrt{n}}\right)$ and h_n its derivative. Using Taylor expansion,

$$H_n(x)=F(x)+f(x)\,\frac{g(x)}{\sqrt{n}}+f'(x+\eta_1(x))\,\frac{g^2(x)}{2n},$$

$$h_n(x)=f(x)+f(x)\,\frac{g'(x)}{\sqrt{n}}+f'(x+\eta_2(x))\left(1+\frac{g'(x)}{\sqrt{n}}\right)\frac{g(x)}{\sqrt{n}}.$$

Consider first S_1.

By the same way as before, one can show that under H_n,

$$S_1=\sqrt{n}\,\bar{X}_n h_n(0)+n^{-1/2}\sum_{1\leqq i\leqq n}\{X_i\leqq 0)-\tfrac{1}{2}\}+o_P(1).$$

The asymptotic mean μ_1 is then seen to be

$$f(0)\int_{-\infty}^{+\infty} x(f(x)\,g'(x)+f'(x)\,g(x))\,dx+\int_{-\infty}^{0}(f(x)\,g'(x)+f'(x)\,g(x))\,dx,$$

while the variance remains the same. Here the monotony of f' and the fact that $\int_{-\infty}^{+\infty} x^2 f(x)\,dx<\infty$, are used to show that in asymptotic considerations, H_n and h_n can be replaced by

$$F(x)+f(x)\frac{g(x)}{\sqrt{n}} \quad \text{and} \quad f(x)+f(x)\frac{g'(x)}{\sqrt{n}}+f'(x)\frac{g(x)}{\sqrt{n}}.$$

By partial integration, μ_1 simplifies to

$$-f(0)\int_{-\infty}^{+\infty} f(x)\,g(x)+f(0)\,g(0).$$

Consider now S_2 and S_3.

Looking at the proof of Theorem II.2 in Antille [1], one sees that, under H_n,

$$\sup\left\{\left|n^{-3/2}\sum_{1\leq i<j\leq n}\left\{I\left(X_i+X_j\leqq\frac{t}{\sqrt{n}}\right)-I(X_i+X_j\leqq 0)\right\}\right.\right.$$
$$\left.\left.-t\int_{-\infty}^{+\infty} f^2(x)\,dx\right| : t\in[-M,+M]\right\}$$

tends to 0 in probability as n goes to infinity, for every fixed number M.

Further it still holds that

$$\sqrt{n}\,M_n=(f(0))^{-1}\,n^{-1/2}\sum_{1\leqq i\leqq n}\{\tfrac{1}{2}-I(X_i\leqq 0)\}+o_P(1).$$

Therefore, by the same argument as before,

$$S_2=\sqrt{n}\,\bar{X}_n\left(\int_{-\infty}^{+\infty} f^2(x)\,dx\right)+n^{-3/2}$$
$$\cdot\sum_{1\leqq i<j\leqq n}\{I(X_i+X_j\leqq 0)-E(I(X_i+X_j\leqq 0))\}$$
$$+n^{-3/2}\,n(n-1)\,2^{-1}\,E(I(X_1+X_2\leqq 0)-\tfrac{1}{2})+o_P(1),$$

and

$$S_3=(f(0))^{-1}\left(\int_{-\infty}^{+\infty} f^2(x)\,dx\right)n^{-1/2}\sum_{1\leqq i\leqq n}\{\tfrac{1}{2}-I(X_i\leqq 0)\}$$
$$+n^{-3/2}\sum_{1\leqq i<j\leqq n}\{I(X_i+X_j\leqq 0)-E(I(X_i+X_j\leqq 0))\}$$
$$+n^{-3/2}\,n(n-1)\,2^{-1}\,E(I(X_1+X_2\leqq 0)-\tfrac{1}{2})+o_P(1).$$

Using Taylor expansions for H_n and h_n and assumptions c) and d) one gets,

$$\mu_2=-\left(\int_{-\infty}^{+\infty} f^2(x)\,dx\right)\left(\int_{-\infty}^{+\infty} f(x)\,g(x)\,dx\right)$$
$$+\tfrac{1}{2}\int_{-\infty}^{+\infty}(F(-x)\,f(x)\,g'(x)+F(-x)\,f'(x)\,g(x)+f^2(x)\,g(-x))\,dx,$$

$$\mu_3=-g(0)\left(\int_{-\infty}^{+\infty} f^2(x)\,dx\right)$$
$$+\tfrac{1}{2}\int_{-\infty}^{+\infty}(F(-x)\,f(x)\,g'(x)+F(-x)\,f'(x)\,g(x)+f^2(x)\,g(-x))\,dx.$$

By partial integration the second term of the right hand sides simplifies to

$$\tfrac{1}{2}\int_{-\infty}^{+\infty} f^2(x)(g(x)+g(-x))\,dx.$$

The asymptotic variance is unchanged and Theorem 4 follows.

III.4. Proof of Theorems 3 and 5

Let $Y_i = X_{(i)} - X_{(i-1)}$, $i=2,\dots,n$, and choose $0<\varepsilon<\frac{1}{2}$ as in Theorem 3. Define $l=[\varepsilon n]+1$, $m=[n/2]$. Then

$$S_{4,\varepsilon}=n^{-1/2}\sum_{l\leqq i\leqq m}\{I(Y_i-Y_{n-i+2}\leqq 0)-\tfrac{1}{2}\}.$$

Step 1. We give first a representation of Y_i through exponentially distributed random variables.

