The Annals of Mathematical Statistics
1969, Vol. 40, No. 2, 445–455

SEQUENTIAL NONPARAMETRIC TWO-WAY CLASSIFICATION WITH A PRESCRIBED MAXIMUM ASYMPTOTIC ERROR PROBABILITY[1]

BY MELVIN N. WOINSKY AND LUDWIK KURZ

Bell Telephone Laboratories, Incorporated and New York University

1. Introduction. Consider the problem of classifying the iid random sample $X_1, \cdots, X_m$ into either of two populations π_1 and π_2, characterized by cdf $F(x)$ and $G(x)$, respectively. It is assumed that $F(x)$ and $G(x)$ are continuous and satisfy the relation $\int (F - G)\, dF > 0$. Further assume that independent reference random samples $Y_1, \cdots, Y_n$ and $Z_1, \cdots, Z_n$ are available from π_1 and π_2, respectively. A classification procedure based on the use of two Mann-Whitney [12] statistics can be defined as follows: Let

$$t_{gn} = (mn)^{-1} \textstyle\sum_{j=1}^{m} \sum_{i=1}^{n} [c(Y_i - X_j) - c(X_j - Z_i)] \tag{1.1}$$

where

$$c(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leqq 0 \end{cases}$$

and classify $(X_1, \cdots, X_m) = \mathbf{X}_m$ into π_1 if $t_{gn} > 0$ and into π_2 if $t_{gn} \leqq 0$. In addition, consider the situation where there is one reference sample $Y_1, \cdots, Y_n$ available from π_1 and the observer knows that $G(x) = F(x - s)$ and *knows* the value of the translation parameter $s > 0$. Then let

$$t_{sn} = (mn)^{-1} \textstyle\sum_{j=1}^{m} \sum_{i=1}^{n} [c(Y_i - X_j) - c(X_j - Y_i - s)] \tag{1.2}$$

and classify $\mathbf{X}_m$ into π_1 if $t_{sn} > 0$. Hudimoto [8] first proposed a classification rule similar to that given in (1.1). Gupta [5] gives a classification rule using magnitudes of Mann-Whitney statistics which is applicable to the two-sided problem.

The decision rules based on t_{gn} and t_{sn} have been studied [15], [16] for application to signal detection in a communication system. A signal detector samples the output of a communication channel which contains one of two stationary stochastic processes, corresponding to the conditions of "noise only" or signal imbedded in noise. The reference samples are stored in the detector and $\mathbf{X}_m$ is obtained by the detector for each bit of information sent. The translation case using t_{sn} represents a constant signal level "s" imbedded in additive noise. The

---

Received 22 January 1968; revised 31 July 1968.
[1] The research reported herein was supported in part by the Doctoral Support Plan of Bell Telephone Laboratories, Incoprorated, and in part by the National Science Foundation, under grant GK-1075. The work was performed while Dr. M. N. Woinsky was on full-time assignment at New York University and is based on part of a thesis submitted to the Graduate Division of the School of Engineering and Science of New York University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

value of the signal s is known at the detector. In communication applications the average error rate is of primary importance. Based on the large-sample probability of error, it was shown [16] that these fixed-sample size classification procedures are about as "efficient" as a two-sample Mann-Whitney test but, because of symmetry in the form of the statistics, are more suitable for applications in which it is desired to minimize the average error. Also, by choosing equal reference sample sizes we enhance the symmetry and improve the performance of decision rule (1.1). For instance, if the *a priori* probability of $\mathbf{X}_m \varepsilon \pi_1$ is one-half and the total number of samples is the same, these classification procedures provide approximately the same average error probability as a two-sample Mann-Whitney test with a critical region set to minimize the average error.

The assumption $\lambda = \int (F - G)\, dF > 0$ or its equivalent $P[Z \geqq Y] > \frac{1}{2}$, is required since the decision rule is based on the one-sided form of the Mann-Whitney statistic. The decision procedure (1.1) is not meaningful for $\lambda \leqq 0$ since the large-sample error probabilities are both equal to or greater than $\frac{1}{2}$.

The purpose of this investigation is to extend the decision procedures previously described so that the maximum probability of misclassification is a preassigned constant $p < \frac{1}{2}$. This is not possible for the fixed sample size procedure (1.1) since λ is unknown. Both large-sample error probabilities are continuous monotonically strictly decreasing functions of λ for $\lambda \varepsilon [0, \frac{1}{2}]$ and the errors equal $\frac{1}{2}$ at $\lambda = 0$. Then for any fixed sample size we can find a sufficiently small λ to make the errors larger than p. Based on the asymptotic normality of $n^{\frac{1}{2}}[t_{gn} - E(t_{gn})]$ and its asymptotic variance, a fixed sample size procedure could achieve the desired result for large n if the nuisance parameter $\lambda/\sigma(t_{gn})$ were known. Our solution to the problem is based on sequentially estimating this parameter. The reference samples are obtained sequentially from π_1 and π_2. The stopping rule is defined by forming a nonparametric estimate of the nuisance parameter and comparing the estimate to a boundary which depends upon p. The sample to be classified $\mathbf{X}_M$ is then taken nonsequentially. The size of this sample M is a rv since it is taken as a fixed proportion of the reference sample size N. The random samples of random size are than used in the decision equations of (1.1) or (1.2). It is shown that the sequential procedure terminates with probability 1 and is asymptotically efficient in the sence of Chow and Robbins [4]. In addition, a general upper bound is found for the expected sample size. The asymptotic distribution of the sequential decision criterion is found as $p \to 0$. The asymptotic distribution indicates that p represents an approximation to the maximum error probability for small values of p. It should be emphasized that except for the bound on the expected sample size the results are asymptotic.

