

**APPROXIMATION OF MULTIVARIATE
DISTRIBUTION FUNCTIONS**

MARGUS PIHLAK



TARTU UNIVERSITY
PRESS

Faculty of Mathematics and Computer Science, University of Tartu, Tartu

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (Ph. D.) in mathematical statistics on December 15, 2006, by the Council of the Faculty of Mathematics and Computer Science, University of Tartu

Supervisor:

Professor, Cand. Sc.

Tõnu Kollo
University of Tartu
Tartu, Estonia

Opponents:

Professor, Ph.D

Albert Satorra
University Pompeu Fabra
Barcelona, Spain

Associate Professor, Dr. habil.

Ivan Zezula
P. J. Safarik University
Kosice, Slovakia

The public defence will take place on January 19, 2007

ISSN 1024-4212

ISBN 978-9949-11-516-7 (trükis)

ISBN 978-9949-11-517-4 (PDF)

Autoriõigus Margus Pihlak, 2007

Tartu Ülikooli Kirjastus

www.tyk.ee

Tellimus nr. 684

Contents

List of original publications	7
Acknowledgements	9
Introduction	10
1 Matrix Integral	14
1.1 Basic matrix operations and notation	14
1.2 Partitioned matrices	15
1.3 Matrix derivative	18
1.4 Matrix integral	21
1.4.1 Definition of matrix integral	21
1.4.2 Properties of matrix integral	23
2 Approximation of multivariate distributions	30
2.1 Approximation of multivariate density functions	30
2.1.1 Cumulants	30
2.1.2 Relation between two density functions	32
2.2 Approximation of distribution functions	34
2.2.1 Integration of the relation between densities	35
2.2.2 General relation between two distribution functions	40

2.3	Approximation with the normal distribution	43
2.3.1	Representation of the unknown distribution function through the normal distribution	43
2.3.2	Simulation	47
2.3.3	Application of the Edgeworth type approximation .	51
3	Approximation by copulas	55
3.1	Some basic results of copula theory	55
3.1.1	Copulas and random variables	56
3.1.2	Dependence in copula theory	57
3.2	Archimedean and Gaussian copulas	59
3.2.1	Archimedean copulas	59
3.2.2	Copulas and density functions	60
3.2.3	Two-dimensional Gaussian copula	61
3.3	Application of copulas for distribution approximation . . .	62
3.3.1	Goodness-of-fit test for Archimedean copulas	63
3.3.2	Multivariate Kolmogorov-Smirnov test	64
3.3.3	Modeling the joint distribution of H and DBH . . .	65
3.4	Comparing different models of distribution	68
3.4.1	Bivariate Kolmogorov-Smirnov test	68
3.4.2	Results	70
3.4.3	Discussion and further development	71
	Bibliography	73
	Kokkuvõte	76
	Curriculum Vitae	78

List of original publications

1. Pihlak, M. (2004) Matrix integral. *Linear Algebra and Its Applications* **388**, 315-325
2. Iital, A., Stålnacke, P., Deelstra, J., Loigu, E., Pihlak, M. (2005). Effects of large-scale changes in emissions on nutrient concentrations in Estonian rivers in the Lake Peipsi drainage basin. *Journal of Hydrology*, **304**, 261-273.
3. Adermann, V., Pihlak, M. (2005) Using copulas for modeling the dependence between tree height and diameter at breast height. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, **9**, 77-85.
4. Pihlak, M. (2006a) Approximation of multivariate distribution functions. *Acta Applicandae Mathematicae*, 16 pages (submitted).
5. Vetemaa, M., Eschbaum, R., Verliin, A., Albert, A., Eero, M., Lillemägi, R., Pihlak, M., Saat, T. (2006b) Annual and seasonal dynamics of fish in the brackish-water Matsalu Bay, Estonia. *Ecology of Freshwater Fish*, **15**, 211-220.
6. Vetemaa, M., Neuman, E., Thoresson, G., Pihlak, M. (2006c) Trade-off between number and growth rate of offspring: intra-ovarian competition of embryos of viviparous blenny *Zoarces viviparous*. *Annales Zoologici Fennici* **43**, 303-308.

Conference abstracts

1. Kollo, T, Pihlak. M. (2002) Approximation of multivariate distribution functions. In: *8th International Vilnius Conference on Probability Theory and Mathematical Statistics*, Abstracts of Communications, Vilnius, Lithuania, June 23 to June 29, 2002, p. 252.
2. Pihlak, M., Tamsalu, R. (2003) Using split-up method on prey-predator dynamical system. In: *International Summer Seminar "Stochastic and Dynamical Systems"*. Abstracts SDS 2003, Sudak, Ukraine , May 30 to June 3, 2003, p. 71-72,
3. Pihlak, M. (2006) Approximation of multivariate distribution functions. In: *9th International Vilnius Conference on Probability Theory and Mathematical Statistics*, Abstracts of Communications, Vilnius, Lithuania, June 26 to June 30, 2006, p. 261.

Acknowledgements

I would like to express my gratitude to my supervisor Professor Tõnu Kollo for his advice and support during the whole process of my PhD studies and during all phases writing this Thesis.

I would like also to thank Veiko Adermann for forestry data and useful suggestions concerning classical sampling methods in forestry.

I am grateful to Meelis Käärik for technical advice when writing Thesis.

Many thanks to my colleagues in Estonian Marine Institute and Tallinn Technical University for moral support.

I would like to express my gratitude to my mother for her support and understanding during my studies.

Finally special thanks to my very good friends Ülo Maiväli, Mait Sepp and Elise Sonn for moral support and interesting discussions about science and philosophy.

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 approximate 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 represent the function G by means of the function F :

$$G(x) = \sum_{l=0}^k a_l \frac{d^l F(x)}{dx^l} + R_k \quad (1)$$

where a_l , $l = 1, 2, \dots, k$, are coefficients depending on the first l moments of X and Y and R_k is the remainder term. If the series (1) converges then we can use the equality (1) as an approximation for the distribution function G . The problem is, how to determine the coefficients a_l in equality (1). This type relation between distribution functions was first considered by R. A. Fisher and E. A. Cornish in 1937.

The method of Cornish and Fisher (1937) includes the following steps. Assume that random variables X and Y have 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 a representation of the probability density function of Y through the density function of X . Integrating this relation gives us an approximation of the unknown distribution function in the form (1) where the coefficients $a_l, l = 1, 2, \dots, k$ are functions of the first k cumulants of random variables X and Y .

This technic was generalized to the multivariate density functions by Traat (1986). In the multivariate case a relation between two densities is obtained by using matrix derivative. Different variants of matrix derivative

(Frechet derivative in matrix form) and related matrix algebra were examined and developed by different authors: Dwyer and MacPhail (1948), Neudecker (1969), MacRae (1974), Kollo (1977). In the papers Kollo and von Rosen (1995), (1998), a method is worked out which enables to present a complicated multivariate density of interest through the known density and cumulants of both distributions under consideration. The dimension of the approximating distribution can be higher than the dimension of the distribution of interest. 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 much more complicated.

In the multivariate density approximations higher order matrix derivatives are represented by matrices with growing dimensionalities. For integration of expansions a new notion - matrix integral is needed. The notion has to be an inverse operation of the matrix derivative. A solution to the problem is given in the paper Pihlak (2004) where matrix integral is introduced and its basic properties studied. Application of the matrix integral for integrating a density expression is possible after introducing several additional notations and providing new results. This technique is given in Pihlak (2006a) and presented in Section 2.2 of the dissertation. The results have been presented at the 9th Vilnius Conference on Probability Theory and Mathematical Statistics (Pihlak, 2006b).

It would be interesting to compare the goodness of obtained approximation with some other method.

The second approach to the approximation of multivariate distributions in this thesis is approximation with copulas. Initiator of the copula theory is A. Sklar who presented foundations of the theory in 1959 (Sklar, 1959). The approach enables to construct multivariate distribution function with given marginals and given dependence structure. Kotz and Seeger (1991) focused to construction of multivariate probability density functions using copula theory. In the framework of copula theory very important class of copulas is known as Archimedean copulas which include several one-parameter families of distributions. Copulas have been useful tools for modeling bivariate probability distributions. In the thesis the relation between height and diameter at breast height of a tree is modeled by means of copulas, following Adermann and Pihlak (2005). Very important property of copulas is their invariance with respect to rank correlation coefficients. Rank correlation plays very important role in non-parametric statistics. For example,

Mann-Kendall test is based on the Kendall's rank correlation coefficient. In article Ital et all (2004) the author of the dissertation has used this test when studying trends of the nutrient concentration in Estonian rivers in the Lake Peipsi drainage basin.

The main purposes of the present theses can be formulated as follows:

- work out a general technique to approximate the unknown distribution function with known distribution function by means of Taylor expansion,
- apply the method to real data to examine approximation of the empirical distribution function with the approximation based on the distribution function of the normal distribution,
- compare the goodness of our method of approximation with approximation by means of copulas.

