VERSITA

Mathematica Slovaca

DOI: 10.2478/s12175-008-0099-7
Math. Slovaca **58** (2008), No. 5, 635–652

APPROXIMATION OF MULTIVARIATE DISTRIBUTION FUNCTIONS

Margus Pihlak

(*Communicated by Miloslav Duchoň*)

ABSTRACT. In the paper the unknown distribution function is approximated with a known distribution function by means of Taylor expansion. For this approximation a new matrix operation — matrix integral — is introduced and studied in [PIHLAK, M.: *Matrix integral*, Linear Algebra Appl. **388** (2004), 315–325]. The approximation is applied in the bivariate case when the unknown distribution function is approximated with normal distribution function. An example on simulated data is also given.

1. Introduction

Let Y be a random variable with unknown distribution function G and let G_n be the empirical distribution function of Y found from the sample $y_1, y_2, \dots, y_n$. Our aim is to present the unknown distribution function G by means of a known distribution function F. Let F be the distribution function of the random variable X. It is assumed that function F is k times continuously differentiable. Then we can approximate the function G by means of the function F as

$$G(x) \approx \sum_{l=0}^{k} a_l \frac{\mathrm{d}^l F(x)}{\mathrm{d}x^l} \tag{1}$$

where a_l, $l = 1, 2, \dots k$, are coefficients depending on the first l moments of X and Y. The problem is how to determine the coefficients a_l, $l = 1, 2, \dots, k$,

---

2000 Mathematics Subject Classification: Primary 62H10; Secondary 15A69.
Keywords: matrix derivative, matrix integral, Edgeworth type expansion.

in equality (1). The idea of this type of approximation was suggested by R. A. Fisher and E. A. Cornish in [2].

The method of Cornish and Fisher [2] includes the following steps. Assume that random variables X and Y have the moments and cumulants up to sufficiently high order k. Firstly the characteristic function of Y is presented through the characteristic function of X as a Taylor series. Then the inverse Fourier transform is used to get from the Taylor expansion of the characteristic function an approximation of the probability density function of Y through the density function of X.

A stimulating result for our study has been a general relation between two multivariate density functions, derived in Kollo, von Rosen [4]. In their paper a general formal multivariate density expansion is presented where complicated density of interest is presented through the approximating density and cumulants of both distributions under consideration. In applications approximation of the distribution function is at least as important as of the density function. In univariate case an expansion of the distribution function can be obtained from a density expansion by integration. In multivariate case the situation is more complicated. The problem of dimensions have to be solved. Kollo and von Rosen have approximated the density function $f_Y(x)$ by the density function $f_X(x)$ using matrix derivative. The notion "matrix integral" has been introduced by Pihlak [7] to make it possible to integrate this relation between the two density function.

The paper is organized in the following way. In Section 2 we study results of matrix algebra necessary in further study. In Section 3 we introduce and study new operation, matrix integral. In Section 4 we apply matrix integral to integration of the relation between two density functions. In Section 5 we present the relation between unknown distribution function and normal distribution function. In Section 6 we give an example of approximation of simulated data.

2. Preparation

Let us denote matrix X with p rows and q columns by $X : p \times q$. The element of matrix X in the ith row and jth column is denoted by $(X)_{ij}$. A $p \times 1$-matrix is called p-vector. The ith coordinate of the p-vector a is denoted by $(a)_i$.

If we handle partitioned matrix X then it is denoted by $[X_{ij}]$ where X_{ij} denotes the block in ith row and jth column of blocks. We use index pairs for indicating rows and columns of a partitioned matrix following [1]. A row of partitioned matrix A is denoted by index (k,l), i.e. this is the lth row of the kth block-row. A column of partitioned matrix A is denoted by index (g,h), i.e. this is the hth column of the gth block-column. The element of the partitioned matrix A in the (k,l)th row and the (g,h)th column is denoted by $(A)_{(k,l)(g,h)}$. The vectorization operation is denoted by vec. For $X\colon p\times q$ matrix $\mathrm{vec}X\colon pq\times 1$ is the following pq-vector:

$$\mathrm{vec}\,X = ((X)_{11},\dots,(X)_{p1},(X)_{12},\dots,(X)_{p2},\dots(X)_{1q},\dots,(X)_{pq})'.$$

A useful operation in multivariate statistics is the Kronecker product. This operation is denoted by $\otimes$. Let us have the matrices $X\colon p\times q$ and $Y\colon r\times s$. Then the Kronecker product $X\otimes Y$ is the $pr\times qs$-matrix which is partitioned into $r\times s$ blocks:

$$X\otimes Y = [(X)_{lj}Y], \qquad l=1,\dots,p; \quad j=1,\dots,q$$

where

$$(X)_{lj}Y = \begin{pmatrix} (X)_{lj}(Y)_{11} & \cdots & (X)_{lj}(Y)_{1s} \\ \vdots & \ddots & \vdots \\ (X)_{lj}(Y)_{r1} & \cdots & (X)_{lj}(Y)_{rs} \end{pmatrix}.$$

Now we are ready to define the matrix derivative. The notion "matrix derivative" has been used for different representations of the *Frechet*' derivative. In multivariate statistics the definition of N e u d e c k e r [6] is mainly used. Let the elements of the matrix $Y\colon r\times s$ be functions of matrix $X\colon p\times q$. Assume that for all $i=1,\dots,p$, $j=1,\dots,q$, $k=1,\dots,r$ and $l=1,\dots,s$ partial derivatives $\frac{\partial (Y)_{kl}}{\partial (X)_{ij}}$ exist and are continuous in the set A. Then the matrix derivative is defined as follows.