2. Fixed-sample size results. Before we define the sequential procedure, the following fixed-sample size results (see Appendix A) are required. Let $m = [2rn/(1 - r)]$ for use in t_{gn} and $m = [rn/(1 - r)]$ for use in t_{sn} with $0 < r < 1$. Then as $n \to \infty$, $n^{\frac{1}{2}}[t_{sn} - E(t_{sn})]$ and $n^{\frac{1}{2}}[t_{gn} - E(t_{gn})]$ both have asymptotic nondegenerate normal distributions for $\mathbf{X}_m \varepsilon \pi_1$ and $\mathbf{X}_m \varepsilon \pi_2$. The moments of the distributions are

$$
\begin{aligned}
&E_1(t_{gn}) = E_1(t_{sn}) = \lambda, \qquad E_2(t_{gn}) = E_2(t_{sn}) = -\lambda, \\
&\sigma_1^2(t_{gn}) = n^{-1}\sigma_{1g}^2 + n^{-2}(1 - r)[\tfrac{1}{6} - (\epsilon^2 - \lambda^2)]/2r, \qquad (2.1)\\
&\sigma_1^2(t_{sn}) = n^{-1}\sigma_{1s}^2 + n^{-2}(1 - r)(\tfrac{1}{3} + \lambda^2 - \lambda)/r
\end{aligned}
$$

where

$$\sigma_{1g}^2 = (1 - 3[\lambda^2 + \epsilon_1 - \epsilon_2 - r(\epsilon_1 - 3\epsilon_2 + 2\lambda - \lambda^2)])/6r, \tag{2.2}$$

$$\sigma_{1s}^2 = (1 - 3[\lambda^2 + (\epsilon_1 - \epsilon_2)(1 - 2r)])/3r, \tag{2.3}$$

and

$$
\begin{aligned}
&\epsilon_1 = \tfrac{1}{3} - \textstyle\int G^2\, dF, \qquad \epsilon_2 = \tfrac{1}{3} - \int (1 - F)^2\, dG, \\
&\lambda = \textstyle\int (F - G)\, dF, \qquad (2.4)\\
&\epsilon^2 = 2\lambda - (\epsilon_1 + \epsilon_2) = \textstyle\int (F - G)^2\, dF.
\end{aligned}
$$

E_1 and E_2 represent the expectation given that $\mathbf{X}_m\, \varepsilon\, \pi_1$ and $\mathbf{X}_m\, \varepsilon\, \pi_2$, respectively. The notation for the variance is similarly defined. Relations for σ_{2g}^2 and σ_{2s}^2 can be obtained from (2.2) and (2.3) by interchanging ϵ_1 and ϵ_2.

Estimators of λ, ϵ_1 and ϵ_2 based on the reference samples are required. For the case where independent reference samples $\mathbf{Y}_n$, $\mathbf{Z}_n$ are available from π_1 and π_2, respectively, we form the empirical cdf's $F_n(x)$ based on $\mathbf{Y}_n$ and $G_n(x)$ based on $\mathbf{Z}_n$. The estimators

$$
\begin{aligned}
&\hat{\lambda} = \tfrac{1}{2} - n^{-2} \textstyle\sum_{i=1}^n \sum_{j=1}^n c(Y_i - Z_j), \\
&\hat{\epsilon}_1 = \tfrac{1}{3} - n^{-3} \textstyle\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n c(Y_i - Z_j)c(Y_i - Z_k), \qquad (2.5)\\
&\hat{\epsilon}_2 = \tfrac{1}{3} + 2\hat{\lambda} - n^{-3} \textstyle\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n c(Z_i - Y_j)c(Z_i - Y_k),
\end{aligned}
$$

are formed by replacing F and G by F_n and G_n in equations (2.4). The estimator $\hat{\lambda}$ is the Mann-Whitney [12] statistic and is [2] the UMV unbiased estimated of λ. Similar estimators can be constructed in an obvious manner for the separate translation case.

