ASYMPTOTIC EXPANSIONS OF THE DISTRIBUTIONS OF SOME CLASSIFICATION STATISTICS AND THE PROBABILITIES OF MISCLASSIFICATION WHEN THE TRAINING SAMPLES ARE DEPENDENT

by

SHIBDAS BANDYOPADHYAY
University of California,
Davis

SOMESH DAS GUPTA*
University of Minnesota,
Minneapolis

(Running title: ASYMPTOTIC DISTRIBUTIONS OF CLASSIFICATION STATISTICS)

Technical Report Number 260
January, 1976

*Partly supported by U. S. Army Research Grant DAAG-29-76-G-0038 at the University of Minnesota.

SUMMARY

Asymptotic expansions of the distributions of some classification statistics and associated probabilities of misclassification are considered for a two population classification problem when the population distributions follow a stationary Gaussian process. Special cases have been considered when the population distributions follow a first order autoregressive process and, in particular, the probabilities of misclassification is studied as a function of the measure of dependence between the two populations.

1. Introduction: Let ω_0 be an experimental unit which is a random outcome from a population π. It is known that π is identical to one of the two specified populations π_1 and π_2, where π_1 and π_2 denote the same population π^* at two different points of time t_1 and t_2, respectively. Let $X_0 = X(\omega_0)$ be a p×1 vector of measurement on the unit ω_0. The problem is to identify π with one of π_1 and π_2 based on X_0 and the knowledge of the distributions of X_0 in π_1 and π_2, which are not completely known. Information about these distributions is obtained from a sample of N units $\omega_1,\ldots,\omega_N$ (called training sample) from π^* with $X_{i\alpha}$ as the X-observation on the unit ω_α observed at time t_i, $\alpha=1,\ldots,N$; $i=1,2$.

Let X_t denote the X-observation at time t. We shall assume that

$$(1.1) \qquad X_t = m_t + U_t,$$

where U_t follows a stationary Gaussian process, and in particular,

$$(1.2) \qquad \begin{pmatrix} U_{t_1} \\ U_{t_2} \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau & \Sigma \end{pmatrix}\right],$$

where Σ is a nonsingular matrix and τ is a p×p symmetric matrix (see Anderson (1971)). Then $(X'_{1\alpha}X'_{2\alpha})$, $(\alpha=1,\ldots,N)$ are i.i.d. and

$$(1.3) \qquad \begin{pmatrix} X_{1\alpha} \\ X_{2\alpha} \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau & \Sigma \end{pmatrix}\right],$$

where

$$(1.4) \qquad \mu_i = m_{t_i}, \qquad i=1,2.$$

A special case of (1.2) will also be treated when U_t follows a first order autoregressive process, i.e.,

(1.5) $$U_t = \lambda U_{t-1} + \epsilon_t, \qquad t=0,\pm 1,\pm 2,\cdots ,$$

where $|\lambda| < 1$ and ϵ_t's are i.i.d. $N_p(0,\Lambda)$. Then we can take (see Anderson (1971))

(1.6) $$\tau = \rho\Sigma, \qquad |\rho| < 1.$$

Let H_i denote the hypothesis that ω_0 is from π_i (i=1,2). When μ_1, μ_2 and Σ are known the form of a likelihood-ratio rule (see Anderson (1958)) is given by the following: Accept H_1 iff

(1.7) $$W^* \equiv ((X_0-\mu_2;\Sigma)) - ((X_0-\mu_1;\Sigma)) > k,$$

where k is a constant, and for a p×1 vector Y and a p×p nonsingular matrix B

(1.8) $$((Y;B)) = Y'B^{-1}Y.$$

When some of the parameters μ_1, μ_2 and Σ in (1.7) are unknown we replace them by their respective estimates, chosen suitably, based on the samples $(X'_{1\alpha}X'_{2\alpha})$, $\alpha=1,\ldots,N$. Such rules will be called plug-in likelihood-ratio (PLR) rules. The notion of PLR rules was first introduced by Wald (1944).

In this paper we assume that μ_1 and μ_2 are not known and we consider several cases depending on the available knowledge about Σ and τ.

The estimates of μ_1 and μ_2 will be taken, respectively, as $\bar{X}_1$ and $\bar{X}_2$, where

(1.9) $$\bar{X}_i = \sum_{\alpha=1}^{N} X_{i\alpha}, \qquad i=1,2.$$

Define

(1.10) $$S_{ij} = \sum_{\alpha=1}^{N} (X_{i\alpha}-\bar{X}_i)(X_{j\alpha}-\bar{X}_j)'; \qquad i,j=1,2.$$

The plug-in version of W* is given in general, by

(1.11) $$W_N = (\bar{X}_1-\bar{X}_2)'B^{-1}(X_0 - \tfrac{1}{2}(\bar{X}_1+\bar{X}_2)),$$

where the matrix B is defined below.

Case (a): $B = \Sigma$, when Σ is known.

