ON A CONSTRUCTION OF CERTAIN OPTIMUM RANK TESTS

Tamura, Ryoji
Shimane University

https://doi.org/10.5109/13010

出版情報：統計数理研究. 11 (1/2), pp.19-24, 1964-03. Research Association of Statistical Sciences
バージョン：
権利関係：

ON A CONSTRUCTION OF CERTAIN OPTIMUM RANK TESTS

By

Ryoji TAMURA

(*Received* September 1, 1963)

§ 1. **Introduction.** We have in the previous paper [6] derived a modification of certain well-known rank sum tests in the sense of raising their asymptotic efficiency. For this purpose, the following forms have been proposed as the test statistics

$$mT^k = \sum_{i=1}^{N} E^k_{N,i} Z_i, \quad k>0 \tag{1}$$

where Z_i is 1 or 0 if the i-th smallest in the combined sample is X or Y and

$$E^k_{N,i} = \begin{cases} (i/N)^k & \text{for locations problems} \\ [(i/N)^k + ((N+1-i)/N)^k - |(i/N)^k - ((N+1-i)/N)^k|]/2 & \text{for scale problems.} \end{cases} \tag{2}$$

We shall also notice that all notations in this paper are followed by [6]. In [6], the asymptotic normality of the statistics T_k has been proved and moreover the asymptotic efficiency has been calculated for some $k>0$ and some alternatives. We shall deal in this paper with more general statistics among the form (1), i. e. $E^k_{N,i}$ is some function of rank, and intend to construct the tests with some optimum properties.

§ 2. **Wilcoxon's type.** Our purpose is to test the hypothesis $\theta=0$ against the alternative $\theta>0$ based on the two samples $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ from the distributions $F(x)$ and $F(x+\theta)$. Then we define the statistics $U[h]$ as an extension of Wilcoxon's U

$$m\tilde{U}[h] = \sum_{i=1}^{N} h(i/1+N) Z_i, \quad N=m+n \ , \ \lambda_N = m/N \tag{3}$$

,where we assume $h(t)$ to satisfy the following conditions

(i) $h(1)=0(\sqrt{N})$

(4) (ii) $|h^{(i)}(t)|\leq K\{t(1-t)\}^{-i-1/2+\delta}$ for $i=0,1,2$ and some $\delta>0$

(iii) $\int_0^1 h(t)\ dt=0,\quad \int_0^1 h^2(t)\ dt=1.$

Assumptions (i) and (ii) are needed for the asmyptotic normality of $\tilde{U}[h]$ and (iii) is only a normalized condition. Following [2], we may prove the asymptotic normality of $\tilde{U}[h]$ by the same tecknique as in [6]. Moreover applying the theory of Chernoff-Savage [2], we may get the efficacy $E^2(\tilde{U})$ of our $\tilde{U}[h]$ tests if the mean value of $\tilde{U}$ has a finite derivative under the hypothesis

$$E^2(\tilde{U})=\lambda(1-\lambda)[\int_{-\infty}^{\infty} f(x)h'[F(x)]\quad dF(x)]^2 \tag{5}$$

,where $h'[F(x)]$ means the derivative with regard to $F(x)$.
If there exists some suitable function $h(t)$ maximizing the value of $E^2(\tilde{U})$, it will give an optimum statistic within the class of our form (3). Then it is sufficient to maximize the integral

$$\begin{aligned} I[h] &= \int_{-\infty}^{\infty} f(x)h'[F(x)]\ dF(x) \\ &= \int_0^1 f[F^{-1}(t)]h'(t)dt. \end{aligned} \tag{6}$$

(a) Normal case. We may use the method of undetermined multipliers in the calculas of variations to maximize $I[h]$ subject to the assumption (iii). Thus from

$$\int_0^1 \{h'(t)\varphi[\Phi^{-1}(t)]-\alpha h^2(t)\}\ dt=\max$$

,where Φ is the standard normal distribution function with density φ and α is an undetermined multiplier, we may obtain the following solution of Eular equation for variations,

$$h(t)=\Phi^{-1}(t). \tag{7}$$

The statistic $\tilde{U}[h]$ with the above $h(t)$ derives what is called as the Normal scores test that has been shown to be locally most powerful rank test for the normal alternative by Dwass [3] and others. Thus it has been shown that an optimum test also results from the point of view of our generalization.