3. Sequential procedure and results. Results are obtained [16] for sequential procedures based both on t_{gN} and t_{sN}. The separate translation case performs better than the general case as is apparent from the variance relations of (2.2) and (2.3). For the same total number of samples and the same number of reference samples in each case we have $\sigma_{1g}^2 - \sigma_{1s}^2 = \sigma_{2g}^2 - \sigma_{2s}^2 = m^{-1}(\epsilon^2 - \lambda^2)$ and by the Schwarz inequality $\epsilon^2 \geqq \lambda^2$ for all (F, G). The limiting results would then indicate that on the average more samples are required in the general case to meet the same error p. For brevity, however, we will only consider the procedure based on t_{gN} and in the latter work *the subscript g is omitted.*

We are interested in the maximum variance or

$$
\begin{aligned}
\sigma^2(\lambda, \epsilon_1, \epsilon_2) &= \max[\sigma_1^2, \sigma_2^2] \\
&= [1 + 6r\lambda - 3(1 + r)\lambda^2 \qquad (3.1)\\
&\qquad - 3(1 - r)(\epsilon_1 + \epsilon_2) + 6(1 - 2r)A]/6r
\end{aligned}
$$

where

$$A = \max(\epsilon_1, \epsilon_2) \quad \text{if} \quad 0 < r \leqq \tfrac{1}{2}$$
$$= \min(\epsilon_1, \epsilon_2) \quad \text{if} \quad \tfrac{1}{2} \leqq r \leqq 1.$$

It can be shown (see Appendix A) that

$$(24r)^{-1} \leqq (1 - 3\lambda^2)/6r \leqq \sigma^2(\lambda, \epsilon_1, \epsilon_2) \leqq B(r)(1 - 3\lambda^2) \leqq B(r) \tag{3.2}$$

where

$$B(r) = (1 - r)/2r \quad \text{if} \quad 0 < r \leqq \tfrac{1}{2}$$
$$= (1 + r)/6r \quad \text{if} \quad \tfrac{1}{2} \leqq r < 1.$$

Relation (3.2) is valid if the parameters λ, ϵ_1 and ϵ_2 are replaced by their estimates. Then $\sigma = \sigma(\lambda, \epsilon_1, \epsilon_2)$ and $\hat{\sigma} = \sigma(\hat{\lambda}, \hat{\epsilon}_1, \hat{\epsilon}_2)$ are positive and bounded.

In the fixed sample size decision procedure if λ, ϵ_1 and ϵ_2 are known then the sample size (continuous)

$$n(p) = \sigma^2(\Phi^{-1}[1 - p])^2/\lambda^2 \tag{3.3}$$

will cause the maximum of the two approximate large-sample misclassification errors to be equal to p. The actual sample size used would be $[n(p)]$ which is the smallest integer equal to or greater than $n(p)$. The function $\Phi^{-1}(\cdot)$ represents the inverse function of $\Phi(x)$, the standard normal cdf. Relation (3.3) will be sequentialized (λ and σ estimated sequentially) noting that $\hat{\lambda} \varepsilon [-\frac{1}{2}, \frac{1}{2}]$ although $\lambda \varepsilon (0, \frac{1}{2}]$.

The sequential stopping rule is defined as follows: Take one sample at a time, from each population π_1 and π_2, and stop with the random reference samples $Y_1, \cdots, Y_N, Z_1, \cdots, Z_N$ where N is the first integer $n \geqq n_0$ such that

$$n^{\frac{1}{2}}(\hat{\lambda}/\hat{\sigma}) \geqq \Phi^{-1}(1 - p) \tag{3.4}$$

with $n_0 \geqq 1$, a fixed integer. After the reference samples are obtained, define a rv $M = [2rN/(1 - r)]$ and nonsequentially obtain the sample $(X_1, \cdots, X_M) = \mathbf{X}_M$ which is to be classified. Form the statistic t_N, as given in Equation (1.1), and classify $\mathbf{X}_M$ into π_1 if $t_N > 0$ and into π_2 if $t_N \leqq 0$. The value p is the prescribed maximum error probability and satisfies $0 < p < \frac{1}{2}$. The ratio r is that proportion of the total number of samples which belongs to the group of samples to be classified. A proper choice for r would depend upon the relative cost of obtaining reference samples and samples to be classified.

Using the upper bounds on σ (3.2), simplified sequential procedures can be constructed which only require the estimation of λ. The simplified procedures, however, require more samples to achieve the same error p.

