Helmut Rieder
(Editor)

Robust Statistics, Data Analysis, and Computer Intensive Methods

In Honor of Peter Huber's 60th Birthday

Springer

Robustness in Discriminant Analysis

Y.S. Kharin *
Belarussian State University

Abstract

The problems of discriminant analysis in $\mathbf{R}^N$ for L classes are considered for the situations, when hypothetical (classical) model of data is distorted. Classification of distortion types is given. Robustness of classical decision rules is evaluated to the distortions of probability density functions of observations to be classified, and robust decision rules are constructed.

Key words and phrases: Discriminant analysis, types of distortions, breakdown point, robust decision rule.
AMS 1991 subject classifications: Primary 62H; secondary 62C.

1 Introduction

The main title of modern statistical data analysis and discriminant analysis, especially, is formulated by P.J. Huber (1981): "Statistical inferences are based only in part upon the observations. An equally important base is formed by prior assumptions about the underlying situation ...". The system of these prior assumptions is called by hypothetical model of data. Unfortunately, our prior assumptions cannot be exact. That is why real data does not correspond to the assumed hypothetical model absolutely: There are some distortions of the model. Because of this fact the classical discriminant decision rules (DR), which are optimal for the hypothetical model, can be non-stable and, in particular, for distorted models lose their optimality. The following topical problems formulated by P.J. Huber have their impact on modern statistics and are investigated in this paper: (A) description and classification of distortion types in discriminant analysis; (B) evaluation of robustness (stability) of classical discriminant decision rules to distortions; (C) evaluation of breakdown points; (D) construction of robust decision rules (RDR), which are not much sensitive to definite distortion-types; (E) construction of special software for robust discriminant analysis.

In Berger (1985, 1990) and Kadane (1978, 1984) the problems (B) and (C) are considered for the cases when prior probabilities and loss functions of general

* These investigations were supported by Belarussian National Found of Science, Grant F40.263

model of statistical decisions theory are distorted. In Kazakos (1982), Lachenbruch (1966), Lawoko (1986), Randless (1978) and Tiku et al. (1986) the problems (B) and (C) are investigated mainly for "contaminations" of multivariate Gaussian p.d.f.. Here and in Kharin (1981, 1983, 1984, 1991, 1992, 1993) we consider the problem (A) and the problems (B)–(E) for different types of distortions of p.d.f. and data.

2 Hypothetical Model of Discriminant Analysis

In $\mathbf{R}^N$, let the random observations $x = (x_j) \in \mathbf{R}^N$ be registered from $L \geq 2$ classes, with prior probabilities $\pi_1, \ldots, \pi_L$ respectively. An observation from the class Ω_i is a random vector $X_i \in \mathbf{R}^N$ with hypothetical conditional p.d.f. $p_i^0(x)$ for $i \in S = \{1, \ldots, L\}$. When the set of densities $p_i^0(\cdot)$ are unknown, they need to be estimated by a so-called training sample $Z \subset \mathbf{R}^{Nn}$ of size n. Usually Z is classified:

$$Z = Z_1 \cup \ldots \cup Z_L,$$

where $Z_i = \{z_{i1}, \ldots, z_{in_i}\}$ is a random sample of size n_i from Ω_i, and the $Z_1, \ldots, Z_L$ are jointly independent. Otherwise, Z is called non-classified training sample. We consider a given loss matrix $W = (w_{ij})$, where $w_{ij} \geq 0$ is the loss incurred if an observation from Ω_i is assigned to Ω_j, where $i, j \in S$. The problem of discriminant analysis is the classification of a random observation $X \in \mathbf{R}^N$ from one of the L classes Ω_i with minimal expected losses.

The randomized form of a DR is defined as a set of L critical functions:

$$\chi = (\chi_i) = \chi(x, Z) = (\chi_i(x, Z)), \qquad i \in S, \ x \in \mathbf{R}^N, \tag{2.1}$$

where $\chi_i(x, Z) \in [0, 1]$ is the probability of the random decision $d = i$ of assigning x to Ω_i; in particular, we have the identity

$$\chi_1(x, Z) + \cdots + \chi_L(x, Z) \equiv 1. \tag{2.2}$$

When $\chi_i(x, Z) \in \{0, 1\}$ for all $i \in S$, the DR (2.1) takes the non-randomized form:

$$d = d(x, Z) = \sum_{i=1}^{L} i \, \mathrm{I}_{V_i}(x), \qquad x \in \mathbf{R}^N, \ d \in S, \tag{2.3}$$

where $\mathrm{I}_A(x)$ is indicator function of a set $A \subset \mathbf{R}^N$ and

$$V_i = V_i(Z) = \{x : \chi_i(x, Z) = 1\} = \{x : d(x, Z) = i\} \subset \mathbf{R}^N$$

is the region in observation space, where the decision $d = i$ is made.

