Kryštof Eben

A generalization of Wishart density for the case when the inverse of the covariance matrix is a band matrix

Mathematica Bohemica, Vol. 119 (1994), No. 4, 337–346

Persistent URL: http://dml.cz/dmlcz/126118

119 (1994) MATHEMATICA BOHEMICA No. 4, 337–346

A GENERALIZATION OF WISHART DENSITY FOR THE CASE WHEN THE INVERSE OF THE COVARIANCE MATRIX IS A BAND MATRIX

KRYŠTOF EBEN, Praha

(Received October 10, 1990)

Summary. In a multivariate normal distribution, let the inverse of the covariance matrix be a band matrix. The distribution of the sufficient statistic for the covariance matrix is derived for this case. It is a generalization of the Wishart distribution. The distribution may be used for unbiased density estimation and construction of classification rules.

Keywords: Band inverse covariance matrix, Wishart distribution, unbiased density estimation, discriminant analysis

AMS classification: 62H10

1. Introduction

Let $\mathbf{X}$ have a p-dimensional normal distribution $N_p(\mu, \Sigma)$, where Σ^{-1} is a band matrix, that is, Σ^{-1} is of the form

$$\Sigma^{-1} = \begin{pmatrix} \sigma^{11} & \dots & \sigma^{1q} & & & \\ \vdots & & & & 0 & \\ \sigma^{q1} & & & \ddots & & \\ & & & & & \\ & & \ddots & & & \sigma^{p-q+1,p} \\ & 0 & & & & \vdots \\ & & & \sigma^{p,p-q+1} & \dots & \sigma^{pp} \end{pmatrix}.$$

The above form of Σ^{-1} occurs, for example, if the coordinates of $\mathbf{X} = (X_1, \dots, X_p)'$ are successive observations of a time-dependent variable and the structure of dependency is Markovian of order $q-1$. Indeed, in this case we have for $j-i>q-1$

$$\begin{aligned}\operatorname{cov}(X_i, X_j|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p) \\ = \operatorname{cov}(X_i, X_j|X_{i+1}, \ldots, X_{j-1}) = 0\end{aligned}$$

and the partial correlation coefficient of X_i and X_j, given all remaining variables, equals zero. But this implies $\sigma^{ij} = 0$.

This type of data occurs in medical experiments, e.g. when measuring a biochemical characteristic of blood in newborns, at 10 fixed time instants, starting from the birth.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be a random sample from the above distribution and put

$$C = \sum_{i=1}^{N} (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'.$$

For the matrix C (and similarly for other matrices), $C_{(q)}$ will denote the band of C with $q-1$ nonzero subdiagonals:

$$C_{(q)} = \begin{pmatrix} c^{11} & \cdots & c^{1q} & & 0 & \\ \vdots & & & & & \\ c^{q1} & & & \ddots & & \\ & & & & & \\ & & \ddots & & & c^{p-q+1,p} \\ & 0 & & & & \vdots \\ & & & c^{p,p-q+1} & \cdots & c^{pp} \end{pmatrix}.$$

If Σ^{-1} is a band matrix, $\operatorname{tr}(\Sigma^{-1}C) = \operatorname{tr}(\Sigma^{-1}C_{(q)})$ and the likelihood of $\mathbf{X}_1, \ldots, \mathbf{X}_N$ equals [1]

$$\begin{aligned}&(2\pi)^{-Np/2}|\Sigma|^{-N/2}\operatorname{etr}\Big\{-\frac{1}{2}\Sigma^{-1}[C + N(\overline{\mathbf{X}} - \mu)(\overline{\mathbf{X}} - \mu)']\Big\} \\ &= (2\pi)^{-Np/2}|\Sigma|^{-N/2}\operatorname{etr}\Big\{-\frac{1}{2}\Sigma^{-1}[C_{(q)} + N(\overline{\mathbf{X}} - \mu)(\overline{\mathbf{X}} - \mu)']\Big\}.\end{aligned}$$

Clearly, $(\overline{\mathbf{X}}, C_{(q)})$ is a sufficient statistic for the parameters (μ, Σ), and for the purposes of inference, the distribution of $C_{(q)}$ is needed. This distribution we find from the characteristic function, thus generalizing the method of Ingham (1933) for deriving the Wishart distribution.

