ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

Density expansions for correlations and eigenvalues of the covariance matrix

TÕNU KOLLO AND ANNE SELART

ABSTRACT. In the paper explicit approximation formulae for the sample correlation coefficient and eigenvalues of the sample covariance matrix are found. The approximations are based on univariate and multivariate normal and skew normal distributions. A simulation experiment has been carried out where the empirical distributions are compared with different approximations.

1. Introduction

Density expansions have become one of the main tools when approximating distributions of complicated statistics in the small sample situation. The theory is well developed in the univariate case (see Field & Ronchetti (1990) or Kolassa (1994), for example). Besides normal approximations several other approaches have been worked out that are based on non-symmetric densities like chi-square approximation (Hall, 1983) or gamma-approximation (Gerber (1979), for instance). In the multivariate case most of the existing results are based on the multivariate normal distribution (Barndorff-Nielsen & Cox (1989), Skovgaard (1986), Traat (1986)). So far the density expansions based on the non-symmetric densities are not frequent. We can refer here only to Tan (1979), (1980) and Kollo & von Rosen (1995) for Wishart approximations and Gupta & Kollo (2003) for the skew normal density expansions. In this paper results of Gupta & Kollo (2003) are developed and applied in two special cases: for eigenvalues of the sample covariance matrix and for an element of the sample correlation matrix. It is known that empirical distributions of these statistics are skewed and

Received October 20, 2003.

2000 *Mathematics Subject Classification.* 62E20, 62H10, 65C60.

Key words and phrases. Skew normal distribution, multivariate density expansion, multivariate cumulants, multivariate Hermite polynomials.

The authors are thankful for the financial support of Estonian Science Foundation Grant 4366 and Target Financed Project 0181776s01.

therefore it would be natural to use also a skewed distribution for the approximation. In our case the multivariate skew normal density is used in this role. The obtained skew normal approximations are compared with the corresponding empirical distributions and with the usual Edgeworth type expansions which are based on normal distribution. Both univariate and multivariate approximations are considered. The used formulae are based on a general relation between two density functions presented in Kollo & von Rosen (1998).

The structure of the paper is the following. In Section 2 basic notation and notions are presented and in Section 3 the necessary results on multivariate skew normal distribution are given. In Section 4 different approximations for eigenvalues of the sample covariance matrix and for an element of the sample correlation matrix are found and compared with the empirical distributions found from the simulation experiment. In Appendix the first three cumulants of the sample correlation matrix have been derived.

2. Basic notions

In the following bold capital letters $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ stand for random vectors and lower-case bold letters $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ for their realizations. Lower-case letters $\mathbf{t}$, $\mathbf{u}$, $\mathbf{v}$,... are used for arbitrary constant vectors, and constant matrices are denoted by capital letters A, B, C, ... Notation A^t stands for the transposed matrix A. If $\mathbf{X}$ is a continuous random p-vector then the density function of $\mathbf{X}$ is denoted by $f_{\mathbf{X}}(\mathbf{x})$, for $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ the density function is $\phi_p(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$.

Let $\mathbf{X}$ be a continuous random p-vector. The characteristic function $\varphi_{\mathbf{X}}(\mathbf{t})$ of $\mathbf{X}$ is defined as the expectation:

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E\exp(i\,\mathbf{t}^t\mathbf{X}) = \int_{\mathbf{R}^p} e^{i\,\mathbf{t}^t\mathbf{x}} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}, \quad \mathbf{t} \in \mathbb{R}^p,$$

and the cumulant function $\psi_{\mathbf{X}}(\mathbf{t})$:

$$\psi_{\mathbf{X}}(\mathbf{t}) = \ln \varphi_{\mathbf{X}}(\mathbf{t}).$$

Moments $m_k(\mathbf{X})$ of the random vector $\mathbf{X}$ are defined as the derivatives

$$m_k(\mathbf{X}) = \frac{1}{i^k}\frac{d^k\varphi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}^k}\bigg|_{\mathbf{t}=\mathbf{0}}, \quad k = 1, 2, \ldots,$$

and central moments $\overline{m}_k(\mathbf{X})$ as

$$\overline{m}_k(\mathbf{X}) = \frac{1}{i^k}\frac{d^k\varphi_{\mathbf{X}-E\mathbf{X}}(\mathbf{t})}{d\mathbf{t}^k}\bigg|_{\mathbf{t}=\mathbf{0}}, \quad k = 1, 2, \ldots.$$

Cumulants $c_k(\mathbf{X})$ are defined as the derivatives

$$c_k(\mathbf{X}) = \frac{1}{i^k}\frac{d^k\psi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}^k}\bigg|_{\mathbf{t}=\mathbf{0}}, \quad k = 1, 2, \ldots. \tag{2.1}$$

For derivation of cumulants we need the notion of matrix derivative. If the elements of an $r \times s$ matrix Y are functions of a $p \times q$ matrix X, the matrix derivative of Y by X is a $pq \times rs$ partitioned matrix $\dfrac{dY}{dX}$ which is defined by the following equality:

$$\frac{dY}{dX} = \frac{\partial}{\partial \mathrm{vec}\, X} \mathrm{vec}^t Y, \tag{2.2}$$

where

$$\frac{\partial}{\partial \mathrm{vec}\, X} = \left(\frac{\partial}{\partial x_{11}}, \dots, \frac{\partial}{\partial x_{p1}}, \frac{\partial}{\partial x_{12}}, \dots, \frac{\partial}{\partial x_{p2}}, \dots, \frac{\partial}{\partial x_{1q}}, \dots, \frac{\partial}{\partial x_{pq}} \right)^t.$$

Higher order matrix derivatives are defined recursively:

$$\frac{d^k Y}{dX^k} = \frac{d}{dX} \left(\frac{d^{k-1} Y}{dX^{k-1}} \right).$$

In the following text the derivative $\dfrac{d^k Y}{dX^k}$ is also denoted by $Y^{(k)}(X)$. For the matrix derivative and related matrix algebra (commutation matrix $K_{p,q}$, vec-operator etc.), see Magnus & Neudecker (1999) or Kollo (1991).

