Sugiura, N.
Osaka Math. J.
15 (1963), 257-268.

ON A GENERALIZATION OF THE WILCOXON TEST FOR CENSORED DATA

By

NARIAKI SUGIURA

1. Introduction

Let two populations Π_i $(i=1, 2)$ be such that

$$\Pi_i:\ \mathrm{P}\{X_i \leq x\} = \begin{cases} 0 & x<0, \\ p_{i-1}+\int_0^x f_{i-1}(t)dt & x\geqq 0. \end{cases} \tag{1.1}$$

When we wish to test the hypothesis $H: \Pi_1=\Pi_2$ by two random samples $(X_1, X_2, \cdots, X_{n_1})$ and $(Y_1, Y_2, \cdots, Y_{n_2})$ taken from Π_1 and Π_2 respectively, ties occuring at the origin prevent us from using the Wilcoxon statistic. As Kruskal and Wallis [4] and Putter [7] considered, however, the concept of midrank is available in this case and we define the test statistic U_m as follows:

$$U_m = \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} m(X_i, Y_j), \tag{1.2}$$

where

$$m(X, Y) = \begin{cases} 1 & X>Y, \\ \frac{1}{2} & X=Y, \\ 0 & X<Y. \end{cases}$$

If we define V_m by interchanging X and Y in (1.2), we can easily see that $U_m+V_m=1$. So we consider only U_m in the following.

The mean and variance of U_m are calculated in section 2 and the consistency as well as unbiasedness of the test based on U_m are shown in section 3. The asymptotic relative efficiency is calculated with respect to the location alternative in section 4 and the asymptotic efficiency relative to Halperin's U_c conditional test [2] in section 5. Finally in section 6 we apply these tests to some data of cleft-palate patients kindly provided by Mr. A. Takayori, Dental School, Osaka University.

2. Mean and variance of U_m

Proposition 1. *Mean and variance of the statistic U_m defined in* (1.2) *are such that*

$$\text{(2.1)}\qquad \mathrm{E}(U_m) = \frac{1}{2}p_0p_1+q_0p_1+\int_0^\infty F_1(t)dF_0(t) \qquad (=p,\ say),$$

$$\text{(2.2)}\qquad \mathrm{Var}(U_m) = \frac{1}{n_1n_2}\left\{pq-\frac{1}{4}p_0p_1+(n_1-1)s_1+(n_2-1)s_2\right\},$$

where

$$q = 1-p, \qquad s_1 = \int_0^\infty (F_0(t)-q+p_0)^2dF_1(t)+p_1\left(q-\frac{1}{2}p_0\right)^2,$$

$$q_0 = 1-p_0,$$

$$F_i(x) = \int_0^x f_i(t)dt \qquad s_2 = \int_0^\infty (F_1(t)-p+p_1)^2dF_0(t)+p_0\left(p-\frac{1}{2}p_1\right)^2.$$

$(i=0,\ 1)$,

Proof. By the definition of U_m in (1.2), we have

$$\begin{aligned}\mathrm{E}(U_m) &= \mathrm{E}[m(X,Y)]\\ &= \mathrm{P}\{X>Y\geqq 0\}+\frac{1}{2}\mathrm{P}\{X=Y=0\}\\ &= \iint_{t_1>t_2>0} f_0(t_1)f_1(t_2)dt_1dt_2+q_0p_1+\frac{1}{2}p_0p_1\end{aligned}$$

to get (2.1). Since U_m is a kind of U statistic due to Hoeffding [3] and Lehmann [5], Problem 8 in Fraser [1, p. 257] is available to calculate its variance, that is,

$$\mathrm{Var}(U_m) = \frac{1}{n_1n_2}\{\zeta_{1,1}+(n_1-1)\zeta_{0,1}+(n_2-1)\zeta_{1,0}\},$$

where

$$\zeta_{1,1} = \mathrm{Var}[m(X,Y)] = pq-\frac{1}{4}p_0p_1,$$

$$\zeta_{0,1} = \mathrm{Var}[f^*_{0,1}(Y)], \qquad f^*_{0,1}(y) = \mathrm{E}[m(X,Y)\mid Y=y],$$

$$\zeta_{1,0} = \mathrm{Var}[f^*_{1,0}(X)], \qquad f^*_{1,0}(x) = \mathrm{E}[m(X,Y)\mid X=x].$$

In our case

$$f^*_{0,1}(y) = \begin{cases} 1 & y<0,\\ q_0+\frac{1}{2}p_0 & y=0,\\ q_0-F_0(y) & y>0,\end{cases}$$

and hence

$$\zeta_{0,1} = \left(q_0+\frac{1}{2}p_0\right)^2 p_1 + \int_0^\infty (q_0-F_0(t))^2 dF_1(t) - p^2 ,$$

which, after some calculations, turns out to be equal to s_1 in (2.2). In the same way, we get s_2 in (2.2) from $\zeta_{1,0}$.