Let $W_1, W_2, \dots$ be a sequence of i.i.d. random variables with exponential distribution and mean 1.

Put

$$U_i=\sum_{1\leqq k\leqq i} W_k \Big/ \sum_{1\leqq k\leqq n+1} W_k.$$

It is well-known (see Breiman [2], p. 285) that the vector $(U_1, \dots, U_n)$ has the same distribution as the vector of order statistics of n i.i.d. random variables with uniform distribution on $[0, 1]$.

Thus the vectors $(X_{(1)}, \dots, X_{(n)})$ and $(\Psi_n(U_1), \dots, \Psi_n(U_n))$ are identically distributed, Ψ_n being the inverse of the distribution function H_n. As in Theorem 5, H_n is given by $F\left(x+\frac{g(x)}{\sqrt{n}}\right)$. We now compute Ψ_n.

Define $y=y(x)$ (depending on n) by

$$y=x+n^{-1/2}g(x).$$

y is strictly increasing in x and varies from $-\infty$ to $+\infty$ for n large enough, since g has a bounded derivative. y can be written as

$$y=x+n^{-1/2}g(y)+n^{-1}K_n(y),$$

with

$$K_n(y)=n^{1/2}(g(x)-g(y)).$$

Differentiating both sides of the last expression with respect to x and applying the mean value theorem we get, with $0<\delta<1$,

$$\begin{aligned} K_n'(y)(1+n^{-1/2}g'(x)) &= n^{1/2}(g'(x)-g'(y)(1+n^{-1/2}g'(x))) \\ &= n^{1/2}g''(x+\delta(y-x))(x-y)-g'(x)\,g'(y) \\ &= g''(x+\delta(y-x))\,g(x)-g'(x)\,g'(y). \end{aligned}$$

Since g' and g'' are bounded, $K_n'(y)$ is, for n large enough, uniformly bounded on compact subsets of $\mathbb{R}$. By definition,

$$H_n(y-n^{-1/2}g(y)-n^{-1}K_n(y))=F(y), \qquad \forall\, y\in\mathbb{R}.$$

Hence $y-n^{-1/2}g(y)-n^{-1}K_n(y)=H_n^{-1}(F(y))$.

Put $y=F^{-1}(t)$, $t\in(0,1)$, to get

$$H_n^{-1}(t)=\Psi(t)-n^{-1/2}g(\Psi(t))-n^{-1}K_n(\Psi(t)),$$

where

$$\Psi(t)=F^{-1}(t).$$

By assumption, Ψ is three times differentiable on (0, 1). Using Taylor expansions,

$$\begin{aligned}X_{(i)}-X_{(i-1)}&=H_n^{-1}(U_i)-H_n^{-1}(U_{i-1})\\&=\Psi'(U_i)(U_i-U_{i-1})-n^{-1/2}g'(\Psi(U_i))\,\Psi'(U_i)(U_i-U_{i-1})+\mathscr{R}_{i,1},\end{aligned}$$

with

$$\begin{aligned}\mathscr{R}_{i,1}&=\tfrac{1}{2}\Psi''(U_i+\delta_1(U_i-U_{i-1}))(U_i-U_{i-1})^2\\&\quad-\tfrac{1}{2}(g\circ\Psi)''(U_i+\delta_2(U_i-U_{i-1}))(U_i-U_{i-1})^2n^{-1/2}\\&\quad-n^{-1}(K_n\circ\Psi)''(U_i+\delta_3(U_i-U_{i-1}))(U_i-U_{i-1}),\end{aligned}$$

where $0<\delta_i<1$, for $i=1,2,3$.

It is well-known that,

$$\max\left\{\left|U_i-\frac{i}{n+1}\right|: i=1,\dots,n\right\}=O_P(n^{-1/2}).$$

Then, since $(K_n\circ\Psi)'$ is uniformly bounded on compact subsets of (0, 1),

$$\begin{aligned}&\max\{|\Psi''(U_i+\delta_1(U_i-U_{i-1}))|: i=l,\dots,m\}=O_P(1),\\&\max\{|(g\circ\Psi)''(U_i+\delta_2(U_i-U_{i-1}))|: i=l,\dots,m\}=O_P(1),\\&\max\{|(K_n\circ\Psi)'(U_i+\delta_3(U_i-U_{i-1}))|: i=l,\dots,m\}=O_P(1).\end{aligned}$$

By definition of U_i,

$$U_i-U_{i-1}=W_i\Big/\sum_{1\leqq k\leqq n+1}W_k.$$

W_i being exponential,

$$\max\{W_i: i=1,\dots,n\}=O_P(\log n).$$

Therefore,

$$\max\{|\mathscr{R}_{i,1}|: i=l,\dots,m\}=O_P((\log n)^2n^{-2}).$$

Further,

$$\Psi'(U_i)=\Psi'\left(\frac{i}{n+1}\right)+\Psi''\left(\frac{i}{n+1}\right)\left(U_i-\frac{i}{n+1}\right)$$
$$+\tfrac{1}{2}\Psi'''\left(\frac{i}{n+1}+\delta_4\left(U_i-\frac{i}{n+1}\right)\right)\left(U_i-\frac{i}{n+1}\right)^2,$$

and

$$n^{-1/2}g'(\Psi(U_i))\,\Psi'(U_i)=n^{-1/2}g'\left(\Psi\left(\frac{i}{n+1}\right)\right)\Psi'\left(\frac{i}{n+1}\right)$$
$$+n^{-1/2}(g\circ\Psi)''\left(\frac{i}{n+1}+\delta_5\left(U_i-\frac{i}{n+1}\right)\right)\left(U_i-\frac{i}{n+1}\right),$$

where $0<\delta_4,\ \delta_5<1$.

