ON THE DISTRIBUTION OF WALD'S CLASSIFICATION STATISTIC

By Harman Leon Harter

Michigan State College

Summary. In this paper we shall consider the exact distribution of Wald's classification statistic V in the univariate case, some theoretical approximations in various multivariate cases, and an empirical distribution in a particular multivariate case. We shall also draw some conclusions as to the potential usefulness of the statistic V and the work which remains to be done.

1. Introduction. In many educational and industrial problems it is necessary to classify persons or objects into one of two categories—those fit and those unfit for a particular purpose. In formulating this problem of classification, Wald [1] assumed that for p tests we know the scores of N_1 individuals known to belong to population Π_1 and of N_2 individuals known to belong to population Π_2, along with those of the individual under consideration, a member of the population Π, where it is known a priori that Π is identical with either Π_1 or Π_2. He assumed moreover that the distribution of the test scores of the individuals making up Π_1 and Π_2 are two p-variate normal distributions which have the same covariance matrix, but are independent of each other. In order to classify the individual in question into either Π_1 or Π_2, Wald introduced the statistic V defined by the relation

$$V = \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} t_{i,n+1} t_{j,n+2} \qquad (n = N_1 + N_2 - 2), \tag{1}$$

where

$$\| s^{ij} \| = \| s_{ij} \|^{-1}, s_{ij} = \frac{\sum_{\alpha=1}^{n} t_{i\alpha} t_{j\alpha}}{n}, \tag{2}$$

and where the variates $t_{i\beta}$ (i $= 1, \cdots, p; \beta = 1, \cdots, n + 2$) are normally and independently distributed with unit variance and with expected values

$$E(t_{i\alpha}) = 0 \ (\alpha = 1, \cdots, n), \qquad E(t_{i,n+1}) = \rho_i, \qquad E(t_{i,n+2}) = \zeta_i, \tag{3}$$

where ρ_i and ζ_i are constants.

2. The exact distribution of V when $p = 1$. In the univariate case, the definition (1) reduces to

$$V = s^{11} t_{1,n+1} t_{1,n+2}, \tag{4}$$

where

$$s^{11} = \frac{1}{s_{11}} = \frac{1}{\sum_{\alpha=1}^{n} t_{1\alpha}^2 / n}, \qquad n = N_1 + N_2 - 2.$$

Thus, in the case $p = 1$,

$$V = \frac{t_{1,n+1}\, t_{1n,+2}}{\sum_{\alpha=1}^{n} t_{1\alpha}^2/n} = \frac{xy}{z}, \tag{5}$$

where

$$x = t_{1,n+1}, \qquad y = t_{1,n+2}, \qquad z = \sum_{\alpha=1}^{n} t_{1\alpha}^2/n.$$

In the degenerate case ($\rho_1 = \zeta_1 = 0$), x and y are normally distributed with zero means, so that their probability laws are

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad P(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}. \tag{6}$$

Because of symmetry we have then

$$P(|\, x\,|) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad P(|\, y\,|) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \tag{7}$$

It is well known that $z = \sum_{\alpha=1}^{n} t_{1\alpha}^2/n$ is distributed as χ^2/n with n degrees of freedom, that is, the probability law for z is

$$P(z) = \frac{n^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \frac{z^{\frac{1}{2}n-1} e^{-\frac{1}{2}nz}}{2^{\frac{1}{2}n}}. \tag{8}$$

Now we proceed to find the probability law of $V = \dfrac{|x| \cdot |\, y\,|}{z}$ in a manner similar to that used by Shrivastava [2] in investigating a different statistic. Let $w = \ln |\, V\,| = \ln |\, x\,| + \ln |\, y\,| - \ln z$. Then the characteristic function of w is given by

$$\phi(t) = \int_0^\infty\int_0^\infty\int_0^\infty e^{iwt} P(|\, x\,|)P(|\, y\,|)P(z)\, dx\, dy\, dz. \tag{9}$$