Denoting expectation by $\mathrm{E}\{\cdot\}$, the risk functional—expected loss—is

$$r = r(\chi; \{p_i^0\}) = \sum_{i,j \in S} \pi_i w_{ij} \, \mathrm{E}\Big\{ \int_{\mathbf{R}^N} \chi_j(x, Z) p_i^0(x) \, dx \Big\} \geq 0. \tag{2.4}$$

A Bayesian decision rule (BDR) minimizes the risk (2.4) for given conditional p.d.f. $\{p_i^0(\cdot)\}$, prior probabilities $\{\pi_i(\cdot)\}$, and loss matrix W, and takes the form (2.3), which is

$$d = d_0(x) = \sum_{i\in S} i \prod_{\substack{k\in S\\ k\neq i}} \mathrm{I}(f_{ki}^0(x) \geq 0), \tag{2.5}$$

where $\mathrm{I}(\cdot \geq 0) = \mathrm{I}_{[0,\infty)}(\cdot)$ is the Heavyside unit function, and

$$f_{ki}^0(x) = \sum_{l\in S} c_{lki} p_l^0(x), \qquad c_{lki} = \pi_l(w_{lk} - w_{li}). \tag{2.6}$$

Its risk—minimal possible value—is called Bayesian risk:

$$r_0 = \sum_{i\in S} \pi_i r_{0i}; \qquad r_{0i} = \sum_{j\in S} w_{ij} \int_{V_j^0} p_i^0(x)\, dx. \tag{2.7}$$

When $\{p_i^0(\cdot)\}$ are unknown, the family of plug-in decision rules

$$\hat{d}(x, Z)\colon \mathbf{R}^N \times \mathbf{R}^{Nn} \longrightarrow S$$

is used in practice: A plug-in DR is defined by substituting any consistent statistical estimator $\hat{p}_i(\cdot)$, based on the training sample Z, for true values $p_i^0(\cdot)$ respectively, in the BDR (2.5) and (2.6).

3 Types of Distortions

The types of distortions of the hypothetical model in discriminant analysis can be classified into three subsets:

D.1 Small-sample effects;

D.2 Distortions of models for observations to be classified;

D.3 Distortions of models for training samples.

D.1 includes the distortions caused by the substitution of $\hat{p}_i(\cdot)$ for the true p.d.f. $p_i^0(\cdot)$ in a BDR. There is a well-known result of Glick (1972) that, under weak conditions, the risk r of a plug-in DR $\hat{d}(\cdot)$ converges to r_0 as $n \to \infty$. But for "small" sample sizes n the deviation $r - r_0 > 0$ can be significant for practice and will depend on the type of estimators: parametric (D.1.1) or non-parametric (D.1.2). For the investigation of the difference $r - r_0$ the method of risk asymptotic expansions used by Okamoto (1963), Raudis (1976), Kharin (1981, 1983) and other authors turned out fruitful.

In D.2 the hypothetical p.d.f. $p_i^0(\cdot)$ are distorted: an observation from Ω_i in reality may have any p.d.f.

$$p_i(\cdot) \in \mathcal{P}_i(\varepsilon_{+i}), \qquad i \in S, \tag{3.1}$$

where $\mathcal{P}_i(\varepsilon_{+i})$ is a suitable set of distorted densities for the class Ω_i with distortion radius $\varepsilon_{+i} \geq 0$; if $\varepsilon_{+i} = 0$, then $\mathcal{P}_i(0) = \{p_i^0(\cdot)\}$ is the one-point set consisting of the hypothetical p.d.f.. Distortions of type D.2.1 and D.2.2 occur if, respectively, a parametric or a non-parametric formulation of the set $\mathcal{P}_i(\varepsilon_{+i})$ is employed.

D.2.1 includes 3 main subsets of distortions:

D.2.1.1 Errors in parameters assignment;

D.2.1.2 Finite mixtures of ε-close distributions;

D.2.1.3 Additive distortions of observations.