[1] etr(.) = exp tr(.)

2. The density of $C_{(q)}$

The matrix C has the Wishart distribution $W_p(n, \Sigma)$ with $n = N - 1$ degrees of freedom and the characteristic function

$$E\exp\left\{\mathrm{i}\sum_{j\leqslant k}\theta_{jk}c_{jk}\right\} = E\exp\left\{\frac{\mathrm{i}}{2}\mathrm{tr}C\tilde{\Gamma}\right\} = |I - i\tilde{\Gamma}\Sigma|^{-n/2}, \tag{2.1}$$

where $\tilde{\Gamma} = (\tilde{\gamma}_{ij}), \tilde{\gamma}_{ij} = (1 + \delta_{ij})\theta_{ij}$ and δ_{ij} is the Kronecker delta. Let $\Gamma = (\gamma_{ij})$ be a symmetric band matrix of order p, i.e. $\gamma_{ij} = 0$ for $|i - j| \geqslant q$. Since tr $C_{(q)}\Gamma =$ tr $C\Gamma$, the characteristic function of $C_{(q)}$ is obtained from (2.1) as

$$\varphi(\Gamma) = E\exp\left\{\frac{\mathrm{i}}{2}\mathrm{tr}C_{(q)}\Gamma\right\} = E\exp\left\{\frac{\mathrm{i}}{2}\mathrm{tr}C\Gamma\right\} = |I - \mathrm{i}\Gamma\Sigma|^{-n/2}.$$

First we have to prove two lemmas concerning the absolute integrability of the characteristic function.

Lemma 1. *Let* $\mathbf{x}$, $\mathbf{y}$ *be vectors of length* p, $A = (a_{ij})$ *and* $\mathbf{y} = A\mathbf{x}$. *Let* $I = \{i_1, \ldots, i_r\} \subset \{1, 2, \ldots, p\}$, $J = \{j_1, \ldots, j_r\} \subset \{1, 2, \ldots, p\}$. *Denote*

$$|A|_{IJ} = \begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} \\ & & \\ a_{i_r j_1} & \cdots & a_{i_r j_r} \end{pmatrix}.$$

Then for the vectors of differentials, obviously $\mathrm{d}\mathbf{y} = A\,\mathrm{d}\mathbf{x}$ *and the wedge product* $\bigwedge_{i\in I} \mathrm{d}y_i$ *of the elements of* $\mathrm{d}\mathbf{y}$ *equals*

$$\bigwedge_{i\in I}\mathrm{d}y_i = \sum_{J\subset\{1,2,\ldots,p\};\mathrm{card}(J)=r} |A|_{IJ}\,\mathrm{d}x_{j_1} \wedge \ldots \wedge \mathrm{d}x_{j_r}\,. \tag{2.2}$$

P r o o f. The assertion (2.2) follows immediately from the antisymmetric property of the wedge product. □

Lemma 2. *The characteristic function of* $C_{(q)}$ *is absolutely integrable for* $n > 2p$.

P r o o f. Clearly

$$\varphi(\Gamma) = |I - i\Sigma^{1/2}\Gamma\Sigma^{1/2}|^{-n/2}.$$

Let $\Sigma^{1/2}\Gamma\Sigma^{1/2} = HLH'$, $L = \mathrm{diag}\{l_1, \ldots, l_p\}$, $l_1 \geqslant l_2 \geqslant \ldots \geqslant l_p$ be the spectral decomposition of $\Sigma^{1/2}\Gamma\Sigma^{1/2}$. Then a straightforward manipulation with differentials (see Muirhead (1982), Theorem 3.2.17) gives

$$\begin{aligned} \mathrm{d}(\Sigma^{1/2}\Gamma\Sigma^{1/2}) &= \Sigma^{1/2}\,\mathrm{d}\Gamma\,\Sigma^{1/2} \\ &= \mathrm{d}H\,LH' + H\,\mathrm{d}L\,H' + HL\,\mathrm{d}H' \end{aligned}$$

and

$$H'\Sigma^{1/2}\,\mathrm{d}\Gamma\,\Sigma^{1/2}H = H'\,\mathrm{d}H\,L - LH'\,\mathrm{d}H + \mathrm{d}L. \tag{2.3}$$

