THE DISTRIBUTION OF WALD'S CLASSIFICATION STATISTIC WHEN THE DISPERSION MATRIX IS KNOWN

By S. JOHN

Indian Statistical Institute, Calcutta

SUMMARY. The discriminant function can be used for classifying an individual as belonging to one or other of two populations provided we know the parameters characterising the two populations. Wald suggested the use of a certain statistic in situations where such knowledge is absent. The exact distribution of this statistic in the case where the dispersion matrix is known, is obtained in this paper.

1. Introduction

Consider the problem of classifying a given collection of individuals as belonging to one or other of two populations P_1, P_2 based on measurements carried out on each individual with respect to p characteristics $x_1, x_2, \ldots, x_p$. Let us assume that among the individuals belonging to P_1, $x_1, x_2, \ldots, x_p$ follow a multivariate normal distribution with means $\mu_1, \mu_2, \ldots, \mu_p$ and variance-covariance matrix Σ and that among individuals belonging to P_2, $x_1, x_2, \ldots, x_p$ follow a multivariate normal distribution with the same variance-covariance matrix Σ but with means $\nu_1, \nu_2, \ldots, \nu_p$. According to a method originally suggested by Fisher (1936) an individual with measurements $y_1, y_2, \ldots, y_p$ is assigned to P_1 or P_2 according as

$$(\mu-\nu)\Sigma^{-1}y' \gtrless \tfrac{1}{2}(\mu-\nu)\Sigma^{-1}(\mu+\nu)'$$

where $\mu = (\mu_1, \ldots, \mu_p)$ and $\nu = (\nu_1, \nu_2, \ldots, \nu_p)$. It will be noted that Fisher's method requires a knowledge of μ, ν and Σ. To remedy this Wald (1944) considered the use of the statistic

$$u = (\bar{x}^{(1)}-\bar{x}^{(2)})S^{-1}y',$$

the individuals being classified as belonging to P_1 when $U < d$, where d is so determined that the critical region is of the desired size; $\bar{x}^{(1)}$ is the vector of means determined from measurements on a sample of n_1 individuals known to belong to P_1; $\bar{x}^{(2)}$ is the vector of means determined from measurements on a random sample of n_2 individuals known to belong to P_2, and S is the variance-covariance matrix estimated from the pooled corrected sums of squares and products from the two samples. The distribution of U turns out to be complicated and Wald has not given an explicit expression for it.

This paper considers the distribution of U in the case when the variance-covariance matrix is known; i.e. the distribution of

$$V = (\bar{x}^{(1)}-\bar{x}^{(2)})\Sigma^{-1}y' \quad \ldots \quad (1.1)$$

where $\bar{x}^{(1)}$, $\bar{x}^{(2)}$, Σ and y have the same meanings as before.

2. Reduction of the problem

Looking at the problem from a more general point of view, it will be seen that the statistic whose distribution is in question is

$$z = t\,\Sigma^{-1}w' \qquad \dots \quad (2.1)$$

where $t = (t_1, t_2, \dots, t_p)$ is a vector of p normal variates following a multivariate normal distribution with variance-covariance matrix Σ and means $a = (a_1, a_2, \dots, a_p)$, and where $w = (w_1, w_2, \dots, w_p)$ is a vector of p normal variates independent of t and following a multivariate normal distribution with variance-covariance matrix Σ and mean $b = (b_1, b_2, \dots b_p)$. The statistic V reduces to z when multiplied by $(n_1n_2/n_1+n_2)^{\frac{1}{2}}$.

In the further reduction of the problem we require the following.

Lemma: *The statistic z is invariant under non-singular transformations of t and w provided the two transformations are the same.*

Proof: Let $x = tC$ and $y = wC$ where C is a $(p\times p)$ non-singular matrix. Let Σ_0 denote the variance-covariance matrix of x (which is the same as the variance-covariance matrix of y). Then

$$\Sigma_0 = C'\Sigma C \text{ and } x\Sigma_0^{-1}y' = tC(C^{-1}\Sigma^{-1}C'^{-1})C'w' = t\,\Sigma^{-1}w'.$$

This proves the lemma.