The presentation is based on the papers by Pihlak (2004), Adermann and Pihlak (2005) and Pihlak (2006a). However, several extensions and new results have also been added.

The structure of the thesis is briefly described below. The thesis consists of three chapters.

In Chapter 1 we introduce main concepts and necessary results of matrix algebra for approximation of distribution functions. In particular, we present basic properties of the Kronecker product, star product and matrix derivative. Then we study a new notion, the matrix integral, introduced in Pihlak (2004). Matrix integral is defined by Definition 1.4.1 as an inverse operation of the matrix derivative. Theorem 1.4.1 gives us general rule how to find the matrix integral.

In Chapter 2 we apply the matrix integral to integrate a relation between two density functions. In the first part of this chapter we present some recent results of approximation of a multivariate density function. In the second part we find a general relation between two distribution functions. In the third part we show how to approximate the unknown distribution function using Edgeworth type expansions. Finally we apply the Edgeworth type expansion to the forestry data. The main results of distribution functions' approximation are presented by Theorems 2.2.3 and 2.3.1.

Chapter 3 deals with modeling using copula theory. Copula is defined by Definition 3.1.1. The relation between copulas and bivariate distribution functions is formulated in Theorem 3.1.1. In the first part we introduce and study main concepts and results of the copula theory. In particular, we

examine the relation between a copula and rank correlation coefficients. In the second part we study Archimedean copulas and present two-dimensional copula by normal distributions. In the third part we apply copulas for approximation of the joint distribution of tree height and tree diameter at breast height following Adermann and Pihlak (2005). The first author of this article has collected data and gives description of classical methods in forestry. Data analysis, comparison of different methods, choice of the best model of approximation and statistical analysis of different models is the contribution of the second author in Adermann and Pihlak (2005). In the end of this chapter we compare the distribution approximations obtained by Edgeworth type expansion and by copulas.

Finally some technical remarks. The end of proofs is denoted by \square . The mean value and variance are denoted as E and D . The set of real numbers is denoted by symbol \mathfrak{R} . Theorems, lemmas, propositions and definitions are enumerated by sections.

Chapter 1

Matrix Integral

In this chapter we introduce the basic notation and results of matrix algebra necessary for approximation of multivariate distributions. Vectorization operator, direct product and star product of matrices are the basic notions under consideration. Also the structure of partitioned matrices and commutation matrices is described. In the second section basic properties of the matrix derivative are given. In the third section the matrix integral is introduced as an inverse operation of the matrix derivative. The notion of matrix integral is introduced in Pihlak (2004). For more detailed presentation of matrix algebra see Schott (1997), Harville (1997), Rao and Rao (1998), Kollo and von Rosen (2005), for example.

1.1 Basic matrix operations and notation

In this section we introduce the main notation and describe the basic matrix operations used for approximation of multivariate distributions.

Let us denote matrix \mathbf{X} with p rows and q columns by $\mathbf{X} : p \times q$. The element of matrix \mathbf{X} in the i -th row and j -th column is denoted by x_{ij} . So we have

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pq} \end{pmatrix}.$$

For matrices \mathbf{X} consisting of complex expressions of matrices the notation $x_{ij} = (\mathbf{X})_{ij}$ is also used. A $p \times 1$ -matrix is called p -vector. The i -th coordinate of the p -vector \mathbf{a} is denoted by a_i .

The $p \times q$ -matrix of zeros is denoted by $\mathbf{0}_{p \times q}$ and a p -vector with zeros as coordinates is denoted as $\mathbf{0}_p$.

The identity matrix of order p is denoted by \mathbf{I}_p .

The transposed matrix \mathbf{X} is denoted by \mathbf{X}' . So we have $(\mathbf{X})_{ij} = (\mathbf{X}')_{ji}$.

Let \mathbf{A} be a $p \times p$ -square matrix. Then trace of the matrix \mathbf{A} is denoted as $\text{tr}(\mathbf{A})$:

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^p a_{ii}.$$

The determinant of a square matrix \mathbf{X} is denoted by $|\mathbf{X}|$.

1.2 Partitioned matrices

Now we describe main matrix operations from so-called newer matrix algebra (Magnus, Neudecker, 1999).

If we handle partitioned matrix \mathbf{X} then its blocks in the i -th row and j -th column of block is denoted by $[\mathbf{X}]_{ij}$.

First we define the partitioned matrix.

Definition 1.2.1. *A $p \times q$ -matrix \mathbf{A} is called partitioned matrix if it consists of $p_i \times q_j$ -submatrices $[\mathbf{A}]_{ij}$ (blocks) so that*

$$\mathbf{A} = \begin{pmatrix} [\mathbf{A}]_{11} & [\mathbf{A}]_{12} & \cdots & [\mathbf{A}]_{1v} \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{A}]_{u1} & [\mathbf{A}]_{u2} & \cdots & [\mathbf{A}]_{uv} \end{pmatrix}, \sum_{i=1}^u p_i = p, \sum_{j=1}^v p_i = q.$$

We use index pairs for indicating rows and columns of a partitioned matrix following Anderson (2003, p. 642-643). A row of partitioned matrix \mathbf{A} is denoted by index (k, l) , if this is the $(\sum_{i=1}^{k-1} p_i + l)$ -th row of matrix \mathbf{A} , i.e. this is the l -th row of the k -th block-row.

A column of partitioned matrix \mathbf{A} is denoted by index (g, h) , if this is the $(\sum_{i=1}^{g-1} p_i + h)$ -th column of matrix \mathbf{A} , i.e. this is the h -th column of the g -th block-column.

The element of a partitioned matrix \mathbf{A} in the (k, l) -th row and the (g, h) -th column is denoted by $a_{(k,l)(g,h)}$.

Definition 1.2.2. The $pq \times pq$ partitioned matrix $\mathbf{K}_{p,q}$ which consists of $q \times p$ -blocks is called commutation matrix, if

$$(\mathbf{K}_{p,q})_{(i,j)(g,h)} = \begin{cases} 1, & \text{if } g = j, \quad h = 1 \quad i, h = 1, 2, \dots, p; j, g = 1, 2, \dots, q, \\ 0, & \text{otherwise.} \end{cases} \quad (1.1)$$

From the equality (1.1) follows that one element in each column and row of $\mathbf{K}_{p,q}$ equals to one and the other elements are zeros. It is also straightforward to detect that $\mathbf{K}_{p,1} = \mathbf{K}_{1,p} = \mathbf{I}_p$.

The vectorization operation is denoted by vec . For matrix $\mathbf{X} : p \times q$ the following pq -vector is denoted by $\text{vec}\mathbf{X}$:

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

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

$$\mathbf{X} \otimes \mathbf{Y} = [x_{lj}\mathbf{Y}], \quad l = 1, 2, \dots, p; j = 1, 2, \dots, q$$

where

$$x_{lj}\mathbf{Y} = \begin{pmatrix} x_{lj}y_{11} & \cdots & x_{lj}y_{1s} \\ \vdots & \ddots & \vdots \\ x_{lj}y_{r1} & \cdots & x_{lj}y_{rs} \end{pmatrix}.$$

The Kroneckerian k -th power of a p -vector \mathbf{a} is the p^k -vector $\mathbf{a}^{\otimes k}$,

$$\mathbf{a}^{\otimes k} = \underbrace{\mathbf{a} \otimes \mathbf{a} \otimes \dots \otimes \mathbf{a}}_{k \text{ times}}, \quad k = 1, 2, \dots$$

For $k = 0$, we define $\mathbf{a}^{\otimes 0} = 1$. Let \mathbf{A} be an $r \times s$ matrix. Then $r^k \times s^k$ -matrix $\mathbf{A}^{\otimes k}$ is called the Kroneckerian k -th power of \mathbf{A} and is defined as k times Kronecker product of \mathbf{A} to itself:

$$\mathbf{A}^{\otimes k} = \underbrace{\mathbf{A} \otimes \mathbf{A} \otimes \dots \otimes \mathbf{A}}_{k \text{ times}}$$

with $\mathbf{A}^{\otimes 0} = 1$.

Main properties of the Kronecker product are presented in the following proposition. For proof see Kollo (1991, p. 24-26), or Neudecker (1969).