Definition 1. The *derivative of the matrix $Y\colon r\times s$ by the matrix $X\colon p\times q$* is the matrix $\frac{\mathrm{d}Y}{\mathrm{d}X}\colon rs\times pq$ expressed as

$$\frac{\mathrm{d}Y}{\mathrm{d}X} = \frac{\mathrm{d}}{\mathrm{d}\,\mathrm{vec}'\,X}\otimes \mathrm{vec}\,Y$$

where

$$\frac{\mathrm{d}}{\mathrm{d}\,\mathrm{vec}'\,X} = \left(\frac{\partial}{\partial (X)_{11}},\dots,\frac{\partial}{\partial (X)_{p1}},\dots,\frac{\partial}{\partial (X)_{1q}},\dots,\frac{\partial}{\partial (X)_{pq}}\right).$$

There exists also another widely used definition of the matrix derivative. The matrix derivative defined by M a c R a e [5] keeps the structure of involved matrices.

Definition 2. *Derivative of the matrix $Y: r\times s$ by the matrix $X: p\times q$* is a $pr\times qs$-matrix $\frac{\partial Y}{\partial X}$ defined by equality

$$\frac{\partial Y}{\partial X}=\frac{\mathrm{d}}{\mathrm{d}X}\otimes Y$$

where

$$\frac{\mathrm{d}}{\mathrm{d}X}=\begin{pmatrix}\frac{\partial}{\partial (X)_{11}} & \cdots & \frac{\partial}{\partial (X)_{1q}}\\ \vdots & \ddots & \vdots\\ \frac{\partial}{\partial (X)_{p1}} & \cdots & \frac{\partial}{\partial (X)_{pq}}\end{pmatrix}.$$

By means of matrix derivative defined by Definition 1 we can define kth order cumulats of a random vector. Let X be a random p-vector and $t\in\Re^p$. Then the characteristic function of X

$$\varphi_X(t)=E(\mathrm{e}^{\mathrm{i}t'X})$$

and the kth order cumulant of X is the kth order matrix derivative:

$$c_k(X)=\frac{1}{i^k}\frac{\mathrm{d}^k\ln(\varphi_X(t))}{\mathrm{d}t^k}\bigg|_{t=0}.$$

It follows straightforwardly that $c_1(X)=E(X')$ and $c_2(X)=D(X)$.

Let X and Y be random p-vectors with probability density functions $f_X(x)$ and $f_Y(y)$ respectively. Assume that $f_X(x)$ is uniformly continuous and continuously differentiable necessary number of times by argument x. Let us denote the kth order derivative of the function $f_X(x)$ by $f_X^{(k)}(x)$. Then in notations explained above the next equality holds ([4]):

$$\begin{aligned}f_Y(x)=&f_X(x)-(E(Y)-E(X))'\operatorname{vec}f_X^{(1)}(x)\\&+\frac{1}{2}\operatorname{vec}'\{D(Y)-D(X)+(E(Y)-E(X))(E(Y)-E(X))'\}\operatorname{vec}f_X^{(2)}(x)\\&-\frac{1}{6}\operatorname{vec}'\{(c_3(Y)-c_3(X))+3\operatorname{vec}'(D(Y)-D(X))\\&\qquad\otimes(E(Y)-E(X))+(E(Y)-E(X))'^{\otimes 3}\}\operatorname{vec}f_X^{(3)}+\dots.\end{aligned}\tag{2}$$

Our aim is to integrate the expression (2). For this integration matrix integral is introduced and studied.

3. Matrix integral

In this part we introduce the matrix integral as the inverse operation of the matrix derivative. The formal definition of the matrix integral is as follows ([7]).

Definition 3. Let $Z\colon rs \times pq$ be a function of $X\colon p \times q$. A matrix $Y(X)\colon r \times s$ is called the *matrix integral* of $Z = Z(X)\colon rs \times pq$, where $X\colon p \times q$, if

$$\frac{\mathrm{d}Y(X)}{\mathrm{d}X} = Z.$$

The fact, that matrix Y is the matrix integral of a matrix Z is denoted as

$$\int_{\Re^{pq}} Z \circ \mathrm{d}X = Y.$$

If Y is the matrix integral of matrix Z, then also $Y + C$ is a matrix integral of Z, where C is a constant matrix with the same dimensions as matrix Y. Definition 3 is used also to define the definite matrix integral.

Definition 4. A difference $\int_A^B Z \circ \mathrm{d}X = Y(B) - Y(A)$ is called the *definite matrix integral of matrix Z from A to B.*

If the matrix derivative increases the dimensions of the matrix, then the matrix integral decreases the dimensions of the matrix. For decreasing the dimensions of a matrix M a c R a e [5] has introduced the star product of matrices. She has denoted this operation as $*$.

Let us have matrix $A : p \times q$ and partitioned-matrix $B\colon pr \times qs$, consisting of $r \times s$ blocks.

Definition 5. The *star product* $A * B\colon r \times s$ is defined as

$$A * B = \sum_{l=1}^{p} \sum_{j=1}^{q} (A)_{lj} [B_{lj}]$$

where the blocks B_{lj} are $r \times s$-matrices.

By means of the star product we can find the matrix integral in the following way ([7]):

$$Y = \int_{\Re^{pq}} \mathrm{d}X * \frac{\partial Y}{\partial X} = \mathrm{d}X * \left(\frac{\mathrm{d}}{\mathrm{d}X} \otimes Y \right) = \int_{\Re^{pq}} \mathrm{d}Y.$$

Now we present some basic properties of the matrix integral which we apply to integration of the equality (2). Proofs can be found in [7].

