CHAPTER X

POPULATION FREE QUALITY OF LINEAR DISCRIMINATION

The quality of the linear discrimination procedure is usually determined by functionals which can be expressed in terms of the conditional expectation and variance of the discriminant function under chosen samples. In this chapter we investigate these random values depending on unknown parameters and our purpose is to find their expectation values and suggest their estimators. In order to separate results from details of derivations, we first set the discriminant problem, formulate results, and discuss conclusions. Then we present proofs.

Problem setting

Let $\mathbf{x}$ be observation vectors with n components in populations $\mathfrak{S}_\nu$ (we write $\mathbf{x} \in \mathfrak{S}_\nu$), $\nu = 1, 2$. Let $\mathfrak{X}_\nu = \{\mathbf{x}_m\}$ be samples from populations $\mathfrak{S}_\nu$ of size $N_\nu > 1$, $\nu = 1, 2$, $N = N_1 + N_2$. The observer starts from sample means and sample covariance matrices

$$\overline{\mathbf{x}}_\nu = \frac{1}{N_\nu} \sum_{\mathbf{x}_m \in \mathfrak{X}_\nu} \mathbf{x}_m, \quad C_\nu = \frac{1}{N_\nu - 1} \sum_{\mathbf{x}_m \in \mathfrak{X}_\nu} (\mathbf{x}_m - \overline{\mathbf{x}}_\nu)(\mathbf{x}_m - \overline{\mathbf{x}}_\nu)^T, \tag{1}$$

where m runs from 1 to N. Define pooled covariance matrix

$$C = \frac{(N_1 - 1)C_1 + (N_2 - 1)C_2}{N - 2} \tag{2}$$

and its resolvent

$$H = H(t) = (I + tC)^{-1}.$$

We consider the class $\mathfrak{K}$ of the discriminant functions

$$w(\mathbf{x}) = (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \Gamma(C)(\mathbf{x} - (\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2)/2), \tag{3}$$

where

$$\Gamma(C) = \int t(I + tC)^{-1}\, d\eta(t) \tag{4}$$

is a matrix that is diagonalized together with C and has $\Gamma(\lambda_i)$ as eigenvalues, where λ_i, $i = 1, \ldots, n$, are eigenvalues of C; assume that the function $\eta : \mathbb{R}^1 \to \mathbb{R}^1$ has a finite variation on $[0, \infty)$ and a sufficient number of moments $\eta_k = \int t^k\, |d\eta(t)|$, $k = 1, 2, \ldots$.

We restrict the populations with the only requirement that the four moments of all variables exist. We have

$$\begin{aligned} G_\nu &= \mathbf{E}_\nu\ (w(\mathbf{x}) \mid \mathfrak{X}_1, \mathfrak{X}_2, \mathbf{x} \in \mathfrak{S}_1) \\ &= (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \Gamma(C)(\mathbf{a}_\nu - (\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2)/2), \quad \nu = 1, 2, \end{aligned} \tag{5}$$

and

$$\begin{aligned} D_\nu &= \mathrm{var}_\nu(\cdot)\ (w(\mathbf{x}) \mid \mathfrak{X}_1, \mathfrak{X}_2, \mathbf{x} \in \mathfrak{S}_\nu) \\ &= (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T \Gamma(C)\Sigma_\nu\Gamma(C)(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2), \quad \nu = 1, 2, \end{aligned} \tag{6}$$

where the conditional expectation $\mathbf{E}_\nu$ and variance $\mathrm{var}_\nu(\cdot)$ are calculated for $\mathbf{x} \in \mathfrak{S}_\nu$ under fixed samples $\mathfrak{X}_1$ and $\mathfrak{X}_2$, $\nu = 1, 2$.

Suppose the discriminant rule is $w(\mathbf{x}) \geq \theta$ against $w(\mathbf{x}) < \theta$, where θ is the classification threshold. If distributions are normal, then, obviously, the (sample dependent) probabilities of errors are

$$\begin{aligned} \alpha_1 &\stackrel{\text{def}}{=} \mathbf{P}(w(\mathbf{x}) \leq \theta \mid \mathbf{x} \in \mathfrak{S}_1) = \Phi(-(G_1 - \theta)/D_1), \\ \alpha_2 &\stackrel{\text{def}}{=} \mathbf{P}(w(\mathbf{x}) > \theta \mid \mathbf{x} \in \mathfrak{S}_2) = \Phi((G_2 - \theta)/D_2). \end{aligned} \tag{7}$$

Denote

$$\mathbf{a}_\nu = \mathbf{E}\ \mathbf{x}, \quad \Sigma_\nu = \mathrm{cov}\ (\mathbf{x}, \mathbf{x})$$

for $\mathbf{x} \in \mathfrak{S}_\nu$, $\nu = 1, 2$, and let

$$\overline{\mathbf{x}} = \overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2, \qquad k(t) = t\overline{\mathbf{x}}^T H(t)\overline{\mathbf{x}}.$$

Define

$$\begin{aligned} g_\nu(t) &= t\ (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T H(t)(\mathbf{a}_\nu - (\overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2)/2), \\ d_\nu(t, t') &= tt'\ \overline{\mathbf{x}}^T H(t)\Sigma_\nu H(t')\overline{\mathbf{x}}, \quad \nu = 1, 2. \end{aligned} \tag{8}$$

The parameters (5) and (6) are equal to

$$G_\nu = \int g_\nu(t)\, d\eta(t), \quad D_\nu = \int d_\nu(t,t')\, d\eta(t) d\eta(t'), \quad \nu = 1,2. \quad (9)$$

Denote

$$y = n/(N-2), \quad y_\nu = n/(N_\nu - 1), \quad \rho_\nu = N_\nu/(N-2), \quad \nu = 1,2,$$
$$h(t) = n^{-1}\text{tr } H(t), \quad s_\nu = s_\nu(t) = 1 - t\mathbf{E} \text{ tr } H(t)C_\nu/(N-2),$$
$$R = (I + t_1 s_1 \Sigma_1 + t_2 s_2 \Sigma_2)^{-1}. \quad (10)$$

We study expectation values and variance of the discriminant function first for normal distributions; then we will perform the generalization using the Normal Evaluation Principle that was developed in Chapter 4.

Leading Parts of Functionals for Normal Populations

To estimate the remainder terms for normal populations, we introduce a single scalar parameter

$$M = \max\, (3\|\Sigma_1\|^2, 3\|\Sigma_2\|^2, |\mathbf{a}_1 - \mathbf{a}_2|^4)$$

(here and in the following, squares of vectors denote squares of their lengths). To be more concise in upper estimates of the remainder terms, denote

$$\tau = \sqrt{M}t, \qquad n_0 = \min\,(n, N_1 - 1, N_2 - 1), \quad \omega_k = a \max\,(1, \tau^k)/n_0, \quad (11)$$

where a and k are non-negative numerical constants.

We start from the results of the investigation of spectral functions of high-dimensional pooled sample covariance matrices that were obtained in Chapter 3. This investigation was carried out under an assumption of the population normality. We also begin with the discriminant problem for normal populations with different (generally speaking) covariance matrices. Then we generalize our results using the Normal Evaluation Principle offered in Chapter 4 to arbitrary populations with four moments of variables. Let us cite some consequences of theorems proved in Chapter 3 in the form of a lemma.

