CHAPTER IV

NORMAL EVALUATION OF QUALITY FUNCTIONS

In this Chapter we prove that for high dimension of variables, most of standard rotation invariant functionals measuring the quality of regularized multivariate procedures can be approximately but reliably evaluated under a hypothesis of population normality. This means that the quality of these procedures proves to be approximately distribution free. Our purpose is to investigate the precise facts concerning this phenomena. We (1) study some classes of functionals of the quality function type for regularized versions of mostly used linear multivariate procedures, (2) single out the leading terms and show that these depend on only two moments of variables, and (3) obtain upper estimates of correction terms accurate up to absolute constants.

Let $\mathbf{x}$ be an observation vector in n-dimensional population $\mathfrak{S}$ (let us write $\mathbf{x} \sim \mathfrak{S}$), and $\mathfrak{K}_4 = \mathfrak{K}_4(M)$ be a class of populations with $\mathbf{E}\mathbf{x} = 0$ and the maximum fourth moment

$$M = \sup \mathbf{E}\ (\mathbf{e}^T \mathbf{x})^4 > 0, \tag{1}$$

where the supremum is calculated over all $\mathbf{e}$, and $\mathbf{e}$ (here and in the following) are non-random vectors of unit length. Denote $\Sigma = \operatorname{cov}(\mathbf{x}, \mathbf{x})$. We consider a sample $\mathfrak{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ from $\mathfrak{S}$ and the statistics

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{m=1}^{N} \mathbf{x}_m, \qquad S = \frac{1}{N} \sum_{m=1}^{N} \mathbf{x}_m \mathbf{x}_m^T,$$
$$C = \frac{1}{N} \sum_{m=1}^{N} (\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T.$$

Here S and C are sample covariance matrices (for known and for unknown expectation vectors).

Measure of Normalizability

We say that function $f : \mathbf{R}^n \to \mathbf{R}^1$ of a random vector $\mathbf{x}$ allows ε-normal evaluation or, shortly, is ε-normalizable (in the square mean) in a class $\mathfrak{K}$ of n-dimensional distributions $\mathfrak{S}$, if for each $\mathfrak{S} \in \mathfrak{K}$, we can choose a normal distribution $\mathbf{y} \sim \mathbf{N}(\mathbf{a}, \Sigma)$ having the same parameters $\mathbf{a} = \mathbf{E}\mathbf{x}$ and $\Sigma = \text{cov}\ (\mathbf{x}, \mathbf{x})$ in $\mathfrak{S}$ such that $\mathbf{E}\ (f(\mathbf{x}) - f(\mathbf{y}))^2 \leq \varepsilon$.

Example 1. Let $n = 1$, $\xi \sim \mathbf{N}(0, 1)$, $x = \xi^3/\sqrt{15}$. Denote the distribution law of x by $\mathfrak{S}$. Then the function $f(x) = x$ allows ε-normal evaluation (by normal $y = \xi$) in $\mathfrak{S}$ with $\varepsilon = 0.45$.

The following statements based on properties of martingale differences show that the effect of normalization may be owed to the uniform dependence on a set of independent variables.

REMARK 1. *Let $f(\mathfrak{X})$ be a function of a set $\mathfrak{X} = \{\mathbf{x}_m\}$ of independent vectors $\mathbf{x}_m$, $m = 1, \ldots, N$. Suppose that functions*

$$f_m(\mathfrak{X}, \theta) = f(\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}, \theta\mathbf{x}_m, \mathbf{x}_{m+1}, \ldots, \mathbf{x}_N)$$

are continuous differentiable with respect to $\theta \in [0, 1]$, $m = 1, \ldots, N$, and

$$\max_m \sup_\theta \mathbf{E} \left| \frac{df_m(\mathfrak{X}, \theta)}{d\theta} \right|^2 \leq c/N^2. \qquad (2)$$

Then var $f(\mathfrak{X}) \leq c/N$.

Example 2. Let $\mathfrak{S} \in \mathfrak{K}_4$ be an n-dimensional population and let $f(t)$ be a continuous differentiable function of $t \geq 0$ that has a derivative not greater b_1 in absolute value. Then $f(\bar{\mathbf{x}}^2)$ satisfies (2) with $c = b_1^2 M(2 + n/N)$ and the function $f(\bar{\mathbf{x}}^2)$ allows ε-normal evaluation in $\mathfrak{S}$ with $\varepsilon = c/N$.

