The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1491–1499

THE CLASSIFICATION STATISTIC W^* IN COVARIATE DISCRIMINANT ANALYSIS

BY AHMED ZOGO MEMON[1] AND MASASHI OKAMOTO[2]

Iowa State University

1. Introduction. Consider the problem of classification of a multivariate observation into one of two normal populations Π_1 and Π_2, where in addition to the knowledge of a discriminator x information is available on a covariate y whose mean is known to be the same in both populations.

Let

$$\begin{pmatrix} x_{i1} \\ y_{i1} \end{pmatrix}, \cdots, \begin{pmatrix} x_{iN_i} \\ y_{iN_i} \end{pmatrix}, \ i = 1, 2,$$

be two random samples drawn independently from Π_i: $N((^{\mu_i}_{\nu}), \Sigma)$, where $(x'_{i\alpha}, y'_{i\alpha}) = (x_{i\alpha 1}, \cdots, x_{i\alpha p}, y_{i\alpha p+1}, \cdots, y_{i\alpha p+q})$, $\alpha = 1, \cdots, N_i$ and the covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is positive definite. Suppose $\hat{B}$ is a sample estimate of the regression matrix B of the discriminator x on the covariate y. Although the covariate has no discriminating power by itself, Cochran and Bliss (1948) and Cochran (1964) still propose to utilize the additional information by replacing x by $x^* = x - \hat{B}y$ in the Anderson discriminant function W. Allowing for the possibility that we have samples from some more multivariate normal populations having the same covariance matrix Σ, the modified criterion is given by

$$W^* = [x^* - \tfrac{1}{2}(\bar{x}_1^* + \bar{x}_2^*)]'(S_{11} - S_{12}S_{22}^{-1}S_{21})^{-1}(\bar{x}_1^* - \bar{x}_2^*), \tag{1.1}$$

where $\bar{x}_i^* = \bar{x}_i - \hat{B}\bar{y}_i$, $i = 1, 2$; $\hat{B} = S_{12}S_{22}^{-1}$, $\bar{x}_i$ and $\bar{y}_i$ denote the sample means and finally

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = [s_{ij}]$$

is the best unbiased estimator of Σ so that nS is distributed according to $W(n, \Sigma)$, a Wishart distribution with n degrees of freedom and covariance matrix Σ. The classification procedure is to assign the given observation to the population Π_1 or Π_2 according as W^* takes a positive or negative value. The Mahalanobis distance between Π_1 and Π_2 is

$$D^2 = (\mu_1 - \mu_2)'(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}(\mu_1 - \mu_2), \tag{1.2}$$

Received September 13, 1968; revised January 1, 1970.

[1] Now at the West Pakistan Agricultural University, Pakistan.

[2] Now at Osaka University, Japan.

and it is easily seen (see, e.g., [1] page 139) that as N_1, N_2, $n \to \infty$ the limiting distribution of W^* is $N(\frac{1}{2}D^2, D^2)$ or $N(-\frac{1}{2}D^2, D^2)$ according as $\binom{x}{y}$ comes from Π_1 or Π_2.

The purpose of this paper is to obtain an asymptotic expansion of the distribution of W^* with respect to N_1^{-1}, N_2^{-1} and n^{-1} as well as the associated probabilities of misclassification of both kinds, which will be more useful than the limiting distribution for moderate sizes of N_1, N_2 and n in providing a better approximation to the true distribution. Since in the special case when $q = 0$ the statistic W^* is reduced to the ordinary Anderson statistic W, this paper is an extension of the paper [5] by one of the authors.

2. The main result. We shall state here the main theorem and its corollaries, leaving the proof of the theorem to Section 4. We shall also give some discussion on the three classification procedures that may be used in the existing situation.

