NONPARAMETRIC EMPIRICAL BAYES ESTIMATION OF THE MATRIX PARAMETER OF WISHART DISTRIBUTION

by

Marianna Pensky
Purdue University

Technical Report # 95-37

August 1995

*

Nonparametric empirical Bayes estimation of the matrix parameter of Wishart distribution *

Marianna Pensky
Department of Statistics
Purdue University

Abstract

Independent pairs $(\mathbf{X}_1, \Sigma_1), (\mathbf{X}_2, \Sigma_2), \ldots, (\mathbf{X}_n, \Sigma_n)$ are considered, where each Σ_i is distributed according to some unknown density function $g(\Sigma)$ and, given $\Sigma_i = \Sigma$, $\mathbf{X}_i$ has conditional density function $q(\mathbf{x}/\Sigma)$ of the Wishart type. In each pair the first component is observable but the second is not. After the $(n+1)$-th observation $\mathbf{X}_{n+1}$ is obtained, the objective is to estimate Σ_{n+1} corresponding to $\mathbf{X}_{n+1}$. This estimator is called the empirical Bayes (EB) estimator of Σ.

An EB estimator of Σ is constructed without any parametric assumptions on $g(\Sigma)$. Its posterior mean square risk is examined, and the estimator is demonstrated to be pointwise asymptotically optimal.

1 Introduction

Symmetric positive definite $k \times k$ matrix $\mathbf{X}$ is said to have a Wishart distribution if its density function has the form

$$q(\mathbf{x}/\Sigma) = C_{k,r}[\det(\Sigma)]^{-\frac{r}{2}}[det(\mathbf{x})]^{\frac{r-k-1}{2}} \exp\left\{-0.5\,\mathrm{tr}(\mathbf{x}\Sigma^{-1})\right\}, \quad r > k, \tag{1.1}$$

where $C_{k,r} = \left[2^{\frac{rk}{2}}\pi^{\frac{k(k-1)}{4}} \prod_{j=1}^{k} \Gamma\left(\frac{r+j-1}{2}\right)\right]^{-1}$, $r > k$, and Σ is symmetric positive definite $k \times k$ matrix. Note that here and in what follows matrices are denoted by bold characters, $(\mathbf{z})_{i,j}$ or $z_{i,j}$ is the (i,j)-th element of a matrix $\mathbf{z}$, $\mathrm{tr}(\mathbf{z}) = \sum_{j=1}^{k}(\mathbf{z})_{j,j}$, $\tilde{\mathbf{z}}$ is the transpose of $\mathbf{z}$.

*The research was carried out during a visit by M. Pensky to the Department of Statistics, Purdue University. The author would like to thank Prof. James Berger and Prof. Shanti Gupta for their help, hospitality and support.

The Wishart distribution is fundamental to multivariate statistical analysis. Many authors were occupied with its investigation. One of the basic problems, the problem of estimation of $\mathbf{\Sigma}$ given observations $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ on $\mathbf{X} \sim q(\mathbf{x}/\mathbf{\Sigma})$, was solved long ago: unbiased estimator of $\mathbf{\Sigma}$ has the form $\hat{\mathbf{\Sigma}}_n = (nr)^{-1} \sum_{j=1}^{n} \mathbf{X}_j$ (see Anderson (1984)). However, sometimes the situation occurs when not only $\mathbf{X}$ but also $\mathbf{\Sigma}$ is random, and we need to make some conclusions about the value of the random parameter. Namely, the following problem arises.

Independent pairs $(\mathbf{X}_1, \mathbf{\Sigma}_1), (\mathbf{X}_2, \mathbf{\Sigma}_2), ..., (\mathbf{X}_n, \mathbf{\Sigma}_n)$ are given where each $\mathbf{\Sigma}_i$ is distributed according to some unknown density function $g(\mathbf{\Sigma})$ and, given $\mathbf{\Sigma}_i = \mathbf{\Sigma}$, $\mathbf{X}_i$ has conditional density function $q(\mathbf{x}/\mathbf{\Sigma})$, $i = 1, ..., n$. In each pair the first component is observable but the second is not. After the $(n+1)$-th observation $\mathbf{X}_{n+1} \equiv \mathbf{Y}$ is obtained, the objective is to estimate $\mathbf{\Sigma}_{n+1} \equiv \mathbf{S}$ corresponding to $\mathbf{Y}$. The problem of estimation of $\mathbf{S}$ is called the **empirical Bayes (EB)** estimation problem.

EB model was very popular recently and EB estimators were constructed for a number of families of conditional distributions. However, the vast majority of papers dealt with the situation when observations are univariate. EB estimation in the multivariate case was conducted by Ghosh, M. (1992), Ghosh M., Shieh G. (1991), (1992), Judge, G.G., Hill, R.S. & Bock, M.E. (1990), Kubokawa, T. & Robert, C. & Caleh, A.K.MD.E. (1992) and Shieh, G. (1993), who obtained various EB estimators for the mean and covariance matrix of normal distribution.

In the present situation both observations and parameters are matrices. An EB estimator of $\mathbf{\Sigma}$ is constructed without any parametric assumptions on $g(\mathbf{\Sigma})$, only some constraints on the moments of $\mathbf{\Sigma}$ are imposed. The estimator is produced using a general technique proposed by the author (see Penskaya (1992), Penskaya(1993)).

Let us denote

$$p(\mathbf{x}) = \int_{\mathcal{A}} q(\mathbf{x}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) d\mathbf{\Sigma}, \tag{1.2}$$

$$\mathbf{F}(\mathbf{x}) = \int_{\mathcal{A}} \mathbf{\Sigma}\, q(\mathbf{x}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) d\mathbf{\Sigma}. \tag{1.3}$$

Here and in what follows $\mathcal{A}$ is a space of symmetric positive definite $k \times k$ matrices, $d\mathbf{\Sigma} = \prod_{j=1}^{k} \prod_{i=1}^{j} d\Sigma_{i,j}$. If we knew prior density $g(\mathbf{\Sigma})$, then under the square loss, the Bayes estimator $\mathbf{S}(\mathbf{Y})$ of $\mathbf{\Sigma}$ would take the form

$$\mathbf{S}(\mathbf{Y}) = \frac{\mathbf{F}(\mathbf{Y})}{p(\mathbf{Y})}. \tag{1.4}$$

An EB estimator $\mathbf{S}_n(\mathbf{Y})$ of $\mathbf{\Sigma}$ is an estimator of $\mathbf{S}(\mathbf{Y})$ from observations $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$. Note that $\mathbf{S}(\mathbf{Y})$ and $\mathbf{F}(\mathbf{Y})$ are matrix-valued functions, whereas $p(\mathbf{Y})$ is a scalar-valued function of the matrix $\mathbf{Y}$.