Denote by $\mathrm{d}S = (\mathrm{d}s_{ij})$ the symmetric matrix of differential forms on the right-hand side of (2.3); then

$$\begin{aligned} \mathrm{d}s_{ij} &= (l_i - l_j)\mathbf{h}_j'\,\mathrm{d}\mathbf{h}_i \quad \text{for } i \neq j \\ &= dl_i \quad \text{for } i = j, \end{aligned}$$

where $\mathbf{h}_1, \ldots, \mathbf{h}_p$ are the columns of H. For any matrix A, let vec (A) be the vector formed by the columns of A. According to the well known relations between the vec function and Kronecker product, we have from (2.3)

$$\begin{aligned} \mathrm{vec}(dS) &= (H'\Sigma^{1/2} \otimes H'\Sigma^{1/2})\mathrm{vec}(d\Gamma), \\ \mathrm{vec}(d\Gamma) &= (H'\Sigma^{1/2} \otimes H'\Sigma^{1/2})^{-1}\mathrm{vec}(\mathrm{d}S) \\ &= (\Sigma^{-1/2}H \otimes \Sigma^{-1/2}H)\mathrm{vec}(\mathrm{d}S) \\ &= (\Sigma^{-1/2} \otimes \Sigma^{-1/2})(H \otimes H)\mathrm{vec}(dS). \end{aligned} \tag{2.4}$$

The differential form needed for integration over symmetric band matrices Γ equals the wedge product of non-null distinct elements of $\mathrm{d}\Gamma$, i.e.

$$\bigwedge_{0 \leqslant j-i<q} \mathrm{d}\gamma_{ij}.$$

Using (2.2) with the set I of double indices corresponding to the upper half-band of Γ, $I = \{(1,1), \ldots, (1,q), (2,2), \ldots, (2,q+1), \ldots, (p-q+1, p-q+1), \ldots, (p-q+1, p), \ldots, (p,p)\}$, and the sets $J \subset \{(1,1), \ldots, (1,p), \ldots, (p,1), \ldots, (p,p)\}$, we obtain from (2.4)

$$\bigwedge_{0 \leqslant j-i<q} \mathrm{d}\gamma_{ij} = \sum_{J \subset \{(1,1),\ldots,(p,p)\};\mathrm{card}\, J = \varepsilon} |(\Sigma^{-1/2} \otimes \Sigma^{-1/2})(H \otimes H)|_{IJ} \bigwedge_{(i,j)\in J} \mathrm{d}s_{ij}, \tag{2.5}$$

where $\varepsilon = p + (p-1) + \ldots + (p-q+1) = \frac{q}{2}(2p-q+1)$.

It follows from Hadamard's inequality (e.g. Rao (1965), 1.e. 3.3) that all minors in (2.5) are bounded by the same constant and thus do not affect the integrability of the form. Further, all wedge products of the forms $\mathbf{h}_j'\,\mathrm{d}\mathbf{h}_i$ are integrable due to the finite volume of the orthogonal group. Now the term of the highest order in l_1 involved in the forms

$$\bigwedge_{(i,j)\in J} ds_{ij} = \bigwedge_{(i,j)\in J; i\neq j} (l_i - l_j)\mathbf{h}_j' d\mathbf{h}_i \wedge \bigwedge_{(i,j)\in J; i=j} dl_i$$

equals, just as for the case of ordinary Wishart distribution,

$$\left(\prod_{i=2}^{p}(l_1 - l_i)\right) \mathrm{d}l_1 ,$$

which in the expression for $|\varphi(\Gamma)| = \prod_i (1 + l_i^2)^{-n/2}$ induces the term

$$\frac{\prod_{i=2}^{p}(l_1 - l_i)}{(1 + l_1^2)^{-n/2}} \mathrm{d}l_1 .$$

This, however, is integrable for $\frac{1}{2}n - (p-1) > 1$, i.e. $n > 2p$. The same argument applies to l_2 etc. □

R e m a r k. For the special case of ordinary Wishart distribution, the characteristic function is absolutely integrable for $n > 2p$ (cf. Herz (1955), p. 482) but the density exists and equals the result of the inverse Fourier transform for $n \geqslant p$. For band matrices, we have the same result concerning absolute integrability, $n > 2p$. We shall see below that the result of the inverse Fourier transform is defined for $n \geqslant q$ and it may be conjectured that it equals the density even in this case. The proof would probably require other methods.

