Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

A.W. Ambergen

Asymptotic distributions of estimators for posterior probabilities in a classification model with both continuous and discrete variables

Department of Mathematical Statistics Report MS-R8409 May

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

ASYMPTOTIC DISTRIBUTIONS OF ESTIMATORS FOR POSTERIOR PROBABILITIES IN A CLASSIFICATION MODEL WITH BOTH CONTINUOUS AND DISCRETE VARIABLES

A.W. AMBERGEN
Centre for Mathematics and Computer Science, Amsterdam

This paper is devoted to the asymptotic distribution of estimators for the posterior probability that an observation vector originates from one of k populations. The estimators are based on training samples. The random vectors contain both continuous and discrete variables. Observation vector and prior probabilities are regarded as given constants. The continuous part of the random vector has conditional on the discrete part a multivariate normal distribution. Several assumptions about homogeneity of the dispersion matrices are considered.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 62H30.
KEY WORDS & PHRASES: estimating posterior probabilities, discriminant analysis.

Report MS-R8409
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. INTRODUCTION

Let us assume that an observation originates from one of k populations. Information about each of the populations is available in the form of a training sample i.e. outcomes of independent random vectors. Each random vector contains both continuous and discrete random variables. So we have the realizations of the independent random vectors

$$(X_{hi1},\dots,X_{hip},W_{hi1},\dots,W_{hiq})^T, \quad i = 1,\dots,N_h; \quad h = 1,\dots,k.$$

where $X_{hi1},\dots,X_{hip}$ are p continuous and $W_{hi1},\dots,W_{hiq}$ q discrete random variables. The discrete variable $W_{hi\ell}$ has a finite range of α_ℓ distinct values or categories. We combine the q discrete variables into one discrete variable D_{hi} with values in the set $1,\dots,d$ where $d = \Pi_{\ell=1}^{q}\ \alpha_\ell$. Writing $X_{hi} = (X_{hi1},\dots,X_{hip})^T$ and assuming conditional normality for the continuous variables we shall concentrate on the random vectors

$$(1.1) \qquad \binom{X_{hi}}{D_{hi}}, \quad i = 1,\dots,N_h; \quad h = 1,\dots,k$$

where

$$(1.2) \qquad X_{hi}\Big|_{D_{hi}=\ell} \sim N_p(\mu_{h\ell},\Sigma_{h\ell}), \quad h = 1,\dots,k; \quad \ell = 1,\dots,d$$

and

$$P(D_{hi}=\ell) = p_{h\ell} > 0 \text{ with } \sum_{\ell=1}^{d} p_{h\ell} = 1, \quad h = 1,\dots,k.$$

Let $f_{h\ell}$ denote the $N_p(\mu_{h\ell},\Sigma_{h\ell})$ probability density function. The given observation vector, denoted by $(x^T,j)^T$, $x \in \mathbb{R}^p$, $j \in \{1,\dots,d\}$, is considered as the result of a drawing from one of k distributions, each associated with one population. The distributions have densities $p_{hj}f_{hj}(x)$, $h = 1,\dots,k$ in the point $(x^T,j)^T$ w.r.t. the product of Lebesgue measure and counting measure. Assume that the prior probabilities of belonging to the populations are $\rho_1,\dots,\rho_k$. The posterior probabilities are defined by

$$(1.4) \qquad \rho_{t|(x,j)} = \frac{\rho_t p_{tj} f_{tj}(x)}{\Sigma_{h=1}^{k} \rho_h p_{hj} f_{hj}(x)} \qquad t = 1,\dots,k.$$

We shall consider $\rho_1,\dots,\rho_k$ and $(x^T,j)^T$ as given constants. The k unknown posterior probabilities are considered as unknown parameters which are estimated from the training samples. Let $R_{t|(x,j)}$ denote an estimator for $\rho_{t|(x,j)}$, $t = 1,\dots,k$. Using the notation

$$(1.5) \qquad \begin{aligned} R_{\cdot|(x,j)} &= (R_{1|(x,j)},\dots,R_{k|(x,j)})^T, \\ \rho_{\cdot|(x,j)} &= (\rho_{1|(x,j)},\dots,\rho_{k|(x,j)})^T \end{aligned}$$

and $N = \Sigma_{h=1}^{k} N_h$ we shall prove that $N^{\frac{1}{2}}(R_{\cdot|(x,j)}-\rho_{\cdot|(x,j)})$ is asymptotically normal with expectation zero and a singular dispersion matrix.
We shall consider four different situations depending on assumptions about homogeneity of the variance-covariance matrices of the multivariate normal distributions.

The literature about estimating posterior probabilities in discriminant analysis deals mostly with the assumptions k = 2 and joint normality of the measured variables. So this paper in which discrete variables are added is an important extension. SCHAAFSMA & VAN VARK (1977) considers the case $p = 1$, $k = 2$. SCHAAFSMA & VAN VARK (1979) deals with $p \geq 1$, $k = 2$. AMBERGEN (1981) considers the case $p \geq 1$, $k \geq 2$ and gives of various stochasts exact moments. AMBERGEN & SCHAAFSMA (1982) considers $p \geq 1$, $k \geq 2$ with normality assumptions as well as only assumptions about continuous densities, in the latter case a nonparametric approach is given. AMBERGEN & SCHAAFSMA (1983[a]) contains an application to physical anthropology. AMBERGEN & SCHAAFSMA (1983[b]) contains a simulation experiment in which the theoretical confidence coefficient for confidence intervals for the posterior probabilities is compared with that obtained by using asymptotically normal approximations. AMBERGEN & SCHAAFSMA (1984) is a revised version of the latter. SCHAAFSMA (1982) has an emphasize on the selection of variables. Apart from the "estimative" methods used in this paper the "predictive" method of GEISSER (1964) has been discussed in the literature. AITCHISON, HABBEMA & KAY (1977) is a comparison of the two methods. McLACHLAN (1977) studies the bias of sample based posterior probabilities. McLACHLAN (1979) compares the bias of classical plug in estimators with that of predictive estimators. RIGBY (1982) constructs credibility intervals for the posterior probabilities in order to compare the estimative and predictive estimators. In KRZANOWSKI (1975) an allocation model for two populations with mixtures of continuous and binary variables is considered. The continuous variables have also conditional on the discrete variables multivariate normal distributions. However, more structure is supposed for estimation the discrete and continuous parameters than in this paper will be done. The argumentation in that paper is obvious because for small training samples the unstructured estimation can be unsatisfactory.

2. RESULTS

In this section we define the estimators and give the asymptotic distributions for the posterior estimators in four different cases. The proofs are given in the following sections. We use the following definitions and notations, in which $h = 1,\dots k$ and $\ell = 1,\dots,d$:

$$\Delta^2_{x;h\ell} = (x-\mu_{h\ell})^T\ \Sigma^{-1}_{h\ell}(x-\mu_{h\ell})$$

$$f_{h\ell}(x) = f(x;\mu_{h\ell},\Sigma_{h\ell}) = |2\pi\Sigma_{h\ell}|^{-\frac{1}{2}}\exp(-\tfrac{1}{2}\Delta^2_{x;h\ell})$$

N_h size of h-th training sample

$$N = \sum_{h=1}^{k} N_h,\quad b_h = \frac{N_h}{N}$$

$$\Delta_{h\ell i} = I(D_{hi}=\ell),\quad N_{h\ell} = \sum_{i=1}^{N_h} \Delta_{h\ell i}$$

$$\hat{p}_{h\ell} = \frac{N_{h\ell}}{N_h}$$

$$\hat{\mu}_{h\ell} = \frac{1}{N_{h\ell}} \sum_{i=1}^{N_h} X_{hi}\Delta_{h\ell i}$$

$$\hat{\Sigma}_{h\ell} = \frac{1}{N_{h\ell}} \sum_{i=1}^{N_h} (X_{hi}-\hat{\mu}_{h\ell})(X_{hi}-\hat{\mu}_{h\ell})\Delta_{h\ell i}$$

$$R_{h|(x,j)} = \frac{\rho_h \hat{p}_{hj} \hat{f}_{hj}(x)}{\Sigma_{s=1}^{k} \rho_s \hat{p}_{sj} \hat{f}_{sj}(x)}$$

$$\hat{f}_{hj}(x) = f_{hj}(x;\hat{\mu}_{hj},\hat{\Sigma}^*_{hj})$$

where $\hat{\Sigma}^*_{hj}$ is an estimator for Σ_{hj} defined in theorem 2.1. We state now