Case (b): When Σ and τ are unknown,

(1.12) $$B = (S_{11}+S_{22})/(2N-2).$$

Case (c): When Σ is unknown and $\tau = \rho\Sigma$ with known ρ,

(1.13) $$B = (S_{11}-\rho S_{12}-\rho S_{21}+S_{22})/(1-\rho^2)(2N-2).$$

Case (d): When Σ is unknown and $\tau = \rho\Sigma$ with unknown ρ,

(1.14) $$B = (S_{11}+S_{22})/(2N-2).$$

Note that in each of the above cases B is an unbiased consistent estimate of Σ. In cases (b) and (c), B is an asymptotically efficient estimate of Σ; however, an asymptotically efficient estimate of Σ in case (d) is difficult to obtain (when $p > 1$). The limiting distribution of W_N as $N \to \infty$ is normal with variance α and mean $\frac{1}{2}\alpha$ if X_0 is from π_1 and mean $-\frac{1}{2}\alpha$ if X_0 is from π_2, where

$$\alpha = ((\mu_1 - \mu_2; \Sigma)). \tag{1.15}$$

We shall assume $\alpha > 0$.

In this paper we derive asymptotic expansions of the distributions of $(W_N - \frac{1}{2}\mathcal{A}_N)/\mathcal{A}_N^{1/2}$ and $(W_N + \frac{1}{2}\mathcal{A}_N)/\mathcal{A}_N^{1/2}$ when $\mathcal{E}X_0 = \mu_1$ and $\mathcal{E}X_0 = \mu_2$, respectively, in powers of $1/N$, where

$$\mathcal{A}_N = ((\bar{X}_1 - \bar{X}_2; B)). \tag{1.16}$$

Anderson (1973) obtained the asymptotic expansions of these distributions when $\tau = 0$ and B is given by (1.11). We shall closely follow Anderson's method; however we have to modify and extend Anderson's results because $\bar{X}_1$ and $\bar{X}_2$ are correlated and the presence of τ leads to a different estimate of Σ.

We shall also derive asymptotic expansions of the distributions of $(W_N - \frac{1}{2}\alpha)/\alpha^{1/2}$ and $(W_N + \frac{1}{2}\alpha)/\alpha^{1/2}$ when $\mathcal{E}X_0 = \mu_1$ and $\mathcal{E}X_0 = \mu_2$, respectively. This would extend the results of Okamoto (1963) who derived these asymptotic distributions when $\tau = 0$ and the usual estimate is taken for Σ. For deriving these results we shall again follow Anderson's method which is much simpler than Okamoto's method.

Finally, asymptotic expansions for $P\ (W_N \leq 0)$ will be obtained under both the hypotheses H_1 and H_2.

In classification theory, usually the training samples from π_1 and π_2 are drawn independently, i.e., on different sets of units. This is the first paper where the classification problem is treated with a dependent training sample which occurs quite often in practice. The results in this paper also reveal the influence of the covariance matrix τ on the distributions and the probabilities of misclassification.

2. Asymptotic Expansions of the Distributions of Studentized W_N

There exists a nonsingular p×p matrix L such that

$$(2.1)\qquad \begin{aligned} L\Sigma L' &= I_p, \\ L\tau L' &= D = \mathrm{diag}(\rho_1,\ldots,\rho_p). \end{aligned}$$

It can be seen that W_N and $\mathcal{Q}_N$ (in all the cases (a)-(d)) are invariant under the transformations

$$(2.2)\qquad \begin{aligned} X_0 &\to L(X_0-\mu_1) \\ \bar{X}_1 &\to L(\bar{X}_1-\mu_1) \\ \bar{X}_2 &\to L(\bar{X}_2-\mu_1) \\ S_{ij} &\to LS_{ij}L'; \qquad i,j=1,2. \end{aligned}$$

Hence, without loss of generality, we shall assume that

$$(2.3)\qquad \begin{aligned} \Sigma &= I_p, \qquad \tau = D = \mathrm{diag}(\rho_1,\ldots,\rho_p), \\ \mu_1 &= 0, \qquad \mu_2 = -\delta, \end{aligned}$$

where

$$(2.4)\qquad \delta'\delta = \alpha.$$

As in Anderson (1973), we define Y, Z and V as follows.

$$(2.5)\qquad \begin{aligned} \bar{X}_1 - \bar{X}_2 &= \delta + Y/n^{1/2}, \qquad \bar{X}_1 = Z/n^{1/2}, \\ B &= I_p + V/n^{1/2}, \qquad n = 2(N-1). \end{aligned}$$

Then

$$(2.6)\qquad \begin{pmatrix} Y \\ Z \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2(n/N)(I_p-D) & (n/N)(I_p-D) \\ (n/N)(I_p-D) & (n/N)I_p \end{pmatrix}\right].$$

Let $Y = (Y_1,\ldots,Y_p)'$, $Z = (Z_1,\ldots,Z_p)'$, $V = [V_{ij}]$. Let J_n be the subset of the sample space defined by

$$(2.7)\qquad J_n = \{|Y_i| < 4(\log n)^{1/2},\ \ |Z_i| < 2(\log n)^{1/2},$$
$$|V_{ij}| < 2\log n;\qquad i,j=1,2,\ldots,p\}.$$

The proof of the following lemma is given in the appendix.

