

## ASYMPTOTIC PROPERTIES OF RANK TESTS OF SYMMETRY UNDER DISCRETE DISTRIBUTIONS

BY DANA VORLÍČKOVÁ

Charles University

The paper deals with problems of rank tests of symmetry when samples are drawn from purely discrete distributions so that ties of zero and non-zero observations may occur. Zero observations are considered in the same way as nonzero ones. Two ways of treatment of ties are used in the paper, randomization of ties and the method of averaged scores. The asymptotic distributions of the statistics are derived under hypothesis of symmetry and under contiguous alternatives of location. The asymptotic power and efficiency of tests are established.

**1. Introduction.** Let  $X_1, \dots, X_N$  be a random sample from a discrete distribution under which random variables  $X_i$ ,  $1 \leq i \leq N$ , may take values  $k = 0, \pm 1, \pm 2, \dots$  only. (The set of integers could be replaced by an arbitrary countable set.)

We say that the observations  $X_1, \dots, X_N$  satisfy the hypothesis of symmetry  $\bar{H}_1$ , if they are independent, equally discretely distributed over the set of all integers  $C$  and their common distribution function satisfies  $F(x - 0) = 1 - F(-x)$ ,  $-\infty < x < +\infty$ .

We also denote by the symbol  $\bar{H}_1$  the system of all  $N$ -dimensional discrete densities (i.e., densities with respect to the  $N$ -dimensional counting measure)  $p$  such that

$$p(x_1, \dots, x_N) = \prod_{i=1}^N p_1(x_i), \quad x_i \in C, 1 \leq i \leq N,$$

where  $p_1(x)$  is an arbitrary symmetric one-dimensional discrete density, that is,  $p_1(x) = p_1(-x)$ ,  $x \in C$ .

Let  $R_i$  be the rank of  $|X_i|$  among  $|X_1|, \dots, |X_N|$ , i.e.,

$$(1) \quad R_i = \sum_{j=1}^N u(|X_i| - |X_j|), \quad 1 \leq i \leq N,$$

where  $u(x) = 1$  or  $0$  according as  $x$  is  $\geq$  or  $<$   $0$ , and  $R$  will stand for a vector of ranks  $(R_1, \dots, R_N)$ . The ordered sample  $|X|^{(\cdot)}$  consists of  $g + 1$  groups of equal observations called ties, the  $j$ th tie ( $0 \leq j \leq g$ ) containing  $\tau_j$  observations;  $\tau = (\tau_0, \tau_1, \dots, \tau_g) = \tau(|X|^{(\cdot)})$ ,  $\tau_0$  denotes the number of zeros in the sample,  $\sum_{j=0}^g \tau_j = N$ .

Let  $a_i$ ,  $1 \leq i \leq N$ , be arbitrary scores and let  $\text{sgn } x = 1, 0$  or  $-1$  according as  $x >, =$  or  $<$   $0$ .

In [2] there are introduced rank tests of hypothesis of symmetry in the continuous case which are based on statistics of the form  $\sum_{i=1}^N \text{sgn } X_i a(R_i)$ . Since there are difficulties caused by ties and zero observations under discrete distributions and the vector  $R$  has not the same distribution under all distributions

---

Received February 17, 1971; revised March 1, 1972.

from  $\bar{H}_1$ , we cannot use the above statistic. Hence we shall treat ties of absolute values of observations (including zeros) as ties of observations in testing hypothesis of randomness (see [5]). We shall consider a more general type of statistics with known regression constants  $c_1, \dots, c_N$ .

Randomization of ties leads to statistics

$$(2) \quad S^* = \sum_{i=1}^N c_i \operatorname{sgn} X_i a(R_i^*),$$

where

$$(3) \quad R_i^* = \sum_{j=1}^N u(|X_i| + U_i - |X_j| - U_j), \quad 1 \leq i \leq N,$$

$U_1, \dots, U_N$  being a supplementary independent sample from the uniform distribution over the interval  $[0, 1]$ .

