The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1610–1614

ON THE TRIMMED MANN-WHITNEY STATISTIC[1]

By Thomas P. Hettmansperger

The Pennsylvania State University

1. Introduction and summary. Consider random samples from two independent distributions with absolutely continuous distribution functions $F(z)$ and $F(z - \theta)$, respectively. For testing the hypotheses $\theta = 0$ against $\theta > 0$, Hodges and Lehmann [6] show that the Pitman asymptotic efficiency of W, the Mann-Whitney form of the Wilcoxon statistic, with respect to the t-statistic is never smaller than .864 and in their 4th Berkeley Symposium paper [7] they indicate this efficiency is almost always greater than or equal to 1 for distributions with tails at least as heavy as those of a normal distribution. Hence for distributions with heavier tails, W is a more robust statistic than the t-statistic.

For the moment, consider a single sample of size n from a distribution with absolutely continuous distribution function $F(z - \theta)$. For distributions with heavier tails some authors have proposed the α-trimmed mean as an estimate of θ; that is, the mean based on the middle $n - 2[n\alpha]$ observations. Tukey [16] and Huber [8] study this statistic when F is a contaminated normal distribution. Bickel [3] studies the asymptotic relative efficiency properties of the α-trimmed mean relative to the mean for the class of continuous distributions with symmetric, unimodal densities. For estimating the location of a Cauchy distribution, Rothenberg, Fisher and Tilanus [12] show the trimmed mean based on the middle 24 percent of the observations is the most efficient trimmed mean relative to the maximum likelihood estimate.

This single sample statistic suggests a Mann-Whitney statistic based on trimmed samples. It is hoped that the effects of contamination by gross errors in the underlying distributions can be considerably reduced by considering such a statistic. We have the benefits of using a simple and well known rank statistic and, at the same time, of being able to increase the efficiency by adjusting the trimming proportions according to the weight in the tails of the underlying distributions. This type of statistic is also related to rank tests for censored data which have been studied by Basu [2], Gastwirth [5] and Sobel [14], [15]. If we have samples size m and n from absolutely continuous distributions corresponding to $F(z)$ and $F(z - \theta)$, respectively, we denote by W_α the Mann-Whitney statistic based on the middle $m - 2[m\alpha]$ and $n - 2[n\alpha]$ observations of the samples. We refer to W_α as the α-trimmed Mann-Whitney statistic. It is the purpose of this paper to investigate the Pitman asymptotic relative efficiency properties of W_α for a sub-class of absolutely continuous distributions. First some definitions are given in Section 2. In Section 3 and Section 4 we establish the asymptotic normality of W_α, derive the efficiency of W_α relative to W and give some examples. The results of Section 5 include a greatest lower bound on this efficiency.

Received 10 July 1967.

[1] This research is part of the author's doctoral theses at the University of Iowa.

It is interesting to note that this bound is the same one found by Bickel [3] for the efficiency of the trimmed mean relative to the mean.

2. Some notation and definitions. Let $X_1 < \cdots < X_m$ and $Y_1 < \cdots < Y_n$ be the order statistics of samples size m and n from two distributions with respective absolutely continuous distribution functions $F(z)$ and $F(z - \theta)$, where F has symmetric continuous density f. We further assume for $0 < \alpha < \frac{1}{2}$ that f is continuously differentiable in some neighborhood of the population quantiles of order α and $1 - \alpha$, respectively. Let K_α denote this class of distribution functions. We will assume throughout this paper that the sample sizes m and n increase in such a way that $\lim (m/(m + n)) = \lambda$, $0 < \lambda < 1$. Let T_{ni} be a two sample statistic such that $(T_{ni} - \mu_{ni}(\theta))/\sigma_{ni}(\theta)$ has an asymptotic normal distribution for all θ, $i = 1, 2$. Following Mood [11], the Pitman asymptotic relative efficiency of T_{n1} relative to T_{n2} is given by

$$e(T_{n1}, T_{n2}) = \lim_{n\to\infty} [\sigma_{n1}^2(0)]^{-1}(\mu'_{n1}(0))^2/[\sigma_{n2}^2(0)]^{-1}(\mu'_{n2}(0))^2$$

where $\mu'_{ni}(0)$ is the derivative of $\mu_{ni}(\theta)$ with respect to θ evaluated at $\theta = 0$, $i = 1, 2$.