Corollary. *Under the null hypothesis H*

$$\mathrm{E}(U_m) = \frac{1}{2}, \tag{2.3}$$

$$\mathrm{Var}\,(U_m) = \frac{1}{12n_1n_2}\{3(1-p_0^2)+(n_1+n_2-2)(1-p_0^3)\} . \tag{2.4}$$

Proof. Since under the null hypothesis $f_0(t)=f_1(t)$ and $p_0=p_1$, we have from (2.1)

$$\mathrm{E}\,(U_m) = \frac{1}{2}p_0^2+p_0q_0+\frac{1}{2}[F_0(t)^2]_0^\infty = \frac{1}{2},$$

and from (2.2)

$$\mathrm{Var}\,(U_m) = \frac{1}{n_1n_2}\left\{\frac{1}{4}(1-p_0^2)+(n_1+n_2-2)s\right\},$$

where

$$\begin{aligned} s &= \int_0^\infty \left(F_0(t)-\frac{1}{4}+p_0\right)^2 dF_0(t)+\frac{1}{4}p_0q_0^2 \\ &= \frac{1}{12}(1-p_0^3) . \end{aligned}$$

This proves (2.4).

3. Consistency and unbiasedness of the U_m test

In this section we consider the following alternative,

$$K:\ p_0+F_0(x) < p_1+F_1(x) \quad \text{for any} \quad x \geqq 0 , \tag{3.1}$$

that is to say, Π_1 is stochastically larger than Π_2. Let the test function determined by U_m, *which will be called the U_m test*, be

$$\phi(X_1, \cdots, X_{n_1};\ Y_1, \cdots, Y_{n_2}) = \begin{cases} 1 & U_m > U_\alpha , \\ 0 & U_m < U_\alpha , \end{cases} \tag{3.2}$$

where the constant U_α is determined such that $\mathrm{E}\phi=\alpha$ under the null hypothesis H.

Theorem 1. *The U_m test of the hypothesis H: $\Pi_1=\Pi_2$ against the alternative K is unbiased and consistent under the limiting condition*:

(3.3) $$n_1+n_2=N\,, \quad n_1=\alpha_1 N\,, \quad n_2=\alpha_2 N\,, \quad \text{aud} \quad N\to\infty\,,$$
$$(\textit{with } \alpha_1 \textit{ and } \alpha_2 \textit{ fixed}).$$

Proof. By the lemma 3.1 in Lehmann [5], it is sufficient, to prove consistency, to show that, under the alternative K, $\mathrm{E}(U_m)>1/2$ and $\mathrm{Var}(U_m)\to 0$ as $N\to\infty$. The former is derived from (2.1) and (3.1), while the latter from (2.2).

Unbiasedness is proved from the following lemma which assures the validity of Theorem 3.1 in Lehmann [5], even when the population distribution is discontinuous at the origin as is the case with (1.1).

Lemma. *If the test function satisfies* $\phi(x_1, \cdots, x_{n_1};\ y_1, \cdots, y_{n_2}) \geqq \phi(x_1, \cdots, x_{n_1};\ z_1, \cdots, z_{n_2})$ *whenever* $y_i \leqq z_i$ $(i=1, 2, \cdots, n_2)$, *then the power function against the alternative K in* (3.1) *satisfies* $\mathrm{E}_{G_0,G_1}(\phi) \geqq \mathrm{E}_{G_0,G_0}(\phi)$ *for G_0 and G_1 representing the distribution function of Π_1 and Π_2 respectively in* (1.1).