THEOREM. *If $D > 0$, then an asymptotic expansion of the distribution function of $D^{-1}(W^* - \frac{1}{2}D^2)$ when $\binom{x}{y}$ comes from Π_1 is given by*

$$(2.1)\quad F(z; D) = \{1 + \textstyle\sum_{i=1}^{3} L_i^*(d; D) + \frac{1}{2}[\sum_{i=1}^{3} L_i^*(d; D)]^2 + \sum_{i\leq j=1}^{3} Q_{ij}^*(d; D) + 0_3\}\Phi(z),$$

where $d = d/dz$, $\Phi(z)$ is the cdf *of $N(0, 1)$, 0_3 is the third order term with respect to N_1^{-1}, N_2^{-1} and n^{-1} and*

$$L_i^*(d; D) = L_i(p), \qquad Q_{ij}^*(d; D) = Q_{ij}(p) \qquad \text{for} \quad i, j = 1, 2,$$

$$(2.2)\quad L_3^*(d; D) = L_3(p+q), \qquad Q_{33}^*(d; D) = Q_{33}(p+q),$$

$$Q_{13}^*(d; D) = Q_{13}(p) + (2N_1 nD^2)^{-1}q[7d^4 + 2p(2d^2 + Dd)],$$

$$Q_{23}^*(d; D) = Q_{23}(p) + (2N_2 nD^2)^{-1}q[(7d^2 - 10Dd + 3D^2)d^2 + 2p(2D^2 - Dd)],$$

$L_i(p)$ and $Q_{ij}(p)$ denoting $L_i(d; D)$ and $Q_{ij}(d; D)$ defined in Okamoto [5], *page* 1287, *and* [6], *which depend on the dimensionality p.*

COROLLARY 2.1. *When W^* is used as the classification criterion, the probability of misclassifying an observation into the population Π_2 when it comes in fact from the population Π_1, is given by evaluating the above distribution function at $z = -\frac{1}{2}D$.*

Since we obtain $-W^*$ by interchanging $\bar{x}_1{}^*$ and $\bar{x}_2{}^*$ in W^*, we have the following result.

COROLLARY 2.2. *The interchange of N_1 and N_2 in the probability of misclassification evaluated in Corollary* 2.1 *gives the other kind of probability of misclassification.*

In a situation as considered in this paper, one may utilize information on (i) discriminator x only, or (ii) whole variate $\binom{x}{y}$, or (iii) residual x^*, to get the discriminant functions

$$W_1 = (x - \tfrac{1}{2}(\bar{x}_1 + \bar{x}_2))' S_{11}^{-1}(\bar{x}_1 - \bar{x}_2),$$

$$W_2 = \begin{pmatrix} x - \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \\ y - \frac{1}{2}(\bar{y}_1 + \bar{y}_2) \end{pmatrix}' S^{-1} \begin{pmatrix} \bar{x}_1 - \bar{x}_2 \\ \bar{y}_1 - \bar{y}_2 \end{pmatrix}$$

and W^*, respectively. The question that naturally arises then is which one of these three classification procedures is the best in the sense of minimization of probabilities of misclassification.

Let D_k^2, $L_i^{(k)}$, $Q_{ij}^{(k)}$ $(i, j, k = 1, 2, 3)$ denote the Mahalanobis distance, the linear and quadratic terms appearing in the asymptotic expansion of probability of misclassification corresponding to the kth procedure. Then it is easily seen that

$$\begin{aligned}
&D_1^2 \leqq D_2^2 = D_3^2, \\
&L_i^{(1)} = L_i^{(3)}, \qquad Q_{ij}^{(1)} = Q_{ij}^{(3)}, \qquad \text{for} \quad i, j = 1, 2 \\
&L_3^{(2)} = L_3^{(3)}, \qquad Q_{33}^{(2)} = Q_{33}^{(3)}, \\
&Q_{i3}^{(1)} \lesseqgtr Q_{i3}^{(3)} \lesseqgtr Q_{i3}^{(2)} \qquad \text{for} \quad i = 1, 2.
\end{aligned}$$

Thus, the third (Cochran and Bliss) procedure has, instead of the nominal dimensionality p, the effective dimensionality p for the terms L_i and Q_{ij} $(i, j = 1, 2)$, $p+q$ for the terms L_3 and Q_{33}, and some values between p and $p+q$ for Q_{12} and Q_{13}. Now we infer from Table 2 of [5] that the probability of misclassification increases, as p increases or D decreases. It, therefore, follows that the third procedure is preferable to the second, while its superiority over the first depends on the balance of the increased Mahalanobis distance and the increased effective dimensionality, which is not treated in this paper.