LEMMA 10.1. (corollary of Theorems 3.1 and 3.2).
If populations $\mathfrak{S}_\nu$ *are normal* $\mathbf{N}(\mathbf{a}_\nu, \Sigma_\nu)$, $\nu = 1, 2$, *and* $y < 1$, *then*

1^o $s_\nu(t) \geq (1 + \tau y)^{-1}$, $\nu = 1, 2$,

$$1 - y + yh(t) = \frac{(N_1 - 1)s_1(t) + (N_2 - 1)s_2(t)}{N - 2};$$

2^o $\mathbf{E}H = R + \Omega$, $\|\Omega\|^2 \leq \omega_6$; var $(\mathbf{e}^T H\mathbf{e}) \leq \tau^2/N$;

3^o $1 - s_\nu(t) = ts_\nu(t)(N - 2)^{-1}\mathrm{tr}\ (\Sigma_\nu R) + \varepsilon_\nu$, $\varepsilon_\nu^2 \leq \omega_8$, $\nu = 1, 2$;

4^o var $(t_\nu \bar{\mathbf{x}}_\nu^T H \bar{\mathbf{x}}_\nu) \leq a(1 + \tau^2)/N_\nu$, $\mathbf{x} \in \mathfrak{S}_\nu$, $\nu = 1, 2$;

where $H = H(t)$ *and* a *is a numerical coefficient.*

We consider estimators of non-random $h(t)$, $s_\nu(t)$, and of functions $g_\nu(t)$ and $d_\nu(t, t')$ of the form

$$\widehat{h}(t) = n^{-1}\mathrm{tr}\ H(t), \quad \widehat{s}_\nu(t) = 1 - t\ \mathrm{tr}\ (H(t)C_\nu)/(N - 2),$$
$$\widehat{g}_\nu(t) = 1/2\ k(t) - (1 - \widehat{s}_\nu(t))/\rho_\nu \widehat{s}_\nu(t),$$
$$\widehat{d}_\nu(t, t') = tt'\overline{\mathbf{x}}^T(I + tC)^{-1}C_\nu(I + t'C)^{-1}\overline{\mathbf{x}}, \quad \nu = 1, 2. \tag{12}$$

The following statements relate these statistics to functions of true covariance matrices.

THEOREM 10.1. *If populations are normal* $\mathfrak{S}_\nu = \mathbf{N}(0, \Sigma_\nu)$, *the inequality holds* $n < N_\nu - 1$ *for* $\nu = 1, 2$, *and* $t \geq 0$, *then*

1^o $\mathbf{E}\ (\widehat{h}(t) - h(t))^2 \leq o_1 = \omega_2$;

2^o $\mathbf{E}\ (\widehat{s}_\nu(t) - s_\nu(t))^2 \leq o_2 = \omega_2$;

3^o $\mathbf{E}\ k(t) = t\mathbf{a}^T R\mathbf{a} + \dfrac{1 - s_1}{\rho_1 s_1} + \dfrac{1 - s_2}{\rho_2 s_2} + o_3$, where $o_3^2 \leq \omega_{10}$;

4^o $2\mathbf{E}\ g_\nu(t) = t\mathbf{a}^T R\mathbf{a} + \dfrac{1 - s_1}{\rho_1 s_1} + \dfrac{1 - s_2}{\rho_2 s_2} + o_4$, where $o_4^2 \leq \omega_{10}$;

5^o $(1 - y_\nu)^2\mathbf{E}\ (\widehat{g}_\nu(t) - g_\nu(t))^2 \leq o_5 = \omega_{10}$;

6^o $(1 - y_\nu)^2\mathbf{E}\ \left|\dfrac{\widehat{d}_\nu(t, t')}{\widehat{s}_\nu(t)\widehat{s}_\nu(t')} - d_\nu(t, t')\right|^2 \leq o_6 = \omega_{12}$, $\nu = 1, 2$.

The problem to estimate two first moments and variance of the discriminant functions defined by (5) and (6) can be solved as follows. We consider the statistics

$$\widehat{G}_\nu = \widehat{G}_\nu(\eta) = \int \widehat{g}_\nu(t)\, d\eta(t),$$
$$\widehat{D}_\nu = \widehat{D}_\nu(\eta) = \iint \frac{\widehat{d}_\nu(t,t')}{\widehat{s}_\nu(t)\widehat{s}_\nu(t')}\, d\eta(t)d\eta(t'), \quad \nu = 1,2.$$

THEOREM 10.2. *If populations are normal* $\mathfrak{S}_\nu = \mathbf{N}(0,\Sigma_\nu)$ *and* $n < N_\nu - 1$, $\nu = 1,2$, *then*

1^o $(1-y_\nu)^2(\mathbf{E}\ \widehat{G}_\nu - \mathbf{E}\ G_\nu)^2 \le o_1 = a\eta_{10}/n_0,$

2^o $\mathrm{var}\ G_\nu \le a\eta_4/n_0, \qquad (1-y_\nu)^2\ \mathrm{var}\ \widehat{G}_\nu \le o_2 = a\eta_4/n_0,$

3^o $(1-y_\nu)^4(\mathbf{E}\ \widehat{D}_\nu - \mathbf{E}\ D_\nu)^2 \le o_3 = a\eta_{12}/n_0,$

4^o $\mathrm{var}\ D_\nu \le a\eta_6/n_0, \qquad \mathrm{var}\ \widehat{D}_\nu \le o_4 = a\eta_4/n_0, \quad \nu = 1,2,$

where a are some numerical coefficients.

Leading Parts of Functionals for Arbitrary Populations

Passing to arbitrary distributions with four moments of variables we redefine the parameters M and γ introduced in Chapter 2. Define

$$M_\nu = \sup_{|\mathbf{e}|=1} \mathbf{E}_\nu\ (\mathbf{e}^T, \mathbf{x} - \mathbf{a}_\nu)^4, \quad \nu = 1,2, \quad M = \max\,(M_1, M_2) > 0, \tag{13}$$

where the expectations $\mathbf{E}_\nu$ are calculated for $\mathbf{x} \in \mathfrak{S}_\nu$, $\nu = 1,2$, and $\mathbf{e}$ are non-random unit vectors;

$$\gamma_\nu = \operatorname*{var}_{\|\Omega\|\le 1}(\mathbf{x}^T\Omega\mathbf{x}/n)/M, \quad \mathbf{x} \in \mathfrak{S}_\nu, \quad \nu = 1,2, \qquad \gamma = \max(\gamma_1,\gamma_2), \tag{14}$$

where Ω are non-random symmetric positive semidefinite matrices of the spectral norm not greater 1.