In Chapter 2 the assumptions were formulated which provide the convergence for entries of resolvents of sample covariance matrices S and C as $n \to \infty$, $N \to \infty$, and $n/N \to \lambda > 0$. The limit values of these resolvents proved to be depending only on limit values of spectral functions of the matrices Σ. By the definition above such resolvents (i.e., the entries of these) allow ε-normal evaluation with $\varepsilon \to 0$ as $n \to \infty$.

Spectral Functions of Large Sample Covariance Matrices

We define the resolvent type matrices $H_0 = H_0(t) = (I+A+tS)^{-1}$ and $H = H(t) = (I + A + tC)^{-1}$, where (and in the following) I denotes the identity matrix and A is a positively semidefinite symmetric matrix of constants. Consider spectral functions

$$h_0(t) = n^{-1}\text{tr } H_0(t), \qquad h(t) = n^{-1}\text{tr } H(t),$$

$$s_0(t) = 1 - y + yh_0(t), \qquad s(t) = 1 - y + yh(t), \quad \text{where } y = n/N.$$

In order to obtain a finite measure of the normalization we use recent results presented in Chapter 2, where the leading parts of spectral functions were singled out for finite n and N. The upper estimates of the remainder terms will be obtained as functions of n, N, M, $t \geq 0$, and of a special measures of quadratic form variance

$$\gamma = \sup_{\|\Omega\| \leq 1} \text{var } (\mathbf{x}^T \Omega \mathbf{x}/n)/M, \tag{3}$$

where the supremum is calculated over positive semidefinite symmetric matrices Ω of constants with spectral norm not greater 1. Note that the parameter γ can serve as a measure of the dependence of variables. Indeed, $\gamma = n^{-2}\text{tr } \Sigma^2/\|\Sigma\|^2$ for $\mathbf{x} \sim \mathbf{N}(0, \Sigma)$ and $\gamma = O(n^{-1})$ for independent components of $\mathbf{x}$ with bounded fourth momenta.

Our resolvents H_0(t) and H(t) differ from those investigated in Chapter 2 by the addition of non-zero matrices A. The generalization can be performed by a formal reasoning as follows. Note that the linear transformation $\mathbf{x}' = B\mathbf{x}$, where $B^2 = (I + A)^{-1}$ takes H_0 to the form $H_0' = B(I + tS')^{-1}B$, where

$$S' = N^{-1} \sum_{m=1}^{N} \mathbf{x}'\mathbf{x}'^T, \qquad \mathbf{x}'_m = B\ \mathbf{x}_m, \quad m = 1, \ldots, N.$$

It can be readily seen that $M' = \sup \mathbf{E}\ (\mathbf{e}^T\mathbf{x}')^4 \leq M$, and the products $M'\gamma' \leq M\gamma$, where γ' is the measure (3) calculated for $\mathbf{x}'$. Let us apply results of Chapter 2 to vectors $\mathbf{x}' = B\mathbf{x}$. The matrix elements of our H_0 can be reduced to matrix elements of the resolvents of S' by the linear transformation B with $\|B\| \leq 1$. The remainder terms for $\mathbf{x}' = B\mathbf{x}$ are not greater than those for $\mathbf{x}$. The same reasoning also holds for the matrices H. A survey of upper estimates obtained in Chapter 2 shows that all of these remain valid for our H_0 and H.

Let us formulate a corollary which will be a starting point of our development below in the form of a lemma. To be more concise in estimates, we denote

$$c_{kl} = c_{k,l} = c_{kl}(t) = a \max(1, \tau^k) \max(1, y^l), \tag{4}$$

where a is a positive numerical constant and $k, l = 0, 1, \ldots, 10$, $\tau = \sqrt{M}t$, and $y = n/N$.

