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

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

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

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

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.