CHAPTER XI

THEORY OF DISCRIMINANT ANALYSIS OF THE INCREASING NUMBER OF INDEPENDENT VARIABLES

Problem Setting

We consider a sequence $\mathfrak{P} = \{\mathfrak{P}_n\}$ of the discrimination analysis problems

$$\left(p, k, l, f(\mathbf{x}, \theta), \theta_1, \theta_2, \widehat{\theta}_1, \widehat{\theta}_2, w(\mathbf{x}), \alpha_1, \alpha_2\right)_n, \quad n = 1, 2, \ldots, \tag{1}$$

where the arguments are as follows (we do not write out the subscripts n for arguments of (1)). The vectors $\mathbf{x}$ are observed in two populations $\mathfrak{S}_1$ and $\mathfrak{S}_2$ which are defined by a parametric family of distribution densities $f(\mathbf{x}, \theta)$ with respect to a sigma-finite measure $\mu(\mathbf{x})$ and vectors $\theta = \theta_1$ and $\theta = \theta_2$. The vectors $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are estimators of $\theta_1,\ \theta_2 \in \mathbb{R}^l$ over independent samples of size $n_1 = n_2 = n$ from populations $\mathfrak{S}_1$ and $\mathfrak{S}_2$. For convenience the index n enumerating problems coincides with the sample size. A discriminant function $w(\mathbf{x})$ is calculated using the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$, and the discrimination rule is used $w(\mathbf{x}) > c$ against $w(\mathbf{x}) \leq c$, where c is an a priori threshold. The random values

$$\alpha_1 = \int\limits_{w(\mathbf{x}) \leqslant c} f\left(\mathbf{x}, \theta_1\right) \mu\left(d\mathbf{x}\right), \quad \text{and} \quad \alpha_2 = \int\limits_{w(\mathbf{x}) > c} f\left(\mathbf{x}, \theta_2\right) \mu\left(d\mathbf{x}\right). \tag{2}$$

are (conditional) probabilities of the discrimination errors of two kinds.

Let us restrict $\mathfrak{P}$ with the following conditions that can lead to a theory of interest for applications.

Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX

A (block dependence of variables).

Let the components of vectors $\mathbf{x}$ and θ be partitioned into k non-intersecting subsets so that

$$\mathbf{x} = (\mathbf{x}^1, \ldots, \mathbf{x}^k), \quad \theta = (\theta^1, \ldots, \theta^k), \quad \text{and } f(\mathbf{x}, \theta) = \prod_{i=1}^{k} f^i(\mathbf{x}^i, \theta^i). \tag{3}$$

For each $i = 1, \ldots, k$, all vectors θ^i are m-dimensional, where $m = c_1$ does not depend on n, $l = km$.

This condition does restrict the dependence between a finite number of variables. In the general case we may expect that the assumption of block independence introduces an inaccuracy comparable with an average measure of dependence between variables.

B (uniform regularity of distribution densities).

We assume that the functions $f(\mathbf{x}, \theta)$ have the following properties.

B1. $f(\mathbf{x}, \theta)$ are positive and the functions $\ln f(\mathbf{x}, \theta)$ are thrice differentiable with respect to all components of θ.

B2. All three (and mixed) derivatives of $\ln f(\mathbf{x}, \theta)$ with respect to components of θ are not greater in absolute value than majorizing non-negative functions $\varphi(\mathbf{x})$ such that

$$\int [\varphi(\mathbf{x})]^4 f(\mathbf{x}, \theta) \; \mu(d\mathbf{x}) < c_2,$$

where c_2 does not depend on n. The functions $\varphi(\mathbf{x})$ can be different for different derivatives (we do not specify these and do not write out subscripts for $\varphi(\cdot)$).

B3. The information matrices

$$I(\theta) = \int \left(\frac{\partial}{\partial \theta} \ln f(\mathbf{x}, \theta) \; \frac{\partial^T}{\partial \theta} \ln f(\mathbf{x}, \theta) \right) f(\mathbf{x}, \theta) \, \mu(d\mathbf{x})$$

are positively definite for all θ and all eigenvalues of these are not less $c_3 > 0$, where c_3 does not depend on n.

In view of condition A the matrices $I(\theta)$ have a block diagonal structure with the blocks $I^i(\theta)$ of size $m \times m$ depending only on the subvectors θ^i, $i = 1, \ldots, k$.

However, the requirements of regularity can be softened. In [32] limit formulas are obtained without the assumption of the existence of the third derivatives. The derivatives are majorized only jointly for a set of variables. Some components of θ_1 and θ_2 may approach to singular points.

C (rate of the increase of the number of blocks).
The limit exists $\lim_{n\to\infty} k/n = \kappa > 0$.

D (uniform approaching of populations in the parameter space).

$$\max_{i=1,\ldots,k} n(\theta_1^i - \theta_2^i)^2 \leq c_4,$$

where c_4 does not depend on n (here and in the following, the squares of vectors denote the squares of their length).

In practice it is required that all blocks provide uniformly small contributions to the discrimination. The requirements of the approaching of populations can be mitigated. In [32] the approaching conditions have a general non-parametric form. Separate variables providing a contribution to the discrimination of the order of magnitude larger than n^{-1} can be taken into account explicitly by the integration.

Define

$$J^i = \int \ln \frac{f^i\left(\mathbf{x}^i, \theta_1^i\right)}{f^i\left(\mathbf{x}^i, \theta_2^i\right)} \left[f\left(\mathbf{x}, \theta_1\right) - f\left(\mathbf{x}, \theta_2\right)\right] \mu(d\mathbf{x}) \geq 0, \quad i = 1, \ldots, k,$$

$$J_n = \sum_{i=1}^{k} J^i. \tag{4}$$

Note that these integrals exist by condition B.

We introduce a description of the set $\{J^i\}$ by means of an empiric distribution function of the quantities $v^i = nJ^i/2, \quad i = 1, \ldots, k,$

$$R_n\left(v\right) = k^{-1}\sum_{i=1}^{n} \text{ind } (v \geq nJ^i/2).$$

The function $R_n(v)$ is defined so that, formally,

$$k^{-1}\sum_{i=1}^{k}\left(\cdot\right) = \int \left(\cdot\right) dR_n\left(v\right). \tag{5}$$

E (existence of a limit distribution for the set $\{J^i\}$). For each $v > 0$, the limit exists

$$\lim_{n\to\infty} R_n(v) = R(v).$$

This condition does not restrict applications of the present limit theory to finite-dimensional problems and is introduced in order to provide the convergence.

Under condition E, there exist the limit

$$J = \lim_{n\to\infty} J_n = 2\kappa \int v dR(v).$$

F (requirements to estimators).

The estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are assumed to be independent and have a standard collection of 'good' asymptotic properties that must hold uniformly with respect to blocks number $i = 1, \ldots, k$.

F1 (uniform asymptotic $\sqrt{n}$-unbiasedness).

$$\lim_{n\to\infty} \max_i \sqrt{n} \left| \mathbf{E}\, \widehat{\theta}^i_\nu - \theta^i_\nu \right| = 0, \quad \nu = 1, 2.$$

F2 (uniform asymptotic efficiency).

$$\lim_{n\to\infty} \max_i \left| n\mathbf{E}\, (\widehat{\theta}^i_\nu - \theta^i_\nu)^T I^i(\theta_\nu)(\widehat{\theta}^i_\nu - \theta^i_\nu) - m \right| = 0, \quad \nu = 1, 2.$$

F3 (uniform n^2-boundedness of the fourth momenta).

$$\max_i n^2\, \mathbf{E}\, |\widehat{\theta}^i_\nu - \theta_\nu|^4 < c_5, \quad \nu = 1, 2,$$

where c_5 does not depend on n.

We introduce random vectors

$$\mathbf{z}^i_\nu = n^{1/2} I^i(\theta^i_\nu)(\widehat{\theta}^i_\nu - \theta^i_\nu), \tag{6}$$

and their distribution functions $F^i_\nu(z), \quad i = 1, \ldots, k, \quad \nu = 1, 2.$

F4 (uniform asymptotic normality).

$$\lim_{n\to\infty} \max_i \sup_t |F_\nu^i(t) - \Phi_m(t)| = 0, \quad \nu = 1, 2,$$

where $\Phi_m(t)$ is the distribution function of $\mathbf{N}(0_m, I_m)$.

It is convenient to renormalize the estimators in order to obtain finite quantities. The conditions F1–F4 can be reformulated for the variables (6) as follows:

$$\begin{aligned}
&\lim_{n\to\infty} \max_i |\mathbf{E}\ \mathbf{z}_\nu^i| = 0,\\
&\lim_{n\to\infty} \max_i |\mathbf{E}\ (\mathbf{z}_\nu^i)^2 - m| = 0,\\
&\lim_{n\to\infty} \max_i \mathbf{E}\ |\mathbf{z}_\nu|^4 < c_6,\\
&\lim_{n\to\infty} \max_i \sup_t |F_\nu^i(t) - \Phi_m(t)| = 0,
\end{aligned} \tag{7}$$

$\nu = 1, 2$, where c_6 does not depend on n.

Let us specify the discrimination rules. We consider the statistics

$$\widehat{J}^i = \int \ln \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)}\ [f(\mathbf{x}, \widehat{\theta}_1) - f(\mathbf{x}, \widehat{\theta}_2)]\ \mu(d\mathbf{x}) \geq 0, \quad i = 1, \ldots, k,$$

$$\widehat{J}_n = \sum_{i=1}^k \widehat{J}^i \tag{8}$$

(the integrals exist by assumptions B1 and B2).

Let $\eta(v)$ be an a priori weighting function for contributions of blocks to the discriminant function. Suppose that $\eta(v) \in \mathfrak{K}$, where $\mathfrak{K}$ is a class of functions of bounded variation on $[0, \infty)$ and continuous everywhere except, perhaps, of a finite number of discontinuity points which do not coincide with the discontinuity points of $R(v)$.

We consider discriminant functions of two forms:

$$g(\mathbf{x}) = \sum_i^k \eta(v^i)\ \ln\ \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)}, \tag{9}$$

where $v^i = nJ^i/2$, with J^i of the form (4), $\quad i = 1, \ldots, k$, and

$$g(\mathbf{x}) = \sum_i^k \eta(u^i)\ \ln\ \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)}, \tag{10}$$

where $u^i = n\widehat{J}^i/2$, with $\widehat{J}^i$ of the form (8), $i = 1, \ldots, k$.

Example 1. Let $\mathbf{x}$ be normal vectors with $p = k$ independent components x_i having unit variances, $m = 1$. Then

$$f(\mathbf{x}, \theta) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}} \exp\left(-(x_i - \theta^i)^2/2\right).$$

The derivatives of $\ln f(\mathbf{x}, \theta)$ satisfy the requirements B2 to the majorants $\varphi(\mathbf{x}) = \text{const}\ (1 + |x_i|)$, $i = 1, \ldots, k$. The information matrix is the identity matrix. If the estimators for θ_1 and θ_2 are sample means then conditions F are satisfied. We have $J^i = (\theta_2^i - \theta_1^i)^2$ and $\widehat{J}^i = (\widehat{\theta}_2^i - \widehat{\theta}_1^i)^2$, $i = 1, \ldots, k$. The discriminant function (10) is

$$w(\mathbf{x}) = \sum_{i}^{k} \eta(n\widehat{J}^i/2)(\widehat{\theta}_1^i - \widehat{\theta}_2^i)^T\ (x_i - \frac{\widehat{\theta}_1^i + \widehat{\theta}_2^i}{2}).$$

Example 2. Random vectors $\mathbf{x}$ with $k = p$ independent components x_i that are normally distributed as $\mathbf{N}(\theta_\nu^i, d^i)$, $\nu = 1, 2$, $d^i \geq c_7 > 0$, $i = 1, \ldots, k$ so that $m = 2$ and

$$f(\mathbf{x}, \theta_\nu) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi}} \exp\left(-(x^i - \theta_\nu^i)^2/2d^i\right), \quad \nu = 1, 2.$$

In this case

$$J^i = (\theta_1^i - \theta_2^i)^2(1/d_1^i - 1/d_2^i)/2 + (d_1^i - d_2^i)^2/(2d_1^i d_2^i), \quad i = 1, \ldots, k.$$

For the standard estimators of θ_ν (sample means and sample variances), it can be readily seen that conditions F are satisfied. The discriminant function $w(\mathbf{x})$ is quadratic with respect to components of the vectors $\mathbf{x}$.