Lemma 4.1. (corollary of theorems in Chapter 2).
If $t \geq 0$ *and* $\mathfrak{S} \in \mathfrak{K}_4(M)$ *then:*

1^o $\mathbf{E}\ H_0(t) = (I + A + ts_0(t)\Sigma) + \Omega_0,$

where $\|\Omega_0\| \leq o_1 = c_{31}\sqrt{\gamma + 1/N}$;

2^o var $(\mathbf{e}^T H_0 \mathbf{e}) \leq o_2 = c_{20}/N$;

3^o $t\ \mathbf{E}\ \bar{\mathbf{x}}^T H_0(t)\bar{\mathbf{x}} = 1 - s_0(t) + o_3,$ $\quad o_3^2 \leq c_{52}(\gamma + 1/N)$;

4^o var $(t\bar{\mathbf{x}}^T H_0(t)\bar{\mathbf{x}}) \leq o_4 = c_{20}/N$;

5^o $\mathbf{E}\ H(t) = (I + A + ts_0(t)\Sigma)^{-1} + \Omega,$

where $\|\Omega\|^2 \leq o_5 = c_{63}(\gamma + 1/N)$;

6^o var $(\mathbf{e}^T H(t)\mathbf{e}) \leq o_6 = c_{20}/N$;

7^o $ts_0(t)\ \mathbf{E}\ \bar{\mathbf{x}}^T H(t)\bar{\mathbf{x}} = 1 - s_0(t) + o_7,$ $\quad |o_7| \leq c_{42}\sqrt{\gamma + 1/N}$;

8^o var $(t\bar{\mathbf{x}}^T H(t)\bar{\mathbf{x}}) \leq o_8 = c_{64}/N.$

Normal Evaluation of Sample Dependent Functionals

We study rotation invariant functionals including well-known quality functions that depend on expectation value vectors, sample means, and population and sample covariance matrices. For sake of generality, let us consider a set of k n-dimensional populations $\mathfrak{S}_1, \ldots, \mathfrak{S}_k$ with expectation vectors $\mathbf{E}\ \mathbf{x} = \mathbf{a}_i$, and covariance matrices cov $(\mathbf{x}, \mathbf{x}) = \Sigma_i$ for $\mathbf{x}$ in $\mathfrak{S}_i$, moments M_i of the form (1) and the parameters γ_i of the form (3), $i = 1, \dots, k$. Let $\mathfrak{X}_i$ be independent samples from $\mathfrak{S}_i$ of size N_i; denote $y_i = n/N_i$,

$$\bar{\mathbf{x}}_i = N_i^{-1} \sum_m \mathbf{x}_m, \qquad S_i = N_i^{-1} \sum_m (\mathbf{x}_m - \mathbf{a}_i)(\mathbf{x}_m - \mathbf{a}_i)^T,$$

$$C_i = N_i^{-1} \sum_m (\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T,$$

where m runs over all numbers of vectors in $\mathfrak{X}_i$, and $i = 1, \dots, k$.

We introduce more general resolvent type matrices

$$H_0 = (I + t_0 A + t_1 S_1 + \dots + t_k S_k)^{-1},$$
$$H = (I + t_0 A + t_1 C_1 + \dots + t_k C_k)^{-1}$$

where $t_0, t_1, \dots, t_k \geq 0$ and A are symmetric positively semidefinite matrices of constants. We consider the following classes of functionals depending on $A, \bar{\mathbf{x}}_i, S_i, C_i, t_i,\ i = 1, \dots, k$.

The class $\mathfrak{L}_1 = \{\Phi_1\}$ of functionals $(k = 1)$ $\Phi_1 = \Phi_1(t_0, t_1)$ of the form

$$n^{-1}\mathrm{tr}\ H_0, \quad \mathbf{e}^T H_0 \mathbf{e}, \quad t_1 \bar{\mathbf{x}}_1^T H_0 \bar{\mathbf{x}}_1, \quad n^{-1}\mathrm{tr}\ H, \quad \mathbf{e}^T H \mathbf{e}, \quad t_1 \bar{\mathbf{x}}_1 H \bar{\mathbf{x}}_1.$$

Note that the matrices $tH = C_\alpha \stackrel{\text{def}}{=} (C + \alpha I)^{-1}$, where $\alpha = 1/t$, can be considered as regularized estimators of the inverse covariance matrix Σ^{-1}.