<u>THEOREM 2.1.</u>

$$LN^{\frac{1}{2}}(R_{\cdot|(x,j)} - \rho_{\cdot|(x,j)}) \to N_k(0,\Psi\Lambda\Psi) \tag{2.1}$$

where

$$\Psi_{t,t} = \tfrac{1}{2}\rho_{t|(x,j)}(1-\rho_{t|(x,j)}) \tag{2.2}$$

$$\Psi_{t,s} = -\tfrac{1}{2}\rho_{t|(x,j)}\rho_{s|(x,j)} \quad t \neq s \tag{2.3}$$

$R_{\cdot|(x,j)}, \rho_{\cdot|(x,j)}$ are defined in (1.5) and for Λ we distinguish four cases dependent on $\hat{\Sigma}^*_{hj}$, the estimator for Σ_{hj}:

<u>Case I.</u>

$$\hat{\Sigma}^*_{hj} = \hat{\Sigma}_{hj} \quad h = 1,\dots,k \tag{2.4}$$

$$\begin{aligned} \Lambda_{t,t} &= \frac{4}{b_t p_{tj}}(1-p_{tj}) + \frac{2}{b_t p_{tj}}(p+\Delta^4_{x;tj}) \\ \Lambda_{t,s} &= 0 \quad t \neq s \end{aligned} \tag{2.5}$$

<u>Case II.</u> Assumption: $\Sigma_{1j} = \dots = \Sigma_{kj}$ $(=\Sigma_j)$

$$\hat{\Sigma}^*_{hj} = \hat{\Sigma}_j = \frac{1}{\hat{p}_{1j}b_1+\dots+\hat{p}_{kj}b_k} \sum_{h=1}^{k} b_k \hat{p}_{hj} \hat{\Sigma}_{hj} \tag{2.6}$$

$$\Lambda_{t,t} = \frac{4}{b_t p_{tj}} (1-p_{tj}) + \frac{4}{b_t p_{tj}} \Delta^2_{x;tj} + 2(\Sigma^k_{h=1} b_h p_{hj})^{-1} \Delta^4_{x;tj}$$

(2.7)

$$\Lambda_{t,s} = 2(\Sigma^k_{h=1} b_h p_{hj})^{-1} \{(x-\mu_{tj})^T \Sigma_j^{-1} (x-\mu_{sj})\}^2 \quad t \neq s$$

Case III. Assumption $\Sigma_{11} = \ldots = \Sigma_{kd} (=\Sigma)$

(2.8) $$\hat{\Sigma}^*_{hj} = \hat{\Sigma} = \sum_{h=1}^{k} b_h \sum_{\ell=1}^{d} \hat{p}_{h\ell} \hat{\Sigma}_{h\ell}$$

(2.9)
$$\Lambda_{t,t} = \frac{4}{b_t p_{tj}} (1-p_{tj}) + \frac{4}{b_t p_{tj}} \Delta^2_{x;tj} + 2\Delta^4_{x;tj}$$
$$\Lambda_{t,s} = 2\{(x-\mu_{tj})^T \Sigma^{-1} (x-\mu_{sj})\}^2 \quad t \neq s$$

Case IV. Assumption: $\Sigma_{h1} = \ldots = \Sigma_{h1} (=\Sigma_h)$, $h = 1,\ldots,k$

(2.10) $$\hat{\Sigma}^*_{hj} = \hat{\Sigma}_h = \sum_{\ell=1}^{d} \hat{p}_{h\ell} \hat{\Sigma}_{h\ell}$$

(2.11)
$$\Lambda_{t,t} = \frac{4(1-p_{tj})}{b_t p_{tj}} (1+\Delta^2_{x;tj}) + \frac{2(p+\Delta^4_{x;tj})}{b_t}$$
$$\Lambda_{t,s} = 0 \quad t \neq s$$

Remark. A special situation occurs if only one discrete state is possible: $p_{hj} = 1$, $h = 1,\ldots,k$; $j = 1$. For the four cases we obtain:

Case I = Case IV

(2.12)
$$\Lambda_{t,t} = \frac{2}{b_t} (p+\Delta^4_{x;t})$$
$$\Lambda_{t,s} = 0 \quad t \neq s$$

Case II = Case III

(2.13)
$$\Lambda_{t,t} = \frac{4}{b_t} \Delta^2_{x;t} + 2\Delta^4_{x;t}$$
$$\Lambda_{t,s} = 2\{(x-\mu_t)^T \Sigma^{-1} (x-\mu_s)\}^2 \quad t \neq s$$

where we dropped the index $j = 1$ in $\Delta^2_{x;t}$. These expressions (2.12) and (2.13) are the same as found earlier in a model with only continuous multivariate normal variables, see e.g. formulas (2.4) and (3.6) in AMBERGEN & SCHAAFSMA (1982).

3. SOME PROPERTIES OF MATRICES

We introduce some definitions and give a summary of properties which enable us to perform computations in a relatively short and elegant way. We closely follow notations in MAGNUS & NEUDECKER (1979).

- If A is a $m\times n$ matrix, $A_{\cdot j}$ the jth column of A, then $\mathrm{vec}(A)$ is the mn columnvector defined by $\mathrm{vec}^T(A) = (A^T_{\cdot 1},\ldots,A^T_{\cdot n})$.
- If A is a $m\times n$ and B a $s\times t$ matrix then the Kronecker product $A\otimes B$ is the (ms,nt) matrix defined by $A\otimes B = (a_{ij}B)$
- $(A_1\otimes B_1)(A_2\otimes B_2)\ldots(A_k\otimes B_k) = (A_1A_2\ldots A_k)\otimes(B_1B_2\ldots B_k)$ provided that $A_1A_2\ldots A_k$ and $B_1B_2\ldots B_k$ exists.
- $(A\otimes B)^T = A^T\otimes B^T$
- $(A\otimes B)\otimes C = A\otimes(B\otimes C)$
- $(A+B)\otimes C = (A\otimes B) + (B\otimes C)$
- $A\otimes(B+C) = A\otimes B + A\otimes C$
- $\mathrm{vec}\ (ABC) = (C^T\otimes A)\mathrm{vec}(B)$
- $A = 1\otimes A = A\otimes 1$
- $\mathrm{trace}(AB) = \mathrm{vec}^T(A^T)\mathrm{vec}(B)$
- For e_i the i-th unit dolumn vector of dimension p we define $K_p = \Sigma^p_{i=1}\Sigma^p_{j=1}(E_{ij}\otimes E_{ji})$ where $E_{ij} = e_ie_j^T$.
- $K_p\mathrm{vec}(A) = \mathrm{vec}(A^T)$
- $K_p = K_p^T$
- $K_p(A\otimes B) = (B\otimes A)K_p$
- $\mathrm{vec}(I)\mathrm{vec}^T(I) = \Sigma_{i,j}E_{ij}\otimes E_{ij}$

For μ a $p\times 1$ vector and Σ a symmetric $p\times p$ matrix we shall use frequently:

- $\mathrm{vec}(\mu\mu^T) = \mu\otimes\mu = I\mu\otimes\mu.1 = (I\otimes\mu)(\mu\otimes 1) = (I\otimes\mu)\mu$
- $(\mu\otimes I)\Sigma = \mu\otimes\Sigma$ because $((\mu\otimes I)\Sigma)^T = \Sigma(\mu^T\otimes I) = (1\otimes\Sigma)(\mu^T\otimes I) = \mu^T\otimes\Sigma = (\mu\otimes\Sigma)^T$
- $(I\otimes\mu)\Sigma = (\Sigma\otimes\mu)^T$ because $((I\otimes\mu)\Sigma)^T = \Sigma(I\otimes\mu^T) = (\Sigma\otimes 1)(I\otimes\mu^T) = \Sigma\otimes\mu^T = (\Sigma\otimes\mu)^T$

- $(\mu\otimes\Sigma)(I\otimes\mu^T) = \mu\otimes\Sigma\otimes\mu^T$ because
  $(\mu\otimes\Sigma)(I\otimes\mu^T) = ((\mu\otimes\Sigma)\otimes 1)(I\otimes\mu^T) = ((\mu\otimes\Sigma)I)\otimes(1.\mu^T) = \mu\otimes\Sigma\otimes\mu^T$
- $(\Sigma\otimes\mu)(\mu^T\otimes I) = \mu^T\otimes\Sigma\otimes\mu$ because
  $((\Sigma\otimes\mu)(\mu^T\otimes I))^T = (\mu\otimes I)(\Sigma\otimes\mu^T) = (\mu\otimes I)(1\otimes(\Sigma\otimes\mu^T)) = (\mu.1)\otimes(I(\Sigma\otimes\mu^T)) =$
  $\mu\otimes\Sigma\otimes\mu^T = (\mu^T\otimes\Sigma\otimes\mu)^T$