Under the method of averaged scores we utilize statistics

$$(4) \quad \bar{S} = \sum_{i=1}^N c_i \operatorname{sgn} X_i a(R_i, \tau)$$

with  $R_i$  given by (1) and  $a(i, \tau)$  defined as

$$(5) \quad a(i, \tau) = \frac{1}{\tau_k} \sum_{j \leq T_{k-1} + 1}^{T_k} a_j, \quad T_{k-1} < i \leq T_k,$$

where  $T_j = \tau_0 + \dots + \tau_j$ ,  $1 \leq j \leq g$ .

Under  $\bar{H}_1$ , vectors  $R^*$  and  $\tau$  are independent,  $R^*$  has the uniform distribution over the space of all permutations  $\mathcal{B}$  and

$$(6) \quad E(S^* | \tau) = 0,$$

$$(7) \quad \operatorname{Var}(S^* | \tau) = N^{-1} \sum_{i=1}^N c_i^2 \sum_{j=\tau_0+1}^N a_j^2,$$

$$(8) \quad E(\bar{S} | \tau) = 0,$$

$$(9) \quad \operatorname{Var}(\bar{S} | \tau) = N^{-1} \sum_{i=1}^N c_i^2 \sum_{j=\tau_0+1}^N a^2(j, \tau)$$

holds.

Our treatment of nonzero ties is the same as Hájek's method [1] and we use statistics of the same type, but zero observations are deleted in [1] so that scores  $a_i$  with  $i > N - \tau_0$  fail in the statistics [1] unlike the statistics  $S^*$  and  $\bar{S}$  in this paper, where scores  $a_i$  with  $i < \tau_0$  do not occur.

In [1] there are also derived asymptotic distributions of the respective test statistics under hypothesis of symmetry, but under different conditions than in this paper. Pratt [3] recommends that we consider an ordered sample with zero observations but he does not construct a theory of test statistics and their distribution. Putter [4] deals with the sign test when zeros are present (in this special case ties of nonzero observations are irrelevant) and considers either deleting zeros or randomly assigning positive and negative signs to zero observations.

**2. Asymptotic distribution of test statistics under  $\bar{H}_1$ .** Let  $X_1, \dots, X_N$  be a random sample satisfying  $\bar{H}_1$ . Denote by  $G$  the common distribution function of random variables  $|X_i|$ ,  $1 \leq i \leq N$ , and by  $G^*$  the distribution function of random variables  $|X_i| + U_i$ ,  $1 \leq i \leq N$ ; put  $U_i^* = G^*(|X_i| + U_i)$ .  $G^*$  is continuous so that  $U_i^*$ ,

$1 \leq i \leq N$ , has uniform distribution over  $[0, 1]$  and  $R_i^*$  defined by (3) stands for the rank of  $U_i^*$ ,  $1 \leq i \leq N$ .

Let  $\varphi(u)$ ,  $0 < u < 1$ , be (throughout the paper) an arbitrary nonconstant square-integrable function and let the scores satisfy

$$(10) \quad \int_0^1 (a(1 + [uN]) - \varphi(u))^2 du \rightarrow 0 \quad \text{as } N \rightarrow \infty .$$

Put

$$(11) \quad T^* = \sum_{i=1}^N c_i \operatorname{sgn} X_i \varphi(U_i^*) .$$

**THEOREM 1.** *Let  $\bar{H}_1$  hold and  $|X_i|$ ,  $1 \leq i \leq N$ , have a distribution function  $G$ . Let scores  $a_i$ ,  $1 \leq i \leq N$ , satisfy (10) with  $\varphi$  such that  $\int_{G(0)}^1 \varphi^2(u) du > 0$ . Then, statistics  $S^*$  defined by (2) are asymptotically normal  $(0, \sigma_N^2)$ , where  $\sigma_N^2$  is given by (7) or*

$$(12) \quad \sigma_N^2 = \int_{G(0)}^1 \varphi^2(u) du \sum_{i=1}^N c_i^2 .$$

Moreover,  $|T^* - S^*|/\sigma_N \rightarrow 0$  as  $N \rightarrow \infty$  in probability  $P \in \bar{H}_1$ ,  $T^*$  given by (11).

