Numerical Tabulations for A Statistic Similar to Student's t*

1. Introduction

In several of his papers (see [6], [7], [8]) John E. Walsh explored tests for the median which are based on order statistics and can be applied under rather general conditions. These papers were written before 1949, at a time when programmed high speed computers were not yet generally available; it is therefore remarkable how much progress he was able to make in the numerical evaluation of these test statistics.

The present paper deals with a statistic similar to that considered by Walsh in [8]. Using contemporary computing equipment it has now become possible to compute tables which make this statistic available for practical use, and the main purpose of this paper is to present such tables.

2. A class of statistics

2.1. Let $X$ be a random variable with the continuous distribution function $F(x) = P(X \leq x)$, and let $X_1, X_2, \ldots, X_{2m+1}$ be a random sample of $X$ of size $2m+1$. We denote by $X(1) \leq X(2) \leq \ldots \leq X(2m+1)$ the corresponding ordered sample.

Let

$$\mu = \inf \{x: F(x) = \frac{1}{2}\} \quad (2.1.1)$$

be the population median of $X$, and

$$V = X_{(m+1)} \quad (2.1.2)$$

the sample median.

For a given integer $r$, $1 \leq r \leq m$, we consider the statistic

$$S_{m,r} = \frac{X_{(m+1)} - \mu}{X_{(m+1+r)} - X_{(m+1-r)}} \quad (2.1.3)$$

* Research supported by the U.S. Office of Naval Research and the National Science Foundation.
This statistic, and a somewhat more general one, was already studied by M.M. Siddiqui in [5] and some of its properties were discussed in [1] and [2]. Its structure is similar to that of the range-midrange statistic proposed by E.S. Pearson in [4] and studied by John E. Walsh in [8], as well as to that of Student's t, in view of the following properties:

a) The numerator is the difference between a location parameter (population median) and an estimate of that parameter (sample median), while the denominator is an estimate (sample interquantile range) of a scale parameter (population interquantile range).

b) The statistic $S_{m,r}$ is invariant under linear transformations and therefore, for given prototype distribution function $F(\cdot)$, the probability distribution of $S_{m,r}$ is independent of location- and scale-parameters, i.e. is the same for all random variables with distribution functions $F(\frac{x-a}{b})$ where $a$ is any real and $b$ any positive constant.

In view of property b) one can choose, for example, for the prototype distribution function

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

and tabulate the probability distributions of $S_{m,r}$ for practically useful values of $1 \leq r \leq m$. Since, for given $r,m$, these probability distributions are the same for all normal random variables, independently of their expectations and variances, they can be used in a manner analogous to that in which Student's t-statistic is used. The statistic $S_{m,r}$ defined by (2.1.3) offers the practical advantage that, contrary to the t-statistic, but similarly to the range-midrange statistic, it can be computed when only very few order statistics are available, even when all the other sample values are "censored" out.

2.2. If the random variable $X$ has probability distribution function $F(x)$ and probability density $f(x) = F'(x)$ then it is not difficult to derive (see [1]) for the probability distribution function of $S_{m,r}$ the expression

$$P\{S_{m,r} \leq \lambda\} = 1 - P(\lambda) \quad (2.2.1)$$

where

$$P(\lambda) = \frac{(2m+1)!}{[(m-r)!r-1]!} \frac{1}{2} \int_{-\infty}^{\infty} \int_{u=\lambda-1}^{\infty} \int_{w=v}^{\infty} f(u) f(v) f(w) du dv dw \quad (2.2.2)$$

$$F(u) [F(v) - F(u)]^{r-1} [F(w) - F(v)]^{r-1} [1 - F(w)]^{m-r} dw du dv$$