Substituting the values of $P(\,|\, x\,|\,)$, $P(\,|\, y\,|\,)$ and $P(z)$ from (7) and (8), and making use of the independence of x, y and z, we have

$$\phi(t) = \frac{n^{\frac{1}{2}n}}{2^{\frac{1}{2}n-1}\Gamma(\frac{1}{2}n)\pi} \int_0^\infty x^{it} e^{-\frac{1}{2}x^2}\, dx \int_0^\infty y^{it} e^{-\frac{1}{2}y^2}\, dy \int_0^\infty z^{\frac{1}{2}n-1-it} e^{-\frac{1}{2}nz}\, dz. \tag{10}$$

Expressing the integrals in (10) in terms of Gamma functions and simplifying, we find

$$\phi(t) = \frac{n^{it}}{\pi\Gamma(\frac{1}{2}n)} \Gamma(\tfrac{1}{2}n - it) \left[\Gamma\left(\frac{it+1}{2}\right)\right]^2. \tag{11}$$

Upon inserting this result in the Levy inversion formula

$$P(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iwt} \phi(t)\, dt \tag{12}$$

and making the substitution $v = it$, we obtain

$$(13) \qquad P(w) = \frac{n^{it}}{2\pi^2 i\Gamma(\frac{1}{2}n)} \int_{-i\infty}^{+i\infty} e^{-vw}\Gamma(\tfrac{1}{2}n - v)\left[\Gamma\left(\frac{v+1}{2}\right)\right]^2 dv.$$

Using a property of the Gamma function given by Whittaker and Watson [3]

$$(14) \qquad \Gamma(z)\Gamma(1 - z) = \frac{\pi}{\sin \pi z}$$

and letting $z = \frac{1}{2}n - v$, we obtain

$$(15) \qquad \Gamma(\tfrac{1}{2}n - v) = \frac{\pi}{\Gamma(v - \frac{1}{2}n + 1)\sin \pi(\frac{1}{2}n - v)}.$$

Substituting this value of $\Gamma(\frac{1}{2}n - v)$ in (13), and simplifying, we find

$$(16) \qquad P(w) = \frac{1}{\Gamma(\frac{1}{2}n)} \cdot \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} e^{-vw} \frac{n^v\left[\Gamma\left(\frac{v+1}{2}\right)\right]^2}{\sin \pi(\frac{1}{2}n - v)\Gamma(v - \frac{1}{2}n + 1)} dv.$$

We shall now perform a contour integration, using as the contour the imaginary axis plus the semicircle in the right half-plane with center at the origin and infinite radius. It can be shown that for $| n/2e^w | < 1$, and hence for $| V | > n/2$, the integral around the semicircular portion of the contour is zero. Hence, under these conditions, the integral on the right side of (16) is equal to $(-2\pi i)$ times the sum of the residues at all the singular points in the right half-plane. The integrand has simple poles at $v = \frac{1}{2}n, \frac{1}{2}n + 1, \frac{1}{2}n + 2, \cdots$, and no other singularities in the right half-plane. Inserting the actual values of the residues, using the fact that $\cos k\pi = (-1)^k$, for k an integer, and letting $v = j + \frac{1}{2}n$, we find

$$(17) \qquad P(w) = \frac{1}{\pi\Gamma(\frac{1}{2}n)} \sum_{j=0}^{\infty} e^{-(j+\frac{1}{2}n)w} n^{j+\frac{1}{2}n}(-1)^j \frac{\left[\Gamma\left(\frac{2j+n+2}{4}\right)\right]^2}{\Gamma(j+1)}.$$

Replacing e^w by $| V |$ and multiplying by $\frac{dw}{d|V|} = \frac{1}{|V|}$, we obtain the probability law for $| V |$

$$(18) \qquad P(| V |) = \frac{1}{\pi\Gamma(\frac{1}{2}n)} \sum_{j=0}^{\infty} n^{j+\frac{1}{2}n} | V |^{-j-\frac{1}{2}n-1}(-1)^j \frac{\left[\Gamma\left(\frac{2j+n+2}{4}\right)\right]^2}{\Gamma(j+1)}.$$

The infinite series on the right side of (18) converges for precisely those values of $| V |$ for which the integral along the semicircular portion of the path is zero, that is for $| V | > \frac{1}{2}n$. Since the values of x and y are symmetric about zero and uncorrelated, the values of V are also symmetric about zero, and hence $P(V) = \frac{1}{2}P(| V |)$.