Now let us introduce the risk function. Consider the (i,j)-th element $(\mathbf{S}_n(\mathbf{Y}))_{i,j}$ of an EB estimator $\mathbf{S}_n(\mathbf{Y})$. Its posterior risk is given by

$$(p(\mathbf{Y}))^{-1}\ \mathbf{E}_{p^n} \int_{\mathcal{A}} \left[(\mathbf{S}_n(\mathbf{Y}))_{i,j} - (\mathbf{S}(\mathbf{Y}))_{i,j}\right]^2 q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) d\mathbf{\Sigma} \tag{1.5}$$

where $\mathbf{E}_{p^n}$ is the mathematical expectation over all possible values of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. It is easy to see that (1.5) can be broaken into two components. The first component

$$(p(\mathbf{Y}))^{-1}\ \mathbf{E}_{p^n} \int_{\mathcal{A}} \left[(\mathbf{S}(\mathbf{Y}))_{i,j} - (\mathbf{\Sigma})_{i,j}\right]^2 q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) d\mathbf{\Sigma}$$

is the posterior risk of the (i,j)-th element $(\mathbf{S}(\mathbf{Y}))_{i,j}$ of the Bayes estimator (1.4), and it is independent of $\mathbf{S}_n(\mathbf{Y})$. . Thus the quality of EB estimator $(\mathbf{S}_n(\mathbf{Y}))_{i,j}$ is described by the second component

$$\Delta_n^{(i,j)}(\mathbf{Y}) = \mathbf{E}_{p^n} \left[(\mathbf{S}_n(\mathbf{Y}))_{i,j} - (\mathbf{S}(\mathbf{Y}))_{i,j}\right]^2.$$

The overall risk is the sum of $\Delta_n^{(i,j)}(\mathbf{Y})$, and therefore we will characterize our EB estimator by its risk

$$\Delta_n(\mathbf{Y}) = \mathbf{E}_{p^n} \left[\mathrm{tr}\left(\tilde{\mathbf{S}}_n(\mathbf{Y}) - \tilde{\mathbf{S}}(\mathbf{Y})\right)(\mathbf{S}_n(\mathbf{Y}) - \mathbf{S}(\mathbf{Y}))\right]. \tag{1.6}$$

For $\Delta_n(\mathbf{Y})$ to converge to zero some constraints on $g(\mathbf{\Sigma})$ must be imposed. We will assume that

$$\int_{\mathcal{A}} [\mathrm{tr}(\mathbf{\Sigma})]^2 g(\mathbf{\Sigma}) d\mathbf{\Sigma} < \infty, \quad \int_{\mathcal{A}} \det(\mathbf{\Sigma}) g(\mathbf{\Sigma}) d\mathbf{\Sigma} < \infty. \tag{1.7}$$

Estimator $\mathbf{S}_n(\mathbf{Y})$ is called *pointwise asymptotically optimal*, if $\Delta_n(\mathbf{Y}) \to 0$ as ns for every $\mathbf{Y}$.

Note that the quality of EB estimator can also be characterized by a prior risk. However, it seems reasonable to use a posterior risk as a measure of quality of an EB estimator. First, using $\Delta_n(\mathbf{Y})$, we calculate mean square error of our estimate at a point $\mathbf{Y}$, which is actually the point of interest. Furthermore, if we apply the risk function (1.6), then the observations having very low probabilities, i.e., the observations that we would never get, stop influence the risk function. Thus we are able to search for more effective methods of estimation at our point of interest $\mathbf{Y}$.

In Section 2 of the present paper we construct estimators $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$ of the numerator $\mathbf{F}(\mathbf{Y})$ and the denominator $p(\mathbf{Y})$ of the Bayes estimator (1.4). We also investigate mean square errors of the estimators $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$. In Section 3 we present an EB estimator of $\mathbf{\Sigma}$ and obtain an upper bound for its risk (1.6), in the form

$$\Delta_n(\mathbf{Y}) = O\left(n^{-\frac{\mu m}{(\mu+1)(m+K)}}\right), \tag{1.8}$$

where $K = 0.5\,k(k+1)$, $m = 2m_1$, μ and m_1 are arbitrary positive integers. Computational aspects of the estimation are discussed in Section 4. Section 5 contains the proofs of the assessments of Section 2 and Section 3.

2 Estimation of $\mathbf{F}(\mathbf{Y})$ and $p(\mathbf{Y})$.

To construct an EB estimator of $\mathbf{\Sigma}$, which is an estimator of (1.4) from observations, we will obtain estimators $p_n(\mathbf{Y})$ and $\mathbf{F}_n(\mathbf{Y})$ of $p(\mathbf{Y})$ and $\mathbf{F}(\mathbf{Y})$, respectively, and then estimate the ratio $\mathbf{F}(\mathbf{Y})/p(\mathbf{Y})$.

It is easy to guess that the usual technique of substitution of estimator $g_n(\mathbf{\Sigma})$ for $g(\mathbf{\Sigma})$ turns out to be extremely complicated here. Actually, we have to estimate density function in $[0.5\,k(k+1)]$-dimensional space from indirect observations. This problem results in a huge system of linear equations which is ill-posed. And after we estimated $g(\mathbf{\Sigma})$, we would need to calculate the integral in a space of positive definite symmetric matrices.

So we need to apply another method of estimation of $\mathbf{F}(\mathbf{Y})$ and $p(\mathbf{Y})$. It was proposed by the author (see Penskaya(1992), (1993)). To derive $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$, we should search for approximate solutions $\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})$ and $\varphi_h(\mathbf{x};\mathbf{Y})$ of equations

$$\int_{\mathcal{A}} q(\mathbf{x}/\mathbf{\Sigma})\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})d\mathbf{x} \simeq \mathbf{\Sigma}\, q(\mathbf{Y}/\mathbf{\Sigma}), \tag{2.1}$$

$$\int_{\mathcal{A}} q(\mathbf{x}/\mathbf{\Sigma})\varphi_h(\mathbf{x};\mathbf{Y})d\mathbf{x} \simeq q(\mathbf{Y}/\mathbf{\Sigma}). \tag{2.2}$$

If we found functions $\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})$ and $\varphi_h(\mathbf{x};\mathbf{Y})$ such that for every $\mathbf{Y}$

$$D_\varepsilon(\mathbf{Y}) = \int_{\mathcal{A}} \mathrm{tr}\left[\tilde{\mathbf{\Phi}}_\varepsilon(\mathbf{x};\mathbf{Y})\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})\right] p(\mathbf{x})d\mathbf{x} < \infty, \tag{2.3}$$

$$d_h(\mathbf{Y}) = \int_{\mathcal{A}} \left[\varphi_h(\mathbf{x};\mathbf{Y})\right]^2 p(\mathbf{x})d\mathbf{x} < \infty, \tag{2.4}$$

and, moreover,

$$\mathbf{B}_\varepsilon(\mathbf{Y}) = \int_{\mathcal{A}} \left[\int_{\mathcal{A}} q(\mathbf{x}/\mathbf{\Sigma})\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})d\mathbf{x} - \mathbf{\Sigma}\, q(\mathbf{Y}/\mathbf{\Sigma})\right] g(\mathbf{\Sigma})d\mathbf{\Sigma} \to 0,\ (\varepsilon \to 0), \tag{2.5}$$

$$b_h(\mathbf{Y}) = \int_{\mathcal{A}} \left[\int_{\mathcal{A}} q(\mathbf{x}/\mathbf{\Sigma})\varphi_h(\mathbf{x};\mathbf{Y})d\mathbf{x} - q(\mathbf{Y}/\mathbf{\Sigma})\right] g(\mathbf{\Sigma})d\mathbf{\Sigma} \to 0,\ (h \to 0), \tag{2.6}$$

then our proposed estimators $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$ would be

$$\mathbf{F}_n(\mathbf{Y}) = n^{-1}\sum_{j=1}^{n} \mathbf{\Phi}_\varepsilon\left(\mathbf{X}_j,\mathbf{Y}\right), \quad \varepsilon = \varepsilon(n); \tag{2.7}$$

$$p_n(\mathbf{Y}) = n^{-1}\sum_{j=1}^{n}\varphi_h(\mathbf{X}_j,\mathbf{Y}), \quad h = h(n). \tag{2.8}$$

Mean square risks of the estimators $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$ are of the form

$$\varrho_n(\mathbf{Y}) = \mathbf{E}_{p^n}\left[\mathrm{tr}(\widetilde{\mathbf{F}}_n(\mathbf{Y}) - \widetilde{\mathbf{F}}(\mathbf{Y}))(\mathbf{F}_n(\mathbf{Y}) - \mathbf{F}(\mathbf{Y}))\right], \tag{2.9}$$

$$\sigma_n(\mathbf{Y}) = \mathbf{E}_{p^n}(p_n(\mathbf{Y}) - p(\mathbf{Y}))^2. \tag{2.10}$$

It is easy to check (see Penskaya(1993)) that under conditions (2.3) and (2.4) mean square errors of estimators (2.7) and (2.8) are bounded by

$$\varrho_n(\mathbf{Y}) \leq \mathrm{tr}\left[\widetilde{\mathbf{B}}_\varepsilon(\mathbf{Y})\mathbf{B}_\varepsilon(\mathbf{Y})\right] + n^{-1}D_\varepsilon(\mathbf{Y}), \quad \varepsilon = \varepsilon(n), \tag{2.11}$$

$$\sigma_n(\mathbf{Y}) \leq b_h^2(\mathbf{Y}) + n^{-1}d_h(\mathbf{Y}), \quad h = h(n). \tag{2.12}$$

Thus, provided (2.5), (2.6) are valid, there exist dependences $h = h(n)$ and $\varepsilon = \varepsilon(n)$ such that the errors (2.11) and (2.12) tend to zero as $n \to \infty$.