<u>Lemma 2.1</u>: $\qquad P(J_n) = 1 - o(n^{-2}).$

Now

$$(2.8)\qquad \mathcal{O}_N = ((\delta+Y/n^{1/2};\ I_p+V/n^{1/2})),$$

and let

$$(2.9)\qquad G_{1N} = (\delta+Y/n^{1/2})'(I_p+V/n^{1/2})^{-1}(Z/n^{1/2}),$$

$$(2.10)\qquad G_{2N} = ((\delta+Y/n^{1/2};\ (I_p+V/n^{1/2})^2)).$$

Assume now $\mathcal{E}X_0 = 0$ (i.e., under H_1). Then

$$(2.11)\qquad P(W_N \leq u) = \mathcal{E}\Phi[(u\mathcal{O}_N^{1/2}+G_{1N})G_{2N}^{-1/2}],$$

where Φ is the c.d.f. of N(0,1). Now

$$(2.12)\qquad \begin{aligned}&|\mathcal{E}\Phi[(u\mathcal{O}_N^{1/2}+G_{1N})/G_{2N}^{1/2}]\chi(J_n)\\ &\quad - \mathcal{E}\Phi[(u\mathcal{O}_N^{1/2}+G_{1N})/G_{2N}^{1/2}]|\\ &\leq \mathcal{E}\ \chi(J_n^{C}) = o(n^{-2}),\end{aligned}$$

where χ stands for the indicator function of a set.

Following Anderson (1973), it can be shown, for sufficiently large n and Y, Z, V in J_n, that

$$\Phi[(u\alpha_N^{1/2}+G_{1N})/G_{2N}^{1/2}] = \Phi(u) + n^{-1/2}\varphi(u)C(Z,V) \tag{2.13}$$
$$+ n^{-1}\varphi(u)[D(Y,Z,V) - (u/2)C^2(Z,V)] + R_n,$$

where $\varphi(u) = d\Phi(u)/du$,

$$C(Z,V) = (u/2\Delta^2)\delta'V\delta + (\delta'Z)/\Delta, \tag{2.14}$$

$$\begin{aligned} D(Y,Z,V) = {} & \Delta^{-1}(Y'Z-\delta'VZ) \\ & + \Delta^{-2}u(\delta'VY-\delta'V^2\delta) \\ & - \Delta^{-3}(\delta'Y\delta'Z-\delta'Z\delta'V\delta) \\ & + u\Delta^{-4}[(7/8)(\delta'V\delta)^2-\delta'Y\delta'V\delta], \end{aligned} \tag{2.15}$$

and

$$\Delta = \alpha^{1/2}.$$

R_n is the sum of three terms R_{1n}, R_{2n} and R_{3n}, where R_{1n} is a homogeneous polynomial (not depending on n) of degree 3 in Y, Z and V multiplied by $n^{-3/2}$, R_{2n} is a homogeneous polynomial (not depending on n) of degree 4 in Y, Z and V multiplied by n^{-2}, and R_{3n} is $O((\log n)^5/n^{5/2})$.

Next, we follow Anderson's method, although we clarify his results and fill in some gaps. First note that the expectation of a polynomial in Y_i's and Z_j's is O(1), and

$$\mathcal{E}V_{ij}^{2k-1} = O(n^{-1/2}), \quad \mathcal{E}V_{ij}^{2k} = O(1) \tag{2.16}$$

for any positive integer k. We shall now obtain the expectation of the right-hand side of (2.13) multiplied by $\chi(J_n)$.

$\mathcal{E}[\delta' Z\chi(J_n)] = 0$ since $\mathcal{E}Z = 0$ and J_n is symmetric about 0 in Z_i and the density of Z_i is symmetric about 0. Moreover, $\mathcal{E}[\delta' V\delta\chi(J_n)] = O(n^{-3/2})$, since $\mathcal{E}V = 0$,

$$|\mathcal{E}(\delta' V\delta)\chi(J_n)| = |\delta'\mathcal{E}V\chi(J_n{}^C)\delta|,$$

and

$$\text{(2.17)} \qquad \begin{aligned} &\mathcal{E}[|V_{ij}|\chi(|V_{ij}| > 2\log n)] \\ &\leq [\mathcal{E}V_{ij}{}^2]^{1/2}P[\chi(|V_{ij}| > 2\log n)] = O(n^{-3/2}) \end{aligned}$$

by taking $k = 3$ in (A.3). Thus

$$\text{(2.18)} \qquad \mathcal{E}[C(Z,V)\chi(J_n)] = O(n^{-3/2}).$$

Next, by Cauchy-Schwarz' inequality

$$\text{(2.19)} \qquad \begin{aligned} &|\mathcal{E}[D(Y,Z,V)-(u/2)C^2(Z,V)] - \mathcal{E}[D(Y,Z,V)-(u/2)C^2(Z,V)]\chi(J_n)| \\ &\leq [\mathcal{E}\{D(Y,Z,V)-(u/2)C^2(Z,V)\}^2]^{1/2}[P(J_n{}^C)]^{1/2} = o(n^{-1}). \end{aligned}$$

Similarly

$$\text{(2.20)} \qquad |\mathcal{E}R_{1n}-\mathcal{E}R_{1n}\chi(J_n)| = O(n^{-2}).$$

Since the third-order moments of the elements of Y, Z and V are either 0 or $O(n^{-1/2})$, $\mathcal{E}R_{1n} = O(n^{-1/2})$. Note now

$$\text{(2.21)} \qquad |\mathcal{E}R_{2n}\chi(J_n)| \leq \mathcal{E}|R_{2n}| = O(1),$$

$$(2.22)\qquad |\mathcal{E}R_{3n}\chi(J_n)| = O((\log n)^5/n^{5/2}) = o(n^{-2}).$$

Combining the above results, we get

$$(2.23)\qquad \begin{aligned}&\mathcal{E}\Phi[(u\alpha_N^{1/2}+G_{1N})G_{2N}^{-1/2}]\\ &= \Phi(u) + n^{-1}\varphi(u)\mathcal{E}[D(Y,Z,V)-(u/2)C^2(Z,V)] + o(n^{-2}).\end{aligned}$$

The following theorem is thus proved.