Proposition 1.2.1. *Let us have matrices $\mathbf{A} : p \times q$, $\mathbf{B} : q \times n$, $\mathbf{C} : r \times t$, $\mathbf{D} : t \times s$ and a scalar c . Then the following equalities hold:*

$$(i) \quad (\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}',$$

$$(ii) \quad (\mathbf{A} + \mathbf{B}) \otimes (\mathbf{C} + \mathbf{D}) = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C} + \mathbf{A} \otimes \mathbf{D} + \mathbf{B} \otimes \mathbf{D},$$

$$(iii) \quad (\mathbf{AB}) \otimes (\mathbf{CD}) = (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}),$$

$$(iv) \quad c(\mathbf{A} \otimes \mathbf{B}) = (c\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (c\mathbf{B}),$$

$$(v) \quad \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C},$$

(vi) *for nonsingular matrices $\mathbf{A} : p \times p$ and $\mathbf{B} : r \times r$*

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}.$$

The vec-operator establishes the following relation between the matrix product and the Kronecker product.

Proposition 1.2.2. *Let \mathbf{C} be an $r \times q$ -matrix, \mathbf{A} a $p \times s$ -matrix and \mathbf{B} a $s \times r$ -matrix. Then*

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\mathbf{B}.$$

It is possible to commute the Kronecker product with help of the commutation matrix.

Proposition 1.2.3. *Let \mathbf{A} be a $p \times q$ and \mathbf{B} be an $r \times s$ matrix. Then*

$$\mathbf{A} \otimes \mathbf{B} = \mathbf{K}_{r,p}(\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{q,s}.$$

Now we study the star product. This operation is introduced in MacRae (1974). She has denoted this operation by $*$. Later the star product is important when studying matrix integral.

Definition 1.2.3. *Let us have matrix $\mathbf{A} : p \times q$ and partitioned-matrix $\mathbf{B} : pr \times qs$, consisting of $r \times s$ blocks. Then the star product $\mathbf{A} * \mathbf{B} : r \times s$ is defined as*

$$\mathbf{A} * \mathbf{B} = \sum_{l=1}^p \sum_{j=1}^q a_{lj} [\mathbf{B}]_{lj}$$

where the blocks $[\mathbf{B}]_{lj}$ are $r \times s$ -matrices.

It is seen that the star product decreases the dimensions of involved matrices.

Next we present some properties of the star product. For proofs the interested reader is referred to Kollo (1991 p. 34-37).

Proposition 1.2.4. *Let \mathbf{A} and \mathbf{B} be the matrices of the same dimensions. Then*

$$\mathbf{A} * \mathbf{B} = \text{tr}(\mathbf{A}'\mathbf{B}).$$

Proposition 1.2.5. *Let us have matrices $\mathbf{A} : p \times q$, $\mathbf{B} : pr \times qs$ and $\mathbf{C} : m \times n$. Then*

$$(\mathbf{A} * \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} * \mathbf{C}).$$

Proposition 1.2.6. *Let us have matrices $\mathbf{A} : p \times q$, $\mathbf{B} : q \times r$ and $\mathbf{C} : r \times s$. Then*

$$\mathbf{ABC} = \mathbf{B} * \text{vec}(\mathbf{A})\text{vec}'(\mathbf{C}') = \mathbf{B}' * (\mathbf{C} \otimes \mathbf{I}_p)\mathbf{K}_{p,s}(\mathbf{A} \otimes \mathbf{I}_s).$$

1.3 Matrix derivative

Matrix derivative has been a useful tool for statisticians for more than 30 years. The notion "matrix derivative" has been used for different representations of the *Frechet' derivative*. In multivariate statistics the definition of H. Neudecker is mainly used (Neudecker, 1969).

Let the elements of the matrix $\mathbf{Y} : r \times s$ be functions of matrix $\mathbf{X} : p \times q$. Assume that for all $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$, $k = 1, 2, \dots, r$ and $l = 1, 2, \dots, s$ partial derivatives $\frac{\partial y_{kl}}{\partial x_{ij}}$ exist and are continuous in an open set A . Then the matrix derivative is defined as follows.

Definition 1.3.1. *The matrix $\frac{d\mathbf{Y}}{d\mathbf{X}} : rs \times pq$ is called matrix derivative of $\mathbf{Y} : r \times s$ by $\mathbf{X} : p \times q$ in a set A , if*

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{d}{d\text{vec}'\mathbf{X}} \otimes \text{vec}\mathbf{Y}$$

where

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

There exists also another widely used form of the matrix derivative. The matrix derivative defined by MacRae (1974) keeps the structure of involved matrices.

Definition 1.3.2. The matrix $\frac{\partial \mathbf{Y}}{\partial \mathbf{X}} : pr \times qs$ is called matrix derivative of $\mathbf{Y} : r \times s$ by $\mathbf{X} : p \times q$ in a set A , if

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

where

$$\frac{d}{d\mathbf{X}} = \begin{pmatrix} \frac{\partial}{\partial x_{11}} & \cdots & \frac{\partial}{\partial x_{1q}} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{p1}} & \cdots & \frac{\partial}{\partial x_{pq}} \end{pmatrix}.$$

We shall use the following metric of a $p \times q$ -matrix \mathbf{X} :

$$\|\mathbf{X}\| = \sqrt{\text{tr}(\mathbf{X}'\mathbf{X})}. \quad (1.2)$$

The spaces $\Re^{p \times q}$ and \Re^{pq} are isometric in the sense of the metric (1.2).

Proposition 1.3.1. The definitions 1.3.1 and 1.3.2 are equivalent in the sense of metric (1.2).

Proof. Let us apply equality (1.2) on $\frac{d\mathbf{Y}}{d\mathbf{X}}$ and $\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$. We get

$$\left\| \frac{d\mathbf{Y}}{d\mathbf{X}} \right\| = \sqrt{\text{tr}\left(\frac{d\mathbf{Y}'}{d\mathbf{X}} \frac{d\mathbf{Y}}{d\mathbf{X}}\right)} = \sqrt{\sum_{i,j=1}^{rs,pq} \left(\frac{d\mathbf{Y}}{d\mathbf{X}}\right)_{ij}} = \sqrt{\sum_{i,j,k,l=1}^{r,s,p,q} \frac{\partial y_{ij}}{\partial x_{kl}}}$$

and

$$\left\| \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \right\| = \sqrt{\text{tr}\left(\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}\right)} = \sqrt{\sum_{i,j=1}^{pr,qs} \left(\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}\right)_{ij}} = \sqrt{\sum_{i,j,k,l=1}^{r,s,p,q} \frac{\partial y_{ij}}{\partial x_{kl}}}.$$

So we have got that

$$\left\| \frac{d\mathbf{Y}}{d\mathbf{X}} \right\| = \left\| \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \right\|.$$

□

Next we give the functional relation between Neudecker's and MacRae's matrix derivatives.

Proposition 1.3.2. Let \mathbf{X} a $p \times q$ -matrix and \mathbf{Y} a $r \times s$ -matrix. Then

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}} = (\mathbf{I}_{rp} \otimes \text{vec}'(\mathbf{I}_q))(\mathbf{I}_r \otimes \mathbf{K}_{p,q} \left(\frac{d\mathbf{Y}}{d\mathbf{X}}\right)' \mathbf{K}_{r,s} \otimes \mathbf{I}_q)(\text{vec}(\mathbf{I}_r) \otimes \mathbf{I}_{qs}). \quad (1.3)$$

and

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \mathbf{K}_{r,s}(\mathbf{I}_{rs} \otimes \text{vec}(\mathbf{I}_q))(\mathbf{I}_s \otimes \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \otimes \mathbf{I}_p)(\text{vec}(\mathbf{I}_s) \otimes \mathbf{I}_{pq})\mathbf{K}_{q,p}. \quad (1.4)$$

The proofs of equalities (1.3) and (1.4) can be found in Kollo (1991, p. 59-60).

Now we present some properties of Neudecker matrix derivative (for proofs the reader is referred to Kollo (1991, p. 53-58) or Magnus and Neudecker (1999, p. 147-157)).

