NONPARAMETRIC TESTS FOR SCALE

Tamura, Ryoji
Shimane University

https://doi.org/10.5109/13020

出版情報：統計数理研究. 12 (1/2), pp.89-94, 1966-03. Research Association of Statistical Sciences

NONPARAMETRIC TESTS FOR SCALE

By

Ryoji Tamura

(*Received* February 1, 1965)

We shall in this paper propose certain nonparametric test statistics for the scale problem which are essentially the same kind as Tamura [6]. Some of them have higher efficiency than Mood [2], Tamura [6], [7] and Sen [4] and one of them is very easy to apply practically.

Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be the samples from the continuous c.d.f. $F(x-\xi)$ and $G(y)=F((y-\eta)/\theta)$ where ξ and η are respectively the population medians. We shall consider to test the hypothesis H: $\theta=1$ against the alternative H': $\theta>1$. Assume first that ξ and η are known, say zero without loss of generality. The auther has proposed the following in [6],

$$(1)\qquad Q_N=\binom{m}{2}^{-1}\binom{n}{2}^{-1}\sum\sum\phi(X_{\alpha_1},\ X_{\alpha_2};\ Y_{\beta_1},\ Y_{\beta_2}),$$

where

$$\phi(x,\ x';\ y,\ y')=\begin{cases}1 & \text{for } y<x,\ x'<y' \text{ or } y'<x,\ x'<y\\ 0 & \text{otherwise}\end{cases}$$

and the summations run over all subscripts α, β such that $1\leqq\alpha_1<\alpha_2\leqq m$, $1\leqq\beta_1<\beta_2\leqq n$.

Now we propose the following as a generalization of Q_N (1),

$$(2)\qquad Q_{N,s}=\binom{m}{s}^{-1}\binom{n}{2}^{-1}\sum\sum\phi(X_{\alpha_1},\ \cdots,\ X_{\alpha_s};\ Y_{\beta_1},\ Y_{\beta_2}),$$

where

$$\phi(x_1,\ \cdots,\ x_s;\ y',\ y)=\begin{cases}1 & \text{for } y<x_1,\ \cdots,\ x_s<y' \text{ or } y'<x_1,\ \cdots,\ x_s<y\\ 0 & \text{otherwise}\end{cases}$$

and the summations are over all subscripts α, β such that $1\leqq\alpha_1<\cdots<\alpha_s\leqq m$, $1\leqq\beta_1<\beta_2\leqq n$.

Denoting the mean value of $Q_{N,s}$ by $\mu_s(\theta)$, then we get

$$(3)\qquad \mu_s(\theta)=2\iint_{x'<x}\{F(x)-F(x')\}^s dG(x)dG(x').$$

For later necessity, we compute the derivative of $\mu_s(\theta)$ at $\theta=1$,

$$(4)\qquad (\partial\mu_s(\theta)/\partial\theta)_{\theta=1}=2\int_{-\infty}^{\infty}[F^s(x)-\{1-F(x)\}^s]xf(x)dF(x).$$

As shown by Dwass [1], $\sqrt{N}(Q_{N,s}-\mu_s)$ is equivalent to the statistic

$$V_N=\frac{s}{\sqrt{\lambda m}}\sum_{i=1}^{m}\{\psi_{10}(X_i)-\mu_s\}+\frac{2}{\sqrt{(1-\lambda)n}}\sum_{j=1}^{n}\{\psi_{01}(Y_j)-\mu_s\},$$

which is asymptotically normally distributed, where

$$\psi_{10}(x_1)=E\phi\,(x_1,\ X_2,\ \cdots,\ X_s;\ Y,\ Y'),\ \lambda=m/N,\ N=m+n$$
$$\psi_{01}(y)=E\phi(X_1,\ \cdots,\ X_s;\ y,\ Y).$$

The asymptotic variance $\sigma_s^2(\theta)$ of $Q_{N,s}$ may be expressed by

(5) $$N\sigma_s^2(\theta)=\lambda^{-1}s^2\sum\mathrm{Var}\,\psi_{10}(X_i)+4(1-\lambda)^{-1}\sum\mathrm{Var}\,\psi_{01}(Y_j).$$