D.2.2 includes 4 main subsets of distortions (Kharin (1992)):

D.2.2.1 Huber type of distortions;

D.2.2.2 Distortions in L_2 (and χ^2) metric;

D.2.2.3 Distortions in Kolmogorov metric;

D.2.2.4 Random distortions of densities.

In D.3 the hypothetical assumptions about the training samples Z, $Z_1, \ldots, Z_L$ are broken:

D.3.1 Parametric ε-nonhomogeneity of Z_i $(i \in S)$;

D.3.2 Dependence of sample elements;

D.3.3 Misclassification of the training sample;

D.3.4 Outliers in the training sample;

D.3.5 Missing values of feature variables.

It should be noticed that the problem of identification of these distortions becomes essential for applied statistical analysis.

4 Decision Rules Robustness to the Distortions of Observations to be Classified

In accordance with the classification scheme given in the previous section, we investigate here the distortions of the following two types: D.2.2.1 and D.2.2.2.

Let the conditional p.d.f. $p_i(\cdot)$ be subjected to Huber type (D.2.2.1) distortions (3.1), where

$$\mathcal{P}_i(\varepsilon_{+i}) = \{p_i(\cdot) \mid p_i(x) = (1-\varepsilon_i)p_i^0(x) + \varepsilon_i h_i(x),\ 0 \leq \varepsilon_i \leq \varepsilon_{+i}\}, \qquad (4.1)$$

where $h_i(\cdot)$ may be an arbitrary density of "contaminating" distribution and $\varepsilon_{+i} < 1$ is the distortion (contamination) radius $(i \in S)$. This means that the class Ω_i of observed objects actually consists of two subclasses:

$\Omega_i = \Omega_i^0 \cup \Omega_i^h$ such that $\Omega_i^0 \cap \Omega_i^h = \emptyset$, where Ω_i^0 is the most frequently observed (well-studied) subclass, whereas Ω_i^h is the seldom observed (non-studied) subclass. An observation from Ω_i^0 follows the known (or at least, estimated) density $p_i^0(\cdot)$, and an observation from Ω_i^h the unknown density $h_i(\cdot)$. An object from Ω_i belongs to Ω_i^0 with probability $1-\varepsilon_i$, and to Ω_i^h with probability ε_i.

To evaluate the stability of any DR χ given by (2.1) and (2.2) we define the functional of maximum risk:

$$r_+ = r_+(\chi) = \sup\Big\{r(\chi; \{p_i(\cdot)\}) \;\Big|\; p_i(\cdot) \in \mathcal{P}_i(\varepsilon_{+i})\Big\}. \tag{4.2}$$

The coefficient of unstability for the BDR (2.5) and (2.6), if $r_0 > 0$, is:

$$\kappa_+ = \kappa_+(\chi^0) = (r_+(\chi^0) - r_0)/r_0 \geq 0. \tag{4.3}$$

For given $\delta > 0$, the δ-admissible distortion radius by definition is:

$$\varepsilon_+(\delta) = \inf\Big\{\alpha \;\Big|\; \sup_{\{0 \leq \varepsilon_{+i} \leq \alpha\}} \kappa_+(\chi^0) \geq \delta\Big\}, \tag{4.4}$$

and the breakdown point is

$$\varepsilon_+^* = \inf\Big\{\alpha \;\Big|\; \sup_{\{0 \leq \varepsilon_{+i} \leq \alpha\}} r_+(\chi^0) \geq r^*\Big\}, \tag{4.5}$$

where $r^* = \sum_{i,j=1}^{L} \pi_i w_{ij}/L$ is the value of risk for equiprobable guesses.

Define a RDR $\chi^* = (\chi_1^*(x), \ldots, \chi_L^*(x))$ as a DR that minimizes the maximum risk:

$$r_+^* = r_+(\chi^*) = \inf_{\chi} r_+(\chi). \tag{4.6}$$

Introduce the notations: $w_{i+} = \max_{j \in S} w_{ij}$, and

$$\overline{w}_+ = \sum_{i \in S} \pi_i w_{i+}, \qquad q(p_i^0; \chi_j) = \int_{\mathbf{R}^N} \chi_j(x) p_i^0(x)\, dx \in [0,1].$$