**PROOF.** Denote  $E(\cdot | \tau)$  by  $E^\tau(\cdot)$ ; then

$$\begin{aligned} E^\tau(S^* - T^*)^2 &\leq \sum_{i=1}^N c_i^2 E^\tau(a(R_i^*) - \varphi(U_i^*))^2 \\ &= E(a(R_1^*) - \varphi(U_1^*))^2 \sum_{i=1}^N c_i^2 . \\ \frac{E^\tau(S^* - T^*)^2}{\operatorname{Var}(S^* | \tau)} &\leq \frac{E(a(R_1^*) - \varphi(U_1^*))^2 \sum_{i=1}^N c_i^2}{\operatorname{Var}(S^* | \tau)} \rightarrow 0 \quad \text{as } N \rightarrow \infty \end{aligned}$$

because  $1/\sum c_i^2 \operatorname{Var}(S^* | \tau) \rightarrow \int_{G(0)}^1 \varphi^2(u) du > 0$  and  $E(a(R_1^*) - \varphi(U_1^*))^2 \rightarrow 0$  for the same reason as in the continuous case (see [2], Theorem V.1.7). Asymptotic normality of  $T^*$  according to Theorem V.1.2 of [2] concludes the proof.  $\square$

**THEOREM 2.** *Under the assumptions of Theorem 1, the statistic  $\sum_{i=1}^N c_i \operatorname{sgn} X_i \times a(R_i^*, \varphi)$  with  $a(i, \varphi) = E(\varphi(U^{(i)}))$ ,  $1 \leq i \leq N$ , is asymptotically normal  $(0, \sigma_N^2)$  with  $\sigma_N^2$  given by (7) or (12).*

If  $G$  is the common distribution function of  $|X_1|, \dots, |X_N|$  under  $\bar{H}_1$  put

$$\begin{aligned} I_k &= [G(k - 1), G(k)) , & k \in C, k \geq 1 , \\ I_0 &= (0, G(0)) . \end{aligned}$$

It is obvious that  $\bigcup_{k=0}^\infty I_k = (0, 1)$ .

Denoting Lebesgue measure by  $\mu$ , define

$$(13) \quad \begin{aligned} \bar{\varphi}(u) &= \frac{1}{\mu I_k} \int_{I_k} \varphi(v) dv , & u \in I_k, \mu I_k > 0 , \\ &= 0 , & u \in I_k, \mu I_k = 0 . \end{aligned}$$

Put

$$(14) \quad \tilde{S} = \sum_{i=1}^N c_i \operatorname{sgn} X_i a(R_i^*, \bar{\varphi}) ,$$

where  $a(i, \bar{\varphi}) = E(\bar{\varphi}(U^{(i)}))$ .

**THEOREM 3.** Let  $\bar{H}_1$  hold and  $|X_i|$ ,  $1 \leq i \leq N$ , have a distribution function  $G$ . Let  $a(i, \tau)$ ,  $1 \leq i \leq N$ , be defined by (5) for  $a(i)$  satisfying (10) with  $\varphi$  such that  $\int_{G(0)}^1 (\tilde{\varphi}(u))^2 du > 0$ . Then, the statistic  $\bar{S}$  defined by (4) is asymptotically normal  $(0, \bar{\sigma}_N^2)$ , where

$$(15) \quad \bar{\sigma}_N^2 = \int_{G(0)}^1 (\tilde{\varphi}(u))^2 du \sum_{i=1}^N c_i^2 .$$