4. ASYMPTOTIC DISTRIBUTION OF A BASIC RANDOM VECTOR

The asymptotic distribution presented in lemma 4.1 will be the corner-stone for the proof of theorem 2.1.
For h = 1,...,k and s = 1,...,d we define

$$U_{h,s} = \sum_{i=1}^{N_h} I(D_{hi}=s) \qquad : 1 \times 1 \text{ r.v.}$$

$$S_{h,s} = \sum_{i=1}^{N_h} X_{hi} I(D_{hi}=s) \qquad : p \times 1 \text{ r.v.}$$

$$T_{h,s} = \sum_{i=1}^{N_h} (X_{hi}\otimes X_{hi}) I(D_{hi}=s) \quad : p^2 \times 1 \text{ r.v.}$$

and formulate the lemma

<u>LEMMA 4.1.</u>

$$N_h^{\frac{1}{2}} \begin{bmatrix} \frac{1}{N_h}U_{h,s} - p_{hs} \\ \frac{1}{N_h}S_{h,s} - p_{hs}\mu_{hs} \\ \frac{1}{N_h}T_{h,s} - p_{hs}(\mathrm{vec}(\Sigma_{hs})+\mu_{hs}\otimes\mu_{hs}) \\ \frac{1}{N_h}U_{h,t} - p_{ht} \\ \frac{1}{N_h}S_{h,t} - p_{ht}\mu_{ht} \\ \frac{1}{N_h}T_{h,t} - p_{ht}(\mathrm{vec}(\Sigma_{ht})+\mu_{ht}\otimes\mu_{ht}) \end{bmatrix} \xrightarrow{L} N_{2(1+p+p^2)}(\underline{0},M_h)$$

with M_h partitioned as

$$\begin{pmatrix} M_{h,ss} & M_{h,st} \\ M_{h,ts} & M_{h,tt} \end{pmatrix}$$

and with an obvious further partioning, we state further that

$$M_{h,ss;1,1} = p_{hs}(1-p_{hs})$$

$$M_{h,ss;1,2} = p_{hs}(1-p_{hs})\mu_{hs}^T$$

$$M_{h,ss;1,3} = p_{hs}(1-p_{hs})\{\mathrm{vec}^T(\Sigma_{hs})+\mu_{hs}^T\otimes\mu_{hs}^T\}$$

$$M_{h,ss;2,2} = p_{hs}\Sigma_{hs} + p_{hs}(1-p_{hs})\mu_{hs}\mu_{hs}^T$$

$$M_{h,ss;2,3} = p_{hs}\{\mu_{hs}^T\otimes\Sigma_{hs}+\Sigma_{hs}\otimes\mu_{hs}^T\}+p_{hs}(1-p_{hs})\{\mu_{hs}\mathrm{vec}^T(\Sigma_{hs}) + \mu_{hs}(\mu_{hs}^T\otimes\mu_{hs}^T)\}$$

$$M_{h,ss;3,3} = p_{hs}\{(\Sigma_{hs}\otimes\Sigma_{hs})(I+K_p)+(\mu_{hs}\mu_{hs}^T)\otimes\Sigma_{hs}+\mu_{hs}^T\otimes\Sigma_{hs}\otimes\mu_{hs} + \mu_{hs}\otimes\Sigma_{hs}\otimes\mu_{hs}^T + \Sigma_{hs}\otimes\mu_{hs}\mu_{hs}^T\} + p_{hs}(1-p_{hs}). \{\mathrm{vec}(\Sigma_{hs})\mathrm{vec}^T(\Sigma_{hs})+\mathrm{vec}(\Sigma_{hs})(\mu_{hs}^T\otimes\mu_{hs}^T)+(\mu_{hs}\otimes\mu_{hs})\mathrm{vec}^T(\Sigma_{hs}) + (\mu_{hs}\otimes\mu_{hs})(\mu_{hs}^T\otimes\mu_{hs}^T)\}$$

and

$$M_{h,ts;1,1} = -p_{ht}p_{hs}$$

$$M_{h,ts;1,2} = -p_{ht}p_{hs}\mu_{hs}^T$$

$$M_{h,ts;1,3} = -p_{ht}p_{hs}\{\mathrm{vec}^T(\Sigma_{hs})+\mu_{hs}^T\otimes\mu_{hs}^T\}$$

$$M_{h,ts;2,1} = -p_{ht}p_{hs}\mu_{ht}$$

$$M_{h,ts;2,2} = -p_{ht}p_{hs}\mu_{ht}\mu_{hs}^T$$

$$M_{h,ts;2,3} = -p_{ht}p_{hs}\mu_{ht}\{\mathrm{vec}^T(\Sigma_{hs})+\mu_{hs}^T\otimes\mu_{hs}^T\}$$

$$M_{h,ts;3,1} = -p_{ht}p_{hs}\{vec(\Sigma_{ht})+\mu_{ht}\otimes\mu_{ht}\}$$

$$M_{h,ts;3,2} = -p_{ht}p_{hs}\{vec(\Sigma_{ht})+\mu_{ht}\otimes\mu_{ht}\}\mu_{hs}^T$$

$$M_{h,ts;3,3} = -p_{ht}p_{hs}\{vec(\Sigma_{ht})vec^T(\Sigma_{hs})+(\mu_{ht}\otimes\mu_{ht})vec^T(\Sigma_{hs}) + vec(\Sigma_{ht})(\mu_{hs}^T\otimes\mu_{hs}^T)+(\mu_{ht}\otimes\mu_{ht})(\mu_{hs}^T\otimes\mu_{hs}^T)\}.$$

In order to prove lemma 4.1 we shall first formulate two other lemmas. For that purpose we shall introduce the short notation $X = X_{h1}$ and $I_s = I(D_{h1}=s)$ which we shall use in the remaining part of this section.

LEMMA 4.2.

(a) $EI_s = p_{hs}$

(b) $\text{var } I_s = p_{hs}(1-p_{hs})$

(c) $EI_sX = p_{hs}\mu_{hs}$

(d) $EI_sXX^T = p_{hs}(\Sigma_{hs}+\mu_{hs}\mu_{hs}^T)$

(e) $EI_sX\otimes X = p_{hs}(vec(\Sigma_{hs})+\mu_{hs}\otimes\mu_{hs})$.

PROOF. (a) and (b) follow from the binomial $(1,p_{hs})$ distribution, (c), (d) and (e) can be derived with use of conditional expectation.

LEMMA 4.3. *If* $U \sim N_p(0,I)$, $X = \mu+\Sigma^{\frac{1}{2}}U$ *with* $\Sigma^{\frac{1}{2}}$ *symmetric positive definite then*

(a) $EU\otimes U^T = I$

(b) $EU\otimes U = vec(I)$

(c) $EUU^T\otimes U = 0$

(d) $EUU^T\otimes U^T = 0$

(e) $EX(X^T\otimes X^T) = \mu(\mu^T\otimes\mu^T)+\mu vec^T(\Sigma)+\mu^T\otimes\Sigma+\Sigma\otimes\mu^T$

(f) $EUU^T\otimes UU^T = K_p+I\otimes I+vec(I)vec^T(I)$

(g) $E(X\otimes X)(X^T\otimes X^T) = \mu\mu^T\otimes\mu\mu^T+\Sigma\otimes\mu\mu^T+(\mu\otimes\mu)vec^T(\Sigma) + \mu^T\otimes\Sigma\otimes\mu+\mu\otimes\Sigma\otimes\mu^T+vec(\Sigma)(\mu^T\otimes\mu^T) + \mu\mu^T\otimes\Sigma+(\Sigma\otimes\Sigma)(I+K_p)+vec(\Sigma)vec^T(\Sigma)$.