Proposition 1.3.3. *For a matrix $\mathbf{X} : p \times q$*

$$\frac{d\mathbf{X}}{d\mathbf{X}} = \mathbf{I}_{pq}$$

and

$$\frac{d\mathbf{X}'}{d\mathbf{X}} = \mathbf{K}_{p,q}.$$

Proposition 1.3.4. *If matrix $\mathbf{Y} = \mathbf{U} + \mathbf{V}$, then*

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{d\mathbf{U}}{d\mathbf{X}} + \frac{d\mathbf{V}}{d\mathbf{X}}.$$

Proposition 1.3.5. *If $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{B}$ where $\mathbf{A} : r \times p$ and $\mathbf{B} : q \times s$ are constant matrices, then*

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \mathbf{A}' \otimes \mathbf{B}.$$

Proposition 1.3.6. *Let matrix $\mathbf{Z} : m \times n$ be a function of matrix $\mathbf{Y} : r \times s$ and matrix \mathbf{Y} be a function of matrix $\mathbf{X} : p \times q$. Then*

$$\frac{d\mathbf{Z}}{d\mathbf{X}} = \frac{d\mathbf{Z}}{d\mathbf{Y}} \frac{d\mathbf{Y}}{d\mathbf{X}}.$$

Proposition 1.3.7. *Let \mathbf{Y} and \mathbf{Z} be $r \times s$ and $s \times n$ matrices respectively. Then*

$$\frac{d(\mathbf{Y}\mathbf{Z})}{d\mathbf{X}} = (\mathbf{Z}' \otimes \mathbf{I}_r) \frac{d\mathbf{Y}}{d\mathbf{X}} + (\mathbf{I}_n \otimes \mathbf{Y}) \frac{d\mathbf{Z}}{d\mathbf{X}}.$$

Proposition 1.3.8. *Let $\mathbf{X} : p \times p$ be nonsingular. Then*

$$\frac{d\mathbf{X}^{-1}}{d\mathbf{X}} = -\mathbf{X}^{-1} \otimes (\mathbf{X}^{-1})'.$$

The higher order derivatives are defined recursively. In the case of Definition 1.3.1 (Kollo 1991, p. 69)

$$\frac{d^k \mathbf{Y}}{d\mathbf{X}^k} = \frac{d}{d\mathbf{X}} \frac{d^{k-1}}{d\mathbf{X}^{k-1}} = \underbrace{\frac{d}{d\text{vec}'\mathbf{X}} \otimes \frac{d}{d\text{vec}\mathbf{X}} \otimes \dots \otimes \frac{d}{d\text{vec}\mathbf{X}}}_{k\text{times}} \otimes \text{vec}\mathbf{Y} \quad (1.5)$$

and

$$\frac{\partial^k \mathbf{Y}}{\partial \mathbf{X}^k} = \frac{\partial}{\partial \mathbf{X}} \frac{\partial^{k-1} \mathbf{Y}}{\partial \mathbf{X}^{k-1}}, \quad k > 1, \quad k \in \mathbb{N}$$

in the case of Definition 1.3.2 (Kollo 1991, p. 70).

1.4 Matrix integral

In this section we introduce matrix integral and study its properties. The notion "matrix integral" has been introduced in Pihlak (2004). Beside basic properties of the matrix integral several examples are given to demonstrate practical usage of the notion. The presentation is based on the paper Pihlak (2004).

1.4.1 Definition of matrix integral

First we define the indefinite matrix integral. The definition is based on the MacRae's matrix derivative.

Definition 1.4.1. Let $\mathbf{Z} : rs \times pq$ be a function of $\mathbf{X} : p \times q$. A matrix $\mathbf{Y}(\mathbf{X}) : r \times s$ is called the matrix integral of $\mathbf{Z} = \mathbf{Z}(\mathbf{X}) : rs \times pq$ where $\mathbf{X} : p \times q$, if

$$\frac{\partial \mathbf{Y}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{Z}.$$

The fact that matrix \mathbf{Y} is the matrix integral of a matrix \mathbf{Z} is denoted as

$$\int_{\mathbb{R}^{pq}} \mathbf{Z} \circ d\mathbf{X} = \mathbf{Y}. \quad (1.6)$$

If \mathbf{Y} is a matrix integral of matrix \mathbf{Z} , then also $\mathbf{Y} + \mathbf{C}$ is a matrix integral of \mathbf{Z} , where \mathbf{C} is a constant matrix with the same dimensions as matrix \mathbf{Y} . Definition 1.4.1 is used also to define the definite matrix integral.

Definition 1.4.2. A difference $\int_{\mathbf{A}}^{\mathbf{B}} \mathbf{Z} \circ d\mathbf{X} = \mathbf{Y}(\mathbf{B}) - \mathbf{Y}(\mathbf{A})$ is called the definite matrix integral of matrix \mathbf{Z} from \mathbf{A} to \mathbf{B} .

When the matrix derivative increases the dimensions of the differentiated matrix, then the matrix integral decreases the dimensions of the integrated matrix.

Next we express matrix integral using the star product. For this we shall introduce the differential operator d :

$$d = \begin{pmatrix} dx_{11} & \cdots & dx_{1q} \\ \vdots & \ddots & \vdots \\ dx_{p1} & \cdots & dx_{pq} \end{pmatrix} * \begin{pmatrix} \frac{\partial}{\partial x_{11}} & \cdots & \frac{\partial}{\partial x_{1q}} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{p1}} & \cdots & \frac{\partial}{\partial x_{pq}} \end{pmatrix}$$

$$= \sum_{l=1}^p \sum_{j=1}^q \frac{\partial}{\partial x_{lj}} dx_{lj}$$

where $*$ denotes the star product of matrices. Now we shall introduce the matrix differential.

Definition 1.4.3. *The matrix $d\mathbf{Y} : r \times s$ is called the matrix differential of matrix $\mathbf{Y}(\mathbf{X})$ if*

$$d\mathbf{Y} = d\mathbf{X} * \left(\frac{d}{d\mathbf{X}} \otimes \mathbf{Y} \right).$$

Theorem 1.4.1. *Let $\mathbf{Z} = \frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$. Then*

$$\int_{\mathbb{R}^{pq}} \mathbf{Z} \circ d\mathbf{X} = \int_{\mathbb{R}^{pq}} d\mathbf{X} * \mathbf{Z}.$$

Proof. By means of operator d we can present the matrix differential $d\mathbf{Y}$ in the following way (Pihlak, 2004):

$$d\mathbf{X} * \mathbf{Z} = d\mathbf{X} * \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} = d\mathbf{X} * \left(\frac{d}{d\mathbf{X}} \otimes \mathbf{Y} \right) = \begin{pmatrix} dx_{11} & \cdots & dx_{1q} \\ \vdots & \ddots & \vdots \\ dx_{p1} & \cdots & dx_{pq} \end{pmatrix}$$

$$* \begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \cdots & \frac{\partial y_{1s}}{\partial x_{11}} & \cdots & \frac{\partial y_{11}}{\partial x_{1q}} & \cdots & \frac{\partial y_{1s}}{\partial x_{1q}} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial y_{r1}}{\partial x_{11}} & \cdots & \frac{\partial y_{rs}}{\partial x_{11}} & \cdots & \frac{\partial y_{r1}}{\partial x_{1q}} & \cdots & \frac{\partial y_{rs}}{\partial x_{1q}} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial y_{11}}{\partial x_{p1}} & \cdots & \frac{\partial y_{1s}}{\partial x_{p1}} & \cdots & \frac{\partial y_{11}}{\partial x_{pq}} & \cdots & \frac{\partial y_{1s}}{\partial x_{pq}} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial y_{r1}}{\partial x_{p1}} & \cdots & \frac{\partial y_{rs}}{\partial x_{p1}} & \cdots & \frac{\partial y_{r1}}{\partial x_{pq}} & \cdots & \frac{\partial y_{rs}}{\partial x_{pq}} \end{pmatrix}$$

$$\begin{aligned}
&= \begin{pmatrix} \sum_{i,j=1}^{p,q} \frac{\partial y_{11}}{\partial x_{ij}} dx_{i,j} & \cdots & \sum_{i,j=1}^{p,q} \frac{\partial y_{1s}}{\partial x_{ij}} dx_{i,j} \\ \vdots & \ddots & \vdots \\ \sum_{i,j=1}^{p,q} \frac{\partial y_{r1}}{\partial x_{ij}} dx_{i,j} & \cdots & \sum_{i,j=1}^{p,q} \frac{\partial y_{rs}}{\partial x_{ij}} dx_{i,j} \end{pmatrix} \\
&= \begin{pmatrix} dy_{11} & \cdots & dy_{1s} \\ \vdots & \ddots & \vdots \\ dy_{r1} & \cdots & dy_{rs} \end{pmatrix} = d\mathbf{Y}.
\end{aligned}$$

The matrix integral of \mathbf{Z} we can write down in the following formal form:

$$\mathbf{Y} = \int_{\mathfrak{R}^{pq}} d\mathbf{Y} = \int_{\mathfrak{R}^{pq}} d\mathbf{X} * \mathbf{Z}.$$

Using equality (1.6) we get the statement of theorem. \square

Example 1.4.1 Let us have $\mathbf{X} = (x_1, x_2)$ and $\mathbf{Y} = \begin{pmatrix} x_1 x_2 \\ x_1^2 + \frac{x_1}{x_2} \end{pmatrix}$. Then

$$\begin{aligned}
d\mathbf{X} * \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} &= (dx_1, dx_2) * \begin{pmatrix} \frac{\partial x_1 x_2}{\partial x_1} & \frac{\partial x_1 x_2}{\partial x_2} \\ \frac{\partial (x_1^2 + \frac{x_1}{x_2})}{\partial x_1} & \frac{\partial (x_1^2 + \frac{x_1}{x_2})}{\partial x_2} \end{pmatrix} \\
&= \begin{pmatrix} x_2 dx_1 + x_1 dx_2 \\ (2x_1 + \frac{1}{x_2}) dx_1 - \frac{1}{x_2^2} dx_2 \end{pmatrix} = \begin{pmatrix} dx_1 x_2 \\ d(x_1^2 + \frac{x_1}{x_2}) \end{pmatrix}.
\end{aligned}$$

Now we have

$$\mathbf{Y} = \int_{\mathfrak{R}^{pq}} \begin{pmatrix} d(x_1 x_2) \\ d(x_1^2 + \frac{x_1}{x_2}) \end{pmatrix}.$$

1.4.2 Properties of matrix integral

Let the dimensions of matrices \mathbf{X} and \mathbf{Y} be the same as in Definition 1.4.1. We assume without loss of generality that matrix $\mathbf{C} = \mathbf{0}_{r \times s}$.