We generalize Theorem 10.2 as follows:

THEOREM 10.3. *If populations $\mathfrak{S}_\nu$ are such that the four moments exist for each component of the observation vector* $\mathbf{x}$*, and* $n < N_\nu - 1,\ \nu = 1, 2$*, then statements of Theorems 10.1 and 10.2 are both valid with the remainder terms* $o_i,\ i = 1, \ldots, 6$*, which satisfy the inequality*

$$\max_i |o_i| \le a(Mt^2)^k (\gamma + 1/n_0)^{-1/4},$$

where $a > 0$ *and* $k > 0$ *are some numerical constants.*

Now, we perform the limit transition as $n \to \infty$, $N \to \infty$, and $n/N \to \lambda^*$. Consider a sequence $\mathfrak{P} = \{\mathfrak{P}_n\}$ of discrimination problems

$$\mathfrak{P}_n = (\mathfrak{S}_1, \mathfrak{S}_2, \mathfrak{X}_1, \mathfrak{X}_2, N_1, N_2, w(\mathbf{x}), G_1, G_2, D_1, D_2)_n \qquad (15)$$

where the running index $n = 1, 2, \ldots$ coincides with the dimension of the observations $\mathbf{x}$ in $\mathfrak{S}_\nu$, $\nu = 1, 2$, the discriminant function $w(\mathbf{x}) = w(\mathbf{x}, \eta)$ of the form (3) is calculated over samples $\mathfrak{X}_1$ and $\mathfrak{X}_2$ of size N_1 and N_2, and the function $\eta = \eta(t)$ defined above does not depend on n. We restrict $\mathfrak{P}$ by the following conditions.

A. The populations $\mathfrak{S}_\nu$ in $\mathfrak{P}$ are such that the observation vectors $\mathbf{x} \in \mathfrak{S}_\nu$ have four moments of all components of $\mathbf{x}$, $\nu = 1, 2$.

B. For each n, $M < c_0$, where c_0 does not depend on n.

C. The parameters $\gamma \to 0$ as $n \to \infty$.

D. In $\mathfrak{P}$, we have $n/N_\nu \to \lambda_\nu^* < 1$, and $N_\nu/N \to \rho_\nu^*$ as $n \to \infty$, where $0 < c_1 \le \rho_\nu^* < 1,\ \nu = 1, 2$.

Remark 1. Under assumptions A–D,

$$\lim_{n\to\infty} \mathbf{E}\,(\widehat{G}_\nu - G_\nu)^2 = 0, \qquad \lim_{n\to\infty} \mathbf{E}\,(\widehat{D}_\nu - D_\nu)^2 = 0,\ \nu = 1, 2.$$

THEOREM 10.4. *Suppose conditions A–D are satisfied and the functions* $n^{-1}\mathrm{tr}\,(I + t\Sigma)^{-1}$ *and* $(\mathbf{a}_1 - \mathbf{a}_2)^T (I + t\Sigma)^{-1}(\mathbf{a}_1 - \mathbf{a}_2)$ *converge as* $n \to \infty$, $t \ge 0$.

Then as $n \to \infty$ the limits in the square mean exist

1^o $s_\nu^* = \underset{n\to\infty}{\text{l.i.m.}}\ s_\nu(t), \quad g_\nu^*(t) = \underset{n\to\infty}{\text{l.i.m.}}\ g_\nu(t) = \underset{n\to\infty}{\text{l.i.m.}}\ \widehat{g}_\nu(t),$

2^o $d_\nu^*(t,t') = \underset{n\to\infty}{\text{l.i.m.}}\ d_\nu(t,t') = \underset{n\to\infty}{\text{l.i.m.}}\ \widehat{d}_\nu(t,t'), \quad t,t' \geq 0,$

3^o $G_\nu^* = \underset{n\to\infty}{\text{l.i.m.}}\ G_\nu = \underset{n\to\infty}{\text{l.i.m.}}\ \widehat{G}_\nu,$

4^o $D_\nu^* = \underset{n\to\infty}{\text{l.i.m.}}\ D_\nu = \underset{n\to\infty}{\text{l.i.m.}}\ \widehat{D}_\nu, \quad \nu = 1,2,$

5^o *if the populations are normal* $\mathbf{N}(\mathbf{a}_\nu, \Sigma_\nu)$, *for each* n, $\nu = 1,2$, *and there exists* $d > 0$ *such that* $\mathbf{P}(D_\nu < d) \to 0$ *as* $n \to \infty$, $\nu = 1,2$, *then in probability*

$$\underset{n\to\infty}{\text{plim}}\ \alpha_1 = \Phi(-(G_1^* - \theta)/D_1^*),$$
$$\underset{n\to\infty}{\text{plim}}\ \alpha_2 = \Phi((G_2^* - \theta)/D_2^*).$$

This theorem makes it possible to study limit characteristics of linear discriminant procedures independently on distributions.

Example 1. Let $\Sigma_1 = \Sigma_2$ for each n in $\mathfrak{P}$. Then the functions $s_1(t)$ and $s_2(t)$ differ by a quantity vanishing in $\mathfrak{P}$ as $n \to \infty$ and tend to a common limit. In this case, we have $|\mathbf{E}\,D_1 - \mathbf{E}\,D_2| \to 0$ and the estimator $\widehat{D} \stackrel{\text{def}}{=} \rho_1\widehat{D}_1 + \rho_2\widehat{D}_2$ can be expressed asymptotically in terms of the integral of $(tk(t') - t'k(t))/(t-t')$. That leads to the second and third statement of Theorem 9.1. Thus the main results of Chapter 9 follow from Theorems 10.1–10.3 as a special case.

Example 2. Let $\Sigma_2 = 0$ and Σ_1 be non-degenerate for each n. Then $D_2 = 0$ and $s_2 \to 1$. Consider a subclass of discriminant functions that are constructed with a step-wise function $\eta(t)$ at the point $t \geq 0$ so that $\Gamma = H(t)$ ('ridge'-estimator of the inverse covariance matrix). Assume, in addition, that vectors $\mathbf{a} = \mathbf{a}_1 - \mathbf{a}_2$ have components a_i in the coordinate system in which Σ_1 is diagonal such that $a_i^2/\lambda_i = J/n$, where λ_i are the corresponding eigenvalues of Σ_1, $i = 1,\ldots,n$ (the case of equal contributions to the square of the 'Mahalanobis distance'). By Theorem 10.2, we have $2G_1 = (J-y_1)(1-h)/\rho_1 s_1 + \varepsilon_1$, $2G_2 = -(J-y_2)(1-h)/\rho_1 s_1 + \varepsilon_2$, $D_1 = t\ dk/dt - k + \varepsilon_3$, $k = (J+y_1)(1-h)/\rho_1 s_1 + \varepsilon_4$, where $h = h(t)$, $k = k(t)$, and $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 \to 0$ as $n \to \infty$.