Theorem 4.1 *If Huber type distortions of form (3.1) and (4.1) occur, then the maximum risk value of a BDR (2.5) and (2.6) is given by*

$$r_+(\chi^0) = r_0 + \sum_{i \in S} \varepsilon_{+i} \pi_i (w_{i+} - r_{0i}).$$

PROOF The proof is by maximization of the risk functional (2.4) in the equivalent form

$$r(\chi^0; \{p_i\}) = r_0 - \sum_{i \in S} \varepsilon_i \pi_i r_{0i} + \sum_{i \in S} \varepsilon_i \pi_i \sum_{j \in S} w_{ij} q(h_i; \chi_j^0)$$

with respect to contaminating densities $h_i(\cdot)$ and radii ε_i. ////

Corollary 4.2 *For a BDR (2.5) and (2.6), the δ-admissible distortion radius (4.4) and breakdown point (4.5) are*

$$\varepsilon_+(\delta) = \min\Big\{1,\ \delta r_0 \Big/ \sum_{i\in S} \pi_i(w_{i+} - r_{0i})\Big\},$$
$$\varepsilon_+^* = (r^* - r_0)/(\overline{w}_+ - r_0).$$

Corollary 4.3 *For a* 0, 1 *-loss matrix, the risk r becomes the classification error probability, and we have*

$$\varepsilon_+(\delta) = \min\{1,\ \delta\, r_0/(1-r_0)\}, \qquad \varepsilon_+^* = 1 - \big(L(1-r_0)\big)^{-1}.$$

Notation: $\delta_{ij} = \mathrm{I}(i=j)$ [Kronecker symbol].

Theorem 4.4 *Under Huber type distortions, for every $x \in \mathbf{R}^N$, a RDR has the following form:*

$$\chi_i^*(x) = \delta_{i,d_*(x)}, \qquad d_*(x) = \arg\min_{k\in S} \sum_{j=1}^{L} \pi_j(1-\varepsilon_{+j})p_j^0(x)w_{jk}.$$

Proof By solving the minimax problem (4.2) and (4.6); for details, the reader is conferred to Kharin (1992; Theorem 5.2 and proof, p 108). ////

Corollary 4.5 *If all classes Ω_i are equidistorted (that is, $\varepsilon_{+i} = \varepsilon_+ \ \forall i \in S$) then the notions of RDR and BDR coincide.*

By the method of risk asymptotic expansion with respect to $\varepsilon_+ = \max \varepsilon_{+i}$ (cf. Kharin (1992)) it can be shown that

$$r_+(\chi^*) = r_+(\chi^0) - \sum_{l,m\in S} \rho_{lm}\pi_l\pi_m\varepsilon_{+l}\varepsilon_{+m} + \mathrm{O}(\varepsilon_+^3), \tag{4.7}$$

where ρ_{lm} are expansion coefficients, not dependent on ε_{+l}. The quadratic form in (4.7) is non-negative by definition, so the gain is

$$b = r_+(\chi^0) - r_+(\chi^*) \geq 0.$$

For example, if the hypothetical model is the linear Fisher model with two classes:

$$p_i^0(x) = \varphi_N(x \mid a_i, B), \qquad i \in S = \{1,2\} \tag{4.8}$$

where a_i and B are the hypothetical mean vector and covariance matrix, respectively, then

$$b = \frac{\sqrt{\pi_1(1-\pi_1)}\,(\varepsilon_{+2} - \varepsilon_{+1})^2}{2\sqrt{2\pi}\,\Delta\exp(\Delta^2/8)}$$

with the Mahalanobis interclass distance $\Delta = \sqrt{(a_2-a_1)^{\mathsf{T}} B^{-1}(a_2-a_1)}$.

Consider now the case D.2.2.2 of distortions (3.1), for $i \in S$, in L_2-metric:

$$\mathcal{P}_i(\varepsilon_{+i}) = \left\{ p_i(\cdot) \,\middle|\, \int_{\mathbf{R}^N} \frac{(p_i(x) - p_i^0(x))^2}{\psi_i(x)}\, dx = \varepsilon_i^2, \ 0 \le \varepsilon_i \le \varepsilon_{+i} \right\}, \qquad (4.9)$$

where $\psi_i(\cdot) \ge 0$ is a non-negative normed weight function,

$$\int_{\mathbf{R}^N} \psi_i(x)\, dx = 1.$$

If $\psi_i(\cdot) \equiv p_i^0(\cdot)$, then we obtain the ε_{+i}-neighbourhood (4.9) of the hypothetical p.d.f. in χ^2-metric.