(b) Uniform case. From the form (6),

$$I[h]=\lim_{t\to 1} h(t)-\lim_{t\to 0} h(t).$$

Thus we must give more weight (positive or negative) at extreme ranks. Now we assume $h(t)$ to be $\Phi^{-1}(t)$ for a time, then the expectation $\mu(\theta)$ has not a finite derivative $\mu'(0)$ at $\theta=0$. On the other hand, the fact that the efficiency of $\tilde{U}_k$ tests approach to infinity when $k\to 0$ as has been shown in [6] leads us to the following consideration. Construct the normalized function $h_0(t)$ corresponding to the case $k\to 0$ in [6] as follows,

(8)
$$\begin{aligned} h_0(t) &= \lim_{k\to 0} \frac{(k+1)\sqrt{2k+1}}{k}\left(t^k-\frac{1}{k+1}\right) \\ &= \log t+1. \end{aligned}$$

However our test statistic

(9)
$$m\tilde{U}[h_0]=\sum_{i=1}^{N}\left(\log\frac{i}{N+1}+1\right)Z_i$$

has not also a mean value with a finite derivative at $\theta=0$. For these cases, we may apply the theory of Hodges-Lehmann [4] that is a slight generalization of Pitman's efficiency.

Consider the sequences of test statistics $\{S_N\}$, $\{T_N\}$ to test the hypothesis $\theta=\theta_0$ whose expectations

$$\mu_N(\theta)=E_\theta(S_N),\quad \nu_N(\theta)=E_\theta(T_N)$$

and suppose that

$$[S_N-\mu_N(\theta)]/b_N \text{ and } [T_N-\nu_N(\theta)]/c_N$$

tend in law to $N(0, 1)$ whenever $\theta\to\theta_0$. Let the sample sizes necessary to achieve the same power β against the same alternative at the same significance level α be respectively r_N and N. Then it has been proved in [4] that

(10)
$$\lim_{N\to\infty}\frac{\nu_N(\theta)-\nu_N(\theta_0)}{\mu_N(\theta)-\mu_N(\theta_0)}\cdot\frac{b_{rN}}{c_N}=1$$

Now let

$$m S_N = m\tilde{U}[h_0] = \sum_{i=1}^{N} \left(\log \frac{i}{N+1} + 1\right) Z_i$$

$$m\, T_N = \Phi^{-1}\left(\frac{i}{N+1}\right) Z_i.$$

We may get from (10) the asymptotic efficiency $e_{L,N}$ of $\tilde{U}[h_0]$ test relative to the Normal scores test

$$(11)\quad e_{L,N} = \lim_{N\to\infty} \frac{N}{r_N} = \left(\lim_{N\to\infty} \frac{\int [\Phi^{-1}\{\lambda_N F(x) + (1-\lambda_N)F(x+\theta)\} - \Phi^{-1}\{F(x)\}]dF(x)}{\int [\log\{\lambda_N F(x) + (1-\lambda_N)F(x+\theta)\} - \log F(x)]\, dF(x)}\right)^2.$$

After some easy calculations of integration and L' Hospital rule, we obtain

$$(12)\quad e_{L,N} = \left(\lim_{N\to\infty} \left[\frac{1}{\theta} + \frac{\lambda_N}{1-\lambda_N}\right] \Big/ \left[\frac{1-\lambda_N}{\varphi[\Phi^{-1}\{(1-\lambda_N)\theta\}]} + \frac{\lambda_N}{\varphi[\Phi^{-1}(1-\lambda_N\theta)]}\right]\right)^2 = \infty$$

Thus it follows that for the uniform alternative (and also exponential case, see (c).) our test with $h_0(t)$ is very much more efficient than the Normal scores test. The similar considerations lead us that the Normal scores test has infinite efficiency with regard to U_k tests with non-zero fixed k (Hodges-Lehmann have dealt with only $k=1$).

(c) Exponential case. Since we get the same results as the uniform case, we omit them.