Let $g(x)$ be a continuous and k times differentiable function of a p-vector x. Let the derivative of kth order of the function $g(x)$ be denoted by $g^{(k)}(x)$. Denote the inner product of p-vectors a and b as (a, b):

$$(a, b) = a'b.$$

For notational conveniences we introduce the vector

$$1_p = \underbrace{(1, \ldots, 1)'}_{p \text{ times}}.$$

Let us introduce the scalar differential operator

$$\frac{\mathrm{d}}{{}_s\mathrm{d}x} = \sum_{l=1}^{p} \frac{\partial}{\partial (x)_l}$$

and the operator

$${}_s\mathrm{d}x := \sum_{l=1}^{p} d(x)_l.$$

Proposition 1. *The next relation holds in the notations, given above*

$$\int\limits_{\Re_p} g^{(k)}(x) \circ \mathrm{d}x = g^{(k-1)'}(x).$$

Proposition 2. *Let a be a constant not depending on vector x. Then*

$$\int\limits_{\Re^p} \left(\frac{\mathrm{d}}{\mathrm{d}x'}, 1_p \right) a g(x) \circ {}_s\mathrm{d}x = a g(x).$$

Proposition 3. *Let a be a constant not depending on vector x. Then*

$$\int\limits_{\Re^p} \left(\frac{\mathrm{d}^{\otimes k}}{\mathrm{d}x'}, 1_{p^k} \right) a g^{(k)}(x) \circ {}_s\mathrm{d}x = \int\limits_{\Re^p} \left(\frac{\mathrm{d}^{\otimes k-1}}{\mathrm{d}x'}, 1_{p^{k-1}} \right) a g^{(k-1)}(x)$$

where $k = 1, 2, \ldots$.

Proposition 4. *Let functions g and G be such that*

$$g(x) = \frac{\partial^p G(x)}{\partial (x)_1 \ldots \partial (x)_p}.$$

Then the following relation is valid

$$\underbrace{\int \ldots \int}_{p} g(x)\, \mathrm{d}(x)_1 \ldots \mathrm{d}(x)_p = \int\limits_{\Re^p} \frac{\mathrm{d}}{{}_s\mathrm{d}x} G(x) \circ {}_s\mathrm{d}x.$$

Proposition 5. *Let a be a constant p-vector. Then*

$$\underbrace{\int \dots \int}_{p} \left(a, \frac{\mathrm{d}^{\otimes k}}{\mathrm{d}x}\right) g(x)\,\mathrm{d}(x)_1 \dots \mathrm{d}(x)_p = \left(a, \frac{\mathrm{d}^{\otimes k}}{\mathrm{d}x}\right) G(x).$$

4. Application of matrix integral

In this sequel we apply the matrix integral to integration of the relation between multivariate density functions $f_X(x)$ and $f_Y(y)$. Let us write the equality (2) in the form

$$f_Y(x) = f_X(x) - a' \operatorname{vec} f_X^{(1)}(x) + \operatorname{vec}' B \operatorname{vec} f_X^{(2)}(x) - \operatorname{vec}' C \operatorname{vec} f_X^{(3)}(x) + \dots \quad (3)$$

where p-vector,

$$a = (E(X) - E(Y))',$$

$p \times p$-matrix

$$B = \frac{1}{2}(D(Y) - D(X) + (E(Y) - E(X))(E(Y) - E(X))')$$

and $p^2 \times p$-matrix

$$C = \frac{1}{6}((c_3(Y) - c_3(X)) + 3(D(Y) - D(X)) \otimes (E(Y) - E(X))) \\ + (E(Y) - E(X))^{\otimes 2}(E(Y) - E(X))').$$

Matrix C can be also considered as partitioned matrix consisting of p blocks where each block is a $p \times p$-matrix.

Let us introduce the operators

$$\frac{\mathrm{d}}{\mathrm{d}x} := \left(\frac{\partial}{\partial(x)_1}, \frac{\partial}{\partial(x)_2}, \dots, \frac{\partial}{\partial(x)_p}\right)',$$

$$\frac{\mathrm{d}^{\otimes 2}}{\mathrm{d}x} := \left(\frac{\partial^2}{\partial(x)_1^2}, \frac{\partial^2}{\partial(x)_1(x)_2} \dots, \frac{\partial^2}{\partial(x)_p^2}\right)'$$

and

$$\frac{\mathrm{d}^{\otimes 3}}{\mathrm{d}x} := \left(\frac{\partial^3}{\partial(x)_1^3}, \frac{\partial^3}{\partial(x)_1^2(x)_2} \dots, \frac{\partial^3}{\partial(x)_p^3}\right)'.$$

We rewrite the equality (3) in the form

$$f_Y(x) = \left(1 - \left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) + \left(\operatorname{vec} B, \frac{\mathrm{d}^{\otimes 2}}{\mathrm{d}x}\right) - \left(\operatorname{vec} C, \frac{\mathrm{d}^{\otimes 3}}{\mathrm{d}x}\right)\right) f_X(x) + \dots.$$

For distribution functions we formulate the following result.

Theorem 1. *Let X and Y be random vectors. Then between probability distribution functions $F_Y(x)$ and $F_X(x)$ the next relation holds:*

$$F_Y(x) = \left(1 - \left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) + \left(\operatorname{vec} B, \frac{\mathrm{d}^{\otimes 2}}{\mathrm{d}x}\right) - \left(\operatorname{vec} C, \frac{\mathrm{d}^{\otimes 3}}{\mathrm{d}x}\right)\right) F_X(x) + \dots \tag{4}$$

P r o o f. Between multivariate density function $f_X(x)$ and $F_X(x)$ the next relation holds:

$$\frac{\partial^p F_X(x)(x)}{\partial(x)_1 \dots \partial(x)_p} = f_X(x) \tag{5}$$

Using Propositions 4, 5 and equality (5) we get

$$\int_{-\infty}^{(x)_1} \cdots \int_{-\infty}^{(x)_p} f_X(u)\, \mathrm{d}(u)_1 \dots \mathrm{d}(u)_p = \frac{\mathrm{d}}{{}_s\mathrm{d}u} F_X(u) \circ {}_s\mathrm{d}u = F_X(x)$$

and

$$\int_{-\infty}^{(x)_1} \cdots \int_{-\infty}^{(x)_p} \left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) f_X(u)\, \mathrm{d}(u)_1 \dots \mathrm{d}(u)_p$$
$$= \left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) \int\limits_{(-\infty,(x)_1)\times\ldots\times(-\infty,(x)_p)} \frac{\mathrm{d}}{{}_s\mathrm{d}u} F_X(u) \circ {}_s\mathrm{d}u = \left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) F_X(x).$$