Theorem 2.1: When H_1 obtains

$$\begin{aligned}&P[(W_N - \tfrac{1}{2}\alpha_N)/\alpha_N^{1/2} \leq u]\\ &= \Phi(u) + n^{-1}\varphi(u)\mathcal{E}[D(Y,Z,V)-(u/2)C^2(Z,V)] + O(n^{-2}).\end{aligned}$$

Note now

$$(2.24)\qquad \begin{aligned}&P_1[(W_N - \tfrac{1}{2}\alpha_N)/\alpha_N^{1/2} \leq u]\\ &= P_2[(W_N + \tfrac{1}{2}\alpha_N)/\alpha_N^{1/2} \geq -u],\end{aligned}$$

where P_i denotes the probability under H_i, $i=1,2$. Thus we get the following corollary.

Corollary 2.1.1: When H_2 obtains

$$\begin{aligned}&P[(W_N + \tfrac{1}{2}\alpha_N)/\alpha_N^{1/2} \leq u]\\ &= \Phi(u) - n^{-1}\varphi(u)\mathcal{E}[D(Y,Z,V)-(u/2)C^2(Z,V)] + O(n^{-2}).\end{aligned}$$

The above two results are similar to those of Anderson's (1973) except that we needed a modified proof to treat our case. The presence of τ now leads to new problems regarding the evaluation of $\mathcal{E}D(Y,Z,V)$ and $\mathcal{E}C^2(Z,V)$. We shall consider this now for different cases.

Case (a): $B = I_p$. In this case $V = 0$ and $|V_{ij}| < 2 \log n$ for all i,j trivially holds.

$$\mathcal{E}D(Y,Z,V) = \Delta^{-1}\mathcal{E}(Y'Z) - \Delta^{-3}\mathcal{E}(\delta'Y\delta'Z)$$

(2.25)

$$= \Delta^{-1}(n/N)\sum_{i=1}^{p}(1-\rho_i) - \Delta^{-3}(n/N)\delta'(I_p-D)\delta.$$

In terms of the original parameters,

$$\mathcal{E}D(Y,Z,V) = \Delta^{-1}(n/N)\mathrm{tr}(I_p-\Sigma^{-1}\tau)$$

(2.26)

$$- \Delta^{-3}(n/N)[\Delta^2-(\mu_1-\mu_2)'\Sigma^{-1}\tau\Sigma^{-1}(\mu_1-\mu_2)].$$

For the special case $\tau = \rho\Sigma$,

(2.27) $$\mathcal{E}D(Y,Z,V) = \Delta^{-1}(n/N)(p-1)(1-\rho).$$

Next

(2.28) $$\mathcal{E}C^2(Z,V) = \Delta^{-2}\mathcal{E}(\delta'Z)^2 = (n/N).$$

For what follows, (n/N) is replaced by 2. Theorem 2.1 now yields the following:

Theorem 2.2: When $B = \Sigma$ and H_1 obtains

$$P[(W_N - \tfrac{1}{2}\alpha_N)/\alpha_N^{1/2} \le u] = \Phi(u) + n^{-1}\varphi(u)[2\Delta^{-1}(p-1) - 2\Delta^{-1}\mathrm{tr}(\Sigma^{-1}\tau)$$
$$+ 2\Delta^{-3}(\mu_1-\mu_2)'\Sigma^{-1}\tau\Sigma^{-1}(\mu_1-\mu_2) - u] + O(n^{-2}),$$

which reduces to

$$\Phi(u) + n^{-1}\varphi(u)[2\Delta^{-1}(p-1)(1-\rho) - u] + O(n^{-2}),$$

when $\tau = \rho\Sigma$.

<u>Case (b)</u>: $B = (S_{11}+S_{22})/(2N-2)$.

Note that

$$\mathcal{E}[D(Y,Z,V)-(u/2)C^2(Z,V)]$$

$$= 2\Delta^{-1}tr(I_p-D) - 2\Delta^{-3}\delta'(I_p-D)\delta - u - u\Delta^{-2}\mathcal{E}(\delta'V^2\delta)$$

(2.29) $$+ [(7/8)u\Delta^{-4}-(u^3/8\Delta^4)]\mathcal{E}(\delta'V\delta)^2 + O(n^{-1}).$$

In this case

(2.30) $$V = (4n)^{-1/2}[B_1+B_2-2nI_p],$$

where

(2.31) $$B_1 = S_{11} + S_{22} + S_{12} + S_{21} \sim W_p[2(I_p+D),n/2],$$

(2.32) $$B_2 = S_{11} + S_{22} - S_{12} - S_{21} \sim W_p[2(I_p-D),n/2],$$

and B_1 and B_2 are independently distributed. Hence

(2.33) $$\mathcal{E}V^2 = (4n)^{-1}[\mathcal{E}(B_1-\mathcal{E}B_1)^2 + \mathcal{E}(B_2-\mathcal{E}B_2)^2].$$

(2.34) $$\mathcal{E}(\delta'V\delta)^2 = (4n)^{-1}[Var(\delta'B_1\delta) + Var(\delta'B_2\delta)].$$

We shall reduce the above expressions using the following lemma.