In this Chapter our first purpose is to show that under conditions A–F the probabilities of the discrimination errors converge as $n \to \infty$. The limit values are expected to be functions of the following characteristics:

- m — the number of parameters in a block;
- κ — limit ratio of the number of blocks to the sample size;
- c — fixed threshold in the discrimination rule;
- $R(v)$ — limit distribution function of (non-random) contributions of the blocks to the distance between the populations; and
- $\eta(v)$ — the weight function of the blocks.

A Priori Weighting of Independent Variables

Here we assume that the arguments of the weight function $\eta(v)$ are chosen by a priori data in dependence on the parameters $v^i = nJ^i/2$ with J^i of the form (4). Let $\bar{\theta}$ without subscripts denote the mean point $\bar{\theta} = (\theta_1 + \theta_2)/2$, and let, accordingly, $\bar{\theta}^i = (\theta_1^i + \theta_2^i)/2$, for components number $i = 1, \ldots, k$, of θ. We introduce the vectors

$$\mathbf{b}^i = n^{1/2}[I^i(\bar{\theta})]^{1/2}(\theta_2^i - \theta_1^i), \quad i = 1, \ldots, k. \tag{11}$$

LEMMA 11.1. *$nJ^i = (\mathbf{b}^i)^2 + o(1)$ as $n \to \infty$ uniformly in i, and the quantities nJ^i are uniformly bounded for $i = 1, \ldots, k$, $n = 1, 2, \ldots$*

Here and in the following the orders of magnitude in the remainder terms are supposed to be uniform in $i = 1, \ldots, k$ as $n \to \infty$.

Proof. We expand $\ln f^i(\mathbf{x}^i, \theta_2^i)$ in the Taylor series near the point θ_1^i up to terms of the second order and integrate. It follows that

$$\begin{aligned}\int \ln \frac{f^i(\mathbf{x}^i, \theta_1^i)}{f^i(\mathbf{x}^i, \theta_2^i)}\, f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}) = \\ = \int (\theta_1^i - \theta_2^i, \nabla) \ln f^i(\mathbf{x}^i, \theta_1^i)\, f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}) \\ - 1/2 \int [(\theta_1^i - \theta_2^i, \nabla)^2 \ln f^i(\mathbf{x}^i, \theta_1^i)] f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}) \\ + 1/6 \int (\theta_1^i - \theta_2^i\, \nabla)^3 \ln f^i(\mathbf{x}^i, \xi^i)\, f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}),\end{aligned} \tag{12}$$

$i = 1, \ldots, k$, where (and below) ∇ is the vector operator of differentiating with respect to components of the vector θ; ξ^i is a vector of the intermediate values of the argument. In this expansion the terms of the first order vanish in view of the normalization of $f(\mathbf{x}, \theta_1)$. In the second order terms we use the well known property of the information matrix

$$\begin{aligned}I(\theta) &= \int [\nabla \ln f(\mathbf{x}, \theta)]\, [\nabla \ln f(\mathbf{x}, \theta)]^T\, f(\mathbf{x}, \theta)\mu(d\mathbf{x}) \\ &= -\int [\nabla \nabla^T \ln f(\mathbf{x}, \theta)\,]\, f(\mathbf{x}, \theta)\mu(d\mathbf{x}).\end{aligned}$$

We find that the contribution of the second order terms to (12) is

$$(\theta_2^i - \theta_1^i) I^i(\theta_1)(\theta_2^i - \theta_1^i)/2. \tag{13}$$

In view of assumptions B and D, the third order terms equal

$$O(|\theta_2^i - \theta_1^i|^3) \int \varphi(\mathbf{x}) f(\mathbf{x}, \theta_1)\mu(dx) = O(n^{-3/2}),$$

where $\varphi(\mathbf{x})$ depends on $\mathbf{x}^i$ only. In (13) we can replace $I(\theta_1)$ by $I(\bar\theta)$ with the accuracy to $o(1)$. Indeed, we have

$$\begin{aligned} I^i(\theta_1) - I^i(\bar\theta) = \\ &= \int [(\bar\theta^i - \theta_1^i, \nabla)\nabla\nabla^T \ln\, f^i(\mathbf{x}^i, \xi^i)]\; f(\mathbf{x}, \xi)\mu(d\mathbf{x}) \\ &\quad + \int [\nabla\nabla^T \ln\, f^i(\mathbf{x}^i, \xi^i)\;](\bar\theta^i - \theta_1^i, \nabla) f(\mathbf{x}, \xi)\mu(d\mathbf{x}) \\ &= O(|\bar\theta^i - \theta_1^i|), \end{aligned}$$

where ξ and ξ^i are some intermediate values of the parameters. By assumption D we have $|\bar\theta^i - \theta_1^i| = O(n^{-1})$.

Thus, the left hand side of (12) equals $(\mathbf{b}^i)^2/2n + o(n^{-1})$. From the symmetry of assumptions, it follows that $J^i = (\mathbf{b}^i)^2/n + o(n^{-1})$. This is the first lemma statement. The second statement immediately follows from assumptions B3 and D. Lemma 11.1 is proved. □

Under assumptions A–F there exists a constant c_8 such that $nJ^i/2 < c_8$, $i = 1, \ldots, k$, for each n, and $R(c_8) = 1$.

Lemma 11.2. *Under assumptions A–F as $n \to \infty$ we have*

$$\begin{aligned} n \int \ln \frac{f^i(\mathbf{x}^i, \widehat\theta_1^i)}{f^i(\mathbf{x}^i, \widehat\theta_2^i)} f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}) = \\ &= (\mathbf{b}^i + \mathbf{z}_2^i)^2/2 - (\mathbf{z}_1^i)^2/2 + nO(\omega^i), \\ n \int \ln^2 \frac{f^i(\mathbf{x}^i, \widehat\theta_1^i)}{f^i(\mathbf{x}^i, \widehat\theta_2^i)} f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}) &= (\mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i)^2 + nO(\omega^i), \\ n \int \ln^4 \frac{f^i(\mathbf{x}^i, \widehat\theta_1^i)}{f^i(\mathbf{x}^i, \widehat\theta_2^i)} f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}) &= nO(|\omega^i|^{4/3}), \end{aligned} \tag{14}$$

where $\mathbf{b}^i \in \mathbb{R}^m$ *are defined by* (11)*, and*

$$\omega^i = (n^{-1/2} + |\widehat{\theta}_1^i - \theta_1^i| + |\widehat{\theta}_2^i - \theta_2^i|)^3, \quad i = 1, \dots, k. \tag{15}$$

Proof. We expand $\ln f^i(\mathbf{x}^i, \widehat{\theta}_1^i)$ and $\ln f^i(\mathbf{x}^i, \widehat{\theta}_2^i)$ in the Taylor series around the point θ_1 up to the terms of the third order. It follows

$$\begin{aligned}
\ln \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)} &= \\
&= (\widehat{\theta}_1^i - \widehat{\theta}_2^i, \nabla) \ln f^i(\mathbf{x}^i, \theta_1^i) \\
&+ (\widehat{\theta}_1^i - \theta_1^i, \nabla)^2 \ln f^i(\mathbf{x}^i, \theta_1^i)/2 - (\widehat{\theta}_2^i - \theta_1^i, \nabla)^2 \ln f^i(\mathbf{x}^i, \theta_1^i)/2 \\
&+ (\widehat{\theta}_1^i - \theta_1^i, \nabla)^3 \ln f^i(\mathbf{x}^i, \xi^i)/6 - (\widehat{\theta}_2^i - \theta_1^i, \nabla)^3 \ln f^i(\mathbf{x}^i, \xi^i)/6,
\end{aligned} \tag{16}$$

where ξ^i is an intermediate value of the argument. We integrate (16) with respect to $f(x, \theta_1)\mu(d\mathbf{x})$. As in the previous lemma the terms of the first order vanish; the terms of the second order are equal to

$$-(\widehat{\theta}_1^i - \theta_1^i)^T I^i(\theta_1)(\widehat{\theta}_1^i - \theta_1^i)/2 + (\widehat{\theta}_2^i - \theta_1^i)^T I^i(\theta_1)(\widehat{\theta}_2^i - \theta_1^i)/2. \tag{17}$$

But we have $I^i(\theta_1) - I^i(\bar{\theta}) = O(n^{-1/2})$. One can see that the arguments θ_1 can be replaced by $\bar{\theta}$ with the accuracy to $O(\omega^i)$. Introducing the vectors $\mathbf{z}_1^i$, $\mathbf{z}_2^i$ and $\mathbf{b}^i$ defined by (6) and (11), we obtain that (17) equals to

$$(\mathbf{b}^i + \mathbf{z}_2^i)^2/2n - (\mathbf{z}_1^i)^2/2n + O(\omega^i).$$

Integrating the terms of the third order in (16) we obtain $O(\omega^i)$. The first statement of our lemma is proved.

Now we expand the square of the logarithm in the Taylor series up to the first order terms. It follows that

$$\ln^2 \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)} = [(\widehat{\theta}_1^i - \widehat{\theta}_2^i, \nabla) \ln \ f^i(\mathbf{x}^i, \xi^i)]^2, \tag{18}$$

where the components of ξ^i lay between the components of vectors $\widehat{\theta}_1^i$ and $\widehat{\theta}_2^i$. Using conditions B we substitute

$$\ln\ f^i(\mathbf{x}^i, \xi^i) = \ln\ f^i(\mathbf{x}^i, \theta^i) + O(|\xi^i - \theta^i|)\ \varphi(\mathbf{x}).$$

Let us integrate (18) with respect to $f^i(\mathbf{x}^i, \xi^i)\mu(d\mathbf{x})$. We obtain

$$(\widehat{\theta}_1^i - \widehat{\theta}_2^i)^T I^i(\bar{\theta})(\widehat{\theta}_1^i - \widehat{\theta}_2^2) + O((\widehat{\theta}_1^i - \widehat{\theta}_2^i)^2|\xi^i - \theta_1^i|). \qquad (19)$$

The leading term is $(\mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i)^2/n$, and the correction term is of the order of $O(\omega^i)$. The second statement of the lemma is proved.

In the third statement we have

$$\left|\ln\ \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)}\right|^4 \leq |\widehat{\theta}_1^i - \widehat{\theta}_2^i|^4 \varphi^4(\mathbf{x}).$$

Integrating with respect to $f^i(\mathbf{x}, \theta^1)\mu(d\mathbf{x})$ we obtain the magnitude of the order of $|\widehat{\theta}_1^i - \widehat{\theta}_2^i|^4$. The lemma is proved. □

THEOREM 11.1. *Under assumptions A–F for the discriminant function* $g(\mathbf{x})$ *of the form* (9) *with the a priori weights of blocks* $\eta_n(v^i)$, *where* $v^i = nJ^i/2,\ i = 1, \ldots, k$, *the convergence in probability holds*

$$G_{n1} \stackrel{\text{def}}{=} \int g(\mathbf{x}) f(\mathbf{x}, \theta_1)\mu(d\mathbf{x}) \to G(\eta), \quad \nu = 1, 2, \qquad (20)$$

$$D_{n\nu} \stackrel{\text{def}}{=} \int [g(\mathbf{x}) - \int g(\mathbf{x}) f(\mathbf{x}, \theta_\nu)\mu(d\mathbf{x})]^2 f(\mathbf{x}, \theta_\nu)\mu(d\mathbf{x}) = D(\eta), \ (21)$$

as $n \to \infty,\ \nu = 1, 2,$ *where*

$$G(\eta) = \kappa \int v\eta(v)\ dR(v) \quad \textit{and} \quad D(\eta) = 2\kappa \int (v + m)\eta^2(v) dR(v). \qquad (22)$$

Proof. Let $\mathbf{E}_1$ be the conditional expectation operator with respect to the measure $f(\mathbf{x}, \theta_1)\mu(d\mathbf{x})$ under chosen samples.
As $n \to \infty$, by Lemma 11.2 we have

$$\mathbf{E}_1 g(\mathbf{x}) = \mathbf{E}\ \sum_i \eta(nJ^i/2)\ [(\mathbf{b}^i + \mathbf{z}_2^i)/2n - (\mathbf{z}_1^i)^2/2n + O(\omega^i)]. \ (23)$$

Denote $v^i = nJ^i/2$. It follows

$$\sum_i \eta(nJ^i/2)(\mathbf{b}^i)^2/2n = \sum_i (\eta(v^i)v^i + o(1))/n. \tag{24}$$

We use (5), conditions B, F2, and Lemma 11.1. The expressions in (24) can be rewritten as

$$\frac{k}{n}\int v\eta_(v)dR_n(v) + o(1) = \kappa \int v\eta(v)dR(v) + o(1).$$

For the first order (with respect to $\mathbf{z}_\nu^i$) terms, the contribution to (23) is

$$O(1)\sum_i (\mathbf{b}^i, \mathbf{z}_2^i)/2n. \tag{25}$$

In view of conditions B, D, C, and E, we have $\mathbf{b}^i = O(1)$, $\|I^i(\theta)\| = O(1)$, $|\mathbf{E}\,\mathbf{z}_2^i| = o(1)$. We conclude that the expectation of (25) tends to 0 as $n \to \infty$. The contribution of the second order terms to $\mathbf{E}_1 g(\mathbf{x})$ is

$$\sum_i \eta(nJ^i/2)\ \mathbf{E}\ [(\mathbf{z}_2^i)^2 - (\mathbf{z}_1^i)^2]/2n. \tag{26}$$

By (7) $\mathbf{E}\ (\mathbf{z}_2^i)^2 = \mathbf{E}\ (\mathbf{z}_1^i)^2 + o(1)$ as $n \to \infty$. We conclude that the expectation of (26) vanish as $n \to \infty$. The contribution of the third order terms to (23) is the sum of ω^i; in view of condition E3, its expectation tends to 0. Thus we have proved that the expectation of G_{n1} tends to $G(\eta)$.