The results can be stated as follows:

$$P[N < \infty] = 1, \tag{3.5}$$

$$\lim_{p \to 0} N/n(p) = 1 \quad \text{a.s.}, \tag{3.6}$$

$$E(N/n(p)) \leqq 6rB(r)(\max([\bar{n}(p)]; n_0)/\bar{n}(p) + Q[p]), \tag{3.7}$$

where

$$Q[p] = [\bar{n}(p)(g_1 - 1)]/\bar{n}(p) + a/\{g_2(1 - \exp(-a(g_1 - 1)/g_1))\},$$

$$g_1 = (1 + 1/g_2 + 2/g_2^{\frac{1}{2}})^{\frac{1}{2}},$$

$$g_2 = 2B(r)(1 - 3\lambda^2)^3[\Phi^{-1}(1 - p)]^2, \tag{3.8}$$

$$a = 2\lambda^2(1 - 3\lambda^2)^2,$$

$$\bar{n}(p) = g_2/a,$$

$$\lim_{p\to 0} E(N/n(p)) = 1, \tag{3.9}$$

$$\lim_{p\to 0} P[n^{\frac{1}{2}}(p)(t_N - \lambda) \leqq x] = \Phi(x/\sigma_1) \quad \text{if} \quad \mathbf{X}_m \,\varepsilon\, \pi_1, \tag{3.10}$$

$$\lim_{p\to 0} P[n^{\frac{1}{2}}(p)(t_N + \lambda) \leqq x] = \Phi(x/\sigma_2) \quad \text{if} \quad \mathbf{X}_m \,\varepsilon\, \pi_2. \tag{3.11}$$

Applying the asymptotic distributions of (3.10) and (3.11), for small values of p, to the sequential dexicion procedure leads to the result that the maximum of the two misclassification errors is equal to p for all $\lambda \,\varepsilon\, (0, \frac{1}{2}]$.

The first term (within the parenthesis) in the upper bound on the expected sample size, relation (3.7), would be equal to one in most applications. The term $Q[p]$ is a positive monotonically increasing function of λ. As an example of the magnitude of $Q[p]$, if $r = \frac{1}{2}$ and $p = .005$ we get $Q[p] \leqq 1.35$ for $\lambda \,\varepsilon\, [0, \frac{1}{4}]$, and $Q[p] \leqq 16.1$ for $\lambda \,\varepsilon\, [0, \frac{1}{2}]$. Relation (3.9) indicates that the sequential procedure is asymptotically efficient. This interpretation follows from the fact that if the maximum error is equal to p and $E(N) = n(p)$, we do as well, on the average, not knowing λ, ϵ_1 and ϵ_2 as would be possible if these parameters were known and a fixed-sample size procedure were used.

4. Proof of the sequential results. A sequence of sequential procedures is constructed by considering a sequence of decreasing prescribed error bounds $\{p_u\}$. Let u be an index for the sequence, i.e., $u = 1, 2, 3, \cdots$ and the sequence $\{p_u\}$ be such that $\lim_{u\to\infty} p_u = 0$. Then $\{n(p_u)\}$ as given by (3.3), or $\{n_u\}$ for brevity, is an increasing sequence of positive numbers and $\lim_{u\to\infty} n_u = \infty$. For each p_u the stopping rule of (3.4) generates the rv N_u. We then have a sequence of rv's $\{N_u\}$ corresponding to the sequence $\{p_u\}$.

THEOREM 1. *If* $\lambda > 0$, *then* $P[N_u < \infty] = 1$.

PROOF. If follows from Sen [13] that $\hat{\lambda} \to \lambda$ a.s. Also, the uniform (on the reals) a.s. convergence of empirical cdf's and the triangle law insure that $\hat{\epsilon}_1 \to \epsilon_1$ a.s., $\hat{\epsilon}_2 \to \epsilon_2$ a.s. and by continuity

$$\hat{\lambda}/\hat{\sigma} \to \lambda/\sigma \quad \text{a.s.} \tag{4.1}$$

From (3.4)

$$P[N_u = \infty] = P\{n^{\frac{1}{2}}(\hat{\lambda}/\hat{\sigma}) < \Phi^{-1}[1 - p_u] \quad \text{for all} \quad n \geqq n_0\}, \tag{4.2}$$

but (4.1) implies that (4.2) is zero if $\lambda > 0$, which completes the proof.

THEOREM 2. $\lim_{u\to\infty} N_u/n_u = 1$ a.s.

PROOF. From stopping rule (3.4) and the lower bound of (3.2), it follows that

$$N_u \geqq (\Phi^{-1}[1 - p_u])^2/6r$$

which implies

$$\lim_{u\to\infty} N_u = \infty \quad \text{a.s.} \tag{4.3}$$

Then following a procedure similar to that used by Chow and Robbins [14], Lemma 1, we get $\lim_{u\to\infty} (n_u/N_u)^{\frac{1}{2}} = 1$ a.s. which implies the theorem.

The following lemma is useful in obtaining the results concerning $E(N_u/n_u)$.