Applying the same principles to $\left(\operatorname{vec}' B, \frac{\mathrm{d}^{\otimes 2}}{\mathrm{d}x}\right) f_X(x)$ and $\left(\operatorname{vec}' C, \frac{\mathrm{d}^{\otimes 3}}{\mathrm{d}x}\right) f_X(x)$ we get

$$\int_{-\infty}^{(x)_1} \cdots \int_{-\infty}^{(x)_p} \left(\operatorname{vec}' B, \frac{\mathrm{d}^{\otimes 2}}{\mathrm{d}x}\right) f_X(u)\, \mathrm{d}(u)_1 \dots \mathrm{d}(u)_p = \left(\operatorname{vec}' B, \frac{\mathrm{d}^{\otimes 2}}{\mathrm{d}x}\right) F_X(x)$$

and

$$\int_{-\infty}^{(x)_1} \cdots \int_{-\infty}^{(x)_p} \left(\operatorname{vec}' C, \frac{\mathrm{d}^{\otimes 3}}{\mathrm{d}x}\right) f_X(u)\, \mathrm{d}(u)_1 \dots \mathrm{d}(u)_p = \left(\operatorname{vec}' C, \frac{\mathrm{d}^{\otimes 3}}{\mathrm{d}x}\right) F_X(x).$$

So we have proved the equality (4). □

We shall examine the equality (4) term by term. That means we study the terms $\left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) F_X(x)$, $\left(\operatorname{vec}' B, \frac{\mathrm{d}^{\otimes 2}}{\mathrm{d}x}\right) F_X(x)$ and $\left(\operatorname{vec}' C, \frac{\mathrm{d}^{\otimes 3}}{\mathrm{d}x}\right) F_X(x)$. For this study we formulate three lemmas.

Lemma 1. *Let the functions $f(x)$ and $F(x)$ be such that (5) holds and let a be a constant p-vector. Then*

$$\int\limits_{\Re^p} \left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) f(x) \circ {}_s\mathrm{d}x = (a)_i f(x) - \sum_{j=1}^{p} ((a)_i - (a)_j) \frac{\partial F(x)}{\partial (x)_j}, \qquad i = 1, 2, \ldots, p.$$

P r o o f. Let us take the ith component of vector a. We can add and subtract to $\left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) f(x)$ the terms $(a)_i \frac{\partial f(x)}{\partial (x)_j}$, $j = 1, 2, \ldots, p$, $j \neq i$. So we get

$$\left(a, \frac{\mathrm{d}}{\mathrm{d}x}\right) f(x) = (a)_i \sum_{j=1}^{p} \frac{\partial f(x)}{\mathrm{d}(x)_j} + \sum_{j=1}^{p} ((a)_j - (a)_i) \frac{\partial f(x)}{\partial (x)_j}. \tag{6}$$

Applying Proposition 2 we get

$$\int\limits_{\Re^p} (a)_i \sum_{j=1}^{p} \frac{\partial f(x)}{\mathrm{d}(x)_j} \circ {}_s\mathrm{d}x = (a)_i f(x). \tag{7}$$

To the second term on right hand side of (6) we apply Proposition 5:

$$\begin{aligned} \int\limits_{\Re^p} \sum_{j=1}^{p} ((a)_j - (a)_i) \frac{\partial f(x)}{\partial (x)_j} \circ {}_s\mathrm{d}x &= \sum_{j=1}^{p} ((a)_i - (a)_j) \frac{\partial}{\partial (x)_j} \int\limits_{\Re^p} f(x) \circ {}_s\mathrm{d}x \\ &= \sum_{j=1}^{p} ((a)_i - (a)_j) \frac{\partial F(x)}{\partial (x)_j}. \end{aligned} \tag{8}$$

After subtracting the equality (8) from equality (7) we get the statement. □

Applying Propositions 4 and 5 and assumption (5) we can prove the next two Lemmas.

Lemma 2. *Let the functions $f(x)$ and $F(x)$ be such that (5) holds and B be a constant $p \times p$-matrix Then*

$$\int\limits_{\Re^p} \operatorname{vec}' B \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} \circ {}_s\mathrm{d}x = \operatorname{vec}' B \frac{\mathrm{d}^2 F(x)}{\mathrm{d}x^2}.$$

Lemma 3. *Let the functions $f(x)$ and $F(x)$ be such that (5) holds and C be a constant $p^2 \times p$-matrix. Then*

$$\int\limits_{\Re^p} \operatorname{vec}' C \frac{\mathrm{d}^3 f(x)}{\mathrm{d}x^3} \circ {}_s\mathrm{d}x = \operatorname{vec}' C \frac{\mathrm{d}^3 F(x)}{\mathrm{d}x^3}.$$

Using the equality (4) and Lemmas 1, 2 and 3 we get

$$\begin{aligned} F_Y(x) = F_X(x) - (a)_i f_X(x) + \sum_{j=1}^{p}((a)_i - (a)_j)\frac{\partial F_X(x)}{\partial (x)_j} \\ + \operatorname{vec}' B \frac{\mathrm{d}^2 F_X(x)}{\mathrm{d}x^2} - \operatorname{vec}' C \frac{\mathrm{d}^3 F_X(x)}{\mathrm{d}x^3} + \dots . \end{aligned} \tag{9}$$

Now we give general relation between two distribution functions using Theorem 1 and Lemmas 1–3.