Let us prove that (23) converges in probability. First, consider its leading part. This random value is a sum of independent variables and its variance is a sum of variances of summands. Majorizing these by the second moments we obtain that the variance of the leading part of (23) is not greater that

$$O(1)\sum_i (\mathbf{E}\ |\mathbf{b}^i + \mathbf{z}_2^i|^4 + \mathbf{E}\ |\mathbf{z}_1^i|^4)/n^2.$$

Here, the expectation values of summands are $O(1)$. Therefore, the variance of the leading part of (23) vanishes, and it follows that this term converges in probability. The absolute values of the remainder terms in (23) tends to 0 in probability since the sum of their expectations vanishes. Thus we have proved the first statement of the lemma.

Further, in view of the first statement we have

$$\begin{aligned}\mathbf{E}_1\left(g\left(\mathbf{x}\right)-G\left(\eta\right)\right)^2 &= \mathbf{E}_1\left(g\left(\mathbf{x}\right)-\mathbf{E}_1 g\left(\mathbf{x}\right)\right)^2+\mathbf{E}_1\left(g\left(\mathbf{x}\right)-G\left(\eta\right)\right)^2\\ &= \sum_i \eta^2(nJ^i/2)[\mathbf{E}_1(g^i)^2-(\mathbf{E}_1 g^i)^2]+\xi_n,\end{aligned} \tag{27}$$

where $\xi_n \to 0$ in probability as $n \to \infty$. Here the subtrahends are calculated in the first of equations (14). From Lemma 11.2 it follows that the expectation of each of these terms tends to a quantity of the order $O(n^{-2})$. It follows that their contribution to (27) tends to 0 in probability as $n \to \infty$. By the second statement of Lemma 11.2, the minuend is

$$\sum_i \eta^2\left(nJ^i/2\right)\left[(\mathbf{b}^i+\mathbf{z}_2^i-\mathbf{z}_1^i)^2/n+O(w^i)\right]. \tag{28}$$

Let us calculate the expectation of the leading part of (28). In view of (7) and D we have

$$\mathbf{E}\left(\mathbf{b}^i,\mathbf{z}_1^i\right)=o\left(1\right),\quad \mathbf{E}\left(\mathbf{z}_1^i,\mathbf{z}_2^i\right)=o\left(1\right),\quad \mathbf{E}\left(\mathbf{z}_\nu^i\right)^2=m+o\left(1\right),$$

and $\mathbf{E}\,\omega^i=o(1)$, where $i=1,\ldots,k,\ \ \nu=1,2$. Thus the expectation of (28) equals

$$\sum_i \eta^2\left(nJ^i/2\right)[(\mathbf{b}^i)^2+2m]/n+o\left(1\right).$$

We transform this expression substituting $v^i=nJ^i/2,$ and $(b^i)^2=nJ^i+o\left(1\right)$. The expectation of (28) can be written as

$$\frac{2k}{n}\int(v+m)\eta^2(v)dR_n(v)+o(1)=2\kappa\int(v+m)\eta^2(v)dR(v)+o(1).$$

Thus we have proved that the expectation of D_{n1} tends to $D(\eta)$. To prove the convergence in probability we notice that the variance of the leading part of (28) is a sum of variances, and each summand has a variance not greater than the second moment that equals

$$O(1)\ \mathbf{E}\ [(\mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i)^2]^2/n^2, \quad i = 1, \ldots, k.$$

By Lemma 11.1 and (7) we have $|\mathbf{b}^i|^4 = O(1)$, $\mathbf{E}\ |\mathbf{z}_\nu^i|^4 = O(1)$ for all i and $\nu = 1, 2$. It follows that this variance is $O(n^{-2})$. Consequently the leading part of (27) converges in the square mean to its expectation value. The sum of the remainder terms in (27) tends to 0 and it follows that the random value $\xi_n \to 0$ in probability as $n \to \infty$. We conclude that D_{n1} converges to $D(\eta)$ in probability. The symmetric statement for $\nu = 2$ follows from the symmetry of assumptions. The proof of the theorem is complete. □

Corollary. If $\eta(v) = 1$ for all $v > 0$, then $G_{n\nu} \to J/2$ and $D_{n\nu} \to J + 2\kappa m$ as $n \to \infty$ in probability, $\nu = 1, 2$, where

$$J = 2\kappa \int v dR(v). \tag{29}$$

Example 3. For $m = 1$ and normal distribution in Example 1 we have $I^i(\theta) = 1$ for each i and the statement of Theorem 11.1 follows with the zero remainder term. The function $g(\mathbf{x})$ is normally distributed for chosen samples and the first statement of Lemma 11.2 follows with the zero remainder term. The random values $G_{n\nu}$ and $D_{n\nu}$ have non-central χ^2-distribution with k degrees of freedom, $\nu = 1, 2$. The variance of these values decreases as n^{-1}, and it follows that $G_{n1} \to G(\eta)$ and $D_{n\nu} \to D(\eta)$ in probability, $\nu = 1, 2$.

To be more concise in the proofs we denote

$$g^i = g^i(\mathbf{x}^i) = \ln \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)}, \quad i = 1, \ldots, k, \tag{30}$$

and let the subscripts i everywhere in sums run over $i = 1, \ldots, k$.

THEOREM 11.2. *Suppose the conditions A–F are satisfied, the discriminant function $g(\mathbf{x})$ has the form (5) with the weights $\eta(nJ^i/2)$ of the blocks, $i = 1, \ldots, k$, and additionally, let $D(\eta) > 0$. Then in probability,*

$$\alpha_1 \to \Phi\left(-\frac{G(\eta) - c}{\sqrt{D(\eta)}}\right), \quad \alpha_2 \to \Phi\left(-\frac{G(\eta) + c}{\sqrt{D(\eta)}}\right),$$

where $G(\eta)$ and $D(\eta)$ are of the form (22).

Proof. Consider the random value $\widehat{\alpha}_1 = F_1(c)$, where $F_1(\cdot)$ is the (conditional) distribution function of $g = g(\mathbf{x})$ under chosen samples and $\mathbf{x}$ in population 1. Notice that g is a weighted sum of independent random values g^i. Denote the (conditional for fixed samples) expectation with respect to the measure $f(\mathbf{x}, \theta_1)\mu(d\mathbf{x})$ by $\mathbf{E}_1$. We have $G_{n1} = \mathbf{E}_1 g, \quad D_{n1} = \mathbf{E}_1(g - \mathbf{E}_1 g)^2$. Define

$$T_{n1} = \sum_i \mathbf{E}_1 |g^i - \mathbf{E}_1 g^i|^3.$$

We apply the Esseen inequality (see in [57]) to $g = g(\mathbf{x})$. If $D_{n1} > 0$, then the distribution function of g is close to the distribution function $\Phi_n(\cdot)$ of the normal law $\mathbf{N}(G_{n1}, D_{n1})$ with the accuracy to $T_{n1}/D_{n1}^{3/2}$. By virtue of Theorem 11.1 $D_{n1} \to D(\eta)$ in probability. It follows that $D_{n1} > D(\eta)/2 > 0$ with the probabilities $p_n \to 1$. Now, the expectation

$$\mathbf{E}\ \widehat{T}_{n1} \leq \sum_i (\mathbf{E}\ |g^i - \mathbf{E}_1\ g^i|^4)^{3/4}. \tag{31}$$

From Lemma 11.2 conditions D and F3 it follows that $\mathbf{E}|g^i|^4 \leq O(n^{-2})$ for each i. Hence the left hand side of (31) is not larger than $O(n^{-1/2})$. By the Esseen inequality, the random value $g = g(\mathbf{x})$ has the (conditional) distribution function approaching the standard error function Φ_n with the probabilities $p_n \to 1$. By virtue of Theorem 11.1 $G_{n1} \to G(\eta)$ and $D_{n1} \to D(\eta) > 0$ in probability. Consequently $F_1(c)$ tends to the distribution function of $\mathbf{N}(G(\eta), D(\eta))$ in probability. We proved the first statement of Theorem 11.2 for $\nu = 1$. The second one follows from the symmetry of assumptions. The proof is complete. □

If $\eta(v) = 1$ for all $v \geq 0$, then $G(\eta) = J/2$, $D(\eta) = J$, and the limits in the formulation of Theorem 11.2 are equal to

$$\Phi\left(-\frac{J \pm 2c}{2\sqrt{J + 2\lambda}}\right),$$

where $\lambda = \kappa m$.

Example 4. Consider two normal populations $\mathbf{N}(\theta_\nu, I)$, $\nu = 1, 2$, with common covariance matrix cov $(\mathbf{x}, \mathbf{x}) = I$. If $\eta(v) = 1$

for all $v \geq 0$, we obtain the discriminant function (mentioned in Introduction)

$$g(\mathbf{x}) = (\widehat{\theta}_1 - \widehat{\theta}_2)^T(\mathbf{x} - (\widehat{\theta}_1 + \widehat{\theta}_2)/2),$$

where $\widehat{\theta}_\nu$ are sample means, $\nu = 1, 2$. In this case,

$$G_{n1} = (\widehat{\theta}_1 - \widehat{\theta}_2)^T(\theta_\nu - (\widehat{\theta}_1 + \widehat{\theta}_2)/2), \quad D_{n\nu} = (\widehat{\theta}_1 - \widehat{\theta}_2)^2, \quad \nu = 1, 2.$$

Theorem 11.1 states that $G_{n1} \to J/2$, $D_{n\nu} \to J$ in probability as $n \to \infty$, $\nu = 1, 2$, where $J = \lim_{n\to\infty}(\theta_1 - \theta_2)^2$.

Minimization of the Limit Error Probability for a Priori Weighting

Obviously, to minimize the limit value of $(\alpha_1 + \alpha_2)/2$ established by Theorem 11.2 we should choose the threshold $c = 0$ (non-trivial optimum thresholds can be obtained for essentially different sample sizes, see Introduction). Thus

$$\min_c \operatorname*{plim}_{n\to\infty} (\alpha_1 + \alpha_2)/2 = \Phi(-\rho(\eta)/2),$$

where, by definition, the 'effective Mahalanobis distance' $\rho = \rho(\eta) = 2G(\eta)/\sqrt{D(\eta)}$.
If $\eta(v) = 1$ for all $v \geq 0$, then we have $\rho = \sqrt{J}$.

THEOREM 11.3. *Varying the function $\eta(\cdot)$ in the class $\mathfrak{K}$ under $c = 0$, fixed m, κ, and $R(v)$ with $g(\mathbf{x})$ of the form* (9)*, we have*

$$\inf_\eta \min_c \operatorname*{plim}_{n\to\infty} (\alpha_1 + \alpha_2)/2 = \Phi(-\rho(\eta_{\text{opt}})/2),$$

where

$$\eta_{\text{opt}} = \eta_{\text{opt}}(v) = \frac{v}{v+m} \quad \text{and} \quad \rho^2(\eta_{\text{opt}}) = 2\kappa \int \frac{v^2}{v+m} dR(v). \tag{32}$$

Proof. We vary $\eta(v)$ for $c = 0$ and obtain the necessary condition of the extremum

$$D(\eta) \int v\delta\eta(v)dR(v) = 2G(\eta) \int (v+m)\eta(v)\delta\eta(v)dR(v).$$

It follows that $\eta(v) = \text{const}\ v/(v+m)$. The proportionality coefficient does not effect the value of $\rho(\eta)$. Set $\eta(v) = \eta_{\text{opt}}(v)$ from (32). Let us show that the value $\rho(\eta_{\text{opt}})$ is not less than $\rho(\eta)$ for any other $\eta(t)$ from $\mathfrak{K}$. Using the Cauchy–Bunyakovskii inequality we obtain

$$\begin{aligned}
2G(\eta) &= 2\kappa \int v\eta(v)dR(v) \\
&\le 2\kappa \left[\int \frac{v^2}{v+m}dR(v) \int (v+m)\eta^2(v)dR(v)\right]^{1/2} \\
&= \rho(\eta_{\text{opt}})\sqrt{D(\eta)}.
\end{aligned}$$

This ends the proof of Theorem 11.3. □

This theorem was first proved in [15].