The asymptotic efficiency e of the third procedure relative to the second may be defined as the ratio of the sample sizes for each procedure which yield the same probability of misclassification. Let (N_1, N_2, n) and (N_1^*, N_2^*, n^*) be numbers of degrees of freedom for Procedures 3 and 2, respectively. The principal terms in the asymptotic expansion of the probabilities of misclassification are the same for both procedures, in fact $\Phi(-\frac{1}{2}D)$. Therefore, if all numbers of degrees of freedom are large, then the probabilities of misclassification when $\binom{x}{y}$ comes from Π_1 are determined by the linear terms, which are given by Corollary 2.1 as

$$N_1^{-1} a_1(p) + N_2^{-1} a_2(p) + n^{-1} a_3(p+q) \tag{2.3}$$ and

$$N_1^{*-1} a_1(p+q) + N_2^{*-1} a_2(p+q) + n^{*-1} a_3(p+q) \tag{2.4}$$

for Procedures 3 and 2, respectively, where the three functions a_1, a_2 and a_3 can be expressed (see Corollary 2 of [5]) as

$$\begin{aligned}
a_1(p) &= (2D^2)^{-1}(d_0^4 + 3pd_0^2), \\
a_2(p) &= (2D^2)^{-1}(d_0^4 - (p-4)\,d_0^2), \\
a_3(p) &= \tfrac{1}{2}(p-1)\,d_0^2
\end{aligned} \tag{2.5}$$

with $d_0^i = d^i\Phi(z)/dz^i\,|_{z=-\frac{1}{2}D}$. Now $N_1^*/N_1 = N_2^*/N_2 = e$ by definition and, if all information about Σ comes from the two samples from Π_1 and Π_2, it follows that $n^*/n = (N_1^* + N_2^* - 2)/(N_1 + N_2 - 2) \doteqdot e$. Equating (2.3) and (2.4), we find that

$$e = \frac{N_1^{-1} a_1(p+q) + N_2^{-1} a_2(p+q) + n^{-1} a_3(p+q)}{N_1^{-1} a_1(p) + N_2^{-1} a_2(p) + n^{-1} a_3(p+q)}. \tag{2.6}$$

In the case $N_1 = N_2$ ($= N$, say) and hence $n = N_1+N_2-2 \doteqdot 2N$, the expression (2.6) is reduced to

$$e = 1+4q/[(p+q)D^2+4(p-1)] \tag{2.7}$$

in view of (2.5). Table 1 shows values of e by the formula (2.7) for some combinations of values of p, q and D.

TABLE 1
Asymptotic efficiency of the Cochran and Bliss discriminant procedure relative to Procedure 2

		p							
q	D	1	2	3	4	5	6	8	10
	1	3.00	1.57	1.33	1.24	1.18	1.15	1.11	1.09
1	2	1.50	1.25	1.17	1.13	1.10	1.08	1.06	1.05
	4	1.13	1.08	1.06	1.04	1.04	1.03	1.02	1.02
	1	3.67	2.00	1.62	1.44	1.35	1.29	1.21	1.17
2	2	1.67	1.40	1.29	1.22	1.18	1.15	1.12	1.10
	4	1.17	1.12	1.09	1.07	1.06	1.05	1.04	1.04
	1	4.00	2.33	1.86	1.63	1.50	1.41	1.31	1.24
3	2	1.75	1.50	1.38	1.30	1.25	1.21	1.17	1.14
	4	1.19	1.14	1.12	1.10	1.08	1.07	1.06	1.05