PROOF. (a)...(d) follow directly from

$$EU_i = EU_i^3 = 0$$

and

$$EU_i^2 = 1, \quad EU_i^4 = 3$$

where U_i is the i-th component of U. To prove (e):

$$EX(X^T \otimes X^T) = E(\mu+\Sigma^{\frac{1}{2}}U)(\mu^T+U^T\Sigma^{\frac{1}{2}})\otimes(\mu^T+U^T\Sigma^{\frac{1}{2}}) =$$

$$\mu(\mu^T\otimes\mu^T) + E\mu(U^T\Sigma^{\frac{1}{2}}\otimes U^T\Sigma^{\frac{1}{2}}) + E\Sigma^{\frac{1}{2}}U(\mu^T\otimes U^T\Sigma^{\frac{1}{2}}) + E\Sigma^{\frac{1}{2}}U(U^T\Sigma^{\frac{1}{2}}\otimes\mu^T)$$

where we have deleted terms with first and thirth moments. Using $\Sigma^{\frac{1}{2}}U = 1\otimes\Sigma^{\frac{1}{2}}U = \Sigma^{\frac{1}{2}}U\otimes 1$ for the last two terms we get

$$\mu(\mu^T\otimes\mu^T) + \mu E(U^T\otimes U^T)(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}}) + \mu^T\otimes\Sigma^{\frac{1}{2}}EUU^T\Sigma^{\frac{1}{2}} + \Sigma^{\frac{1}{2}}EUU^T\Sigma^{\frac{1}{2}}\otimes\mu^T$$

With $EUU^T = I$, $EU^T\otimes U^T = \mathrm{vec}^T(I)$ and the property $\mathrm{vec}(ABC) = (C^T\otimes A)\mathrm{vec}(B)$ using the special choice $A = C = \Sigma^{\frac{1}{2}}$ and $B = I$ (e) is proved.

proof of (f):

By defining $T_{ij} = E_{ij} + E_{ji}$ we obtain

$$EU_iU_jUU^T = T_{ij} + \delta_{ij}I$$

where δ_{ij} is the Kronecker delta: $\delta_{ij} = 1$ if $i = j$ and $= 0$ if $i \neq j$.
Now

$$EUU^T\otimes UU^T = E\Sigma_{i,j}\, U_iU_jE_{ij}\otimes UU^T =$$

$$\Sigma_{i,j}(E_{ij}\otimes(T_{ij}+\delta_{ij}I)) = \Sigma_{i,j}E_{ij}\otimes T_{ij} + \Sigma_{i,j}E_{ij}\otimes\delta_{ij}I =$$

$$\Sigma_{i,j}E_{ij}\otimes E_{ij} + \Sigma_{i,j}E_{ij}\otimes E_{ji} + (\Sigma_iE_{ii})\otimes I =$$

$$\mathrm{vec}(I)\mathrm{vec}^T(I) + K_p + I\otimes I$$

proof of (g):

$$E(X\otimes X)(X^T\otimes X^T) = EXX^T\otimes XX^T =$$

$$E(\mu+\Sigma^{\frac{1}{2}}U)(\mu^T+U^T\Sigma^{\frac{1}{2}})\otimes(\mu+\Sigma^{\frac{1}{2}}U)(\mu^T+U^T\Sigma^{\frac{1}{2}})$$

deleting terms with first and fifth moments this becomes

$$\mu\mu^T\otimes\mu\mu^T + \Sigma^{\frac{1}{2}}EUU^T\Sigma^{\frac{1}{2}}\otimes\mu\mu^T + E\mu U^T\Sigma^{\frac{1}{2}}\otimes\mu U^T\Sigma^{\frac{1}{2}} +$$
$$E\Sigma^{\frac{1}{2}}U\mu^T\otimes\mu U^T\Sigma^{\frac{1}{2}} + E\mu U^T\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}}U\mu^T +$$
$$E\Sigma^{\frac{1}{2}}U\mu^T\otimes\Sigma^{\frac{1}{2}}U\mu^T + \mu\mu^T\otimes\Sigma^{\frac{1}{2}}EUU^T\Sigma^{\frac{1}{2}} +$$
$$E\Sigma^{\frac{1}{2}}UU^T\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}}UU^T\Sigma^{\frac{1}{2}}.$$

Now, use $EUU^T = I$, $EU^T\otimes U^T = \mathrm{vec}^T(I)$ and $\mathrm{vec}^T(I)(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}}) = \mathrm{vec}^T(\Sigma)$ then the first three terms become

$$\mu\mu^T\otimes\mu\mu^T + \Sigma\otimes\mu\mu^T + (\mu\otimes\mu)\mathrm{vec}^T(\Sigma).$$

Because $EU\otimes U^T = I$ we obtain for the fourth term

$$(\Sigma^{\frac{1}{2}}\otimes\mu)I(\mu^T\otimes\Sigma^{\frac{1}{2}}) = (1\otimes(\Sigma^{\frac{1}{2}}\otimes\mu))(\mu^T\otimes\Sigma^{\frac{1}{2}}) =$$
$$(1.\mu^T)\otimes((\Sigma^{\frac{1}{2}}\otimes\mu)\Sigma^{\frac{1}{2}}) = \mu^T\otimes(\Sigma^{\frac{1}{2}}\otimes\mu)(\Sigma^{\frac{1}{2}}\otimes 1) = \mu^T\otimes\Sigma\otimes\mu$$

The fifth term is the transpose of the fourth term and is $\mu\otimes\Sigma\otimes\mu^T$. Further

$$(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})\mathrm{vec}(I)(\mu^T\otimes\mu^T) = \mathrm{vec}(\Sigma)(\mu^T\otimes\mu^T)$$

and

$$\mu\mu^T\otimes\Sigma^{\frac{1}{2}}EUU^T\Sigma^{\frac{1}{2}} = \mu\mu^T\otimes\Sigma$$

The last term can be written as

$$(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})EUU^T\otimes UU^T(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})$$

now using $(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})K_p = K_p(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})$ and $(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})\mathrm{vec}(I) = \mathrm{vec}(\Sigma)$ the proof of (g) and thus of lemma 4.3 is finished. □

Proof of lemma 4.1. With the results of lemma 4.2 and lemma 4.3 it is now easy to compute the components of the partitioned matrices $M_{h,ss}$ and $M_{h,ts}$. We have

$$M_{h,ss;1,1} = \mathrm{var}(N_h^{-\frac{1}{2}}U_{h,s}) = N_h^{-1}\mathrm{cov}(\Sigma_i I(D_{hi}=s),\ \Sigma_j I(D_{hj}=s)) =$$
$$= N_h^{-1}\ \Sigma_i\ \mathrm{var}(I(D_{hi}=s)) = \mathrm{var}(I_s)$$

since $I(D_{hi}=s)$ and $I(D_{hj}=s)$ for $i \neq j$ are independent. Further, without complete proofs we summarize:

$$M_{h,ss;1,2} = EI_s X^T - EI_s EI_s X^T$$

$$M_{h,ss;1,3} = EI_s X^T \otimes X^T - EI_s EI_s X^T \otimes X^T$$

$$M_{h,ss;2,2} = EI_s XX^T - EI_s XEI_s X^T$$

$$M_{h,ss;2,3} = EI_s X(X^T \otimes X^T) - EI_s XEI_s X^T \otimes X^T$$

$$M_{h,ss;3,3} = EI_s (X \otimes X)(X^T \otimes X^T) - EI_s X \otimes XEI_s X^T \otimes X^T$$

The components of the partitioned matrix $M_{h,ts}$ are derived in a similar way. We have

$$M_{h,ts;1,1} = \mathrm{cov}(N_h^{-\frac{1}{2}} U_{h,t}, N_h^{-\frac{1}{2}} U_{h,s}) = N_h^{-1} \mathrm{cov}(\Sigma_i I(D_{hi}=s), \Sigma_j I(D_{hj}=t))$$

$$= \mathrm{cov}(I_s, I_t) = EI_s I_t - EI_s EI_t = -EI_s EI_t$$

where we have used the independence between $I(D_{hi}=s)$ and $I(D_{hj}=t)$ for $i \neq j$ and the fact that $EI_s I_t = 0$. This is because $I_s I_t$, the product of the two variables I_s and I_t, can only have the value 0. So that $EI_s I_t Y = 0$ for any random variable Y. Deleting terms with such an expectation it is easy to verify that

$$M_{h,ts;1,2} = -EI_t EI_s X^T$$

$$M_{h,ts;1,3} = -EI_t EI_s X^T \otimes X^T$$

$$M_{h,ts;2,1} = -EI_t XEI_s$$

$$M_{h,ts;2,2} = -EI_t XEI_s X^T$$

$$M_{h,ts;2,3} = -EI_t XEI_s X^T \otimes X^T$$

$$M_{h,ts;3,1} = -EI_t X \otimes XEI_s$$

$$M_{h,ts;3,2} = -EI_t X \otimes XEI_s X^T$$

$$M_{h,ts;3,3} = -EI_t X \otimes XEI_s X^T \otimes X^T.$$

Application of lemma 4.2 and lemma 4.3 gives the earlier mentioned expressions.

5. ASYMPTOTIC DISTRIBUTION OF SOME ESTIMATORS

The following lemma, "the δ-method", will play a dominant role in the remaining part of this paper. It can be found in SERFLING (1980), §3.3, theorem A. We will refer to it as "the δ-method".