The class $\mathfrak{L}_2 = \{\Phi_2\}$ of functionals $\Phi_2 = \Phi_2(t_0, t_1, \dots, t_k)$ of the form

$$n^{-1}\mathrm{tr}\ H_0, \quad \mathbf{e}^T H_0 \mathbf{e}, \quad t_i \bar{\mathbf{x}}_i H_0 \bar{\mathbf{x}}_i,$$
$$n^{-1}\mathrm{tr}\ H, \quad \mathbf{e}^T H \mathbf{e}, \quad t_i \bar{\mathbf{x}}_i H \bar{\mathbf{x}}_i, \quad i = 1, \dots, k.$$

The class $\mathfrak{L}_3 = \{\Phi_3\}$ of functionals of the form $\Phi_3 = D_m \Phi_2$ and $\partial/\partial t_0\ D_m \Phi_2$, where $\Phi_3 = \Phi_3(t_0, t_1, \dots, t_k)$, and D_m is the partial differential operator of the mth order

$$D_m = \frac{\partial^m}{\partial z_{i_1} \dots \partial z_{i_m}},$$

where $z_j = \ln t_j$, $t_j \geq 0$, $j = 0, 1, \ldots, k$, and $i_1, i_2, \ldots, i_m$ are numbers from $\{0, 1, \ldots, k\}$;

Note that by differentiation of resolvents one can obtain functionals with the matrices A, S, and C in the numerator. This class includes a variety of functionals which are used as quality functions of multivariate procedures, for example:

$$\bar{\mathbf{x}}_i^T A \bar{\mathbf{x}}_i, \quad t_i \bar{\mathbf{x}}_i^T H \bar{\mathbf{x}}, \quad n^{-1} \mathrm{tr}\,(AH), \quad t_i n^{-1} \mathrm{tr}\,(HC_i H), \quad \mathbf{e}^T HAH \mathbf{e},$$
$$t_i \bar{\mathbf{x}}_i HC_i H \bar{\mathbf{x}}_i, \quad t_0 \mathbf{e}^T HAHAH \mathbf{e}, \quad \text{etc.}$$

The class $\mathfrak{L}_4 = \{\Phi_4\}$ of functionals of the form

$$\Phi_4 = \Phi_4(\eta_0, \eta_1, \ldots, \eta_k) = \iint \Phi_3(t_0, t_1, \ldots, t_k)\; d\eta_0(t_0)\; d\eta_1(t_1) \ldots d\eta_k(t_k),$$

where $\eta_i(t)$ are functions of $t \geq 0$ with the variation not greater 1 on $[0, \infty)$, $i = 0, 1, \ldots, k$, having a sufficient number of moments

$$\beta_j = \int t^j |d\eta_i(t)|, \quad i = 1, \ldots, k, \quad j = 0, 1, \ldots,$$

where the functions Φ_3 are extended by continuity to zero values of arguments. This class presents a number of functionals constructed using arbitrary linear combinations and linear transformations of regularized estimators of the inverse covariance matrices with different ridge parameters, for example, such as sums of $\alpha_i (I + t_i C)^{-1}$, functions $n^{-1} \mathrm{tr}\,(I + tC)^{-k}$, $k = 1, 2, \ldots,$ $\exp\,(-tC)$, etc.. Such functionals will be used in Chapters 5–10 to construct regularized approximately unimprovable statistical procedures.

The class $\mathfrak{L}_5 = \{\Phi_5\}$ of functionals $\Phi_5 = \Phi_5(z_1, \ldots, z_p)$, where $z_1, \ldots, z_p$ are functionals from $\mathfrak{L}_4$, and the Φ_5 are continuously differentiable with respect to all arguments with partial derivatives bounded in absolute value by a constant $a_5 \geq 0$.

Obviously, $\mathfrak{L}_1 \in \mathfrak{L}_2 \in \mathfrak{L}_3 \in \mathfrak{L}_4 \in \mathfrak{L}_5$.

To be more concise, we redefine the quantities N, y, τ, and γ as follows:

$$N = \min_i N_i, \quad y = n/N, \quad \gamma = \max_i \nu_i / M_i, \; \tau = \max_i \sqrt{M_i} t_i, \quad (5)$$

$i = 1, \ldots, k$. Definition (4) will be used with these new parameters.

Denote $\omega = \sqrt{\gamma + 1/N}$.