Example 3. Consider the populations from Example 2 and assume, additionally, that these populations are normal. Then $\alpha_1 =$

$\Phi(-(J-y_1)/\rho_1 s_1\sqrt{D_1})$. and $\alpha_2 = 1/2$. Here the dependence $\alpha_1 = \alpha_1(t)$ is determined by the spectral function $h = h(t)$ of the matrices Σ_1. To evaluate these functions explicitly, we choose a special case of limit spectra of Σ_1 given by the 'ρ-model' that was considered in Chapter 2: the eigenvalues of Σ_1 lie on a segment $[c_1, c_2]$ as $n \to \infty$, where $c_1 = \gamma^2(1-\sqrt{\rho})^2$, $c_2 = \sigma^2(1+\sqrt{\rho})^2$, $\sigma > 0$, $\rho < 1$, and the limit spectral density of $\Sigma = \Sigma_1$ is

$$f(u) = (2\pi)^{-1/2}(1-\rho)u^{-2}\sqrt{(c_2-u)(u-c_1)}.$$

Then as $n \to \infty$ we have $y_1 \to y_1^*$, $h \to h^*$ and

$$(1-h^*)(1-\rho h^*) = \rho_1^*\sigma^2(1-\rho)^2 h^* s_1^* t,$$

where $s_1^* = 1 - y_1^*(1-h^*)$. Let $J \to J_*$. Then we find that the minimum of $\operatorname*{plim}_{n\to\infty} \alpha_1$ is attained for $t = \rho/[\rho_1^*\sigma^2(1-\rho)^2 y_1^*]$ and is equal to $\Phi(-\sqrt{J_*^{\text{opt}}})$, where

$$J_*^{\text{opt}} = {J_*}^2(J_* + y_1^*)^{-1}(1 - \rho y_1^*/(\rho + y_1^*)).$$

In contrast to the case of equal covariance matrices considered in Chapter 9, the argument of the error function is twice as large and the parameter y_1^* is determined by the first sample only.

Discussion

Theorems 10.1–10.3 single out the leading parts of the error probabilities in the discriminant analysis using pooled covariance matrix and the discriminant functions from the class $\mathfrak{K}$. The results hold for any populations with four moments of variables having different, in general, covariance matrices under fixed sample sizes N_1 and N_2. Upper estimates of the remainder terms are obtained with the accuracy to absolute constants. For normal populations these remainder terms prove to be of the order of magnitude of $n_0^{-1/2}$, where $n_0 = \min\,(n, N_1 - 1, N_2 - 1)$. In the general case, for any populations with four moments of variables, the remainder terms are of the order of magnitude of $(\gamma + 1/n_0)^{-1/4}$. This estimate is weakened owing to the application of the Normal Evaluation Principle.

It seems to be possible to sharpen it and to obtain the stronger estimate $(\gamma+1/n_0)^{-1/2}$. The parameter γ is $O(n^{-1})$ for non-degenerate normal distributions if eigenvalues of the covariance matrices are bounded, and is $O(n^{-1})$ for any distributions of independent variables with bounded moments.

Theorem 9.2 cannot be generalized for non-normal distributions, since linear discriminant function may be not normally distributed, and therefore the expression of the error probabilities in terms of moments is not applicable. In the general case, to estimate the quality of the discrimination procedure one can use the distribution-free measure offered by R. Fisher [17]

$$F = \frac{(\mathbf{E}_1 w(\mathbf{x}) - \mathbf{E}_2 w(\mathbf{x}))^2}{\operatorname{var}_1(w(\mathbf{x})) + \operatorname{var}_2(w(\mathbf{x}))} = \frac{(G_1 - G_2)^2}{D_1 + D_2}$$

where $\mathbf{E}_1$ and $\mathbf{E}_2$ are conditional expectation operators and $\operatorname{var}_1(\cdot)$ and $\operatorname{var}_2(\cdot)$ are conditional variances for fixed samples. Substituting $w(\mathbf{x}) = \mathbf{k}^T\mathbf{x} + l$ and maximizing $F = F(\mathbf{k})$ we obtain the standard solution $\mathbf{k} = C^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$, where C is the pooled covariance matrix. For $\mathbf{k} = \Gamma(C)\ (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$, where $\Gamma(C) = \Gamma(C, \eta)$ is the matrix (4), we obtain the extremum problem $F = F(\eta) \to \max$, where now the maximum is sought within a special class of regularized discriminant function. Theorems 10.1–10.2 single out non-random leading parts of G_1, G_2, D_1, D_2 and provide estimators of these and, thereby, the estimators of $F(\eta)$. Theorems 10.3 and 10.4 make it possible to compare linear discriminant problems independently on distributions and estimate the quality of these by the Fischer criterion. Maximizing the limit value of the functional F using Theorem 10.4, one can obtain unimprovable-in-the-limit discriminant procedures.

Proofs

In order to prove Theorems 10.1–10.3, first we investigate expectation values and variances of the functions $k(t)$, $g_\nu(t)$, and $d_\nu(t, t')$ for normal populations $\mathbf{N}(\mathbf{a}_\nu, \Sigma_\nu)$, $\nu = 1, 2$. To sharpen estimates of the remainder terms we define the coefficient ω with two subscripts

$$\omega_{kl} = \omega_{k,l} = \max\ (1, \tau^k) \max\ (1, y_1^l, y_2^l)\ (\gamma + 1/N_0),$$

where $N_0 = \min(N_1, N_2) - 1$.

LEMMA 10.2. *If $t \geq 0$, then*

$$\mathbf{E}\ k(t) = t\mathbf{a}^T(I + t_1 s_1 \Sigma_1 + t_2 s_2 \Sigma_2)^{-1}\mathbf{a} + \frac{1-s_1}{\rho_1 s_1} + \frac{1-s_2}{\rho_2 s_2} + o_6,$$

where $\mathbf{a} = \mathbf{a}_1 - \mathbf{a}_2$, $s_\nu = s_\nu(t)$, $\nu = 1, 2$, $o_6^2 \leq \omega_{10,6}$;
the variance var $k(t) \leq \omega_{42}$.

Proof. Denote the centered values $\overset{o}{\mathbf{x}}_\nu = \bar{\mathbf{x}}_\nu - \mathbf{a}_\nu$, $\nu = 1, 2$. In view of the independence of H from $\bar{\mathbf{x}}$ for normal distributions, we first calculate the expectation under fixed H. It follows that

$$\mathbf{E}\ k(t) = t\mathbf{a}^T H\mathbf{a} + t_1 N_1^{-1}\text{tr}\ (\Sigma_1 H) + t_2 N_2^{-1}\text{tr}\ (\Sigma_2 H).$$

Now we apply the Helmert transformation of sample vectors and reduce matrices of the form C to more simple matrices of the form S. Namely, we define

$$S_\nu = (N_\nu - 1)^{-1} \sum_{m=1}^{N_\nu - 1} \mathbf{x}_m \mathbf{x}_m^T, \quad \nu = 1, 2,$$

for the transformed vectors $\mathbf{x}_m \sim \mathbf{N}(0, \Sigma_\nu)$, $\nu = 1, 2$, and

$$S = \frac{(N_1 - 1)S_1 + (N_2 - 1)}{N_1 + N_2 - 2}, \quad \text{and} \quad H_0 = (I + tS)^{-1}.$$

As a result we have $C = S$, $H = H_0$, and $s_\nu(t) = s_{0\nu}(t)$, $\nu = 1, 2$. A problem formulated in terms of S and H_0 will be called the *adjoint* problem.

By Lemma 10.1, we have $\mathbf{E}\ H_0 = R + \Omega$, where $\|\Omega\| \leq \omega_{42}$. Form statement 3 of Lemma 10.1 it follows that $N_\nu^{-1}\text{tr}\ (\Sigma_\nu R) = (1 - s_\nu)/\rho_\nu s_\nu$, where $|o| \leq \omega_{53}$, $\nu = 1, 2$. We obtain the first statement of our lemma.