3. Lemmas. Let

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

be a $(p+q)\times(p+q)$ nonsingular and symmetric matrix and $\partial_{rs} = \partial_{sr} = \frac{1}{2}(1+\delta_{rs})\,\partial/\partial\sigma_{rs}$ for $r \leqq s$ a differential operator, where δ_{rs} denotes Kronecker's delta. For any function of this matrix we denote by the symbol $(\ )_{rs}$ an effect of ∂_{rs} on the function and by $]_0$ the value of the function at $\Sigma = I$. The meaning of $(\ )_{rs,tu}$, etc., will be obvious. Write

$$\begin{aligned} \Sigma_* &= [\sigma_*^{ij}] = \Sigma_{11.2}^{-1}, \\ \Sigma_{**} &= [\sigma_{**}^{ij}] = \Sigma_*(I+BB')\Sigma_*, \end{aligned} \tag{3.1}$$

where $\Sigma_{11.2} = \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $B = [\beta_{ij}] = \Sigma_{12}\Sigma_{22}^{-1}$.

LEMMA 3.1.

$$\begin{aligned} (\sigma_*^{ij})_{rs} &= -\tfrac{1}{2}(\sigma_*^{ir}\sigma_*^{sj}+\sigma_*^{is}\sigma_*^{rj}) && \textit{if}\quad r, s \leqq p, \\ &= \tfrac{1}{2}\textstyle\sum_{k=1}^{p}\beta_{k,s-p}(\sigma_*^{ir}\sigma_*^{kj}+\sigma_*^{ik}\sigma_*^{rj}) && \textit{if}\quad r \leqq p, s > p, \\ &= -\tfrac{1}{2}\textstyle\sum_{k,l=1}^{p}\beta_{k,r-p}\beta_{l,s-p}(\sigma_*^{ik}\sigma_*^{lj}+\sigma_*^{il}\sigma_*^{kj}) && \textit{if}\quad r, s > p. \end{aligned} \tag{3.2}$$

$$(\sigma_*^{ij})_{rs}]_0 = \tfrac{1}{2}(\sigma_{**}^{ij})_{rs}]_0 = -\tfrac{1}{2}(\delta_{ir}\delta_{sj}+\delta_{is}\delta_{rj}) \qquad \textit{for any}\quad i, j, r, s. \tag{3.3}$$

PROOF. From the identity $\Sigma_* \Sigma_{11.2} = I$, we have

$$(\Sigma_*)_{rs} = -\Sigma_*(\Sigma_{11.2})_{rs}\Sigma_*. \tag{3.4}$$

Let E_{rs} be a matrix such that the only non-zero element is the (r, s) element and that it is 1. Substituting

$$\begin{aligned}(\Sigma_{11.2})_{rs} &= \tfrac{1}{2}(E_{rs}+E_{sr}) && \text{if } r, s \leqq p,\\ &= -\tfrac{1}{2}(E_{r,s-p}B'+BE_{s-p,r}) && \text{if } r \leqq p,\ s > p,\\ &= \tfrac{1}{2}B(E_{r-p,s-p}+E_{s-p,r-p})B' && \text{if } r, s > p\end{aligned}$$

into (3.4), we obtain (3.2). Also as

$$(\sigma_{**}^{ij})_{rs} = \textstyle\sum_{g,h=1}^{p}\left[(\sigma_*{}^{ig})_{rs}c_{gh}\sigma_*{}^{hj}+\sigma_*{}^{ig}(c_{gh})_{rs}\sigma_*{}^{hj}+\sigma_*{}^{ig}c_{gh}(\sigma_*{}^{hj})_{rs}\right],$$

where $[c_{gh}] = I+BB'$, the proof of (3.3) follows immediately from above.