Let $\mathcal{B} = \{\Gamma = (\gamma_{ij});\ \gamma_{ij} \in (-\infty, \infty), |i - j| < q\}$ denote the set of all symmetric band matrices Γ. Lemma 2 allows us, for $n > 2p$, to obtain the desired density of $C_{(q)}$ as the result of the inverse Fourier transform:

$$\begin{aligned} f(C_{(q)}) &= \frac{1}{2^p(2\pi)^\varepsilon} \int_{\mathcal{B}} \mathrm{etr}\Big\{ -\frac{\mathrm{i}}{2} C_{(q)}\Gamma\Big\}\varphi(\Gamma)\mathrm{d}\Gamma \\ &= \frac{1}{2^p}|\Sigma|^{-n/2}\frac{1}{(2\pi)^\varepsilon} \int_{\mathcal{B}} \mathrm{etr}\Big\{ -\frac{\mathrm{i}}{2} C_{(q)}\Gamma\Big\}|\Sigma^{-1} - \mathrm{i}\Gamma|^{-n/2}\,\mathrm{d}\Gamma , \end{aligned} \tag{2.6}$$

where ε is as in (2.5). Put $A = \frac{1}{2}C_{(q)}$, $k = \frac{n}{2}$. Similarly as Ingham (1933), we transform (2.6) to a complex integral by changing the sign of Γ and using the substitution $V = \Sigma^{-1} + \mathrm{i}\Gamma$:

$$\begin{aligned} f(C_{(q)}) &= \frac{1}{2^p}|\Sigma|^{-k}\frac{1}{(2\pi)^\varepsilon} \int_{\mathcal{B}} \mathrm{etr}\{i\Gamma A\}|\Sigma^{-1} + i\Gamma|^{-k}\mathrm{d}\Gamma \\ &= 2^{-p}|\Sigma|^{-k}\mathrm{etr}\{-\Sigma^{-1}A\}J(A,p,q), \end{aligned} \tag{2.7}$$

where

$$J(A,p,q) = \frac{1}{(2\pi\mathrm{i})^\varepsilon} \int_{V\in\{\Sigma^{-1}+i\Gamma,\gamma_{ij}\in(-\infty,\infty),|i-j|<q\}} \mathrm{etr}\{AV\}|V|^{-k}\mathrm{d}V. \tag{2.8}$$

Remark. $J(A,p,q)$ is a further generalization of the formula

$$\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} e^{cs}s^{-k}\,ds = \frac{c^{k-1}}{\Gamma(k)}, \qquad c>0,\ a>0,\ k>1 \tag{2.9}$$

whose multivariate analog was considered by Ingham (1933).

For any matrix B, B_j will denote the j-th diagonal block of order q:

$$B_j = \begin{pmatrix} b_{jj} & \cdots & b_{j,j+q-1} \\ \\ b_{j+q-1,j} & \cdots & b_{j+q-1,j+q-1} \end{pmatrix}, \quad j=1,\dots,p-q+1. \tag{2.10}$$

Further, let $\tilde{B}_j$ be the lower right-hand subblock of order $q-1$ taken from the matrix B_j:

$$\tilde{B}_j = \begin{pmatrix} b_{j+1,j+1} & \cdots & b_{j+1,j+q-1} \\ \\ b_{j+q-1,j+1} & \cdots & b_{j+q-1,j+q-1} \end{pmatrix}. \tag{2.11}$$

Lemma 3. *The integral $J(A,p,q)$ does not depend on Σ (if only Σ is positive definite) and equals*

$$J(A,p,q) = \frac{1}{2^{\alpha}\pi^{(p-q)(q-1)/2}(\Gamma(k-\frac{q-1}{2}))^{p-q}\Gamma_q(k)}\,\frac{\prod_{j=1}^{p-q+1}|A_j|^{k-\frac{q+1}{2}}}{\prod_{j=1}^{p-q}|\tilde{A}_j|^{k-\frac{q}{2}}},$$

where $\alpha=(p-q)(q-1)+\frac{q(q-1)}{2}$ and

$$\Gamma_q(k) = \pi^{q(q-1)/4}\prod_{j=1}^{q}\Gamma\Big(k-\frac{j-1}{2}\Big)$$

is the multivariate gamma function.