3. Multivariate skew normal distribution

The skew normal distribution is an extension of the normal distribution. The symmetry of the normal distribution is distorted with one extra parameter called the shape parameter. In the case of a scalar random variable the sign of the shape parameter indicates whether the heavy tail is in direction of negative or positive values. Shape parameter value 0 corresponds to the normal distribution. The scalar version of the skew normal distribution was first introduced by Azzalini (1985) and generalized to the multivariate case in Azzalini & Dalla Valle (1996) and Azzalini & Capitanio (1999). We shall follow Gupta & Kollo (2003) in the presentation of the multivariate skew normal distribution.

Random p-vector $\mathbf{Z}$ is said to have multivariate skew normal distribution with parameters $\boldsymbol{\mu}$, Σ and $\boldsymbol{\alpha}$, if the density function of $\mathbf{Z}$ is of the form

$$f_{\mathbf{Z}}(\mathbf{z}) = 2\phi_p(\mathbf{z}\,; \boldsymbol{\mu}, \Sigma)\Phi(\boldsymbol{\alpha}^t(\mathbf{z} - \boldsymbol{\mu})), \qquad \mathbf{z} \in \mathbb{R}^p,$$

where $\phi_p(\mathbf{z}; \boldsymbol{\mu}, \Sigma)$ is the density function of the normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$, $\Phi(\cdot)$ is the distribution function of the standard normal distribution $N(0,1)$ and $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ are constant p-vectors.

The constant vector $\boldsymbol{\mu} : p \times 1$ is called the location parameter and $\boldsymbol{\alpha} : p \times 1$ is called the shape parameter. If the distribution of $\mathbf{Z} : p \times 1$ is skew normal with parameters $\boldsymbol{\mu}$, Σ and $\boldsymbol{\alpha}$, we use the following notation: $\mathbf{Z} \sim SN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$. If the location parameter $\boldsymbol{\mu} = \mathbf{0}$ then instead $SN_p(\mathbf{0}, \Sigma, \boldsymbol{\alpha})$ the notation $SN_p(\Sigma, \boldsymbol{\alpha})$ is used.

Skew normal distributions share several good properties with normal distributions. For example, the marginal distributions of a skew normal vector are skew normal. Cumulants of skew normal and normal distributions are also similar. As skew normal distributions have one additional parameter, the cumulants have additional terms comparing to cumulants of a normal distribution as can be seen from the following table (Gupta & Kollo, 2003).

Table 1: Cumulants of the normal and skew normal distributions

$\mathbf{Z} \sim SN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$	$\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$
$c_1(\mathbf{Z}) = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\nu}$	$c_1(\mathbf{X}) = \boldsymbol{\mu}$
$c_2(\mathbf{Z}) = \Sigma - \frac{2}{\pi}\boldsymbol{\nu}\boldsymbol{\nu}^t$	$c_2(\mathbf{X}) = \Sigma$
$c_3(\mathbf{Z}) = \sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\boldsymbol{\nu}(\boldsymbol{\nu})^{t\otimes 2}$	$c_3(\mathbf{X}) = 0$

In Table 1 we use notation

$$\boldsymbol{\nu} = \frac{\Sigma\boldsymbol{\alpha}}{\sqrt{1 + \boldsymbol{\alpha}^t\Sigma^{-1}\boldsymbol{\alpha}}}.$$

The derivatives of the skew normal and the normal density functions have again some similarities. Both of them can be expressed through the density function itself and multivariate Hermite polynomials. The multivariate Hermite polynomial of order k is defined by the equality

$$H_k(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (-1)^k \frac{\phi_p^{(k)}(\mathbf{x}; \boldsymbol{\mu}, \Sigma)}{\phi_p(\mathbf{x}; \boldsymbol{\mu}, \Sigma)},$$

where $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and $\phi_p^{(k)}(\mathbf{x}; \boldsymbol{\mu}, \Sigma)$ stands for the k-th derivative of $N_p(\boldsymbol{\mu}, \Sigma)$ density function. The expressions of a first three multivariate Hermite polynomials are of the following form (Kollo, 1991, p. 141):

$$\begin{aligned}
H_1(\mathbf{z}; \boldsymbol{\mu}, \Sigma) &= \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu}); \\
H_2(\mathbf{z}; \boldsymbol{\mu}, \Sigma) &= \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^t\Sigma^{-1} - \Sigma^{-1}; \\
H_3(\mathbf{z}; \boldsymbol{\mu}, \Sigma) &= \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu})((\mathbf{z} - \boldsymbol{\mu})^t\Sigma^{-1})^{\otimes 2} - \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu}) \operatorname{vec}{}^t\Sigma^{-1} \\
&\quad - \Sigma^{-1} \otimes (\mathbf{z} - \boldsymbol{\mu})^t\Sigma^{-1} - (\mathbf{z} - \boldsymbol{\mu})^t\Sigma^{-1} \otimes \Sigma^{-1}.
\end{aligned}$$

From Gupta & Kollo (2003) we get the expressions of the first derivatives of a skew normal density. Let $\mathbf{Z} \sim SN_k(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$. Then the first three derivatives of the density function $f_{\mathbf{Z}}(\mathbf{z})$ are of the form:

$$
\begin{aligned}
f^{(1)}_{\mathbf{Z}}(\mathbf{z}) &= \left[-H_1(\mathbf{z};\boldsymbol{\mu},\Sigma) + \boldsymbol{\alpha}\frac{\phi(\boldsymbol{\alpha}^t\mathbf{z})}{\Phi(\boldsymbol{\alpha}^t\mathbf{z})}\right] f_{\mathbf{Z}}(\mathbf{z});\\
f^{(2)}_{\mathbf{Z}}(\mathbf{z}) &= \Bigg[H_2(\mathbf{z};\boldsymbol{\mu},\Sigma)\\
&\quad - \frac{\phi(\boldsymbol{\alpha}^t\mathbf{z})}{\Phi(\boldsymbol{\alpha}^t\mathbf{z})}\left(\boldsymbol{\alpha}H_1^t(\mathbf{z};\boldsymbol{\mu},\Sigma) - \boldsymbol{\alpha}h_1(\boldsymbol{\alpha}^t\mathbf{z})\boldsymbol{\alpha}^t + H_1(\mathbf{z};\boldsymbol{\mu},\Sigma)\boldsymbol{\alpha}^t\right)\Bigg] f_{\mathbf{Z}}(\mathbf{z});\\
f^{(3)}_{\mathbf{Z}}(\mathbf{z}) &= \Bigg[-H_3(\mathbf{z};\boldsymbol{\mu},\Sigma) + \frac{\phi(\boldsymbol{\alpha}^t\mathbf{z})}{\Phi(\boldsymbol{\alpha}^t\mathbf{z})}\Big(\boldsymbol{\alpha}\mathrm{vec}^t H_2(\mathbf{z};\boldsymbol{\mu},\Sigma) + H_2(\mathbf{z};\boldsymbol{\mu},\Sigma)\otimes\boldsymbol{\alpha}^t\\
&\quad + \boldsymbol{\alpha}^t\otimes H_2(\mathbf{z};\boldsymbol{\mu},\Sigma) + \boldsymbol{\alpha}h_2(\boldsymbol{\alpha}^t\mathbf{z})(\boldsymbol{\alpha}^t\otimes\boldsymbol{\alpha}^t)\\
&\quad + h_1(\boldsymbol{\alpha}^t\mathbf{z})[\boldsymbol{\alpha}H_1^t(\mathbf{z};\boldsymbol{\mu},\Sigma)\otimes\boldsymbol{\alpha}^t\\
&\quad + H_1(\mathbf{z};\boldsymbol{\mu},\Sigma)(\boldsymbol{\alpha}^t\otimes\boldsymbol{\alpha}^t) + \boldsymbol{\alpha}\boldsymbol{\alpha}^t\otimes H_1^t(\mathbf{z};\boldsymbol{\mu},\Sigma)]\Big)\Bigg] f_{\mathbf{Z}}(\mathbf{z}),
\end{aligned}
\tag{3.1}
$$

where $h_i(\cdot)$ are univariate and $H_i(\mathbf{z};\boldsymbol{\mu},\Sigma)$ multivariate Hermite polynomials, and ϕ, Φ are the density and distribution functions of $N(0,1)$.

4. Expansions for eigenvalues and correlations

A general relation between two density functions is given in the paper Kollo & von Rosen (1998) as Corollary 3.1. An unknown density is approximated by a formal density expansion which includes cumulants of unknown density and cumulants and derivatives of a known density function.

We are going to approximate first the density function of the statistic

$$Y_i = \sqrt{n}(d_i - \delta_i),$$

where δ_i is the i-th eigenvalue of the population covariance matrix Σ and d_i is the i-th eigenvalue of a sample covariance matrix S. From simulation experiments it is known that empirical distribution of this statistic is skewed. If the sample is from the normal population $N_r(\boldsymbol{\mu},\Sigma)$, then the cumulants of Y_i can be expressed as follows (Siotani, Hayakawa & Fujikoshi (1985), p. 454):

$$
\begin{aligned}
c_1(Y_i) &= \frac{a_i}{\sqrt{n}} + \mathcal{O}(n^{-\frac{3}{2}});\\
c_2(Y_i) &= 2\delta_i^2 + \frac{2b_i}{n} + \mathcal{O}(n^{-2});\\
c_3(Y_i) &= \frac{8\delta_i^3}{\sqrt{n}} + \mathcal{O}(n^{-\frac{3}{2}});\\
c_k(Y_i) &= o(n^{-\frac{1}{2}}), \quad k \geq 4,
\end{aligned}
$$

where

$$a_i = \sum_{k\neq i}^{p} \frac{\delta_i\delta_k}{\delta_i - \delta_k}; \quad b_i = \sum_{k\neq i}^{p} \frac{\delta_i^2\delta_k^2}{(\delta_i - \delta_k)^2}. \tag{4.1}$$

Set $i = 1$, then $Y = Y_1$ is determined by the largest eigenvalues of the population and the sample covariance matrices.

Our approximations will be based on skew normal and normal distributions. Skew normal distribution can be used for approximation if we consider the shape parameter $\boldsymbol{\alpha}$ dependent on the sample size n in the following way

$$\boldsymbol{\alpha} = \frac{\boldsymbol{\alpha}^\star}{\sqrt{n}}, \tag{4.2}$$

where $\boldsymbol{\alpha}^\star = \mathcal{O}(1)$. In this case the terms in derivatives of a skew normal density function $f_{\mathbf{Z}}$ are of diminishing order (Gupta & Kollo, 2003)

$$f^{(1)}_{\mathbf{Z}}(\mathbf{z}) = \mathcal{O}(1) + \mathcal{O}(n^{-\frac{1}{2}});$$
$$f^{(2)}_{\mathbf{Z}}(\mathbf{z}) = \mathcal{O}(1) + \mathcal{O}(n^{-\frac{1}{2}}) + \mathcal{O}(n^{-1});$$
$$f^{(3)}_{\mathbf{Z}}(\mathbf{z}) = \mathcal{O}(1) + \mathcal{O}(n^{-\frac{1}{2}}) + \mathcal{O}(n^{-1}) + \mathcal{O}(n^{-3/2})$$

and further

$$f^{(k)}_{\mathbf{Z}}(\mathbf{z}) = \mathcal{O}(1) + o(1), \quad k \geq 4.$$

In the expansions we shall exclude later on these terms of $f^{(i)}_{\mathbf{Z}}$ $(i = 1, 2, 3)$ which are of higher order.