§ **3. Ansari-Bradley's type.** In this section, we concern with the scale problems and generalize Ansari-Bradley's statistic. We first define the statistics $\tilde{S}[h]$

$$(13)\qquad m\,\tilde{S}[h] = \sum_{i=1}^{N} J_{N,i} Z_i$$

,where $J_{N,i} = \left[h\left(\frac{i}{N+1}\right) + h\left(\frac{N+1-i}{N+1}\right) - \left| h\left(\frac{i}{N+1}\right) - h\left(\frac{N+1-i}{N+1}\right)\right|\right]/2.$

We assume that $h(t)$ and $J_{N,i}$ respectively satisfy the assumption (17) and the regularity conditions of Chernoff-Savage. Then its asymptotic normality is established and the efficacy is expressed if the expectation of $\tilde{S}[h]$ has a finite derivative under the hypothesis

$$(14)\quad E^2(\tilde{S}) = \left[\int_{-\infty}^{0} xf(x)h'[F(x)]\, dF(x) - \int_{0}^{\infty} xf(x)h'[1-F(x)]dF(x)\right]^2.$$

When $F(x)$ is symmetrical, the expression $I[h]$ in bracket of (14) becomes

$$(15)\qquad I[h] = 2\int_{0}^{1/2} h'(t)\, F^{-1}(t)\, f[F^{-1}(t)]dt.$$

(a) Uniform case. Easy computation shows that

$$I[h]=-2\lim_{t\to 0} h(t).$$

Though both $h(t)=\Phi^{-1}(t)$ and $h_1(t)=\log 2t+1$ give very much large weight at extreme ranks, we prefer $h_1(t)$ to the other $h(t)=\Phi^{-1}(t)$ as the similar results as in section 2. However it may be possible to exist more favourable $h(t)$ than $h_1(t)$.

(b) Double exponential case. Since the density is

$$f(x)=\frac{1}{2}\exp(-|x|),$$

we may get

$$I[h]=2\int_0^{1/2} t\, h'(t) \log 2t\ dt \tag{16}$$

which must be maximized subject to the normalized restrictions

$$\int_0^{1/2} h(x)dx=0, \quad \int_0^{1/2} h^2(x)\ dx=1/2. \tag{17}$$

Following the method of variations, we may determine the form (18) as the solution of Eular equation,

$$h(t)=\log\ 2t+1. \tag{18}$$

This form may be also attained from our S_k tests in [6] by the similar calculation as (8). Capon [1] has also derived the same form by another point of view and proved to be locally best for the double exponential case.

(c) Normal case. $I[h]$ may be expressed as follows,

$$I[h]=2\int_0^{1/2} h'(t)\Phi^{-1}(t)\varphi[\Phi^{-1}(t)]\,dt. \tag{19}$$

Eular equation under the conditions for $h(t)$ is given as

$$h(t)-\frac{d}{dt}\Phi^{-1}(t)\varphi[\Phi^{-1}(t)]=0 \tag{20}$$

and it is solved as

$$h(t)=\frac{1}{\sqrt{2}}[\Phi^{-1}(t)^2-1]. \tag{21}$$

The corcesponding statistic $\tilde{S}[h]$ has thus an optimum property and it has been shown to be locally best rank test for the normal case by Klotz [5].

References

[1] CAPON, J.: *Asymptotic efficiency of certain locally most powerful rank tests.* Ann. Math. Statist. 32 (1961), 88–100.

[2] CHERNOFF, H. and SAVAGE, I. R.: *Asymptotic normality and efficiency of certain nonparametric test statistics.* Ann. Math, Statist. 29 (1958), 972–994.

[3] DWASS, M.: *The large-sample power of rank order tests in the two sample problem.* Ann. Math. Statist. 27 (1956), 352–374.

[4] HODGES, Jr. J. L. and LEHMANN, E. L.: *Comparision of the Normal scores and Wilcoxen tests.* Proc. Fourth Berkeley Sympo. Math. Stat. and Proba. I, (1961), 307–317.

[5] KLOTZ, J.: *Non-darametric tests for scale.* Ann. Math. Statist. 33 (1962), 498–512.

[6] TAMURA, R.: *On a modification of ccrtain rank tests.* Ann. Math. Statist. 34 (1963), 1101–1103.

SHIMANE UNIVERSITY.