Commentationes Mathematicae Universitatis Carolinae

Zuzana Prášková

Asymptotic expansion and a local limit theorem for the signed Wilcoxon statistic

Commentationes Mathematicae Universitatis Carolinae, Vol. 17 (1976), No. 2, 335--344

Persistent URL: http://dml.cz/dmlcz/105698

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

17,2 (1976)

ASYMPTOTIC EXPANSION AND A LOCAL LIMIT THEOREM FOR THE SIGNED WILCOXON STATISTIC

Zuzana PRÁŠKOVÁ, Praha

Abstract: In the present paper, the signed Wilcoxon statistic is investigated. Under the hypothesis of symmetry, a local limit theorem with the Edgeworth expansion is proved and an asymptotic expansion of the distribution function is derived.

Key words and phrases: The signed(one-sample) Wilcoxon statistic, characteristic function, local limit theorem, Edgeworth expansion.

AMS: Primary 62G10 Ref. Ž.: 9.74
Secondary 60E05, 60F05

---

1. Introduction. Let $X_1,\dots,X_N$ be independent random variables with a continuous distribution function (df) F, symmetric about zero. Let $Z^{(1)}< Z^{(2)}< \dots < Z^{(N)}$ denote the ordered $|X_j|$. Define ranks $D_1,\dots,D_N$ by

$$|X_{D_j}| = Z^{(j)}, \quad j = 1,\dots,N.$$

Further define random variables $T_1,\dots,T_N$ by

$$T_j = 1 \text{ if } X_{D_j} > 0,$$
$$= 0 \text{ otherwise.}$$

Under the hypothesis of symmetry, the T_j are independent with $P(T_j = 1) = 1/2$. Consider the Wilcoxon signed statistic

(1.1) $$S_N = \sum_{j=1}^{N} jT_j.$$

This statistic is a sum of independent nonidentically distributed random variables and its distribution is asymptotically normal with natural parameters

$$ES_N = N(N + 1)/4,$$

$$\text{var } S_N = N(N + 1)(2N + 1)/24.$$

The problem of Edgeworth expansion for the distribution function of S_N was investigated numerically (without studying the rate of convergence) by Fellingham and Stoker (1964) and by Claypool and Holbert (1974). A valid expansion for df to three terms can be obtained as a special case of results of Albers, Bickel and van Zwet (1974).

In the present paper, the Edgeworth expansion to p terms, $p \geq 2$, is established both for the probabilities $P(S_N = k)$ as well as for the df of S_N, in the latter case with additional terms due to the lattice character of the distribution.

2. Basic notation and main results. Let Φ and φ be the df and the density of $\mathcal{N}(0,1)$, let H_m be the Hermite polynomial of degree m. Let $\varkappa_j$, $j \geq 1$, be the cumulants of S_N, $\varkappa_1$, $\varkappa_2$ being denoted as μ, σ^2, respectively. Put

(2.1) $$\tilde{\varkappa}_j = \varkappa_j/\sigma^j.$$

For every integer $\nu \geq 1$ put

(2.2) $$Q_{2\nu}(x) = \Sigma^* H_{2\nu+2\Sigma k_j}(x) \prod_{j=2}^{\nu+1} (k_j!)^{-1}(\tilde{\varkappa}_{2j}/(2j)!)^{k_j},$$

and

(2.3) $$\bar{Q}_{2\nu}(x) = -\Sigma^* H_{2\nu+2\Sigma k_j-1}(x) \prod_{j=2}^{\nu+1} (k_j!)^{-1} \times$$
$$\times (\tilde{\varkappa}_{2j}/(2j)!)^{k_j},$$

where the summation Σ^* extends over all nonnegative integers $k_2,\ldots,k_{\nu+1}$ such that $\sum_{j=2}^{\nu+1} (j-1)k_j = \nu$ and Σ in the subscripts stands for $\sum_{j=2}^{\nu+1}$.