Proof.* Let

(3.4) $$g(x)=\begin{cases} G_1^{-1}(G_0(x)) & G_0^{-1}(p_1)\leqq x\,, \\ 0 & G_0^{-1}(p_1)>x\,, \end{cases}$$

then the distribution function of $g(x)$ under Π_1 is $G_1(z)$. From (3.4) and (3.1), $g(x)\leqq x$ for all $x\geqq 0$. Hence

$$\begin{aligned}\mathrm{E}_{G_0,G_1}[\phi(X_1, \cdots, X_{n_1};\ Y_1, \cdots, Y_{n_1})] &= \mathrm{E}_{G_0,G_0}[\phi(X_1, \cdots, X_{n_1};\ g(Y_1), \cdots, g(Y_{n_2}))] \\ &\geqq \mathrm{E}_{G_0,G_0}[\phi(X_1, \cdots, X_{n_1};\ Y_1, \cdots, Y_{n_2})]\,.\end{aligned}$$

4. Efficiency of the U_m test for the location alternative

When the hypothesis $H: \Pi_1=\Pi_2$ and the alternative K in (3.1) differ only in location such that

(4.1) $$\Pi_1: \begin{cases} p_0=\int_{-\infty}^0 f(t-\theta_0)dt\,, \\ F_0(x)=\int_0^x f(t-\theta_0)dt\,, \end{cases} \qquad \Pi_2: \begin{cases} p_1=\int_{-\infty}^0 f(t-\theta)dt\,, \\ F_1(x)=\int_0^x f(t-\theta)dt\,, \end{cases}$$

then we are concerned with testing $H: \theta=\theta_0$ against $K: \theta<\theta_0$. Suppose there exist the maximum likelihood estimators of θ_0 and θ denoted by $\hat\theta_0=\hat\theta_0\ (X_1, \cdots, X_{n_1})$ and $\hat\theta=\hat\theta\ (Y_1, \cdots, Y_{n_2})$ and let the test function ψ be

(4.2) $$\psi(X_1, \cdots, X_{n_1};\ Y_1, \cdots, Y_{n_2})=\begin{cases} 1 & \hat\theta-\hat\theta_0<c_\alpha\,, \\ 0 & \hat\theta-\hat\theta_0>c_\alpha\,, \end{cases}$$

* This lemma is also proved from the lemma 1 of chapter 3 and the lemma 2 of chapter 5 in Lehmann, "Testing Statistical Hypothesis", John Wiley & Sons, Inc. 1959.

where the constant c_α is determined such that $E\psi=\alpha$ under H.

Theorem 2. *If the maximum likelihood estimators of θ_0 and θ denoted by $\hat{\theta}_0$ and $\hat{\theta}$ exist and are distributed asymptotically normal and efficient, the asymptotic efficiency of the test U_m defined by* (3.2) *relative to the test ψ defined by* (4.2) *is*

$$(4.3)\qquad e_{\phi,\psi}=\frac{\left\{p_0 f(-\theta_0)+2\int_{-\theta_0}^{\infty} f(t)^2dt\right\}^2}{\left(p_0q_0+\frac{1}{3}q_0^3\right)\left\{-f'(-\theta_0)+\frac{1}{p_0}f(-\theta_0)^2-\mathrm{E}\left(\frac{\partial^2\log f(X-\theta_0)}{\partial\theta_0^2}\right)\right\}}\,.$$

Proof. Put $\theta=\theta_0-kN^{-1/2}$ and consider the limiting condition (3.3). As U_m is distributed asymptotically normal under K by Lehmann [5], the asymptotic power of the test ϕ is $\Phi[(\mathrm{E}_\theta(U_m)-U_\alpha)/\mathrm{Var}_\theta(U_m)^{1/2}]$, where Φ is the distribution function of standardized normal distribution. From Proposition 1 we have

$$\left.\frac{\partial \mathrm{E}_\theta(U_m)}{\partial\theta}\right|_{\theta=\theta_0}=-\frac{1}{2}p_0f(-\theta_0)-\int_{-\theta_0}^{\infty}f(t)^2dt\,,$$

and

$$\mathrm{Var}_\theta(U_m)=\frac{1-p_0^3}{12\alpha_1\alpha_2N}+0(N^{-2})\,.$$

Hence the asymptotic power of ϕ is

$$\Phi\left(\frac{\mathrm{E}_\theta(U_m)-U_\alpha}{\mathrm{Var}_\theta(U_m)^{1/2}}\right)=\Phi\left(a+\left.\frac{\partial\mathrm{E}_\theta(U_m)}{\partial\theta}\right|_{\theta=\theta_0}\mathrm{Var}_{\theta_0}(U_m)^{-1/2}(\theta-\theta_0)+0(N^{-1/2})\right)$$
$$=\Phi(a+kc+0(N^{-1/2}))\,,$$

where

$$(4.4)\qquad c=\sqrt{\alpha_1\alpha_2}\,\frac{p_0f(-\theta_0)+2\int_{-\theta_0}^{\infty}f(t)^2dt}{(p_0q_0+q_0^3/3)^{1/2}}\quad\text{and}\quad\Phi(a)=\alpha\,.$$