Using again the boundedness of the functions involved,

$$\max\left\{\left|\Psi'''\left(\frac{i}{n+1}+\delta_4\left(U_i-\frac{i}{n+1}\right)\right)\left(U_i-\frac{i}{n+1}\right)^2\right|: i=l,\dots,m\right\}=O_P(n^{-1}),$$
$$\max\left\{\left|n^{-1/2}(g\circ\Psi)''\left(\frac{i}{n+1}+\delta_5\left(U_i-\frac{i}{n+1}\right)\right)\left(U_i-\frac{i}{n+1}\right)\right|: i=l,\dots,m\right\}$$
$$=O_P(n^{-1}).$$

Thus, putting everything together,

$$Y_i=X_{(i)}-X_{(i-1)}$$
$$=\Psi'\left(\frac{i}{n+1}\right)(U_i-U_{i-1})+\Psi''\left(\frac{i}{n+1}\right)\left(U_i-\frac{i}{n+1}\right)(U_i-U_{i-1})$$
$$-n^{-1/2}g'\left(\Psi\left(\frac{i}{n+1}\right)\right)\Psi'\left(\frac{i}{n+1}\right)(U_i-U_{i-1})+\mathscr{R}_{i,2},$$

where

$$\max\{|\mathscr{R}_{i,2}|: i=l,\dots,m\}=O_P((\log n)^2 n^{-2}).$$

We now introduce some notation:

Define,

$$\overline{W}_k=W_{n-k+2},\qquad k=1,\dots,n+1,$$
$$\bar{\bar{U}}_i=\sum_{1\leq k\leq i}\overline{W}_k\Big/\sum_{1\leq k\leq n+1}\overline{W}_k=\sum_{1\leq k\leq i}\overline{W}_k\Big/\sum_{1\leq k\leq n+1}W_k.$$

Then $U_{n-i+2}=1-\bar{\bar{U}}_{i-1}$ and $U_{n-i+1}=1-\bar{\bar{U}}_i$.

Therefore,

$$Y_{n-i+2}=\Psi_n(1-\bar{\bar{U}}_{i-1})-\Psi_n(1-\bar{\bar{U}}_i).$$

Since $\Psi(1-t)=-\Psi(t)$,

$$\Psi_n(1-t)=\Psi(1-t)-n^{-1/2}g(\Psi(1-t))-n^{-1}K_n(\Psi(1-t))$$
$$=-\Psi(t)+n^{-1/2}\bar{g}(\Psi(t))-n^{-1}K_n(\Psi(1-t)),$$

with $\bar{g}(x)=-g(-x)$.

Thus,

$$\begin{aligned}Y_{n-i+2}=&\Psi(\bar{U}_i)-\Psi(\bar{U}_{i-1})-n^{-1/2}\bar{g}(\Psi(\bar{U}_i))\\&+n^{-1/2}\bar{g}(\Psi(\bar{U}_{i-1}))+n^{-1}K_n(\Psi(1-\bar{U}_i))\\&-n^{-1}K_n(\Psi(1-\bar{U}_{i-1})).\end{aligned}$$

Then interchanging U_i and $\bar{U}_i$, g and $\bar{g}$, in the expansion for Y_i, we obtain a similar one for Y_{n-i+2}.

Now

$$I(Y_i-Y_{n-i+2}\leqq 0)=I\left((Y_i-Y_{n-i+2})\left(\sum_{1\leqq k\leqq n+1}W_k/\Psi'\left(\frac{i}{n+1}\right)\right)\leqq 0\right),$$

since $\Psi'\left(\frac{i}{n+1}\right)>0$, by assumption. The asymptotic distribution of $S_{4,\varepsilon}$ is thus the same as the asymptotic distribution of $n^{-1/2}\sum_{l\leqq i\leqq m}\{I(Z_i\leqq 0)-\frac{1}{2}\}$, where, by our representation,

$$\begin{aligned}Z_i&=(Y_i-Y_{n-i+2})\left(\sum_{1\leqq k\leqq n+1}W_k/\Psi'\left(\frac{i}{n+1}\right)\right)\\&=W_i+\frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)}W_i\left(U_i-\frac{i}{n+1}\right)-n^{-1/2}g'\left(\Psi\left(\frac{i}{n+1}\right)\right)W_i\\&\quad-\bar{W}_i-\frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)}\bar{W}_i\left(\bar{U}_i-\frac{i}{n+1}\right)+n^{-1/2}\bar{g}'\left(\Psi\left(\frac{i}{n+1}\right)\right)\bar{W}_i+\mathscr{R}_{i,3},\end{aligned}$$

with

$$\max\{|\mathscr{R}_{i,3}|: i=l,\ldots,m\}=O_P(n^{-1}(\log n)^2).$$

This is the desired representation.