Notice that

(2.4) $$\frac{d}{dx} (\varphi(x)\bar{Q}_{2\nu}(x)) = \varphi(x)Q_{2\nu}(x).$$

Further define

(2.5) $$\Phi_p(x) = \Phi(x) + \varphi(x) \sum_{\nu=1}^{p-1} \bar{Q}_{2\nu}(x), \quad p \geq 2,$$

(2.6) $$B_{2\lambda}(x) = \sum_{m=1}^{\infty} \cos(2\pi mx)(2^{2\lambda-1}(\pi m)^{2\lambda})^{-1}, \quad \lambda \geq 1,$$

(2.7) $$B_{2\lambda+1}(x) = \sum_{m=1}^{\infty} \sin(2\pi mx)(2^{2\lambda}(\pi m)^{2\lambda+1})^{-1}, \quad \lambda \geq 0.$$

Throughout the paper, c and C denote positive constants, the values of which are not specified and may differ in different formulas or in different places in the same formula. Now, we shall formulate the results.

<u>Theorem 2.1.</u> Denote $x = (k - \mu)/\sigma$, $k = 0,1,\ldots$ $\ldots,N(N+1)/2$. Then

(2.8) $$\sigma P(S_N = k) = \varphi(x)(1 + \sum_{\nu=1}^{p-1} Q_{2\nu}(x)) + O(N^{-p})$$

uniformly in k.

Theorem 2.2. Let

$$h_\lambda = 1, \quad \lambda = 4m + 1,\ 4m + 2,$$

$$h_\lambda = -1, \quad \lambda = 4m + 3,\ 4m,$$

where m is an integer. Then

$$(2.9)\quad P(S_N - \mu)\sigma^{-1} < x) = \Phi_p(x) + \sum_{\lambda=1}^{r-1} h_\lambda\, \sigma^{-1} B_\lambda(\sigma x + \mu) \frac{d^\lambda}{dx^\lambda}\Phi_p(x) + O(N^{-p})$$

uniformly in x.

Remarks. 1. The sum on the right-hand side of (2.9) includes terms, which are of higher order than N^{-p}; this is due to the fact that both the expression and the proof are more easy to handle in the present form (see Esseen (1945)).

2. It can be shown that Theorems 2.1 and 2.2 are valid for the two-sample Wilcoxon statistic under the hypothesis of randomness. Theorem 2.2 for this statistic gives a generalization of a similar result of Rogers (1971). (See [11] and [9]).

3. Proofs of theorems. Let f_N, $\bar{f}_N$, $\tilde{f}_N$ be the characteristic functions (cf) of S_N, $S_N - \mu$, $(S_N - \mu)/\sigma$, respectively.

Lemma 3.1. The following inequalities hold true:

$$(3.1)\qquad |\bar{f}_N(t)| \le e^{-t^2\sigma^2/5} \text{ for } |t| \le \pi/N,$$

$$(3.2)\qquad |\bar{f}_N(t)| \le c\, e^{-N/48} \text{ for } \pi/N < |t| \le 2\pi/N,$$

(3.3) $$|\bar{f}_N(t)| < e^{-N/8} \text{ for } 2\pi/N < |t| \le \pi .$$

Proof. It can be easily shown that

(3.4) $$\bar{f}_N(t) = \prod_{j=1}^{N} \cos (tj/2).$$

The symmetry of $\bar{f}_N$ allows us to consider $0 \le t \le \pi$ only. The proof is given in three steps as follows:

1. Let $0 < t \le \pi/N$. Then for $1 \le j \le N$

$$|\cos (tj/2)| = \cos (tj/2) \le 1 - t^2 j^2 (2\pi^2)^{-1}$$

and

$$|\bar{f}_N(t)| \le \prod_{j=1}^{N} (1 - t^2 j^2 (2\pi^2)^{-1} \le$$
$$\le \exp (-t^2 \sum_{j=1}^{N} j^2 (2\pi^2)^{-1}) = e^{-2t^2 \sigma^2 \pi^{-2}} \le e^{-t^2 \sigma^2/5}.$$

2. Let $\pi/N < t \le 2\pi/N$, let K denote the integral part of π/t. Then

$$|\bar{f}_N(t)| \le \prod_{j=1}^{K} |\cos (tj/2)| \le \prod_{j=1}^{K} (1 - t^2 j^2 (2\pi^2)^{-1}) \le$$
$$\le \exp (-t^2 \sum_{j=1}^{K} j^2 (2\pi^2)^{-1}) \le e^{-t^2 K^3 (6\pi^2)^{-1}}.$$