Since by assumption $\hat{\theta}-\hat{\theta}_0$ is distributed asymptotically normal with mean $\theta-\theta_0$ and variance

$$-n_1^{-1}\left\{p_0\frac{d^2\log p_0}{d\theta_0^2}+\mathrm{E}\left(\frac{\partial^2\log f(X-\theta_0)}{\partial\theta_0^2}\right)\right\}^{-1}$$
$$-n_2^{-1}\left\{p_1\frac{d^2\log p_1}{d\theta^2}+\mathrm{E}\left(\frac{\partial^2\log f(X-\theta)}{\partial\theta^2}\right)\right\}^{-1},$$

the asymptotic power of the test ψ at $\theta=\theta_0-k^*N^{-1/2}$ is

$$(4.5)\qquad \Phi(a+k^*c^*+0(N^{-1/2}))\,,$$

where

$$(4.6)\qquad c^* = \sqrt{\alpha_1\alpha_2}\left\{-f'(-\theta_0)+\frac{1}{p_0}f(-\theta_0)^2-\mathrm{E}\left(\frac{\partial^2 \log f(X-\theta_0)}{\partial\theta_0^2}\right)\right\}^{1/2}.$$

We can get the asymptotic relative efficiency from (4.4), (4.5), and $e_{\theta,\psi}=(k^*/k)^2=(c/c^*)^2$.

EXAMPLE 1. Normal distribution. When $f(x)=(2\pi)^{-1/2}e^{-x^2/2}$ in Theorem 2, the efficiency becomes

$$(4.7)\qquad e_{\phi,\psi}=\frac{\{\Phi(-\theta_0)f(\theta_0)+\pi^{-1/2}\Phi(\sqrt{2}\,\theta_0)\}^2\Phi(-\theta_0)}{(\Phi(-\theta_0)+\Phi(\theta_0)^2/3)\{\Phi(-\theta_0)(\Phi(\theta_0)-\theta_0 f(\theta_0))+f(\theta_0)^2\}\Phi(\theta_0)}.$$

Some numerical values of $e_{\phi,\psi}$ are shown in the following Table 1.

Table 1. Efficiency for the normal distribution.

θ_0	$-\infty$	-1	0	1	∞
$e_{\phi,\psi}$	1	0.970	0.972	0.969	0.955 ($=3/\pi$)

As θ_0 tends to plus infinity, $e_{\phi,\psi}$ tends to the efficiency $3/\pi$ for the ordinary Wilcoxon test relative to the Student t-test (see Mood [6]). It is interesting that the efficiency is nearly equal to 1 irrespective of the value of θ_0.

EXAMPLE 2. Exponential distribution. When $f(x)$ is equal to e^{-x} for $x\geqq 0$ and zero otherwise, the condition concerning $\hat{\theta}_0$ and $\hat{\theta}$ stated in Theorem 2 is not satisfied. Calculating directly, we get $\hat{\theta}_0=\log(1-r_1/n_1)$ and $\hat{\theta}=\log(1-r_2/n_2)$. Using the asymptotic normality of r_1/n_1 and r_2/n_2, we can conclude that $\hat{\theta}_0$ is distributed asymptotically normal with mean $\log q_0$ and variance p_0/n_1q_0, and $\hat{\theta}$ with mean $\log q_1$ and variance p_1/n_2q_1. From this we can get the asymptotic power (4.5) of the test ψ in (4.2) with c^* in (4.6) as follows:

$$c^*=\sqrt{\alpha_1\alpha_2}\frac{e^{\theta_0/2}}{(1-e^{\theta_0})^{1/2}}.$$

This turns out to be equal to the right side of (4.6), and hence the efficiency may be calculated by (4.3), i.e.

$$(4.8)\qquad e_{\phi,\psi}=\frac{3(1-e^{\theta_0})}{3-3e^{\theta_0}+e^{2\theta_0}}.$$

From (4.8) we can see that the efficiency decreases monotonically from one to zero, as θ_0 changes from $-\infty$ to zero. Some numerical values are shown below.