Let us have a random p-vector X with the density function $f(x)$ and the distribution function $F(x)$. We introduce the next notations:

$$u_{-j} = ((u)_1, \dots, (u)_{j-1}, (u)_{j+1}, \dots, (u)_p)$$

and

$$\mathrm{d}u_{-j} = \mathrm{d}(u)_1 \dots \mathrm{d}(u)_{j-1} \mathrm{d}(u)_{j+1} \dots \mathrm{d}(u)_p.$$

Let $f_j((x)_j)$ be the marginal density function of $(X)_j$. Let $f_{ij}((x)_i, (x)_j)$ be the marginal joint density function of $(X)_i$ and $(X)_j$ and $f_{ijk}((x)_i, (x)_j, (x)_k)$ be the joint density function of $(X)_i$, $(X)_j$ and $(X)_k$. Let $f((x)_j|x_{-j})$ be the conditional density function of $(X)_j$. The number of combinations of n elements by k elements is denoted as C_k^n.

Using notations given above we formulate the following theorem.

Theorem 2. *Let X be a random p-vector with known distribution function $F(x)$ and density function $f(x)$. Let Y be a random p-vector with unknown distribution function $F_Y(x)$. Then the distribution function $F_Y(x)$ is expressed through the distribution of X as follows:*

$$\begin{aligned} F_Y(x) = F(x) - (a)_i f(x) + \sum_{j=1}^{p}((a)_i - (a)_j) f_j((x)_j) F(x_{-j}|(x)_j) \\ + \sum_{i=1}^{p}(B)_{ii}\frac{\mathrm{d}f_i((x_i)}{\mathrm{d}(x)_i}F(x_{-i}|(x)_i) + f_i((x)_i)\frac{\partial F(x_{-i}|(x)_i)}{\partial (x)_i} \\ + 2\sum_{i,j=1}^{p}(B)_{ij} f_{ij}((x)_i, (x)_j) F(x_{-i-j}|(x)_i, (x)_j) \\ - \sum_{i=1}^{p}(C)_{(i,i)(1,i)}\Bigg(\frac{\mathrm{d}^2 f_i((x_i)}{\mathrm{d}(x)_i^2}F(x_{-i}|(x)_i) + 2\frac{\mathrm{d}f_i((x_i)}{\mathrm{d}(x)_i}\frac{\partial F(x_{-i}|(x)_i)}{\partial (x)_i} \\ + f_i((x)_i)\frac{\partial^2 F(x_{-i}|(x)_i)}{\partial (x)_i^2}\Bigg) \end{aligned}$$

$$
\begin{aligned}
&- 3\mathrm{C}_2^p \sum_{i,j=1}^{p} ((C)_{(j,j)(1,i)} + (C)_{(j,i)(1,j)} + (C)_{(i,j)(1,j)}) \times \\
&\times \Bigg(\frac{\partial f_{ij}((x)_i,(x)_j)}{\partial (x)_j} F(x_{-i-j}|(x)_i,(x)_j) \\
&\qquad\qquad + f_{ij}((x)_i,(x)_j) \frac{\partial F(x_{-i-j}|(x)_i,(x)_j)}{\partial (x)_j} \Bigg) \\
&- 3\mathrm{C}_2^p \sum_{i,j=1}^{p} ((C)_{(i,i)(1,j)} + (C)_{(i,j)(1,i)} + (C)_{(j,i)(1,i)}) \times \\
&\times \Bigg(\frac{\partial f_{ij}((x)_i,(x)_j)}{\partial (x)_i} F(x_{-i-j}|(x)_i,(x)_j) \\
&\qquad\qquad + f_{ij}((x)_i,(x)_j) \frac{\partial F(x_{-i-j}|(x)_i,(x)_j)}{\partial (x)_i} \Bigg) \\
&- 6\mathrm{C}_3^p \sum_{i,j,k=1}^{p} (C)_{(i,j)(1,k)} f_{ijk}((x)_i,(x)_j,(x)_k) F(x_{-i-j-k}|(x)_i,(x)_j,(x)_k) \\
&\qquad\qquad + \cdots .
\end{aligned}
\tag{10}
$$

P r o o f. The statement is obtained by applying consequently Lemmas 1, 2 and 3. Let us start from equality (9). Firstly we find the partial derivatives $\frac{\partial F(x)}{\partial (x)_j}$. We get

$$
\begin{aligned}
\frac{\partial F(x)}{\partial (x)_j} &= \int_{-\infty}^{(x)_1} \cdots \int_{-\infty}^{(x)_{j-1}} \int_{-\infty}^{(x)_{j+1}} \cdots \int_{-\infty}^{(x)_p} f((x)_i, u_{-j}) \, \mathrm{d}u_{-j} \\
&= f_j((x)_j) \int_{-\infty}^{(x)_1} \cdots \int_{-\infty}^{(x)_{j-1}} \int_{-\infty}^{(x)_{j+1}} \cdots \int_{-\infty}^{(x)_p} f(u_{-j}|(x)_i) \, \mathrm{d}u_{-j} \\
&= f_j((x)_j) F(x_{-j}|(x)_j), \qquad j = 1, 2, \ldots, p.
\end{aligned}
$$

It is straightforward to detect that

$$
\begin{aligned}
\operatorname{vec}' B \frac{\mathrm{d}^2 F(x)}{\mathrm{d}x^2} &= \sum_{i,j=1}^{p} (B)_{ij} \frac{\partial^2 F(x)}{\partial^2 (x)_i (x)_j}, \\
\operatorname{vec}' C \frac{\mathrm{d}^3 F(x)}{\mathrm{d}x^3} &= \sum_{i,j,k=1}^{p} (C)_{(i,j)(1,k)} \frac{\partial^3 F(x)}{\partial^2 (x)_i (x)_j (x)_k}.
\end{aligned}
$$