Now let $\zeta_{ij} = 1$, if $Y_i > X_j$, and 0 otherwise, $i = 1, \cdots, n$, $j = 1, \cdots, m$, then for $0 < \alpha < \frac{1}{2}$ we define

$$W_\alpha = ((m - 2[m\alpha])(n - 2[n\alpha]))^{-1} \sum{}^* \zeta_{ij}$$

where $\sum^*$ implies the summation is extended over all i and j such that $[m\alpha] + 1 \leqq j \leqq m - [m\alpha]$ and $[n\alpha] + 1 \leqq i \leqq n - [n\alpha]$. We denote the sample quantiles $X_{[m\alpha]+1}$, $X_{m-[m\alpha]}$, $Y_{[n\alpha]+1}$ and $Y_{n-[n\alpha]}$ by X_α, $X_{1-\alpha}$, Y_α and $Y_{1-\alpha}$ respectively. The corresponding population quantiles which we assume to be unique, are denoted by b_α, $b_{1-\alpha}$, c_α and $c_{1-\alpha}$, respectively. If we let $\mathbf{Z} = (m^{\frac{1}{2}}(X_\alpha - b_\alpha), m^{\frac{1}{2}}(X_{1-\alpha} - b_{1-\alpha}), n^{\frac{1}{2}}(Y_\alpha - c_\alpha), n^{\frac{1}{2}}(Y_{1-\alpha} - c_{1-\alpha}))$, then, conditional on $\mathbf{Z}$, W_α is distributed like a Mann-Whitney statistic based on samples size $m - 2[m\alpha]$ and $n - 2[n\alpha]$ from distributions with densities $f(\zeta)/(F(X_{1-\alpha}) - F(X_\alpha))$ if $X_\alpha < \zeta < X_{1-\alpha}$ and 0 otherwise, and $f(\zeta - \theta)/(F(Y_{1-\alpha} - \theta) - F(Y_\alpha - \theta))$ if $Y_\alpha < \zeta < Y_{1-\alpha}$ and 0 otherwise. Finally we define

$$R_n = ((m - 2[m\alpha])(n - 2[n\alpha]))^{-1} \sum{}^* (\zeta_{ij} - E(\zeta_{ij} \mid \mathbf{Z})).$$

3. Asymptotic theory. Lehmann [9] shows the conditional distribution of $n^{\frac{1}{2}}R_n$, given $\mathbf{Z}$, is asymptotically normal for all θ and Cramer [4], p. 369, shows the joint density of $\mathbf{Z}$ converges pointwise to the multivariate normal density. It follows from Theorem 2 of Sethuraman [13] that

$$(n^{\frac{1}{2}}R_n, m^{\frac{1}{2}}(X_\alpha - b_\alpha), m^{\frac{1}{2}}(X_{1-\alpha} - b_{1-\alpha}), n^{\frac{1}{2}}(Y_\alpha - c_\alpha), n^{\frac{1}{2}}(Y_{1-\alpha} - c_{1-\alpha}))$$

has an asymptotic normal distribution for all θ. In case $\theta = 0$, the covariances are:

$\sigma_{11} = 1/12\lambda(1 - 2\alpha)$, $\sigma_{1j} = 0$ if $j \neq 1$, $\sigma_{22} = \sigma_{33} = \alpha(1 - \alpha)/f^2(b_\alpha)$, $\sigma_{23} = \alpha^2/f^2(b_\alpha)$, $\sigma_{44} = \sigma_{55} = \alpha(1 - \alpha)/f^2(c_\alpha - \theta)$ and $\sigma_{45} = \alpha^2/f^2(c_\alpha - \theta)$.