LEMMA 5.1. *suppose* $Lm^{\frac{1}{2}}(Y_m-\eta) \to N_p(0,\Sigma)$ *for some sequence of random variables* Y_m *assuming outcomes in* $\mathbb{R}^p$, *suppose moreover that* $h = (h_1,\ldots,h_q): \mathbb{R}^p \to \mathbb{R}^q$ *is differentiable at* η,

$$\nabla_h = \left(\frac{\partial h_1}{\partial x}(\eta),\ldots,\frac{\partial h_q}{\partial x}(\eta)\right) \text{ where } \frac{\partial}{\partial x} = \left(\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_p}\right)^T$$

then

$$Lm^{\frac{1}{2}}(h(Y_m)-h(\eta)) \to N_q(0,(\nabla_h)^T\Sigma(\nabla_h)).$$

Let us now define

$$\hat{p}_{hs} = \frac{1}{N_h} U_{h,s}$$

$$\hat{\mu}_{hs} = \frac{S_{h,s}}{U_{h,s}} = \frac{1/N_h\, S_{h,s}}{1/N_h\, U_{h,s}}$$

$$\mathrm{vec}(\hat{\Sigma}_{hs}) = \frac{T_{h,s}}{U_{h,s}} - \frac{S_{h,s}\otimes S_{h,s}}{U_{h,s}\otimes U_{h,s}} = \frac{1/N_h\, T_{h,s}}{1/N_h\, U_{h,s}} - \frac{1/N_h\, S_{h,s}\otimes 1/N_h\, S_{h,s}}{1/N_h\, U_{h,s}\;\; 1/N_h\, U_{h,s}}$$

We formulate the lemma

LEMMA 5.2.

$$N_h^{-1}\begin{bmatrix} \hat{p}_{hs}-p_{hs} \\ \hat{\mu}_{hs}-\mu_{hs} \\ \mathrm{vec}(\hat{\Sigma}_{hs})-\mathrm{vec}(\Sigma_{hs}) \\ \hat{p}_{ht}-p_{ht} \\ \hat{\mu}_{ht}-\mu_{ht} \\ \mathrm{vec}(\hat{\Sigma}_{ht})-\mathrm{vec}(\Sigma_{ht}) \end{bmatrix} \overset{L}{\to} N_{2(1+p+p^2)}\left(0,\begin{pmatrix} B_{h,ss} & B_{h,st} \\ B_{h,ts} & B_{h,tt}\end{pmatrix}\right)$$

where

$$B_{h,ss} = \begin{bmatrix} p_{hs}(1-p_{hs}) & 0 & 0 \\ 0 & \frac{1}{p_{hs}}\Sigma_{hs} & 0 \\ 0 & 0 & \frac{1}{p_{hs}}(I+K_p)(\Sigma_{hs}\otimes\Sigma_{hs}) \end{bmatrix}$$

and

$$B_{h,ts} = \begin{bmatrix} -p_{ht}p_{hs} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

while $B_{h,st} = (B_{h,ts})^T$ and $B_{h,tt}$ is $B_{h,ss}$ with s replaced by t.

<u>PROOF</u>. The lemma can be proved with lemma 5.1 and lemma 4.1. So we need the matrix ∇_h of partial derivatives in the point $\eta_h = (p_{hs},\mu_{hs}^T,\text{vec}^T(\Sigma_{hs}), p_{ht},\mu_{ht}^T,\text{vec}^T(\Sigma_{ht}))^T$. Write

$$\nabla_h = \begin{bmatrix} \nabla_{h,ss} & \nabla_{h,st} \\ \nabla_{h,ts} & \nabla_{h,tt} \end{bmatrix}$$

then $\nabla_{h,ss}$ is $(\hat{p}_{hs},\hat{\mu}_{hs}^T,\text{vec}^T(\hat{\Sigma}_{hs}))$ differentiated to

$$\left(\frac{\partial}{\partial(1/N_h\ U_{hs})},\left(\frac{\partial}{\partial(1/N_h\ S_{hs})}\right)^T,\left(\frac{\partial}{\partial(1/N_h\ T_{hs})}\right)^T\right)^T$$

in the point $(p_{hs},\mu_{hs}^T,\text{vec}^T(\Sigma_{hs}))$:

$$\nabla_{h,ss} = \begin{pmatrix} 1 & -\frac{1}{p_{hs}}\mu_{hs}^T & -\frac{1}{p_{hs}}\text{vec}^T(\Sigma_{hs}) + \frac{1}{p_{hs}}\mu_{hs}^T\otimes\mu_{hs}^T \\ 0 & \frac{1}{p_{hs}}I_p & -\frac{1}{p_{hs}}(\mu_{hs}^T\otimes I_p) - \frac{1}{p_{hs}}(I_p\otimes\mu_{hs}^T) \\ 0 & 0 & \frac{1}{p_{hs}}I_{p^2} \end{pmatrix}$$

$\nabla_{h,tt}$ is $\nabla_{h,ss}$ with s replaced by t and it is easy to see that $\nabla_{h,ts} = \nabla_{h,st} = 0$. With use of the properties presented in section 3 the computation is straightforward.

6. CASE I. NO ASSUMPTION ABOUT HOMOGENEITY OF VARIANCE-COVARIANCE MATRICES

In order to derive the asymptotic distribution of $N^{\frac{1}{2}}(R_{\cdot|(x,j)}-\rho_{\cdot|(x,j)})$, in which (x,j) is the observation vector, it is easiest first to find out the asymptotic distribution of

$$N_h^{\frac{1}{2}}(\hat{p}_{hj}\hat{f}_{hj}(x) - p_{hj}f_{hj}(x)).$$

Remember that

$$\hat{f}_{hj}(x) = \frac{1}{(2\pi)^{p/2}} |\hat{\Sigma}_{hj}|^{-\frac{1}{2}} \exp(-\tfrac{1}{2}(x-\hat{\mu}_{hj})^T\hat{\Sigma}_{hj}^{-1}(x-\hat{\mu}_{hj}))$$

where $\hat{\mu}_{hj}$ and $\hat{\Sigma}_{hj}$ have been defined in section 2. Application of the δ-method requires that we need

$$\frac{\partial\hat{p}_{hj}\hat{f}_{hj}(x)}{\partial\hat{p}_{hj}}(\eta) = f_{hj}(x)$$

$$(6.1)\qquad \frac{\partial\hat{p}_{hj}\hat{f}_{hj}(x)}{\partial\hat{\mu}_{hj}}(\eta) = p_{hj}f_{hj}(x)\Sigma_{hj}^{-1}(x-\mu_{hj}) \quad \text{and}$$

$$\frac{\partial\hat{p}_{hj}\hat{f}_{hj}(x)}{\partial\mathrm{vec}(\hat{\Sigma}_{hj})}(\eta) = p_{hj}f_{hj}(x)\{-\tfrac{1}{2}\mathrm{vec}(\Sigma_{hj}^{-1})+\tfrac{1}{2}(\Sigma_{hj}^{-1}(x-\mu_{hj}))\otimes(\Sigma_{hj}^{-1}(x-\mu_{hj}))\}$$

where $\eta = (p_{hj},\mu_{hj}^T,\mathrm{vec}^T(\Sigma_{hj}))^T$.

Now with lemma 5.2 in which it is sufficient only to consider that component of the random vector which has $B_{h,jj}$ as the asymptotic variance, we obtain:

<u>LEMMA 6.1.</u>

$$N_h^{\frac{1}{2}}(\hat{p}_{hj}\hat{f}_{hj}(x)-p_{hj}f_{hj}(x)) \xrightarrow{L} N(0,(p_{hj}(1-p_{hj})+\tfrac{1}{2}p_{hj}(p+\Delta_{x;hj}^4))f_{hj}^2(x))$$

<u>PROOF</u>. With the remarks preceding lemma 6.1 we compute immediately the variance

$$p_{hj}(1-p_{hj})f_{hj}^2(x) + p_{hj}f_{hj}^2(x)\Delta_{x;hj}^2 + \tfrac{1}{4}p_{hj}f_{hj}^2(x)\Big[\Big\{-\mathrm{vec}^T(\Sigma_{hj}^{-1}) +$$

$$+(\Sigma_{hj}^{-1}(x-\mu_{hj}))^T\otimes(\Sigma_{hj}^{-1}(x-\mu_{hj}))^T\Big\}(I+K_p)(\Sigma_{hj}\otimes\Sigma_{hj})\Big\{-\mathrm{vec}(\Sigma_{hj}^{-1}) +$$

$$(\Sigma_{hj}^{-1}(x-\mu_{hj}))\otimes(\Sigma_{hj}^{-1}(x-\mu_{hj}))\Big\}\Big].$$