Step 2. We now approximate the statistic $n^{-1/2}\sum_{l\leqq i\leqq m}\{I(Z_i\leqq 0)-\frac{1}{2}\}$, by sums of independent random variables. To do this we use the fact that W_i and U_i are becoming more and more independent with i and n increasing. The idea is to replace U_i by a different random variable and thus to enforce independence.

For this, choose $\frac{1}{2}<\gamma<1$ and define

$$\alpha_n=[n^\gamma].$$

Then define integers $N, \beta_0, \ldots, \beta_N$ as follows:

Let $\beta_0=l-1$, and for $i=1, 2, \ldots, N-1$, put $\beta_{i+1}=\beta_i+\alpha_n$, where N is such that $\beta_{N-1}<m$ and $m-\beta_{N-1}<\alpha_n$. Let $\beta_N=m$.

We now introduce some new variables. For $j=1,\ldots,N$, define

$$V_j=\sum_{1\leqq k\leqq\beta_{j-1}}W_k/\Big(\sum_{1\leqq k\leqq n+1}W_k-\sum_{\beta_{j-1}+1\leqq k\leqq\beta_j}(W_k+\bar{W}_k)+2(\beta_j-\beta_{j-1})\Big)-\frac{\beta_{j-1}}{n+1}.$$

$\overline{V}_j$ is defined similarly by replacing W_k by $\overline{W}_k$. Finally, define, for any $l \leqq i \leqq m$, $i \in \mathbb{N}$, the integer $j(i)$ by

$$j(i)=j \quad \text{if } \beta_{j-1}+1 \leqq i \leqq \beta_j, \quad j=1, \ldots, N.$$

We now want to estimate the magnitude of the error for our statistic if in the definition of Z_i, $U_i-\frac{i}{n+1}$ and $\overline{U}_i-\frac{i}{n+1}$ are replaced by $V_{j(i)}$ and $\overline{V}_{j(i)}$.

We have:

$$\begin{aligned}
&\left|U_i-\frac{i}{n+1}-V_{j(i)}\right| \\
&\leqq\left|\frac{\sum\limits_{1 \leqq k \leqq i} W_k}{\sum\limits_{1 \leqq k \leqq n+1} W_k}-\frac{\sum\limits_{1 \leqq k \leqq \beta_{j(i)-1}} W_k}{\sum\limits_{1 \leqq k \leqq n+1} W_k}-\frac{i-\beta_{j(i)-1}}{n+1}\right| \\
&\quad+\frac{\sum\limits_{1 \leqq k \leqq \beta_{j(i)-1}} W_k \cdot\left|\sum\limits_{\beta_{j(i)-1}+1 \leqq k \leqq \beta_{j(i)}}(W_k+\overline{W}_k)-2(\beta_{j(i)}-\beta_{j(i)-1})\right|}{\left(\sum\limits_{1 \leqq k \leqq n+1} W_k\right)\left(\sum\limits_{1 \leqq k \leqq n+1} W_k-\sum\limits_{\beta_{j(i)-1}+1 \leqq k \leqq \beta_{j(i)}}(W_k+\overline{W}_k)+2(\beta_{j(i)}-\beta_{j(i)-1})\right)} \\
&\leqq \frac{\left|\sum\limits_{\beta_{j(i)-1}+1 \leqq k \leqq i} W_k-(i-\beta_{j(i)-1})\right|}{\sum\limits_{1 \leqq k \leqq n+1} W_k}+\frac{\alpha_n\left|\sum\limits_{1 \leqq k \leqq n+1} W_k-(n+1)\right|}{(n+1) \sum\limits_{1 \leqq k \leqq n+1} W_k} \\
&\quad+\frac{\left|\sum\limits_{\beta_{j(i)-1}+1 \leqq k \leqq \beta_{j(i)}} W_k-(\beta_{j(i)}-\beta_{j(i)-1})\right|}{\sum\limits_{1 \leqq k \leqq n+1} W_k} \\
&\quad+\frac{\left|\sum\limits_{\beta_{j(i)-1}+1 \leqq k \leqq \beta_{j(i)}} \overline{W}_k-(\beta_{j(i)}-\beta_{j(i)-1})\right|}{\sum\limits_{1 \leqq k \leqq n+1} W_k}
\end{aligned}$$

The second term is of order $n^{\gamma-3/2}$ by the central limit theorem. The three other terms can be handled similarly. So consider the first expression (say).