PROPOSITION 1. *Functionals* $\Phi_1 \in \mathfrak{L}_1$ *allow* ε*-normal evaluation in the class of populations* $\mathfrak{K}_4 = \mathfrak{K}_4(M)$ *with*
$\varepsilon = \varepsilon_1 \stackrel{\text{def}}{=} c_{10,6}\ (\gamma + 1/N)$.

Proof. Let $\tilde{\mathfrak{S}}$ denote a normal population $\mathbf{N}(0, \Sigma)$ with a matrix $\Sigma = \Sigma_1 = \operatorname{cov}(\mathbf{x}, \mathbf{x})$ that is identical in $\mathfrak{S}$ and $\tilde{\mathfrak{S}}$. We set
$N = N_1,\ \ y = y_1 = n/N,\ \ t_0 = 1,\ \ t_1 = t$.

Let $\mathbf{E}$ and $\tilde{\mathbf{E}}$ denote the expectation operators for $\mathbf{x} \sim \mathfrak{S}$ and $\tilde{\mathfrak{S}}$, respectively, and, by definition , let

$$h_0(t) = \mathbf{E}\ n^{-1}\mathrm{tr}\ \ H_0, \quad s_0(t) = 1 - y + \mathbf{E}\ N^{-1}\mathrm{tr}\ (I + A)H_0,$$
$$\tilde{h}_0(t) = \tilde{\mathbf{E}}\ n^{-1}\mathrm{tr}\ H_0, \quad \tilde{s}_0(t) = 1 - y + \tilde{\mathbf{E}}\ N^{-1}\mathrm{tr}\ \ (I + A)H_0,$$
$$G_0 = (I + A + ts_0(t)\Sigma)^{-1}, \quad \tilde{G}_0 = (I + A + t\tilde{s}_0(t)\Sigma)^{-1}.$$

Statement 1 of Lemma 4.1 implies that

$$|h_0(t) - \tilde{h}_0(t)|\ (1 + tN^{-1}\mathrm{tr}\ G_0\Sigma\tilde{G}_0) \le o_1,$$

where o_1 is defined by Lemma 4.1. The trace in the parenthesis is non-negative. From statement 2 of Lemma 4.1, the inequality follows var $(n^{-1}\mathrm{tr}\ H_0) \le o_2$ both in $\mathfrak{S}_1$ and in $\tilde{\mathfrak{S}}_1$. We conclude that n^{-1} tr H_0 allows ε-normal evaluation with $\varepsilon \le 4\sigma_1^2 + 2\sigma_2 \le c_{62}\omega^2$. Also, from statement 1 it follows that

$$\|\mathbf{E}\ H_0 - \tilde{\mathbf{E}}\ H_0\| \le t\ |s_0(t) - \tilde{s}_0(t)|\ \|\mathbf{G}_0\Sigma\tilde{\mathbf{G}}_0\| + 2\sigma_1,$$

where $|s_0(t) - \tilde{s}_0(t)| \le 2o_1 y$ and $\|G_0\Sigma\tilde{G}_0\| \le \sqrt{M}$. Thus the norm in the left hand side is not greater than $2(1 + \tau y)o_1$. From Lemma 4.1 it follows that var $(\mathbf{e}^T H_0 \mathbf{e}) \le o_2$ both for $\mathfrak{S}_1$ and for $\tilde{\mathfrak{S}}_1$. We conclude that $\mathbf{e}^T H_0 \mathbf{e}$ allows ε-normal evaluation with $\varepsilon = c_{84}\omega^2$.

Further, by the third statement of Lemma 4.1

$$|t\mathbf{E}\ \bar{\mathbf{x}}^T H_0 \bar{\mathbf{x}}| - |t\tilde{\mathbf{E}}\ \bar{\mathbf{x}}^T H_0 \bar{\mathbf{x}}| \le |s_0(t) - \tilde{s}_0(t)| + 2|o_3|,$$

where the summands in the right hand side are not greater than $c_{32}\omega$ and $c_{31}\omega$, respectively. In view of statement (4) from Lemma 4.1 we have var $(t\bar{\mathbf{x}}^T H_0\bar{\mathbf{x}}) \le c_{20}/N$ both for $\mathfrak{S}_1$ and for $\tilde{\mathfrak{S}}_1$. We conclude that $t\bar{\mathbf{x}}^T H_0\bar{\mathbf{x}}$ allows ε-normal evaluation with $\varepsilon = c_{64}\omega^2$.