To obtain a series for $P(\mid V \mid)$ which converges when $\mid V \mid < \frac{1}{2}n$, it would be necessary to perform a contour integration around the left half-plane, which is considerably more difficult, since the presence of $\left[\Gamma\left(\frac{v+1}{2}\right)\right]^2$ in the integrand of (16) introduces double poles at $v = -1, -3, -5, \cdots$.

If we drop the restriction $\zeta_1 = 0$, but keep $\rho_1 = 0$, V will still be distributed symmetrically about zero, since x is distributed symmetrically about zero and is independent of y. The probability laws for x and z will be the same as in the degenerate case, but $P(\mid V \mid)$ will be different, due to a change in $P(\mid y \mid)$. Since the mean of the distribution of y's is now $\zeta_1 \neq 0$, we have

$$P(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\zeta_1)^2}, \tag{19}$$

which yields

$$P(\mid y \mid) = \frac{1}{\sqrt{2\pi}} [e^{-\frac{1}{2}(y-\zeta_1)^2} + e^{-\frac{1}{2}(-y-\zeta_1)^2}] = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}\zeta_1^2} e^{-\frac{1}{2}y^2} \sum_{r=0}^{\infty} \frac{(y\zeta_1)^{2r}}{(2r)!}. \tag{20}$$

Proceeding in the same manner as for the degenerate case, we find as the characteristic function of $w = \ln \mid V \mid$

$$\phi(t) = \frac{n^{it} e^{-\frac{1}{2}\zeta_1^2}}{\pi\Gamma(\frac{1}{2}n)} \Gamma\left(\frac{it+1}{2}\right) \Gamma(\tfrac{1}{2}n - it) \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{it+1}{2}\right). \tag{21}$$

Again using the Levy inversion formula (12), and letting $v = it$, we have

$$\begin{aligned} P(w) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{2\pi^2 i\Gamma(\frac{1}{2}n)} \int_{-i\infty}^{+i\infty} n^v e^{-vw} \Gamma\left(\frac{v+1}{2}\right) \\ \cdot \Gamma(\tfrac{1}{2}n - v) \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{v+1}{2}\right) dv. \end{aligned} \tag{22}$$

This integral may be evaluated by integrating around the same contour as in the degenerate case. Performing the contour integration and simplifying, we obtain

$$P(w) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{\pi\Gamma(\frac{1}{2}n)} \sum_{v=\frac{1}{2}n}^{\infty} \frac{n^v e^{-vw} \Gamma\left(\frac{v+1}{2}\right)}{\Gamma(v - \frac{1}{2}n + 1)} (-1)^{v-\frac{1}{2}n} \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{v+1}{2}\right). \tag{23}$$

Replacing e^w by $\mid V \mid$ and multiplying by $\frac{dw}{d\mid V \mid} = \frac{1}{\mid V \mid}$, we find

$$P(\mid V \mid) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{\pi\Gamma(\frac{1}{2}n)} \sum_{v=\frac{1}{2}n}^{\infty} \frac{n^v \Gamma\left(\frac{v+1}{2}\right) (-1)^{v-\frac{1}{2}n}}{\mid V \mid^{v+1} \Gamma(v - \frac{1}{2}n + 1)} \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!} \Gamma\left(r + \frac{v+1}{2}\right). \tag{24}$$

Letting $v = j + \frac{1}{2}n$, this may be written in the form

$$(25) \qquad P(|V|) = \frac{e^{-\frac{1}{2}\zeta_1^2}}{\pi\Gamma(\frac{1}{2}n)} \sum_{j=0}^{\infty} \frac{(-1)^j n^{j+\frac{1}{2}n}\Gamma\left(\frac{2j+n+2}{4}\right)}{|V|^{j+\frac{1}{2}n+1}\Gamma(j+1)} \cdot \sum_{r=0}^{\infty} \frac{(2\zeta_1^2)^r}{(2r)!}\Gamma\left(r + \frac{2j+n+2}{4}\right).$$

This expression is valid (since the integral vanishes along the semicircular portion of the contour) and converges for precisely the same values of V as in the degenerate case, that is for $|V| > n/2$.