In order to compute the value $\sigma_{s,0}^2$ of $\sigma_s^2(\theta)$ under the hypothesis H, we first get the following

$$\begin{aligned}\psi_{10}(x)&=2\Pr(Y<x,\ X_2,\ \cdots,\ X_s<Y'\,|\,H)\\&=2\iint_{y<x<y'}\{F(y')-F(y)\}^{s-1}dF(y)dF(y')\\&=2s^{-1}(1+s)^{-1}[1-F^{s+1}(x)-\{1-F(x)\}^{s+1}].\end{aligned}$$

Thus we get

$$\mathrm{Var}\ \psi_{10}(X)=8s^{-2}(s+1)^{-2}\Big[\frac{1}{2s+3}-\frac{2}{(s+2)^2}+\frac{(s+1)!^2}{(2s+3)!}\Big].$$

Similarly we get

$$\mathrm{Var}\ \psi_{01}(Y)=2(s+1)^{-2}\Big[\frac{1}{2s+3}-\frac{2}{(s+2)^2}+\frac{(s+1)!^2}{(2s+3)!}\Big].$$

Therefore it follows that

(6) $$\sigma_{s,0}^2=\Big(\frac{1}{m}+\frac{1}{n}\Big)8(s+1)^{-2}\Big[\frac{1}{2s+3}-\frac{2}{(s+2)^2}+\frac{(s+1)!^2}{(2s+3)!}\Big].$$ (1)

By using (4) and (6), we may get the asymptotic efficiency e_s of the $Q_{N,s}$ test with regard to the $Q_{N,1}$ test as follows,

(7) $$e_s=\frac{(s+1)^2}{360\Big[\frac{1}{2s+3}-\frac{2}{(s+2)^2}+\frac{(s+1)!^2}{(2s+3)!}\Big]}\left[\frac{\int_{-\infty}^{\infty}\{F^s-(1-F)^s\}xf\,dF}{\int_{-\infty}^{\infty}(2F-1)xf\,dF}\right]^2.$$

For $s=2$, which is corresponding to the Q_N test (1), we get $e_s=1$. *Namely it has been proved that the $Q_{N,1}$ test is asymptotically equivalent to the Q_N test.* Assume that $F(x)$ be normal, then we get after some computations

(8) $$e_s=\frac{2s+3}{10\Big[s^2-2+\frac{(s+2)!^2}{(2s+2)!}\Big]}\left[\frac{EX_{1\,s+2}^2-1}{EX_{1|3}^2-1}\right]^2$$

, where X_{1s} expresses the smallest variable in the ordered sample of size s from $N(0,\ 1)$. The values may be computed by the Ruben's table [3].

(1) This result has been suggested by Dr. N. Sugiura, Osaka University.

Table

s	2	3	4	5	6	8	10	12	14	15	16
e_s	1	1.03	1.09	1.11	1.14	1.18	1.21	1.221	1.225	1.225	1.220

In the case where medians are unknown, we again propose the similar statistic $\tilde{Q}_{N,s}$ as $\tilde{Q}_N$ in [6]

(9) $$\tilde{Q}_{N,s}=\binom{m}{s}^{-1}\binom{n}{2}^{-1}\sum\phi(X_{\alpha_1}-\tilde{X},\cdots,\ X_{\alpha_s}-\tilde{X}\,;\ Y_{\beta_1}-\tilde{Y},\ Y_{\beta_2}-\tilde{Y})$$

, where $\tilde{X}$ and $\tilde{Y}$are sample medians.
From the Sukhatme's theorem [5], we define the following

(10) $$\begin{aligned} &A_s(t_1,\ t_2)=E\phi(X_1-t_1,\ \cdots,\ X_s-t_1\,;\ Y-t_2,\ Y'-t_2)\\ &W_s(t_1,\ t_2)=\phi(X_1-t_1,\ \cdots,\ X_s-t_1\,;\ Y-t_2,\ Y'-t_2)-A_s(t_1,\ t_2). \end{aligned}$$

First we prove that the partial derivatives of A_s respective to t_i vanish in $t_1=t_2=0$ under symmetric $F(x)$. Now

$$\begin{aligned}(\partial A_s/\partial t_i)_{t_1=t_2=0}&=(-1)^{i-1}2s\iint_{x<x'}\{F(x')-F(x)\}^{s-1}\{f(x')-f(x)\}dG(x)dG(x')\\ &=(-1)^{i-1}2s\Big[\iint_{x<x'}\{F(x')-F(x)\}^{s-1}f(x')dG(x')dG(x)\\ &\qquad-\iint_{x<x'}\{F(x')-F(x)\}^{s-1}f(x)dG(x')dG(x)\Big].\end{aligned}$$