Now we define:

$$h(t) = n^{-1}\mathrm{tr}\ H, \quad s(t) = 1 - y + \mathbf{E}\ N^{-1}\mathrm{tr}\ \ (I + A)H,$$
$$\tilde{h}(t) = \tilde{\mathbf{E}}\ n^{-1}\mathrm{tr}\ \ H, \quad \tilde{s}(t) = 1 - y + \tilde{\mathbf{E}}\ N^{-1}\mathrm{tr}\ \ H,$$
$$G = (I + A + ts(t)\Sigma)^{-1}, \quad \tilde{G} = (I + A + t\tilde{s}(t)\Sigma)^{-1}.$$

Statement 5 of Lemma 4.1 implies that

$$|h(t) - \tilde{h}(t)| \le tN^{-1}\mathrm{tr}\ \ (G\Sigma\tilde{G})\ |s_0(t) - \tilde{s}_0(t)| + \|\Omega\| \le c_{32}\omega.$$

By statement 6 of Lemma 4.1 we have var $(n^{-1}\mathrm{tr}\ H) \le o_2/N$ both in $\mathfrak{S}_1$ and in $\tilde{\mathfrak{S}}_1$. We conclude that $n^{-1}\mathrm{tr}\ H$ allows ε-normal evaluation with $\varepsilon = c_{64}\omega$.

From statement 5 of Lemma 4.1 it follows that

$$\|\mathbf{E}\ H - \tilde{\mathbf{E}}\ H\| \le t\ |s_0(t) - \tilde{s}_0(t)|\ \|G\Sigma\tilde{G}\| + 2\|\Omega\|\ \le c_{32}\omega.$$

By statement 6 of Lemma 4.1, we have var $(\mathbf{e}^T H\mathbf{e}) \le c_{53}/N$ both in $\mathfrak{S}_1$ and in $\tilde{\mathfrak{S}}_1$. It follows that $\mathbf{e}^T H\mathbf{e}$ allows ε-normal evaluation with $\varepsilon = c_{64}\omega^2$.

Further, using statements 1 and 7 of Lemma 4.1 we obtain that min $(s_0(t), \tilde{s}_0(t)) \ge (1 + \tau y)$, and

$$|t\mathbf{E}\ \bar{\mathbf{x}}^T H\bar{\mathbf{x}}| - |t\tilde{\mathbf{E}}\ \bar{\mathbf{x}}^T H\bar{\mathbf{x}}| \le |1/s_0(t) - 1/\tilde{s}_0(t)| + o,$$

where $|o| \le c_{53}\omega$. The first summand in the right hand side is not greater $c_{54}\omega$. From Lemma 4.1 it follows that var $(t\bar{\mathbf{x}}^T H\bar{\mathbf{x}}) \le c_{64}/N$. We conclude that $t\bar{\mathbf{x}}^T H\bar{\mathbf{x}}$ allows ε-normal evaluation with $\varepsilon = c_{10,6}\ \omega$. This ends the proof of Proposition 1. □

Proposition 2. *Functionals $\Phi_2 \in \mathfrak{L}_2$ allow ε-normal evaluation in $\mathfrak{K}_4(M)$ with $\varepsilon = \varepsilon_2 = k^2\varepsilon_1$.*

Proof. We consider normal populations $\tilde{\mathfrak{S}}_i = \mathbf{N}(\mathbf{a}_i, \Sigma_i)$ with $\mathbf{a}_i = \mathbf{E}\ \mathbf{x}$ and $\Sigma_i =$ cov $(\mathbf{x}, \mathbf{x})$, for $\mathbf{x}$ in $\mathfrak{S}_i$, $i = 1, \dots, k$. Let $\mathbf{E}_i$ be the expectation operator for the random vectors

$$\mathbf{x}_1 \sim \mathfrak{S}_1, \dots, \ \ \mathbf{x}_i \sim \mathfrak{S}_i, \ \ \mathbf{x}_{i+1} \sim \tilde{\mathfrak{S}}_{i+1}, \dots, \ \ \mathbf{x}_k \sim \tilde{\mathfrak{S}}_k,$$