Now our goal is to find $\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})$ and $\varphi_h(\mathbf{x};\mathbf{Y})$ satisfying (2.1) – (2.6). After a quick look at equations (2.1) and (2.2) and at the density function (1.1), we come to the conclusion that $\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})$ and $\varphi_h(\mathbf{x};\mathbf{Y})$ have the following forms

$$\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y}) = [\det(\mathbf{Y})]^{\frac{r-k-1}{2}}[\det(\mathbf{x})]^{-\frac{r-k-1}{2}}\mathbf{\Psi}_\varepsilon(\mathbf{x}-\mathbf{Y})\,I(\mathbf{x}-\mathbf{Y}\in\mathcal{A}), \tag{2.13}$$

$$\varphi_h(\mathbf{x};\mathbf{Y}) = [\det(\mathbf{Y})]^{\frac{r-k-1}{2}}[\det(\mathbf{x})]^{-\frac{r-k-1}{2}}\psi_h(\mathbf{x}-\mathbf{Y})\,I(\mathbf{x}-\mathbf{Y}\in\mathcal{A}). \tag{2.14}$$

Then equations (2.1) and (2.2) appear as

$$\int \exp\left\{-0.5\ \mathrm{tr}\left[(\mathbf{x}-\mathbf{Y})\mathbf{\Sigma}^{-1}\right]\right\}\mathbf{\Psi}_\varepsilon(\mathbf{x}-\mathbf{Y})d\mathbf{x} \simeq \mathbf{\Sigma}, \tag{2.15}$$

$$\int \exp\left\{-0.5\ \mathrm{tr}\left[(\mathbf{x}-\mathbf{Y})\mathbf{\Sigma}^{-1}\right]\right\}\psi_h(\mathbf{x}-\mathbf{Y})d\mathbf{x} \simeq 1. \tag{2.16}$$

The last two integrals are evaluated over all values of $\mathbf{x}$ such that $\mathbf{x}\in\mathcal{A}$ and $\mathbf{x}-\mathbf{Y}\in\mathcal{A}$. Note that $\mathbf{Y}$ is positive definite matrix, therefore the intersection of the two sets coincides with the set $(\mathbf{x} : \mathbf{x}-\mathbf{Y}\in\mathcal{A})$.

We say that matrix $\mathbf{z} = \sqrt{\mathbf{u}}$ is a **square root** of symmetric positive definite matrix $\mathbf{u}$ if $\mathbf{z}$ is a matrix of upper triangular type with positive diagonal elements such that $\tilde{\mathbf{z}}\mathbf{z} = \mathbf{u}$. It can be shown that there always exists an unique square root of a symmetric positive definite matrix.

Let us introduce new variable $\mathbf{z} = \sqrt{\mathbf{x} - \mathbf{Y}}$. Note that $\mathbf{z}$ here is a positive definite matrix of an upper triangular type and range of $\mathbf{z}$ is

$$\mathcal{B} = \{\mathbf{z} : \ z_{i,j} \in R, \ z_{j,j} \geq 0, \ j = 1, ..., k, \ i \leq j\}. \tag{2.17}$$

Jacobian of transformation is

$$J = 2^k \prod_{j=1}^{k} \mathbf{z}_{j,j}^{k-j+1}. \tag{2.18}$$

Taking into account the fact that $\mathrm{tr}(\mathbf{ab}) = \mathrm{tr}(\mathbf{ba})$ for any matrices $\mathbf{a}$ and $\mathbf{b}$, we write integral equations (2.15) and (2.16) as

$$\int_{\mathcal{B}} \exp\left\{-0.5 \ \mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\tilde{\mathbf{z}})\right\} \mathbf{V}_\varepsilon(\mathbf{z}) d\mathbf{z} \simeq 2^{-k}\mathbf{\Sigma}, \tag{2.19}$$

$$\int_{\mathcal{B}} \exp\left\{-0.5 \ \mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\tilde{\mathbf{z}})\right\} v_h(\mathbf{z}) d\mathbf{z} \simeq 2^{-k}. \tag{2.20}$$

Here $\mathbf{V}_\varepsilon(\mathbf{z}) = \mathbf{\Psi}_\varepsilon(\tilde{\mathbf{z}}\mathbf{z}) \prod_{j=1}^{k} z_{j,j}^{k-j+1}$, $v_h(\mathbf{z}) = \psi_h(\tilde{\mathbf{z}}\mathbf{z}) \prod_{j=1}^{k} z_{j,j}^{k-j+1}$.

Combination of formulas (2.13) – (2.20) leads to the following result: as soon as $\mathbf{V}_\varepsilon(\mathbf{z})$ and $v_h(\mathbf{z})$ are solutions of the equations (2.19) and (2.20), respectively, estimators $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$ are presented by formulas (2.7) and (2.8) with $\mathbf{\Phi}_\varepsilon(\mathbf{x}; \mathbf{Y})$ and $\varphi_h(\mathbf{x}; \mathbf{Y})$ given by

$$\mathbf{\Phi}_\varepsilon(\mathbf{x}; \mathbf{Y}) = [\det(\mathbf{Y}\mathbf{x}^{-1})]^{\frac{r-k-1}{2}} \mathbf{V}_\varepsilon\left(\sqrt{\mathbf{x} - \mathbf{Y}}\right) \prod_{j=1}^{k} \left[\left(\sqrt{\mathbf{x} - \mathbf{Y}}\right)_{j,j}\right]^{j-k-1} I(\mathbf{x} - \mathbf{Y} \in \mathcal{A}), \tag{2.21}$$

$$\varphi_h(\mathbf{x}; \mathbf{Y}) = [\det(\mathbf{Y}\mathbf{x}^{-1})]^{\frac{r-k-1}{2}} v_h\left(\sqrt{\mathbf{x} - \mathbf{Y}}\right) \prod_{j=1}^{k} \left[\left(\sqrt{\mathbf{x} - \mathbf{Y}}\right)_{j,j}\right]^{j-k-1} I(\mathbf{x} - \mathbf{Y} \in \mathcal{A}). \tag{2.22}$$

Here $\left(\sqrt{\mathbf{x} - \mathbf{Y}}\right)_{j,j}$ is the j-th diagonal element of the matrix $\sqrt{\mathbf{x} - \mathbf{Y}}$. Let us put

$$K = k(k+1)/2$$

and pick up an even number $m = 2m_1$, $m_1 \geq 1$. Formulas (2.21) and (2.22) contain functions $\mathbf{V}_\varepsilon(\mathbf{z})$ and $v_h(\mathbf{z})$ which are to be found. Function $\mathbf{V}_\varepsilon(\mathbf{z})$ is determined by the Lemma 3 which is based on Lemma 1 and Lemma 2. Lemma 4 provides a form for $v_h(\mathbf{z})$.