From expression (4.2) we get that $\boldsymbol{\alpha} = \mathcal{O}(n^{-\frac{1}{2}})$ and

$$\frac{1}{\sqrt{1 + \boldsymbol{\alpha}^t\Sigma\boldsymbol{\alpha}}} = \mathcal{O}(1).$$

Thus the cumulants of $\mathbf{Z} \sim SN_r(\boldsymbol{\mu}, \Sigma, \boldsymbol{\alpha})$ can be expressed as

$$c_1(\mathbf{Z}) = \boldsymbol{\mu} + \frac{1}{\sqrt{n}}\sqrt{\frac{2}{\pi}}\frac{\Sigma\boldsymbol{\alpha}^\star}{\sqrt{1 + \frac{1}{n}\boldsymbol{\alpha}^{\star t}\Sigma\boldsymbol{\alpha}^\star}} = \boldsymbol{\mu} + \frac{1}{\sqrt{n}}A_1(\mathbf{Z});$$

$$c_2(\mathbf{Z}) = \Sigma - \frac{1}{n}\frac{2}{\pi}\frac{\Sigma\boldsymbol{\alpha}^\star\boldsymbol{\alpha}^{\star t}\Sigma}{1 + \frac{1}{n}\boldsymbol{\alpha}^{\star t}\Sigma\boldsymbol{\alpha}^\star} = \Sigma - \frac{1}{n}A_2(\mathbf{Z});$$

$$c_3(\mathbf{Z}) = \frac{1}{n^{\frac{3}{2}}}\sqrt{\frac{2}{\pi}}\left(\frac{4}{\pi} - 1\right)\frac{\Sigma\boldsymbol{\alpha}^\star(\boldsymbol{\alpha}^{\star t}\Sigma)^{\otimes 2}}{\left(1 + \frac{1}{n}\boldsymbol{\alpha}^{\star t}\Sigma\boldsymbol{\alpha}^\star\right)^{\frac{3}{2}}} = \frac{1}{n^{\frac{3}{2}}}A_3(\mathbf{Z}),$$

where $A_1(\mathbf{Z})$, $A_2(\mathbf{Z})$ and $A_3(\mathbf{Z})$ are of order $\mathcal{O}(1)$. From here we get the order of the cumulants: $c_1(\mathbf{Z}) = \mathcal{O}(1) + \mathcal{O}(n^{-\frac{1}{2}})$, $c_2(\mathbf{Z}) = \mathcal{O}(1) + \mathcal{O}(n^{-1})$

and $c_3(\mathbf{Z}) = \mathcal{O}(n^{-\frac{3}{2}})$. By Gupta & Kollo (2003)

$$c_k(\mathbf{Z}) = \mathcal{O}(n^{-\frac{k}{2}}), \quad k \geq 4.$$

From the text above it can be concluded that in expansions the terms which include differences of cumulants and derivatives of $f_{\mathbf{Z}}$ are of diminishing order and the expansion converges if the term which includes the difference of the second cumulants equals 0. The similar convergence holds if we use normal density $f_{\mathbf{X}}$ instead of skew normal. Then $c_2(\mathbf{X}) = \mathcal{O}(1)$ and $c_k(\mathbf{X}) = 0$ if $k \geq 3$.

We are going to examine closer four expansions. The first two are based on Theorem 4.1 and the other two are based on Theorem 4.2 from article Kollo & von Rosen (1998). These two theorems are presented jointly in the following statement.

Theorem. *Let $\mathbf{Y}$ be a random p-vector and let $\mathbf{X}$ be a random r-vector, $p \leq r$. Let $D\mathbf{X}$ be a non-singular matrix with different eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_r$.*

(i) *Let $D\mathbf{Y}$ be non-singular and let $(D\mathbf{Y})^{\frac{1}{2}}$ be any square root of $D\mathbf{Y}$ so that $(D\mathbf{Y})^{\frac{1}{2}}(D\mathbf{Y})^{\frac{1}{2}} = D\mathbf{Y}$. Then*

$$f_{\mathbf{Y}}(\mathbf{y}_0) = |D\mathbf{X}|^{\frac{1}{2}}|D\mathbf{Y}|^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}(r-p)} \times \left\{ f_{\mathbf{X}}(\mathbf{x}_0) - \frac{1}{6}\mathrm{vec}^t\big(P^t c_3(\mathbf{Y})P^{\otimes 2} - c_3(\mathbf{X})\big)\mathrm{vec}\, f_{\mathbf{X}}^{(3)}(\mathbf{x}_0) + r_3 \right\},$$

where $\mathbf{x}_0 = P^t\mathbf{y}_0 - P^tE\mathbf{Y} + E\mathbf{X}$, $P = (D\mathbf{Y})^{-\frac{1}{2}}V^t$, $V = (\mathbf{v}_1, \ldots, \mathbf{v}_p)$ is an $r \times p$ matrix, with columns being the first p eigenvalue-normed eigenvectors ($\mathbf{v}_i^t\mathbf{v}_i = \lambda_i$) of $D\mathbf{X}$, and r_3 denotes the remainder term.

(ii) *Let $m_2(\mathbf{Y})$ be non-singular and let $(m_2(\mathbf{Y}))^{\frac{1}{2}}$ be any square root of $m_2(\mathbf{Y})$ so that $(m_2(\mathbf{Y}))^{\frac{1}{2}}(m_2(\mathbf{Y}))^{\frac{1}{2}} = m_2(\mathbf{Y})$. If $E\mathbf{X} = \mathbf{0}$, then*

$$f_{\mathbf{Y}}(\mathbf{y}_0) = |D\mathbf{X}|^{\frac{1}{2}}|m_2(\mathbf{Y})|^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}(r-p)} \times \Big\{ f_{\mathbf{X}}(\mathbf{x}_0) - E\mathbf{Y}^tP\mathrm{vec}\, f_{\mathbf{X}}^{(1)}(\mathbf{x}_0) - \frac{1}{6}\left[\mathrm{vec}^t(P^t c_3(\mathbf{Y})P^{\otimes 2} - c_3(\mathbf{X}))\right]\mathrm{vec}\, f_{\mathbf{X}}^{(3)}(\mathbf{x}_0) + r_3 \Big\},$$

where $\mathbf{x}_0 = P^t\mathbf{y}$, $P = (m_2(\mathbf{Y}))^{-\frac{1}{2}}V^t$, $V = (\mathbf{v}_1, \ldots, \mathbf{v}_p)$ is an $r \times p$ matrix, with columns being the first p eigenvalue-normed eigenvectors ($\mathbf{v}_i^t\mathbf{v}_i = \lambda_i$) of $m_2(\mathbf{X})$, and r_3 denotes the remainder term.