Proof. For $p=q$, the matrices A and V are no more band matrices and it suffices to compare (2.7) with the density of $W_q(n,\Sigma)$.

Let $p=q+1$ and let us partition A and V as

$$A = \left(\begin{array}{c|c} a & \mathbf{b}'\ 0 \\ \hline \mathbf{b} & \\ 0 & B \end{array}\right), \quad V = \left(\begin{array}{c|c} v & \mathbf{w}'\ 0 \\ \hline \mathbf{w} & \\ 0 & W \end{array}\right),$$

where $B = A_2$, $W = V_2$. Then

$$\operatorname{etr}\{AV\}|V|^{-k} = \exp\{av + 2\mathbf{b}'\mathbf{w}\}\operatorname{etr}\{BW\}v^{-k}\left|W - \frac{1}{v}\begin{pmatrix}\mathbf{w}\\0\end{pmatrix}\begin{pmatrix}\mathbf{w}\\0\end{pmatrix}'\right|^{-k}.$$

As the upper left-hand block of order $q-1$ in B equals $\tilde{A}_1$,

$$J(A, q+1, q) = \left(\frac{1}{2\pi i}\right)^q \int_{v,\mathbf{w}} \exp\left\{av + 2\mathbf{b}'\mathbf{w} + \frac{1}{v}\mathbf{w}'\tilde{A}_1\mathbf{w}\right\}v^{-k}$$

$$\times \left[\left(\frac{1}{2\pi i}\right)^{\varepsilon-q} \int_W \operatorname{etr}\left\{B\left[W - \frac{1}{v}\begin{pmatrix}\mathbf{w}\\0\end{pmatrix}\begin{pmatrix}\mathbf{w}\\0\end{pmatrix}'\right]\right\}\left|W - \frac{1}{v}\begin{pmatrix}\mathbf{w}\\0\end{pmatrix}\begin{pmatrix}\mathbf{w}\\0\end{pmatrix}'\right|^{-k} dW\right] dv\, d\mathbf{w},$$

the integrals being according to (2.8) complex integrals taken over lines parallel to the imaginary axis; e.g. $v \in \{v = \sigma^{11} + i\gamma_{11}\,;\ \gamma_{11} \in (-\infty, \infty)\}$.

Again, B and W are ordinary symmetric matrices; the expression in brackets does not depend on v, $\mathbf{w}$ and equals $J(B, q, q) = J(A_2, q, q)$. It remains to carry out integrations, first for the coordinates of $\mathbf{w}$ and finally for v. We have

$$\begin{aligned}\int_{\mathbf{w}} \exp\left\{av + 2\mathbf{b}'\mathbf{w} + \frac{1}{v}\mathbf{w}'\tilde{A}_1\mathbf{w}\right\} d\mathbf{w} &= (i\sqrt{\pi})^{q-1}|\tilde{A}_1|^{-1/2}v^{\frac{q-1}{2}}\exp\{v(a - \mathbf{b}'\tilde{A}_1^{-1}\mathbf{b})\}\\ &= (i\sqrt{\pi})^{q-1}|\tilde{A}_1|^{-1/2}v^{\frac{q-1}{2}}\exp\left\{v\frac{|A_1|}{|\tilde{A}_1|}\right\}\end{aligned}$$

(Cauchy's theorem allows to replace every line parallel to the imaginary axis by the imaginary axis itself, then formula (5) of Wishart and Bartlett (1933) is applied) and

$$\frac{1}{2\pi i}\int_v \exp\{v\frac{|A_1|}{|\tilde{A}_1|}\}v^{-k+\frac{q-1}{2}}dv = \left(\frac{|A_1|}{|\tilde{A}_1|}\right)^{k-\frac{q-1}{2}-1}\frac{1}{\Gamma(k - \frac{q-1}{2})},$$

which follows from (2.9). Thus $J(A, q+1, q)$ does not depend on Σ and equals the desired expression; proceeding by induction completes the proof. □

Remark. The previous proof partially rephrases the original derivation of Ingham (1933), where more details are to be found.