Lemma 1 . *If*

$$w_{j,m}(t) = \frac{2}{\sqrt{2\pi}} \exp\left\{-\frac{1}{t^2}\right\} \sum_{l=0}^{m-1} \frac{t^{-l-j}}{l!} H_{l+j}\left(\frac{1}{t}\right), \tag{2.23}$$

then

$$\left|\varepsilon^{-j}\int_0^\infty w_{j,m}\left(\frac{t\sqrt{2}}{\varepsilon}\right)\exp\left(-\frac{t^2}{2\theta^2}\right)dt-\theta^{1-j}\right|\le\varepsilon^m\theta^{1-j-m}. \tag{2.24}$$

Here $H_r(t)$ are Hermite polynomials

$$H_r(t)=(-1)^r e^{t^2}\frac{d^r}{dt^r}\left(e^{-t^2}\right),\quad r=0,1,2,...$$

Lemma 2. *Let functions $Q_{j,m}(t), j=1,2,...,k$, be given by the expressions*

$$Q_{j,m}(t)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{t^2}\right\}\sum_{s=0}^{m+j-2}\left[a_s^{(j)}t^{-(2s+2)}\right], \tag{2.25}$$

where coefficients $a_s^{(j)}$, $s=0,...,m+j-2$, are the solutions of the system of the linear equations

$$\sum_{s=max(0,l-1)}^{j+m-2}\left[a_s^{(j)}b_{l,s}^{(j)}\right]=\frac{I(l\ge j-1)}{(l-j+1)!},\quad l=0,...,j+m-2, \tag{2.26}$$

with

$$b_{l,s}^{(j)}=\frac{2^l\,(2s-l)!}{l!\,(s-l)!}(j-1)I(l\le s)+\frac{(2s-l)!\,2^{l-1}}{(s+1-l)!\,(l-2)!}I(l\ge 2), \tag{2.27}$$

if $s\ge l-1$, $l=0,...,j+m-2$, and $b_{l,s}^{(j)}=1$, if $l=1,s=0$.

Then for some $C>0$ independent of ε and θ, the following inequality is valid

$$\left|\varepsilon^{-j}\int_0^\infty Q_{j,m}\left(\frac{t\sqrt{2}}{\varepsilon}\right)\exp\left(-\frac{t^2}{2\theta^2}\right)\left[(j-1)+\frac{t^2}{\theta^2}\right]dt-\theta^{1-j}\right|\le C\varepsilon^m\theta^{1-j-m}, j=1,2,...,k. \tag{2.28}$$

Lemma 3. *Approximate solution $\mathbf{V}_\varepsilon(\mathbf{z})$ of equation (2.19) has the form*

$$\mathbf{V}_\varepsilon(\mathbf{z})=\tilde{\mathbf{z}}\,\mathbf{U}(\mathbf{z};\varepsilon)\,\mathbf{z}, \tag{2.29}$$

where $\mathbf{U}(\mathbf{z};\varepsilon)$ is the diagonal matrix with elements

$$U_{j,j}(\mathbf{z};\varepsilon)=2^{-k}\varepsilon^{-K}Q_{j,m}\left(\frac{z_{j,j}\sqrt{2}}{\varepsilon}\right)\prod_{\substack{i=1\\i\ne j}}^{k}w_{i,m}\left(\frac{z_{i,i}\sqrt{2}}{\varepsilon}\right). \tag{2.30}$$

Here functions $w_{i,m}(t)$ and $Q_{j,m}(t)$ are defined by formulas (2.23) and (2.25) – (2.27), respectively.

Lemma 4. *Approximate solution $v_h(\mathbf{z})$ of equation (2.20) has the form*

$$v_h(\mathbf{z}) = h^{-K} \prod_{j=1}^{k} u_j\left(\frac{z_{j,j}}{h}\right) \prod_{i=1}^{j-1} u_{k+1}\left(\frac{z_{i,j}}{h}\right), \tag{2.31}$$

where functions $u_j(t)$, $j = 1, ..., k, k+1$, are even and satisfy the following conditions

$$\int_{-\infty}^{\infty} u_j(t) t^l dt = \begin{cases} 1, & \text{if } l = 0, \\ 0, & \text{if } l = 1, ..., m-1, \\ \lambda_j \neq 0, & \text{if } l = m, \ m = 2m_1, \end{cases} \tag{2.32}$$

$$\int_0^{\infty} u_j^2(t)\, t^{j-k-1} dt < \infty, \ j = 1, ..., k, k+1. \tag{2.33}$$

Validity of Lemma 1 can be verified by mere substitution of $w_{j,m}$ into (2.24) and application of formulas 7.386 and 8.950 of Gradshtein & Ryzhik(1980). Lemmas 2, 3 and 4 are proved in Section 5.

Based on Lemma 3 and Lemma 4, we formulate the following Theorem.

Theorem 1. *Estimator $\mathbf{F}_n(\mathbf{Y})$ of $\mathbf{F}(\mathbf{Y})$ is given by (2.7), where $\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})$ has the form (2.21) with $\mathbf{V}_\varepsilon(\mathbf{z})$ determined by Lemma 3. Estimator $p_n(\mathbf{Y})$ of $p(\mathbf{Y})$ is (2.8), where $\varphi_h(\mathbf{x};\mathbf{Y})$ is defined by (2.22) with $v_h(\mathbf{z})$ of the form (2.31). Here parameters $\varepsilon = \varepsilon(n)$ and $h = h(n)$ have the following asymptotic form*

$$\varepsilon(n) \sim h(n) \sim n^{-\frac{1}{2m+2K}}. \tag{2.34}$$

Mean square errors (2.9) and (2.10) of the estimators $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$ are bounded by

$$\varrho_n(\mathbf{Y}) \leq C_1 n^{-\frac{m}{m+K}} \left\{ \left[\int_A q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) \ tr(\mathbf{\Sigma}) \left(det(\mathbf{\Sigma})\right)^{1/2} \sum_{j=1}^{k} (\sqrt{\mathbf{\Sigma}})_{j,j}^{-m} d\mathbf{\Sigma} \right]^2 + \right.$$
$$\left. + \ \int_A q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) [tr(\mathbf{\Sigma})]^2 \left[1 + \sum_{j=1}^{k} H_{j,j}^{-4}\right] d\mathbf{\Sigma}, \right\}; \tag{2.35}$$

$$\sigma_n(\mathbf{Y}) \leq C_2 n^{-\frac{m}{m+K}} \left\{ p(\mathbf{Y}) + \left[\int_A q(\mathbf{Y}/\mathbf{\Sigma}) [tr(\mathbf{\Sigma}^{-1})]^m g(\mathbf{\Sigma}) d\mathbf{\Sigma} \right]^2 \right\}, \tag{2.36}$$

respectively, where $K = k(k+1)/2$ and constants C_1 and C_2 are independent of n and $\mathbf{Y}$.

Theorem 1 is proved in Section 5.

3 EB estimation of $\mathbf{\Sigma}$

Now, as we constructed estimators $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$ of $\mathbf{F}(\mathbf{Y})$ and $p(\mathbf{Y})$, respectively, our objective is to obtain an estimator of $\mathbf{S}(\mathbf{Y})$. For this purpose we select some positive constants μ and ν, and

$$\delta_n \sim n^{-\frac{\mu m}{2(\mu+1)(m+K)}}, \tag{3.1}$$

and put

$$p_{n\delta}(\mathbf{Y}) = p_n(\mathbf{Y})[1+\delta_n(p_n(\mathbf{Y}))^{-\mu}]^{\nu}. \tag{3.2}$$

Then our proposed EB estimator of $\mathbf{\Sigma}$ turns out to be

$$\mathbf{S}_n(\mathbf{Y}) = [p_{n\delta}(\mathbf{Y})]^{-1}\,\mathbf{F}_n(\mathbf{Y}), \tag{3.3}$$

where $\mathbf{F}_n(\mathbf{Y})$ and $p_n(\mathbf{Y})$ are defined in Theorem 1.