LEMMA 1. *Let*

$$S(a, b) = b^{-1} \textstyle\sum_{j=0}^{\infty} \exp\{-a((b + j)^{\frac{1}{2}} - b^{\frac{1}{2}})^2\} \tag{4.4}$$

with a and b positive. Then

$$\begin{aligned} S(a, b) \leqq S_U(a, b) &= b^{-1}[b^{\frac{1}{2}}((b + d)^{\frac{1}{2}} - b^{\frac{1}{2}})] \\ &\quad + b^{-1}(1 - \exp(-a(1 - (d/b + 1)^{-\frac{1}{2}})))^{-1} \end{aligned} \tag{4.5}$$

is a valid upper bound for any value of $d \in (0, \infty)$. *In* (4.5) *the brackets have the same meaning as stated with respect to* $[n(p)]$. *The* $\min_d \{S_U(a, b)\}$ *occurs when* d *is chosen as the solution of* $\cosh(g) = (a - g)^2/(2ab) + 1$ *where* $g = a(1 - (d/b + 1)^{-\frac{1}{2}})$. *The solution for* $2b/a \gg 1$ is $d \doteq 1/a + 2(b/a)^{\frac{1}{2}}$. *Using this value of d in* (4.4) *gives*

$$S(a, b) \leqq S_U(a, b) = O((ab)^{-\frac{1}{2}}) \quad \text{as} \quad ab \to \infty. \tag{4.6}$$

PROOF. Let $C(j, b) = ((b + j)^{\frac{1}{2}} - b^{\frac{1}{2}})^2$ and consider j a continuous variable. The curve of $C(j, b)$ vs j is positive and concave upward. Draw a tangent to the curve at the point $j = d$. The equation of the tangent is

$$T(j, b, d) = (1 - (d/b + 1)^{-\frac{1}{2}})(j - b^{\frac{1}{2}}((b + d)^{\frac{1}{2}} - b^{\frac{1}{2}}))$$

and it follows that $C(j, b) \geqq C'(j, b)$ where

$$\begin{aligned} C'(j, b) &= 0 && \text{for} \quad 0 \leqq j < [b^{\frac{1}{2}}((b + d)^{\frac{1}{2}} - b^{\frac{1}{2}})] \\ &= T(j, b, d) && \text{for} \quad j \geqq [b^{\frac{1}{2}}((b + d)^{\frac{1}{2}} - b^{\frac{1}{2}})]. \end{aligned} \tag{4.7}$$

Then

$$S(a, b) \leqq S_U(a, b) = b^{-1} \textstyle\sum_{j=0}^{\infty} \exp\{-aC'(j, b)\}$$

and from (4.7)

$$S_U(a,b) = b^{-1}[b^{\frac{1}{2}}((b + d)^{\frac{1}{2}} - b^{\frac{1}{2}})] + b^{-1} \textstyle\sum_{i=0}^{\infty} \exp\{-a(1 - (d/b + 1)^{-\frac{1}{2}})i\} \tag{4.8}$$

where the index has been translated in obtaining the second term of (4.8). The remaining series is geometric so that (4.5) follows directly. The approximate solution for d is obtained by using the first two terms of the Maclaurin expansion for cosh (g) when $2b/a \gg 1$. Letting $d = 1/a + 2(b/a)^{\frac{1}{2}}$ in (4.5) yields a complicated expression which for $(ab)^{\frac{1}{2}} \gg 2$ reduces to $S_U(a, b) \doteq 2(ab)^{-\frac{1}{2}} + (3/2)(ab)^{-1} + \frac{1}{2}(ab)^{-3/2}$ which implies (4.6).

THEOREM 3. *An upper bound on* $E(N_u/n_u)$ *is given by* (3.7) *and* (3.8) *and in addition*

$$\lim_{u\to\infty} E(N_u/n_u) = 1.$$

PROOF. Consider a different sequential stopping rule based on the upper bound of (3.2). Let

$$\bar{n}_u = B(r)(\Phi^{-1}[1 - p_u])^2/D^2(\lambda) \tag{4.9}$$

where

$$D(\lambda) = \lambda/(1 - 3\lambda^2)^{\frac{1}{2}}$$

and $\bar{N}_u$ be the first integer $n \geqq n_0$ such that

$$n^{\frac{1}{2}}D(\hat{\lambda}) \geqq \bar{n}_u^{\frac{1}{2}}D(\lambda). \tag{4.10}$$

Then from (3.2) it follows that $N_u \leqq \bar{N}_u$ for *every* u and

$$N_u/n_u \leqq (\bar{n}_u/n_u)(\bar{N}_u/\bar{n}_u) \leqq 6rB(r)(\bar{N}_u/\bar{n}_u). \tag{4.11}$$

We have $E(\bar{N}_u/\bar{n}_u) = n_0/\bar{n}_u + (1/\bar{n}_u)\sum_{n=n_0}^{\infty} P[\bar{N}_u > n]$ and it follows easily that

$$E(\bar{N}_u/\bar{n}_u) \leqq \max([\bar{n}_u]; n_0)/\bar{n}_u + Q'[p_u] \tag{4.12}$$

where

$$Q'[p_u] = (1/\bar{n}_u)\sum_{n=[\bar{n}_u]}^{\infty} P[\bar{N}_u > n]. \tag{4.13}$$