Lemma 2.2: Let A be a $p \times p$ random matrix distributed as $W_p(\Lambda,m)$, where Λ is a nonsingular diagonal matrix. Then

(a) $$\mathcal{E}(A-\mathcal{E}A)^2 = m[\Lambda^2 + \Lambda(\mathrm{tr}\ \Lambda)],$$

and for any vector $\delta \neq 0$

(b) $$\mathrm{Var}(\delta'A\delta) = 2m(\delta'\Lambda\delta)^2.$$

The proof of this lemma is omitted.

Thus

$$\mathcal{E}V^2 = (4n)^{-1}[4(I_p+D)^2+4(I_p+D)\{\mathrm{tr}(I_p+D)\} + 4(I_p-D)^2 + 4(I_p-D)\{\mathrm{tr}(I_p-D)\}](n/2)$$

$$= [(p+1)I_p + (\mathrm{tr}\ D)D + D^2], \tag{2.35}$$

$$\mathcal{E}(\delta'V^2\delta) = (p+1)\Delta^2 + \delta'D^2\delta + \delta'D\delta(\mathrm{tr}\ D), \tag{2.36}$$

$$\mathcal{E}(\delta'V\delta)^2 = 2[\Delta^4 + (\delta'D\delta)^2]. \tag{2.37}$$

From Theorem 2.1 and the above results we get the following theorem.

Theorem 2.3: When $B = (S_{11}+S_{22})/(2N-2)$ and H_1 obtains

$$P[(W_N - \tfrac{1}{2}\alpha_N)/\alpha_N^{1/2} \leq u] = \Phi(u) + n^{-1}\varphi(u)[2\Delta^{-1}\mathrm{tr}(I_p-\Sigma^{-1}T) - u$$

$$- 2\Delta^{-3}(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}T\Sigma^{-1})(\mu_1-\mu_2) - u(p+1)$$

$$- u\Delta^{-2}(\mu_1-\mu_2)'\Sigma^{-1}T\Sigma^{-1}T\Sigma^{-1}(\mu_1-\mu_2)$$

$$- u\Delta^{-2}(\mu_1-\mu_2)'\Sigma^{-1}\tau\Sigma^{-1}(\mu_1-\mu_2)\mathrm{tr}(\tau\Sigma^{-1}) + ((7/4)u - u^3/4)$$

$$+ [(7/4)u\Delta^{-4} - (u^3/4\ \Delta^4)][(\mu_1-\mu_2)'\Sigma^{-1}\tau\Sigma^{-1}(\mu_1-\mu_2)]^2 + O(n^{-2}),$$

which reduces to

$$\Phi(u) + n^{-1}\varphi(u)[(2/\Delta)(p-1)(1-\rho) - u\{p + 1/4 + (p-3/4)\rho^2\}$$

$$- (u^3/4)(1+\rho^2)] + O(n^{-2}),$$

when $\tau = \rho\Sigma$.

<u>Case (c)</u>: $B = (S_{11}-\rho S_{12}-\rho S_{21}+S_{22})/(1-\rho^2)(2N-2)$, and $D = \rho I_p$.

In this case $2(N-1)B \sim W_p(I_p,n)$, and

$$V = n^{1/2}[B-\mathcal{E}(B)].$$

Hence, from Lemma 2.2

$$\mathcal{E}(\delta'V^2\delta) = (p+1)\Delta^2, \tag{2.38}$$

$$\mathcal{E}(\delta'V\delta)^2 = 2\Delta^4. \tag{2.39}$$

Theorem 2.1 now yields the following.

<u>Theorem 2.4</u>: When $\tau = \rho\Sigma$, $B = (S_{11}-\rho S_{12}-\rho S_{21}+S_{22})/(1-\rho^2)(2N-2)$ and H_1 obtains

$$P[(W_n - \tfrac{1}{2}\mathcal{a}_N)/\mathcal{a}_N^{1/2} \le u]$$

$$= \Phi(u) + n^{-1}\varphi(u)[2\Delta^{-1}(1-\rho)(p-1) - u(p+1/4) - u^3/4] + O(n^{-2}).$$

3. Asymptotic Expansions of the Distributions of Normalized W_n.

First note that under H_1

$$P[(W_N - \tfrac{1}{2}\alpha)/\alpha^{1/2} \leq u] = \mathcal{E}\Phi[(w + \tfrac{1}{2}G_{3n})G_{2N}^{-1/2}], \tag{3.1}$$

where

$$w = u\Delta + \Delta^2/2, \tag{3.2}$$

$$G_{3n} = (\bar{X}_1+\bar{X}_2)'B^{-1}(\bar{X}_1-\bar{X}_2), \tag{3.3}$$

and G_{2n} is given by (2.10). Define $U = (U_1,\ldots,U_p)'$ by

$$(\bar{X}_1+\bar{X}_2) = U/n^{1/2} - \delta, \tag{3.4}$$

and Y and V by (2.5). Then

$$\begin{pmatrix} Y \\ U \end{pmatrix} \sim N_{2p}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2(n/N)(I_p-D) & 0 \\ 0 & 2(n/N)(I_p+D) \end{pmatrix}\right]. \tag{3.5}$$