Proposition 1.4.1. *Let \mathbf{X} be a $p \times q$ -matrix. Then*

$$\int_{\mathfrak{R}^{pq}} \text{vec} \mathbf{I}_p \text{vec}' \mathbf{I}_q \circ d\mathbf{X} = \mathbf{X}.$$

Proof. According to Definition 1.3.2, matrix $\text{vec} \mathbf{I}_p \text{vec}' \mathbf{I}_q$ is the matrix derivative $\frac{\partial \mathbf{X}}{\partial \mathbf{X}}$. According to Definition 1.4.1 matrix \mathbf{X} is the matrix integral of the matrix $\text{vec} \mathbf{I}_p \text{vec}' \mathbf{I}_q$. \square

Proposition 1.4.2. *Let $\mathbf{Y} = \mathbf{U} + \mathbf{V}$. Then*

$$\int_{\mathbb{R}^{pq}} \mathbf{Y} \circ d\mathbf{X} = \int_{\mathbb{R}^{pq}} \mathbf{U} \circ d\mathbf{X} + \int_{\mathbb{R}^{pq}} \mathbf{V} \circ d\mathbf{X}.$$

Proof. From distributivity of the Kronecker product:

$$(\mathbf{A} + \mathbf{B}) \otimes (\mathbf{C} + \mathbf{D}) = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{A} \otimes \mathbf{D}) + (\mathbf{B} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{D})$$

we obtain

$$\frac{d}{d\mathbf{X}} \otimes \mathbf{Y} = \frac{d}{d\mathbf{X}} \otimes (\mathbf{U} + \mathbf{V}) = \frac{d}{d\mathbf{X}} \otimes \mathbf{U} + \frac{d}{d\mathbf{X}} \otimes \mathbf{V}$$

what completes the proof. \square

Proposition 1.4.3. *Let $\mathbf{A} : pq \times pq$ be a constant diagonal matrix and let $\mathbf{X} : p \times q$. Then*

$$\int_{\mathbb{R}^{pq}} \mathbf{A} \circ d\mathbf{X} = \begin{pmatrix} a_{11}x_{11} \\ a_{22}x_{21} \\ \vdots \\ a_{pp}x_{p1} \\ a_{p+1,p+1}x_{12} \\ \vdots \\ a_{pq,pq}x_{pq} \end{pmatrix}.$$

Proof. The statement follows from the definition of the Kronecker product and from Definition 1.4.1. \square

Next properties are useful when integrating the relation between two density functions. In these properties we assume, that the matrix derivative is defined by Definition 1.3.1.

Let $g(\mathbf{x})$ be a continuous function of a p -vector \mathbf{x} . Assume that $g(\mathbf{x})$ is k times continuously differentiable in an open set A . Denote the derivative of k -th order of the function $g(\mathbf{x})$ by $g^{(k)}(\mathbf{x})$. Let us denote the inner product of p -vectors \mathbf{a} and \mathbf{b} as (\mathbf{a}, \mathbf{b}) :

$$(\mathbf{a}, \mathbf{b}) = \mathbf{a}'\mathbf{b}.$$

For notational conveniences we introduce the vector

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

Let us introduce the scalar differential operator

$$\frac{d}{s d\mathbf{x}} = \sum_{l=1}^p \frac{\partial}{\partial x_l} \quad (1.7)$$

and the operator

$$s d\mathbf{x} := \sum_{l=1}^p dx_l. \quad (1.8)$$

In a set A next properties are valid.

Proposition 1.4.4. *The next relation holds in the notations, given above*

$$\int_{\mathfrak{R}_p} g^{(k+1)}(\mathbf{x}) \circ d\mathbf{x} = (g^{(k)}(\mathbf{x}))'.$$

Proof. If $k = 1$, then the equality follows straightforwardly from Definition 1.4.1. Let the statement be valid for every positive integer k . We have to show that the statement is valid for $k + 1$. By means of equality (1.5), Definition 1.3.1 and Definition 1.4.1 we get

$$\begin{aligned} \int_{\mathfrak{R}_p} g^{(k+1)}(\mathbf{x}) \circ d\mathbf{x} &= \int_{\mathfrak{R}_p} \frac{dg^{(k)}(\mathbf{x})}{d\mathbf{x}} \circ d\mathbf{x} = \int_{\mathfrak{R}_p} \frac{d}{d\mathbf{x}'} g^{(k)}(\mathbf{x}) * d\mathbf{x} \\ &= \frac{d}{d\mathbf{x}'} \int_{\mathfrak{R}_p} g^{(k)}(\mathbf{x}) * d\mathbf{x} = \frac{d}{d\mathbf{x}'} \int_{\mathfrak{R}_p} g^{(k)}(\mathbf{x}) \circ d\mathbf{x} = \frac{d}{d\mathbf{x}'} (g^{(k-1)}(\mathbf{x}))' = (g^{(k)}(\mathbf{x}))'. \end{aligned}$$

□

Proposition 1.4.5. *Let $\mathbf{A} : n \times p^{k-1}$ be a constant matrix and $A_l, l = 1, 2, \dots, n$ the l -th row-vector of this matrix. Then*

$$\int_{\mathfrak{R}^p} \mathbf{A} g^{(k)}(\mathbf{x}) \circ d\mathbf{x}' = \begin{pmatrix} (A_1, \text{vec} g^{(k-1)}(\mathbf{x})) \\ \vdots \\ (A_n, \text{vec} g^{(k-1)}(\mathbf{x})) \end{pmatrix}.$$

Proof. Let us apply the operator $\frac{d}{d\mathbf{x}'}$ to the right hand side of the statement. According to Definition 1.4.1 we have to get matrix $\mathbf{A} g^{(k)}(\mathbf{x})$ as a result. Notice that the matrices

$$\frac{d}{d\mathbf{x}'} \otimes \begin{pmatrix} (A_1, \text{vec} g^{(k-1)}(\mathbf{x})) \\ \vdots \\ (A_n, \text{vec} g^{(k-1)}(\mathbf{x})) \end{pmatrix}$$

and $\mathbf{A}g^{(k)}$ have the same dimensions: they both are $n \times p$ -matrices. Now we must show that the corresponding elements are equal. We obtain that

$$\left(\frac{d}{d\mathbf{x}'} \otimes \begin{pmatrix} (A_1, \text{vec}g^{(k-1)}(\mathbf{x})) \\ \vdots \\ (A_n, \text{vec}g^{(k-1)}(\mathbf{x})) \end{pmatrix}\right)_{lm} = \sum_{j=1}^{p^{k-1}} a_{lj}(g^{(k)}(\mathbf{x}))_{jm} = (\mathbf{A}g^{(k)}(\mathbf{x}))_{lm}$$

where $l = 1, 2, \dots, n$ and $m = 1, 2, \dots, p$. Thus the matrices are equal. \square

Proposition 1.4.6. *Let a be a scalar not depending on vector \mathbf{x} . Then*

$$\int_{\mathbb{R}^p} \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_p\right) ag(\mathbf{x}) \circ ({}_s d\mathbf{x}) = ag(\mathbf{x}).$$

Proof. The scalar differential operator (1.7) can be presented as an inner product:

$$\frac{d}{d\mathbf{x}} = \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_p\right).$$

Thus applying equalities (1.7) and (1.8) we get

$$\int_{\mathbb{R}^p} \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_p\right) ag(\mathbf{x}) * ({}_s d\mathbf{x}) = \int_{\mathbb{R}^p} \frac{d}{d\mathbf{x}} ag(\mathbf{x}) * ({}_s d\mathbf{x}) = \int_{\mathbb{R}^p} dag(\mathbf{x}) = ag(\mathbf{x}).$$

\square

The next property is a generalization of Proposition 1.4.6.