By independence, for $r>0$,

$$\begin{aligned}
&P\left\{\max\left\{\left|\sum_{\beta_{j(i)-1}+1 \leqq k \leqq i} W_k-(i-\beta_{j(i)-1})\right|: i=l, \ldots, m\right\} \leqq n^r\right\} \\
&=\prod_{1 \leqq j \leqq N} P\left\{\max\left\{\left|\sum_{\beta_{j(i)-1}+1 \leqq k \leqq i} W_k\right.\right.\right. \\
&\qquad \left.\left.\left.-(i-\beta_{j(i)-1})\right|: \beta_{j(i)-1}<i \leqq \beta_{j(i)}\right\} \leqq n^r\right\} \\
&\geqq P\left\{\max\left\{\left|\sum_{1 \leqq k \leqq i} W_k-i\right|: i=1, \ldots, \alpha_n\right\} \leqq n^r\right\}^N .
\end{aligned}$$

By a lemma due to Skorokhod (see Breiman [2], p. 45),

$$\begin{aligned}
&P\left\{\max\left\{\left|\sum_{1 \leqq k \leqq i} W_k-i\right|: i=1, \ldots, \alpha_n\right\}>n^r\right\} \\
&\leqq \frac{1}{1-c} P\left\{\left|\sum_{1 \leqq k \leqq \alpha_n} W_k-\alpha_n\right|>n^r / 2\right\},
\end{aligned}$$

with

$$c=\sup\{P\{|\sum_{i\leqq k\leqq\alpha_n} W_k-(\alpha_n-i)|>n^r\}: i=1,\ldots,\alpha_n\}.$$

By Tschebyscheff inequality,

$$c\leqq n^{-2r}\sup\{\mathrm{Var}(\sum_{i\leqq k\leqq\alpha_n} W_k-(\alpha_n-i)): i=1,\ldots,\alpha_n\}=n^{-2r}\alpha_n\sim n^{\gamma-2r}.$$

Now choose $r>\frac{\gamma}{2}$. Then $c\leqq\frac{1}{2}$, for n large enough, and

$$\begin{aligned}P\{\max\{|\sum_{1\leqq k\leqq i} W_k-i|: i=1,\ldots,\alpha_n\}>n^r\}\\ \leqq 2P\{n^{-\gamma/2}|\sum_{1\leqq k\leqq\alpha_n} W_k-\alpha_n|>n^{r-\gamma/2}/2\}\\ \leqq 2P\{n^{-\gamma/2}|\sum_{1\leqq k\leqq\alpha_n} W_k-\alpha_n|>\alpha_n^{(r-\frac{\gamma}{2})/\gamma}/2\}.\end{aligned}$$

Now use the large deviation theorem (see Feller [3], p. 549) for exponentially distributed variables. For this, let $\gamma'=\min\left\{1/7,\left(r-\frac{\gamma}{2}\right)/\gamma\right\}$. Then, for large n, if $r>\gamma/2$,

$$P\{n^{-\gamma/2}|\sum_{1\leqq k\leqq\alpha_n} W_k-\alpha_n|>\alpha_n^{(r-\frac{\gamma}{2})/\gamma}/2\}\leqq\exp(-\alpha_n^{2\gamma'})\sim\exp(-n^{2\gamma\gamma'}).$$

Thus,

$$\begin{aligned}P\{\max\{|\sum_{\beta_{j(i)-1}+1\leqq k\leqq i} W_k-(i-\beta_{j(i)-1})|: i=l,\ldots,m\}\leqq n^r\}\\ \geqq(1-2\exp(-n^{2\gamma\gamma'}))^N\xrightarrow[n\to\infty]{}1,\qquad\text{since } N\leqq n.\end{aligned}$$

Putting everything together we get:

$$\begin{aligned}\max\left\{\left|U_i-\frac{i}{n+1}-V_{j(i)}\right|: i=l,\ldots,m\right\}\\ =O_P(n^{r-1})+O_P(n^{\gamma-3/2}),\qquad\text{for all } r>\gamma/2.\end{aligned}$$

Obviously, the same is true, if we replace U_i and $V_{j(i)}$ by $\overline{U}_i$ and $\overline{V}_{j(i)}$.

Now define:

$$\begin{aligned}\tilde{Z}_i=W_i+\frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)}W_iV_{j(i)}-n^{-1/2}g'\left(\Psi\left(\frac{i}{n+1}\right)\right)W_i\\ -\overline{W}_i-\frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)}\overline{W}_i\overline{V}_{j(i)}+n^{-1/2}g'\left(\Psi\left(\frac{i}{n+1}\right)\right)\overline{W}_i.\end{aligned}$$

Then $Z_i = \tilde{Z}_i + \mathscr{R}_{i,4}$, and since $\gamma < 1$,

$$\max\{|\mathscr{R}_{i,4}|: i=l, \ldots, m\} = O_P(n^{-s}),$$

with $s > 1/2$ (and s of course depending on γ). We mention here for later use that our results also show, that

$$\max\{|V_{j(i)}|: i=l, \ldots, m\} = O_P(n^{-1/2}).$$

Step 3. We prove here that the statistic $n^{-1/2} \sum_{l \leqq i \leqq m} \{I(Z_i \leqq 0) - \frac{1}{2}\}$ is asymptotically equivalent to

$$n^{-1/2} \sum_{l \leqq i \leqq m} \{I(\tilde{Z}_i \leqq 0) - \tfrac{1}{2}\}.$$

Let $1/2 < s' < s$ and define,

$$A_n = \{\max\{|\mathscr{R}_{i,4}|: i=l, \ldots, m\} \leqq n^{-s'}\},$$

$$B_i = \left\{ \left| \frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} \right| (|V_{j(i)}| + |\overline{V}_{j(i)}|) \leqq \tfrac{1}{2} \right\},$$

$$C_n = \bigcap_{l \leqq i \leqq m} B_i \cap A_n.$$

Then $P\{C_n\} \to 1$, for $n \to \infty$.