Further, using the independence of $\bar{\mathbf{x}}$ and H, we obtain

$$\begin{aligned}\text{var}\ k(t) &= (t\mathbf{E}\ \bar{\mathbf{x}}^T(H - \mathbf{E}\ H)\bar{\mathbf{x}})^2 + \mathbf{E}\ \text{var}\ (t\bar{\mathbf{x}}^T\Omega\bar{\mathbf{x}}) \\ &\leq \mathbf{E}\ [t^2(\bar{\mathbf{x}}^2)^2\text{var}'(\mathbf{e}^T H\mathbf{e}) + t^2\text{var}\ (2\mathbf{a}^T\Omega\overset{o}{\bar{\mathbf{x}}} + \overset{o}{\bar{\mathbf{x}}}{}^T\Omega\overset{o}{\bar{\mathbf{x}}}),\end{aligned}$$

where $\mathbf{e} = \bar{\mathbf{x}}/|\bar{\mathbf{x}}|$, $\Omega = \mathbf{E}\ H$, and the conditional variance var $'(\cdot)$ is calculated under fixed $\bar{\mathbf{x}}$. The vector $\overset{o}{\mathbf{x}} = \bar{\mathbf{x}} - \mathbf{a} \sim \mathbf{N}(0, \Sigma)$, where $\Sigma = \Sigma_1/N_1 + \Sigma_2/N_2$. We calculate

$$\mathbf{E}\ (\bar{\mathbf{x}}^2)^2 = (\mathbf{a}^2)^2 + 2\mathbf{a}^2 \text{ tr } \Sigma + \text{ tr}^2\ \Sigma + 2\text{tr } \Sigma^2 \leq aM(1 + y_1^2 + y_2^2), \quad (16)$$

where a is a number. We obtain that var $k(t)$ is not greater than

$$t^2\mathbf{E}\ (\bar{\mathbf{x}}^2)^2\tau^2/(N-2) + 4t^2\mathbf{a}^T\Omega\Sigma\Omega\mathbf{a}/N_0 + 2t^2 \text{ tr } (\Sigma\Omega\Sigma\Omega)/N_0^2$$
$$\leq a\tau^4(1 + y_1^2 + y_2^2)/(N-2) + 4\tau^2(1 + y_1 + y_2)/N_0,$$

where $N_0 = \min\ (N_1, N_2) - 1$. The required estimate of var $k(t)$ follows. Thus Lemma 10.2 is proved. □

Denote $\mathbf{g}_\nu = g_\nu(t),\ \nu = 1, 2,\quad k = k(t)$.

LEMMA 10.3. *If $t \geq 0$ then*

1^o $\mathbf{E}\ g_\nu = t\mathbf{a}^T(I + t_1 s_1 \Sigma_1 + t_2 s_2 \Sigma_2)^{-1}\mathbf{a} - (1 - s_1)/\rho_1 s_1$
$+ (1 - s_2)/\rho_2 s_2 + o$, where $o^2 \leq \omega_{10,6}$;

2^o $\mathbf{E}\ g_\nu = \mathbf{E}\ k/2 - (1 - s_\nu)/\rho_\nu s_\nu + \varepsilon_\nu$, where $|\varepsilon_\nu| \leq \omega_{10,6}$;

3^o var $g_\nu \leq \omega_{42}$, $\nu = 1, 2$.

Proof. Denote $\overset{o}{\mathbf{x}}_\nu = \bar{\mathbf{x}}_\nu - \mathbf{a}_\nu$, $\nu = 1, 2$. Let $\nu = 1$. We have

$$g_1 = t(\mathbf{a}^T H\mathbf{a} - \overset{o}{\mathbf{x}}_1 H\overset{o}{\mathbf{x}}_1 + \overset{o}{\mathbf{x}}_2 H\overset{o}{\mathbf{x}}_2 - \mathbf{a}^T H\overset{o}{\mathbf{x}}_2)/2. \quad (17)$$

We calculate the expectation value using the independence of H and $\bar{\mathbf{x}}$ and pass to the adjoint problem. It follows that

$$2t\ \mathbf{E}\ g_1 = t\mathbf{E}\ [\mathbf{a}^T H_0\mathbf{a} - N_1^{-1}\text{tr } (\Sigma_1 H_0) + N_2^{-1}\text{tr } (\Sigma_2 H_0)]/2.$$

This expression differs by a sign from one of summands of the expressions for $\mathbf{E}\ k(t)$. Acting similiarly to the proof of Lemma 10.2, we obtain the statement 1. At the same time, by definition (8) we have

$$\mathbf{E}\ g_1 = \mathbf{E}\ k(t)/2 - \mathbf{E}\ t\overset{o}{\mathbf{x}}_1^T H\overset{o}{\mathbf{x}}_1.$$

Note that here $\overset{o}{\mathbf{x}}_1$ does not depend on H and the subtrahend equals $tN_1^{-1}\mathrm{tr}\ (\Sigma_1 H_0)$, (here $H_0 = H$). The statement 2 of our lemma follows.

To estimate var (g_1) we estimate the variance of summands in (17). By Lemma 10.1, the variance of the first summand is not greater than $\tau^4/(N-2)$. In the second and in the third summand, taking into account the independence of H and $\dot{\tilde{\mathbf{x}}}_\nu$ we obtain

$$\mathrm{var}\ (\overset{o}{\mathbf{x}}{}_\nu^T H\tilde{\mathbf{x}}_\nu) = \mathbf{E}\ (\overset{o}{\mathbf{x}}{}_\nu^2)^2 \mathrm{var}\ '(\mathbf{e}^T H\mathbf{e}) + \mathrm{var}\ (\overset{o}{\mathbf{x}}{}_\nu^T(\mathbf{E}\ H)\overset{o}{\mathbf{x}}_\nu),$$

where the variance var'$(\cdot)$ in the right hand side is calculated under fixed $\overset{o}{\mathbf{x}}_\nu$, and $\mathbf{e}$ is a unit vector in direction of $\overset{o}{\mathbf{x}}_\nu$, $\nu = 1, 2$. Here $\mathbf{E}\ (\overset{o}{\mathbf{x}}{}_\nu^2)^2 \leq 3My_\nu^2$, $\nu = 1, 2$, var $(\mathbf{e}^T H\mathbf{e}) \leq \tau^4/(N-2)$. Thus the first summand equals

$$2N_\nu^{-2}\mathrm{tr}\ (\mathbf{E}\ H\Sigma_\nu \mathbf{E}\ H\Sigma_\nu) \leq 2My_\nu/N_\nu, \quad \nu = 1, 2.$$

Therefore the second and the third term in (17) provide a contribution to var (g_1) not greater ω_{42}. The fourth term in (17) contributes an amount not larger than

$$t^2\mathbf{a}^T\mathbf{E}\ H_0\Sigma_2\mathbf{E}\ H_0\mathbf{a}/N_2 \leq \omega_{20}.$$

The sum of all these contributions to var (g_1) is not greater $a(1+\tau^4)(1+y_1^2+y_2^2)/\min\ (N_1-1, N_2-1)$, where a is a number. It follows that var $(g_1) \leq \omega_{42}$. The symmetric statement for $\nu = 2$ follows from the symmetry of assumptions. Lemma 10.3 is proved. □

Now we are able to construct estimators of functions $s_\nu(t)$, $g_\nu(t)$, and $d_\nu(t, t')$ with small bias and small variance (of the order of magnitude of the remainder terms).