Redefine J_n by J_n^*, where

$$\begin{aligned} J_n^* = \{|Y_i| < 4(\log n)^{1/2},\ |U_i| < 4(\log n)^{1/2}, \\ |V_{ij}| < 2\log n;\quad i,j=1,\ldots,p\}. \end{aligned} \tag{3.6}$$

Then it can be shown as in Lemma 2.1 that

$$P(J_n^*) = 1 - o(n^{-2}). \tag{3.7}$$

Replacing J_n by J_n^* and proceeding as in Section 2 it can be shown that under H_1

$$P[(W_N - \tfrac{1}{2}\alpha)/\alpha^{1/2} \le u]$$

$$= \Phi(u) + n^{-1}\varphi(u)\mathcal{E}[D^*(Y,U,V) - (u/2)C^{*2}(Y,U,V)] + O(n^{-2}), \tag{3.8}$$

where

$$\begin{aligned} D^*(Y,U,V) &= (2\Delta)^{-1}[U'Y+\delta'VY-\delta'V^2\delta-U'V\delta] \\ &\quad - (2\Delta^3)^{-1}[(\delta'Y-\delta'V\delta)(U'\delta+\delta'V\delta-\delta'Y)] \\ &\quad -u[(2\Delta^2)^{-1}(3\delta'V^2\delta+Y'Y-4\delta'VY) - 3(2\Delta^4)^{-1}(\delta'Y-\delta'V\delta)^2], \end{aligned} \tag{3.9}$$

$$C^*(Y,U,V) = (2\Delta)^{-1}(U'\delta+\delta'V\delta-\delta'Y) - u\Delta^{-2}/\delta'Y-\delta'V\delta). \tag{3.10}$$

Now, in terms of the original parameters

$$\begin{aligned} \mathcal{E}D^*(Y,U,V) &= \mathcal{E}(\delta'V\delta)^2[(2\Delta^3)^{-1} + 3u(2\Delta^4)^{-1}] \\ &\quad - \mathcal{E}(\delta'V^2\delta)[(2\Delta)^{-1}+3u(2\Delta^2)^{-1}] \\ &\quad + (2/\Delta^3)(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}\tau\Sigma^{-1})(\mu_1-\mu_2) \\ &\quad - (2u/\Delta^2)\mathrm{tr}(I_p-\Sigma^{-1}\tau) \\ &\quad + (6u/\Delta^4)(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}\tau\Sigma^{-1})(\mu_1-\mu_2) + O(1/n), \end{aligned} \tag{3.11}$$

$$\begin{aligned} \mathcal{E}C^{*2}(Y,U,V) &= 2 + 4(u^2/\Delta^4+u/\Delta^3)(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}\tau\Sigma^{-1})(\mu_1-\mu_2) \\ &\quad + [(1/2\Delta)^2 + (u^2/\Delta^4) + (u/\Delta^3)]\mathcal{E}(\delta'V\delta)^2 + O(1/n). \end{aligned} \tag{3.12}$$

<u>Case (a)</u>: $B = I_p$.

Here $V = 0$ and we get the following theorem from (3.8) to (3.12).

<u>Theorem 3.1</u>: When Σ is known, $B = \Sigma$, and H_1 obtains

$$P[(W_N - \tfrac{1}{2}\alpha)/\alpha^{1/2} \leq u] = \Phi(u) + n^{-1}\varphi(u)[2\Delta^{-3}(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}\tau\Sigma^{-1})(\mu_1-\mu_2)$$

$$- u - 2u\Delta^{-2}\mathrm{tr}(I_p-\Sigma^{-1}\tau)$$

$$-2(u^3\Delta^{-4}+u^2\Delta^{-3})(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}\tau\Sigma^{-1})(\mu_1-\mu_2)$$

$$+ 6u\Delta^{-4}(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}\tau\Sigma^{-1})(\mu_1-\mu_2)] + O(n^{-2}),$$

which reduces to

$$\Phi(u) + n^{-1}\varphi(u)[2(1-\rho)\Delta^{-1}-u-2(p-3)(1-\rho)u\Delta^{-2}$$

$$- 2(1-\rho)(u^3\Delta^{-2}+u^2\Delta^{-1})] + O(n^{-2}),$$

when $\tau = \rho\Sigma$.

<u>Case (b)</u>: <u>Σ and τ are unknown, and B is given by (1.11)</u>. In this case, $P[(W_N - \frac{1}{2}\alpha)/\alpha^{1/2} \leq u]$ under H_1 can be obtained from (3.8), (3.11), (3.12), (2.36) and (2.37); due to its lengthy form the explicit expression is omitted.

<u>Case (c)</u>: <u>Σ is unknown, $\tau = \rho\Sigma$ with known ρ, and B is given by (1.12)</u>. From (3.8), (3.11), (3.12), (2.38) and (2.39) we get the following.