Example 5 (The case of a portion r of non-informative blocks of variables).
Let $R(v) = r \ge 0$ for $0 \le v \le b \le c_8$ and $R(b) = 1$, $r \le 1$. If $\eta(v) = 1$ for all $v \ge 0$, then

$$G(1) = J/2 = \kappa(1-r)b, \quad D(1) = J + 2\kappa m = 2\kappa[(1-r)b + m].$$

If $\eta(v) = \eta_{\text{opt}}(v) = v/(v+m)$, then $\rho^2(\eta_{\text{opt}}) = 2\kappa b^2(1-r)/(b+m)$. The ratio

$$\frac{\rho^2(\eta_{\text{opt}})}{\rho^2(1)} = \frac{(1-r)b+m}{(1-r)(b+m)} \ge 1.$$

Example 6. Let $R(v) = v/b$ for $v \le b \le c_8$. If $\eta(v) = 1$ for all $v \ge 0$, then $G = G(1) = J/2 = \kappa b/2$, $D(1) = \kappa(b+2m)$. If $\eta(v) = \eta_{\text{opt}}(v) = v/(v+m)$, the value

$$\rho^2(\eta_{\text{opt}}) = \kappa(b - 2m + 2m^2 b^{-1}\ln(1 + mb^{-1})) \ge \rho^2(1).$$

If $b \gg m$, then $\rho^2(1) \cong 1/2\ \kappa b^2/m$ whilst $\rho^2(\eta_{\text{opt}}) \cong 2/3\ \kappa b^2/m$.

Note that, under assumptions A–F, the contributions of blocks to the discrimination are $(\theta_1^i - \theta_2^i)^2 = O(n^{-1})$, whilst the bias of the estimators $(\widehat{\theta}_1^i - \widehat{\theta}_2^i)^2$ is of the same order of magnitude. This means that the true contributions are substantially different from their estimators. Thus a practical problem arises of improving the discrimination by weighting of blocks using the estimation characteristics.

Weighting of Independent Variables by Estimators

We consider now a more realistic problem when the weights in (10) are calculated by the estimators $\widehat{J}^i$ of the quantities J^i, $i = 1, \ldots, k$. Denote $\bar{\theta} = (\theta_1 + \theta_2)/2$ and

$$\widehat{\mathbf{b}}^i = n^{1/2}[I^i(\bar{\theta})]^{1/2}(\widehat{\theta}_2^i - \widehat{\theta}_1^i) = \mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i,$$

$$\widehat{g}^i = \widehat{g}^i(\mathbf{x}^i) = \ln \frac{f^i(\mathbf{x}^i, \widehat{\theta}_1^i)}{f^i(\mathbf{x}^i, \widehat{\theta}_2^i)}, \quad 1 = 1, \ldots, k. \tag{33}$$

Lemma 11.3. *$n\widehat{J}^i = (\widehat{\mathbf{b}}^i)^2 + n\omega^i$, $i = 1, \ldots, k$, where $\mathbf{b}^i$ and ω^i are defined by* (11) *and* (15).

Proof. Let $\mathbf{E}_\nu$ denote the conditional expectation calculated by the integration with respect to the measure $f(\mathbf{x}, \theta_\nu)d\mu(\mathbf{x})$, $\nu = 1, 2$. We have $\widehat{J}^i = \mathbf{E}_1\widehat{g}^i - \mathbf{E}_2\widehat{g}^i$, $i = 1, \ldots, k$. Expanding $\widehat{g}^i$ with respect to $(\widehat{\theta}_1^i - \widehat{\theta}_2^i)$ near the point $\widehat{\theta}_1^i$ to the first order terms, we obtain

$$\mathbf{E}_1\widehat{g}^i = 1/2\ (\widehat{\theta}_1^i - \widehat{\theta}_2^i)^T I^i(\widehat{\theta}_1^i)(\widehat{\theta}_1^i - \widehat{\theta}_2^i) + O(|\widehat{\theta}_1^i - \widehat{\theta}_2^i|^3)\ \mathbf{E}_1\varphi(\mathbf{x})$$

with $\varphi(\mathbf{x})$ from B2. But $I^i(\widehat{\theta}^i) - I^i(\bar{\theta}^i) = O(|\widehat{\theta}^i - \bar{\theta}^i|)$ for each i. Therefore,

$$\mathbf{E}_1\widehat{g}^i = 1/2\ (\widehat{\theta}_1^i - \widehat{\theta}_2^i)^T I(\bar{\theta}^i)(\widehat{\theta}_1^i - \widehat{\theta}_2^i) + \omega^i,$$

$i = 1, \ldots, k$. From the symmetry of assumptions, it follows that to calculate $\mathbf{E}_2\widehat{g}^i$ it suffices to change the sign and transpose the subscripts 1 and 2. We obtain the required statement. Lemma 11.3 is proved. □

Theorem 11.4. *Under assumptions A–F, the distribution functions $F^i(u)$ of $n\widehat{J}^i/2$ are such that for almost all $u \geq 0$*

$$\lim_{n\to\infty} \max_i \ |F^i(u) - F_m^\beta(u)|^2 = 0,$$

where $\beta = \beta(i) = |\mathbf{b}^i|/\sqrt{2}$, $i = 1, \ldots, k$, and $F_m^\beta(u)$ are the non-central χ^2 -distribution functions with m degrees of freedom and the non-centrality parameters $\beta^2 = |\mathbf{b}^i|^2/2$, $i = 1, \ldots, k$.

Proof. By Lemma 11.3,

$$n\widehat{J}^i/2 = (\mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i)^2 + O(n\omega_i), \quad i = 1, \ldots, k,$$

where $\mathbf{z}_2^i$ and $\mathbf{z}_1^i$ are independent vectors that, in view of F4, are asymptotically normal as $n \to \infty$ uniformly with respect to $i = 1, \dots, k$, and to the arguments of the distribution functions. It follows that the vectors $(\mathbf{z}_2^i - \mathbf{z}_1^i)/\sqrt{2}$ are also asymptotically normal and the distribution function of these tends to the distribution function of $\mathbf{N}(0_m, I_m)$. Consequently the random values $u^i = (\mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i)^2/2$ have the distribution function equal to $F_m^\beta(u) + o(1)$, where $\beta^2 = (\mathbf{b}^i)^2/2$ for each i. But the convergence in probability $nJ^i/2 \to u^i$ implies the convergence in distribution. It follows that $F^i(u) \to F_m^\beta(u)$ for almost all $u \geq 0$. This proves Theorem 11.4. □

Let us write out some well-known relations for non-central χ^2-distribution needed in the following. Denote the density

$$f_m^\beta(v) = \frac{dF_m^\beta(v)}{dv} = \exp\left(-\frac{\beta^2}{2}\right) \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\beta^2}{2}\right)^k f_{m+2k}^0(v),$$

$m = 0, 1, 2, \dots$.

We have the integral representation

$$f_m^\beta(v) = (2\pi)^{-m/2} \int \exp\left(-(\vec{z} - \vec{\beta})^2/2\right) \delta(v - |\vec{z}|^2) d\vec{z}, \qquad (34)$$

where β is the length of the vector $\vec{\beta} \in \mathbb{R}^m$, and $\delta(\cdot)$ is the Dirac δ-function. The characteristic function of the non-central χ^2-distribution is

$$\chi_m^\beta(t) = \exp\left(\frac{it\beta^2}{1 - 2it}\right)(1 - 2it)^{-m/2}. \qquad (35)$$

We have the recurrent relations:

$$F_m^\beta(u) - F_{m+2}^\beta(u) = 2f_{m+2}^\beta(u),$$
$$u f_m^\beta(u) = m f_{m+2}^\beta(u) + \beta^2 f_{m+4}(u), \quad u \geq 0. \qquad (36)$$

The derivatives are:

$$2u \frac{df_m^\beta(u)}{du} = \beta^2 f_{m+2}^\beta(u) + (m - u - 2) f_m^\beta(u),$$
$$\frac{\partial F_m^\beta(u)}{\partial \beta^2} = -f_{m+2}^\beta(u), \quad u \geq 0. \qquad (37)$$

Expanding $f_m^\beta(u)$ in the series with respect to small $u > 0$, we obtain

$$f_m^\beta(u) = \frac{u^{m/2-1}}{2^{m/2}\Gamma(m/2)}\exp\,(-\beta^2/2)[1 + (\beta^2/m - 1)u/2 + O(u^2)].$$

In special cases,

$$f_1^\beta(\tau^2) = \frac{1}{\sqrt{2\pi}}\frac{1}{2\tau}\left[\exp\,(-(\tau-\beta)^2/2) + \exp\,(-(\tau+\beta)^2/2)\right],$$

$$f_3^\beta(\tau^2) = \frac{1}{\sqrt{2\pi}}\frac{1}{2\beta}\left[\exp\,(-(\tau-\beta)^2/2) - \exp\,(-(\tau+\beta)^2/2)\right]. \tag{38}$$

Lemma 11.4. *Under assumptions A–F for any measurable bounded function* $\Psi(\mathbf{z}_1, \mathbf{z}_2)$ *of two* m*-dimensional vectors* $\mathbf{z}_1$ *and* $\mathbf{z}_2$ *defined by* (6), *we have*

$$\mathbf{E}\ \Psi(\mathbf{z}_1^i, \mathbf{z}_2^i)(\mathbf{z}_\nu^i)^2 =$$

$$= (2\pi)^{-m}\iint \Psi(\mathbf{z}_1, \mathbf{z}_2)(\mathbf{z}_1)^2 \exp\left(-\frac{\mathbf{z}_1^2 + \mathbf{z}_2^2}{2}\right) d\mathbf{z}_1 d\mathbf{z}_2 + o(1) \tag{39}$$

as $n \to \infty$, *where* $\mathbf{z}_\nu \in \mathbb{R}^m$ *are of the form* (6), $\nu = 1, 2$.

Proof. Let $\nu = 1$. Denote the distribution function of two independent normal vectors $\mathbf{N}(0_m, I_m)$ by $\Phi_m(\cdot)$. Let F^i denote the joint distribution function of $\mathbf{z} = (\mathbf{z}_1^i, \mathbf{z}_2^i)$. From Theorem 11.4 it follows that $F^i(\mathbf{z}) \to \Phi_m(\mathbf{z})$ as $n \to \infty$ uniformly with respect to $\mathbf{z}$. Denote $\Delta(\mathbf{z}) = F^i(\mathbf{z}) - \Phi_m(\mathbf{z})$. It suffices to prove that $\int \mathbf{z}^2 d\Delta(\mathbf{z}) \to 0$ as $n \to \infty$. Choose a positive $T > 0$. Then

$$\left|\int \mathbf{z}^2 d\Delta(\mathbf{z})\right| \le T^2 \sup_{|\mathbf{z}|<T} |F^i(\mathbf{z}) - \Phi_m(\mathbf{z})| + T^{-2}\left|\int_{|\mathbf{z}|\ge T} |\mathbf{z}|^4 d\Delta(\mathbf{z})\right|.$$

The first summand vanishes as $n \to \infty$ for any chosen T. The second summand is $O(T^{-2})$ in view of (7). It follows that the left hand side tends to 0 as $n \to \infty$. Relation (39) for $\nu = 2$ follows from assumptions. This proves the lemma. □

Define

$$G_{n\nu} = \int g(\mathbf{x}) f(\mathbf{x}, \theta_\nu)\mu(d\mathbf{x}),$$

$$D_{n\nu} = \int [g(\mathbf{x}) - G_{n\nu}]^2 f(\mathbf{x}, \theta_\nu)\mu(d\mathbf{x}), \quad \nu = 1, 2.$$

LEMMA 11.5. *Under assumptions A–F if the weights* $\eta(n\widehat{J}^i/2)$ *are used in the discriminant function* $g(\mathbf{x})$ *of the form* (10), *we have as* $n \to \infty$

$$\begin{aligned} &\mathbf{E}\ G_{n1} \to G(\eta) = \kappa \int \beta^2 [\int \eta(u) f^{\beta}_{m+2}(u) du]\ dR(\beta^2), \\ &\mathbf{E}\ D_{n\nu} \to D(\eta) = 2\kappa \int [\int \eta^2(u) f^{\beta}_{m}(u) u du]\ dR(\beta^2) \end{aligned} \tag{40}$$

$\nu = 1, 2$.