Passing to the adjoint problem we define

$$\psi_\nu = \mathbf{x}_\nu^T H_0 \mathbf{x}_\nu/(N-2), \quad \nu = 1, 2,$$

where the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are from the samples $\mathfrak{X}_1$ and $\mathfrak{X}_2$ of the adjoint problem.

Remark 2.

$$t\psi_\nu \leq 1, \quad \mathbf{E}\ t\psi_\nu = 1 - s_\nu, \quad \mathrm{var}\ \ (t\psi_\nu) \leq \delta,$$

where $\delta = \max_{\nu=1,2} \text{var } (t\psi_\nu) \le \omega_{42}$ (this follows from Lemma 2.3).

Remark 3. The statistics $\widehat{s}_\nu = \widehat{s}_\nu(t) \ge 1 - y_\nu$ are non-biased estimators of $s_\nu = s_\nu(t)$ with the variance var $\widehat{s}_\nu \le y^2\omega_{40}$, $\nu = 1, 2$.

Let us prove this statement. The unbiasedness of $\widehat{s}_\nu$ and the inequality $1 - \widehat{s}_\nu < y_\nu$ follow from the definition,$\nu = 1, 2$. We have

$$\begin{aligned} \text{var } \widehat{s}_\nu &= t^2 \text{var } [\text{tr } (H_0 S_\nu)/(N-2)] \\ &= t^2(\mathbf{E}\ [\psi_\nu \text{ tr } (H_0 S_\nu)/(N-2)] - (\mathbf{E}\ \psi_\nu)^2 \\ &\le (\text{var } [\text{tr } (tH_0 S_\nu)/(N-2)])^{1/2} [\text{var } (t\psi_\nu)]^{1/2}]. \end{aligned}$$

Here var $(t\psi_{\nu\nu}) \le \delta$, and consequently, var $(\widehat{s}_\nu)$ also is not greater than this quantity. Remark 3 is justified. □

LEMMA 10.4. *For $y_\nu < 1$, $\nu = 1, 2$, the statistics $\widehat{g}_\nu = \widehat{g}_\nu(t)$ can serve as approximations to $g_\nu = g_\nu(t)$ such that*

$$(1 - y_\nu)^2 |\mathbf{E}\ \widehat{g}_\nu - \mathbf{E}\ g_\nu|^2 \le \omega_{10,6},$$

$$\text{var } (g_\nu) \le \omega_{42}, \qquad (1 - y_\nu)^2 \text{ var } (\widehat{g}_\nu) \le \omega_{42}, \quad \nu = 1, 2. \tag{18}$$

Proof. Let $\nu = 1, 2$. We compare $\mathbf{E}\ \widehat{g}_\nu$ with the definition of $k = k(t)$. We have $\widehat{s}_\nu \ge 1 - y_\nu$, $\nu = 1, 2$. It follows that

$$2(\mathbf{E}\ \widehat{g}_\nu - \mathbf{E}\ g_\nu)^2 \le \rho_\nu^{-2}(1 - y_\nu)^{-2} \text{var } (\widehat{s}_\nu) + \omega_{10,6}. \tag{19}$$

Here, by Remark 3 the variance of $\widehat{s}_\nu$ is not greater than $y^2\omega_{40}$. Since $y/\rho_\nu = y_\nu$, the right hand side of (19) is not greater than $(1-y_\nu)^{-2}\omega_{42}$, $\nu = 1, 2$. We obtain the first statement of our lemma.

Since var $\widehat{g}_\nu \le$ var k + var $(1/\widehat{s}_\nu)$, where var $k \le \omega_{42}$, we obtain

$$\text{var } (1/\widehat{s}_\nu) \le (1 - y_\nu)^{-2} \rho_\nu^{-2} \text{var } \widehat{s}_\nu \le (1 - y_\nu)^{-2} \omega_{42}.$$

The last statement of our lemma follows. □

Now we consider the random values $d_\nu(t, t')$, $\nu = 1, 2$. Our purpose is to find approximating statistics for these values. Let us apply the method of alternative elimination of variables to expressions quadratic in the resolvent. We pass to the adjoint problem and enumerate sample vectors in such a way that the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are from the samples $\mathfrak{X}_1$ and $\mathfrak{X}_2$ of size $N_1 - 1$, $N_2 - 1$, respectively. Denote

$$S^\nu = S - N^{-1}\mathbf{x}_\nu\mathbf{x}_\nu^T, \quad H_0^\nu = H_0^\nu(t) = (I + tS^\nu)^{-1},$$
$$v_\nu = v_\nu(t) = \mathbf{e}^T H_0^\nu \mathbf{x}_\nu, \quad u_\nu = u_\nu(t) = \mathbf{e}^T H_0 \mathbf{x}_\nu.$$

By these definitions,

$$H_0 = H_0^\nu - tH_0^\nu\mathbf{x}_\nu\mathbf{x}_\nu^T H_0/(N-2), \quad H_0 = \mathbf{x}_\nu = (1 - t\psi_\nu)H_0^\nu\mathbf{x}_\nu,$$
$$u_\nu = (1 - t\psi_\nu)v_\nu, \quad \nu = 1, 2. \tag{20}$$

Note that the random value

$$\xi = [1 - t\psi_{\nu\nu}(t)][1 - t'\psi_{\nu\nu}(t')] = s_\nu(t)s_\nu(t') + \varepsilon_\nu,$$

where $\mathbf{E}\ \varepsilon_\nu^2 \leq 4\delta$.

LEMMA 10.5. *For $t \geq t' \geq 0$, we have*

$$\sqrt{tt'}|\mathbf{E}\ \mathbf{e}^T H_0(t)\Sigma_\nu H_0(t')\mathbf{e} - \mathbf{E}\ \mathbf{e}^T H_0^\nu(t)\Sigma_\nu H_0^\nu(t')\mathbf{e}| \leq \omega_{20},$$

and

$$\sqrt{tt'}|\mathbf{E}\ \mathbf{e}^T H_0(t)S_\nu H_0(t')\mathbf{e} - s_\nu(t)s_\nu(t')\mathbf{E}\ \mathbf{e}^T H_0(t)\Sigma_\nu H_0(t')\mathbf{e}|$$
$$\leq \sqrt{\omega_{62}}, \quad \nu = 1, 2,$$

where a is a numerical coefficient.