3. Approximate distributions of V in various multivariate cases. Wald [1] has shown that the distribution of the statistic V is the same as that of the statistic

$$(26) \qquad \bar{V} = -n\frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)},$$

where the joint distribution of m_1, m_2 and m_3 is known. Since m_1, m_2 and m_3 are of the order $1/n$ in the probability sense, the denominator of (26) is near -1 nearly always for sufficiently large n. Accordingly, Wald has suggested that even for moderately large n, V is distributed approximately as nm_3. By integrating out m_1 and m_2 over the domain for which the joint distribution is real and $\geqq 0$, it is possible to find the distribution of m_3, and from it the distribution of nm_3, which is approximately the distribution of V, for sufficiently large n. We restrict ourselves to values of n and p satisfying the relation $1 < p \leqq n$. Four cases have to be considered: (1a) n even, p odd; (1b) n even, p even; (2a) n odd, p even; and (2b) n odd, p odd.

For the degenerate case $\rho_i = \zeta_i = 0$, it can be shown that the joint distribution of m_1, m_2 and m_3 given by Wald [1] reduces to

$$(27) \qquad C[(1 - m_1)(1 - m_2) - m_3^2]^{(n-1-p)/2}[m_1m_2 - m_3^2]^{(p-3)/2}dm_1dm_2\,dm_3,$$

where C is a constant. In integrating out m_1 and m_2, we must be careful to integrate over only the domain for which the joint distribution (27) is real and $\geqq 0$. This requires that the following inequalities hold:

$$(28) \qquad m_1m_2 - m_3^2 \geqq 0, \qquad (1 - m_1)(1 - m_2) - m_3^2 \geqq 0.$$

From these it follows that the limits for m_1 and m_2 are

$$(29) \qquad \frac{m_3^2}{m_2} \leqq m_1 \leqq 1 - \frac{m_3^2}{1 - m_2}, \qquad \frac{1 - \sqrt{1 - 4m_3^2}}{2} \leqq m_2 \leqq \frac{1 + \sqrt{1 - 4m_3^2}}{2}.$$

For Case 1a (n even, p odd), let $p = 3 + 2c$, where c = an integer $\geqq 0$. The distribution function $G_{n,p}(m_3)$ can then be expressed as a double integral, as follows:

$$G_{n,3+2c}(m_3) = C \int_{(1-\sqrt{1-4m_3^2})/2}^{(1+\sqrt{1-4m_3^2})/2} \int_{m_3^2/m_2}^{1-m_3^2/(1-m_2)} [(1-m_1)(1-m_2) - m_3^2]^{(n-4)/2-c} \cdot [m_1 m_2 - m_3^2]^c \, dm_1 dm_2. \tag{30}$$

Expanding repeatedly by the binomial theorem and integrating out m_1, then expanding again and integrating out m_2, we find

$$\begin{aligned} G_{n,3+2c}(m_3) &= C \sum_{j=0}^{(n-4)/2-c} \binom{\frac{n-4}{2} - c}{j} \\ &\cdot \sum_{k=0}^{j} \binom{j}{k} \sum_{q=0}^{c} (-1)^{j+q} \binom{c}{q} \frac{2}{n-2-2j-2q} \\ &[A_{j,k,q}(m_3) + B_{j,k,q}(m_3) - C_{j,k}(m_3) - D_{j,k}(m_3)], \end{aligned} \tag{31}$$

where

$$\begin{aligned} A_{j,k,q}(m_3) &= \sum_{r=0}^{\min[(n-2)/2-j-q,\,(n-4)/2-c-k]} \binom{\frac{n-2}{2} - j - q}{r} m_3^{2(k+q+r)} \\ &\cdot \sum_{t=0}^{(n-4)/2-c-k-r} (-1)^t \binom{\frac{n-4}{2} - c - k - r}{t} \frac{2}{n-2-2k-2q-2r-2t} \\ &\cdot \left[\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r-t} - \left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r-t}\right], \end{aligned} \tag{32}$$