Let us show that the estimator (3.3) is pointwise asymptotically optimal, i.e., its posterior risk $\Delta_n(\mathbf{Y})$ tends to zero as $n \to \infty$ for every $\mathbf{Y}$. For this purpose we establish the relationship between $\Delta_n(\mathbf{Y})$ and $\varrho_n(\mathbf{Y})$ and $\sigma_n(\mathbf{Y})$ (see (1.6),(2.35), (2.36)).

Theorem 2. *There exists positive C_3 independent of n and $\mathbf{Y}$, such that for each value of $\mathbf{Y}$*

$$\begin{aligned}\Delta_n(\mathbf{Y}) \le C_3 \Big\{ & \delta_n^{-\frac{2}{\mu}}\varrho_n(\mathbf{Y}) + \delta_n^2\,\mathrm{tr}[\widetilde{\mathbf{F}}(\mathbf{Y})\mathbf{F}(\mathbf{Y})](p(\mathbf{Y}))^{-2-2\mu} + \\ & + \sigma_n(\mathbf{Y})\,\mathrm{tr}[\widetilde{\mathbf{F}}(\mathbf{Y})\mathbf{F}(\mathbf{Y})]p^{-2}(\mathbf{Y})\left[p^{-2}(\mathbf{Y}) + \delta_n^{-\frac{2}{\mu}}\right]\Big\}.\end{aligned} \tag{3.4}$$

Substituting δ_n, $\varrho_n(\mathbf{Y})$ and $\sigma_n(\mathbf{Y})$ into (3.4), we finally get that there exists a function $C(\mathbf{Y})$ such that

$$\Delta_n(\mathbf{Y}) \le C(\mathbf{Y})\, n^{-\frac{\mu m}{(\mu+1)(m+K)}}. \tag{3.5}$$

The last inequality means that estimator (3.3) is poinwise asymptotically optimal for every $\mathbf{Y}$.

Inequality (3.5) contains the function $C(\mathbf{Y})$ which is finite for every $\mathbf{Y}$. Formula (3.4) gives us the idea of the relationship between $C(\mathbf{Y})$ and the value of μ. Let $\|\mathbf{Y}\|$ be a norm of the matrix $\mathbf{Y}$. Since $p(\mathbf{Y}) \to 0$ as $\|\mathbf{Y}\| \to \infty$, then for big values of $\|\mathbf{Y}\|$, the smaller μ we pick up, the smaller $\Delta_n(\mathbf{Y})$ we obtain. On the other hand, the smaller μ is selected, the smaller rate of convergence we get (compare with (3.5)). It is likely that the bound (3.4) can be further improved by the choice of δ_n depending on $\mathbf{Y}$, i.e.,

$$\delta_n = \delta_n(\mathbf{Y}).$$

However, this choice, as well as examination of the dependence between $C(\mathbf{Y})$ and parameter m, is the matter of future studies.

4 Computational aspects

As it was mentioned in the Introduction, EB estimation of the parameter $\mathbf{\Sigma}$ of the Wishart distribution involves estimation in the $k(k+1)/2$-dimensional space, therefore it could be computationally intractable. However, in this paper we succeed in the rejection of the amount of computations. Actually, estimator $\mathbf{S}_n(\mathbf{Y})$ is expressed explicitly in terms of $\mathbf{V}_\varepsilon(\mathbf{z})$ and $v_h(\mathbf{z})$ which, in turn, have the known forms (see Lemma 3 and Lemma 4) with some parameters to be evaluated.

Matrix function $\mathbf{V}_\varepsilon(\mathbf{z})$ contains coefficients $a_s^{(j)}, j = 1, ..., k,\ s = 0, 1, ..., m+j-2$, in its presentation. These coefficients are provided by k systems of $(m+j-1)$ linear equations with $(m+j-1)$ unknown values each. It is easy to notice that matrices of the systems are almost triangle (see (2.27)), namely, $b_{l,s}^{(j)} = 0$, if $s \le l-2$, and thus some special methods can be applied to their solution.

Scalar function $v_h(\mathbf{z})$ is the product of the functions $u_j(t),\ j = 1, 2, ..., k, k+1$, that satisfy the conditions (2.32) and (2.33). Conditions similar to (2.32) and (2.33) are very common in kernel density estimation. For the sake of construction of functions $u_j(t),\ j = 0, 1, ..., k$, we choose a system of linear-independent even functions $\{\varphi_i(t)\},\ i = 0, 2, ..., m-1$, and put

$$u_j(t) = |t|^{\tau_j} \sum_0^{m-1} \alpha_i^{(j)} \varphi_i(t). \tag{4.1}$$

Here $\tau_j = \mathrm{Int}(0.5(k+2-j))$ and $\mathrm{Int}(t)$ is the integer part of t. Coefficients $\alpha_i^{(j)}$ are the solutions of the systems of m linear equations with m unknown values

$$\sum_{i=0}^{m-1} \alpha_i^{(j)} \beta_{i,l}^{(j)} = \gamma_l^{(j)}, \quad l = 0, ..., m-1, \quad j = 1, ..., k, k+1, \tag{4.2}$$

with

$$\gamma_l^{(j)} = \begin{cases} 1, & \text{if } l = 0, \\ 0, & \text{if } l = 1, ..., m-1, \end{cases}$$

$$\beta_{i,l}^{(j)} = \begin{cases} \int_{-\infty}^{\infty} |t|^{\tau_j + l} \varphi_i(t) dt, & \text{if } l \text{ is even}, \\ 0, & \text{otherwise}. \end{cases}$$

Usually solution of the systems of the form (4.2) is a standard procedure, and, since $m << 0.5\,k(k+1)$, the size of the system is not high.

5 Proofs

Proof of Lemma 2. To prove Lemma 2, we apply formula 3.471.9 of Gradshtein & Ryzhik(1980)

$$\int_0^\infty u^{\nu-1} \exp\left\{-\frac{1}{u} - \gamma u\right\} du = 2\left(\frac{\beta}{\gamma}\right)^{\nu/2} K_\nu\left(2\sqrt{\beta\gamma}\right)$$

with $K_\nu(t)$ being the Bessel function of the first kind (see 8.402 of Gradshtein & Ryzhik (1980)), $\gamma = \theta^{-2}$, $\beta = \varepsilon^2/4$, $\nu = -(s+0.5)$. We will also use the fact that $K_\nu(t) = K_{-\nu}(t)$ for any ν and positive t. Changing variables $u = t^2$, we obtain

$$\int_0^\infty t^{-(2s+2)} \exp\left\{-\frac{\varepsilon^2}{2t^2} - \frac{t^2}{2\theta^2}\right\} dt = (\theta\varepsilon)^{-(s+0.5)} K_{s+0.5}\left(\frac{\varepsilon}{\theta}\right). \quad (5.1)$$

Applying series expansion for $K_{s+0.5}(t)$ (see 8.468, Gradshtein & Ryzhik(1980))

$$K_{s+0.5}(t) = \sqrt{\frac{\pi}{2t}} e^{-t} \sum_{l=0}^{s} \frac{(s+l)!}{l!(s-l)!(2t)^l}$$

and combining two equalities of the form (5.1), we get

$$\frac{1}{\varepsilon}\int_0^\infty \left(\frac{t}{\varepsilon}\right)^{-(2s+2)} \exp\left\{-\frac{\varepsilon^2}{2t^2} - \frac{t^2}{2\theta^2}\right\}\left[(j-1) + \frac{t^2}{\theta^2}\right] dt = \exp\left\{-\frac{\varepsilon}{\theta}\right\}\sum_{l=0}^{s+1} b_{l,s}^{(j)}\left(\frac{\theta}{\varepsilon}\right)^{-l}, \quad (5.2)$$

where $b_{l,s}^{(j)}$ are defined by formula (2.27). Now we multiply both parts of (5.2) by $a_s^{(j)}$ and take a sum over $s = 0, 1, ..., m+j-2$. After changing the order of summation and recalling that $a_s^{(j)}$, $s = 0, ..., m+j-2$, are the solutions of the system of linear equations (2.26), we arrive at the following relationship

$$\left|\varepsilon^{-j}\int_0^\infty \exp\left(-\frac{t^2}{2\theta^2}\right) Q_{m,j}\left(\frac{t\sqrt{2}}{\varepsilon}\right)\left[(j-1) + \frac{t^2}{\theta^2}\right] dt - \theta^{1-j}\right| \le$$

$$\le \varepsilon^m \theta^{1-j-m}\left\{1 + \sum_{s=max(0,l-1)}^{j+m-2} a_s^{(j)} b_{m+j-1,s}^{(j)}\right\},$$

which proves Lemma 2.