But

$$P\{\tilde{Z}_i \leqq 0, Z_i > 0, C_n\} \leqq P\{-n^{-s'} \leqq \tilde{Z}_i \leqq 0, B_i\}$$
$$= \int_{B_i} P\{-n^{-s'} \leqq \tilde{Z}_i \leqq 0 \mid V_{j(i)}, \overline{V}_{j(i)}\}\, dP.$$

Since W_i, $\overline{W}_i$ are independent of $(V_{j(i)}, \overline{V}_{j(i)})$ by construction, one can easily compute the conditional distribution of $\tilde{Z}_i$ given $V_{j(i)}$ and $\overline{V}_{j(i)}$. The density is given by

$$\frac{1}{\lambda_1 + \lambda_2} e^{-\lambda_1^{-1} x}, \quad x \geqq 0,$$

$$\frac{1}{\lambda_1 + \lambda_2} e^{-\lambda_2^{-1} x}, \quad x \leqq 0,$$

where

$$\lambda_1 = 1 + \frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} V_{j(i)} - n^{-1/2} g'\left(\Psi\left(\frac{i}{n+1}\right)\right),$$

$$\lambda_2 = 1 + \frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} \overline{V}_{j(i)} - n^{-1/2} \bar{g}'\left(\Psi\left(\frac{i}{n+1}\right)\right).$$

On B_i, $\lambda_i \geqq 1/4$, for $i=1,2$ and n large enough. Therefore, on B_i,

$$P\{-n^{-s'} \leqq \tilde{Z}_i \leqq 0 \mid V_{j(i)}, \bar{V}_{j(i)}\} \leqq 2n^{-s'}.$$

Hence

$$P\{\tilde{Z}_i \leqq 0, Z_i > 0, C_n\} \leqq 2n^{-s'}.$$

Now

$$\begin{aligned} n^{-1/2} \sum_{l \leqq i \leqq m} \{I(Z_i \leqq 0) - \tfrac{1}{2}\} &\geqq n^{-1/2} \sum_{l \leqq i \leqq m} \{I(\tilde{Z}_i \leqq 0) - \tfrac{1}{2}\} - n^{-1/2} \sum_{l \leqq i \leqq m} I(\tilde{Z}_i \leqq 0, Z_i > 0, C_n) \\ &\quad - n^{+1/2} I(C_n^c). \end{aligned}$$

(A^c means the complement of the set A.) The last two random variables converge stochastically to zero. Similarly one can show that

$$n^{-1/2} \sum_{l \leqq i \leqq m} \{I(Z_i \leqq 0) - \tfrac{1}{2}\} \leqq n^{-1/2} \sum_{l \leqq i \leqq m} \{I(\tilde{Z}_i \leqq 0) - \tfrac{1}{2}\} + o_P(1).$$

Step 4. We first introduce new random variables. For $i=1, \ldots, m$, let

$$L_i = I(\tilde{Z}_i \leqq 0, W_i - \bar{W}_i > 0) - I(\tilde{Z}_i > 0, W_i - \bar{W}_i \leqq 0).$$

Then,

$$\sum_{l \leqq i \leqq m} I(\tilde{Z}_i \leqq 0) = \sum_{l \leqq i \leqq m} \{I(W_i - \bar{W}_i \leqq 0) + L_i\}.$$

Further define,

$$M_i = E(L_i \mid V_{j(i)}, \bar{V}_{j(i)}).$$

($E(X \mid Y)$ means the conditional expectation of X given Y.)

We now show that our statistic is asymptotically equivalent to

$$n^{-1/2} \sum_{l \leqq i \leqq m} \{I(W_i - \bar{W}_i \leqq 0) - 1/2 + M_i\}.$$

Let $d > 0$ and define

$$D_i = \left\{ \left| \frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} \right| (|V_{j(i)}| + |\bar{V}_{j(i)}|) \leqq d\, n^{-1/2} \right\},$$

and

$$\bar{D}_n = \bigcap_{l \leqq i \leqq m} D_i.$$

For any $\eta>0$ we can choose d so large, that

$$\liminf P\{\tilde{D}_n\} \geqq 1-\eta.$$

Thus it is sufficient to chow that

$$\frac{1}{n}\int_{\tilde{D}_n}\Big(\sum_{l\leqq i\leqq m} L_i-M_i\Big)^2\,dP\to 0, \qquad \text{as } n \text{ tends to } \infty.$$

By Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{1}{n}\int_{\tilde{D}_n}\Big(\sum_{l\leqq i\leqq m} L_i-M_i\Big)^2\,dP &= \frac{1}{n}\int_{\tilde{D}_n}\Big(\sum_{1\leqq j\leqq N}\ \sum_{\beta_{j-1}+1\leqq i\leqq \beta_j}(L_i-M_i)\Big)^2\,dP\\ &\leqq \frac{1}{n}N\sum_{1\leqq j\leqq N}\int_{D_i}\Big(\sum_{\beta_{j-1}+1\leqq i\leqq\beta_j} L_i-M_i\Big)^2\,dP.\end{aligned}$$