Proposition 1.4.7. *Let a be a scalar not depending on vector \mathbf{x} . Then*

$$\int_{\mathbb{R}^p} \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_{p^k}\right)^{\otimes k} ag^{(k)}(\mathbf{x}) \circ ({}_s d\mathbf{x}) = \int_{\mathbb{R}^p} \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_{p^{k-1}}\right)^{\otimes k-1} ag^{(k-1)}(\mathbf{x})$$

where $k = 1, 2, \dots$

Proof. Using Propositions 1.4.4 and 1.4.6 we get

$$\begin{aligned} \int_{\mathbb{R}^p} \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_{p^k}\right)^{\otimes k} ag^{(k)}(\mathbf{x}) \circ ({}_s d\mathbf{x}) &= \int_{\mathbb{R}^p} \left(\frac{d}{d\mathbf{x}'} \otimes \frac{d}{d\mathbf{x}'}^{\otimes k-1}, \mathbf{1}_{p^k}\right) ag^{(k)}(\mathbf{x}) \circ ({}_s d\mathbf{x}) \\ &= \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_{p^{k-1}}\right)^{\otimes k-1} \int_{\mathbb{R}^p} \frac{d}{d\mathbf{x}} ag^k(\mathbf{x}) * ({}_s d\mathbf{x}) = \left(\frac{d}{d\mathbf{x}'}, \mathbf{1}_{p^{k-1}}\right)^{\otimes k-1} ag^{(k-1)}(\mathbf{x}). \end{aligned}$$

So we have got the statement of proposition. \square

Proposition 1.4.8. *Let functions g and G be such that*

$$g(\mathbf{x}) = \frac{\partial^p G(\mathbf{x})}{\partial x_1 \dots \partial x_p}.$$

Then the following equality holds

$$\underbrace{\int \dots \int}_p g(\mathbf{x}) dx_1 \dots dx_p = \int_{\mathbb{R}^p} \frac{d}{{}_s d\mathbf{x}} G(\mathbf{x}) \circ ({}_s d\mathbf{x}).$$

Proof. By integrating the left hand side of the statement we obtain

$$\underbrace{\int \dots \int}_p \frac{\partial G(\mathbf{x})}{\partial x_1 \dots \partial x_p} dx_1 \dots dx_p = G(\mathbf{x}).$$

Applying the matrix integral to the right hand side of the statement we obtain

$$\int_{\mathbb{R}^p} \frac{d}{{}_s d\mathbf{x}} G(\mathbf{x}) * ({}_s d\mathbf{x}) = G(\mathbf{x}).$$

□

From Proposition 1.4.8 follows the next property.

Proposition 1.4.9. *Let \mathbf{a} be a constant p -vector. Then*

$$\underbrace{\int \dots \int}_p \left(\mathbf{a}, \frac{d}{d\mathbf{x}}^{\otimes k} \right) g(\mathbf{x}) dx_1 \dots dx_p = \left(\mathbf{a}, \frac{d}{d\mathbf{x}}^{\otimes k} \right) G(\mathbf{x}).$$

Proof. From statements of Propositions 1.4.9 and 1.4.8 we get

$$\begin{aligned} \underbrace{\int \dots \int}_p \left(\mathbf{a}, \frac{d}{d\mathbf{x}}^{\otimes k} \right) g(\mathbf{x}) dx_1 \dots dx_p &= \left(\mathbf{a}, \frac{d}{d\mathbf{x}}^{\otimes k} \right) \int_{\mathbb{R}^p} \frac{d}{{}_s d\mathbf{x}} G(\mathbf{x}) * ({}_s d\mathbf{x}) \\ &= \left(\mathbf{a}, \frac{d}{d\mathbf{x}}^{\otimes k} \right) G(\mathbf{x}). \end{aligned}$$

So we have proved the statement. □

Let us demonstrate how to find matrix integral. Firstly we give an example of indefinite matrix integral.

Example 1.4.2 Let $\mathbf{X} = (x_1, x_2)$ and

$$\mathbf{Z} = \begin{pmatrix} x_1^2 & x_2^3 \\ 2x_1 & 2x_2 \end{pmatrix}$$

be the matrix derivative of matrix \mathbf{Y} by matrix \mathbf{X} . We can find matrix \mathbf{Y} as the integral:

$$\begin{aligned} \mathbf{Y} &= \int_{\mathfrak{R}^2} d\mathbf{X} * \mathbf{Z} = \int_{\mathfrak{R}^2} (dx_1, dx_2) * \begin{pmatrix} x_1^2 & x_2^3 \\ 2x_1 & 2x_2 \end{pmatrix} \\ &= \int_{\mathfrak{R}^2} \begin{pmatrix} x_1^2 dx_1 + x_2^3 dx_2 \\ 2x_1 dx_1 + 2x_2 dx_2 \end{pmatrix} = \begin{pmatrix} \frac{x_1^3}{3} + \frac{x_2^4}{4} + c_1 \\ x_1^2 + x_2^2 + c_2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{x_1^3}{3} + \frac{x_2^4}{4} \\ x_1^2 + x_2^2 \end{pmatrix} + \mathbf{C} \end{aligned}$$

where

$$\mathbf{C} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

In the next example a definite matrix integral is found.

Example 1.4.3 Let us have $\mathbf{X} = (x_1, x_2, x_3)$,

$$\mathbf{Z} = \begin{pmatrix} x_3 \cos x_1 \cos x_2 & -x_3 \sin x_1 \sin x_2 & \sin x_1 \cos x_2 \\ x_3 \cos x_1 \sin x_2 & x_3 \sin x_1 \cos x_2 & \sin x_1 \sin x_2 \\ -x_3 \sin x_1 & 0 & \cos x_1 \end{pmatrix}$$

$\mathbf{A} = (-\frac{\pi}{2}, -\frac{\pi}{2}, -1)$ and $\mathbf{B} = (\frac{\pi}{2}, \frac{\pi}{2}, 1)$. We shall find the definite matrix integral $\int_{\mathbf{A}}^{\mathbf{B}} \mathbf{Z} \circ d\mathbf{X}$. We obtain

$$\begin{aligned} &\int_{\mathbf{A}}^{\mathbf{B}} \mathbf{Z} \circ d\mathbf{X} \\ &= \int_{\mathbf{A}}^{\mathbf{B}} (dx_1, dx_2, dx_3) * \begin{pmatrix} x_3 \cos x_1 \cos x_2 & -x_3 \sin x_1 \sin x_2 & \sin x_1 \cos x_2 \\ x_3 \cos x_1 \sin x_2 & x_3 \sin x_1 \cos x_2 & \sin x_1 \sin x_2 \\ -x_3 \sin x_1 & 0 & \cos x_1 \end{pmatrix} \\ &= \int_{\mathbf{A}}^{\mathbf{B}} \begin{pmatrix} x_3 \cos x_1 \cos x_2 dx_1 - x_3 \sin x_1 \sin x_2 dx_2 + \sin x_1 \cos x_2 dx_3 \\ x_3 \cos x_1 \sin x_2 dx_1 + x_3 \sin x_1 \cos x_2 dx_2 + \sin x_1 \sin x_2 dx_3 \\ -x_3 \sin x_1 dx_1 + \cos x_1 dx_3 \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} \sin \frac{\pi}{2} \cos \frac{\pi}{2} \\ \sin \frac{\pi}{2} \sin \frac{\pi}{2} \\ 2 \cos \frac{\pi}{2} \end{pmatrix} - \begin{pmatrix} -\sin(-\frac{\pi}{2}) \cos(-\frac{\pi}{2}) \\ -\sin(-\frac{\pi}{2}) \sin(-\frac{\pi}{2}) \\ -2 \cos(-\frac{\pi}{2}) \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix}.$$

Let us have functions $g(\mathbf{x}) = x_1^2 + x_2^2$ and $G(\mathbf{x}) = \frac{1}{3}(x_1^3 x_2 + x_2^3 x_1)$. It is easy to detect that

$$g(\mathbf{x}) = \frac{\partial^2 G(\mathbf{x})}{\partial x_1 \partial x_2}.$$

In the next examples we take $\mathbf{C} = \mathbf{0}_2$.