From part (i) of the previous theorem we get the following normal approximation:

$$\begin{aligned} f_Y(y_0) &= |\Sigma|^{\frac{1}{2}}(2\delta_1^2)^{-\frac{1}{2}}(2\pi)^{r-1}\phi_r(\mathbf{x}_0;\Sigma,\boldsymbol{\mu}) \\ &\times \left\{1+\frac{4}{3\sqrt{n}}\mathrm{vec}^t(\mathbf{v}_1(\mathbf{v}_1^t)^{\otimes 2})\mathrm{vec}\, H_3(\mathbf{x}_0;\Sigma,\boldsymbol{\mu})+r_3\right\}, \end{aligned} \tag{4.3}$$

where $\mathbf{x}_0=\mathbf{v}_1(2\delta_1^2)^{-\frac{1}{2}}(y_0-\frac{1}{n}a_1)+\boldsymbol{\mu}$, $\mathbf{v}_1$ is the eigenvector of Σ corresponding to the largest eigenvalue δ_1 and a_1 is given in (4.1). The second expansion is based on a skew normal distribution

$$\begin{aligned} f_Y(y_0) &= |\Sigma|^{\frac{1}{2}}(2\delta_1^2)^{-\frac{1}{2}}(2\pi)^{r-1}f_{\mathbf{Z}}(\mathbf{x}_0) \\ &\times \left\{1+\frac{1}{6\sqrt{n}}\mathrm{vec}^t\big(2\sqrt{2}\mathbf{v}_1(\mathbf{v}_1^t)^{\otimes 2}-\frac{1}{n}A_3(\mathbf{Z})\big)\mathrm{vec}\, H_3(\mathbf{x}_0;\Sigma)+r_3\right\}, \end{aligned} \tag{4.4}$$

where $\mathbf{x}_0=\mathbf{v}_1(2\delta_1^2)^{-\frac{1}{2}}(y_0-n^{-\frac{1}{2}}a_1)+A_1(\mathbf{Z})$ and $\mathbf{Z}\sim SN_r(\Sigma,\boldsymbol{\alpha})$. To determine the parameter $\boldsymbol{\alpha}$ we assume that the moments of P^tY and $\mathbf{Z}$ are close. Set $E\mathbf{Z}=P^tEY$ or

$$\sqrt{\frac{2}{\pi}}\Sigma\boldsymbol{\alpha}\approx P^tE\mathbf{Y},$$

from where we get

$$\boldsymbol{\alpha}\approx\Sigma^{-1}\mathbf{v}_1(2\delta_1^2)^{-\frac{1}{2}}a_1\sqrt{\frac{\pi}{2}}n^{-\frac{1}{2}}.$$

As the third cumulant $c_3(\mathbf{Z})$ is of order $\mathcal{O}(n^{-\frac{3}{2}})$ it could be excluded from the expansions but we shall keep it to include additional skewness term into the approximations.

In the univariate case we get from part (ii) of Theorem the following two expansions: the first one is based on the normal distribution

$$f_Y(y_0)=\phi(y_0;0,2\delta_1^2)\left\{1+\frac{a_1}{\sqrt{n}}h_1(y_0;2\delta_1^2)+\frac{4\delta_1^3}{3\sqrt{n}}h_3(y_0;2\delta_1^2)+r_3\right\} \tag{4.5}$$

and the other one is based on the skew normal distribution

$$\begin{aligned} f_Y(y_0) = f_Z(y_0)\Big\{1 &+ \frac{a_1}{\sqrt{n}}h_1(y_0;2\delta_1^2,\mu) \\ &+ \frac{1}{\sqrt{n}}\left(\frac{4\delta_1^3}{3}-\frac{1}{n}A_3(Z)\right)h_3(y_0;2\delta_1^2,\mu)+r_3\Big\}. \end{aligned} \tag{4.6}$$

If we assume $EZ=0$ as in Theorem (ii), we must use distribution $SN(\mu,\sigma^2,\alpha)$ (Azzalini, 1985). In the case of $SN(\sigma^2,\alpha)$ the condition $EZ=0$ is satisfied only if $\alpha=0$ i.e. the normal distribution. To determine μ and α we choose again

$$\alpha\approx\sigma^{-1}a_1\sqrt{\frac{\pi}{2}}n^{-\frac{1}{2}}$$

and $\mu = a_1 n^{-\frac{1}{2}} = -EY$, to have the condition $EZ = 0$ satisfied. The parameter σ^2 is chosen to be $2\delta_1^2$ and the obtained result is expansion (4.6).

To study these expansions numerically we consider the normal population $N_3(\boldsymbol{\mu}, \Sigma)$, where

$$\Sigma = \begin{pmatrix} 0.9 & 0.2 & 2.0 \\ 0.2 & 3.0 & 0.5 \\ 2.0 & 0.5 & 5.0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}. \tag{4.7}$$

To find the empirical density of $Y = \sqrt{n}(d_1 - \delta_1)$ we generated 2000 samples with sample size 9.

As can be seen from Figure 1 the approximation (4.3) gives the best fit in the middle part and on the left tail of the empirical distribution, approximations (4.4)–(4.6) have too heavy tails. At the same time it is difficult to decide which approximation is the best on the right tail.