Although the matrix functions involved are considerably different, we can obtain the following

LEMMA 3.2. *All formulas in Okamoto's Lemma* 4 *in* [5] *hold true with* σ *and* σ_2 *replaced by* σ_* *and* σ_{**}, *respectively.*

4. Proof of the main theorem. We can use the Fourier transform (e.g., Cramér [4]) for obtaining the distribution function of $D^{-1}(W^*-\frac{1}{2}D^2)$ for $\binom{x}{y}\in\Pi_1$ from its characteristic function

$$\phi(t) = E\{\exp[itD^{-1}(W^*-\tfrac{1}{2}D^2)] \mid \Pi_1\}. \tag{4.1}$$

In order to evaluate $\phi(t)$ we first consider the conditional characteristic function

$$\begin{aligned}&\psi(\bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2, S)\\ &\quad= E\{\exp[itD^{-1}(W^*-\tfrac{1}{2}D^2)] \mid \bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2, S; \Pi_1\},\end{aligned} \tag{4.2}$$

which is related to $\phi(t)$ by the equation

$$\phi(t) = E[\psi(\bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2, S)]. \tag{4.3}$$

Since the statistic W^* is invariant under any nonsingular linear transformation of the type

$$\binom{x}{y} \to \begin{pmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{pmatrix}\binom{x}{y}+\begin{pmatrix}B_1\\ B_2\end{pmatrix}. \tag{4.4}$$

we may suppose $\mu_1 = 0$, $\mu_2 = \mu_0$, $v = 0$, $\Sigma = I$, where μ_0 is a p-vector with the first component D and the others 0. Then, we find that

$$\begin{aligned}&\psi(\bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2, S)\\ &\quad= \exp\{\tfrac{1}{2}D\theta+\tfrac{1}{2}D^{-1}\theta[\bar{x}_1+\bar{x}_2-\hat{B}(\bar{y}_1+\bar{y}_2)]'\\ &\qquad\cdot S_{11.2}^{-1}[\bar{x}_1-\bar{x}_2-\hat{B}(\bar{y}_1-\bar{y}_2)]\\ &\qquad+\tfrac{1}{2}D^{-2}\theta^2[\bar{x}_1-\bar{x}_2-\hat{B}(\bar{y}_1-\bar{y}_2)]'\\ &\qquad\cdot S_{11.2}^{-1}(I+\hat{B}\hat{B}')S_{11.2}^{-1}[\bar{x}_1-\bar{x}_2-\hat{B}(\bar{y}_1-\bar{y}_2)]\},\end{aligned} \tag{4.5}$$

where $\theta = -it$ and $S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$. By using now the Taylor expansion for the multi-dimensional case we have from (4.3)

$$\phi(t) = \Theta\psi(\mu_1, \mu_2, \nu_1, \nu_2, \Sigma)]_0, \tag{4.6}$$

where

$$\Theta = E(\exp M), \tag{4.7}$$

$$M = \sum_{i=1}^{p} \bar{x}_{1i}\frac{\partial}{\partial\mu_{1i}} + \sum_{i=p+1}^{p+q} \bar{y}_{1i}\frac{\partial}{\partial\nu_{1i}} + \sum_{i=1}^{p} (\bar{x}_{2i}-\mu_{0i})\frac{\partial}{\partial\mu_{2i}} + \sum_{i=p+1}^{p+q} \bar{y}_{2i}\frac{\partial}{\partial\nu_{2i}} + \sum_{i,j=1}^{p+q} (s_{ij}-\delta_{ij})\frac{\partial}{\partial\sigma_{ij}}, \tag{4.8}$$

$\mu_i' = (\mu_{i1}, \cdots, \mu_{ip})$, $\nu_i' = (\nu_{ip+1}, \cdots, \nu_{ip+q})$ for $i = 1, 2$ and $]_0$ means the value when

$$\mu_1 = 0, \qquad \mu_2 = \mu_0, \qquad \nu_1 = \nu_2 = 0, \qquad \Sigma = I. \tag{4.9}$$