Example 1.4.4 Using the equality (1.5) we get

$$g^{(2)}(\mathbf{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

and

$$g^{(1)}(\mathbf{x}) = (2x_1 \quad 2x_2).$$

Applying the star product we get

$$\int_{\mathbb{R}^p} g^{(2)}(\mathbf{x}) \circ d\mathbf{x} = \int_{\mathbb{R}^p} dx * g^{(2)}(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = (g^{(1)}(\mathbf{x}))'.$$

Example 1.4.5 Let us have matrix $\mathbf{A} = (1 \quad 2 \quad 3)'$. Then

$$\int_{\mathbb{R}^p} \mathbf{A} g^{(1)}(\mathbf{x}) \circ d\mathbf{x}' = \int_{\mathbb{R}^p} dx' * \mathbf{A} g^{(1)}(\mathbf{x}) = \begin{pmatrix} x_1^2 + x_2^2 \\ 2x_1^2 + 2x_2^2 \\ 3x_1^2 + 3x_2^2 \end{pmatrix}.$$

Example 1.4.6 Let us have vector $\mathbf{a} = (1, -1)$. Then

$$\begin{aligned} \int \int (\mathbf{a}, \frac{d}{d\mathbf{x}}) g(\mathbf{x}) dx_1 dx_2 &= \left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2} \right) \int \int (x_1^2 + x_2^2) dx_1 dx_2 \\ &= \left(\frac{\partial}{\partial x_1} - \frac{\partial}{\partial x_2} \right) \frac{1}{3} (x_1^3 x_2 + x_2^3 x_1) = \frac{\partial G(\mathbf{x})}{\partial x_1} - \frac{\partial G(\mathbf{x})}{\partial x_2} = (\mathbf{a}, \frac{d}{d\mathbf{x}}) G(\mathbf{x}). \end{aligned}$$

Chapter 2

Approximation of multivariate distributions

In this chapter we study approximation of multivariate distributions by means of Taylor series. The first section presents a general method for density approximation developed by Kollo and von Rosen (1995, 1998). The second and third section are based on results of Pihlak (2006a). These results are presented at 9th International Vilnius Conference on Probability Theory and Mathematical Statistics (Pihlak, 2006b). In the second section we apply matrix integral to integration of the equality between two density functions. In the third section we use Hermite matrix-polynomials to approximate an unknown distribution function through the distribution function of the normal distribution. Finally we give some simulation examples to examine the goodness of approximations.

2.1 Approximation of multivariate density functions

2.1.1 Cumulants

First we study cumulants of a random vector. As will be seen later, the coefficients in equality (1) in Introduction depend on cumulants.

Let us have a random p -vector \mathbf{X} with coordinates X_i , $i = 1, 2, \dots, p$. Let \mathbf{x} be a realization of this vector. The characteristic function of the random

vector \mathbf{X} is defined as follows:

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E(e^{it'\mathbf{X}}), \quad \mathbf{t} \in \mathfrak{R}^p.$$

The cumulant function of the random vector \mathbf{X} is given by equality:

$$\phi_{\mathbf{X}}(\mathbf{t}) = \ln(\varphi_{\mathbf{X}}(\mathbf{t})), \quad \mathbf{t} \in \mathfrak{R}^p.$$

The k -th order cumulant $c_k(\mathbf{X})$ of \mathbf{X} is the k -th matrix derivative of the cumulant function (Kollo and von Rosen, 2005 p. 181):

$$c_k(\mathbf{X}) = \frac{1}{i^k} \frac{d^k \phi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}^k} \Big|_{\mathbf{t}=\mathbf{0}_p}. \quad (2.1)$$

Applying Taylor expansion of function $g(\mathbf{x})$ in matrix form (Kollo, 1991, p. 74)

$$g(\mathbf{x}) = g(\mathbf{x}_0) + \sum_{k=1}^m ((\mathbf{x} - \mathbf{x}_0)')^{\otimes k-1} \frac{d^k g(\mathbf{x})}{d\mathbf{x}^k} \Big|_{\mathbf{x}=\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) + r_m$$

we get representation of $\phi_{\mathbf{X}}(\mathbf{t})$ through cumulants:

$$\phi_{\mathbf{X}}(\mathbf{t}) = 1 + \sum_{k=1}^m \frac{i^k}{k!} (\mathbf{t}')^{\otimes k-1} c_k(\mathbf{X}) \mathbf{t} + r_m$$

where r_m is the remainder term of the series.

Using properties of matrix derivative and assuming that $\varphi_{\mathbf{X}}(\mathbf{t})$ is continuously differentiable by \mathbf{t} we get

$$\begin{aligned} c_1(\mathbf{X}) &= \frac{1}{i} \frac{d \ln(\varphi_{\mathbf{X}}(\mathbf{t}))}{d\mathbf{t}} \Big|_{\mathbf{t}=\mathbf{0}_p} \\ &= \frac{1}{i} \frac{d \ln(\varphi_{\mathbf{X}}(\mathbf{t}))}{d\varphi_{\mathbf{X}}(\mathbf{t})} \frac{dE(e^{it'\mathbf{X}})}{d\mathbf{t}} \Big|_{\mathbf{t}=\mathbf{0}_p} = \frac{1}{i} \varphi_{\mathbf{X}}^{-1}(\mathbf{t}) E\left(\frac{de^{it'\mathbf{X}}}{d\mathbf{t}}\right) \Big|_{\mathbf{t}=\mathbf{0}_p} \\ &= \frac{1}{i} \varphi_{\mathbf{X}}^{-1}(\mathbf{t}) i E(e^{it'\mathbf{X}} \mathbf{X}') \Big|_{\mathbf{t}=\mathbf{0}_p} = E(\mathbf{X}'). \end{aligned}$$

The second cumulant is obtained in the similar way:

$$c_2(\mathbf{X}) = \frac{1}{i^2} \frac{d(\varphi_{\mathbf{X}}^{-1} \frac{d\varphi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}})}{d\mathbf{t}} \Big|_{\mathbf{t}=\mathbf{0}_p}$$

$$\begin{aligned}
&= \frac{1}{i^2} \left(\left(\frac{d\varphi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}} \right)' \frac{d(\varphi_{\mathbf{X}}^{-1})}{d\mathbf{t}} + (\mathbf{I}_p \otimes \varphi_{\mathbf{X}}^{-1}(\mathbf{t})) \frac{d^2\varphi_{\mathbf{X}}(\mathbf{t})}{d\mathbf{t}^2} \right) \Big|_{\mathbf{t}=\mathbf{0}_p} \\
&= \frac{1}{i^2} \left((\mathbf{I}_p \otimes \varphi_{\mathbf{X}}^{-1}(\mathbf{t})) - E(e^{i\mathbf{t}'\mathbf{X}} i\mathbf{X}') \varphi_{\mathbf{X}}^{-2} E(e^{i\mathbf{t}'\mathbf{X}} i\mathbf{X}') \right) \Big|_{\mathbf{t}=\mathbf{0}_p} \\
&= \frac{1}{i^2} i^2 (\mathbf{I}_p E(\mathbf{X}\mathbf{X}') - E(\mathbf{X})E(\mathbf{X}')) = E(\mathbf{X}\mathbf{X}') - E(\mathbf{X})E(\mathbf{X}') = D(\mathbf{X}).
\end{aligned}$$

So we have got that

$$c_1(\mathbf{X}) = E(\mathbf{X}') \quad (2.2)$$

and

$$c_2(\mathbf{X}) = D(\mathbf{X}). \quad (2.3)$$

2.1.2 Relation between two density functions

Let us consider a scalar function $g(\mathbf{X})$. Assume that the function g is m times continuously differentiable. Then we can express function $g(\mathbf{X})$ at point \mathbf{X}_0 as a Taylor expansion (Kollo, 1991 p. 95):

$$g(\mathbf{X}) = g(\mathbf{X}_0) + \sum_{k=1}^m ((\mathbf{X} - \mathbf{X}_0)')^{\otimes k-1} \frac{d^k g(\mathbf{X})}{d\mathbf{X}^k} \Big|_{\mathbf{X}=\mathbf{X}_0} (\mathbf{X} - \mathbf{X}_0) + R_m$$

where R_m is the remainder term of the series.

Let $f_{\mathbf{X}}(\mathbf{x})$ be the probability density function of a random vector \mathbf{X} . Assume that $f_{\mathbf{X}}(\mathbf{x})$ is uniformly continuous and continuously differentiable necessary number of times by argument \mathbf{x} .

The next two lemmas give us relations between the characteristic function $\varphi_{\mathbf{X}}(\mathbf{t})$ and the probability density function $f_{\mathbf{X}}(\mathbf{x})$.