Notation: Let $\mathrm{E}_{\psi_i}\{\cdot\}$ denote expectation w.r.t. the density $\psi_i(\cdot)$ and, for $i \in S$, define radii $\varepsilon_i^* \ge 0$ by

$$\varepsilon_i^* = \left| \mathrm{E}_{\psi_i} \left\{ \Big(\sum_{k \in S} w_{ik}(\chi_k(x) - q(\psi_i; \chi_k)) \Big)^2 \right\} \right|^{1/2} \times \left| \left\{ \inf_x \Big(\frac{\psi_i(x)}{p_i^0(x)} \sum_{l \in S} w_{il}(\chi_l(x) - q(\psi_i; \chi_l)) \Big) \right\}^- \right|^{-1}$$

where y^-, for of any number y, denotes the negative part.

Theorem 4.6 *If distortions (3.1) and (4.9) are present with $\varepsilon_{+i} \le \varepsilon_i^*$ $(i \in S)$, then for an arbitrary DR $\chi(\cdot)$ the maximum risk value (4.2) is*

$$r_+(\chi) = r(\chi; \{p_i^0\}) + \sum_{i \in S} \varepsilon_{+i} \pi_i \left| \mathrm{E}_{\psi_i} \left\{ \Big(\sum_{k \in S} w_{ik}(\chi_k(x) - q(\psi_i; \chi_k)) \Big)^2 \right\} \right|^{1/2}.$$

Proof Maximization in (4.2) under the constraints (4.9) by the method of indefinite Lagrange multipliers. Then maximize w.r.t. $\varepsilon_i \in [0, \varepsilon_{+i}]$. For details, the reader is conferred to Kharin (1992; Theorem 5.4 and proof, pp 115–116).////

Corollary 4.7 *In the case of the χ^2-metric and for radii*

$$\varepsilon_{+i} \le \varepsilon_i^* = \sqrt{\sum_{j \in S} \Big(\frac{w_{ij}}{r_{0i}}\Big)^2 q(p_i^0; \chi_j) - 1} \qquad (i \in S),$$

the coefficient of unstability (4.3) for a BDR is given by

$$\kappa_+(\chi^0) = \sum_{i \in S} \pi_i \varepsilon_{+i} \sqrt{\sum_{j \in S} \Big(\frac{w_{ij} - r_{0i}}{r_0} \Big)^2 q(p_i^0; \chi_j^0)} \ge 0.$$

Corollary 4.8 *For a* $0,1$ *-loss matrix* W *, under the conditions of Corollary 4.7, we have*

$$\varepsilon_i^* = \sqrt{q(p_i^0;\chi_i^0)/(1-q(p_i^0;\chi_i^0))}, \qquad i \in S,$$
$$\kappa_+(\chi^0) = \frac{1}{r_0}\sum_{i\in S}\pi_i\varepsilon_{+i}\sqrt{q(p_i^0;\chi_i^0)(1-q(p_i^0;\chi_i^0))}\ ,$$

and breakdown point

$$\varepsilon_+^* = (1-L^{-1}-r_0)\Big(\sum_{i\in S}\pi_i\sqrt{q(p_i^0;\chi_i^0)(1-q(p_i^0;\chi_i^0))}\Big)^{-1}.$$

Corollary 4.9 *For a BDR* $\chi^0(\cdot)$ *, the hypothetical Gaussian model (4.8), and a* $0,1$ *-loss matrix, in case* $L=2$ *and if the classes* Ω_1 *and* Ω_2 *are equiprobable* $(\pi_1=\pi_2=\frac{1}{2})$ *and equidistorted* $(\varepsilon_{+1}=\varepsilon_{+2})$ *, then the* δ *-admissible distortion radius and breakdown point are*

$$\varepsilon_+(\delta) = \delta\big(1/\Phi(\Delta/2)-1\big)^{1/2},$$
$$\varepsilon_+^* = \big(\Phi(\Delta/2)-1/2\big)\Big(\Phi(\Delta/2)\big(1-\Phi(\Delta/2)\big)\Big)^{-1/2}.$$

Introduce the notations:

$$a_{kj}(x;\chi) = \sum_{i\in S}\pi_i\varepsilon_{+i}w_{ik}w_{ij}c_i(\chi)\psi_i(x),$$
$$b_j(x;\chi) = \sum_{i\in S}\pi_i w_{ij}\Big(p_i^0(x)-\varepsilon_{+i}c_i(\chi)\psi_i(x)\sum_{k\in S}w_{ik}q(\psi_i;\chi_k)\Big),$$
$$c_i(\chi) = \Big|\mathrm{E}_{\psi_i}\Big\{\Big(\sum_{k\in S}w_{ik}\big(\chi_k(x)-q(\psi_i;\chi_k)\big)\Big)^2\Big\}\Big|^{1/2}, \qquad i,j,k\in S. \quad (4.10)$$