For the terms between the square brackets we obtain after the cross-multiplication as first term:

$$\mathrm{vec}^T(\Sigma_{hj}^{-1})(I+K_p)(\Sigma_{hj}\otimes\Sigma_{hj})\mathrm{vec}(\Sigma_{hj}^{-1})$$

now, using $(\Sigma_{hj}\otimes\Sigma_{hj})\mathrm{vec}(\Sigma_{hj}^{-1}) = \mathrm{vec}(\Sigma_{hj})$

$$K_p\mathrm{vec}(\Sigma_{hj}) = \mathrm{vec}(\Sigma_{hj})$$

and

$$\mathrm{vec}^T(\Sigma_{hj}^{-1})\mathrm{vec}(\Sigma_{hj}) = \mathrm{tr}(\Sigma_{hj}^{-1}\Sigma_{hj}) = \mathrm{tr}(I) = p,$$

this first term becomes $2p$. As second term we get

$$-\mathrm{vec}^T(\Sigma_{hj}^{-1})(I+K_p)(\Sigma_{hj}\otimes\Sigma_{hj})\{(\Sigma_{hj}^{-1}(x-\mu_{hj}))\otimes(\Sigma_{hj}^{-1}(x-\mu_{hj}))\},$$

by using

$$K_p((x-\mu_{hj})\otimes(x-\mu_{hj})) = K_p\mathrm{vec}((x-\mu_{hj})(x-\mu_{hj})^T) =$$
$$= \mathrm{vec}((x-\mu)(x-\mu)^T) = (x-\mu)\otimes(x-\mu)$$

this term becomes

$$-2\mathrm{vec}^T(\Sigma_{hj}^{-1})\{(x-\mu_{hj})\otimes(x-\mu_{hj})\} = -2\mathrm{vec}^T((x-\mu_{hj})\Sigma_{hj}^{-1}(x-\mu_{hj})) =$$
$$= -2\Delta_{x;hj}^2.$$

With the same technique we find for the thirth term $-2\Delta_{x;hj}^2$ and for the fourth $2\Delta_{x;hj}^4$. Together we obtain for the variance

$$p_{hj}(1-p_{hj})f_{hj}^2(x) + p_{hj}f_{hj}^2(x)\Delta_{x;hj}^2 + \tfrac{1}{2}p_{hj}f_{hj}^2(x)\{p-2\Delta_{x;hj}^2 + \Delta_{x;hj}^4\}$$

which is equal to $\{p_{hj}(1-p_{hj}) + \frac{1}{2}p_{hj}(p+\Delta_{x;hj}^4)\}f_{hj}^2(x)$. This finishes the proof.

If we extend our considerations to all k populations, remembering that $N = N_1 + \ldots + N_k$ and $b_1 = \frac{N_1}{N}, \ldots, b_k = \frac{N_k}{N}$ as introduced in section 2, we obtain

$$N^{\frac{1}{2}}\begin{bmatrix}\hat{p}_{1j}\hat{f}_{1j}(x) - p_{1j}f_{1j}(x)\\ \vdots \\ \hat{p}_{kj}\hat{f}_{kj}(x) - p_{kj}f_{kj}(x)\end{bmatrix} \xrightarrow{L} N_k(0,D)$$

where D is a diagonal matrix with

$$D_h = \frac{1}{b_h}\{p_{hj}(1-p_{hj}) + \tfrac{1}{2}p_{hj}(p+\Delta^4_{x;hj})\}f^2_{hj}(x)$$

Once again using the δ-method we get for the estimator $R_{\cdot|(x,j)}$ for $\rho_{\cdot|(x,j)}$ that

$$N^{\frac{1}{2}}(R_{\cdot|(x,j)}-\rho_{\cdot|(x,j)}) \xrightarrow{L} N_k(0,\Psi^T\Lambda\Psi)$$

in which

(6.2)
$$\Psi_{t,t} = \tfrac{1}{2}\rho_{t|(x,j)}(1-\rho_{t|(x,j)})$$
$$\Psi_{t,s} = -\tfrac{1}{2}\rho_{s|(x,j)}\rho_{t|(x,j)} \qquad s \neq t$$

and Λ a diagonal matrix with

$$\Lambda_h = \frac{1}{b_h p_{hj}}\{4(1-p_{hj}) + 2(p+\Delta^4_{x;hj})\}$$

7. CASE II. ASSUMPTION: $\Sigma_{1j} = \ldots = \Sigma_{kj} (=\Sigma_j)$

In this case the variance-covariance matrix Σ_j is estimated by

$$\hat{\Sigma}_j = \frac{1}{N_{1j}+\ldots+N_{kj}} \sum_{h=1}^{k}\sum_{i=1}^{N_h} (X_{hi}-\hat{\mu}_{hj})(X_{hi}-\hat{\mu}_{hj})^T I(D_{hi}=j)$$

$$= \sum_{h=1}^{k} \frac{N_{hj}}{N_{1j}+\ldots+N_{kj}}\hat{\Sigma}_{hj} = \frac{1}{\hat{p}_{1j}b_1+\ldots+\hat{p}_{kj}b_k}\sum_{h=1}^{k} b_h\hat{p}_{hj}\hat{\Sigma}_{hj}.$$

First we formulate a lemma

<u>LEMMA 7.1.</u> *If* $\hat{A}_m$ *are symmetric positive definite random* p×p *matrices with*

$$Lm^{\frac{1}{2}}(\text{vec}(\hat{A}_m)-\text{vec}(A)) \to N(0,M)$$

where A *is also symmetric positive definite then*

$$Lm^{\frac{1}{2}}(\mathrm{vec}(\hat{A}_m^{-1})-\mathrm{vec}(A^{-1})) \to N(0,(A^{-1}\otimes A^{-1})M(A^{-1}\otimes A^{-1})).$$

PROOF. By noting that

$$\frac{\partial a^{ij}}{\partial a_{\alpha\beta}} = -a^{i\alpha}a^{\beta j} \quad \text{or} \quad \frac{\mathrm{vec}^T(A^{-1})}{\mathrm{vec}(A)} = -A^{-1}\otimes A^{-1},$$

where we used the notation $a_{ij} = A_{ij}$ and $a^{ij} = (A^{-1})_{ij}$, the lemma follows directly with the δ-method.

The theorem of section 5 and the independence between the samples for the k populations gives

$$N^{\frac{1}{2}}\begin{bmatrix} \hat{p}_{1j}-p_{1j} \\ \hat{\mu}_{1j}-\mu_{1j} \\ \mathrm{vec}(\hat{\Sigma}_{1j})-\mathrm{vec}(\Sigma_j) \\ \cdot \\ \cdot \\ \cdot \\ \hat{p}_{kj}-p_{kj} \\ \hat{\mu}_{kj}-\mu_{kj} \\ \mathrm{vec}(\hat{\Sigma}_{kj})-\mathrm{vec}(\Sigma_j) \end{bmatrix} \xrightarrow{L} N(0,T)$$

where T is a block-diagonal matrix with on the diagonal the blocks

$$\{\frac{1}{b_h}p_{hj}(1-p_{hj}),\frac{1}{b_h}\frac{1}{p_{hj}}\Sigma_j,\frac{1}{b_h}\frac{1}{b_{hj}}(I+K_p)(\Sigma_j\otimes\Sigma_j)\}, \quad h = 1,\dots,k$$

The partial derivatives of $\mathrm{vec}^T(\hat{\Sigma}_j)$ with respect to $\hat{p}_{tj}$, $\hat{\mu}_{tj}$ and $\mathrm{vec}(\hat{\Sigma}_{tj})$ in $\eta_t = (p_{tj},\mu_{tj}^T,\mathrm{vec}^T(\Sigma_j))^T$ for $t = 1,\dots,k$ are

$$\frac{\partial \mathrm{vec}^T(\hat{\Sigma}_j)}{\partial \hat{p}_{tj}}(\eta_t) = 0, \quad \frac{\partial \mathrm{vec}^T(\hat{\Sigma}_j)}{\partial \hat{\mu}_{tj}}(\eta_t) = 0$$

$$\frac{\partial \mathrm{vec}^T(\hat{\Sigma}_j)}{\partial \mathrm{vec}(\hat{\Sigma}_{tj})}(\eta_t) = \frac{b_t p_{tj}}{b_1 p_{1j}+\dots+b_k p_{kj}} I_{p^2}.$$

Thus the δ-method gives

$$N^{\frac{1}{2}}(\mathrm{vec}(\hat{\Sigma}_j)-\mathrm{vec}(\Sigma_j)) \xrightarrow{L} N(0,\frac{1}{(\Sigma_{h=1}^{k} b_h p_{hj})}(I+K_p)(\Sigma_j \otimes \Sigma_j)).$$