For the higher order partial derivatives of $F(x)$ we get

$$\frac{\partial^2 F(x)}{\partial (x)_i^2} = \frac{\mathrm{d}f_i((x_i)}{\mathrm{d}(x)_i} F(x_{-i}|(x)_i) + f_i((x)_i) \frac{\partial F(x_{-i}|(x)_i)}{\partial (x)_i}$$

and

$$\begin{aligned} \frac{\partial^3 F(x)}{\partial (x)_i^3} = \frac{\mathrm{d}^2 f_i((x_i)}{\mathrm{d}(x)_i^2} F(x_{-i}|(x)_i) + 2 \frac{\mathrm{d}f_i((x_i)}{\mathrm{d}(x)_i} \frac{\partial F(x_{-i}|(x)_i)}{\partial (x)_i} \\ + f_i((x)_i) \frac{\partial^2 F(x_{-i}|(x)_i)}{\partial (x)_i^2}, \qquad i = 1, 2, \ldots, p. \end{aligned}$$

Now we study mixed derivatives of $F(x)$. In the case of the second order mixed derivatives we get

$$\begin{aligned} \frac{\partial^2 F(x)}{\partial (x)_i (x)_j} &= \frac{\partial f_i((x_i) F(x_{-i}|(x)_i)}{\partial (x)_j} \\ &= f_{ij}((x)_i, (x)_j) F(x_{-i-j}|(x)_i, (x)_j), \qquad i \neq j, \quad i, j = 1, 2, \ldots p. \end{aligned}$$

For the third order mixed derivatives we get

$$\begin{aligned} \frac{\partial^3 F(x)}{\partial (x)_i (x)_j^2} &= \frac{\partial f_{ij}((x)_i, (x)_j) F(x_{-i-j}|(x)_i, (x)_j)}{\partial (x)_j} \\ = \frac{\partial f_{ij}((x)_i, (x)_j)}{\partial (x)_j} & F(x_{-i-j}|(x)_i, (x)_j) + f_{ij}((x)_i, (x)_j) \frac{\partial F(x_{-i-j}|(x)_i, (x)_j)}{\partial (x)_j} \end{aligned}$$

$$\begin{aligned} \frac{\partial^3 F(x)}{\partial (x)_i (x)_j (x)_k} = f_{ijk}((x)_i, (x)_j, (x)_k) F(x_{-i-j-k}|(x)_i, (x)_j, (x)_k), \\ i \neq j \neq k, \quad i, j, k = 1, 2, \ldots p. \end{aligned}$$

Replacing partial derivatives, the second and the third order derivatives and all mixed derivatives into equality (9) we get the statement of the theorem. $\square$

Let us apply the statement (10) in the two-dimensional case. Let $X = ((X)_1, (X)_2)'$ and $Y = ((Y)_1, (Y)_2)'$ be random vectors with known and unknown distribution functions respectively. Then from Theorem 2 we get

Corollary 1.

$$
\begin{aligned}
&F_Y(x) = \\
&= F(x) - (a)_1 f(x) + ((a)_1 - (a)_2) f_2((x)_2) F((x)_1|(x)_2) \\
&\quad + 2(B)_{12} f(x) + (B)_{11} \frac{\partial f_1((x)_1) F((x)_2|(x)_1)}{\partial (x)_1} + (B)_{22} \frac{\partial f_2((x)_2) F((x)_1|(x)_2)}{\partial (x)_2} \\
&\quad - ((C)_{((1,1)(1,2)} + (C)_{(2,1)(1,1)} + (C)_{(1,2)(1,1)}) \frac{\partial f(x)}{\partial (x)_1} \\
&\quad - ((C)_{(2,2)(1,1)} + (C)_{(2,1)(1,2)} + (C)_{(1,2)(1,2)}) \frac{\partial f(x)}{\partial (x)_2} \\
&\quad - \left((C)_{(1,1)(1,1)} \frac{\partial^2 f_1((x)_1) F((x)_2|(x)_1)}{\partial^2 (x)_1} + (C)_{(2,2)(1,2)} \frac{\partial^2 f_2((x)_2) F((x)_1|(x)_2)}{\partial^2 (x)_2} \right) \\
&\qquad\qquad\qquad + \ \ldots .
\end{aligned}
$$

5. Approximation with the normal distribution

In this section we apply the results of the previous section in the case when we approximate the unknown distribution function through the distribution function of the normal distribution $N(0_2, \Sigma)$.

We introduce first the Hermite matrix-polynomials for p-vector x. By means of these functions we can easily approximate the unknown distribution with the normal distribution. The approximation by Hermite polynomials is first time used in [2]. Hermite matrix polynomials are defined by means of *Neudecker* matrix derivative (Definition 1).

Definition 6. The matrix $H_k(x, \Sigma)$ is called *Hermite matrix-polynomial* if it is defined by the equality

$$
\frac{\mathrm{d}^k f_X(x)}{\mathrm{d}x^k} = (-1)^k H_k(x, \Sigma) f_X(x), \qquad k = 1, 2, \ldots
$$

where $f_X(x)$ is the density function of normal distribution $N(0_p, \Sigma)$.

The Hermite matrix polynomials up to the third order are given by equalities in [3]:

$$
\begin{aligned}
&H_0(x, \Sigma) = 1; \\
&H_1(x, \Sigma) = x'\Sigma^{-1}; \\
&H_2(x, \Sigma) = \Sigma^{-1} x x' \Sigma^{-1} - \Sigma^{-1}; \\
&H_3(x, \Sigma) = (\Sigma^{-1} x)^{\otimes 2} x' \Sigma^{-1} - \operatorname{vec} \Sigma^{-1} (x' \Sigma^{-1}) - (\Sigma^{-1} \otimes x \otimes \Sigma^{-1}).
\end{aligned}
$$

In the univariate case when $X \sim N(0, \sigma^2)$ the Hermitian polynomials $h_1(x)$ and $h_2(x)$ take the following form:

$$h_1(x) = x\sigma^{-2}$$

and

$$h_2(x) = x^2\sigma^{-4} - \sigma^{-2}.$$

Now we apply Theorem 2 in the case if X is a bivariate random vector, $X \sim N(0_2, \Sigma)$. Let ρ be the Pearson correlation coefficient between $(X)_1$ and $(X)_2$. Let us use the following notation:

$$g((x)_1) = \frac{\frac{(x)_2}{\sqrt{(\Sigma)_{22}}} - \frac{(x)_1}{\sqrt{(\Sigma)_{11}}}\rho}{\sqrt{1-\rho^2}}.$$

Let Φ be the distribution function of the standard normal distribution. In the next theorem we present a formal expression of the distribution function of the bivariate random vector.