Proof of Lemma 3 . Since $\mathbf{\Sigma}$ is a symmetric positive definite matrix, then there exists matrix $\mathbf{H} = \sqrt{\mathbf{\Sigma}}$. By definiion, $\mathbf{\Sigma} = \widetilde{\mathbf{H}}\mathbf{H}$, so that $\mathbf{\Sigma}^{-1} = \mathbf{H}^{-1}\widetilde{\mathbf{H}}^{-1}$. Let us change variables $\mathbf{z} = \mathbf{tH}$ in (2.19). Jacobian of the transformation is $\prod_{j=1}^k H_{j,j}^j$, and formula (2.19) takes the form

$$\int_{\mathcal{B}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{t}\widetilde{\mathbf{t}})\right\} \mathbf{V}_\varepsilon(\mathbf{tH}) dt \simeq 2^{-k}\left(\prod_{j=1}^k H_{jj}^{-j}\right)\widetilde{\mathbf{H}}\mathbf{H}. \quad (5.3)$$

Substituting $\mathbf{V}_\varepsilon(\mathbf{z})$ of the form (2.29) into (5.3), we obtain the system of k equations

$$\int_0^\infty \exp\left(-\frac{t^2}{2}\right) Q_{j,m}(tH_{jj},\varepsilon)[(j-1)+t^2]dt \prod_{\substack{i=1\\ i\neq j}}^{k} \int_0^\infty \exp\left(-\frac{t^2}{2}\right) w_{i,m}(tH_{ii},\varepsilon)dt \simeq$$

$$\simeq 2^{-k}(2\pi)^{-\frac{k(k-1)}{4}} \prod_{i=1}^{k}(H_{ii})^{-i},\ j=1,...,k. \tag{5.4}$$

Each of the last equtions, in its turn, may be broaken into

$$\begin{cases} \int_0^\infty \exp\left(-\frac{t^2}{2\theta^2}\right) w_{i,m}(t,\varepsilon)dt \simeq \theta^{1-i}, \quad i\neq j,\ \theta = H_{ii}, \\ \\ \int_0^\infty \exp\left(-\frac{t^2}{2\theta^2}\right) Q_{j,m}(t,\varepsilon)[(j-1)+\frac{t^2}{\theta^2}]dt \simeq 2^{-k}(2\pi)^{-\frac{k(k-1)}{4}}\theta^{1-j},\ \theta = H_{jj}. \end{cases} \tag{5.5}$$

It follows from Lemma 1 and Lemma 2 that equations (5.5) are asymptotically valid, which completes the proof.

Proof of Lemma 4. To prove the Lemma we should simply verify that the difference

$$\beta_h(\mathbf{\Sigma}) = \int_{\mathcal{A}} \exp\left\{-0.5\,\mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\tilde{\mathbf{z}})\right\} v_h(\mathbf{z})d\mathbf{z} - 2^{-k} \tag{5.6}$$

converges to zero as $h \to 0$. Actually, changing variables $h^{-1}\mathbf{z} \to \mathbf{z}$ and using Taylor expansion with Lagrange's reminder for the exponential function, we obtain

$$\beta_h(\mathbf{\Sigma}) = \sum_{i=0}^{m-1}\left[\int_{\mathcal{B}} \frac{h^{2i}}{2^i i!}\left[\mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\tilde{\mathbf{z}})\right]^i \prod_{j=1}^{k} u_j(z_{j,j}) \prod_{i<j} u_{k+1}(z_{i,j})d\mathbf{z}\right] +$$

$$+\int_{\mathcal{B}} \frac{h^m}{2^{m_1}(m_1)!}\left[\mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\tilde{\mathbf{z}})\right]^{m_1} \exp\{-0.5\xi\} \prod_{j=1}^{k} u_j(z_{j,j}) \prod_{i<j} u_{k+1}(z_{i,j})d\mathbf{z} - 2^{-k},$$

where $0<\xi<\mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\tilde{\mathbf{z}})$; $m_1 = m/2$.

According to the conditions (2.32) and (2.33), the first component of the sum is equal to 2^{-k} and all the others vanish, so that

$$|\beta_h(\mathbf{\Sigma})| \leq \frac{h^m}{2^{m_1}(m_1)!}\int_{\mathcal{B}}\left[\mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\tilde{\mathbf{z}})\right]^{m_1} v_h(\mathbf{z})d\mathbf{z}.$$

The last integral is dominated by

$$|\beta_h(\mathbf{\Sigma})| \leq \frac{h^m}{2^{m_1}(m_1)!}\sum_{j=1}^{k}\left\{(\lambda_{k+1}(j-1)+\lambda_j)\left[(\mathbf{\Sigma}^{-1})_{jj}\right]^{m_1}\right\}. \tag{5.7}$$

The right part of inequality (5.7) tends to zero as $h \to 0$, which proves the statement.

Proof of Theorem 1. Let us first calculate mean square risk $\varrho_n(\mathbf{Y})$ of the estimator $\mathbf{F}_n(\mathbf{Y})$. It has the form (2.11), and so we need to find upper bounds for $\mathbf{B}_\varepsilon(\mathbf{Y})$ and for $D_\varepsilon(\mathbf{Y})$.

Changing variables $\mathbf{t} = \left(\sqrt{\mathbf{x}-\mathbf{Y}}\right)\mathbf{H}^{-1}$ and taking into account the equality $\mathbf{\Sigma} q(\mathbf{Y}/\mathbf{\Sigma}) = \tilde{\mathbf{H}} q(\mathbf{Y}/\mathbf{\Sigma})\mathbf{H}$, we present $\mathbf{B}_\varepsilon(\mathbf{Y})$ as

$$\mathbf{B}_\varepsilon(\mathbf{Y}) = \int_{\mathcal{A}} q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) \tilde{\mathbf{H}}\, \mathbf{R}_\varepsilon(\mathbf{H})\, \mathbf{H}\, d\mathbf{\Sigma}.$$

Here $\mathbf{H} = \sqrt{\mathbf{\Sigma}}$ and $\mathbf{R}_\varepsilon(\mathbf{H})$ is the matrix-valued function of the form

$$\mathbf{R}_\varepsilon(\mathbf{H}) = 2^k \left(\prod_{j=1}^{k} H_{j,j}^{j}\right) \left[\int_{\mathcal{B}} \exp\{-0.5 \sum_{i\le j} t_{i,j}^2\}\, \tilde{\mathbf{t}}\, \mathbf{U}(\mathbf{tH};\varepsilon)\, \mathbf{t} d\mathbf{t} - \mathbf{E}\right]. \tag{5.8}$$

Since $\mathbf{U}(\mathbf{tH};\varepsilon)$ is a diagonal matrix, elements of which depend on the products $t_{j,j}H_{j,j}$, $j = 1, ..., k$, only, then from symmetric considerations we obtain that the integral in the right part of equality (5.8) is also a diagonal matrix. It implies that $\mathbf{R}_\varepsilon(\mathbf{H})$ is the diagonal matrix with j-th diagonal element equal to the difference between the left and the right part of the formula (5.4) times $\left(2^k \prod_{i=1}^{k} H_{i,i}^{i}\right)$. For the sake of construction of upper bounds for the diagonal elements of the matrix $\mathbf{R}_\varepsilon(\mathbf{H})$, we need the following apparent assertion.