<u>Theorem 3.2</u>: When H_1 obtains and B is given by (1.12)

$$P[(W_N - \tfrac{1}{2}\alpha)/\alpha^{1/2} \leq u] = \Phi(u) + n^{-1}\varphi(u)[2(1-\rho)\Delta^{-1}-u(3p-1)/2$$

$$- \Delta(p-1)/2 - 2(p-3)(1-\rho)u\Delta^{-2} - 2(1-\rho)(u^3\Delta^{-2}+u^2\Delta^{-1})$$

$$-u(\Delta^2/4 + u^2+u\Delta)] + O(n^{-2}).$$

Similar results can be obtained when H_2 obtains using the following relation:

$$P_1[(W_N - \tfrac{1}{2}\alpha)/\alpha^{1/2} \le u] = P_2[(W_N + \tfrac{1}{2}\alpha)/\alpha^{1/2} \ge -u], \tag{3.13}$$

where P_i is the probability under H_i $(i=1,2)$.

4. Asymptotic Expansions of Probabilities of Misclassification.

Let us consider the rule which classifies X_0 into π_1 iff

$$W_N \ge 0, \tag{4.1}$$

where W_N is given by (1.11). Then the probability of the misclassification (PMC) for this rule under H_1 is given by

$$P_1[(W_N - \tfrac{1}{2}\alpha)/\alpha^{1/2} < -\Delta/2], \tag{4.2}$$

and its asymptotic expansion can be obtained from (3.8) with $u = \Delta/2$. It follows from (3.13) that the PMC for this rule under H_2 is the same as (4.2). Using results in Section 3 we shall now state the asymptotic expansions of (4.2) for different cases under consideration.

When Σ is known and B is taken as Σ, (4.2) reduces to

$$\begin{aligned}&\Phi(-\Delta/2) + n^{-1}\varphi(\Delta/2)[\Delta/2 + \Delta^{-1}\mathrm{tr}(I_p-\Sigma^{-1}\tau)\\ &\quad -(\mu_1-\mu_2)'(\Sigma^{-1}-\Sigma^{-1}\tau\Sigma^{-1})(\mu_1-\mu_2)\{\Delta^{-3}+(4\Delta)^{-1}\}] + O(n^{-2}).\end{aligned} \tag{4.3}$$

In particular, when $\tau = \rho\Sigma$, (4.3) reduces to

$$\Phi(-\Delta/2) + n^{-1}\varphi(\Delta/2)[\Delta(1+\rho)/4 + \Delta^{-1}(p-1)(1-\rho)] + O(n^{-2}). \tag{4.4}$$

When $\tau = \rho\Sigma$ with known ρ but unknown Σ, and B is given by (1.13), (4.2) becomes

$$(4.5)\qquad \Phi(-\Delta/2) + n^{-1}\varphi(\Delta/2)[\Delta(p-1)/4 + \Delta(1+\rho)/4 + (p-1)(1-\rho)\Delta^{-1}] + O(n^{-2}).$$

When $\tau = \rho\Sigma$ with both ρ and Σ unknown, and B is given by (1.15), (4.2) reduces to

$$(4.6)\qquad \begin{aligned} &\Phi(-\Delta/2) + n^{-1}\varphi(\Delta/2)[\Delta(1+\rho)/4 + \Delta(1+\rho^2)(p-1)/4 \\ &\qquad + (p-1)(1-\rho)\Delta^{-1}] + O(n^{-2}). \end{aligned}$$

Remarks:

(i) Assume $p > 1$. Then the PMC of the rule given by (4.1) can be studied as a function of ρ, to the order of approximations indicated, when $\tau = \rho\Sigma$. It follows from (4.4) and (4.5) that the PMC's in these two cases increase (or decrease) with ρ when $\Delta^2 >$ (or $<$) $4(p-1)$. It follows from (4.6) that the PMC in this case increases (or decreases) with ρ according as $\rho >$ (or $<$) $[2\Delta^{-2} - (2p-2)^{-1}]$.

(ii) When $p = 1$, the PMC's in all the cases are increasing functions of ρ. However, it is shown in Bandyopadhyay (1974) that the exact PMC decreases as ρ decreases when Δ is large and it increases as ρ decreases when Δ is small.

(iii) When $\tau = 0$, the results in Section 2 yield Anderson's results (1973) and the results in Sections 3 and 4 yield Okamoto's results (1963).

APPENDIX

Proof of Lemma 2.1:

$$P[|Y_i| > 4(\log n)^{1/2}] = (\tfrac{2}{\pi})^{1/2} \int_{t_n}^{\infty} e^{-1/2x^2} dx$$

where

$$t_n = 4(\log n)^{1/2}[2(n/N)(1-\rho_i)]^{-1/2}.$$

By Mill's ratio inequality

$$P[|Y_i| > 4(\log n)^{1/2}] \leq (\tfrac{2}{\pi})^{1/2} e^{-1/2t_n^2} t_n^{-1} .$$