$$\begin{aligned} B_{j,k,q}(m_3) &= \sum_{r'=(n-2)/2-c-k}^{(n-2)/2-j-q} \binom{\frac{n-2}{2} - j - q}{r'} m_3^{2(k+q+r')} \\ &\cdot \left\{ \binom{c-q}{r' + c + k - \frac{n-2}{2}} \ln \frac{1-\sqrt{1-4m_3^2}}{1+\sqrt{1-4m_3^2}} \right. \\ &+ \sum_{\substack{t'=0 \\ t' \neq r'+c+k-(n-2)/2}}^{c-q} (-1)^{t'-r'-c-k+(n-2)/2} \binom{c-q}{t'} \frac{2}{2t'-2r'-2c-2k+n-2} \\ &\left. \cdot \left[\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r'-t'} - \left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{(n-2)/2-k-q-r'-t'}\right] \right\}, \end{aligned} \tag{33}$$

$$\begin{aligned} C_{j,k}(m_3) &= \sum_{\substack{s=0 \\ s \neq j-k}}^{(n-4)/2-c-k} (-1)^s \binom{\frac{n-4}{2} - c - k}{s} \frac{1}{j-k-s} m_3^{n-2+2k-2j} \\ &\cdot \left[\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{j-k-s} - \left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{j-k-s}\right], \end{aligned} \tag{34}$$

$$D_{j,k}(m_3) = (-1)^{j-k+1}\binom{\frac{n-4}{2}-c-k}{j-k} m_3^{n-2+2k-2j} \ln \frac{1-\sqrt{1-4m_3^2}}{1+\sqrt{1-4m_3^2}}, \tag{35}$$

the terms involving natural logarithms having the value zero when $m_3 = 0$. As a numerical example we have, after normalization,

$$\begin{aligned} G_{10,3}(m_3) = \frac{180}{\pi}\Bigg[&(\tfrac{1}{16} + \tfrac{59}{24}m_3^2 + \tfrac{1}{24}m_3^4 + \tfrac{1}{4}m_3^6)\sqrt{1-4m_3^2} \\ &- (m_3^2 + \tfrac{3}{2}m_3^4 - \tfrac{1}{2}m_3^8) \ln \frac{1+\sqrt{1-4m_3^2}}{1-\sqrt{1-4m_3^2}}\Bigg]. \end{aligned} \tag{36}$$

For Case 1b (n even, p even), let $p = 2 + 2c$, where c = an integer $\geqq 0$. The distribution function $G_{n,p}(m_3)$ can then be expressed as a double integral, as follows:

$$\begin{aligned} G_{n,2+2c}(m_3) = C \int_{(1-\sqrt{1-4m_3^2})/2}^{(1+\sqrt{1-4m_3^2})/2} \int_{m_3^2/m_2}^{1-m_3^2/(1-m_2)} &[(1-m_1)(1-m_2) - m_3^2]^{(n-3)/2-c} \\ &\cdot [m_1 m_2 - m_3^2]^{c-\frac{1}{2}}\, dm_1\, dm_2. \end{aligned} \tag{37}$$

This double integration can be performed by the use of certain formulas given by Peirce [4], and after evaluation we have

$$\begin{aligned} G_{n,2+2c}(m_3) &= C\cdot 2\pi(-1)^{(n-2)/2-c} \\ &\cdot \frac{(2c-1)(2c-3)\cdots 1(n-2c-3)(n-2c-5)\cdots 1}{(n-2)(n-4)\cdots 2} \\ &\cdot \Bigg[\sum_{j=0}^{(n-2)/2} \sum_{\substack{k=0 \\ (j-k)\leqq c}}^{(n-2)/2-j} (-1)^{(n-2)/2+k-c-j}\binom{\frac{n-2}{2}}{j}\binom{\frac{n-2}{2}-j}{k} A'_{j,k}(m_3) \\ &+ \sum_{j=0}^{(n-2)/2} \sum_{\substack{k=0 \\ (j-k)>c}}^{(n-2)/2-j} (-1)^{(n-2)/2+k-c-j}\binom{\frac{n-2}{2}}{j}\binom{\frac{n-2}{2}-j}{k} B'_{j,k}(m_3)\Bigg], \end{aligned} \tag{38}$$