Making use of the inequalities $-1 + \pi/t \le K \le \pi/t$ and $N/2 \le \pi/t < N$ we obtain for $\pi/N < t \le 2\pi/N$

$$|\bar{f}_N(t)| \le e^{-t^2 (N/2-1)^3 (6\pi^2)^{-1}} \le c\, e^{-N/48}.$$

3. For $2\pi/N < t \le \pi$ we shall consider the inequalities

(3.5) $$|\cos x| = (1 - (1 - \cos 2x)/2)^{1/2} \le 1 - (1 - \cos 2x)/4,$$

(3.6) $\sin x \geq 2x/\pi$ valid for $0 \leq x \leq \pi/2$

and the formula

(3.7) $$\sum_{j=1}^{N} \cos tj = \sin(Nt/2)\cos((N+1)t/2)(\sin(t/2))^{-1}.$$

From (3.5) we obtain

$$|\bar{f}_N(t)| \leq \prod_{j=1}^{N} (1-(1-\cos tj)/4) \leq \exp(-N/4 + \sum_{j=1}^{N} \cos tj/4)$$

and utilizing (3.7) and (3.6) we have

$$\sum_{j=1}^{N} \cos tj \leq (\sin(t/2))^{-1} \leq \pi/t < N/2.$$

Hence

$$|\bar{f}_N(t)| < e^{-N/8} .$$

<u>Lemma 3.2.</u> For $|t| < \sigma^{-1+\alpha}$, $0 < \alpha < 1/3$, the following expansion holds:

(3.8) $$\log \bar{f}_N(t) = \sum_{n=1}^{r} (it)^{2n} \varkappa_{2n}/(2n)! + R(t),$$

where

(3.9) $$\varkappa_{2n} = (2^{2n} - 1) \beta_{2n} \sum_{j=1}^{N} j^{2n}(2n)^{-1},$$

(3.10) $$|R(t)| \leq C N^{2r+3} t^{2r+2}$$

and β_{2n} are the Bernoulli numbers.

<u>Proof</u>. We shall use the expansion

(3.11) $$\log \cos t = \sum_{n=1}^{\infty} (-1)^n 2^{2n}(2^{2n} - 1) \beta_{2n} t^{2n}((2n)!2n)^{-1},$$

valid for $|t| < \pi/2$. Suppose that $|t| < \pi/N$. Then we have

$$\log \bar{f}_N(t) = \sum_{j=1}^{N} \sum_{n=1}^{\infty} (-1)^n 2^{2n}(2^{2n}-1) \beta_{2n}(tj/2)^{2n}((2n)!2n)^{-1} =$$

$$= \sum_{n=1}^{r}(-1)^n(2^{2n}-1)\,\beta_{2n}t^{2n}((2n)!\,2n)^{-1}\sum_{j=1}^{N} j^{2n} + R(t) =$$

$$= \sum_{n=1}^{r}(it)^{2n}\,\varkappa_{2n}((2n)!)^{-1} + R(t),$$

where

$$R(t) = \sum_{j=1}^{N}\sum_{n=r+1}^{\infty}(-1)^n(2^{2n}-1)\,\beta_{2n}(tj)^{2n}((2n)!2n)^{-1}.$$

Utilizing the inequality $|\beta_{2n}| \leq 4(2n)!(2\pi)^{-2n}$ we have

$$|R(t)| \leq \sum_{j=1}^{N}\sum_{n=r+1}^{\infty}(tj/\pi)^{2n} \leq \sum_{j=1}^{N}(tj/\pi)^{2r+2}(1 -$$

$$-(tj/\pi)^2)^{-1} \leq N(tN/\pi)^{2r+2}(1-(tN/\pi)^2)^{-1}.$$

Now, let us suppose $|t| < \sigma^{-1+\alpha}$, $0 < \alpha < 1/3$. Then

$$|Nt/\pi| \leq N\pi^{-1}\sigma^{-2/3} \leq 12^{1/3}\pi^{-1} < 1,$$

so that

$$|R(t)| \leq CN^{2r+3}\,t^{2r+2}.$$

Lemma 3.3. For $|t| < \sigma^{\alpha}$, $0 < \alpha < 1/6$,

$$\tilde{f}_N(t) = e^{-t^2/2}\left(1 + \sum_{\nu=1}^{r-1} P_{2\nu}(it)\right) + Z(t), \tag{3.12}$$