**PROOF.** According to Theorem 2,  $\bar{S}$  is asymptotically normal  $(0, \bar{\sigma}_N^2)$ . Now, if we prove that  $E[(\bar{S} - \tilde{S})^2 | \tau] / \bar{\sigma}_N^2 \rightarrow 0$  as  $N \rightarrow \infty$  in probability  $P \in \bar{H}_1$ ,  $(\bar{S} - \tilde{S}) / \bar{\sigma}_N \rightarrow 0$  in probability and  $\bar{S} / \bar{\sigma}_N$  has the same asymptotic distribution as  $\tilde{S} / \bar{\sigma}_N$ .

$$E[(\bar{S} - \tilde{S})^2 | \tau] = \int_{\tau_0/N}^1 (a(1 + [uN], \tau) - a(1 + [uN], \tilde{\varphi}))^2 du \sum_{i=1}^N c_i^2 .$$

Now, we obtain the desired result by following the pattern of the proof of Theorem 3.3 of [5].  $\square$

**THEOREM 4.** Under the assumptions of Theorem 3, the conditional distribution of the statistic  $\bar{S}$  given  $\tau$  is under  $\bar{H}_1$  asymptotically normal  $(0, \text{Var}(\bar{S} | \tau))$ , where  $\text{Var}(\bar{S} | \tau)$  is given by (9).

**PROOF.** Follow the pattern of the proof of Theorem 4 of [5].  $\square$

**3. Asymptotic distribution of the statistics under alternatives.** We shall restrict ourselves to contiguous alternatives (see in [2], e.g., the definition of contiguity). It enables us to use LeCam’s lemmas published in [2], Chapter VI, e.g.

Let  $\mathcal{S}$  be an open interval containing zero, a family of densities  $d(x, \Delta) = d(x - \Delta)$ ,  $\Delta \in \mathcal{S}$ , satisfying the following conditions:

- (a)  $d(x) = d(-x)$ ,  $x \in C$ .
- (16) (b)  $d(x, \Delta)$  is absolutely continuous in  $x$ ,
- (c)  $d(x) = \lim_{\Delta \rightarrow 0} \Delta^{-1} [d(x, \Delta) - d(x)]$  exists for all  $x \in C$ ;
- (d)  $0 < \lim_{\Delta \rightarrow 0} \sum_{k=0}^{\infty} [d'(k - \Delta)]^2 / d(k - \Delta) = \sum_{k=0}^{\infty} [d'(k)]^2 / d(k) < +\infty$ .

$\sum_{k=0}^{\infty} [d'(k)]^2 / d(k)$  is Fisher’s information of a symmetric discrete density  $d(x)$  and we denote it by  $I(d)$ .

Let us consider a sequence of pairs of hypotheses and alternatives  $\{\bar{H}_{1n}, q_{\Delta_n n}\}$ , where  $\bar{H}_{1n}$  is  $\bar{H}_1$  applied to  $N_n$  observations and  $q_{\Delta_n n}$  is given by  $N_n$ -dimensional discrete density

$$(17) \quad q_{\Delta_n n}(x_1, \dots, x_{N_n}) = \prod_{i=1}^{N_n} d(x_i - \Delta_n), \quad x_i \in C, \Delta_n \in \mathcal{S},$$

where  $d(x)$  is a known symmetric discrete density satisfying (16). We shall choose to the density  $q_{\Delta_n n}$  a density  $p_n$  in  $\bar{H}_{1n}$ ,

$$(18) \quad p_n(x_1, \dots, x_{N_n}) = \prod_{i=1}^{N_n} d(x_i) .$$

It will be shown that under certain conditions,  $q_{\Delta_n n}$  is contiguous to  $p_n$  and consequently to  $\bar{H}_{1n}$ . To simplify notation we shall suppress the index  $n$  in the sequel.