where the tilde denotes the probability distribution in the corresponding population, $i = 1, \dots, k-1$. Let $\mathbf{E}_0$ denote the expectation when all populations are normal, $\mathbf{x}_i \sim \tilde{\mathfrak{S}}_i$, $i = 1, \dots, k$, and let $\mathbf{E}_k$ be the expectation operator for $\mathbf{x}_i \sim \mathfrak{S}_i$, $i = 1, \dots, k$. Clearly, for each random f having the required expectations,

$$\mathbf{E}_0 f - \mathbf{E} f = \sum_{i=1}^{k} (\mathbf{E}_{i-1} f - \mathbf{E}_i f).$$

Let us estimate the square of this sum as a sum of k^2 terms. We set $f = \Phi_2$ for the first three forms of the functionals Φ_2 (depending on H_0). Choose some $i : 1 \le i \le k$. In view of the independence of $\mathbf{x}_i$ chosen from different populations, each summand can be estimated by Proposition 1 with $H_0 = H_{0i} = (I + B^i + t_i S_i)^{-1}$, where

$$B^i = I + t_0 A + \sum_{j=1}^{i-1} t_j S_j + \sum_{j=i+1}^{k} t_j S_j$$

is considered to be non-random for this i, $i = 1, \dots, k$. By Proposition 1 each summand allows ε-normal evaluation with $\varepsilon = \varepsilon_1$. We conclude that $(\mathbf{E}\ f_0 - \mathbf{E}\ f)^2 \le k^2 \varepsilon_1$. Similiar arguments hold for f depending on H. This completes the proof of Proposition 2. □

PROPOSITION 3. *Functionals $\Phi_3 \in \mathfrak{L}_3$ allow ε-normal evaluation in $\mathfrak{K}_4(M)$ with $\varepsilon_3 = a_{mk}(1 + \|A\|)^2 (1 + \tau y) \varepsilon_2^{1/2(m+1)}$, where a_{mk} are numerical constants and τ and y are defined by* (5).

Proof. Let $\mathfrak{X}$ denote a collection of samples from populations $(\mathfrak{S}_1, \dots, \mathfrak{S}_k)$ and $\tilde{\mathfrak{X}}$ a collection of samples from normal populations $(\tilde{\mathfrak{S}}_1, \dots, \tilde{\mathfrak{S}}_k)$ with the same first two moments, respectively. Let us compare $\Phi_3(\mathfrak{X}) = D_m \Phi_2(\mathfrak{X})$ and $\tilde{\Phi}_3(\mathfrak{X}) = D_m \tilde{\Phi}_2(\tilde{\mathfrak{X}})$, where Φ_2 and $\tilde{\Phi}_2$ are functionals from $\mathfrak{L}_2$. Note that

$$\frac{\partial H_0}{\partial z_i} = -H_0 t_i S_i H_0 \quad \text{and} \quad \frac{\partial H}{\partial z_i} = -H t_i C_i H,$$

where $z_i = \ln t_i$, $t_i > 0$, $i = 1, \dots, k$. The differential operator D_m transforms H_0 into sums (and differences) of the matrices $T_r = H_0 t_i S_i H_0 \dots t_j S_j H_0$, $i, j = 0, 1, \dots, k$ with different numbers r of the multiples H_0, $1 \le r \le m+1$, $T_1 = H_0$. Note that $\|T_r\| \le 1$, as is easy to see from the fact that the inequalities $\|H_0\Omega_1\| \le 1$ and $\|H_0\| \le 1$ hold for $H_0 = (I + \Omega_1 + \Omega_2)^{-1}$ and for any symmetric positively semidefinite matrices Ω_1 and Ω_2. Now, $\partial/\partial z_i\ T_r$ is a sum of r summands of the form of T_{r+1} plus $r-1$ terms of the form T_r, no more that $2r-1$ summands in total. We can conclude that each derivative $\partial/\partial z_i\ D_m\Phi_2$ is a sum of at most $(2m+1)!!$ terms each of these being bounded by 1 or by $t_j\tilde{\mathbf{x}}_j^2$ for some $j = 1, \dots, k$, depending on Φ_2. But $\mathbf{E}\ (t_j\tilde{\mathbf{x}}^2)^2 \le \tau^2 y^2$, where $j = 1, \dots, k$, and $y = n/N$. It follows that