Theorem 3. *Let $X \sim N(0_2, \Sigma)$ with the distribution function of $F_X(x)$ and $F_Y(x)$ be the unknown distribution function of bivariate random vector Y. Let $f_1((x)_1)$ and $f_2((x)_2)$ be the marginal density functions of $(X)_1$ and $(X)_2$, respectively. Then*

$$\begin{aligned}
F_Y(x) = {} & F_X(x) + \{(a)_2 + 2(B)_{12} + (C_{12}, H_1(x,\Sigma)\} f_X(x) \\
& + \{((a)_1 - (a)_2) f_2((x)_2)\} \Phi(g((x)_2)) \\
& - (B)_{11}\{h_1((x)_1) - g'((x)_1) f_1((x)_1)\} \Phi(g((x)_1)) \\
& - (B)_{22}\{h_1((x)_2) - g'((x)_2) f_2((x)_2)\} \Phi(g((x)_2)) \\
& - C_{(1,1)(1,1)}\{h_2((x)_1) f_1((x)_1) \Phi(g((x)_1)) \\
& - 2h_1((x)_1) f_1((x)_1) f_1(g((x)_1)) g'((x)_1)) \\
& - f_1((x)_1)) h_1(g((x)_1)) f_1(g((x)_1)) g'((x)_1)^2\} \\
& - C_{(2,2)(1,2)}\{h_2((x)_2) f_2((x)_2) \Phi(g((x)_2)) \\
& - 2h_1((x)_2) f_2((x)_2) f_1(g((x)_2)) g'((x)_2)) \\
& - f_2((x)_2)) h_1(g((x)_2)) f_2(g((x)_2)) g'((x)_2)^2\} + \dots
\end{aligned} \tag{11}$$

where

$$C_{12} = \begin{pmatrix} (C)_{(1,1)(1,2)} + (C)_{(1,2)(1,1)} + (C)_{(2,1)(1,1)} \\ (C)_{(2,2)(1,1)} + (C)_{(2,1)(1,2)} + (C)_{(1,2)(1,2)} \end{pmatrix}.$$

P r o o f. Let us start from Corollary 1. Firstly we show that

$$F((x)_2|(x)_1) = \Phi(g((x)_1). \tag{12}$$

We get by integration

$$\int_{-\infty}^{(x)_2} f((x)_1,(u)_2)\,\mathrm{d}(u)_2 = f_1((x)_1)\int_{-\infty}^{(x)_2} f((u)_2|(x)_1)\,\mathrm{d}(u)_2$$
$$= f_1((x)_1)F((x)_2|(x)_1).$$

Let us find the integral $\int\limits_{-\infty}^{(x)_2} f((x)_1,(u)_2)\,\mathrm{d}(u)_2$. We get

$$\int_{-\infty}^{(x)_2} f((x)_1,(u)_2)\,\mathrm{d}(u)_2$$
$$= \frac{1}{2\pi|\Sigma|^{1/2}}\int_{-\infty}^{(x)_2} \exp\left(\frac{1}{2}((x)_1,(u)_2)\Sigma^{-1}((x)_1,(u)_2)\right)\,\mathrm{d}(u)_2$$
$$= \frac{1}{2\pi|\Sigma|^{1/2}}\int_{-\infty}^{(x)_2} \exp\left(\frac{(\Sigma)_{22}(x)_2^2 - 2(\Sigma)_{12}(x)_1(u)_2 + (\Sigma)_{11}(u)_2^2}{2((\Sigma)_{11}(\Sigma)_{22} - (\Sigma)_{12}^2)}\right)\,\mathrm{d}(u)_2$$
$$= f_1((x)_1)\int_{-\infty}^{(x)_2} \exp\left(-\frac{\left(\sqrt{\frac{(\Sigma)_{22}}{(\Sigma)_{11}}}\rho(x)_1 - (u)_2\right)^2}{2(\Sigma)_{22}(1-\rho^2)}\right)\,\mathrm{d}(u)_2.$$

We have got

$$F((x)_2|(x)_1) = \frac{1}{\sqrt{\Sigma_{22}(1-\rho^2)}}\int_{-\infty}^{(x)_2} \exp\left(-\frac{(\sqrt{\frac{(\Sigma)_{22}}{(\Sigma)_{11}}}\rho(x)_1 - (u)_2)^2}{2(\Sigma)_{22}(1-\rho^2)}\right)\,\mathrm{d}(u)_2.$$

It is easy detect that

$$D\left(\sqrt{\frac{(\Sigma)_{22}}{(\Sigma)_{11}}}\rho(X)_1 - (X)_2\right) = (\Sigma)_{22}(1-\rho^2)$$

and

$$g((x)_1) = \frac{(x)_2 - \sqrt{\frac{(\Sigma)_{22}}{(\Sigma)_{11}}}\rho(x)_1}{\sqrt{(\Sigma)_{22}(1-\rho^2)}}.$$