In the second integration, we peform the translation $x=-y'$, $x'=-y$ and apply $F(x)=1-F(-x)$, then the second term is transformed as

$$\iint_{y<y'}\{F(y')-F(y)\}^{s-1}f(y')dG(y)dG(y).$$

Therefore it holds

(11) $$(\partial A_s/\partial t_i)_{t_1=t_2=0}=0.$$

It is remained to show

(12) $$E|W_s(t_1,\ 0)-W_s(0,\ 0)|\leqq M_1t_1,\ \ E|W_s(0,\ t_2)-W_s(0,\ 0)|\leqq M_2t_2$$

, where M_i are constants. For this,

$$\begin{aligned}E|W_s(t_1,0)-W_s(0,\ 0)|\leqq E|&\phi(X_1-t_1,\ \cdots,\ X_s-t_1\,;\ Y,\ Y')\\ &-\phi(X_1,\ \cdots,\ X_s\,;\ Y',\ Y)|+|A_s(t_1,\ 0)-A_s(0,\ 0)|.\end{aligned}$$

Then

$$\begin{aligned}|A_s(t_1,\ 0)-A_s(0,\ 0)|\leqq 2s\iint_{-\infty}^{\infty}|&F(x'+t_1)-F(x')\\ &-F(x+t_1)+F(x)|dG(x)dG(x')\end{aligned}$$

$$\leq 4s\int_{-\infty}^{\infty} F(x+t_1)-F(x)\ dG(x)$$
$$\leq 4sat_1$$

if $f(x)$ is bounded in absolute value by a.

From the definition of ϕ, we may also compute as follows,

$$\begin{aligned} &E\ \phi(X_1-t_1,\ \cdots,\ X_s-t_1;\ Y,\ Y')-\phi(X_1,\ \cdots,\ X_s;\ Y,\ Y') \\ &\quad =\Pr\{\phi(X_1-t_1,\ \cdots,\ X_s-t_1;\ Y,\ Y')=1 \text{ and } \phi(X_1,\ \cdots,\ X_s;\ Y,\ Y')=0\} \\ &\quad +\Pr\{\phi(X_1-t_1,\ \cdots,\ X_s-t_1;\ Y,\ Y')=0 \text{ and } \phi(X_1,\ \cdots,\ X_s;\ Y,\ Y')=1\}. \end{aligned}$$

The first probability is given by summations of the probabilities of the events of the following four types

(i) $Y<X_1-t_1,\ \cdots,\ Y'-t_1,\ \cdots,\ X_s-t_1<Y'$

(ii) $Y'-t_1<Y<X_1-t_1,\ \cdots,\ X_s-t_1<Y'$

(iii) $Y'<X_1-t_1,\ \cdots,\ Y-t_1,\ \cdots,\ X_s-t_1<Y$

(iv) $Y-t_1<Y'<X_1-t_1,\ \cdots,\ X_s-t_1<Y.$

The probability of the event (i) is transformed to the following

$$\sum_{r=0}^{s}\binom{s}{r}\iint_{x<x'-t_1}\{F(x')-F(x+t_1)\}^r\{F(x'+t_1)-F(x')\}^{s-r}dG(x)dG(x')$$
$$\leq \text{const.}\int_{-\infty}^{\infty}\{F(x+t_1)-F(x)\}dG(x)\leq \text{const. } at_1.$$

It is similarly shown that the other probabilities are smaller than const. at_1. Thus the first half of (12) has been proved. The techniques are similar for the latter half. We may conclude that $\tilde{Q}_{N,s}$ has the same asymptotic normal distribution as $Q_{N,s}$ by adding to (11) and (12) the property that $\sqrt{m}(\tilde{X}-\xi)$ and $\sqrt{n}(\tilde{Y}-\eta)$ are both bounded in probability.