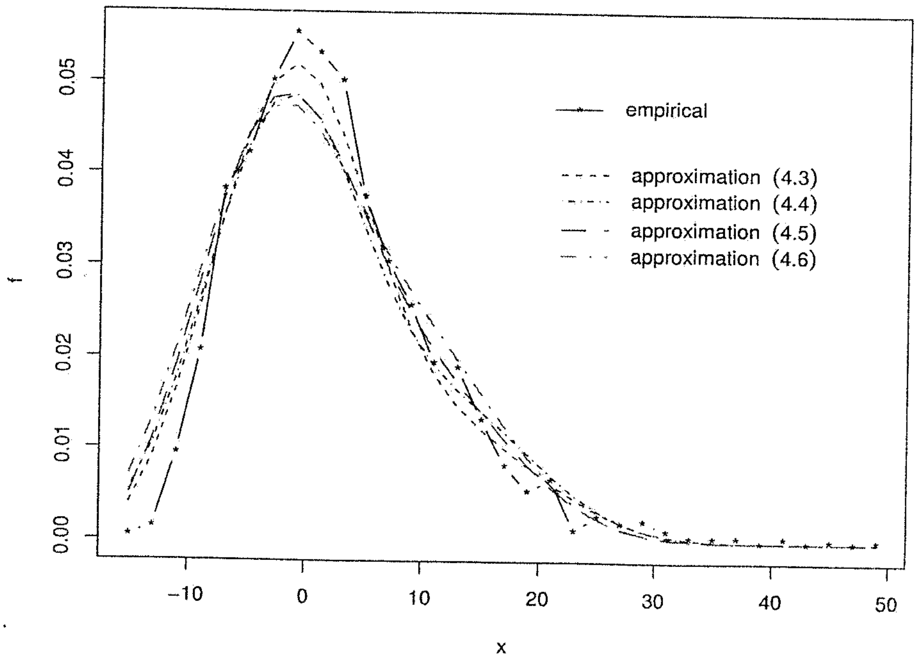

Figure 1. Density approximations for $Y = \sqrt{n}(d_1 - \delta_1)$, sample size $n = 9$.

Another statistic of interest is $Y_{12} = \sqrt{n}(r_{12} - \omega_{12})$, where r_{12} and ω_{12} are elements of the sample correlation matrix $R = S_d^{-\frac{1}{2}} S S_d^{-\frac{1}{2}}$ and population correlation matrix Ω, respectively. At first we need the cumulants of

$\mathbf{Y} = \sqrt{n}\mathrm{vec}\,(R - \Omega)$. The expressions of the first three cumulants of $\mathrm{vec}\,R$ are found in Appendix. If the population distribution is normal $N(\boldsymbol{\mu}, \Sigma)$ then (see Kollo, 1991, p. 106, for example)

$$\overline{m}_2(\mathrm{vec}\,S) = \frac{1}{n}\Pi_N = \frac{1}{n}(I_{r^2} + K_{r,r})(\Sigma \otimes \Sigma)$$

and

$$c_1(\mathbf{Y}) = \frac{1}{2\sqrt{n}}(I_{r^2} \otimes \mathrm{vec}\,{}^t\Pi_N)K_{r^4,r^2}\mathrm{vec}\,D_2;$$

$$c_2(\mathbf{Y}) = D_1\Pi_N D_1^t - \frac{1}{4n}(I_{r^2} \otimes \mathrm{vec}\,{}^t\Pi_N)K_{r^4,r^2}\mathrm{vec}\,D_2\mathrm{vec}\,{}^tD_2K_{r^2,r^4}(\mathrm{vec}\,\Pi_N \otimes I_{r^2});$$

$$c_3(\mathbf{Y}) = \frac{1}{2\sqrt{n}}\Big[- (I_{r^2} \otimes \mathrm{vec}\,{}^t\Pi_N)K_{r^4,r^2}\mathrm{vec}\,D_2\mathrm{vec}\,{}^tD_1\Pi_N D_1^t - (\mathrm{vec}\,{}^tD_2K_{r^2,r^4}(\mathrm{vec}\,\Pi_N \otimes I_{r^2}) \otimes D_1\Pi_N D_1^t)(I_{r^2} + K_{r^2,r^2}) + \frac{1}{2n}(I_{r^2} \otimes \mathrm{vec}\,{}^t\Pi_N)K_{r^4,r^2}\mathrm{vec}\,D_2(\mathrm{vec}\,{}^tD_2K_{r^2,r^4}(\mathrm{vec}\,\Pi_N \otimes I_{r^2}))^{\otimes 2}\Big],$$

where $D_1 = \left(\dfrac{dR}{dS}\right)^t_{S=\Sigma}$ and $D_2 = \left(\dfrac{d^2R}{dS^2}\right)^t_{S=\Sigma}$.

To find the cumulants of Y_{12} we can consider $Y_{12} = \mathbf{e}_2^t\mathbf{Y}$, where $\mathbf{e}_2$ is an r^2-dimensional basis vector, so that all other elements of $\mathbf{Y}$ are eliminated except Y_{12}. Hence $c_1(Y_{12}) = \mathbf{e}_2^tc_1(\mathbf{Y})$, $c_2(Y_{12}) = \mathbf{e}_2^tc_2(\mathbf{Y})\mathbf{e}_2$ and $c_3(Y_{12}) = \mathbf{e}_2^tc_3(\mathbf{Y})\mathbf{e}_2^{\otimes 2}$. The cumulants can be expressed in a following way:

$$c_1(Y_{12}) = \mathbf{e}_2^tc_1(\mathbf{Y}) = \frac{1}{\sqrt{n}}C_1;$$

$$c_2(Y_{12}) = \mathbf{e}_2^tc_2(\mathbf{Y})\mathbf{e}_2 = C + \frac{1}{n}C_2;$$

$$c_3(Y_{12}) = \mathbf{e}_2^tc_3(\mathbf{Y})\mathbf{e}_2^{\otimes 2} = \frac{1}{\sqrt{n}}C_3 + o(n^{-1}),$$

where C and C_i $(i = 1, 2, 3)$ do not depend on the sample size n. Analogously to the approximations (4.3)–(4.6) we get the following four expansions. The first one is based on the multivariate normal distribution