By construction W_i, $\overline{W}_i$, $i=\beta_{j-1}+1,\ldots,\beta_j$, are independent of $V_{j(i)}$, $\overline{V}_{j(i)}$. Thus

$$\begin{aligned}&\int_{D_i}\Big(\sum_{\beta_{j-1}+1\leqq i\leqq\beta_j} L_i-M_i\Big)^2\,dP\\ &\quad=\int_{D_i} E\Big(\Big(\sum_{\beta_{j-1}+1\leqq i\leqq\beta_j} L_i-M_i\Big)^2\,\Big|\,V_{j(i)},\overline{V}_{j(i)}\Big)\,dP\\ &\quad=\int_{D_i}\sum_{\beta_{j-1}+1\leqq i\leqq\beta_j} E((L_i-M_i)^2\mid V_{j(i)},\overline{V}_{j(i)})\,dP\\ &\quad\leqq\int_{D_i}\sum_{\beta_{j-1}+1\leqq i\leqq\beta_j} E(L_i^2\mid V_{j(i)},\overline{V}_{j(i)})\,dP.\end{aligned}$$

But

$$\begin{aligned}&E(L_i^2\mid V_{j(i)},\overline{V}_{j(i)})\\ &\quad=P\{\tilde{Z}_i\leqq 0,\ W_i-\overline{W}_i>0\mid V_{j(i)},\overline{V}_{j(i)}\}\\ &\qquad+P\{\tilde{Z}_i>0,\ W_i-\overline{W}_i\leqq 0\mid V_{j(i)},\overline{V}_{j(i)}\}.\end{aligned}$$

Using independence, we obtain for the first term of the right-hand side:

$$\begin{aligned}&P\{\tilde{Z}_i\leqq 0,\ W_i-\overline{W}_i>0\mid V_{j(i)},\overline{V}_{j(i)}\}\,I(D_i)\\ &\quad\leqq P\Big\{W_i-\Big(d+\Big|g'\Big(\Psi\Big(\frac{i}{n+1}\Big)\Big)\Big|\Big)\,n^{-1/2}\log n\\ &\qquad-\overline{W}_i-\Big(d+\Big|\bar{g}'\Big(\Psi\Big(\frac{i}{n+1}\Big)\Big)\Big|\Big)\,n^{-1/2}\log n\leqq 0,\ W_i-\overline{W}_i>0\Big\}\\ &\qquad+P\{W_i>\log n\}+P\{\overline{W}_i>\log n\}\\ &\quad\leqq C\,n^{-1/2}\log n+2\,n^{-1}, \qquad \text{for all } i=l,\ldots,m,\end{aligned}$$

with a constant C not depending on n. Here we used the fact that $W_i-\overline{W}_i$ has the density $\frac{1}{2}e^{-|x|}$. The same bound applies to the second term. Therefore,

$$\frac{1}{n}\int_{\bar{D}_n}\Big(\sum_{l\leq i\leq m} L_i - M_i\Big)^2\,dP$$
$$\leq n^{-1} N \sum_{1\leq j\leq N}\ \sum_{\beta_{j-1}+1\leq i\leq \beta_j} (2\,C\,n^{-1/2}\log n + 4\,n^{-1})$$
$$\leq N\,\tilde{C}\,n^{-1/2}\log n \sim n^{1-\gamma}\,\tilde{\tilde{C}}\,n^{-1/2}\log n.$$

Since $\gamma > \frac{1}{2}$, the last expression converges to zero and our statement is proved.

Step 5. We now compute the variables M_i to get the desired approximation of $S_{4,\varepsilon}$ by sums of independent variables. By independence,

$$\begin{aligned} M_i &= P\{\tilde{Z}_i \leq 0,\ W_i - \overline{W}_i > 0 \mid V_{j(i)}, \overline{V}_{j(i)}\} \\ &\quad - P\{\tilde{Z}_i > 0,\ W_i - \overline{W}_i \leq 0 \mid V_{j(i)}, \overline{V}_{j(i)}\} \\ &= P\{\lambda_1 W_i - \lambda_2 \overline{W}_i \leq 0,\ W_i - \overline{W}_i > 0\} \\ &\quad - P\{\lambda_1 W_i - \lambda_2 \overline{W}_i > 0,\ W_i - \overline{W}_i \leq 0\} \\ &= P\{\lambda_1 W_i - \lambda_2 \overline{W}_i \leq 0\} - P\{W_i - \overline{W}_i \leq 0\}, \end{aligned}$$

with λ_1, λ_2 as defined before.

We may assume $\lambda_1, \lambda_2 > 0$, since this is true on a set of large probability for all $i = l, \ldots, m$. Now, assuming $\lambda_1, \lambda_2 > 0$, one can easily show, that

$$M_i = \frac{\lambda_2 - \lambda_1}{2(\lambda_1 + \lambda_2)}$$
$$= \frac{\dfrac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)}(\overline{V}_{j(i)} - V_{j(i)}) - n^{-1/2}\left(\bar{g}^{-1}\left(\Psi\left(\frac{i}{n+1}\right)\right) - g'\left(\Psi\left(\frac{i}{n+1}\right)\right)\right)}{4 + \mathscr{R}_{i,5}}$$

where $\max\{|\mathscr{R}_{i,5}| : i = l, \ldots, m\} = O_P(n^{-1/2})$. Since the numerator is of the same order, $\mathscr{R}_{i,5}$ can be neglected. We may also replace $(\overline{V}_{j(i)} - V_{j(i)})$ by $(\bar{U}_i - U_i)$, as proved above.