We note that $n^{\frac{1}{2}}(W_\alpha - E(W_\alpha)) = n^{\frac{1}{2}}R_n + n^{\frac{1}{2}}(E(\zeta \mid \mathbf{Z}) - E(W_\alpha))$, where

$$(3.1) \quad E(\zeta \mid \mathbf{Z}) = P(Y > X \mid \mathbf{Z}) = \int_{X_\alpha}^{X_{1-\alpha}} \int_v^{Y_{1-\alpha}} [f(w - \theta)f(v)]/[(F(Y_{1-\alpha} - \theta) - F(Y_\alpha - \theta))(F(X_{1-\alpha}) - F(X_\alpha))]\, dw\, dv.$$

An application of the theorem [1], p. 76, shows that $n^{\frac{1}{2}}(W_\alpha - E(W_\alpha))$ is asymptotically normally distributed.

It remains to calculate the asymptotic parameters in order to determine the efficiency. We first consider the asymptotic variance under the assumption that $\theta = 0$. In this case $E(W_\alpha) = \frac{1}{2}$ and $\operatorname{var}(n^{\frac{1}{2}}(W_\alpha - \frac{1}{2})) = \operatorname{var}(n^{\frac{1}{2}}R_n) + 2\operatorname{cov}(n^{\frac{1}{2}}R_n, n^{\frac{1}{2}}(E(\zeta \mid \mathbf{Z}) - \frac{1}{2})) + \operatorname{var}(n^{\frac{1}{2}}(E(\zeta \mid \mathbf{Z}) - \frac{1}{2}))$. Now, from 3.1,

$$E(\zeta \mid \mathbf{Z}) = (2F(Y_{1-\alpha}) - F(X_{1-\alpha}) - F(X_\alpha))/2(F(Y_{1-\alpha}) - F(Y_\alpha)).$$

If this is expanded in a Taylor Series about $(b_\alpha, b_{1-\alpha}, c_\alpha, c_{1-\alpha})$, we have

$$\operatorname{var}(n^{\frac{1}{2}}(W_\alpha - \tfrac{1}{2})) = (m + n)/12m(1 - 2\alpha) + \alpha(m + n)/2m(1 - 2\alpha)^2 + o(1/n)$$

Hence $\lim_{n\to\infty} \operatorname{var}(n^{\frac{1}{2}}(W_\alpha - \frac{1}{2})) = (1 + 4\alpha)/12\lambda(1 - 2\alpha)^2$. For any θ the asymptotic mean is $\int_{b_\alpha}^{b_{1-\alpha}} \int_v^{c_\alpha} f(w - \theta)f(v)(1 - 2\alpha)^{-2}\,dw\,dv$ and the derivative of this expression at $\theta = 0$ is $\int_{b_\alpha}^{b_{1-\alpha}} f^2(v)(1 - 2\alpha)^{-2}\,dv$. Note with $\alpha = 0$ this answer gives the corresponding result for the Mann-Whitney statistic.

THEOREM. *For the class K_α of distribution functions defined in Section 2 and for $0 < \alpha < \frac{1}{2}$ the asymptotic efficiency of W_α relative to W is*

$$e(\alpha) = (\int_{b_\alpha}^{b_{1-\alpha}} f^2(v)\,dv)^2/(1 + 4\alpha)(1 - 2\alpha)^2(\int_{-\infty}^{\infty} f^2(v)\,dv)^2.$$