Table 2. Efficiency for the exponential distribution.

θ_0	$-\infty$	-2	-1	-0.5	-0.2	-0.1	0
$e_{\phi,\psi}$	1	0.99	0.93	0.76	0.45	0.26	0

EXAMPLE 3. Uniform distribution in [0, 1]. In this case we take the test function ψ corresponding to (4.2) as follows:

$$\psi(X_1, \cdots, X_{n_1};\ Y_1, \cdots, Y_{n_2}) = \begin{cases} 1 & \frac{r_1}{n_1} - \frac{r_2}{n_2} < c_\alpha, \\ 0 & \frac{r_1}{n_1} - \frac{r_2}{n_2} > c_\alpha. \end{cases}$$

Then the asymptotic power of the test ψ is given by (4.5) with $c^* = \sqrt{\alpha_1\alpha_2}\{-\theta_0(1+\theta_0)\}^{-1/2}$. Hence

$$e_{\phi,\psi} = \frac{-3\theta_0(2+\theta_0)^2}{1-\theta_0+\theta_0^2}. \tag{4.9}$$

From (4.9) we can see that the curve of efficiency is unimodal with the maximum value $3(4\sqrt{2}-5)$ at $\theta_0 = 1-\sqrt{2}$.

Table 3. Efficiency for the uniform distribution.

θ_0	-1	-0.8	-0.6	-0.4	-0.2	-0.1	-0.05	0
$e_{\phi,\psi}$	1	1.42	1.80	1.97	1.57	0.98	0.54	0

5. Efficiency of the U_m test relative to Halperin's U_c conditional test

Halperin [2] proposed the following U_c conditional test: Put

$$U_c = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} c(X_i, Y_j), \tag{5.1}$$

where

$$c(X, Y) = \begin{cases} 1 & X > Y \geqq 0, \\ 0 & Y \geqq X \geqq 0, \end{cases}$$

and let r_1 and r_2 be the number of zeroes appearing in the X's and the Y's respectively, then Halperin [2] showed the conditional asymptotic normality of U_c under the null hypothesis H for given r $(=r_1+r_2)$ and considered the test (3.2) with U_m replaced by U_c, *which will be denoted by* ϕ'. The relation between two statistics U_m and U_c is given by

$$U_c = U_m - \frac{r_1 r_2}{2n_1 n_2}. \tag{5.2}$$

Theorem 3. *Suppose* p_0 *and* p_1 *in* (1.1) *are such that* $p_0=p(\theta_0)$ *and* $p_1=p(\theta)$ *with the function* $p(\theta)$ *differentiable in some neighbourhood of* $\theta=\theta_0$, *then* $(U_c-E)V^{-1/2}$ *is distributed asymptotically normal with mean zero and variance one, for given* r, *under the alternative* $K: \theta<\theta_0$ *with* $\theta=\theta_0+0(N^{-1/2})$ *and under the limiting condition*:

$$n_1=\alpha_1 N, \quad r=Np_0+0(N^{1/2}),$$
$$n_2=\alpha_2 N, \quad \text{and} \quad N\to\infty,$$
$$n_1+n_2=N,$$

where

$$(5.3) \qquad E=\frac{(n_1+n_2-r)}{(n_1+n_2)^2}\left\{(n_1+n_2-r)\int_0^\infty F_1^* dF_0^*+r\right\} + \left(\frac{dp}{d\theta}\right)_{\theta=\theta_0}\frac{\theta-\theta_0}{n_1+n_2}\left\{(n_2-n_1)q_0\int_0^\infty F_1^* dF_0^*+n_1q_0+n_2p_0\right\}+0(N^{-1}),$$

$$(5.4) \qquad V=\frac{p_0q_0}{n_1n_2}\left\{(n_2-n_1)q_0\int_0^\infty F_1^* dF_0^*+n_1q_0+n_2p_0\right\}^2 + \frac{q_0^3}{n_1n_2}\left\{n_1\int_0^\infty\left(F_0^*-\int_0^\infty F_0^* dF_1^*\right)^2 dF_1^*+n_2\int_0^\infty\left(F_1^*-\int_0^\infty F_1^* dF_0^*\right)^2 dF_0^*\right\} +0(N^{-3/2}),$$