Since it has been proved that the $Q_{N,1}$ and $Q_{N,2}$ tests are asymptotically equivalent, that is, they have the same asymptotic efficiency, their behaviours must be discussed in small sample. We cannot unfortunately deal with them in the general form, therefore we only check in the simple and special cases. When $m=n=4$, the orderings of X's and Y's which have larger values of $Q_{N,2}$ or $Q_{N,1}$ are respectively given as follows,

$Q_{N,2}$		$Q_{N,1}$	
ordering	value of $\binom{4}{2}\binom{4}{2}Q_{N,2}$	ordering	value of $4\binom{4}{2}Q_{N,1}$
$YYXXXXYY$	24	$YYXXXXYY$	16
$YXYXXXYY$	18	$YXYXXXYY$	15
$YYXXXYXY$	18	$YYXXXYXY$	15
$YXXXXYYY$	18	$YXXYXXYY$	14
$YYYXXXXY$	18	$YYXXYXXY$	14
⋮		$YXYXXYXY$	14
		⋮	

Let the size α of test be 1/70, then the critical regions contain both only an ordering $YYXXXXYY$ and therefore the tests have the equal powers. In the case $\alpha=5/70$, the critical regions are respectively constructed by the above five orderings (in the $Q_{N,1}$ test, six in the randomized form). Since they have three in common, we may compare only the probabilities of the remainders. When $F(x)$ is symmetrical, it holds that

$$\Pr(YYYXXXXY)=\Pr(YXXXXYYY),$$
$$\Pr(YXXYXXYY)=\Pr(YYXXYXXY).$$

Now we assume that $F(x)$ be the uniform distribution in $(-1/2, 1/2)$, then

$$\Pr(YXXXXYYY)=\frac{4!}{2!}\int_{y1<y2}\{1-G(y_2)\}^2\{F(y_2)-F(y_1)\}^4 dG(y_1)dG(y_2)$$
$$=\frac{12}{\theta^2}\left[\int_{1/2}^{\theta/2}\left(\frac{1}{2}-\frac{y}{\theta}\right)^2 dy_2\left\{\int_{-1/2}^{1/2}\left(\frac{1}{2}-y_1\right)^4 dy_1+\int_{-\theta/2}^{-1/2}dy_1\right\}\right.$$
$$\left.+\int_{-1/2}^{1/2}\left(\frac{1}{2}-\frac{y}{\theta}\right)^2 dy_2\left\{\int_{-1/2}^{y_2}(y_2-y_1)^4 dy_1+\int_{-\theta/2}^{-1/2}\left(y_2+\frac{1}{2}\right)^4 dy_1\right\}\right].$$

After some computations, we get

$$\Pr(YXXXXYYY)=\frac{1}{70\theta^4}\left[\frac{35}{2}(\theta-1)^4+28(\theta-1)^3\right.$$
$$\left.+21(\theta-1)^2+8(\theta-1)+1\right]. \tag{13}$$

From the similar computations,

$$\Pr(YXXYXXYY)=\frac{1}{70\theta^4}[21(\theta-1)^3+21(\theta-1)^2+8(\theta-1)+1],$$
$$\Pr(YXYXXYXY)=\frac{1}{70\theta^4}[14(\theta-1)^2+8(\theta-1)+1]. \tag{14}$$

Under the alternative $\theta>1$, (13) is larger than (14). Thus the power of the $Q_{N,2}$ test is larger than that of the $Q_{N,1}$ test in this case. Consider the case $\alpha=17/252$ when $m=n=5$ and denote (power of $Q_{N,2}$ test) $-$ (power of $Q_{N,1}$ test) by d. Then we may get the following by the similar considerations,

$$d=\frac{5(\theta-1)^2}{336\theta^5}[21(\theta-1)^3+63(\theta-1)^2+40(\theta-1)+6]. \tag{15}$$

In this case it may also result the superiority of the $Q_{N,2}$ test.

I should like to express my gratitude to Dr. N, Sugiura, Osaka University, for much help and criticism.

References

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

[2] MOOD, A. M. (1954). *On the asymptotic efficiency of certain nonparametric two sample tests.* Ann. Math. Statist. **26.** 514-522.

[3] RUBEN, H. (1954). *On the moments of order statistics in sample from normal populations.* Biometrika **41.** 200-227.

[4] SEN, P. K. (1963). *On weighted rank-sum tests for dispersion.* A. I. S. M. **15.** 117-135.

[5] SUKHATME, B. V. (1958). *Testing thc hypothesis that two populations differ only in location.* Ann. Math. Statist. **29.** 60-78.

[6] TAMURA, R. (1960). *On the nonparametric tests based on certain U-statistics.* Bull. Math. Statist. **9.** 61-67.

[7] TAMURA, R. (1963). *On a modification of certain rank tests.* Ann. Math. Statist. **34.** 1101-1103.

Shimane University