4. Examples. If $f(v) = (2\pi)^{-\frac{1}{2}} \exp(-v^2)$, $-\infty < v < \infty$, the standard normal density function and $F(v)$ is the corresponding distribution function then

$$e(\alpha) = (2F(2^{\frac{1}{2}}b_{1-\alpha}) - 1)^2/(1 + 4\alpha)(1 - 2\alpha)^2$$

for $0 < \alpha < \frac{1}{2}$. In this case $e(\alpha) < 1$ for all $0 < \alpha < \frac{1}{2}$ and $e(\alpha)$ decreases to $\frac{2}{3}$ as α approaches $\frac{1}{2}$. If $f(v) = (\frac{1}{2})\exp(-|v|)$, $-\infty < v < \infty$, the Laplacian density, then

$$e(\alpha) = (1 + 2\alpha)^2/(1 + 4\alpha)$$

for $0 < \alpha < \frac{1}{2}$. Now $e(\alpha) > 1$ for all $0 < \alpha < \frac{1}{2}$ and $e(\alpha)$ increases to 4/3 as α approaches $\frac{1}{2}$. Finally let $f(v) = (\pi(1 + v^2))^{-1}$, $-\infty < v < \infty$, the Cauchy density, then

$$e(\alpha) = ((1/\pi)\sin(\pi(1 - 2\alpha)) + (1 - 2\alpha))^2/(1 + 4\alpha)(1 - 2\alpha)^2$$

for $0 < \alpha < \frac{1}{2}$. For this example we find that the most efficient trimmed Mann-Whitney statistic occurs for α approximately equal to .375; this requires the use of the middle 25 percent of each sample. We also note that the efficiency curve is quite flat around the maximum.

The following table with entries $e(\alpha)$ provides some illustrative calculations.

α	.05	.10	.25	.35	.40	.45	.49
normal	.99	.94	.92	.81	.75	.70	.66
Laplace	1.01	1.03	1.13	1.20	1.25	1.29	1.32
Cauchy	1.03	1.09	1.34	1.43	1.44	1.40	1.35

5. A bound on the efficiency. Bickel [3] shows the asymptotic efficiency of the α-trimmed mean, relative to the mean, is:

$$e^*(\alpha) = (1 - 2\alpha)^2 \int_{-\infty}^{\infty} v^2 f(v)\, dv / (\int_{b_\alpha}^{b_{1-\alpha}} v^2 f(v)\, dv + 2\alpha b_\alpha^2).$$

Moreover, he shows the greatest lower bound of $e(\alpha)$ is $1/(1 + 4\alpha)$ for the class of continuous distributions with symmetric, unimodal densities and the bound is achieved for any uniform distribution in the class. A similar result is now given for the α-trimmed Mann-Whitney statistic.

THEOREM. *For the subclass of K_α, $0 < \alpha < \frac{1}{2}$, with unimodal densities, the efficiency $e(\alpha)$ of the α-trimmed Mann-Whitney statistic, relative to the Mann-Whitney statistic, satisfies*

$$1/(1 + 4\alpha) \leqq e(\alpha) \leqq 1/(1 + 4\alpha)(1 - 2\alpha)^2.$$

The greatest lower bound $1/(1 + 4\alpha)$ is achieved by any uniform distribution in the class.

PROOF. To minimize $e(\alpha)$ we need only minimize

$$k(\alpha) = \int_{b_\alpha}^{b_{1-\alpha}} f^2(v)\, dv / \int_{-\infty}^{\infty} f^2(v)\, dv.$$

Fix

$$b_{1-\alpha} = \beta, \qquad f(\beta) = \delta, \qquad \int_{-\beta}^{\beta} f^2(v)\, dv = \varphi,$$

and first consider

$$\int_{-\beta}^{\beta} f^2(v)\, dv / \int_{-\infty}^{\infty} f^2(v)\, dv = \varphi/(\varphi + 2\int_{\beta}^{\infty} f^2(v)\, dv).$$