We denote by  $L_\Delta$  a likelihood ratio,

$$(19) \quad L_\Delta = \prod_{i=1}^N d(X_i - \Delta) / d(X_i) ,$$

and introduce statistics  $W_\Delta$ ,

$$(20) \quad W_\Delta = 2 \sum_{i=1}^N \{ [d(X_i - \Delta)/d(X_i)]^{\frac{1}{2}} - 1 \}.$$

and  $T$ ,

$$(21) \quad T = \sum_{i=1}^N \operatorname{sgn} X_i d'(|X_i|)/d(|X_i|).$$

Let us assume that

$$(22) \quad N\Delta^2 \rightarrow b^2, \quad 0 < b^2 < +\infty.$$

LEMMA. If  $\operatorname{Var}$  denotes a variance with respect to  $p$ , then, under (22), for  $W_\Delta$  given by (20) and  $T$  given by (21),

$$(23) \quad \lim_{N \rightarrow \infty} \operatorname{Var} (W_\Delta - \Delta T) = 0$$

holds.

PROOF. Let us denote by  $\sum^*$  a sum over such  $k, k \in C$ , that  $d(k) > 0$ . After some reflections, we obtain

$$\begin{aligned} \operatorname{Var}(W_\Delta - \Delta T) &= 4\Delta^2 \sum_{i=1}^N \operatorname{Var} \left\{ \frac{1}{d^{\frac{1}{2}}(X_i)} \left[ \frac{d^{\frac{1}{2}}(X_i - \Delta) - d^{\frac{1}{2}}(X_i)}{\Delta} - \operatorname{sgn} X_i \frac{d'(|X_i|)}{2d^{\frac{3}{2}}(|X_i|)} \right] \right\} \\ &= 4N\Delta^2 \sum_{k=-\infty}^{+\infty} \left[ \frac{d^{\frac{1}{2}}(k - \Delta) - d^{\frac{1}{2}}(k)}{\Delta} - \frac{d'(k)}{2d^{\frac{3}{2}}(k)} \right]^2, \end{aligned}$$

which tends to zero for the same reason as in the proof of Lemma 2 of [5].  $\square$

Proceeding analogically to the proof of Theorem 4.2 of [5] we attain the following assertion.

THEOREM 5. Let (10) and (22) hold, and let  $\int_{G(0)}^1 \varphi^2(u) du > 0$ . Then, under  $q_\Delta$ , statistics  $S^*$  given by (2) are asymptotically normal with parameters

$$(24) \quad \mu_\Delta = N\Delta \sum_{k=1}^\infty \frac{d'(k)}{d(k)} \int_{G(k-1)}^{G(k)} \varphi(v) dv$$

and  $\sigma_N^2$  given by (12).

THEOREM 6. Let (10) and (22) hold and let  $\int_{G(0)}^1 (\tilde{\varphi}(u))^2 du > 0$  for  $\tilde{\varphi}$  defined by (13). Then, under  $q_\Delta$ , statistics  $\bar{S}$  defined by (4) are asymptotically normal with parameters (24) and (15).

PROOF. Follow the pattern of the proof of Theorem 4.3 from [5].  $\square$

**4. Asymptotic power and efficiency of tests.** Put  $c_1 = c_2 = \dots = 1$  in  $S^*$ ,  $\bar{S}$  respectively. Asymptotic normality  $(0, \operatorname{Var}(S^* | \tau))$  of the statistic  $S^*$  and asymptotic normality  $(0, \operatorname{Var}(\bar{S} | \tau))$  of the conditional distribution of the statistics  $\bar{S}$  enable us to establish the asymptotic power of tests of  $\bar{H}_1$  against  $q_\Delta$  based on  $S^*$  and  $\bar{S}$  (the  $S^*$ -test,  $\bar{S}$ -test, respectively).

Under the assumptions of Theorem 5, the asymptotic power of the  $S^*$ -test attains the value  $1 - \Phi(k_{1-\alpha} - \mu_\Delta \sigma_N^{-1})$ ,  $\mu_\Delta$  given by (24) and  $\sigma_N^2$  by (12). Similarly,

under the assumptions of Theorem 6, we find the asymptotic power of the  $\bar{S}$ -test as  $1 - \Phi(k_{1-\alpha} - \mu_\Delta(\bar{\sigma}_N)^{-1})$ ,  $\bar{\sigma}_N^2$  given by (15).