Proof. Let $\nu = 1$. We find that $\mathbf{e}^T H_0(t)\Sigma_1 H_0(t')\mathbf{e}$ is equal to

$$\mathbf{e}^T H_0^1(t)\Sigma_1 H_0^1(t')\mathbf{e} - t\mathbf{e}^T H_0^1(t)\Sigma_1 H_0^1(t')\mathbf{x}_1(t')u_1(t')/(N-2)$$
$$- tu_1(t)\mathbf{x}_1^T H_0^1(t)\Sigma_1 H_0^1(t')\mathbf{e}/(N-2) +$$
$$+ tt'\ v_1(t)v_1(t')\mathbf{x}_1^T H_0(t)\Sigma_1 H_0(t')\mathbf{x}_1^2/(N-2)^2. \tag{21}$$

We apply the Schwarz inequality. The second term in the right hand side of (21) has the form $(\mathbf{f}^T\mathbf{x}_1)u_1/(N-2)$, where the vector $\mathbf{f}$ does not depend on $\mathbf{x}_1$ and, consequently, $\mathbf{E}\ (\mathbf{f}^T\mathbf{x}_1)^2 \leq \sqrt{M}\mathbf{f}^2$.

Here $\mathbf{f}^2 \leq \|\Sigma_1\|^2 t^2 \leq \tau^2$ and $t\mathbf{E}\ u_1^2 \leq \sqrt{M}t \leq \tau$. We conclude that the contribution of this term to (21) is not greater $\tau^2/(N-2)$. The contribution of the third term can be estimated similiarly. The last term in (21) is the product of the form $v_1(t)v_1(t')r(t,t')$, where $\mathbf{E}\ v_1^4 \leq M$ and $r(t,t')$ is some function such that $r^2(t,t') \leq r(t,t)r(t',t')$ and

$$\mathbf{E}\ r^2(t,t) \leq t^4 \mathbf{E}\ \mathbf{x}_1^T H_0 \Sigma_1 H_0 S_1 H_0 \Sigma_1 H_0 \mathbf{x}_1/(N-2)^4.$$

Since $t_1\|H_0 S_1 H_0\| \leq 1$ it follows that the right hand side is not larger than $Mt^3\mathbf{E}\ \mathbf{x}_1^T H_0 \mathbf{x}_1/(N-2)^3 \leq Mt^3 \mathbf{E}\ \psi_{11}/(N-2)^2 \leq \tau^2/(N-2)^2$. We conclude that the contribution of the fourth term to (21) is not greater $\tau^2/(N-2)$. The first statement of our lemma is proved.

Further, we find

$$\begin{aligned}\mathbf{E}\ \mathbf{e}^T H_0(t) S_1 H_0(t')\mathbf{e} &= \mathbf{E}\ \mathbf{e}^T H_0(t)\mathbf{x}_1\mathbf{x}_1^T H_0(t')\mathbf{e} = \mathbf{E}\ u_1(t)u_1(t') \\ &= \mathbf{E}\ [1 - t\psi_1(t)][1 - t'\psi_1(t)]\ v_1(t)v_1(t') \\ &= s_1(t)s_1(t')\mathbf{E}\ \mathbf{e}^T H_0^1(t)\Sigma_1 H_0^1(t')\ v_1(t)v_1(t') + o,\end{aligned}$$

where $o^2 \leq 4\tau^2\delta\ \mathbf{E}\ v_1^2(t)v_1^2(t') \leq 4M\tau^2\delta$. We obtain the second statement of our lemma for $\nu = 1$. The proof for $\nu = 2$ follows from assumptions. Lemma 10.5 is proved. □

LEMMA 10.6. *If* $t \geq t' \geq 0$, *then*

$$\begin{aligned}&1^o\ tt'\ \text{var}\ (\mathbf{e}^T H_0(t)\Sigma_\nu H_0(t')\mathbf{e}) \leq \omega_{40}, \\ &2^o\ tt'\ \text{var}\ (\mathbf{e}^T H_0(t) S_\nu H_0(t')\mathbf{e}) \leq \omega_{20},\end{aligned}$$

$\nu = 1, 2$, *where a is a numerical coefficient.*

Proof. We use the martingale Lemma 2.2. Let $\nu = 1$. Let us single out alternatively vectors from samples $\mathfrak{X}_1$ and $\mathfrak{X}_2$. In view of the identical dependence on the sample vectors, we find that

$$\text{var}\ (\mathbf{e}^T H_0(t)\Sigma_1 H_0(t')\mathbf{e}) \leq N_1\delta_1 + N_2\delta_2,$$

where δ_ν are expectation values of the sum of three last terms in (21), $\nu = 1, 2$. Using the Schwarz inequality, from the definition of M and τ we find that $\delta_\nu \leq a\tau^4/N^2,\ \nu = 1, 2$. We obtain the first lemma statement for $\nu = 1$. The symmetric statement for $\nu = 2$ follows from assumptions.

To estimate the second variance we again use Lemma 2.2. Let us single out alternatively vectors from $\mathfrak{X}_1$ and then from $\mathfrak{X}_2$. Denote $S_\nu^\nu = S_\nu - N^{-1}\mathbf{x}_\nu \mathbf{x}_\nu^T$, $\nu = 1, 2$. We have

$$\begin{aligned}
\mathbf{e}^T H_0(t) S_\nu H_0(t')\mathbf{e} = \\
&= \mathbf{e}^T H_0^\nu(t) S_\nu^\nu H_0^\nu(t')\mathbf{e} - u_\nu(t) u_\nu(t')/(N-2) \\
&- t'\mathbf{e}^T H_0^\nu(t) S_\nu^\nu H_0^\nu(t')\mathbf{x}_\nu u_\nu(t')/(N-2) \\
&- t u_\nu(t)\mathbf{x}_\nu^T H_0^\nu(t) S_\nu^\nu H_0^\nu(t')\mathbf{e}/(N-2) \\
&+ tt'\ v_\nu(t) v_\nu(t')\mathbf{x}_\nu^T H_0(t) S_\nu^\nu H_0(t')\mathbf{x}_\nu/(N-2)^2, \quad \nu = 1, 2.
\end{aligned} \tag{22}$$

We replace here $tH_0^\nu(t)S_\nu^\nu = I - H_0^\nu(t)$ and majorize $\|H_0^\nu\| \le 1$ for the arguments t and $t = t'$. By Lemma 2.2, we have

$$\operatorname{var}(\mathbf{e}^T H_0(t) S_\nu H_0(t')\mathbf{e}) \le (N_1 - 1)\delta_1 + (N_2 - 1)\delta_2,$$

where the quantities δ_1 and δ_2 are squares of the sum of last four terms and the sum of last three terms, respectively, with S_1^1 replaced by S_2 and the vector $\mathbf{x}_1$ eliminated only in S_1. Using the relation $tH_0^\nu(t)S_\nu^\nu = I - H_0^\nu$, $\nu = 1, 2$, we obtain the second statement. Lemma 10.6 is proved. □

LEMMA 10.7. *For* $0 \le t' \le t$

$$s_\nu(t) s_\nu(t')\mathbf{E}\ d_\nu(t, t') = \mathbf{E}\ \widehat{d}_\nu(t, t') + o,$$

where $o^2 \le \omega_{84}$.