Since

$$\int_{\beta}^{\infty} f^2(v)\, dv \leqq f(\beta) \int_{\beta}^{\infty} f(v)\, dv = \alpha\delta = \int_{\beta}^{\beta+\alpha/\delta} f^2(\beta)\, dv,$$

to minimize $k(\alpha)$ we must choose f such that $f(v) = \delta$ if $\beta \leqq |v| \leqq \beta + \alpha/\delta$ and 0 if $|v| \geqq \beta + \alpha/\delta$. For fixed values of β and δ we next minimize $\varphi/(\varphi + 2\delta\alpha) = 1 - 2\delta\alpha/(\varphi + 2\delta\alpha)$, or equivalently, minimize $\varphi = \int_{-\beta}^{\beta} f^2(v)\, dv$. Clearly f must satisfy $f(v) = \delta_1$ if $|v| \leqq \beta$, where δ_1 is some constant. Hence for fixed β and δ the function which minimizes $k(\alpha)$ is

$$\begin{aligned} f(v) &= \delta_1 \quad |v| \leqq \beta \\ &= \delta \quad \beta \leqq |v| \leqq \beta + \alpha/\delta \\ &= 0 \quad |v| \geqq \beta + \alpha/\delta. \end{aligned}$$

We now consider $2\delta_1^2\beta/(2\delta_1^2\beta + 2\delta\alpha) = 1/(1 + \delta\alpha/\delta_1^2\beta)$. Since $\delta_1\beta = \frac{1}{2}(1 - 2\alpha)$ and f unimodal implies $\delta_1 \geqq \delta$, $k(\alpha)$ is minimized by taking $\delta_1 = \delta$ and $\min k(\alpha) = (1 - 2\alpha)$. Hence $1/(1 + 4\alpha) \leqq e(\alpha)$ and equality is attained by any uniform distribution in the class. The inequality $e(\alpha) \leqq 1/(1 + 4\alpha)(1 - 2\alpha)^2$ is clear from the conclusion of the Theorem of Section 3 since

$$\int_{b_\alpha}^{b_{1-\alpha}} f^2(v)\, dv \leqq \int_{-\infty}^{\infty} f^2(v)\, dv.$$

Acknowledgment. I wish to express my thanks to Professors Hogg and Robert-

son at the University of Iowa for their help and guidance in the preparation of this paper.

REFERENCES

[1] ANDERSON, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] BASU, A. P. (1967). On the large sample properties of a Generalized Wilcoxon-Mann-Whitney Statistic. *Ann. Math. Statist.* **38** 905–915.

[3] BICKEL, P. J. (1965). On some robust estimates of location. *Ann. Math. Statist.* **36** 847–858.

[4] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[5] GASTWIRTH, J. L. (1965). Asymptotically most powerful rank tests for the two sample problem with censored data. *Ann. Math. Statist.* **36** 1243–1247.

[6] HODGES, J. L. and LEHMANN, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27** 324–335.

[7] HODGES, J. L. and LEHMANN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 307–317. Univ. of California Press.

[8] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

[9] LEHMANN, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *Ann. Math. Statist.* **22** 165–179.

[10] MANN, H. B. and WHITNEY, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18** 50–60.

[11] MOOD, A. M. (1954). On the asymptotic efficiency of certain non-parametric two-sample tests. *Ann. Math. Statist.* **25** 514–522.

[12] ROTHENBERG, T., FISHER, F. and TILANUS, C. (1964). A note on estimation from a Cauchy sample. *J. Amer. Statist. Assoc.* **59** 460–463.

[13] SETHURAMAN, J. (1961). Some limit theorems for joint distributions. *Sankhyā* **23** 379–386.

[14] SOBEL, M. (1965). On a generalization of Wilcoxon's rank sum test for censored data. Technical Report No. 69, Univ. of Minnesota.

[15] SOBEL, M. (1966). On a generalization of Wilcoxon's rank sum test for censored data. Technical Report No. 69 (Revised) Univ. of Minnesota.

[16] TUKEY, J. (1960). A survey of sampling from contaminated distributions. *Contribution to Prob. and Statist.* Stanford Univ. Press. 448–486.

[17] WILCOXON, F. (1945). Individual comparisons by ranking methods. *Biometrics Bull.* **1** 80–83.