Now

$$e^{-t_n^2/2} t_n^{-1}$$

$$= e^{-8 \log n/(n-N)(1-\rho_i)} [4(\log n)^{1/2}\{2(n/N)(1-\rho_i)\}^{-1/2}]^{-1}$$

$$= [n^{8N/n(1-\rho_i)} (\log n)^{1/2}]^{-1}[2(1-\rho_i)(n/N)]^{1/2}(1/4)$$

$$\leq [n^2(\log n)^{1/2}]^{-1}(2)^{-1/2} = o(n^{-2}).$$

Hence

(A.1) $$P[|Y_i| > 4(\log n)^{1/2}] \leq o(n^{-2}).$$

Similarly

(A.2) $$P[|Z_i| > 2(\log n)^{1/2}] \leq o(n^{-2}).$$

From (1.12) and (1.13) it follows that $B = B_1 + B_2$, $(N-1)B_k \sim W_p[I_p, N-1]$, $(k=1,2)$ but B_1 and B_2 are not necessarily independent. Let $B_k = ((B_{ijk}))$. Now

$$V_{ii} > 2 \log n$$

$$\Leftrightarrow (\sqrt{n}/2)[B_{1ii} + B_{2ii} - 2] > 2 \log n$$

$$\Leftrightarrow (n/2)[B_{1ii} + B_{2ii}] > 2\sqrt{n} \log n + n.$$

Hence, for $\theta > 0$

$$P[V_{ii} > 2 \log n] \leq e^{-\theta(\sqrt{n} \log n + n/2)} \cdot E\, e^{(\theta/2)(B_{1ii} + B_{2ii})(N-1)}$$

$$\leq e^{-\theta(\sqrt{n} \log n + n/2)} \cdot E \left[e^{\theta(N-1)B_{1ii}} + e^{\theta(N-1)B_{2ii}} \right](1/2)$$

$$= e^{-\theta(\sqrt{n} \log n + n/2)}(1-2\theta)^{-n/4}, \quad 0 < \theta < 1/2.$$

Let

(A.3) $$\theta = k/\sqrt{n}.$$

Fix k and let n be sufficiently large such that $k/\sqrt{n} = \theta < 1/2$. Then

$$P[V_{ii} > 2 \log n] \leq e^{-k\sqrt{n}/2} e^{-k \log n}(1-2k/\sqrt{n})^{-n/4}.$$

But, since $n/4 = (\sqrt{n}/2k)k\sqrt{n}/2$, we have

$$P[V_{ii} > 2 \log n] \leq \text{Constant } e^{-k \log n} = O(n^{-k}).$$

Similarly

$$P[-V_{ii} > 2 \log n] \leq O(n^{-k}).$$

Hence

$$(A.4) \qquad P[|V_{ii}| > 2 \log n] \leq O(n^{-k}).$$

Now let $i \neq j$.

$$V_{ij} > 2 \log n$$

$$\Leftrightarrow (n/4)(B_{1ij} + B_{2ij}) > \sqrt{n} \log n.$$

Hence

$$P[V_{ij} > 2 \log n] \leq e^{-\theta\sqrt{n} \log n} E\, e^{(n/4)\theta(B_{1ij} + B_{2ij})}$$

$$\leq e^{-\theta\sqrt{n} \log n} E \left[e^{(n/2)\theta B_{1ij}} + e^{(n/2)\theta B_{2ij}} \right] (1/2)$$

$$= e^{-\theta\sqrt{n} \log n}(1-\theta^2)^{n/4}, \qquad 0 < \theta < 1/2.$$

Again, for fixed k and for θ as in (A.3) and for sufficiently large n such that $\theta = k/\sqrt{n} < 1/2$, we have

$$P[V_{ij} > 2 \log n] \leq e^{-k \log n}(1-k^2/n)^{-n/4} = O(n^{-k}).$$

Similarly,

$$P[-V_{ij} > 2 \log n] \leq O(n^{-k}).$$

Hence,

(A.5) $$P[|V_{ij}| > 2 \log n] \leq O(n^{-k}).$$

Lemma 2.1 follows from (A.1), (A.2), (A.4) and (A.5) for $k > 2$.

REFERENCES

ANDERSON, T. W. (1958): An Introduction to Multivariate Statistical Analysis, Wiley, New York.

______________ (1971): The Statistical Analysis of Time Series, Wiley, New York.

______________ (1973): An asymptotic expansion of the distribution of studentized classification statistic W. Ann. Stat., 1, 964-972.

BANDYOPADHYAY, S. (1974): Classification with dependent training sample. Doctoral dissertation, School of Statistics, University of Minnesota.

OKAMOTO, M. (1963): An asymptotic expansion for the distribution of linear discriminant function. Ann. Math. Stat., 34, 1286-1301. (Correction: 39, 1358-1359.)

WALD, A. (1944): On a statistical problem arising in the classification of an individual into one of two groups. Ann. Math. Stat., 15, 145-162.

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report No. 260	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Asymptotic expansions of the distributions of some classification statistics and the probabilities of misclassification when the training samples are dependent	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Somesh Das Gupta and Shibdas Bandyopadhyay	8. CONTRACT OR GRANT NUMBER(s): DAAG-29-76-G-0038
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Theoretical Statistics, University of Minnesota, Minneapolis, MN 55455	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Research Office, Mathematics Division - P.O. Box 12211, Research Triangle Park, NC 27709	12. REPORT DATE: 22 January, 1976
	13. NUMBER OF PAGES: 23 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Reproduction in whole or part is permitted for any purpose of the U.S. Government. Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Classification, normal distribution, dependent training sample, asymptotic expansion, probability of misclassification.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Asymptotic expansions of the distributions of some classification statistics and associated probabilities of misclassification are considered for a two population classification problem when the population distributions follow a stationary Gaussian process. Special cases have been considered when the population distributions follow a first order autoregressive process and, in particular, the probabilities of misclassification is studied as a function of the measure of dependence between the two populations.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)