So the equality (12) is proven. In the same way we can prove that

$$F((x)_1|(x)_2) = \Phi(g((x)_2). \tag{13}$$

Secondly we find higher order partial derivatives of $F_X(x)$. It follows straightforwardly that

$$\frac{\partial^2 F_X(x)}{\partial(x)_1(x)_2} = f_X(x), \tag{14}$$

$$\frac{\partial^3 F_X(x)}{\partial(x)_1^2(x)_2} = (H_1(x,\Sigma))_1 f_X(x), \tag{15}$$

$$\frac{\partial^3 F_X(x)}{\partial(x)_1(x)_2^2} = (H_1(x,\Sigma))_2 f_X(x). \tag{16}$$

Using univariate Hermite polynomials we get

$$\begin{aligned}&\frac{\partial f_1((x)_1)F((x)_2|(x)_1)}{\partial(x)_1}\\ &= -h_1((x)_1)f_1((x)_1)\Phi(g((x)_1)) + f_1((x)_1)f_1(g((x)_1))g'((x)_1)\end{aligned} \tag{17}$$

and

$$\begin{aligned}&\frac{\partial^2 f_1((x)_1)F((x)_2|(x)_1)}{\partial(x)_1^2}\\ &= h_2((x)_1)f_1((x)_1)\Phi((g((x)_1)) - 2h_1((x)_1)f_1((x)_1)f_1(g((x)_1))g'((x)_1)\\ &\qquad + f_1((x)_1)h_1(g((x)_1)f_1((g((x)_1))g'((x)_1)^2.\end{aligned} \tag{18}$$

In the same way

$$\begin{aligned}&\frac{\partial f_2((x)_2)F((x)_1|(x)_2)}{\partial(x)_2}\\ &= -h_1((x)_2)f_2((x)_2)\Phi(g((x)_2)) + f_2((x)_2)f_2(g((x)_2))g'((x)_2)\end{aligned} \tag{19}$$

and

$$\begin{aligned}&\frac{\partial^2 f_2((x)_2)F((x)_1|(x)_2)}{\partial(x)_2^2}\\ &= h_2((x)_2)f_2((x)_2)\Phi((g((x)_2)) - 2h_1((x)_2)f_2((x)_2)f_2(g((x)_2))g'((x)_2)\\ &\qquad + f_2((x)_2)h_1(g((x)_2)f_2((g((x)_2))g'((x)_2)^2.\end{aligned} \tag{20}$$

Replacing the expressions (12)-(20) and the statement of Definition 3 into the equality of Corollary 1 we get the statement. □

6. Simulation

Let us generate model data from the random vector

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

where $\ln(X_1)$ and X_2 have the standard normal distribution and the Pearson correlation coefficient $\rho \approx 0.8$. Let $F_X(x)$ be the distribution function of X and S be the sample covariance matrix calculated from data. Let $F(x)$ and $f(x)$ be the distribution and density functions of normal distribution $N(0_2, S)$, respectively. We approximate the function $F_X(x)$ with the function $F(x)$ using the equality (11). Let $c_3(X) = E(X \otimes X' \otimes X)$. Then we get

$$\begin{aligned} F_X(x) =& F(x) + \frac{1}{6}[(C1_3, H_1(x', S)f(x) \\ & - (c_3(X))_{11}\{h_2(x_1))f_1(x_1)\Phi(g(x_1)) - 2h_1(x_1)f_1(x_1)f_1(g(x_1))g'(x_1) \\ & - g(x_1)f_1(g(x_1))f_1(x_1)g'(x_1)^2\} \\ & - (c_3(X))_{42}\{h_2(x_2)f_2(x_2)\Phi(g(x_2)) - 2h_1(x_2)f_2(x_2)f_2(g(x_2)g'(x_2)) \\ & - g(x_2)f_2(g(x_2))f_2(x_2)g'(x_2)^2\}] + \dots \end{aligned} \tag{21}$$

where

$$C1_3 = \begin{pmatrix} (c_3(X))_{12} + (c_3(X))_{21} + (c_3(X))_{31} \\ (c_3(X))_{22} + (c_3(X))_{32} + (c_3(X))_{41} \end{pmatrix}$$

and $f_1(x_1)$ and $f_2(x_2)$ are marginal density functions of $(Y)_1$ and $(Y)_2$ respectively.

The goodness of approximation is estimated by

$$d = \frac{\sum_{i=1}^{k}(F_k(x_i) - F(x_i))^2}{k} \tag{22}$$

where x_i is the ith value of X, $F_k(x_i)$ is the empirical distribution function, $F(x_i)$ is the theoretical distribution function and sample size $k = 200$.

We apply equality (22) to the distribution functions $F_X(x)$ and $F(x)$. This procedure is repeated 10 times. The value $\overline{d}$ is the calculated average. We get that in the case of normal distribution $\overline{d} = 0.0168$ and in the case when applying the equality (21) we have $\overline{d} = 0.00202$. We can conclude that the expression (21) corrects essentially the fit of distribution function.

REFERENCES

[1] ANDERSON, T. W.: *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 2003.
[2] CORNISH, E. A.—FISHER, R. A.: *Moments and cumulants in the specification of distribution*, Rev. Inst. Int. Stat. **5** (1937), 307–322.
[3] KOLLO, T.: *Matrix Derivative in Multivariate Statistics*, Tartu University Press, Tartu, 1991 (Russian).
[4] KOLLO, T.—VON ROSEN, D.: *Approximating by the Wishart distribution*, Ann. Inst. Statist. Math. **47** (1995), 767–783.
[5] MACRAE, E. C.: *Matrix derivatives with an applications to an adaptive linear decision problem*, Ann. Statist. **7** (1974), 381–394
[6] NEUDECKER, H.: *Some theorems on matrix differentiations with special reference to Kronecker matrix products*, J. Amer. Statist. Assoc. **64** (1969), 953–963.
[7] PIHLAK, M.: *Matrix integral*, Linear Algebra Appl. **388** (2004), 315–325

Received 14. 2. 2007
Revised 12. 4. 2007

Institute of Mathematical Statistics
J. Liivi 2-513
EE–50409 Tartu
ESTONIA
E-mail: margusp@staff.ttu.ee