Note that for notational convenience we will mean the operator $\frac{1}{2}(1+\delta_{rs})\partial/\partial\sigma_{rs}$ whenever we write $\partial/\partial\sigma_{rs}$. We know that $\binom{\bar{x}_1}{\bar{y}_1}$, $\binom{\bar{x}_2}{\bar{y}_2}$ and nS are distributed independently according to $N(\binom{0}{0}, N_1^{-1}I)$, $N(\binom{\mu_0}{0}, N_2^{-1}I)$ and $W(n, I)$, respectively. Hence,

$$E_{\bar{x}_{1i}}\left\{\exp\left(\bar{x}_{1i}\frac{\partial}{\partial\mu_{1i}}\right)\right\} = \exp\left(\tfrac{1}{2}N_1^{-1}\frac{\partial^2}{\partial\mu_{1i}^2}\right),$$

$$E_{\bar{x}_{2i}}\left\{\exp\left[(\bar{x}_{2i}-\mu_{0i})\frac{\partial}{\partial\mu_{2i}}\right]\right\} = \exp\left(\tfrac{1}{2}N_2^{-1}\frac{\partial^2}{\partial\mu_{2i}^2}\right),$$

$$E_{\bar{y}_{ji}}\left\{\exp\left(\bar{y}_{ji}\frac{\partial}{\partial\nu_{ji}}\right)\right\} = \exp\left(\tfrac{1}{2}N_j^{-1}\frac{\partial^2}{\partial\nu_{ji}^2}\right) \qquad \text{for } j = 1, 2,$$

and

$$E_S\left\{\exp\left[\sum_{i,j=1}^{p+q} (s_{ij}-\delta_{ij})\frac{\partial}{\partial\sigma_{ij}}\right]\right\} = \exp\left\{n^{-1}tr\left[\frac{\partial}{\partial\sigma_{ij}}\right]^2 + \frac{4}{3}n^{-2}tr\left[\frac{\partial}{\partial\sigma_{ij}}\right]^3 + \cdots\right\}.$$

If we write

$$\mu_{ir} = \nu_{ir} \qquad \text{for} \quad p < r \leqq p+q,\ i = 1, 2, \tag{4.10}$$

then it follows from (4.7) and (4.8) that

$$\Theta = \exp\left[\frac{1}{2N_1}\sum_{r=1}^{p+q}\frac{\partial^2}{\partial\mu_{1r}^2} + \frac{1}{2N_2}\sum_{r=1}^{p+q}\frac{\partial^2}{\partial\mu_{2r}^2} + \frac{1}{n}\sum_{r,s=1}^{p+q}\frac{\partial^2}{\partial\sigma_{rs}^2} + \frac{4}{3n^2}\sum_{r,s,t=1}^{p+q}\frac{\partial^3}{\partial\sigma_{rs}\partial\sigma_{st}\partial\sigma_{tr}} + \cdots\right], \tag{4.11}$$

which can be put in the form of equation (3.17) in Okamoto [5], where each subscript runs over the range $1, \cdots, p+q$. We may rewrite (4.6) as

$$\phi(t) = \Theta \exp A(\mu_1, \mu_2, \nu_1, \nu_2, \Sigma)]_0, \tag{4.12}$$

where

$$\begin{aligned} A(\mu_1, \mu_2, \nu_1, \nu_2, \Sigma) &= \tfrac{1}{2}D\theta + \tfrac{1}{2}D^{-1}\theta[\mu_1+\mu_2-B(\nu_1+\nu_2)]'\Sigma_*[\mu_1-\mu_2-B(\nu_1-\nu_2)] \\ &\quad + \tfrac{1}{2}D^{-2}\theta^2[\mu_1-\mu_2-B(\nu_1-\nu_2)]'\Sigma_{**}[\mu_1-\mu_2-B(\nu_1-\nu_2)], \end{aligned} \tag{4.13}$$

Σ_* and Σ_{**} being defined in (3.1).

We shall now evaluate the coefficients in the expansiòn of $\phi(t)$. First, on substituting $\Sigma_{12} = 0$ into (4.13) it is found that the resulting expression is independent of ν_1, ν_2 and Σ_{22}. The coefficients of N^0, N_1^{-1}, N_2^{-1}, N_1^{-2}, $(N_1N_2)^{-1}$ and N_2^{-2} therefore remain the same as those in [5] for the Anderson statistic W, which implies that $L_i^*(d; D) = L_i(p)$, $Q_{ij}^*(d; D) = Q_{ij}(p)$ for $i, j = 1, 2$.