Theorem 4.10 *Under the assumptions of Theorem 4.6, the set of robust critical functions* $\chi^* = \{\chi_j^*(x)\}$ *are the solution of the following optimization problem:*

$$\int_{\mathbb{R}^N}\Big(\sum_{k,j=1}^{L}a_{kj}(x;\chi)\chi_j(x)\chi_k(x)+\sum_{j=1}^{L}b_j(x;\chi)\chi_j(x)\Big)dx \to \min_{\chi} \qquad (4.11)$$

under the constraints (2.2).

PROOF Putting the "least favourable density" found in Theorem 4.6 into the risk functional (2.4) and using the notations (4.10), we get the minimization problem (4.6) in equivalent form (4.11). ////

The optimization problem (4.11) can be considered as perturbation of the quadratic optimization problem w.r.t. the coefficients $\{a_{kj},\ b_j\})$. To find the RDR χ^* as solution to (4.11), we use the method of successive approximations.

Our results of robustness investigation in discriminant analysis are used in the applied program package ROSTAN (RObust STatistical ANalysis) for IBM PC/AT published by the Department of Mathematical Modeling and Data Analysis in Belarussian State University.

References

[1] Berger, J.O. (1985): *Statistical decision theory and Bayesian analysis.* Springer-Verlag, New York.

[2] Berger, J.O. (1990): Robust Bayesian analysis. *J. Statist. Plann. Inference* **25** 303–328.

[3] Glick, N. (1972): Sample-based classification procedures derived from density estimators. *J. Amer. Statist. Assoc.* **67** 116–122.

[4] Huber, P.J. (1981): *Robust Statistics.* John Wiley, New York.

[5] Kadane, J.B. et al. (1978): Stable decision problems. *Ann. Statist.* **6** 1095–1110.

[6] Kadane, J.B. (ed., 1984): *Robustness of Bayesian Analysis.* North-Holland, New York.

[7] Kazakos, D. (1982): Statistical discrimination using inaccurate models. *IEEE Trans.* **IT-28** 720–728.

[8] Kharin, Y.S. (1981): About statistical classification accuracy at MC-estimators using. *Probability Theory and Its Applications* **26** 866–867.

[9] Kharin, Y.S. (1983): Investigation and optimization of Rosenblatt-Parzen classificator by asymptotic expansions. *Automatics and remote control* 1 91–100.

[10] Kharin, Y.S. (1983): About decision rule robustness under misclassifications presence in training samples. *Automatics and remote control* 11 100–110.

[11] Kharin, Y.S. (1984): Robustness investigation for the decision rules by risk asymptotic expansion method. In *Proceedings of the Third Prague Symposium on Asymptotic Statistics (P. Mandl ed.)*, 309–317. H.X.-Oxford, Elsevier, Amsterdam.

[12] Kharin, Y.S. et al. (1991): Asymptotic robustness of discriminant procedures for dependent and non-homogeneous observations. In *Probability Theory and Mathematical Statistics* 602–610. Proc. of the 5 Vilnius Conf., Utrecht, Netherlands.

[13] Kharin, Y.S. (1992): *Robustness in Statistical Pattern Recognition.* Minsk, Universitetskoje.

[14] Kharin, Y.S. and Zhuk, E.E. (1993): Asymptotic robustness in cluster-analysis for the case of Tukey-Huber distortions. In *Information and Classification* (Opitz, O., Lausen, B., Klar, R., eds.), 31–39. Springer-Verlag, New York.

[15] Lachenbruch, P.A. (1966): Discriminant analysis when the initial samples are misclassified. *Technometrics* **8** 657–662.

[16] Lawoko, C.R.O. et al. (1986): Asymptotic error rates of the W and Z statistics when the training observations are dependent. *Pattern recognition* **18** 467–471.

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

[18] Randless, R.H. (1978): Generalized linear and quadratic discriminant functions using robust estimates. *J. Amer. Statist. Assoc.* **73** 564–568.

[19] Raudis, S.H. (1976): Sample size finiteness in classification problems. *Statistical control problems* **18** 1-180.

[20] Tiku, M.L. et al. (1986): *Robust Inference*. Marcel Dekker, New York.