$$\mathbf{E}\ \left|\frac{\partial}{\partial z_i}\ D_m\Phi\right|^2 \le (1+\tau^2y^2)[(2m+3)!!]^2$$

for any $i = 1, \dots, k$. We introduce a displacement $\delta > 0$ of $z_i = \ln\, t_i$ being the same for all $i = 1, \dots, k$, and replace the derivatives by finite differences. Let Δ_m be a finite difference operator corresponding to D_m . We obtain

$$\Delta_m\Phi_2 = D_m\Phi_2 + \delta\sum_{i=1}^{k}\frac{\partial}{\partial z_i}\ D_m\Phi_2\,|_{z=\xi_i}$$

where ξ_i are some intermediate values of z_i, $i = 1, \dots, k$. By Proposition 2, the function $\Delta_m\Phi_2$ allows ε-normal evaluation with $\varepsilon = \varepsilon' = 2\varepsilon_2 2^{m+1}/\delta^m$. The quadratic difference

$$\mathbf{E}\ (\Delta_m\Phi_2 - D_m\Phi_2)^2 \le \varepsilon'' = \delta^2k^2(1+\tau^2y^2)[(2m+3)!!]^2.$$

We conclude that $D_m\Phi_2$ allows ε-normal evaluation with $\varepsilon = \sigma \stackrel{\text{def}}{=} \varepsilon' + \varepsilon''$. Choosing $\delta = 2\varepsilon_2^{1/(2+m)}(1+\tau^2y^2)^{1/(2+m)}$, we obtain that $\sigma < a\varepsilon_2^{1/(1+m)}$, where the numerical coefficient a depends on m and k. We proved Proposition 3 for the functionals $D_m\Phi_2$.

Now consider functionals of the form $\partial/\partial t_0\ D_m\Phi_2$. Let us replace the derivative by a finite difference with the displacement δ of the argument. An additional differentiation with respect to t_0 and the transition to finite differences transforms each term T_r into $2r$ summands, where $r \le m+1$. Each summand can be increased by a factor of no more than $(1 + \|A\|)$. Choose $\delta = \sqrt{\sigma}$. Reasoning similiarly, we conclude that $\partial/\partial t_0\ D_m\Phi_2$ allows ε-normal evaluation with $\varepsilon = 2(m+1)^2(1+\|A\|)^2\sqrt{\sigma}$. This ends the proof of Proposition 3. □

The next two statements follow immediately.

PROPOSITION 4. *The functionals* $\Phi_4 \in \mathfrak{L}_4$ *allow* ε*-normal evaluation in* $\mathfrak{K}_4(M)$ *with* $\varepsilon = \varepsilon_4 = \varepsilon_3$.

PROPOSITION 5. *The functionals* $\Phi_5 \in \mathfrak{L}_5$ *allow* ε*-normal evaluation in* $\mathfrak{K}_4(M)$ *with* $\varepsilon = \varepsilon_5 \le p^2 a_5^2 \varepsilon_4$.

Discussion

Thus, on the basis of the theory of spectral properties of large sample covariance matrices which was developed in Chapter 2 it is possible to suggest a method of estimating of a number of sample-dependent functionals, including most popular quality functions for regularized procedures independently on distributions. If the moment (1) is bounded and the measure of the quadrics variance γ is small, then commonly used quality functions prove to be insensitive to details of distributions and depend mainly on the first two moments of variables. The remainder terms are of the order of magnitude of $\sqrt{\gamma + 1/N}$, where N is the sample size and γ is small for a large number of restrictively dependent variables. Thus the normality assumption can be used for an approximate evaluation of quality functions of linear regularized multivariate procedures and, consequently, for a comparison of these procedures and their better choice. It immediately follows that the traditional theory of linear multivariate analysis, developed conventionally for normal populations, has a wider range of applicability, at least for regularized versions of procedures. A number of improved multivariate procedures developed earlier for normal populations can be justified and considered as a natural step of an asymptotic development with the guaranteed accuracy.

It also follows that for essentially multivariate problems a 'Normal Evaluation Principle' can be offered: to prove theorems first for normal distributions and then to estimate corrections resulting from non-normality by formulas obtained in the above.