Proof. As in the above let $\mathbf{E}_1$ denote the expectation with respect to the measure $f(\mathbf{x}, \theta_1)\mu(d\mathbf{x})$. By Lemma 11.2 $\mathbf{E}_1 g^i = (\mathbf{y}^2 + 2\mathbf{y}^T\mathbf{z}_1^i) + O(\omega^i)$, where $\mathbf{y} = \mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i$, $\ i = 1, \dots, k$. Applying Lemma 11.4 we obtain that

$$\begin{aligned} &\mathbf{E}\ G_{n1} = \mathbf{E}\mathbf{E}_1 g(\mathbf{x}) = \\ &= \frac{1}{2n} \sum_i \frac{1}{(2\pi)^m} \iint \eta(u^i)(\mathbf{y}^2 + 2\mathbf{y}^T\mathbf{z}) \\ &\qquad \times \exp\ [-(\mathbf{y} - \mathbf{b}^i + \mathbf{z})^2/2 - \mathbf{z}^2/2]\ d\mathbf{y} d\mathbf{z} + o(1). \end{aligned} \tag{41}$$

Here $u^i = \mathbf{y}^2/2 + n\omega_i$, where ω^i is defined by (15). Let us estimate the contribution of the difference $\eta(u^i) - \eta(\mathbf{y}^2/2)$. The bounded function $\eta(\cdot)$ from the class $\mathfrak{K}$ has only a finite number of the discontinuity points. The ε-neighbourhoods of these contribute $O(\varepsilon)$ to $\mathbf{E}\mathbf{E}_1 g(\mathbf{x})$. The contribution of large $|\mathbf{y}|$ and $|\mathbf{z}|$ can be made arbitrarily small. In the remaining finite region the function $\eta(\cdot)$ is continuous and differentiable. The measure of argument values for which the derivatives are large can be made arbitrarily small. It follows that we can assume that $\eta(u^i) - \eta(\mathbf{y}^2/2) = O(n\omega^i)$. Therefore we can replace $\eta(u^i)$ in (41) by $\eta(\mathbf{y}^2/2)$ with the accuracy to $o(1)$.

Now we note that for $\mathbf{z},\ \mathbf{b},\ \mathbf{y} \in \mathbb{R}^m$ we have identically

$$\begin{aligned} &(\mathbf{y}^2 + 2\mathbf{y}^T\mathbf{z})\ \exp\ (-(\mathbf{y} - \mathbf{b} + \mathbf{z})^2/2) = \\ &\quad = (2\mathbf{y}^T \frac{\partial}{\partial \mathbf{b}} + 2\mathbf{b}^T\mathbf{y} - \mathbf{y}^2)\ \exp\ (-(\mathbf{y} - \mathbf{b} + \mathbf{z})^2/2). \end{aligned} \tag{42}$$

Calculating $\mathbf{E}\mathbf{E}_1 g^i$ we remove the differentiation with respect to $\mathbf{b}$ to outside of the integral with respect to $\mathbf{z}_1$ in (41) with $\mathbf{b} = \mathbf{b}^i$. Integrating with respect to $\mathbf{z}_1$ and replacing formally the summation over i by the integration we obtain

$$\begin{aligned}
\mathbf{E}\mathbf{E}_1 g(\mathbf{x}) = \\
&= \frac{\kappa}{2^{m+1}\pi^{m/2}} \int [\int (2\mathbf{y}^T \frac{\partial}{\partial \mathbf{b}} + 2\mathbf{b}^T\mathbf{y} - \mathbf{y}^2) \\
&\quad \times \exp\,(-(\mathbf{b}-\mathbf{y})^2/4)\; \eta(\mathbf{y}^2/2) d\mathbf{y}]\; dR_n(\mathbf{b}^2/2) + o(1) \\
&= 2^{-m-1}\pi^{-m/2}\kappa \int [\int \eta(\mathbf{y}^2/2)\mathbf{b}^T\mathbf{y} \\
&\quad \times \exp\,(-(\mathbf{b}-\mathbf{y})^2/4) d\mathbf{y}]\; dR(\mathbf{b}^2/2) + o(1). \qquad (43)
\end{aligned}$$

We notice that

$$\mathbf{b}^T\mathbf{y}\,\exp\,\left(-(\mathbf{b}-\mathbf{y})^2/4\right) = [\mathbf{b}^2 + 2\mathbf{b}^T \frac{\partial}{\partial \mathbf{b}}]\,\exp\,\left(-(\mathbf{b}-\mathbf{y})^2/4\right).$$

Singling out the integral of the form (34) in (43) we find that

$$\mathbf{E}\mathbf{E}_1 g(\mathbf{x}) \to \kappa \iint \eta(u)\; [\frac{\mathbf{b}^2}{2} + \mathbf{b}^T \frac{\partial}{\partial \mathbf{b}}]\; f_m^\beta(u) du\; dR(\beta^2),$$

where $\beta^2 = \mathbf{b}^2/2$. Using (37) and the recurrent relations for $f_m^\beta(u)$ we have

$$\begin{aligned}
&[\frac{\mathbf{b}^2}{2} + \mathbf{b}^T \frac{\partial}{\partial \mathbf{b}}] f_m^\beta(u) = \\
&= \beta^2[1 + 2\frac{\partial}{\partial \beta^2}] f_m^\beta(u) = \beta^2[f_m^\beta(u) - \frac{\partial}{\partial u}\left(F_m^\beta(u) - F_{m+2}^\beta(u)\right)] \\
&\qquad = \beta^2 f_{m+2}^\beta(u).
\end{aligned}$$

Substituting into (43) we obtain

$$\mathbf{E}\mathbf{E}_1 g(\mathbf{x}) = \kappa \iint \eta(u)\beta^2 f_{m+2}^\beta(u) du dR(\beta^2) + o(1).$$

The first statement of the lemma is proved.

For the variances we find

$$\begin{aligned}\mathbf{E}\mathbf{E}_1(g(\mathbf{x})-G(\eta))^2 &= \\ &= \mathbf{E}\mathbf{E}_1(g(\mathbf{x}) - \mathbf{E}\mathbf{E}_1 g(\mathbf{x}))^2 + o(1) \\ &= \sum_i \eta^2(n\widehat{J}^i/2)[\mathbf{E}\mathbf{E}_1(g^i)^2 - (\mathbf{E}\mathbf{E}_1 g^i)^2)] + o(1).\end{aligned} \tag{44}$$

By Lemma 11.2 $\mathbf{E}\mathbf{E}_1 g^i = O(n^{-1})$, the subtrahends contribute no more than $O(n^{-1})$ to (44), and we can rewrite (44) in the form

$$n^{-1}\mathbf{E}\sum_i \eta^2\left(\frac{|\mathbf{y}^i|^2 + O(n\omega^i)}{2}\right)|\mathbf{y}^i|^2 + o(1), \tag{45}$$

where $\mathbf{y}^i = \mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i$. As in the proof of Lemma 11.5 we find that the argument of $\eta(\cdot)$ can be replaced with the same accuracy by $|\mathbf{y}^i|^2/2$. Using (5) we can rewrite (45) in the form of an integral and apply Lemma 11.4. Introducing normal densities we obtain that (45) equals

$$\frac{\kappa}{(2\pi)^{m/2}}\int[\int \eta^2(\mathbf{z}^2)\mathbf{z}^2 \exp\,(-(\mathbf{z}-\vec{\beta})^2/4)d\mathbf{z}]\; dR_n(\vec{\beta}^2) + o(1), \tag{46}$$

Since the inner integral is bounded and continuous with respect to the variable $\vec{\beta}^2$, we can replace the integration with respect to the measure $R_n(v)$ by the integration with respect to $R(v)$. Let us single out the integral representation of $f_m^\beta(u)$. The expression (46) can be rewritten in the form

$$\begin{aligned}\frac{\kappa}{(2\pi)^{m/2}}\iint u\eta^2(u)\int \delta(u-\mathbf{z}^2)\exp\,(-\frac{(\mathbf{z}-\vec{\beta})^2}{2})d\mathbf{z}dudR(\vec{\beta}^2) + o(1) \\ = \kappa\iint \eta^2(u)uf_m^\beta(u)dudR(\beta^2) + o(1).\end{aligned} \tag{47}$$

This completes the proof of Lemma 11.5. □

THEOREM 11.5. *Under assumptions A–F using $g(\mathbf{x})$ of the form (10) with the weight coefficients $\eta(n\widehat{J}^i/2)$ we have the convergence in probability $G_{n1} \to G(\eta)$ and $D_{n\nu} \to D(\eta)$, $\nu = 1, 2$.*

Proof. In view of Lemma 11.5 it suffices to prove that the variances of G_{n1} and $D_{n\nu}$ vanish as $n \to \infty$. Note that G_{n1} and $D_{n\nu}$ are sums of k independent random values and that their variance is the sum of variances of summands. We obtain the upper estimates of the variances of each summand replacing them by the second moments. In view of the boundedness of $\eta(\cdot)$ we have the same majorizing expressions as in Theorem 11.1. Using the same arguments we find that var $(G_{n1}) \to 0$ and var $(D_{n\nu}) \to 0$ as $n \to \infty$, $\nu = 1, 2$. Theorem 11.5 is proved. □

Denote

$$\sigma(u) = \int \beta^2 f_{m+2}^{\beta}(u) dR(\beta^2), \quad \pi(u) = \int u f_m^{\beta}(u) dR(\beta^2), \quad u \geq 0. \tag{48}$$

The functions in (40) can be rewritten in the form

$$G(\eta) = \kappa \int_0^{\infty} \sigma(u)\eta(u)du,$$
$$D(\eta) = 2\kappa \int_0^{\infty} \pi(u)\eta^2(u)du. \tag{49}$$

THEOREM 11.6. *Let assumptions A–F hold and $D(\eta) > 0$. Then, using $g(\mathbf{x})$ of the form* (10) *with weight coefficients $\eta(n\widehat{J}^i/2)$ we have*

$$\operatorname*{plim}_{n\to\infty} \alpha_1 = \Phi\left(-\frac{G(\eta) - c}{\sqrt{D(\eta)}}\right),$$
$$\operatorname*{plim}_{n\to\infty} \alpha_2 = \Phi\left(-\frac{G(\eta) + c}{\sqrt{D(\eta)}}\right), \tag{50}$$

where $G(\eta)$ and $D(\eta)$ are defined by (40).

Proof. Let $\widehat{F}(\cdot)$ denote the conditional distribution function of the random value $g = g(\mathbf{x})$ under chosen $\widehat{\theta}_1$ and $\widehat{\theta}_2$. Obviously, $\alpha_1 = \widehat{F}(c)$. The function $g(\mathbf{x})$ is a weighted sum of independent random values $\widehat{g}^i$ of the form (33). Denote by $\widehat{G}_{n1}$, $\widehat{D}_{n1}$, $\widehat{T}_{n1}$ the sum of the first moments, variance, and the third absolute moment (conditional under fixed samples) of the summands. Applying the Esseen inequality we find that the distribution function of $g = g(\mathbf{x})$ differs from the distribution function of $\mathbf{N}(\widehat{G}_{n1}, \widehat{D}_{n1})$ by $O(\widehat{T}_{n1}/\widehat{D}_{n1}^{3/2})$, Since $\widehat{D}_{n1} \to D(\eta) > 0$ and $\widehat{T}_{n1} \to 0$ in probability, we can conclude that the distribution function of $g = g(\mathbf{x})$ tends to the distribution function of $\mathbf{N}(G(\eta), D(\eta))$ in probability. The symmetric conclusion for $\nu = 2$ follows form assumptions. Theorem 11.6 is proved. □

Obviously, the minimum of the limit half-sum of error probabilities is reached for $c = 0$ and is

$$\alpha_*(\eta) \stackrel{\text{def}}{=} \operatorname*{plim}_{n\to\infty} (\alpha_1 + \alpha_2)/2 = \Phi(-\rho(\eta)/2), \tag{51}$$

where the effective Mahalanobis distance $\rho(\eta)$ is such that

$$\rho^2(\eta) = 2\kappa[\int_0^\infty \sigma(u)\eta(u)du]^2 \Big/ \int_0^\infty \pi(u)\eta^2(u)du. \tag{52}$$

Example 7. Suppose that the contributions J^i of all blocks to the distance between the populations are identical. Then, for some β^2, $R(v) = 0$ for $v < \beta^2$ and $R(v) = 1$ for $v \geq \beta^2$. We have

$$\sigma(u) = \beta^2 f_{m+2}^\beta(u), \quad \pi(u) = u f_m^\beta(u),$$
$$G(\eta) = \kappa\beta^2 \int \eta(u) f_{m+2}^\beta(u)du.$$

If $\eta(u) = 1$ for all $u > 0$, then $G(1) = \kappa\beta^2$ and using the properties of the χ^2-distribution we find that $D(1) = 2\kappa(m + \beta^2)$.