$$f_{Y_{12}}(y_0) = |\Sigma|^{\frac{1}{2}}C^{-\frac{1}{2}}(2\pi)^{r-1}\phi_r(\mathbf{x}_0; \Sigma, \boldsymbol{\mu}) \times \left\{1 + \frac{1}{6\sqrt{n}}\mathrm{vec}\left(\mathbf{v}_1C_3C^{-3/2}\mathbf{v}_1^{t\otimes 2}\right)\mathrm{vec}\,H_3(\mathbf{x}_0; \Sigma, \boldsymbol{\mu}) + r_3\right\}, \tag{4.8}$$

where $\mathbf{x}_0 = \mathbf{v}_1 C^{-\frac{1}{2}}(y_0 - \frac{1}{n}C_1) + \boldsymbol{\mu}$ and V_1 is the eigenvector of Σ which corresponds to its largest eigenvalue δ_1. The second is based on the multivariate skew normal distribution

$$
\begin{aligned}
f_{Y_{12}}(y_0) &= |\Sigma|^{\frac{1}{2}} C^{-\frac{1}{2}} (2\pi)^{r-1} f_{\mathbf{Z}}(\mathbf{x}_0) \\
&\times \left\{ 1 + \frac{1}{6\sqrt{n}} \operatorname{vec}^t \left(\mathbf{v}_1 C_3 C^{-3/2} (\mathbf{v}_1^t)^{\otimes 2} - \frac{1}{n} A_3(\mathbf{Z}) \right) \operatorname{vec} H_3(\mathbf{x}_0; \Sigma) + r_3 \right\},
\end{aligned} \tag{4.9}
$$

where $\mathbf{x}_0 = \mathbf{v}_1 C^{-\frac{1}{2}}(y_0 - n^{-\frac{1}{2}} C_1) + A_1(\mathbf{Z})$ and $\mathbf{Z} \sim SN_r(\Sigma, \boldsymbol{\alpha})$. Here $\boldsymbol{\alpha} \approx \Sigma^{-1} V_1 C^{-\frac{1}{2}} C_1 \sqrt{\frac{\pi}{2}} n^{-\frac{1}{2}}$. The third is based on the univariate normal distribution

$$
f_{Y_{12}}(y_0) = \phi(y_0; 0, C) \left\{ 1 + \frac{C_1}{\sqrt{n}} h_1(y_0; C) + \frac{1}{6\sqrt{n}} C_3 h_3(y_0; C) + r_3 \right\}. \tag{4.10}
$$

The fourth approximation is based on the univariate skew normal distribution, where again $\alpha \approx \sigma^{-1} C_1 \sqrt{\frac{\pi}{2}} n^{-\frac{1}{2}}$ and $\mu = C_1 n^{-\frac{1}{2}} = -EY$, i.e. the condition $EZ = 0$ is satisfied. The parameter σ^2 we choose equal to C:

$$
\begin{aligned}
f_{Y_{12}}(y_0) = f_Z(y_0) \Bigg\{ 1 &+ \frac{C_1}{\sqrt{n}} h_1(y_0; \mu, C) \\
&+ \frac{1}{6\sqrt{n}} \left(C_3 - \frac{1}{n} A_3(Z) \right) h_3(y_0; \mu, C) + r_3 \Bigg\}.
\end{aligned} \tag{4.11}
$$

For a numerical study we use the same population distribution as in the previous example, i.e $N_3(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ and Σ are given by (4.7). The population correlation matrix is in this case

$$
\Omega = \begin{pmatrix} 1 & 0.12 & 0.94 \\ 0.12 & 1 & 0.13 \\ 0.94 & 0.13 & 1 \end{pmatrix}.
$$

In Figure 2 the graphs of all approximations practically coincide and do not give a good fit with the empirical distribution.

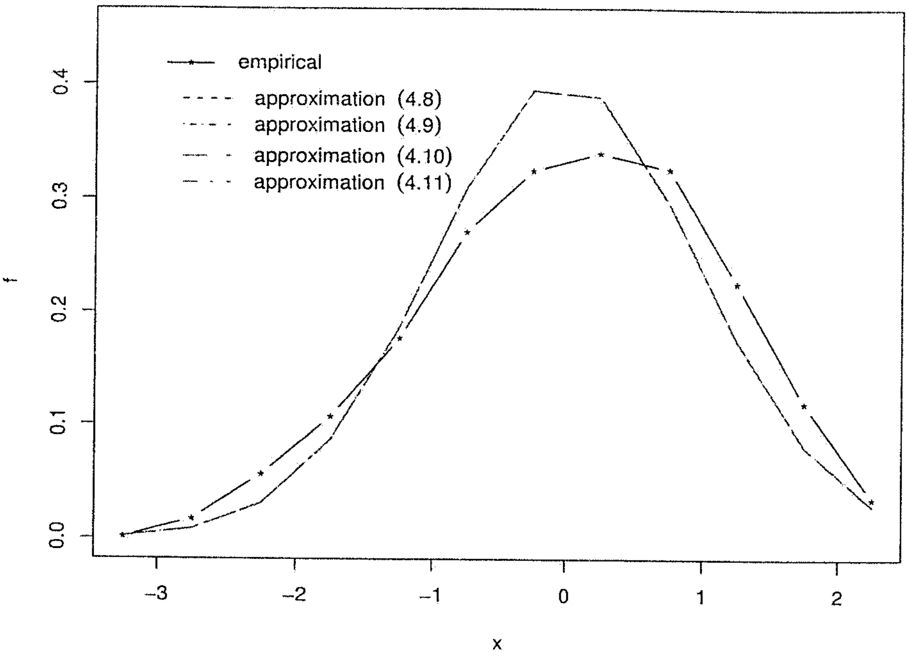

Figure 2. Density approximations for $Y_{12} = \sqrt{n}(r_{12} - \omega_{12})$, sample size $n = 9$.