Lemma 5. *For any $\gamma_i(\varepsilon)$ and γ_i*

$$\left|\prod_{i=1}^{k} \gamma_i(\varepsilon) - \prod_{i=1}^{k} \gamma_i\right| \le \sum_{j=1}^{k} \left\{ |\gamma_i(\varepsilon) - \gamma_i| \prod_{\substack{i=1\\ i\ne j}}^{k} \left(|\gamma_i| + |\gamma_i(\varepsilon) - \gamma_i|\right)\right\}.$$

Applying Lemma 5 with (see Lemma 1 and Lemma 2)

$$\gamma_i(\varepsilon) = \int_0^\infty \varepsilon^{-j} w_{i,m}\left(\frac{t\sqrt{2}}{\varepsilon}\right) \exp\left(-\frac{t^2}{2\theta^2}\right) dt,$$

$\gamma_i = \theta^{1-i}$, $\theta = H_{i,i}$ for $i \ne j$, and

$$\gamma_j(\varepsilon) = \varepsilon^{-j} \int_0^\infty \exp\left(-\frac{t^2}{2\theta^2}\right) Q_{m,j}\left(\frac{t\sqrt{2}}{\varepsilon}\right) \left[(j-1) + \frac{t^2}{\theta^2}\right] dt,$$

$\gamma_j = \theta^{1-j}$, $\theta = H_{j,j}$, we obtain that there exists $C_5 > 0$ such that each of the elements of the matrix $\mathbf{R}_\varepsilon(\mathbf{H})$ is bounded by $C_5 \varepsilon^m \prod_{j=1}^{k} H_{j,j} \sum_{j=1}^{k} H_{j,j}^{-m}$. Therefore

$$|(\mathbf{B}_\varepsilon(\mathbf{Y}))_{i,j}| \le C_5\, \varepsilon^m \sum_{l=1}^{min(i,j)} \int_{\mathcal{A}} q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) |H_{l,i}||H_{l,j}| \prod_{j=1}^{k} H_{j,j} \sum_{j=1}^{k} H_{j,j}^{-m} d\mathbf{\Sigma}.$$

The last inequality implies that there exists $C_6 > 0$ such that

$$\mathrm{tr}[\widetilde{\mathbf{B}_\varepsilon(\mathbf{Y})}\mathbf{B}_\varepsilon(\mathbf{Y})] \leq C_6 \varepsilon^{2m} \left[\int_{\mathcal{A}} q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) \mathrm{tr}(\mathbf{\Sigma}) \det(\mathbf{\Sigma}^{1/2}) \sum_{j=1}^{k} (\sqrt{\mathbf{\Sigma}})_{j,j}^{-m} d\mathbf{\Sigma} \right]^2 . \quad (5.9)$$

Note that, provided the second assumption in (1.7) holds, the integral in (5.9) is finite. It happens because for any positive definite $\mathbf{Y}$ and any $m > 0$

$$\sup_{\mathbf{Y}} \left| q(\mathbf{Y}/\mathbf{\Sigma}) (\sqrt{\mathbf{\Sigma}})_{ii}^{-m} \right| < \infty, \quad i = 1, ..., k.$$

To finish the proof of the inequality (2.35), it remains to find upper bounds for $D_\varepsilon(\mathbf{Y})$. For this purpose in formula (2.3) we substitute (2.21) and (1.2) for $\mathbf{\Phi}_\varepsilon(\mathbf{x};\mathbf{Y})$ and $p(\mathbf{Y})$, respectively, and then make change of variables $\mathbf{t} = \left(\sqrt{\mathbf{x} - \mathbf{Y}}\right)\mathbf{H}^{-1}$, where $\mathbf{H} = \sqrt{\mathbf{\Sigma}}$. So we obtain that

$$D_\varepsilon(\mathbf{Y}) = 2^k \int_{\mathcal{A}} q(\mathbf{Y}/\mathbf{\Sigma}) g(\mathbf{\Sigma}) \widehat{D}_\varepsilon(\mathbf{\Sigma}) d\mathbf{\Sigma}$$

with

$$\widehat{D}_\varepsilon(\mathbf{\Sigma}) = \int_{\mathcal{B}} \mathrm{tr}[\tilde{\mathbf{t}}\mathbf{U}(\mathbf{tH},\varepsilon)\mathbf{tH}\tilde{\mathbf{H}}\tilde{\mathbf{t}}\mathbf{U}(\mathbf{tH},\varepsilon)\mathbf{tH}\tilde{\mathbf{H}}] \prod_{j=1}^{k} \left(t_{j,j}^{j-k-1} H_{j,j}^{2j-k-1}\right) \exp\{-\frac{1}{2}\sum_{i\leq j} t_{i,j}^2\} d\mathbf{t}. \quad (5.10)$$

Since for any symmetric positive definite matrix $\mathbf{a}$, the inequality holds (see 13.215.6.ii of Gradshtein & Ryzhik(1980))

$$a_{ij} \leq \sqrt{a_{ii}a_{jj}},$$

then for any symmetric positive definite matrices $\mathbf{a}$ and $\mathbf{b}$

$$\mathrm{tr}(\mathbf{ab}) \leq \mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b}).$$

Applying the last inequality to (5.10) with $\mathbf{a} = \mathbf{H}\tilde{\mathbf{H}}$ and $\mathbf{b} = \tilde{\mathbf{t}}\mathbf{U}(\mathbf{tH},\varepsilon)\mathbf{t}$ and taking into account the identity $\mathrm{tr}(\mathbf{H}\tilde{\mathbf{H}}) = \mathrm{tr}(\mathbf{\Sigma})$, we get

$$\widehat{D}_\varepsilon(\mathbf{\Sigma}) \leq [\mathrm{tr}(\mathbf{\Sigma})]^2 \int_{\mathcal{B}} \prod_{j=1}^{k} \left(t_{j,j}^{j-k-1} H_{j,j}^{2j-k-1}\right) \left[\mathrm{tr}(\tilde{\mathbf{t}}\mathbf{U}(\mathbf{tH},\varepsilon)\mathbf{t})\right]^2 \exp\{-\frac{1}{2}\sum_{i\leq j} t_{i,j}^2\} d\mathbf{t}. \quad (5.11)$$

Then we calculate the integral in (5.11) in veiw of

$$\left(\mathrm{tr}(\tilde{\mathbf{t}}\mathbf{U}\mathbf{t})\right)^2 = \left(\sum_{i\leq j} t_{i,j}^2 U_{i,i}\right)^2 \leq k\,[\mathrm{tr}(\mathbf{U}^2)] \sum_{i\leq j} t_{i,j}^4, \quad \mathbf{U} \equiv \mathbf{U}^2(\mathbf{tH},\varepsilon).$$

So, we arrive at

$$\widehat{D}_\varepsilon(\mathbf{\Sigma}) \leq C_7\varepsilon^{-2K}[\mathrm{tr}(\mathbf{\Sigma})]^2 \left[1 + \varepsilon^4 \sum_{j=1}^{k} H_{j,j}^{-4}\right],$$

where the constant C_7 depends on dimensionality k only. Thus, returning to $D_\varepsilon(\mathbf{Y})$, we obtain

$$D_\varepsilon(\mathbf{Y}) \leq C_7\varepsilon^{-2K} \int_{\mathcal{A}} q(\mathbf{Y}/\mathbf{\Sigma})g(\mathbf{\Sigma})[\mathrm{tr}(\mathbf{\Sigma})]^2[1 + \sum_{j=1}^{k} H_{j,j}^{-4}]d\mathbf{\Sigma}.$$

Replacing ε by the expression (2.34), we get (2.35).