Calculation of the terms of order n^{-1} and n^{-2} is based solely on the lemmas in Section 3, which are formally identical with the lemmas given in [5]. Hence these terms can be derived from those in [5], or rather its corrected version [6], by changing the dimensionality p into $p+q$. This means $L_3^*(d; D) = L_3(p+q)$, $Q_{33}^*(d; D) = Q_{33}(p+q)$.

We shall now determine the coefficients of $(N_1 n)^{-1}$ and $(N_2 n)^{-1}$ in $\phi(t)$. Similarly with the equation (5.27) in [5], we have

$$\begin{aligned} &\Sigma_{r,s,t} \frac{\partial^4 e^A}{\partial\mu_{1t}^2 \partial\sigma_{rs}^2}\Big]_0 \\ &= \Sigma_{r,s,t}\left\{\left[\frac{\partial^2 A}{\partial\mu_{1t}^2} + \left(\frac{\partial A}{\partial\mu_{1t}}\right)^2\right]_0 \left[\frac{\partial^2 A}{\partial\sigma_{rs}^2} + \left(\frac{\partial A}{\partial\sigma_{rs}}\right)^2\right]_0 + R_{rst}\right\} e^{\frac{1}{2}\theta^2} \\ &= [2N_1 n L_1^*(\theta; D) L_3^*(\theta; D) + \textstyle\sum_{r,s,t=1}^{p+q} R_{rst}]\, e^{\frac{1}{2}\theta^2}, \end{aligned} \tag{4.14}$$

where

$$\begin{aligned} R_{rst} = 4 &\frac{\partial A}{\partial\mu_{1t}} \frac{\partial A}{\partial\sigma_{rs}} \frac{\partial^2 A}{\partial\mu_{1t}\partial\sigma_{rs}} + 2\left(\frac{\partial^2 A}{\partial\mu_{1t}\partial\sigma_{rs}}\right)^2 \\ &+ 2\left(\frac{\partial A}{\partial\mu_{1t}} \frac{\partial^3 A}{\partial\mu_{1t}\partial\sigma_{rs}^2} + \frac{\partial A}{\partial\sigma_{rs}} \frac{\partial^3 A}{\partial\mu_{1t}^2\partial\sigma_{rs}}\right) + \frac{\partial^4 A}{\partial\mu_{1t}^2 \partial\sigma_{rs}^2}\Big]_0. \end{aligned} \tag{4.15}$$

Now it holds that

$$\begin{aligned} \textstyle\sum_{r,s,t=1}^{p+q} R_{rst} &= \textstyle\sum_{r,s=1}^{p} \sum_{t=1}^{p+q} R_{rst} + 2\sum_{r,t=1}^{p} \sum_{s=p+1}^{p+q} R_{rst} \\ &\quad + 2\textstyle\sum_{r=1}^{p} \sum_{s,t=p+1}^{p+q} R_{rst} + \sum_{r,s=p+1}^{p+q} \sum_{t=1}^{p+q} R_{rst}, \end{aligned} \tag{4.16}$$

which can be evaluated from the following results which will be proved later.

$$\begin{aligned}
\Sigma_{r,s,t} R_{rst} &= D^{-2}[4\theta^4(2\theta^2-D\theta)+2(5p+7)\theta^4-D^2\theta^2+(p^2+p)(3\theta^2+D\theta)] && \text{if } 1 \leqq r, s \leqq p,\\
&= D^{-2}[3q\theta^4+\tfrac{1}{2}pq(3\theta^2+D\theta)] && \text{if } 1 \leqq r, t \leqq p,\ p < s \leqq p+q,\\
&= D^{-2}[\tfrac{1}{2}q\theta^4+\tfrac{1}{2}pq(\theta^2+D\theta)] && \text{if } 1 \leqq r \leqq p,\ p < s, t \leqq p+q,\\
&= 0 && \text{if } p < r, s \leqq p+q.
\end{aligned} \tag{4.17}$$