and

$$F_i^*=\frac{F_i(t)}{q_i} \qquad (i=0,\ 1).$$

Proof*. The conditional distribution of r_1 for given r is

$$(5.5) \qquad \frac{\binom{n_1}{r_1}\binom{n_2}{r-r_1}p_0^{r_1}q_0^{n_1-r_1}p_1^{r-r_1}q_1^{n_2-r+r_1}}{\sum_k\binom{n_1}{k}\binom{n_2}{r-k}p_0^k q_0^{n_1-k}p_1^{r-k}q_1^{n_2-r+k}}.$$

Using the normal approximation of r_1 and r_2 in (5.5), we find that under the condition for r's being given, $w=(r_1-\mathrm{E}(r_1|r))\mathrm{V}(r_1|r)^{-1/2}$ is distributed asymptotically normal with mean zero and variance one, where

$$(5.6) \qquad \mathrm{E}(r_1|r)=\frac{n_1n_2p_0q_0p_1q_1}{n_1p_0q_0+n_2p_1q_1}\left(\frac{1}{q_0}-\frac{1}{q_1}+\frac{r}{n_2p_1q_1}\right) =\frac{n_1r}{n_1+n_2}-\frac{n_1n_2}{n_1+n_2}\left(\frac{dp}{d\theta}\right)_{\theta=\theta_0}(\theta-\theta_0)+0(1),$$

* Prof. M. Okamoto, Osaka University, remarks that this proof is heuristic and seems to be improved and simplified by generalizing the Theorem of Steck [8]. This point will be discussed in another occasion.

$$V(r_1|r)=\frac{n_1n_2}{n_1+n_2}p_0q_0(1+0(N^{-1/2}))\,. \tag{5.7}$$

On the other hand,

$$U_c=\frac{(n_1-r_1)(n_2-r+r_1)}{n_1n_2}U^*+\frac{(n_1-r_1)(r-r_1)}{n_1n_2}, \tag{5.8}$$

where U^* is calculated from U_c in (5.1) by excluding the r_1+r_2 zeroes in two samples. Since U^* is the ordinary Wilcoxon statistic for given r and r_1, $t=(U^*-E^*)V^{*-1/2}$ is distributed asymptotically normal with mean zero and variance one under the same condition, where

$$\begin{aligned} E^* &= \int_0^\infty F_1^*dF_0^*\,, \\ V^* &= \frac{1}{n_2-r+r_1}\int_0^\infty\left(F_0^*-\int_0^\infty F_0^*dF_1^*\right)^2 dF_1^* \\ &\quad+\frac{1}{n_1-r_1}\int_0^\infty\left(F_1^*-\int_0^\infty F_1^*dF_0^*\right)^2 dF_0^*. \end{aligned} \tag{5.9}$$

Rewriting (5.8) in terms of w and t instead of r_1 and U^* as in Halperin [2] and noticing $r=(n_1+n_2)p_0+0(N^{1/2})$, we have

$$\begin{aligned} U_c &= E+w\{p_0q_0/n_1n_2(n_1+n_2)\}^{1/2}\{(n_1-n_2)E^*-n_2p_0-n_1q_0\} \\ &\quad+tq_0^{3/2}\left\{\frac{1}{n_2}\int_0^\infty\left(F_0^*-\int_0^\infty F_0^*dF_1^*\right)^2 dF_1^*\right. \\ &\quad\left.+\frac{1}{n_1}\int_0^\infty\left(F_1^*-\int_0^\infty F_1^*dF_0^*\right)^2 dF_0^*\right\}^{1/2}+0(N^{-1})\,, \end{aligned} \tag{5.10}$$

whence follows Theorem 3.

In particular we get from Theorem 3,

Corollary. *Under the null hypothesis H, U_c is distributed asymptotically normal with mean E and variance V for given r, where*

$$\begin{aligned} E &= \frac{(n_1+n_2+r)(n_1+n_2-r)}{2(n_1+n_2)^2}\,, \\ V &= \frac{p_0q_0}{4n_1n_2(n_1+n_2)}\{n_1+n_2+(n_2-n_1)p_0\}^2+\frac{(n_1+n_2)q_0}{12n_1n_2}\,. \end{aligned} \tag{5.11}$$

From Theorem 3, we can calculate the asymptotic efficiency of the U_m test relative to the U_c conditional test for the location alternative.