If scores  $a(i, \tau)$ ,  $1 \leq i \leq N$ , in  $\bar{S}$  are generated by scores  $a_i$ ,  $1 \leq i \leq N$ , satisfying (10) with  $\varphi(u) = \varphi^+(u, d, 0)$ , where

$$\varphi^+(u, d, t) = \frac{d'(D^{-1}(\frac{1}{2}u + \frac{1}{2}, t), t)}{d(D^{-1}(\frac{1}{2}u + \frac{1}{2}, t), t)}, \quad 0 < u < 1,$$

$d$  is a density satisfying (16),  $D(x, t) = \sum_{k \leq x} d(k, t)$ ,  $D^{-1}(u, t) = \inf\{x : D(x, t) \geq u\}$ , the  $\bar{S}$ -test is asymptotically the most powerful among all tests of  $\bar{H}_1$  against  $q_\Delta$  given by (17). It can be seen from the following considerations: Statistic (21) may be written as  $T = \sum_{i=1}^N \text{sgn } X_i \cdot \varphi^+(U_i^*, d, 0)$ , so that it follows from Theorem 1 after some reflections that  $T \sim \tilde{S}_d$ , where  $\tilde{S}_d = \sum_{i=1}^N \text{sgn } X_i a(R_i^*, \tilde{\varphi}^+(\cdot, d, 0))$  (see (14)). In the proof of Theorem 3 there is derived that  $\tilde{S}_d \sim \bar{S}$ . Using the proof of Theorem 5 we obtain  $\bar{S} \sim (1/\Delta) \log L_\Delta + (1/2\Delta)I(d)b^2$ . The most powerful test of  $p$  (18) against  $q_\Delta$  (17) is based on  $\log L_\Delta$ . Let us denote its power by  $\beta(\alpha, p, q_\Delta)$  and the power of the most powerful test of  $\bar{H}_1$  against  $q_\Delta$  by  $\beta(\alpha, \bar{H}_1, q_\Delta)$ . Obviously,

$$\lim \beta(\alpha, p, q_\Delta) \geq \limsup \beta(\alpha, \bar{H}_1, q_\Delta)$$

holds. The  $\bar{S}$ -test also tests  $\bar{H}_1$  against  $q_\Delta$  because it has the asymptotic size  $\alpha$  for all densities from  $\bar{H}_1$ . Then,

$$\lim \beta(\alpha, p, q_\Delta) \leq \liminf \beta(\alpha, \bar{H}_1, q_\Delta).$$

Now, the above two inequalities imply our assertion.

Defining the asymptotic efficiency  $e$  of the  $S^*$ -test as

$$e = (\mu_\Delta \bar{\sigma}_N / \mu_\Delta \sigma_N)^2 = \bar{\sigma}_N^2 / \sigma_N^2,$$

we obtain

$$e = \int_{G(0)}^1 (\tilde{\varphi}(u))^2 du / \int_{G(0)}^1 \varphi^2(u) du.$$

Obviously,  $e \leq 1$  always holds.

**Acknowledgment.** The author thanks the referee for his helpful suggestion.

#### REFERENCES

- [1] HÁJEK, J. (1969). *A Course in Nonparametric Statistics*. Holden-Day, San Francisco.
- [2] HÁJEK, J. and ŠIDÁK, ZB. (1967). *Theory of Rank Tests*. Academia, Praha.
- [3] PRATT, J. W. (1959). Remark on zeros and ties in the Wilcoxon signed rank procedures. *J. Amer. Statist. Assoc.* **54** 655-667.
- [4] PUTTER, J. (1955). The treatment of ties in some nonparametric tests. *Ann. Math. Statist.* **26** 368-386.
- [5] VORLÍČKOVÁ, D. (1970). Asymptotic properties of rank tests under discrete distributions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 275-289.

DEPARTMENT OF STATISTICS  
CHARLES UNIVERSITY  
SOKOLOVSKÁ 83  
PRAHA 8-CZECHOSLOVAKIA