Combining (4.14), (4.16) and (4.17), we obtain

$$(2N_1 n)^{-1} \sum_{r,s,t} \frac{\partial^4 e^A}{\partial \mu_{1t}^2 \partial \sigma_{rs}^2}\Bigg]_0 = [L_1{}^*(\theta; D)L_3{}^*(\theta; D)+Q_{13}^*(\theta; D)]\, e^{\frac{1}{2}\theta^2}, \tag{4.18}$$

where

$$Q_{13}^*(\theta; D) = (2N_1 nD^2)^{-1}[4\theta^4(2\theta^2-D\theta)+(10p+7q+14)\theta^4 - D^2\theta^2+p(3p+4q+3)\theta^2+p(p+2q+1)D\theta], \tag{4.19}$$

which can be rewritten as in (2.2). The expression for $Q_{23}^*(\theta; D)$ is derived in a similar fashion.

We mention here briefly the method for finding the value of $\Sigma_{r,s,t} R_{rst}$ in the four situations in (4.17). Since differentiation appearing in (4.15) is not concerned with Σ_{12} in the first and fourth situations, we may substitute $\Sigma_{12} = 0$ into (4.13) before application of the differential operators. In case of the second and third situations we may put $\Sigma_{11} = I$ and $\Sigma_{22} = I$ in (4.13) and then apply the differential operators. For example, in the third situation, on substituting $\mu_1 = 0$, $\mu_2 = \mu_0$, $v_2 = 0$ as well, we have

$$A = \tfrac{1}{2}D\theta-\tfrac{1}{2}D^{-1}\theta a'\Sigma_* b+\tfrac{1}{2}D^{-2}\theta^2 b'\Sigma_{**} b,$$

where $a' = (a_1, a_2, \cdots, a_p)$, $b' = (b_1, b_2, \cdots, b_p)$ and

$$a_i = D\,\delta_{i1} - \textstyle\sum_{j=p+1}^{p+q} \sigma_{ij} v_{1j},$$

$$b_i = D\,\delta_{i1} + \textstyle\sum_{j=p+1}^{p+q} \sigma_{ij} v_{1j}.$$

Since all partial derivatives of a_i, b_i, $\sigma_*{}^{ij}$ and σ_{**}^{ij} by v_{1t} or σ_{rs} vanish at the point (4.9), it follows that

$$\frac{\partial A}{\partial v_{1t}}\Bigg]_0 = \frac{\partial A}{\partial \sigma_{rs}}\Bigg]_0 = 0. \tag{4.20}$$

Next, using

$$\frac{\partial^2 a_i}{\partial v_{1t} \partial \sigma_{rs}} = -\frac{\partial^2 b_i}{\partial v_{1t} \partial \sigma_{rs}} = -\tfrac{1}{2}\,\delta_{ir}\,\delta_{st},$$

we have

$$\frac{\partial^2 A}{\partial v_{1t} \partial \sigma_{rs}}\Bigg]_0 = \tfrac{1}{2}D^{-1}\theta^2\,\delta_{1r}\,\delta_{st}, \quad \frac{\partial^4 A}{\partial v_{1t}^2 \partial \sigma_{rs}^2}\Bigg]_0 = \tfrac{1}{2}D^{-2}(\theta^2+D\theta)\,\delta_{st}. \tag{4.21}$$

The value of $\Sigma_{r,s,t} R_{rst}$ in case of the third situation can now be easily obtained from (4.15), (4.20) and (4.21).

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] COCHRAN, W. G. (1964). Comparison of two methods of handling covariates in discriminatory analysis. *Ann. Inst. Statist. Math.* **16** 43–53.

[3] COCHRAN, W. G. and BLISS, C. I. (1948). Discriminant functions with covariance. *Ann. Math. Statist.* **19** 151–176.

[4] CRAMÉR, H. (1964). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[5] OKAMOTO, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.* **34** 1286–1301.

[6] OKAMOTO, M. (1968). Correction to 'An asymptotic expansion for the distribution of the linear discriminant function.' *Ann. Math. Statist.* **39** 1358–1359.