Theorem 4. *Let two polulations Π_1 and Π_2 be defined by* (4.1), *then the asymptotic efficiency of the U_m test relative to the U_c conditional test at $r=(n_1+n_2)p_0$ is given by*

(5.12) $$e_{\phi,\phi'} = \frac{\left\{p_0 f(-\theta_0)+2\int_{-\theta_0}^{\infty} f(t)^2 dt\right\}^2 \left\{p_0(q_0+xp_0)^2+\frac{1}{3}q_0^2\right\}}{\left\{p_0 f(-\theta_0)x+2\int_{-\theta_0}^{\infty} f(t)^2 dt\right\}^2 \left(p_0+\frac{1}{3}q_0^2\right)},$$

where

$$x = \frac{2}{1+\lim \frac{n_1}{n_2}}.$$

Proof. The method of proof is the same as in Theorem 2. Corresponding to c^* in (4.5) and (4.6), we get from Theorem 3,

$$\begin{aligned} c^* &= -\left(\frac{dE}{d\theta}\right)_{\theta=\theta_0} V^{-1/2} \\ &= \sqrt{\alpha_1\alpha_2}\left((xf(-\theta_0)+2\int_{-\theta_0}^{\infty} f(t)^2 dt\right)\Big/\left\{p_0 q_0(q_0+xp_0)^2+\frac{1}{3}q_0^3\right\}^{1/2}. \end{aligned}$$

With c in (4.4), we get $e_{\phi,\phi'}=(c/c^*)^2$ in (5.12).

Resolving $e_{\phi,\phi'}$ into partial fractions with respect to x $(0\leqq x\leqq 2)$, we have

(5.13) $$e_{\phi,\phi'} = \frac{p_0 f(-\theta_0)+2\int_{-\theta_0}^{\infty} f(t)^2 dt}{\left(p_0+\frac{1}{3}q_0^2\right)f(-\theta_0)^2}\left\{p_0+\frac{2p_0 D}{p_0 f(-\theta_0)x+2\int_{-\theta_0}^{\infty} f(t)^2 dt} + \frac{p_0 D^2+\frac{1}{3}f(-\theta_0)^2 q_0^2}{\left(p_0 f(-\theta_0)x+2\int_{-\theta_0}^{\infty} f(t)^2 dt\right)^2}\right\},$$

where

(5.14) $$D = q_0 f(-\theta_0)-2\int_{-\theta_0}^{\infty} f(t)^2 dt.$$

Regarding $e_{\phi,\phi'}$ as a function of x, we get the following:

(i) When $D\geqq 0$, $e_{\phi,\phi'}(x)$ is nonincreasing. Hence $\max e_{\phi,\phi'}=e(0)$ and $\min e_{\phi,\phi'}=e(2)$.

(ii) When $D<0$, we put $x_0=-q_0 p_0^{-1}(1+f(-\theta_0)/3D)$, and

(a) if $x_0<0$, $\max e_{\phi,\phi'}=e(2)$ and $\min e_{\phi,\phi'}=e(0)$,

(b) if $0\leqq x_0\leqq 2$, $\max e_{\phi,\phi'}=\max(e(0), e(2))$ and

$$\min e_{\phi,\phi'}=p_0 q_0^2\left(p_0 f(-\theta_0)+2\int_{-\theta_0}^{\infty} f(t)^2 dt\right)^2\Big/\left(p_0+\frac{1}{3}q_0^2\right)(3p_0 D^2+q_0^2 f(-\theta_0)^2),$$

(c) if $x_0>2$, $\max e_{\phi,\phi'}=e(0)$ and $\min e_{\phi,\phi'}=e(2)$. From these facts we can get the following:

Corollary. *The asymptotic relative efficiency* $e_{\phi,\phi'}$ *in Theorem* 4 *is one* *when* $\lim n_1/n_2=1$ *and takes both larger values and smaller values than one as* $\lim n_1/n_2$ *changes, except when* $x_0=1$ *and* $D<0$.