Now we ought tocalculate $\sigma_n(\mathbf{Y})$. Substituting the expression (2.22) into formulas (2.4) and (2.6), we derive that

$$b_h(\mathbf{Y}) = 2^k \int_{\mathcal{A}} \beta_h(\mathbf{\Sigma})q(\mathbf{Y}/\mathbf{\Sigma})g(\mathbf{\Sigma})d\mathbf{\Sigma}, \tag{5.12}$$

$$d_h(\mathbf{Y}) \leq 2^k h^{-2K} \int_{\mathcal{A}} q(\mathbf{Y}/\mathbf{\Sigma})g(\mathbf{\Sigma})\widehat{d}_h(\mathbf{\Sigma})d\mathbf{\Sigma},$$

where

$$\widehat{d}_h(\mathbf{\Sigma}) = \int_{\mathcal{B}} \exp\left\{-\frac{\mathrm{tr}(\mathbf{z}\mathbf{\Sigma}^{-1}\widetilde{\mathbf{z}})}{2}\right\} \prod_{j=1}^{k} \left[z_{j,j}^{j-k-1} u_j^2\left(\frac{z_{j,j}}{h}\right) \prod_{i=1}^{j-1} u_{k+1}^2\left(\frac{z_{i,j}}{h}\right)\right] d\mathbf{z}. \tag{5.13}$$

Now formulas (5.7) and (5.12) provide

$$b_h(\mathbf{Y}) \leq C_8 h^m \int_{\mathcal{A}} \left[\mathrm{tr}(\mathbf{\Sigma}^{-1})\right]^m q(\mathbf{Y}/\mathbf{\Sigma})g(\mathbf{\Sigma})d\mathbf{\Sigma}. \tag{5.14}$$

Changing variables $h^{-1}z_{i,j} \to z_{i,j}$ in (5.13) and taking into account (2.33), we also obtain

$$d_h(\mathbf{Y}) \leq C_9 h^{-2K} p(\mathbf{Y}). \tag{5.15}$$

Here constants C_8 and C_9 are independent of $\mathbf{Y}$ and h. Selecting h of the form (2.34) and combining (5.14) and (5.15), we derive (2.36), which is our objective.

Proof of Theorem 2. First, it is worth noting that since for any $x > 0$ inequality $x^{-1}(1 + \delta x^{-\mu})^{-\nu} \leq [(\mu\nu - 1)\delta]^{1/\mu}$ is valid, then

$$|p_{n\delta}(\mathbf{Y})|^{-1} \leq (\mu\nu - 1)^{1/\mu}\delta^{1/\mu} \equiv C_\mu \delta^{1/\mu} \tag{5.16}$$

for every $\mathbf{Y}$. Let us present the (i,j)-th component of the difference $\mathbf{S}_n(\mathbf{Y}) - \mathbf{S}(\mathbf{Y})$ in a form

$$(\mathbf{S}_n(\mathbf{Y}))_{i,j} - (\mathbf{S}(\mathbf{Y}))_{i,j} = p_{n\delta}^{-1}(\mathbf{Y})\left[(\mathbf{F}_n(\mathbf{Y}))_{i,j} - \mathbf{F}(\mathbf{Y})_{i,j}\right] + (\mathbf{F}(\mathbf{Y}))_{i,j}\left[p_{n\delta}^{-1}(\mathbf{Y}) - p^{-1}(\mathbf{Y})\right].$$

Multiplying each part of the equality by its transpose, taking the trace of both parts and using the Cauchy inequality and (5.16), we get (see (2.9))

$$\Delta_n(\mathbf{Y}) \leq 2\left\{C_\mu^2\, \delta^{-2/\mu}\, \varrho_n(\mathbf{Y}) \;+\; \mathrm{tr}[\tilde{\mathbf{F}}(Y)\mathbf{F}(\mathbf{Y})]\; \mathbf{E}_{p^n}[p_{n\delta}^{-1}(\mathbf{Y}) - p^{-1}(\mathbf{Y})]^2\right\}. \qquad (5.17)$$

So, now we ought to find upper bounds for $\mathbf{E}_{p^n}[p_{n\delta}^{-1}(\mathbf{Y}) - p^{-1}(\mathbf{Y})]^2$. To do that we partition $\Omega \times \mathcal{A}$ into two parts $\mathcal{G} = \{(\omega, \mathbf{Y}) : |p_n(\mathbf{Y}) - p(\mathbf{Y})| < 0.5p(\mathbf{Y})\}$ and its complement $\overline{\mathcal{G}}$. Thus, mathematical expectation also breaks up into two components

$$\mathbf{E}_{p^n}\left[p_{n\delta}^{-1}(\mathbf{Y}) - p^{-1}(\mathbf{Y})\right]^2 \leq 2\int_{\mathcal{G}}\left[p_{n\delta}^{-1}(\mathbf{Y}) - p^{-1}(\mathbf{Y})\right]^2 dP +$$

$$+2\left[C_\mu \delta_n^{-1/\mu} + p^{-1}(\mathbf{Y})\right]^2 P\left\{|p_n(\mathbf{Y}) - p(\mathbf{Y})| > 0.5p(\mathbf{Y})\right\}.$$

Now using Taylor expansion of $p_{n\delta}(y) = p_n(\mathbf{Y})[1+\delta_n(p_n(\mathbf{Y}))^{-\mu}]^\nu$ at a point $p_n(\mathbf{Y}) = p(\mathbf{Y})$ and $\delta_n = 0$ and applying Chebishev inequality, we obtain

$$\mathbf{E}_{p^n}(p_{n\delta}^{-1}(\mathbf{Y}) - p^{-1}(\mathbf{Y}))^2 \leq 4p^{-2}(\mathbf{Y})\left\{\sigma_n(\mathbf{Y}) + \delta_n^2(p(\mathbf{Y}))^{-2\mu} + \left[C_\mu\delta_n^{-1/\mu} + p^{-1}(\mathbf{Y})\right]^2 \sigma_n(\mathbf{Y})\right\}.$$

Combination of the last inequality and (5.17) completes the proof.

References

[1] ANDERSON, T.W. (1984). *An introduction to multivariate statistical analysis.* Wiley, New York.

[2] GHOSH, M. (1992). Hierarchical and empirical Bayes multivariate estimation. In *Current issues in statistical inference. IMS Lecture Notes, Monograph Ser.*, , Vol. 17, pp. 151-177.

[3] GHOSH, M., AND SHIEH G. (1991). Empirical Bayes minimax estimators of matrix normal means. *J. Multivariate Anal.* **38** 305-318.

[4] GHOSH, M., AND SHIEH, G. (1992). Empirical Bayes minimax estimators of the matrix normal means. *Nonparametric Statistics and related topics*, 231-251.

[5] GRADSHTEIN, I.S., AND RYZHIK, I.M. (1980). *Tables of integrals, series, and products.* Academic Press, New York.

[6] JUDGE, G.G., HILL, R.S., AND BOCK, M.E. (1990). An adaptive empirical Bayes estimator of the multivariate normal mean under quadratic loss. *J. Econometrics* **44** 189-213.

[7] KUBOKAWA, T., ROBERT, C., AND CALEH, A.K.MD.E. (1992). Empirical Bayes estimation of the covariance matrix of a normal distribution with unknown mean under the entropy loss. *Sankhyā Ser.A* **54** 402-410.

[8] PENSKAYA, M. (1992). On empirical Bayes estimation of a shift parameter. *Theory Probab. Appl.* **37** 732-734.

[9] PENSKAYA, M.Y. (1993). *A general approach to empirical Bayes estimation.* University of Lund Stat. Research Report, **6**, pp. 1-29.

[10] PENSKAYA, M. (1995). On mean square consistent estimation of a ratio. *Scand. J. Statist.* **22** 129-137.

[11] SHIEH, G. (1993). Empirical Bayes minimax estimators of matrix normal means for arbitrary quadratic loss and unknown covariance matrix. *Statist. Decisions* **11** 317-341.