Therefore, our statistic is asymptotically equivalent to

$$n^{-1/2}\sum_{l\leq i\leq m}\{I(W_i - \overline{W}_i \leq 0) - \tfrac{1}{2}\}$$
$$+ n^{-1/2}\,\tfrac{1}{4}\sum_{l\leq i\leq m}\frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)}(\bar{U}_i - U_i)$$
$$+ n^{-1}\,\tfrac{1}{4}\sum_{l\leq i\leq m}\left(g'\left(\Psi\left(\frac{i}{n+1}\right)\right) - \bar{g}'\left(\Psi\left(\frac{i}{n+1}\right)\right)\right).$$

The last term converges to $\frac{1}{4}\int_{\varepsilon}^{1/2}(g'(\Psi(t)) - g'(-\Psi(t)))\,dt$, using $\bar{g}'(x) = g'(-x)$.

On the other hand,

$$\sum_{l\leq i\leq m} \frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} (\bar{U}_i - U_i)$$

$$= \frac{1}{\sum_{1\leq k\leq n+1} W_k} \sum_{l\leq i\leq m} \left(\frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} \sum_{1\leq k\leq i} (\overline{W}_k - W_k) \right)$$

$$\sim \frac{1}{n} \left(\sum_{1\leq k\leq l-1} \sum_{l\leq i\leq m} \frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} (\overline{W}_k - W_k) \right.$$

$$\left. + \sum_{l\leq k\leq m} \sum_{k\leq i\leq m} \frac{\Psi''\left(\frac{i}{n+1}\right)}{\Psi'\left(\frac{i}{n+1}\right)} (\overline{W}_k - W_k) \right)$$

$$\sim \sum_{1\leq k\leq l-1} \int_{\varepsilon}^{1/2} \frac{\Psi''(t)}{\Psi'(t)} dt (\overline{W}_k - W_k)$$

$$+ \sum_{l\leq k\leq m} \int_{k/n}^{1/2} \frac{\Psi''(t)}{\Psi'(t)} dt (\overline{W}_k - W_k)$$

$$= \sum_{1\leq k\leq l-1} (\ln \Psi'(\tfrac{1}{2}) - \ln \Psi'(\varepsilon))(\overline{W}_k - W_k)$$

$$+ \sum_{l\leq k\leq m} \left(\ln \Psi'(\tfrac{1}{2}) - \ln \Psi'\left(\frac{k}{n}\right)\right)(\overline{W}_k - W_k).$$

Theorems 3 and 5 then follow from the central limit theorem. The calculation of the variance and covariance are easy and thus left to the reader.

III.5. The Neyman-Pearson-Test

We show here how the mean and the variance of the Neyman-Pearson statistic can be computed for testing symmetry with normal, logistic, double exponential or Cauchy distribution against the alternative given by

$$H_n(x) = F\left(x + \frac{x}{\sqrt{n}}\right) I(x \geqq 0) + F(x)\, I(x < 0).$$

Consider the case where

$$f(x) = F'(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Let h_n be the density of H_n.

By Taylor expansion,

$$h_n(x) \sim \left(f(x) + \frac{f(x)}{\sqrt{n}} + \frac{f'(x)\,x}{\sqrt{n}}\right) I(x \geqq 0) + f(x)\, I(x < 0).$$

Therefore,

$$\sum_{1 \leqq i \leqq n} \log \frac{h_n(x_i)}{f(x_i)} \sim n^{-1/2} \sum_{1 \leqq i \leqq n} \left(1 + \frac{f'(x_i)}{f(x_i)}\, x_i\right) I(x_i \geqq 0)$$
$$= n^{-1/2} \sum_{1 \leqq i \leqq n} (1 - x_i^2)\, I(x_i \geqq 0).$$

When the X_i are distributed according to $F(x)$, the last expression is asymptotically normal with mean 0 and variance $\sigma_5^2 = 1$.

Under the alternative, the variance remains the same, while the mean μ_5 is given by

$$\int_0^\infty (1 - x^2)\, f'(x)\, x\, dx = 1.$$

The same method applies to the other distributions. Since simple integrals only are involved, computation is left to the reader.

References

1. Antille, A.: Linéarité asymptotique d'une statistique de rang. Z. Wahrscheinlichkeitstheorie verw. Gebiete **24**, 309–324 (1972)
2. Breiman, L.: Probability. Reading, Mass.: Addison-Wesley 1968
3. Feller, W.: An introduction to probability theory and its applications, Vol. 2. New York: Wiley 1966
4. Gastwirth, J.L.: On the sign test for symmetry. J. Amer. statist. Assoc. **66**, 821–823 (1971)
5. Hàjek, J.: Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. **39**, 325–346 (1968)

Received June 22, 1976