Returning back to $C_{(q)}$ and $n/2 = k$, we obtain

$$f(C_{(q)}) = \frac{1}{2^{\beta}\pi^{\frac{(p-q)(q-1)}{2}}(\Gamma(\frac{n-q+1}{2}))^{p-q}\Gamma_q(\frac{n}{2})}$$

$$\times |\Sigma|^{-\frac{n}{2}}\operatorname{etr}\left\{-\frac{1}{2}\Sigma^{-1}C_{(q)}\right\}\frac{\prod_{j=1}^{p-q+1}|C_j|^{\frac{n-q-1}{2}}}{\prod_{j=1}^{p-q}|\tilde{C}_j|^{\frac{n-q}{2}}}, \tag{2.12}$$

where $\beta = p - (p-q)(q-1)\frac{n-q-2}{2} + \frac{q(q-1)}{2} + (p-q+1)\,q\frac{n-q-1}{2}$.

For $p = q$, the expression (2.12) yields the Wishart density. For arbitrary p and $q = 1$, the density is that of a vector of p independent chi-squares (the matrices $\tilde{C}_j$ degenerate and drop out).

3. An application to density estimation and discriminant analysis

The density (2.12) may be used for the construction of an unbiased estimate of the normal density in the case when Σ^{-1} is a band matrix. Unbiased estimates of densities are based on the fact that for a density function g and a sufficient statistic T for the corresponding distribution, an unbiased estimate of $g(x)$ is

$$\hat{g}(\mathbf{x}) = g(\mathbf{x} \mid T) = \frac{g(\mathbf{x}, T)}{g(T)}, \tag{3.1}$$

since obviously

$$\mathrm{E}\hat{g}(\mathbf{x}) = \int g(\mathbf{x} \mid t) g(t)\, \mathrm{d}t = g(\mathbf{x}).$$

Moreover, for exponential-type densities an unbiased estimate expressed in terms of a sufficient statistic is unique and the estimate is the best unbiased estimate (see Vapnik (1982) and the references therein). For the multivariate normal density with general Σ, the estimates have been derived by Lumelskiĭ and Sapozhnikov (1969) (see also Vapnik (1982), Ch. 3, Section 10). Classification rules based on these estimates appear in Abusev and Lumelskiĭ (1980).

Clearly $g(\mathbf{x} \mid T)$ does not depend on the parameters and the estimates can be obtained from the right-hand side of (3.1) with any suitable values of the parameters, e.g. $\boldsymbol{\mu} = 0$ and $\Sigma = I$ for the normal distribution.

Let $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_N$ be a random sample from $N_p(\boldsymbol{\mu}, \Sigma)$ of size $N + 1 = n + 2$ and let Σ^{-1} be a band matrix with $q - 1$ nonzero subdiagonals. The sufficient statistic for the parameters $\boldsymbol{\mu}$, Σ is

$$T = \left(\sum_{i=0}^{N} \mathbf{X}_i, \left(\sum_{i=0}^{N} \mathbf{X}_i \mathbf{X}_i' \right)_{(q)} \right). \tag{3.2}$$

We can now follow the lines of Abusev and Lumelskiĭ (1980) with the sufficient statistic (3.2), i.e. using the density (2.12) with $\Sigma = I$ instead of the Wishart density $W_p(n, I)$.

Proposition. *Let* $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_N$ *and* Σ *be as above. Let* $g(\mathbf{X})$ *be the density of* $N_p(\boldsymbol{\mu}, \Sigma)$, $\bar{\mathbf{X}} = (\mathbf{X}_1 + \ldots + \mathbf{X}_N)/N$ *and* S *the sample covariance matrix*

$$S = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'.$$

Put $\mathbf{d} = (\delta_1, \ldots, \delta_p)' = \mathbf{X}_0 - \bar{\mathbf{X}}$, $\mathbf{d}_j = (\delta_j, \ldots, \delta_{j+q-1})'$, $\tilde{\mathbf{d}}_j = (\delta_{j+1}, \ldots, \delta_{j+q-1})'$. *Finally, let* S_j *and* $\tilde{S}_j$ *be defined from* S *according to* (2.10) *and* (2.11). *Then an unbiased estimate of the density* $g(.)$ *at the point* $\mathbf{X}_0$ *is*