Since, when Σ is positive definite, there always exists a non-singular matrix C such that

$$C'\Sigma C = 1$$

this lemma reduces our problem to a consideration of the statistic

$$T = u_1v_1+u_2v_2+\dots+u_pv_p \qquad \dots \quad (2.2)$$

where $u_1, v_1, u_2, v_2, \dots, u_p, v_p$ are independent normal variates with unit variance. Without loss of generality we may assume that

$$E(u_1) = E(u_2) = \dots = E(u_p) = 0. \qquad \dots \quad (2.3)$$

Let

$$E(v_i) = m_i (i = 1, 2, \dots, p). \qquad \dots \quad (2.4)$$

The expectation and variance of T are easily found.

$$E(T) = \sum_{i=1}^{p} E(u_i)E(v_i) = 0$$

since, by our assumption $\quad E(u_i) = 0 \quad (i = 1, 2, \dots, p)$

$$V(T) = \sum_{i=1}^{p} V(u_iv_i) = \sum_{i=1}^{p} E(u_i^2v_i^2) = \sum_{i=1}^{p} (1+m_i^2) = p+\sum_{i=1}^{p} m_i^2.$$

To find the distribution of T we adopt the method of characteristic functions.

3. The characteristic function of T

The characteristic function of $u_i v_i$ is

$$\varphi_{u_i v_i} = E(e^{i\theta u_i v_i})$$

$$= E(e^{-\frac{1}{2} v_i^2 \theta^2})$$

$$= \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} v_i^2 \theta^2 - \frac{1}{2}(v_i - m_i)^2} dv_i$$

$$= (1+\theta^2)^{-\frac{1}{2}} \exp\left\{-\frac{m_i^2 \theta^2}{2(1+\theta^2)}\right\}. \quad \dots \quad (3.1)$$

and hence we get for the characteristic function of T

$$\varphi(\theta) = \varphi_T(\theta) = (1+\theta^2)^{-p/2} \exp\left\{-\frac{m\theta^2}{1+\theta^2}\right\}$$

where $m = \frac{1}{2} \sum_{i=1}^{p} m_i^2$.

If we denote by $p(T)$ the density function of T, we then have, by inversion,

$$p(T) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\theta T} \varphi(\theta) d\theta. \quad \dots \quad (3.2)$$

For the purpose of inversion we distinguish two cases viz (1) p even and (2) p odd.

Inversion: *Case* (*i*), $p = 2n$.

In this case,

$$\varphi(\theta) = (1+\theta^2)^{-n} \exp\left\{-\frac{m\theta^2}{1+\theta^2}\right\}$$

$$= e^{-m}\left[\frac{1}{(1+\theta^2)^n} + \frac{1}{1!}\frac{m}{(1+\theta^2)^{n+1}} + \frac{1}{2!}\frac{m^2}{(1+\theta^2)^{n+2}} + \dots\right]. \quad \dots \quad (3.3)$$

Therefore,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\theta T}\varphi(\theta)d\theta = \frac{e^{-m}}{2\pi}\int_{-\infty}^{+\infty}\left[\frac{e^{-i\theta T}}{(1+\theta^2)^n} + \frac{m}{1!}\cdot\frac{e^{-i\theta T}}{(1+\theta^2)^{n+1}} + \dots\right]d\theta. \quad \dots \quad (3.4)$$

The series occurring as integrand in (3.4) being uniformly convergent, permits term by term integration and we may write,

$$p(T) = e^{-m} \sum_{r=0}^{\infty} \frac{m^r}{r!} g_{n+r}(T) \qquad \dots \quad (3.5)$$

where

$$g_r(T) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-i\theta T}}{(1+\theta^2)^r} d\theta$$

$$= \frac{1}{2^{2r-1}(r-1)!} \cdot e^{-|T|} \left\{ (2|T|)^{r-1} + \frac{1}{1!} r(r-1)(2|T|)^{r-2} + \frac{1}{2!} (r+1)r(r-1)(r-2)(2|T|)^{r-3} + \right.$$

$$\left. + \dots + \frac{(2r-2)!}{(r-1)!} \right\}. \qquad \dots \quad (3.6)$$

(The last step is obtained by contour integration).

Case (*ii*), $p = 2n+1$.