Minimization of the Limit Error Probability for Weighting by Estimators

Let us find the maximum of the function $\rho(\eta)$ defined by (52).

THEOREM 11.7. *Varying the quantity c and the function $\eta(\cdot)$ in the class $\mathfrak{K}$ under fixed m, κ, and $R(v)$, for $g(\mathbf{x})$ of the form* (10), *we have*

$$\inf_\eta \min_c \operatorname*{plim}_{n\to\infty} (\alpha_1 + \alpha_2)/2 = \Phi(-\rho(\eta_{\text{opt}})/2),$$

where

$$\eta_{\text{opt}}(u) = \frac{\sigma(u)}{\pi(u)} \quad \textit{and} \quad \rho^2(\eta_{\text{opt}}) = 2\kappa \int_{+0}^\infty \frac{\sigma^2(u)}{\pi(u)} du. \tag{53}$$

Proof. We seek the extremum of (52) by the variation of $\eta(u)$. The necessary condition of the extremum is

$$D(\eta)\int_0^\infty \sigma(u)\delta\eta(u)du = G(u)\int_0^\infty \pi(u)\eta(u)\delta\eta(u)du.$$

Hence $\eta(u) = \text{const}\ \sigma(u)/\pi(u)$. The constant coefficient does not affect the value of $\rho(\eta)$. Set $\eta_{\text{opt}}(u) = \sigma(u)/\pi(u)$. Let us prove that $\eta_{\text{opt}}(u)$ is bounded. By Lemma 11.1 the supports of the distributions $R_n(\beta^2)$ and $R(\beta^2)$ are bounded. For the χ^2-densities, we have the inequality $uf_m^\beta(u) \geq mf_{m+2}^\beta(u)$. It follows that

$$\sigma(u) = \int \beta^2 f_{m+2}^\beta(u)dR(\beta^2) \leq \frac{u}{m}\int \beta^2 f_m^\beta(u)dR(\beta^2) \leq \frac{c_8}{m}\pi(u),$$

$u \geq 0$. We see that the function $\eta_{\text{opt}}(u)$ is bounded and continuous for $u > 0$ and belongs to the class $\mathfrak{K}$. Substituting this function in (52) we obtain the second relation in (53). For any other $\eta(u)$ from $\mathfrak{K}$, using the integral Cauchy–Bunyakovskii inequality we find:

$$\begin{aligned}2G(\eta) &= 2\kappa\int_0^\infty \sigma(u)\eta(u)du \\ &\leq 2\kappa\left(\int_{+0}^\infty \frac{\sigma^2(u)}{\pi(u)}du\int_0^\infty \pi(u)\eta^2(u)du\right)^{1/2} = \sqrt{D(\eta)}\rho(\eta_0).\end{aligned}$$

This ends the proof of Theorem 11.7. □

Example 8. Let $\gamma > 0$ and

$$\begin{aligned}dR(\beta^2) &= (\frac{\gamma}{2\pi})^{m/2}\int \exp\,(-\frac{\gamma\vec{\beta}^2}{2})\delta(\beta^2-\vec{\beta}^2)d\vec{\beta} \\ &= \left(\frac{\gamma}{2}\right)^{m/2}\left[\Gamma\left(\frac{m}{2}\right)\right]^{-1}\exp\,(-\frac{\gamma\beta^2}{2})\beta^{m-2}d\beta^2, \quad \vec{\beta}\in\mathbb{R}^m.\end{aligned}$$

Using the integral representation of $f_m^\beta(u)$ we find that

$$\int f_m^\beta(u)dR(\beta^2) = \left(\frac{g}{2}\right)^{m/2}\left[\Gamma\left(\frac{m}{2}\right)\right]^{-1}\exp\,\left(-\frac{gu}{2}\right)u^{m/2-1},$$

where $g = \gamma/(1+\gamma)$ and

$$\int \beta^2 f^\beta_{m+2}(u) dR(\beta^2) = u \int f^\beta_m(u) dR(\beta^2)/(1+\gamma).$$

Hence $\sigma(u) = \pi(u)/(1+\gamma)$. If $\eta(u) = 1$ for all $u > 0$, we have

$$D(1) = 2(1+\gamma)G(1), \quad \rho^2(1) = \frac{2\kappa}{1+\gamma} \int v dR(v).$$

For the best weight function $\eta(u) = \eta_{\text{opt}}(u)$ we obtain

$$\rho^2(\eta_{\text{opt}}) = \frac{2\kappa}{1+\gamma} \int \sigma(u) du = \frac{2\kappa}{1+\gamma} \int v dR(v).$$

Thus $\rho(\eta_{\text{opt}}) = \rho(1)$. We conclude that the weighting does not diminish the limit half-sum of the error probabilities for this special case.

Example 9. Let all blocks contribute identically to the distance between populations, $(\theta^i_1 - \theta^i_2)^2 = J/k, \quad i = 1, \dots, k$. Then $dR(\beta^2) \neq 0$ only at the point $\beta^2 = J/2$. The best weighting function is

$$\eta_{\text{opt}}(u) = \frac{\beta^2 f^\beta_{m+2}(u)}{u f^\beta_m(u)}.$$

Since

$$\int \frac{[f^\beta_{m+2}(u)]^2}{u f^\beta_m(u)} du \geq \frac{1}{m+\beta^2},$$

we have $\rho(\eta_{\text{opt}}) \geq \rho(1)$. Let $m = 1$. Then, in view of (38) we have $\eta_{\text{opt}}(u) = \beta \cdot \text{th}\ (\beta\sqrt{u})/\sqrt{u}$, where $\beta^2 = J/2$. In spite of the identical contributions of blocks to J, the optimal weighting provides the increase of $\rho(\eta)$ and the decrease of $\operatorname*{plim}_{n\to\infty} (\alpha_1 + \alpha_2)/2$ owing to the effect of a suppression of large deviations of estimators.

THEOREM 11.8. *If assumptions A–F hold and the $\rho_{\text{opt}}(u) = 1$ for all $u \geq 0$, then there exists $\gamma > 0$ such that*

$$\frac{dR(v)}{dv} = \left(\frac{\gamma}{2}\right)^{m/2} \left[\Gamma\left(\frac{m}{2}\right)\right]^{-1} \exp\ \left(-\frac{\gamma v}{2}\right) v^{m/2-1}. \tag{54}$$

Proof. We compare (52) and (53). The inequality $\rho(\eta^0) \geq \rho(1)$ is the Cauchy–Bunyakovskii inequality for the functions $\sqrt{\pi(u)}$ and $\sigma(u)/\sqrt{\pi(u)}$. The case of the equality implies $\sqrt{\pi(u)} = \sigma(u)/\sqrt{\pi(u)}$ almost for all $u > 0$. In view of the continuity we obtain $\sigma(u) = C_1\pi(u)$, where $C_1 > 0$. Substituting the expressions (48) for $\sigma(u)$ and $\pi(u)$ we find that

$$C_1\pi(u) = \int (2 + u - m) f_m^\beta(u) dR(\beta^2) + 2u \int \frac{\partial f_m^\beta(u)}{\partial u} dR(\beta^2)$$
$$= \pi(u) + (2 - m)\pi(u)/u + 2\pi'(u), \quad u > 0.$$

Integrating this differential equation we obtain

$$\pi(u) = C_2 u^{m/2} \exp\,(-C_3 u), \quad C_2, C_3 > 0.$$

We substitute $\pi(u)$ from (48), divide by u, and perform the Fourier transformation of the both parts of this equality. It follows that

$$\int \frac{\pi(u)}{u} \exp\,(iut)dt = \int \chi_m^\beta(t) dR(\beta^2) = \frac{\text{const}}{(C_3 - it)^{m/2}}, \quad t \geq 0.$$

Denote $s = t/(1 - 2it)$. Substituting the expression for $\pi(u)$ we obtain

$$\int \exp\,(is\beta^2) dR(\beta^2) = \frac{\text{const}}{C_3 + is(2C_3 - 1)^{m/2}} \;.$$

This relation holds, in particular, on an interval of s Im $s = 1/4$, $|\text{Re } s| < 1/4$. The analytical continuation to Im $s \geq 0$ makes it possible to perform the inverse Fourier transformation

$$\frac{dR(v)}{dv} = C_4 \int_{-\infty}^{+\infty} \frac{\exp\,(-ivs)}{[C_3 - is(2C_3 - 1)]^{m/2}} ds$$
$$= C_5\; v^{m/2-1} \exp\,\left(-\frac{\gamma v}{2}\right),$$

where $C_4, C_5 > 0$ and $\gamma = 2C_3/(1 - 2C_3) > 0$. Normalizing we obtain (54). The proof is complete. □

Thus the distribution (54) is an only limit distribution for which the effect of weighting by Theorem 11.7 gives no gain.

Statistics to Estimate Probabilities of Errors

Usually the observer does not know, even approximately, the true values of either the parameters θ_1 θ_2, J^i, or the distribution $R(\cdot)$. The functions $\sigma(u)$ and $\pi(u)$ involved in Theorems 11.6, 11.7, and 11.8 also are not known. Let us construct their estimators.

THEOREM 11.9. *Under assumptions A–F there exists the limit*

$$\widehat{J}_* = \operatorname*{plim}_{n\to\infty} \sum_{i=1}^{k} \widehat{J^i} = J + 2\kappa m,$$

where $\widehat{J^i}$ *are defined by* (8) *and* J *is defined by* (29).

Proof. By Lemma 11.3, the sum

$$\sum_i \widehat{J^i} = n^{-1} \sum_i (\mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i)^2 + n\omega^i.$$

Using (7) and (5) we rewrite this expression as

$$n^{-1} \sum_i \, [(\mathbf{b}^i)^2 + 2m] + o(1) = \frac{2\kappa}{n} \int (\beta^2 + m) dR_n(\beta^2) + o(1).$$

In the right hand side $R_n(\cdot) \to R(\cdot)$, and we obtain the limit $J + 2\kappa m$ as $n \to \infty$.

The variance of the sum of $\widehat{J^i}$ is a sum of variances which are not greater than $O(n^{-2})$ $E(\mathbf{b}^i + \mathbf{z}_2^i - \mathbf{z}_1^i)^2$ plus the sum of the remainder terms which tends to 0 in probability since its expectation vanishes. The sum of variances of the leading terms vanishes as $O(n^{-1})$. It follows that the sum of $\widehat{J^i}$ tends to $J + 2\kappa m$ in probability. Theorem 11.9 is proved. □.

To choose a better weighing we begin with constructing of an estimator for $R(\cdot)$. Let us describe the set of the observed values $\widehat{J^i}$ of the form (8) in terms of the function

$$Q_n(u) = k^{-1} \sum_i \text{ind } (u \geq n\widehat{J^i}/2), \tag{55}$$

THEOREM 11.10. *Under assumptions A–F there exists a distribution function $Q(u)$ such that*

$$Q(u) = \operatorname*{plim}_{n\to\infty} Q_n(u) = \int F_m^\beta(u) dR(\beta^2), \quad u \geq 0.$$

Proof. The expectation

$$\mathbf{E}\ Q_n(u) = k^{-1} \sum_i \mathbf{P}(n\widehat{J}^i/2 < u).$$

By Theorem 11.4 these probabilities equal $F_m^\beta(u) + o(1)$ as $n \to \infty$. It follows that $\mathbf{E}\ Q_n(u) \to Q(u)$. The variance of $Q_n(u)$ is a sum of variances. One can readily see that it is $O(n^{-1})$. This proves Theorem 11.10.□

From Theorem 11.10 it follows that the function $Q(u)$ is monotone, continuous, increases as $u^{m/2}$ for small $u > 0$, and $Q(u) \to 1$ exponentially as $u \to \infty$. The following integral relation holds:

$$\int_0^\infty (1 - Q(u))du = \int (m + \beta^2) dR(\beta^2).$$

Remark 1. The functions $\sigma(u)$ and $\pi(u)$ can be expressed in terms of the derivatives of $Q(u)$ as follows:

$$\sigma(u) = 2uQ''(u) + (2 + u - m)Q'(u), \quad \pi(u) = uQ'(u). \qquad (56)$$

We can see that the relations (56) can be used to substitute $\sigma(u)$ and $\pi(u)$ in the expressions (49) for $G(\eta)$ and $D(\eta)$ in terms of the known functions $Q(u)$ and $\eta(u)$.