where

$$
\begin{aligned}
A'_{j,k}(m_3) = m_3^{2j} \Bigg[m_3 \sum_{q=0}^{c} (-1)^q & \\
\cdot \frac{(2m-2c+3)(2m-2c+1)\cdots(2m-2c-2q+5)}{(2c-1)(2c-3)\cdots(2c-2q-1)} & \\
\cdot \left\{ \frac{\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{m+\frac{1}{2}}}{\left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{c-q}} - \frac{\left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{m+\frac{1}{2}}}{\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{c-q}} \right\} & \\
+ (-1)^c \frac{(2m-2c+3)(2m-2c+1)\cdots(2m+3)}{(2c-1)(2c-3)\cdots 1} & \\
\cdot \Bigg((-1)^{m+\frac{1}{2}} \frac{2m(2m-2)\cdots 1}{(2m+1)(2m-1)\cdots 2} \left(-\sin^{-1}\sqrt{1-4m_3^2}\right) & \\
- m_3 \sum_{r=0}^{m-\frac{1}{2}} \frac{2m(2m-2)\cdots(2m-2r+2)}{(2m+1)(2m-1)\cdots(2m-2r+1)} & \\
\cdot \left\{ \left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{m-r-\frac{1}{2}} - \left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{m-r-\frac{1}{2}} \right\} \Bigg) \Bigg], & \qquad (39)
\end{aligned}
$$

$$
\begin{aligned}
B'_{j,k}(m_3) = m_3^{2j} \Bigg[m_3 \sum_{q'=0}^{c} & \\
\cdot \frac{(2m'+2c-3)(2m'+2c-5)\cdots(2m'+2c-2q'-1)}{(2c-1)(2c-3)\cdots(2c-2q'-1)} & \\
\cdot \left\{ \frac{\left(\frac{2}{1+\sqrt{1-4m_3^2}}\right)^{m'-\frac{1}{2}}}{\left(\frac{1-\sqrt{1-4m_3^2}}{2}\right)^{c-q'}} - \frac{\left(\frac{2}{1-\sqrt{1-4m_3^2}}\right)^{m'-\frac{1}{2}}}{\left(\frac{1+\sqrt{1-4m_3^2}}{2}\right)^{c-q'}} \right\} & \qquad (40) \\
+ \frac{(2m'+2c-3)(2m'+2c-5)\cdots(2m'-3)}{(2c-1)(2c-3)\cdots 1} & \\
\cdot m_3 \sum_{r'=0}^{m'-\frac{3}{2}} (-1)^{r'+1} \frac{(2m'-3)(2m'-5)\cdots(2m'-2r'-1)}{(2m'-2)(2m'-4)\cdots(2m'-2r'-2)} & \\
\cdot \left\{ \left(\frac{2}{1+\sqrt{1-4m_3^2}}\right)^{m'-r'-\frac{1}{2}} - \left(\frac{2}{1-\sqrt{1-4m_3^2}}\right)^{m'-r'-\frac{1}{2}} \right\} \Bigg], &
\end{aligned}
$$

$$
m = k + c - j - \tfrac{1}{2}, \qquad m' = j - k - c + \tfrac{1}{2}. \qquad (41)
$$

As a numerical illustration we have, after normalization,

$$
(42)\quad
\begin{aligned}
G_{10,2}(m_3) &= \left(\frac{55125}{16384} + \frac{23625}{256} m_3^2 + \frac{4725}{64} m_3^4\right) \sin^{-1} \sqrt{1 - 4m_3^2} \\
&\quad - \left(\frac{313515}{8192} + \frac{99825}{1024} m_3^2 - \frac{465}{64} m_3^4 - \frac{45}{8} m_3^8\right) | m_3 | \sqrt{1 - 4m} .
\end{aligned}
$$

In Cases 2a and 2b, infinite series of elliptic integrals occur, and it appears that approximate integration is the best than can be done.

The author plans a later paper on the distribution for the nondegenerate case $\rho_i = 0, \zeta_i \neq 0$.