Proof. Let $\nu = 1, 2$. We use the independence of H and $\bar{\mathbf{x}}$, and, at first, fix $\bar{\mathbf{x}}$. Passing to the adjoint problem we obtain

$$\mathbf{E}\ \widehat{d}_\nu(t, t') = tt'\ \mathbf{E}\ \bar{\mathbf{x}}^2 \mathbf{E}\ \mathbf{e}^T H_0(t) S_\nu H_0(t')\mathbf{e}, \quad \nu = 1, 2,$$

where $\mathbf{e} = \bar{\mathbf{x}}/|\bar{\mathbf{x}}|$. In view of Lemma 10.5, it follows that $\mathbf{E}\ \widehat{d}_\nu(t, t')$ equals

$$(tt')^{1/2}\mathbf{E}\ \bar{\mathbf{x}}^2[(tt')^{1/2} s_\nu(t) s_\nu(t')\mathbf{E}'\mathbf{e}^T H_0(t)\Sigma_\nu H_0(t')\mathbf{e} + \varepsilon_\nu],$$

where $\mathbf{E}'$ is conditional expectation operator under chosen $\bar{\mathbf{x}}$, and $\varepsilon_\nu^2 \le \omega_{62}$. We can see that the leading part of this expression is $tt's(t)s(t')\mathbf{E}\ d_\nu(t, t')$. Using the inequality $\mathbf{E}\ \bar{\mathbf{x}}^2 \le \sqrt{M}(1 + y_1 + y_2)$, we obtain that the remainder terms are not larger ω_{84}. This proves the lemma. □

Lemma 10.8. *If* $0 \le t' \le t$ *and* $y_\nu < 1$, *then*

$$(1-y_\nu)^2|\mathbf{E}\ d_\nu(t,t') - \mathbf{E}\ \widehat{d}_\nu(t,t')/\widehat{s}_\nu(t)\widehat{s}_\nu(t')| \le o,$$

$\nu = 1,2$, *where* $o^2 \le \omega_{12,8}$.

Proof. Let $\nu = 1,2$. Denote for brevity $d = d_\nu(t,t')$, $\widehat{d} = \widehat{d}_\nu(t,t')$, $s = s_\nu(t)$, $s' = s_\nu(t')$, $\widehat{s} = \widehat{s}_\nu(t)$, $\widehat{s}' = \widehat{s}_\nu(t')$. We find that the absolute value of the difference in the lemma formulation is not larger than the absolute value of the expectation value of

$$|d - \widehat{d}/ss'| + |\widehat{d}(1/ss' - 1/\widehat{s}\widehat{s}')|.$$

But we have $s, s' \ge (1+\tau y)^{-1}$ and $\widehat{s}, \widehat{s}' \ge (1-y_\nu)$. The first summand is not greater $\omega_{53}(1+\tau y)^2 \le \sqrt{\omega_{12,8}}$ and the second one is not greater than

$$(1+\tau y)^2(1-y_\nu)^{-2}[\mathbf{E}\ \widehat{d}^2(2\ \text{var}\ (\widehat{s}) + 2\ \text{var}\ (\widehat{s}'))]^{1/2}.$$

Estimating with respect to the norm, we have $\|t_\nu H_0 S_\nu\| \le 1$, $\nu = 1,2$. But var $\widehat{s} \le \omega_{40}y^2$. It follows that var $\widehat{s}$ $\mathbf{E}\ \widehat{d}^2 \le \omega_{64}$ and the contribution of the second summand is not greater than $\sqrt{\omega_{10,8}}$. The required statement follows. □

Lemma 10.9. *If* $0 \le t' \le t$, *then*

$$\text{var}\ d_\nu(t,t') \le \omega_{64} \quad \text{and} \quad \text{var}\ \widehat{d}_\nu(t,t') \le \omega_{42}, \quad \nu = 1,2.$$

Proof. Using the independence of H and $\bar{\mathbf{x}}$, we obtain

$$\begin{aligned}\text{var}\ &[tt'\ \bar{\mathbf{x}}^T H(t)\Sigma_\nu H(t')\bar{\mathbf{x}}] \\ &= t^2\mathbf{E}\ [(\bar{\mathbf{x}}^2)^2 tt'\ \text{var}\ (\mathbf{e}^T H(t)\Sigma_\nu H(t')\mathbf{e})] + \text{var}\ (t^2\bar{\mathbf{x}}^T\Omega\bar{\mathbf{x}}),\end{aligned}$$

$\nu = 1,2$, where the first variance in the right hand side is conditional under chosen $\bar{\mathbf{x}}$, $\mathbf{e} = \bar{\mathbf{x}}/|\bar{\mathbf{x}}|$, and $\Omega = \mathbf{E}\ H(t)\Sigma_\nu H(t')$. In the first summand of the right hand side, we use the inequality $\mathbf{E}\ (t\bar{\mathbf{x}}^2)^2 \le a\tau(1+y_1^2+y_2^2)$, where a is a number; the first summand proves to be not greater ω_{21}. In the second summand, $\bar{\mathbf{x}} \sim \mathbf{N}(a,\bar{\Sigma})$, where

$$\bar{\Sigma} = \Sigma/(N_1-1) + \Sigma_2/(N_2-1).$$

and the second summand is not greater in absolute value than

$$t^4 \ [\mathbf{a}^T \Omega \Sigma \Omega \mathbf{a} + 2 \ \mathrm{tr} \ (\Sigma \Omega \Sigma \Omega)] \leq \omega_{41}.$$

Thus the second summand is not larger ω_{41}. This proves the first statement of our lemma.

Further, we use the independence of H and $\bar{\mathbf{x}}$ and split the expression for $\widehat{d_\nu}(t, t')$ to two summands. We pass to the adjoint problem and obtain that var $\widehat{d_\nu}(t, t')$ equals

$$t^2 t'^2 \left(\mathrm{var} \ (\bar{\mathbf{x}}^T \Omega \bar{\mathbf{x}}) + \mathbf{E} \ [\bar{\mathbf{x}}^2 \ \mathrm{var} \ (\mathbf{e}^T H_0(t) S_\nu H_0(t') \mathbf{e})] \right),$$

where $\mathbf{e}$ is a unit vector in the direction of $\bar{\mathbf{x}}$. Here $tt' \|\Omega\|^2 \leq 1$, and $\bar{\mathbf{x}} \sim \mathbf{N}(\mathbf{a}, \Sigma)$, The contribution of the first term is not greater ω_{42}. The second summand also is not greater ω_{42}. It follows that

$$\mathrm{var} \ [tt' \ \bar{\mathbf{x}}^T H(t) C_\nu H(t') \bar{\mathbf{x}}] \leq \omega_{42}.$$

The last statement of our lemma is proved. □

Now we are able to prove the basic theorems of this chapter.

Proof of Theorem 10.1. We pass to the adjoint problem with N_ν less by unit, $\nu = 1, 2$. Then $H = H_0$, $C_\nu = S_\nu$, $\nu = 1, 2$. The first and the second statements of Theorem 10.1 follow from statement 2 of Lemma 10.1. The third and the fourth statements follows from Lemma 10.3, and the fifth one was proved in Lemma 10.4. Statement 6 can be derived from Lemma 10.8. This proves Theorem 10.1.

Theorem 10.2 is a corollary of Theorem 10.1 and Lemmas 10.4 and 10.9.

Theorems 10.3 and 10.4 immediately follow from Theorems 10.1 and Theorem 10.2.