Remark 2. If the function $\eta(u)$ is twice differentiable for $u \geq 0$, then we have

$$\begin{aligned} G(\eta) &= \kappa \int [(u - m)\eta(u) - 2u\eta'(u)] dQ(u) \\ &= \kappa \int_0^\infty [\eta(u) + (u - m - 2)\eta'(u) - 2u\eta''(u)](1 - Q(u))du - k_0, \end{aligned}$$

where $k_0 = \kappa m\eta(0)$, and

$$D(\eta) = 2\kappa \int u\eta^2(u)dQ(u)$$
$$= 2\kappa \int_0^\infty [\eta(u) + 2u\eta'(u)]\eta(u)(1 - Q(u))du. \qquad (57)$$

For a special case when $\eta(u) = 1$ for all $u \geq 0$ we obtain

$$G(1) = J/2 = \kappa \int (u - m)dQ(u),$$

$$D(1) = \widehat{J}_* = 2\kappa \int udQ(u) = 2\kappa \int_0^\infty (1 - Q(u))du,$$

where J is defined by (4) and $\widehat{J}_*$ is defined in Theorem 11.9.

Remark 3. The function of the 'best in the limit' weighting $\eta_{\text{opt}}(u)$ can be written in the form

$$\eta_{\text{opt}}(u) = \frac{2uQ''(u) + (2 + u - m)Q'(u)}{uQ'(u)}$$

and the best in the limit effective Mahalanobis distance is

$$\rho(\eta_{\text{opt}}) = 2\kappa \int_{+0}^\infty \frac{[2uQ''(u) + (2 + u - m)Q'(u)]^2}{uQ'(u)} du.$$

The minimum of the limit half-sum of the error probabilities is

$$\inf_\eta \min_c \operatorname*{plim}_{n\to\infty} (\alpha_1 + \alpha_2)/2 = \Phi(-\rho(\eta_{\text{opt}})/2).$$

Contributions of Variables to Discrimination

If the observer knows exactly the parameters of the populations, then to minimize the discrimination errors one needs all variables. For a sample discrimination rule with no weighting of variables the limit probability of the discrimination error is not minimum, and the problem arises of choosing the best subset of variables minimizing the discrimination errors.

Under the problem setting accepted above, sums of the increasing number of variables produce non-random limit contributions to the discrimination. However, the contributions of separate blocks of variables remain essentially random. To obtain stable recommendations concerning the selection we gather blocks with neighbouring values of J^i and $\widehat{J}^i$ into groups that are sufficiently large to have stable characteristics. Let us investigate the effect of an exclusion of an increasing number $p \leq k$ blocks from the consideration. More precisely, consider two sequences (1) of the discrimination problems such that the second sequence is different by $\tilde{k} = k - p$ blocks of variables with the same other arguments. Assume that $p/k \to \gamma \geq 0$ as $n \to \infty$. In this section we suppose that γ is small. Let us mark the characteristics of the second sequence by the sign tilde.

Contributions by a Priori Information

Suppose that all the excluded blocks have close values of J^i, $J^i \approx J_0$, so that the new limit function is

$$\tilde{R}(v) = \frac{R(v) - \gamma \text{ ind } (v > v_0)}{1 - \gamma},$$

where $v_0 = nJ_0/2$. Let $\delta z = \tilde{z} - z$ denote the change of a value z when we pass from the first sequence to the second one. Assume that γ is small and let us study the effect of the exclusion of variables keeping only first order terms in $\gamma \to 0$. Then

$$\begin{aligned} \gamma = -\delta\kappa/\kappa, \quad \delta R(v) = [\text{ ind } (v - v_0) - R(v)]\delta\kappa/\kappa, \\ \delta G(\eta) = v_0\eta(v_0)\delta\kappa, \qquad \delta D(\eta) = 2(v_0 + m)\eta^2(v_0)d\kappa, \end{aligned} \tag{58}$$

where $G(\eta)$ and $D(\eta)$ are defined by (22). We set $\eta(v) = 1$ for all $v \geq 0$. For a problem with the known set $\{J^i\}$ of contributions we obtain

$$\delta\rho^2(\eta) = \frac{2J}{(J + 2\kappa m)^2}[(J + 4\kappa m)v_0 - mJ]\ \delta\kappa.$$

Remark 4. Let assumptions A–F hold, $\eta(v) = 1$ for all $v \geq 0$, and a small portion of blocks be excluded by a priori values of J^i so that

(58) holds. Then the limit error probability $\alpha_*(\eta) = \Phi(-\rho(\eta)/2)$ is decreased by the exclusion of variables if and only if the inequality holds $v_0 < (J + 4\kappa m)^{-1} mJ$.

Thus we have shown that to decrease the limit probability of the discrimination error, it can be recommended to exclude blocks with

$$J^i < \frac{2m}{n} \frac{J_n}{J_n + 4m\kappa}, \quad \text{where} \quad J_n = \sum_{i=1}^{k} J^i.$$

Contributions by Estimators

We investigate the contribution of variables when a number q of blocks with close sample characteristics are excluded. We assume that $p/k \to \gamma \geq 0$, where γ is small.

Denote

$$\gamma = -\delta\kappa/\kappa, \qquad \delta Q(u) = [\text{ ind } (u \geq u_0) - Q(u)]\delta\kappa/\kappa, \tag{59}$$

where $u_0 = \operatorname*{plim}_{n\to\infty} n\widehat{J}^i/2$ for all excluded blocks. Keeping only first order terms we have

$$\delta G(\eta) = [(u_0 - m)\eta(u_0) - 2u_0\eta'(u_0)]\delta\kappa,$$

$$\delta D(\eta) = 2\delta\kappa \int_0^{u_0} [\eta(u) + 2u\eta'(u)]\eta(u)du = 2u_0\eta^2(u_0)\delta\kappa.$$

Let $\eta(u) = 1$ for all $u > 0$. Then we have

$$\delta J = 2\delta G(1) = 2(u_0 - m)\delta\kappa, \qquad \delta\widehat{J}_* = \delta D(1) = 2u_0\delta\kappa$$

where J and $\widehat{J}_*$ are defined by (4) and Theorem 11.9. We obtain

$$\delta\rho^2(1) = \frac{2J}{(J + 2\kappa m)^2}[u_0(J + 4\kappa m) - 2m(J + 2\kappa m)]\delta\kappa.$$

Remark 5. Let assumptions A–F hold, $\eta(u) = 1$ for all $u > 0$, and a small portion of blocks be excluded by estimators $\widehat{J}^i$ so that (59) holds. Then $\alpha^*(\eta) = \Phi(-\rho(\eta)/2)$ decreases if and only if

$$u_0 < 2m\,\frac{J + 2\kappa m}{J + 4\kappa m}.$$

Thus we have shown that to decrease the limit probability of the discrimination error, it can be recommended to exclude blocks with

$$\widehat{J}^i < \frac{4m}{n}\frac{\widehat{J}_n}{\widehat{J}_n + 2\kappa m}, \quad \text{where} \quad \widehat{J}_n = \sum_{i=1}^{k} \widehat{J}^i.$$

Note that Remarks 4 and 5 are asymptotically equivalent.

Selection of a Large Number of Independent Variables

We now consider a selection of a substantial portion of variables Let q be the number of variables left for the discrimination that were selected by the rule $nJ^i/2 > \tau^2$ and $n\widehat{J}^i > \tau^2$, where τ^2 is an a priori selection threshold, $q \le k$. To investigate the selection effect in our consideration, we take the weighting function of the form $\eta(u) = \text{ind}\ (u \ge \tau^2)$.

Selection by a Priori Information

Remark 6. Under assumptions A–F for $\eta(u) = \text{ind}\ (u \ge \tau^2)$ with $u = nJ^i/2$, the limits exist

$$\delta = \lim_{n\to\infty} \frac{q}{k} = 1 - R(\tau^2).$$

Indeed by (5) we have

$$\frac{q}{k} = k^{-1}\sum_i \eta(nJ^i/2) = \int \eta(v)dR_n(v) = 1 - R_n(\tau^2),$$

where $R_n(v) \to R(v)$ for all $v > 0$.

To treat the selection problem let us redefine

$$G(\delta) = G(\eta), \quad D(\delta) = D(\eta), \quad \rho(\delta) = \rho(\eta),$$
$$J(\delta) = 2\kappa \int v\eta(v)dR(v).$$

By (22) we have

$$G(\delta) = \kappa \int_{v>\tau^2} vdR(v), \quad D(\delta) = 2\kappa \int_{v>\tau^2} (v+m)dR(v).$$

Let us find the threshold τ^2 and the limit portion δ of blocks left for the discrimination that minimize $\alpha_*(\delta) = \Phi(-\rho(\delta)/2)$.

Let b_1 and b_2 denote the left and right boundaries of the distribution $R(v)$: $b_1 = \inf\{v \ge 0\ : R(v) > 0\}$ $\quad b_2 = \inf\{v \ge 0\ : R(v) = 1\}$.

Theorem 11.11. *Suppose assumptions A–F hold and the discrimination function* (9) *is used with the selection coefficients of the form* ind $(nJ^i/2 \geq \tau^2)$. *Then, under the variation of δ with fixed m, κ, and $R(v)$, the condition*

$$b_1 < mJ(1)/(J(1) + 4\kappa m) \tag{60}$$

is sufficient for $\alpha_*(\delta) = \Phi(-\rho(\delta)/2)$ to reach a minimum for δ such that $0 < \delta \leq 1$. The derivative $\alpha'_*(\delta)$ exists almost everywhere for $0 < \delta \leq 1$ and its sign coincides with the sign of the expression $mJ(\delta) - \tau^2(J(\delta) + 4\kappa m\delta)$.

Proof. For $\delta = 1$ we have $\tau^2 = 0$ and $\eta(v) = 1$ for all $v \geq 0$. Obviously,

$$(J(\delta) - J(1))/2\kappa = -\int (1 - \eta(v))v dR(v) \geq -(1 - \delta)\tau^2.$$

The minimum of $\alpha_*(\delta)$ is attained if $\rho^2(\delta) = J^2(\delta)/(J(\delta) + 2m\kappa\delta)$. We find

$$\rho^2(\delta) - \rho^2(1) \geq \text{const } [mJ^2(1) - \tau^2 J(\delta)(J(1) + 2\kappa m) - 2\kappa m\tau^2 J(\delta)].$$

If $\delta \to 1$ then $\tau^2 \to b_1$, $J(\delta) \to J(1)$, and if (60) holds, then there exists $\delta < 1$ such that $\rho^2(\delta) \geq \rho^2(1)$. On the other hand, $\delta \to 0$ implies $J(\delta) \to 0$ and $\rho^2(\delta) \leq J(\delta) \to 0$. The function $J(\delta)$ is continuous for all δ, $0 < \delta \leq 1$. Therefore $\rho^2(\delta)$ reaches a minimum for $0 < \delta < 1$. The derivatives $\alpha'_*(\delta)$ and $\rho'(\delta)$ exist at all points where $\dfrac{d\tau^2}{d\delta}$ exists, i.e., at all points where $dR(\tau^2) > 0$. For these δ,

$$\frac{dJ(\delta)}{d\delta} = 2\kappa\tau^2,$$

$$\frac{d\rho^2(\delta)}{d\delta} = \frac{2\kappa J(\delta)}{(J(\delta) + 2\kappa m\delta)^2}[\tau^2(J(\delta) + 4\kappa m\delta) - mJ(\delta)]. \tag{61}$$

The second statement of the theorem follows. The proof of Theorem 11.11 is complete. □

Example 10. Consider the case of a portion $r \geq 0$ of non-informative blocks. The function $R(v) = r \geq 0$ for $0 \leq v < b_2$

and $R(b_2) = 1$. The value $J(1) = 2\kappa(1-r)b_2$. For $\delta < 1-r$ we have $\tau^2 = b_2$, $J(\delta) = 2\kappa\delta b_2$, $\rho^2(\delta) = \rho^2(1)\delta$, and the increase of δ decreases $\alpha_*(\delta)$. For $r > 0$ and $\delta \geq 1-r$ we have $b_1 = 0$, and (60) is valid. We have $J(\delta) = 2\kappa(1-r)b_2$ independently on δ, and the decrease of δ decreases $\alpha_*(\delta)$. For $r = 0$ we have $b_1 = b_2$, relation (60) does not hold, $J(\delta) = 2\kappa\delta b_2$, and the decrease of δ increases $\alpha^*(\delta)$. The selection is not purposeful if $r = 0$.