Using the lemma of this section we get

$$N^{\frac{1}{2}}(\mathrm{vec}(\hat{\Sigma}_j^{-1})-\mathrm{vec}(\Sigma_j^{-1})) \xrightarrow{L} N(0,\frac{1}{(\Sigma_{h=1}^{k} b_h p_{hj})}(I+K_p)(\Sigma_j^{-1} \otimes \Sigma_j^{-1}))$$

which is, as can easily be seen, asymptotically independent of the estimators $\hat{p}_{hj}$, $\hat{\mu}_{hj}$, $h = 1,\ldots,k$. Hence

$$N^{\frac{1}{2}}\begin{bmatrix} \hat{p}_{1j}-\hat{p}_{1j} \\ \hat{\mu}_{1j}-\hat{\mu}_{1j} \\ \cdot \\ \cdot \\ \cdot \\ \hat{p}_{kj}-p_{kj} \\ \hat{\mu}_{kj}-\mu_{kj} \\ \mathrm{vec}(\hat{\Sigma}_j^{-1})-\mathrm{vec}(\Sigma_j^{-1}) \end{bmatrix} \xrightarrow{L} N(0,V).$$

where V is a block-diagonal matrix with on the diagonal the blocks

$$\text{(7.1)}\qquad \begin{gathered}\{(\frac{1}{b_h}p_{hj}(1-p_{hj}),\frac{1}{b_h}\frac{1}{b_{hj}}\Sigma_j)h=1,\ldots,k\},\\ \frac{1}{(\Sigma_{h=1}^{k} b_h p_{hj})}(I+K_p)(\Sigma_j^{-1}\otimes\Sigma_j^{-1}).\end{gathered}$$

For the partial derivatives of the estimator for the Mahalanobis squared distance $\Delta^2_{x;hj}$ with respect to $\hat{\mu}_{tj}$ and $\mathrm{vec}(\hat{\Sigma}_j^{-1})$ for $h,t = 1,\ldots,k$ in the point $\eta = (p_{1j},\mu_{1j},\ldots,p_{kj},\mu_{kj},\mathrm{vec}^T(\Sigma_j^{-1}))^T$ we get

$$\frac{\partial\hat{\Delta}^2_{x;hj}}{\partial\hat{\mu}_{tj}}(\eta) = -2\Sigma_j^{-1}(x-\mu_{hj})\delta_{ht}$$

$$\frac{\partial\hat{\Delta}^2_{x;hj}}{\partial\mathrm{vec}(\hat{\Sigma}_j^{-1})}(\eta) = (x-\mu_{hj})\otimes(x-\mu_{hj})$$

where δ_{ht} is de KRonecker delta notation.
Again applying the δ-method gives:

$$N^{\frac{1}{2}}\begin{bmatrix} \hat{p}_{1j}-p_{1j} \\ \cdot \\ \cdot \\ \cdot \\ \hat{p}_{kj}-p_{kj} \\ \hat{\Delta}^2_{x;1j}-\Delta^2_{x;1j} \\ \cdot \\ \cdot \\ \cdot \\ \hat{\Delta}^2_{x;kj}-\Delta^2_{x;kj} \end{bmatrix} \xrightarrow{L} N_{2k}\left(0,\begin{pmatrix} \Gamma_p & \Gamma_0 \\ \Gamma_0 & \Gamma_\Delta \end{pmatrix}\right).$$

Where

$$\Gamma_{p;tt} = \frac{1}{b_t}\, p_{tj}(1-p_{tj})$$

$$\Gamma_{p;ts} = 0 \quad t \neq s$$

and

$$\Gamma_{\Delta;tt} = \frac{4}{b_t p_{tj}}\, \Delta^2_{x;tj} + \frac{2}{(\Sigma^k_{h=1} b_h p_{hj})}\, \Delta^4_{x;tj}$$

$$\Gamma_{\Delta;ts} = \frac{2}{(\Sigma^k_{h=1} b_h p_{hj})}\, \{(x-\mu_{tj})^T \Sigma_j^{-1}(x-\mu_{sj})\}^2 \quad t \neq s$$

and $\Gamma_0 = \underline{0}$.
Now, note that for $t = 1,\ldots,k$

$$R_{t|(x,j)} = \frac{\rho_t \hat{p}_{tj} \exp(-\frac{1}{2}\hat{\Delta}^2_{x;tj})}{\Sigma^k_{h=1} \rho_h \hat{p}_{hj} \exp(-\frac{1}{2}\hat{\Delta}^2_{x;hj})}$$

and

$$\rho_{t|(x,j)} = \frac{\rho_t p_{tj} \exp(-\frac{1}{2}\Delta^2_{x;tj})}{\Sigma^k_{h=1} \rho_h p_{hj} \exp(-\frac{1}{2}\Delta^2_{x;hj})}$$

We have the following partial derivatives:

$$\frac{\partial R_{t|(x,j)}}{\partial \hat{p}_{tj}}(\eta) = \frac{1}{p_{tj}}\rho_{t|(x,j)}(1-\rho_{t|(x,j)})$$

$$\frac{\partial R_{t|(x,j)}}{\partial \hat{p}_{sj}}(\eta) = \frac{-1}{p_{sj}}\rho_{s|(x,j)}\rho_{t|(x,j)} \qquad s \neq t$$

and

$$\frac{\partial R_{t|(x,j)}}{\partial \hat{\Delta}^2_{x;tj}}(\eta) = -\tfrac{1}{2}\rho_{t|(x,j)}(1-\rho_{t|(x,j)})$$

$$\frac{\partial R_{t|(x,j)}}{\partial \hat{\Delta}^2_{x;sj}}(\eta) = +\tfrac{1}{2}\rho_{s|(x,j)}\rho_{t|(x,j)} \qquad s \neq t$$

So we obtain

$$N^{\frac{1}{2}}(R_{\cdot|(x,j)}-\rho_{\cdot|(x,j)}) \xrightarrow{L} N(0,\Psi\Lambda\Psi).$$

where Ψ is defined in the previous section and Λ is defined by

$$\Lambda_{tt} = \frac{4}{b_t p_{tj}}(1-p_{tj}) + \frac{4}{b_t p_{tj}}\Delta^2_{x;tj} + \frac{2}{(\Sigma^k_{h=1} b_h p_{hj})}\Delta^4_{x;tj}$$

(7.2)

$$\Lambda_{t,s} = \frac{2}{(\Sigma^k_{h=1} b_h p_{hj})}\{(x-\mu_{tj})^T\Sigma_j^{-1}(x-\mu_{sj})\}^2 \qquad t \neq s$$

8. CASE III. ASSUMPTION $\Sigma_{11} = \ldots = \Sigma_{kd}(=\Sigma)$

In this section we assume that the variance-covariance matrices for each population and for each possible outcome of the discrete random variable are the same, namely Σ.

As estimator for Σ we use

$$\hat{\Sigma} = \frac{1}{N}\sum_{h=1}^{k}\sum_{\ell=1}^{d}\sum_{i=1}^{N_h}(X_{hi}-\hat{\mu}_{h\ell})(X_{hi}-\hat{\mu}_{h\ell})I(D_{hi}=\ell)$$

$$= \sum_{h=1}^{k} b_h \sum_{\ell=1}^{d}\hat{p}_{h\ell}\hat{\Sigma}_{h\ell}$$

where $\hat{\Sigma}_{h\ell}$ is defined in section 2. Write $\hat{\Psi}_h = \sum_{\ell=1}^{d} \hat{p}_{h\ell}\hat{\Sigma}_{h\ell}$ and so

$$\hat{\Sigma} = \sum_{h=1}^{k} b_h \hat{\Psi}_h .$$

For the partial derivatives of $\hat{p}_{hs}$ and $\mathrm{vec}^T(\hat{\Psi}_h)$ with respect to $\hat{p}_{ht}$ and $\mathrm{vec}(\hat{\Sigma}_{ht})$ in $\eta_h = (p_{h1},\dots,p_{hd},\mathrm{vec}^T(\Sigma))^T$ we obtain

$$\frac{\partial \hat{p}_{hs}}{\partial \hat{p}_{ht}}(\eta_s) = \delta_{st} \qquad \frac{\partial \hat{p}_{hs}}{\partial \mathrm{vec}(\hat{\Sigma}_{ht})}(\eta_h) = 0$$

$$\frac{\partial \mathrm{vec}^T(\hat{\Psi}_h)}{\partial \hat{p}_{hs}}(\eta_h) = \mathrm{vec}^T(\Sigma_{hs}) \qquad \frac{\partial \mathrm{vec}^T(\hat{\Psi}_h)}{\partial \mathrm{vec}(\hat{\Sigma}_{hs})}(\eta_h) = p_{hs} I_{p^2} .$$

Application of the theorem and lemma of section 5 gives

$$N_h^{\frac{1}{2}} \begin{bmatrix} \hat{p}_{hs} - p_{hs} \\ \hat{\mu}_{hs} - \mu_{hs} \\ \mathrm{vec}(\hat{\Psi}_h) - \mathrm{vec}(\Sigma) \end{bmatrix} \xrightarrow{L} N\left(0, \begin{bmatrix} p_{hs}(1-p_{hs}) & 0 & 0 \\ 0 & \frac{1}{p_{hs}}\Sigma & 0 \\ 0 & 0 & (I+K_p)(\Sigma\otimes\Sigma) \end{bmatrix}\right).$$

For the asymptotic distributions dealing with $\hat{\Sigma}$ and $\hat{\Sigma}^{-1}$ we obtain in an easy way

$$N^{\frac{1}{2}}(\mathrm{vec}(\hat{\Sigma}) - \mathrm{vec}(\Sigma)) \to N_{p^2}(0,(I+K_p)(\Sigma\otimes\Sigma))$$

and

$$N^{\frac{1}{2}}(\mathrm{vec}(\hat{\Sigma}^{-1}) - \mathrm{vec}(\Sigma^{-1})) \to N_{p^2}(0,(I+K_p)(\Sigma^{-1}\otimes\Sigma^{-1})),$$

where we have used the lemma of section 7 for the last transition.