$$\hat{g}(\mathbf{X}_0) = c(N) \frac{\prod_{j=1}^{p-q+1} |S_j|^{-1/2} (1 + \frac{1}{N+1} \mathbf{d}_j' S_j^{-1} \mathbf{d}_j)^{-\frac{N-q-1}{2}}}{\prod_{j=1}^{p-q} |\tilde{S}_j|^{-1/2} (1 + \frac{1}{N+1} \tilde{\mathbf{d}}_j' \tilde{S}_j^{-1} \tilde{\mathbf{d}}_j)^{-\frac{N-q}{2}}},$$

where

$$c(N) = \Big(\frac{N+1}{n^2}\Big)^{p/2} \frac{(\Gamma(\frac{N-q+1}{2}))^{p-q}}{\Big(\Gamma\big(\frac{N-q}{2}\big)\Big)^{p-q+1}} \Gamma(\frac{N}{2}).$$

The estimator $\hat{g}(\mathbf{X}_0)$ may be used for the construction of a classification rule in a usual manner. Suppose we want to decide whether $\mathbf{X}_0$ comes from $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ or $N_p(\boldsymbol{\mu}_2, \Sigma_2)$ and Σ_1^{-1}, Σ_2^{-1} are band matrices. Let the parameters be unknown and let us have a training sample from each of the two populations. Denote by $\hat{g}(\mathbf{X}_0\,;\, \mathbf{d}^1, S^1)$ the estimator $\hat{g}(\mathbf{X}_0)$ computed from $\mathbf{X}_0$ and the training sample from $N_p(\boldsymbol{\mu}_1, \Sigma_1)$, similarly $\hat{g}(\mathbf{X}_0\,;\, \mathbf{d}^2, S^2)$. Then we decide that $\mathbf{X}_0$ comes from $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ if

$$\frac{\hat{g}(\mathbf{X}_0; \mathbf{d}^2, S^2)}{\hat{g}(\mathbf{X}_0; \mathbf{d}^1, S^1)} \leqslant k, \tag{3.3}$$

where k is a suitable constant depending on the prior probabilities and costs of misclassification.

The rule (3.3) with $q = 2$ has been tested on several data. Even if the matrix Σ^{-1} is not a band matrix, the rule may be applicable. Thus, for the famous Fisher Iris data, our rule performed two percent worse than the quadratic discriminant function (the probabilities of misclassification have been estimated by means of the usual one-leaving-out estimate). On the other hand, it seems that a similar phenomenon occurs as in the case of ridge rules; namely, neglecting some non-diagonal elements in the sample covariance matrix has a similar effect as growing its diagonal. In both cases, the resulting rule should be more robust against ill-conditioned covariance matrices. The performance of the rule will be subject to further investigation.

References

[1] *Abusev, R.A. and Lumelskiĭ, Ya.P.*: Unbiased estimators and classification problems for multivariate normal populations. Theor. Prob. Appl. *25* (1980), 381–389.

[2] *Herz, C.S*: Bessel functions of matrix arguments. Ann. Math. *61* (1955), 474–523.

[3] *Ingham, A.E.*: An integral which occurs in statistics. Proc. Cambridge Phil. Soc. *29* (1933), 271–276.

[4] *Lumelskiĭ, Ya.P. and Sapozhnikov, P.N*: Unbiased estimates of density functions. Theor. Prob. Appl. *14* (1969), 357–365.

[5] *Muirhead, R.J.*: Aspects of Multivariate Statistical Theory. Wiley, New York, 1982.

[6] *Rao, C.R.*: Linear Statistical Inference and Its Applications. Wiley, New York, 1965.

[7] *Vapnik, V.N.*: Estimation of Dependences Based on Empirical Data. Springer Verlag, New York, 1982.

[8] *Wishart, J. and Bartlett, M.S.*: The generalised product moment distribution in a normal system. Proc. Cambridge Phil. Soc. *29* (1933), 260–270.

Author's address: *Kryštof Eben*, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Praha 8, Czech Republic;
e-mail **eben@uivt.cas.cz**.