EXAMPLE 4. Normal distribution in Example 1. When $\theta_0=0$, $D=\frac{1-\sqrt{2}}{2}$ and $x_0=0.6$. Hence this is the case (ii) (b). So we have

$$\max e_{\phi,\phi'} = (15+10\sqrt{2})/28 = 1.04\,,$$
$$\min e_{\phi,\phi'} = (57+40\sqrt{2})/343 = 0.99\,.$$

EXAMPLE 5. Exponential distribution in Example 2. Since D becomes always zero, this is the case (i). Hence

$$\max e_{\phi,\phi'} = 1+\frac{1-e^{2\theta_0}}{3(1-e^{\theta_0})+e^{2\theta_0}}\,,$$
$$\min e_{\phi,\phi'} = 1-\frac{e^{2\theta_0}(3-e^{\theta_0})(1-e^{\theta_0})}{(3-3e^{\theta_0}+e^{2\theta_0})(2-e^{\theta_0})^2}\,.$$

EXAMPLE 6. Uniform distribution in Example 3. Since $D=-(1+\theta_0)<0$, this is the case (ii). We have $x_0=1+\frac{4}{3\theta_0}+\frac{1}{3\theta_0^2}$. When

$$\theta_0=-\frac{1}{4}\,,\quad x_0=1\,,\quad \text{and}\quad \max e_{\phi,\phi'}=49/48\,,$$
$$\min e_{\phi,\phi'}=1\,,$$
$$\theta_0=-\frac{1}{2}\,,\quad x_0<0\,,\quad \text{and}\quad \max e_{\phi,\phi'}=261/224\,,$$
$$\min e_{\phi,\phi'}=45/56\,.$$

It is interesting that in case $\theta_0=-1/4$ the U_m test is better than the U_c conditional test irrespective of the value of $\lim n_1/n_2$.

6. Application

The following table shows the ratio of nasal to oral leakage at the time of blowing for each one of 38 cleft-palate patients classified according to their ages at operation.

age at operation											
1-3.	0,	0,	0,	0,	0,	0,	0,	0,	0.25,	0.46,	0.50,
	0.55,	0.62,	0.75,	0.84,	1.00,	1.70					
16-.	0,	0,	0,	0.11,	0.32,	0.47,	0.58,	0.70,	0.81,	0.83,	0.86,
	0.94,	1.01,	1.39,	1.39,	1.40,	1.44,	1.62,	1.85,	2.01,	2.50	

From these data, we want to test whether the ratio is stochastically larger in the group of operation age above 16 than the group of age

1-3. After numerical calculation we get $U_m=269/(21\times 17)$ and $(U_m-\mathrm{E}(U_m))/\mathrm{Var}(U_m)^{1/2}=2.72$ from (1.2), (2.3) and (2.4). From (5.2) and (5.11), we also get $U_c=257/(21\times 17)$ and $(U_c-E)/V^{1/2}=2.87$. Noticing the asymptotic normality of U_m and U_c, we can conclude that there is a significant difference between two groups ditected either by the U_m or the U_c test.

Acknowledgement: The author wishes to express his gratitude to Prof. M. Okamoto, Osaka University, for his valuable suggestion and instruction and also to Prof. G. Ishii, Osaka City University, for his kind advice and his calling author's attention to Halperin's paper. At the same time, great thanks are due to Mr. A. Takayori, Dental School, Osaka University, for his providing the data.

OSAKA UNIVERSITY

(Received September 30, 1963)

References

[1] D. A. S. Fraser: Nonparametric Methods in Statistics, New York, John Wiley & Sons, Inc. 1957.

[2] M. Halperin: *Extension of the Wilcoxon-Mann-Whitney test to samples censored at the same fixed point*, J. Amer. Statist. Assoc. **55** (1960), 125-138.

[3] W. Hoeffding: *A class of statistics with asymptotically normal distribution*, Ann. Math. Statist. **19** (1948), 293-325.

[4] W. H. Kruskal and W. A. Wallis: *Use of ranks in one criterion variance analysis*, J. Amer. Statist. Assoc. **47** (1952), 583-621.

[5] E. L. Lehmann: *Consistency and unbiasedness of certain nonparametric tests*, Ann. Math. Statist. **22** (1951), 165-179.

[6] A. M. Mood: *On the asymptotic efficiency of certain nonparametric two-sample tests*, Ann. Math. Statist. **25** (1954), 514-522.

[7] J. Putter: *The treatment of ties in some nonparametric tests*, Ann. Math. Statist. **26** (1955), 368-386.

[8] G. P. Steck: *Limit Theorems for Conditional Distributions*, University of California Publications in Statistics **2** (1957), 237-284.