For a fixed integer n, if inequality (4.10) is satisfied it follows that the sequential procedure must have terminated in at most n samples. We then have

$$P[\bar{N}_u \leqq n] \geqq P[n^{\frac{1}{2}}D(\hat{\lambda}) \geqq \bar{n}_u^{\frac{1}{2}}D(\lambda)] \tag{4.14}$$

or

$$P[\bar{N}_u > n] \leqq P[D(\lambda) - D(\hat{\lambda}) > D(\lambda)(n^{\frac{1}{2}} - \bar{n}_u^{\frac{1}{2}})/n^{\frac{1}{2}}]. \tag{4.15}$$

By the mean value theorem

$$D(\lambda) - D(\lambda') \leqq k(\lambda)(\lambda - \lambda') \quad \text{if} \quad \lambda \geqq \lambda', \tag{4.16}$$

where

$$k(\lambda) = \max_{0 \leqq \xi \leqq \lambda < \frac{1}{2}} \{dD(\lambda_1)/d\lambda_1 |_{\lambda_1 = \xi}\} = D^3(\lambda)/\lambda^3.$$

Since $(1 - 3\hat{\lambda}^2)^{\frac{1}{2}} > 0$ for all $\hat{\lambda}\varepsilon[-\frac{1}{2}, \frac{1}{2}]$, $D(\hat{\lambda})$ can be replaced by $\hat{\lambda}$ in (4.10) without effecting the rv $\bar{N}_u$. This was taken into account in obtaining (4.16) and $k(\lambda)$. Since $D(\lambda_1)$ is monotonic strictly increasing it follows from (4.15) and (4.16) that

$$P[\bar{N}_u > n] \leqq P[\lambda - \hat{\lambda} > D(\lambda)(n^{\frac{1}{2}} - \bar{n}_u^{\frac{1}{2}})/n^{\frac{1}{2}}k(\lambda)] \quad \text{for } n \geqq \bar{n}_u. \tag{4.17}$$

From Hoeffding's [7] bound on 2-sample U-statistics we have

$$P[\lambda - \hat{\lambda} \geqq h] \leqq \exp(-2nh^2) \quad \text{for } h \geqq 0. \tag{4.18}$$

When (4.17) and (4.18) are used in conjunction with (4.13) and a shift is made in the summing index, we get

$$Q'[p_u] \leqq Q[p_u] = S_U(a, b) \tag{4.19}$$

where $S_U(a, b)$ is defined in Lemma 1, Equation (4.4), and

$$a = 2D^2(\lambda)/k^2(\lambda) = 2\lambda^2(1 - 3\lambda^2)^2, \qquad b = \bar{n}_u.$$

The bound of (3.8) is obtained by using (4.11) and by setting $d = 1/a + 2(b/a)^{\frac{1}{2}}$ in (4.5). This gives a good bound in the region of interest since $b = \bar{n}_u$ will normally be large. From (4.6), $Q[p_u] = O[(a\bar{n}_u)^{-\frac{1}{2}}]$ so that $\lim_{u\to\infty} Q[p_u] = 0$, which in conjunction with (4.12) implies that

$$\limsup_{u\to\infty} E(\bar{N}_u/\bar{n}_u) \leqq 1. \tag{4.20}$$

From the proof of Theorem 2 it is clear that $\bar{N}_u/\bar{n}_u \to 1$ a.s. and from Loéve [11], Section 11.4A, this implies that $\liminf_{u\to\infty} E(\bar{N}_u/\bar{n}_u) \geqq 1$ so that from (4.20), $\lim_{u\to\infty} E(\bar{N}_u/\bar{n}_u) = 1$ and the family $(\bar{N}_u/\bar{n}_u : 0 < p_u < \frac{1}{2})$ is uniformly integrable. It follows from (4.11) that N_u/n_u is also uniformly integrable so that from Theorem (2), $\lim_{u\to\infty} E(N_u/n_u) = 1$ which completes the proof.

THEOREM 4. *In the limit as $u \to \infty$, the distribution of $n_u^{\frac{1}{2}}(t_{N_u} - \lambda)$ is $N(0, \sigma_1^2)$ if $\mathbf{X}_m \varepsilon \pi_1$, and the distribution of $n_u^{\frac{1}{2}}(t_{N_u} + \lambda)$ is $N(0, \sigma_2^2)$ if $\mathbf{X}_m \varepsilon \pi_2$.*

PROOF. From Section 2 we have that as $n \to \infty$, $n^{\frac{1}{2}}(t_n - \lambda)$ is $N(0, \sigma_1^2)$ if $\mathbf{X}_m \varepsilon \pi_1$ and $n^{\frac{1}{2}}(t_n + \lambda)$ is $N(0, \sigma_2^2)$ if $\mathbf{X}_m \varepsilon \pi_2$ and from Theorem 2, $(N_u/n_u) \to 1$ as $u \to \infty$. Theorem 4 then follows from Anscombe [1], Theorem 1, if the statistic t_n satisfies condition (C2) of reference [1]. This condition can be stated as follows: Given any small positive δ_1 and δ_2, there is a large v and small positive δ_3 such that for any $n > v$, $P\{n^{\frac{1}{2}} |t_{n'} - t_n| < \delta_1$ simultaneously for all integers n' such that $|n' - n| < \delta_3 n\} > 1 - \delta_2$.