Appendix

The cumulant function of R is (Kollo & Ruul, 2003)

$$\psi_R(T) = i\mathrm{vec}^t(T)\mathrm{vec}\,\Omega + \ln\{1 - \frac{1}{2}\mathrm{vec}^t(T)M\mathrm{vec}\,(T) + \frac{i}{2}vec^t(T)N\},$$

where

$$M = D_1\overline{m}_2(\mathrm{vec}\,S)D_1^t;$$
$$N = \{I_{r^2} \otimes \mathrm{vec}^t(\overline{m}_2(\mathrm{vec}\,S))\}K_{r^4,r^2}\mathrm{vec}\,D_2$$

and D_1, D_2 are defined as in Section 4. To find the cumulants of $\mathbf{Y} = \sqrt{n}\mathrm{vec}\,(R - \Omega)$ the derivatives of ψ_R have to be found as indicated by formula (2.1). Denote

$$G = \left(1 - \frac{1}{2}\mathrm{vec}^t(T)M\mathrm{vec}\,(T) + \frac{i}{2}vec^t(T)N\right)^{-1};$$
$$H = \frac{i}{2}N - M\mathrm{vec}\,(T).$$

Then using formula (2.2) the first-order derivative is

$$\begin{aligned}\frac{d\psi_R(T)}{dT} =& \operatorname{vec}\Omega i + \frac{dG^{-1}}{dT}G = \operatorname{vec}\Omega i - \frac{1}{2}\big[\big(M\operatorname{vec}(T) + M\operatorname{vec}(T)\big) + iN\big]G \\ =& \operatorname{vec}\Omega i + HG.\end{aligned}$$

From here the derivate at $T = 0$ equals

$$\left.\frac{d\psi_R(T)}{dT}\right|_{T=0} = \operatorname{vec}\Omega i + \frac{i}{2}N$$

and the first cumulant

$$c_1(\operatorname{vec} R) = \operatorname{vec}\Omega + \frac{1}{2}N.$$

For the second-order derivative

$$\frac{d^2\psi_R(T)}{dT^2} = \frac{dH}{dT}(G \otimes I_r^2) + \frac{dG}{dT}H^t$$

we have to find first $\frac{dH}{dT}$ and $\frac{dG}{dT}$:

$$\frac{dH}{dT} = -M^t; \quad \frac{dG}{dT} = (\frac{i}{2}N - M\operatorname{vec}(T))(-1)G^2 = -H\,G^2.$$

If we replace now obtained results into the expression above, we get:

$$\frac{d^2\psi_R(T)}{dT^2} = -M(G \otimes I_r^2) - H\,G^2H^t = -G \otimes M - H\,G^2H^t$$

and so from the second order derivate at $T = 0$

$$\left.\frac{d^2\psi_R(T)}{dT^2}\right|_{T=0} = -M - \frac{i^2}{4}NN^t$$

we get the second cumulant

$$c_2(\operatorname{vec} R) = M - \frac{1}{4}NN^t.$$

Finally the third order derivative and cumulant are as follows:

$$\begin{aligned}\frac{d^3\psi_R(T)}{dT^3} =& -\frac{dG}{dT} \otimes \operatorname{vec}^t M - \frac{dH}{dT}(G^2H^t \otimes I_{r^2}) \\ & - \left(\frac{dG^2}{dT}H^t + \frac{dH^t}{dT}G^2\right)(I_{r^2} \otimes H^t) \\ =& HG^{-2} \otimes \operatorname{vec}^t M + G^2H^t \otimes M + H2G^3(H^t \otimes H^t) + M \otimes G^2H^t;\end{aligned}$$

$$\left.\frac{d^3\psi_R(T)}{dT^3}\right|_{T=0} = \frac{i}{2}N \otimes \operatorname{vec}^t M + \frac{i}{2}N^t \otimes M + \frac{i^3}{4}N(N^t \otimes N^t) + M \otimes \frac{i}{2}N^t;$$

$$c_3(\operatorname{vec} R) = \frac{1}{2}\left[\frac{1}{2}N(N^t)^{\otimes 2} - (N^t \otimes M)(I_{r^2} + K_{r^2,r^2}) - N\operatorname{vec}^t M\right].$$

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **86**, 715–726.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 579–602.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.

Field, C. and Ronchetti, E. (1990). *Small Sample Asymptotics*. Institute of Mathematical Statistics, Hayward, CA.

Gerber, H. U. (1979). *An Introduction to Mathematical Risk Theory*. S. S. Huebner Foundation for Insurance Education, Philadelphia.

Hall, P. (1983). Chi-squared approximations to the distribution of a sum of independent random variables. *Ann. Probab.* **11**, 1028–1036.

Kolassa, J. E. (1994). *Series Approximation Methods in Statistics*. Springer, New York.

Kollo, T. (1991). *Matrix Derivative for Multivariate Statistics*. Tartu University Press, Tartu. (In Russian)

Kollo, T. and Gupta, A. K. (2003). Density expansions based on the multivariate skew normal distribution. *Sankhyā* **65**, 821–835.

Kollo, T. and von Rosen, D. (1995). Approximating by the Wishart distribution. *Ann. Inst. Statist. Math.* **47**, 767–783.

Kollo, T. and von Rosen, D. (1998). A unified approach to the approximation of multivariate densities. *Scand. J. Statist.* **25**, 93–109.

Kollo, T. and Ruul, K. (2003). Approximations to distribution of the sample correlation matrix. *J. Multivariate Anal.* **55**, 318–334.

Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Second edition. Wiley, Chichester.

Siotani, M., Hayakawa, T., Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*. American Sciences Press, Columbus.

Skovgaard, I. (1986). On the multivariate Edgeworth expansions. *International Statistical Review* **54**, 169–186.

Tan, W. Y. (1979). On the approximation of noncentral Wishart distribution by Wishart distribution. *Metron* **37**, 50–58.

Tan, W. Y. (1980). On approximating multivariate distributions. In: *Multivariate Statistical Analysis, Proceedings of the Research Seminar at Dalhouse University*, North Holland, Amsterdam, 237–249.

Traat, I. (1986). Matrix calculus for multivariate distributions. *Tartu Riikl. Ül. Toimetised* **733**, 64–85.

INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITY OF TARTU, J. LIIVI 2, 50409 TARTU, ESTONIA

E-mail address: tonu.kollo@ut.ee

E-mail address: annes@ut.ee