Selection of Blocks by Estimators

We consider a selection of $q \leq k$ blocks of variables with sufficiently large values of the statistics $\widehat{J}^i$ of the form (8): $\widehat{J}^i \geq 2\tau^2/n$, where τ^2 is a threshold. This problem is reduced to the discrimination with the weight coefficients $\eta(u) = \text{ind}\,(u \geq \tau^2)$, where $u = n\widehat{J}^i/2, \quad i = 1, \ldots, k$. The number of blocks left in the discrimination function is

$$q = \sum_i \eta(n\widehat{J}^i/2).$$

LEMMA 11.6. *If assumptions A–F hold and the discriminant function* (10) *is used with the weights* ind $(n\widehat{J}^i/2 \geq \tau^2)$ $i = 1, \ldots, k$, *then the limit exists*

$$\delta = \operatorname*{plim}_{n\to\infty} \frac{q}{k} = 1 - Q(\tau^2), \tag{62}$$

where $Q(\cdot)$ is defined in Theorem 11.10.

Proof. Using Theorem 11.3 we find

$$\begin{aligned}
\mathbf{E}\ \frac{q}{k} &= k^{-1}\sum_i \mathbf{E}\ \eta(n\widehat{J}^i/2)k^{-1}\sum_i \int \eta(u)dF_m^{\beta i}(u) + o(1) \\
&= \iint \eta(u)dF_m^{\beta}(u)dR(\beta^2) + o(1) \\
&= \int \eta(u)dQ(u) + o(1) = 1 - Q(\tau^2) + o(1),
\end{aligned}$$

where $\beta^i = \sqrt{J^i/2}$. The ratio q/k is a sum of independent random values, and its variance is not greater than $1/k \to 0$. This proves the lemma. □

One can see that τ^2 is a monotone function of δ decreasing with δ. The function $\eta(\cdot)$ is determined by the value of δ uniquely.

Let us redefine $G(\delta) = G(\eta)$, $D(\delta) = D(\eta)$, $\rho(\delta) = \rho(\eta)$, and $\alpha_*(\delta) = \alpha_*(\eta)$. From (40) and (61) it follows:

Remark 7. Under assumptions A–F with $\eta(u)$ of the form (61) we have

$$G(\delta) = \kappa \int_{\tau^2}^{\infty} \sigma(u)du = \kappa \int \beta^2(1 - F^{\beta}_{m+2}(\tau^2))\ dR(\beta^2),$$

$$D(\delta) = 2\kappa \int_{\tau^2}^{\infty} \pi(u)du = 2\kappa \int [\int_{u>\tau^2} uf^{\beta}_m(u)du]\ dR(\beta^2)$$

$$= 2\kappa \int [m(1 - F^{\beta}_{m+2}(\tau^2)) + \beta^2(1 - F^{\beta}_{m+4}(\tau^2)]\ dR(\beta^2).$$

We use Theorem 11.10 to express these values in terms of the function $Q(u)$.

Remark 8. Under assumptions A–F
with $\eta(u) = \text{ind}\ (n\widehat{J}^i/2 \geq \tau^2)$ we have

$$G(\delta) = \kappa \int_{\tau^2}^{\infty} (u - m)Q'(u)du - 2\tau^2 Q(\tau^2)$$

$$= \kappa(\tau^2 - m)(1 - Q(\tau^2)) + \kappa \int_{\tau^2}^{\infty} (1 - Q(u))du,$$

$$D(\delta) = 2\kappa \int_{\tau^2}^{\infty} uQ'(u)du = 2\kappa(1 - Q(\tau^2)) + \kappa \int_{\tau^2}^{\infty} (1 - Q(u))du.$$

By virtue of Theorem 11.2 the random values α_1 and α_2 converge in probability to the limits (50).

Remark 9. Let us consider the problem of the influence of an informational noise on the discrimination. We modify the sequence of problems (1) by adding a block number $i = 0$ of independent variables and assume that the random vector $\mathbf{x}^0$ is distributed as $\mathbf{N}(\theta^0_\nu, I_r)$, where $\theta^0_\nu \in \mathbb{R}^r$, $\nu = 1, 2$, and I_r is the identity matrix of size $r \times r$. To simplify formulas suppose that θ_1^0 and θ_2^0 do not depend on n. Denote $J^0 = (\theta_1^0 - \theta_2^0)^2$. The discriminant function is modified by an addition of a normal variable distributed as $\mathbf{N}(\pm J^0/2, J^0)$. Suppose that all remaining variables are non-informative: let $|b^i| = 0$ for all $i > 0$ so that $R(v) = 1$ for all $v > 0$. In this case as $n \to \infty$,

$$\alpha_*(\delta) \to \Phi\left(-\frac{J^0}{2\sqrt{J^0 + D(\delta)}}\right)$$

with $D(\delta)$ given by Remark 7. We also have $Q(u) = F_m^0(u), \quad \delta = 1 - F_m^0(\tau^2)$, and $D(\rho) = 2\kappa m\delta + 4\kappa m f_{m+2}^0(\tau^2)$. Here the second summand is added to the variance of $g(\mathbf{x})$; this additional term is produced by the selection of those variables which have the greater deviations of estimators (this effect was analyzed in detail in [39]). For a small portion δ of variables left the selection essentially increases (as $\ln 1/\delta$) the effect of the informational noise.

THEOREM 11.12. *Suppose assumptions A–F hold and the discriminant function* (10) *is used with the weighting coefficients of the form* ind $(n\widehat{J}^i/2 \geq \tau^2)$. *Then, under the variation of* δ *with fixed* m, κ, *and* $R(v)$,

1^0 *the condition*

$$2\int (m+u)dR(u) \int u \exp\,(-u/2)dR(u) \\ < m\int \exp\,(-u/2)dR(u)\int udR(u) \qquad (63)$$

is sufficient for the value $\alpha_*(\delta)$ *to have a minimum for* $0 < \delta < 1$;

2^0 *the derivative* $\alpha_*'(\delta)$ *exists for* $0 < \delta \leq 1$ *and the sign of* $\alpha_*'(\delta)$ *coincides with the sign of the difference*

$$\pi(\tau^2)\int_{\tau^2}^{\infty} \sigma(u)du - 2\sigma(\tau^2)\int_{\tau^2}^{\infty} \pi(u)du. \qquad (64)$$

Proof. If $\delta = 1$, then $\tau^2 = 0$, $G(1) = \kappa\int \beta^2 dR(\beta^2)$, and $D(1) = 2\kappa\int(\beta^2+m)dR(\beta^2)$. Using (48) we calculate the derivatives at the point $\tau^2 = 0$

$$\frac{df_m^\beta(\tau^2)}{d\tau^{m-2}} = a \exp\,\left(-\beta^2/2\right),$$

$$\frac{dG(\delta)}{d\tau^{m+2}} = -\kappa\; a\left(\frac{m}{2}+1\right)^{-1}\int \frac{\beta^2}{m}\exp\,\left(-\beta^2/2\right)dR(\beta^2),$$

$$\frac{dD(\delta)}{d\tau^{m+2}} = -2\kappa\; a\left(\frac{m}{2}+1\right)^{-1}\int \exp\,\left(-\beta^2/2\right)dR(\beta^2),$$

where $a = 2^{-m/2}[\Gamma(m/2)]^{-1}$. Calculating the derivative $\dfrac{d\alpha_*(\delta)}{d\tau^{m+2}}$ at the point $\tau^2 = 0$ we find that this derivative is positive if condition (63) holds. It follows that τ^2 is monotone depending on δ and $\alpha_*(\delta) < \alpha_*(1)$ for some $\delta < 1$, $\tau^2 > 0$. The first statement is proved.

Further, we notice that

$$\rho^2(\delta) = 2\kappa[\int_{\tau^2}^{\infty} \sigma(u)du]^2 \Big/ \int_{\tau^2}^{\infty} \pi(u)du\;.$$

Differentiating $\rho^2(\delta)$ we obtain the second statement of the theorem. This completes the proof. □

Example 11. The function $R(v) = r$ for $0 \le v < b = \beta^2$ and $R(b) = 1$. The condition (62) has the form

$$r > \left(2b/m + \exp\ (b/2) - 1\right)^{-1}(2b/m + 1)\,.$$

The selection is purposeful for sufficiently large r and large b. For the blocks with identical non-zero contributions we have $r = 0$ and (63) does not hold. In this case we find that

$$\frac{\rho^2(\delta)}{\rho^2(1)} = \frac{(1 - F^{\beta}_{m+2}(\tau^2))^2(m+\beta)^2}{m(1 - F^{\beta}_{m}(\tau^2)) + \beta^2(1 - F^{\beta}_{m+4}(\tau^2))}.$$

But $F^{\beta}_{m+4}(\tau^2) \le F^{\beta}_{m+2}(\tau^2)$. Replacing F^{β}_{m+4} by F^{β}_{m+2} we obtain the inequality

$$\frac{\rho^2(\delta)}{\rho^2(1)} \le 1 - F^{\beta}_{m+2}(\tau^2) < 1.$$

The minimum of $\alpha_*(\delta)$ is attained for $\delta = 1$, that is, using all variables.

Example 12. Consider the special limit distribution $R(\cdot)$ when the derivative $R'(v)$ exists and is defined by (54). In this case, $\sigma(u) = \pi(u)/(1+\gamma)$. In the inequality (63) the left hand side is equal to the right hand side, and the sufficient condition for the selection to be purposeful is not satisfied. The value

$$\rho^2(\delta) = \frac{2\kappa}{1+\gamma}\int\limits_{\tau^2}^{\infty} \sigma(u)du.$$

One can see that $\alpha_*(\delta)$ is strictly monotone decreasing with the decrease of δ and the increase of τ^2. The minimum of $\alpha_*(\delta)$ is attained when all variables are used.

Remark 10. Let us rewrite the selection conditions (60) and (63) in the form

$$b < Jm/(J + 4\kappa m),$$

$$2\int v\ \exp\ (-v/2)dR(v) < Jm/(J + 2\kappa m)\ \int \exp\ (-v/2)dR(v).$$

The left hand side of the second inequality has a sense of the mean contribution of weakly discriminant variables. It can be compared with the first inequality. One can see that under a ‘good’ discrimination when $\int v dR(v) \gg m$, the boundary of the purposefulness of a selection using estimators is twice as less than under the selection by parameters.

The sufficient condition (60) for the selection to be purposeful involves values which are usually unknown to the observer. Let us rewrite it in the form of limit functions of estimators. Denote

$$w(u) = 2 \ln \left[u^{1-m/2} \exp\ \left(\frac{u}{2} \right) Q'(u) \right].$$

Remark 11. Suppose conditions A–F are satisfied and the discrimination function (10) is used with the weights ind $(n\widehat{J}^i/2 \geq \tau^2)$ of the blocks. Then, under the variation of δ as m, κ, and $R(\cdot)$ are fixed, the condition

$$(1 - 2w'(0)) \int_0^\infty u dQ(u) > m \tag{65}$$

is sufficient for the $\alpha_*(\delta)$ to attain the minimum for $0 < \delta < 1$.

Indeed, from (63) it follows that for small u we have

$$w'(0) = m^{-1} \int \beta^2 \exp\,(-\beta^2/2) dR(\beta^2) / \int \exp\,(-\beta^2/2) dR(\beta^2).$$

The relation (65) readily follows.

Remark 12. Suppose conditions A–F are satisfied and the discrimination function (10) is used with the weights ind $(n\widehat{J}^i/2 \geq \tau^2)$ of the blocks. Then, under the variation of δ with m, κ, and $R(\cdot)$ fixed, the minimum of $\alpha_*(\delta)$ is attained for $\delta = \delta_{\text{opt}}$ and $\tau = \tau_{\text{opt}}$ such that

$$\delta_{\text{opt}} = 1 - Q(\tau^2_{\text{opt}}) \quad \text{and} \quad w'(\tau^2_{\text{opt}}) = G(\delta_{\text{opt}})/D(\delta_{\text{opt}}).$$

Thus the investigation of the empirical distribution $Q_n(u)$ of the form (55) makes it possible to estimate the effect of the selection of variables by estimators. If inequality (65) holds then the selection is purposeful in the limit. Using equation (64) we can choose the best limit selection threshold τ^2_{opt} and the best limit portion δ_{opt} of chosen variables.