The statistic t_n is a special case of a simplified form of generalized U-statistic [10] which can be written as,

$$U_n = (mnn_1)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{n_1} \varphi(X_i, Y_j, Z_k) \tag{4.21}$$

where $m = [c_1 n]$, $n_1 = [c_2 n]$, and $0 < c_{1,2} < \infty$. In (4.21), the three samples have arbitrary cdf's but we assume $E[\varphi(X, Y, Z)] = \theta$. We will show that if $E[\varphi(X, Y, Z) - \theta]^4 < \infty$, then U_n satisfies condition (C2). Since for t_n, $\varphi(X, Y, Z)$ is bounded this latter condition is automatically satisfied and this will complete the proof.

Let $\varphi_1(x) = E[\varphi(x, Y, Z)]$, $\varphi_2(y) = E[\varphi(X, y, Z)]$, and $\varphi_3(z) = E[\varphi(X, Y, z)]$ and form

$$U_m^{(1)} = m^{-1} \sum \varphi_1(X_i), \qquad U_n^{(2)} = n^{-1} \sum \varphi_2(Y_j), \qquad U_{n_1}^{(3)} = n_1^{-1} \sum \varphi_3(Z_k) \tag{4.22}$$

and

$$(4.23) \quad d_n = n^{\frac{1}{2}}(U_n - \theta) - [n^{\frac{1}{2}}(U_m^{(1)} - \theta) + n^{\frac{1}{2}}(U_n^{(2)} - \theta) + n^{\frac{1}{2}}(U_{n_1}^{(3)} - \theta)].$$

It follows from the statistics of (4.22) and results of Anscombe [1], that if $d_n \rightarrow 0$ a.s., as $n \rightarrow \infty$, then U_n satisfies condition (C2). In proving asymptotic normality of U-statistics it has been shown [6], [10] that $d_n \rightarrow 0$ in probability as $n \rightarrow \infty$ if $E[\varphi(X, Y, Z) - \theta]^2 < \infty$. By a similar but lengthy procedure it can be shown [16] that if $E[\varphi(X, Y, Z) - \theta]^4 < \infty$ then $E[d_n^4] = O(n^{-2}) < \infty$. From Loéve [11] this latter result implies that $d_n \rightarrow 0$ a.s., as $n \rightarrow \infty$.

5. Acknowledgment. The authors are grateful to the referee for making some helpful suggestions, and in particular for suggesting a method of estimating all three of the parameters instead of only estimating λ.

APPENDIX A

Moments for Fixed-Sample Size Statistics

The statistic t_{gn} of (1.1) can be generalized slightly by letting the reference samples from the two populations be of different size. Let $Y_1, \cdots, Y_{n_1}$ and $Z_1, \cdots, Z_{n_2}$ be independent random samples from π_1 and π_2, respectively. Then

$$(A1) \quad \begin{aligned} t_g &= (mn_1)^{-1} \textstyle\sum_{j=1}^{m} \sum_{i=1}^{n_1} c(Y_i - X_j) \\ &\quad - (mn_2)^{-1} \textstyle\sum_{j=1}^{m} \sum_{i=1}^{n_2} c(X_j - Z_i) \\ &= U_1 - U_2 \end{aligned}$$

where $c(x)$ is defined in (1.1). The moments of t_g follow directly from Mann and Whitney [12] once we have the correlation coefficient $\rho(U_1, U_2)$. It is then only necessary to find

$$(A2) \quad E(U_1U_2) = (n_1n_2m^2)^{-1} \textstyle\sum E[c(Y_i - X_j)c(X_k - Z_v)].$$

If $\mathbf{X}_m \varepsilon \pi_1$, we have Y_i and $X_j \sim F(x)$, $Z_v \sim G(x)$ so that

$$\begin{aligned} E_1[c(Y_i - X_j)c(X_k - Z_v)] &= P[Y_i > X_j] \cdot P[X_k > Z_v] = \tfrac{1}{2} \textstyle\int G\,dF, \quad (j \neq k), \\ E_1[c(Y_i - X_j)c(X_j - Z_v)] &= P[Y_i > X_j, X_j > Z_v] \\ &= \textstyle\int P[Y_i > x_j]P[Z_v < x_j]\,dF(x_j) \\ &= \textstyle\int (1 - F)G\,dF \end{aligned}$$

and

$$E_1(U_1U_2) = (m - 1)(2m)^{-1} \textstyle\int G\,dF + m^{-1} \textstyle\int (1 - F)G\,dF,$$

which leads to

$$(A3) \quad \rho_1(U_1, U_2) = -\frac{[1 - 6(\lambda - \epsilon_2)]n_1^{\frac{1}{2}}n_2^{\frac{1}{2}}}{(m + n_1 + 1)^{\frac{1}{2}}[m + n_2 + 1 - 12\lambda^2(m + n_2 - 1) + 12(\lambda - \epsilon_1)(n_2 - 1) + 12(\lambda - \epsilon_2)(m - 1)]^{\frac{1}{2}}}$$

where λ, ϵ_1, and ϵ_2 are defined in Section 2. The coefficient $\rho_2(U_1, U_2)$ can be obtained from (A3) by interchanging ϵ_1 and ϵ_2. If we let $\lambda = \epsilon_1 = \epsilon_2 = O(F(x) = G(x))$, Equation (A3) reduces to a result of Whitney [14].