For small values of n, Wald's approximation nm_3 is not applicable. One can obtain a fair approximation by replacing $1/[m_3^2 - (1 - m_1)(1 - m_2)]$ in (26) by its average with respect to m_1 and m_2 over the domain, taking account of the joint distribution function (27). This yields

$$
(43)\qquad V \doteq \frac{C_{n,p}}{C_{n-2,p}} nm_3 \frac{G_{n-2,p}(m_3)}{G_{n,p}(m_3)},
$$

where $C_{n-2,p}$ and $C_{n,p}$ are the constants in the joint distribution of m_1, m_2 and m_3 for the values of n and p involved. The approximation (43), while rather crude, is better than Wald's nm_3 for small values of n, and asymptotically equivalent to it as $n \rightarrow \infty$.

4. An empirical distribution of V. A sampling experiment was performed in order to obtain an empirical distribution of 1000 values of V for $n = 10$, $p = 3$, $\rho_i = \zeta_i = 0$. Ten thousand wooden beads were stamped with two digit numbers whose distribution approximates as nearly as possible that of a normal population with mean 50 and standard deviation 10. One thousand sets of values $x_{i\alpha}(i = 1, 2, 3; \alpha = 1, 2, \cdots, 12)$ were obtained by sampling with replacement from this population. The values $x_{i\alpha}$ were expressed in standard units $t_{i\alpha}$, using

$$
(44)\qquad t_{i\alpha} = \frac{x_{i\alpha} - 50}{10}.
$$

From the standard variables $t_{i\alpha}$, one thousand values of V were calculated by means of (1) and (2), using IBM equipment. The resulting empirical distribution is given in Table 1. This distribution was compared with the theoretical approximation (43), which is, for $n = 10$, $p = 3$, $\rho_i = \zeta_i = 0$

$$
(45)\qquad V \doteq \frac{150}{7} m_3 \frac{G_{8,3}(m_3)}{G_{10,3}(m_3)}.
$$

The approximation fits the observed distribution fairly well for the central classes, but underestimates the frequencies of large values of $| V |$ quite badly.

5. Conclusions. The statistic V is potentially very useful, but much work remains to be done in obtaining the necessary information about its distribution, especially in the small sampling case, and tabulating the associated probabilities. Even in the univariate case, where the exact distribution is known, the amount of labor involved in determining probabilities is very great and a simple approximation is needed, unless a high speed computing device is available. For the multivariate small sampling case, only a crude approximation to the distribution of V is available, and the exact distribution or a better approximation is needed.

TABLE 1

Frequency distribution of 1000 empirical values of V for $n = 10$, $p = 3$, $\rho_i = \zeta_i = 0$ (Class marks integers)

Class mark	Frequency f	Class mark	Frequency f	Class mark	Frequency f
76	1	12	3	−8	15
...		11	3	−9	12
44	1	10	3	−10	6
...		9	6	−11	3
39	1	8	10	−12	2
...		7	16	−13	2
30	1	6	11	−14	3
29	1	5	15	−15	4
...		4	33	...	
24	1	3	54	−18	2
23	1	2	85	...	
...		1	140	−23	1
20	1	0	181	−24	2
19	2	−1	134	...	
18	1	−2	101	−28	1
...		−3	52	−29	1
16	1	−4	26	...	
15	2	−5	17	−36	1
14	1	−6	23	−37	1
13	4	−7	12		

$$\overline{V} = -.0700, \qquad \sigma_V = 5.938$$

The author wishes to express his sincere thanks to the Office of Naval Research for the grant which made this work possible, and to Professor Carl F. Kossack of Purdue University for his helpful suggestions and patient guidance.

REFERENCES

[1] A. Wald, "On a statistical problem arising in the classification of an individual into one of two groups," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 145–162.

[2] M. P. Shrivastava, "On the D^2-statistic," *Bull. Calcutta Math. Soc.*, Vol. 33 (1941), pp. 71–86.

[3] E. T. Whittaker and G. N. Watson, *Modern Analysis*, 4th ed., Cambridge University Press, London, 1940, p. 239.

[4] B. O. Peirce, *A Short Table of Integrals*, Ginn and Co., Boston, 1929, pp. 18–19, Formulas 113, 118, 119, 120, 121, 123.