The characteristic function of $Z = xy$ where x and y are independent normal variates with unit variance and means zero and b respectively would be

$$\phi_{xy}(\theta) = (1+\theta^2)^{-1/2} \exp \left\{ -a \frac{\theta^2}{1+\theta^2} \right\} \qquad \dots \quad (3.7)$$

where $a = \frac{1}{2}b^2$. Also, the density function of Z is

$$\frac{e^{-a}}{\pi} \left[K_0 + \frac{2|z|}{2!} K_1 a + \frac{(2z)^2}{4!} K_2 a^2 + \frac{(2|z|)^3}{6!} K_3 a^3 + \dots \right] \qquad \dots \quad (3.8)$$

where $$K_\nu = K_\nu(z) = \frac{1}{2} \left(\frac{z}{2} \right)^\nu \int_0^\infty e^{-t-\frac{z^2}{4t}} \frac{dt}{t^{\nu+1}} \quad \text{[see Craig (1936)].} \qquad \dots \quad (3.9)$$

Using the inversion theorem for characteristic functions we get from (3.7) and (3.8)

$$\frac{e^{-a}}{2\pi} \int_{-\infty}^{+\infty} e^{-i\theta z} \frac{e^{a/(1+\theta^2)}}{(1+\theta^2)^{\frac{1}{2}}} d\theta = \frac{e^{-a}}{\pi} \left[K_0 + \frac{K_1}{1!} |2z| a + \frac{K_2}{4!} (2z)^2 a^2 + \dots \right].$$

Therefore,

$$\int_{-\infty}^{+\infty} e^{-i\theta z} \frac{e^{a/(1+\theta^2)}}{(1+\theta^2)^1} d\theta = 2 \left[K_0 + \frac{K_1}{2!} |2z| a + \frac{k_2}{4!} (2z)^2 a^2 + \dots \right] \quad \dots \quad (3.10)$$

$$= \psi(a, z) \quad \text{(say).}$$

Differentiating both sides of (3.10) n times with respect to a, we get

$$\int_{-\infty}^{+\infty} e^{-i\theta z} \frac{e^{a/(1+\theta^2)}}{(1+\theta^2)^1} d\theta = \frac{\partial^n}{\partial a^n} \psi(a, z). \quad \dots \quad (3.11)$$

[We note that differentiation within the integral sign in (3.11) is permissible since

$$\int_{-\infty}^{+\infty} e^{-i\theta z} \frac{e^{-a/(1+\theta^2)}}{(1+\theta^2)^{n+1}} d\theta$$

is uniformly convergent]. Since the characteristic function of T is

$$(1+\theta^2)^{-(n+1)} \exp\left\{ -\frac{m\theta^2}{1+\theta^2} \right\} = e^{-m} (1+\theta^2)^{-(n+1)} \exp\{m(1+\theta^2)^{-1}\}$$

it follows from (3.11) that the density function of T is

$$p(T) = \frac{e^{-m}}{2\pi} \left[\frac{\partial^n}{\partial a^n} \psi(a, T) \right]_{a=m} \quad \dots \quad (3.12)$$

Note : The derivatives of $\psi(a, T)$ required in (3.12) can be obtained from the relation

$$\psi(a, T) = 2 \left[K_0 + \frac{2|T|}{2!} K_1 a + \frac{(2T)^2}{4!} K_2 a^2 + \dots \right]. \quad \dots \quad (3.13)$$

For any given value of T, the series in (3.13) can be regarded as a power series in 'a'. The series is easily seen to be convergent for all values of 'a' from the ratio ρ_ν of the $(\nu+1)$-th term to the ν-th term

$$\rho_\nu = \frac{|T|}{\nu(2\nu-1)} \frac{K_\nu}{K_{\nu-1}} a = \frac{c_\nu a}{2\nu-1} \quad \text{(say).}$$

Craig in his paper referred to above, proves that $|c_\nu| < 3$ for all sufficiently large values of ν. Therefore $\rho_\nu \to 0$ as $\nu \to \infty$. Thus the power series in (3.13) is convergent for all values of 'a'. Therefore the derivatives of $\psi(a, T)$ *w.r.t.* a can be obtained by term by term differentiation of

$$2\left[K_0+\frac{2|T|}{2!}K_1 a+\frac{(2T)^2}{4!}K_2 a^2+\ldots\right].$$

Tables of $K_\nu(T)$ can be found in Watson's book 'Theory of Bessel Functions'.

2. Some approximations to the distribution of T has been considered. Details will be given elsewhere.

REFERENCES

CRAIG, C. C. (1936): On the frequency function of xy. *Ann. Math. Stat.*, 7, 1.

FISHER, R. A. (1936): The use of multiple measurements in taxonomic problems. *Ann. Eugen*, 7, 179.

Wald, A. (1944): On a statistical problem arising in the classification of an individual into one of two groups. *Ann. Math. Stat.*, 15, 145.

WATSON, G. N. (1944): *Theory of Bessel Function*, Cambridge University Press.

Paper received : October, 1958.

Revised : April, 1959.