So we obtain

$$N^{\frac{1}{2}}\begin{bmatrix} \hat{p}_{1j}-p_{1j} \\ \hat{\mu}_{1j}-\mu_{1j} \\ \cdot \\ \cdot \\ \cdot \\ \hat{p}_{kj}-p_{kj} \\ \hat{\mu}_{kj}-\mu_{kj} \\ \mathrm{vec}(\hat{\Sigma}^{-1})-\mathrm{vec}(\Sigma^{-1}) \end{bmatrix} \xrightarrow{L} N(0,W)$$

where W is a block-diagonal matrix with on the diagonal the blocks

$$\{(\frac{1}{b_h}p_{hj}(1-p_{hj}), \frac{1}{b_h}\frac{1}{b_{hj}}\Sigma)\ h=1,\ldots,k\},\quad (I+K_p)(\Sigma^{-1}\otimes\Sigma^{-1}).$$

Note that we have now obtained a formula similar to formula (7.1). Since from this point on the computation are the same we can directly use formula (7.2) by replacing Σ_j by Σ and $\Sigma_{h=1}^{k} b_h p_{hj}$ by 1.
Hence

$$N^{\frac{1}{2}}(R_{\cdot|(x,j)}-\rho_{\cdot|(x,j)}) \xrightarrow{L} N(0,\Psi^T\Lambda\Psi)$$

where Ψ is defined in section 6 and Λ by

$$\Lambda_{t,t} = \frac{4}{b_t p_{tj}}(1-p_{tj}) + \frac{4}{b_t p_{tj}}\Delta^2_{x;tj} + 2\Delta^4_{x;tj}$$

$$\Lambda_{t,s} = 2\{(x-\mu_{tj})^T\Sigma^{-1}(x-\mu_{sj})\}^2 \quad t \neq s.$$

9. CASE IV. ASSUMPTION $\Sigma_{h1} = \ldots = \Sigma_{hd} (=\Sigma_h)$, $h = 1,\ldots,k$

Here we have

$$\hat{\Sigma}_h = \frac{1}{N_h}\sum_{\ell=1}^{d}\sum_{i=1}^{N_h}(X_{hi}-\hat{\mu}_{h\ell})(X_{hi}-\hat{\mu}_{h\ell})^T I(D_{hi}=\ell)$$

$$= \sum_{\ell=1}^{d}\hat{p}_{h\ell}\hat{\Sigma}_{h\ell} = \hat{\Psi}_h, \quad \text{defined in section 8.}$$

Therefore we can use the result of expression 8 if we replace $\hat{\Psi}_h$ by $\hat{\Sigma}_h$:

$$N_h^{\frac{1}{2}}\begin{bmatrix} \hat{p}_{hs}-p_{hs} \\ \hat{\mu}_{hs}-\mu_{hs} \\ \mathrm{vec}(\hat{\Sigma}_h)-\mathrm{vec}(\Sigma_h)\end{bmatrix} \xrightarrow{L} N(0,\begin{bmatrix} p_{hs}(1-p_{hs}) & 0 & 0 \\ 0 & \frac{1}{p_{hs}}\Sigma_h & 0 \\ 0 & 0 & (I+K_p)(\Sigma_h\otimes\Sigma_h)\end{bmatrix}).$$

If we now look back to the beginning of section 6 we see that the partial derivatives presented in formula (6.1) are also valid for the case of this section if we replace Σ_{hj} by Σ_h.
In similar way as in lemma 6.1 we obtain that

$$N_h^{\frac{1}{2}}(\hat{p}_{hj}\hat{f}_{hj}(x)-p_{hj}f_{hj}(x)) \xrightarrow{L} N(0,\{p_{hj}(1-p_{hj})+\tfrac{1}{2}p_{hj}^2(p+\Delta_{x;hj}^4)+ \\ + p_{hj}(1-p_{hj})\Delta_{x;hj}^2\}f_{hj}^2(x))$$

and hence

$$N^{\frac{1}{2}}(R_{\cdot|(x,j)}-\rho_{\cdot|(x,j)}) \xrightarrow{L} N(0,\Psi^T\Lambda\Psi)$$

where Ψ is defined in formula (6.2) and Λ is diagonal with

$$\Lambda_h = \frac{4(1-p_{hj})}{b_h p_{hj}}(1+\Delta_{x;hj}^2) + \frac{2(p+\Delta_{x;hj}^4)}{b_h} .$$

REFERENCES

[1] AITCHISON, J., J.D.F. HABBEMA & J.W. KAY (1977), *A critical comparison of two methods of statistical discrimination*. Appl. Statist., 26, no. 1, pp. 15-25.

[2] AMBERGEN, A.W. (1981), *Approximate confidence intervals for posterior probabilities*. Report TW-224. Depart. of Math., Postbox 800, Groningen.

[3] AMBERGEN, A.W. & W. SCHAAFSMA (1982), *The asymptotic variance of estimators for posterior probabilities*. Report SW 86/82, CWI, postbox 4079, Amsterdam.

[4] AMBERGEN, A.W. & W. SCHAAFSMA (1983[a]), *Interval estimates for posterior probabilities, applications to Border Cave.* In Multivariate Statistical Methods, (G.N. van Vark & W.W. Howells, Ed.), Reidel, Dordrecht, The Netherlands.

[5] AMBERGEN, A.W. & W. SCHAAFSMA (1983[b]), *Interval estimates for posterior probabilities in a multivariate normal classification model.* Report SW 96/83, CWI, postbox 4079, Amsterdam.

[6] AMBERGEN, A.W. & W. SCHAAFSMA (1984), *Interval estimates for posterior probabilities in a multivariate normal classification model,* To be published in Journal of Multivariate Analysis.

[7] GEISSER, S. (1964), *Posterior odds for Multivariate Normal Classification.* J. Roy. Statist. Soc. Ser. B26, pp. 69-76.

[8] MAGNUS, J.R. & H. NEUDECKER (1979), *The commutation matrix: some properties and applications.* The Ann. of Stat., Vol. 7, No. 2, pp. 381-394.

[9] McLACHLAN, G.L. (1977), *The bias of sample based posterior probabilities.* Biom. J. vol. 19, no. 6, pp. 421-426.

[10] McLACHLAN, G.L. (1979), *A comparison of the estimative and predictive methods of estimating posterior probabilities.* Comm. Statist. A-Theory Methods, 8(9), pp. 919-929.

[11] KRZANOWSKI, W.J. (1975), *Discrimination and classification using both binary and continuous variables.* J.A.S.A., Vol. 70, pp. 782-790.

[12] RIGBY, R.A. (1982), *A credibility interval for the probability that a new observation belongs to one of two multivariate normal populations.* J.R. Statist. Soc. B, 44, no. 2, pp. 212-220.

[13] SCHAAFSMA, W. (1982), *Selecting variables in discriminant analysis for improving upon classical procedures.* Kanal, L. and Krishnaiah, P.R. (eds.). Handbook of Statistics 2. North Holland, Amsterdam.

[14] SCHAAFSMA, W. & G.N. VAN VARK (1977), *Classification and discrinination problems with applications, part I,* Statistica Neerlandica 31, 25-45.

[15] SCHAAFSMA, W. & G.N. VAN VARK (1979), *Classification and discrimination problems with applications, part II^a,* Statistica Neerlandica 33, 91-126.

[16] SERFLING, R.J. (1980), *Approximation theorems of mathematical statistics.* Wiley, New York.

[17] VAN VARK, G.N. & W.W. HOWELLS (1984), *Multivariate Statistical Methods in Physical Anthropology.* Reidel, Dordrecht, The Netherlands.