Asymptotic normality follows from the fact that t_g is a generalized U-statistic [10]. From Lehmann [9], the terms of $O(m^{-1})$ in $\sigma^2(U_1)$ and $\sigma^2(U_2)$ cannot both be simultaneously zero. It then follows, from (A1), that the limiting distribution is nondegenerate if $\rho_1(U_1, U_2)$ and $\rho_2(U_1, U_2)$ are nonpositive. This is the case since, we shall show in the next paragraph that $\lambda - \epsilon_{1,2} \leqq \frac{1}{6}$. Results for t_{sn} can be obtained in a similar manner.

From the Schwarz inequality, $\int (F - G)^2 \, dF \geqq [\int (F - G) \, dF]^2$ and clearly $G^2 \leqq G$ and $(1 - F)^2 \leqq 1 - F$. Using these inequalities in relations (2.4) yields,

$$\epsilon_1 + \epsilon_2 \leqq 2\lambda - \lambda^2 \tag{A4}$$

$$\lambda - \tfrac{1}{6} \leqq \epsilon_{1,2} \leqq \tfrac{1}{3}. \tag{A5}$$

By replacing F and G by F_n and G_n, respectively we obtain the same relations for $\hat{\lambda}$, $\hat{\epsilon}_1$ and $\hat{\epsilon}_2$. Birnbaum and Klose [3] give stronger bounds on the parameters but their results cannot be used since F_n and G_n do not satisfy the assumptions of [3].

The bounds on $\sigma^2(\lambda, \epsilon_1, \epsilon_2)$ given in (3.2) can now be obtained. Note that $\min(\epsilon_1, \epsilon_2) \leqq (\epsilon_1 + \epsilon_2)/2 \leqq \max(\epsilon_1, \epsilon_2)$, so that from (3.1) we have $6r\sigma^2(\lambda, \epsilon_1, \epsilon_2) \geqq 1 + 6r\lambda - 3(1 + r)\lambda^2 - 3r(\epsilon_1 + \epsilon_2)$ and using (A4) gives the lower bound of (3.2). From (A5), $\epsilon_1 + \epsilon_2 \geqq 2\lambda - \frac{1}{3}$ and $A = \min(\epsilon_1, \epsilon_2) \geqq \lambda - \frac{1}{6}$ for $\frac{1}{2} \leqq r < 1$. From (A4), $\epsilon_{1,2} \leqq 2\lambda - \lambda^2 - \epsilon_{2,1}$ so that $A = \max(\epsilon_1, \epsilon_2) \leqq 2\lambda - \lambda^2 - \min(\epsilon_1, \epsilon_2) \leqq \lambda - \lambda^2 + \frac{1}{6}$ for $0 < r \leqq \frac{1}{2}$. Using these relations in (3.1) gives the upper bound of (3.2).

REFERENCES

[1] Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

[2] Birnbaum, Z. W. (1956). On a use of the Mann-Whitney statistic. *Proc. Third Berkeley Symp. on Math. Statist. and Prob.* **1** 13–17. Univ. California Press.

[3] Birnbaum, Z. W. and Klose, O. M. (1957). Bounds for the variance of the Mann-Whitney statistic. *Ann. Math. Statist.* **28** 933–945.

[4] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

[5] Gupta, S. D. (1964). Nonparametric classification rules. *Sankhyā* **26** 25–30.

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

[7] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[8] Hudimoto, H. (1964). On a distribution-free two-way classification. *Ann. Inst. Statist. Math. Tokyo* **16** 247–253.

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

[10] Lehmann, E. L. (1963). Robust estimation in analysis of variance. *Ann. Math. Statist.* **34** 957–965.

[11] Loéve, M. (1963). *Probability Theory.* VanNostrand, Princeton.

[12] 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.

[13] SEN, P. K. (1960). On some convergence properties of U-Statistics. *Calcutta Statist. Assoc. Bull.* **10** 1–18.

[14] WHITNEY, D. R. (1951). A bivariate extension of the U-statistic. *Ann. Math. Statist.* **22** 274–282.

[15] WOINSKY, M. N. and KURZ, L. (1967). An efficient nonparametric detector based on a three-sample classification model. *IEEE Trans. Information Theory* **14** 744–751.

[16] WOINSKY, M. N. (1967). Detection of sampled and continuous data imbedded in statistically undefined noise. Ph.D. Thesis, New York Univ.