where $P_{2\nu}$ are polynomials of degree 4ν in it, coefficients of which are of order $N^{-\nu}$, or, explicitly, in the notation of Section.2,

$$P_{2\nu}(it) = \sum{}^{*}\prod_{j=2}^{\nu+1}(k_j!)^{-1}(\tilde{\varkappa}_{2j}/(2j)!)^{k_j}(it)^{2\nu+2\sum k_j} \tag{3.13}$$

and the remainder Z satisfies the inequality

$$|Z(t)| \leq c\,e^{-t^2/2}N^{-p}|t|^{p}|Z_p(t)|, \tag{3.14}$$

where Z_p is a polynomial in t depending only on p.

Proof. It follows from Lemma 3.2 and from [9, Lemma 3.2 and Lemma 3.3] .

Proof of Theorem 2.1. Making use of the formula

$$P(S_N = k) = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-ivk} f_N(v)dv$$

and putting $v = t/\sigma$ and $x = (k - \mu)/\sigma$, we have for $0 < \alpha < 1/6$

$$\sigma P(T_N = k) = (2\pi)^{-1} \int_{-\pi\sigma}^{\pi\sigma} e^{-itx} \tilde{f}_N(t)dt =$$

$$= (2\pi)^{-1}\left(\int_{-\infty}^{\infty} - \int_{|t|\geq\sigma^\alpha}\right) e^{-itx-t^2/2} \left(1 + \sum_{\nu=1}^{r-1} P_{2\nu}(it)\right)dt +$$

$$+ (2\pi)^{-1} \int_{|t|<\sigma^\alpha} e^{-itx-t^2/2} Z(t)dt +$$

$$+ (2\pi)^{-1} \int_{\sigma^\alpha\leq|t|\leq\pi\sigma} e^{-itx} \tilde{f}_N(t)dt.$$

The Fourier transform implies

$$(2\pi)^{-1} \int_{-\infty}^{\infty} e^{-itx-t^2/2} \left(1 + \sum_{\nu=1}^{r-1} P_{2\nu}(it)\right)dt =$$

$$= \varphi(x)\left(1 + \sum_{\nu=1}^{r-1} Q_{2\nu}(x)\right),$$

$Q_{2\nu}$ being defined by (2.2). The rest of the proof follows from Lemmas 3.1 and 3.3.

Proof of Theorem 2.2. The proof is similar to that of Theorem 2.2 in [9] and therefore is omitted.

References

[1] ALBERS W., BICKEL P.J., van ZWET W.R.: Asymptotic

expansions for the power of distribution free tests in the one-sample problem, Amsterdam 1974 (preprint).

[2] BICKEL P.J.: Edgeworth expansions in nonparametric statistics, Ann. Statist. 2(1974),1-20.

[3] CLAYPOOL P.L., HOLBERT D.: Accuracy of normal and Edgeworth approximations to the distribution of the Wilcoxon signed rank statistic , J. Amer. Stat. Assoc. 69(1974), 255-258.

[4] DANILOV V.L.: A Survey of Mathematical Analysis I (in Czech), SNTL (Prague, 1968).

[5] ESSEEN C.G.: Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law, Acta Mathematica 77(1945), 1-125.

[6] FELLINGHAM S.A., STOKER D.J.: An approximation for the exact distribution of Wilcoxon test for symmetry, J. Amer. Stat. Assoc. 59(1964), 899-905.

[7] HÁJEK J.: Nonparametric Statistics, Holden-Day (San Francisco, 1969).

[8] PETROV V.V.: The Sums of Independent Random Variables (in Russian), Nauka (Moscow, 1972).

[9] PRÁŠKOVÁ-VÍZKOVÁ Z.: Asymptotic expansion and a local limit theorem for a function of the Kendall rank correlation coefficient, to appear in Ann. Statist. 4(1976).

[10] ROGERS W.F.: Exact null distributions and asymptotic expansions for rank test statistics, Technical Report (Stanford University, 1971).

[11] VÍZKOVÁ Z.: Local limit theorems for some rank test statistics (in Czech), Ph D thesis (Prague, 1974).

Matematicko-fyzikální fakulta
Karlova universita
Sokolovská 83, 18600 Praha 8
Československo

(Oblatum 25.2. 1976)