Lemma 2.1.1. *Assume that $f_{\mathbf{X}}(\mathbf{x})$ is k times differentiable. Then the next relation is valid between derivatives of the probability density function $f_{\mathbf{X}}(\mathbf{x})$ and the characteristic function $\varphi_{\mathbf{X}}(\mathbf{t})$:*

$$\varphi_{\mathbf{X}}(\mathbf{t})(i\mathbf{t})^{\otimes k-1}(i\mathbf{t})' = (-1)^k \int_{\mathfrak{R}^p} e^{i\mathbf{t}'\mathbf{x}} \frac{d^k f_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^k} d\mathbf{x}.$$

Proof. We have to check first that the dimensions of the matrices on the right and left hand side are equal, and then to show that corresponding elements of these matrices are equal. According to the definitions of multiplication and Kronecker product of matrices, both sides of the equality are

$p^{k-1} \times p$ -matrices. To check the equality of the corresponding elements of the matrices we can show that

$$\varphi_{\mathbf{X}}(\mathbf{t}) i^k t_{l_1} t_{l_2} \dots t_{l_k} = (-1)^k \int_{\mathfrak{R}_p} e^{i\mathbf{t}'\mathbf{x}} \frac{\partial^k f_{\mathbf{X}}(\mathbf{x})}{\partial x_{l_1} \dots \partial x_{l_k}} d\mathbf{x}.$$

The detailed proof can be found in Kollo (1991, p. 91-92). \square

Lemma 2.1.2. *Let \mathbf{X} and \mathbf{a} be random p -vector and a non-random p -vector, respectively. Then the Fourier inverse transform of the quantity $\mathbf{a}'(i\mathbf{t})^{\otimes k} \varphi_{\mathbf{X}}(\mathbf{t})$ is of the form:*

$$(-1)^k \mathbf{a}' \text{vec} \frac{d^k f_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^k} = (2\pi)^{-p} \int_{\mathfrak{R}_p} \varphi_{\mathbf{X}}(\mathbf{t}) \mathbf{a}'(i\mathbf{t})^{\otimes k} e^{-i\mathbf{t}'\mathbf{x}} d\mathbf{t}.$$

Proof. According to Lemma 2.1.1 the product $\varphi_{\mathbf{X}}(\mathbf{t})(i\mathbf{t})^{\otimes k-1}(i\mathbf{t})'$ is the Fourier transform of the quantity $(-1)^k \frac{d^k f_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^k}$. After vectorization we get that $\varphi_{\mathbf{X}}(\mathbf{t})(i\mathbf{t})^{\otimes k}$ is the Fourier transform of the vector $(-1)^k \text{vec} \frac{d^k f_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^k}$. The inverse transform action of this vector can be expressed as follows:

$$(-1)^k \text{vec} \frac{d^k f_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^k} = (2\pi)^{-p} \int_{\mathfrak{R}_p} \varphi_{\mathbf{X}}(\mathbf{t})(i\mathbf{t})^{\otimes k} e^{-i\mathbf{t}'\mathbf{x}} d\mathbf{t}.$$

Multiplying both sides of this equality by vector \mathbf{a}' , we get the statement of the lemma. \square

These two lemmas are the key results for finding a relation between two probability density functions. Let \mathbf{X} and \mathbf{Y} be random p -vectors with probability density functions $f_{\mathbf{X}}(\mathbf{x})$ and $f_{\mathbf{Y}}(\mathbf{y})$ respectively. Let these random vectors have the characteristic functions $\varphi_{\mathbf{X}}(\mathbf{t})$ and $\varphi_{\mathbf{Y}}(\mathbf{t})$ where $\mathbf{t} \in \mathfrak{R}^p$. Let us denote the k th order derivative of the function $f_{\mathbf{X}}(\mathbf{x})$ by $f_{\mathbf{X}}^{(k)}(\mathbf{x})$. In these notations the next result holds (Kollo and von Rosen, 1995).

Theorem 2.1.1. *The probability density function $f_{\mathbf{Y}}(\mathbf{x})$ is expressed through the probability density function $f_{\mathbf{X}}(\mathbf{x})$ as follows:*

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{x}) &= f_{\mathbf{X}}(\mathbf{x}) - (E(\mathbf{Y}) - E(\mathbf{X}))' \text{vec} f_{\mathbf{X}}^{(1)}(\mathbf{x}) \\ &+ \frac{1}{2} \text{vec}' \{D(\mathbf{Y}) - D(\mathbf{X}) + (E(\mathbf{Y}) - E(\mathbf{X}))(E(\mathbf{Y}) - E(\mathbf{X}))'\} \text{vec} f_{\mathbf{X}}^{(2)}(\mathbf{x}) \\ &- \frac{1}{6} \{ \text{vec}' [(c_3(\mathbf{Y}) - c_3(\mathbf{X})) + 3 \text{vec}' (D(\mathbf{Y}) - D(\mathbf{X})) \otimes (E(\mathbf{Y}) - E(\mathbf{X}))] \\ &\quad + (E(\mathbf{Y}) - E(\mathbf{X}))'^{\otimes 3} \} \text{vec} f_{\mathbf{X}}^{(3)}(\mathbf{x}) + \dots \end{aligned} \quad (2.4)$$

To demonstrate the idea of getting the result we shall present also the proof.

Proof. Let us start from a trivial identity

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = \frac{\varphi_{\mathbf{Y}}(\mathbf{t})}{\varphi_{\mathbf{X}}(\mathbf{t})} \varphi_{\mathbf{X}}(\mathbf{t}). \quad (2.5)$$

Using Taylor expansions of the characteristic functions $\varphi_{\mathbf{X}}(\mathbf{t})$ and $\varphi_{\mathbf{Y}}(\mathbf{t})$ and equality (2.1) we get from identity (2.5):

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{t}) \prod_{k=1}^{\infty} e^{(\frac{1}{k!}(\mathbf{t}')^{\otimes k-1}(c_k(\mathbf{Y})-c_k(\mathbf{X}))\mathbf{t})}.$$

Applying the Taylor expansion of the exponential function we get the next relation after ordering terms by i^k :

$$\begin{aligned} \varphi_{\mathbf{Y}}(\mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{t}) \{ & 1 + i(c_1(\mathbf{Y}) - c_1(\mathbf{X}))\mathbf{t} + \frac{i^2}{2}\mathbf{t}'[c_2(\mathbf{Y}) - c_2(\mathbf{X}) \\ & + (c_1(\mathbf{Y}) - c_1(\mathbf{X}))'(c_1(\mathbf{Y}) - c_1(\mathbf{X}))]\mathbf{t} \\ & + \frac{i^3}{6}[(\mathbf{t}')^{\otimes 2}(c_3(\mathbf{Y}) - c_3(\mathbf{X}) + (c_1(\mathbf{Y}) - c_1(\mathbf{X}))' \\ & \otimes (c_1(\mathbf{Y}) - c_1(\mathbf{X}))(c_1(\mathbf{Y}) - c_1(\mathbf{X})))\mathbf{t} \\ & + 3(c_1(\mathbf{Y}) - c_1(\mathbf{X}))\mathbf{t}\mathbf{t}'(c_2(\mathbf{Y}) - c_2(\mathbf{X}))\mathbf{t}] + \dots \}. \end{aligned}$$

Using equality (2.1) we get

$$\begin{aligned} \varphi_{\mathbf{Y}}(\mathbf{t}) = \varphi_{\mathbf{X}}(\mathbf{t}) \{ & 1 + i(c_1(\mathbf{Y}) - c_1(\mathbf{X}))\mathbf{t} + \frac{i^2}{2}\text{vec}'(c_2(\mathbf{Y}) - c_2(\mathbf{X})) \\ & + (c_1(\mathbf{Y}) - c_1(\mathbf{X}))'(c_1(\mathbf{Y}) - c_1(\mathbf{X}))\mathbf{t}^{\otimes 2} + \frac{i^3}{6}\{\text{vec}'[(c_3(\mathbf{Y}) - c_3(\mathbf{X})) \\ & + 3\text{vec}'(c_2(\mathbf{Y}) - c_2(\mathbf{X})) \otimes (c_1(\mathbf{Y}) - c_1(\mathbf{X}))] + (c_1(\mathbf{Y}) - c_1(\mathbf{X}))^{\otimes 3}\mathbf{t}^{\otimes 3} + \dots \}. \end{aligned}$$

Applying the Fourier inverse transform from Lemma 2.1.2 to this equality we get the relation between density functions. From equalities (2.2) and (2.3) we have $c_1(\mathbf{X}) = E(\mathbf{X}')$ and $c_2(\mathbf{X}) = D(\mathbf{X})$. Replacing these equalities into the expansion we get the statement of the theorem. \square

2.2 Approximation of distribution functions

In this section we study integration of equality (2.4). For integration we use the matrix integral.

2.2.1 Integration of the relation between densities

In this section we apply the matrix integral to integration of the relation (2.4) between multivariate density functions $f_{\mathbf{X}}(\mathbf{x})$ and $f_{\mathbf{Y}}(\mathbf{y})$. Let us write the equality (2.4) in the form

$$f_{\mathbf{Y}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) - \mathbf{a}' \text{vec} f_{\mathbf{X}}^{(1)}(\mathbf{x}) + \text{vec}' \mathbf{B} \text{vec} f_{\mathbf{X}}^{(2)}(\mathbf{x}) - \text{vec}' \mathbf{C} \text{vec} f_{\mathbf{X}}^{(3)}(\mathbf{x}) + \dots \quad (2.6)$$

where p -vector

$$\mathbf{a} = (E(\mathbf{X}) - E(\mathbf{Y})), \quad (2.7)$$

$p \times p$ -matrix

$$\mathbf{B} = \frac{1}{2}[(\mathbf{Y}) - D(\mathbf{X}) + (E(\mathbf{Y}) - E(\mathbf{X}))(E(\mathbf{Y}) - E(\mathbf{X}))'] \quad (2.8)$$

and $p^2 \times p$ -matrix

$$\begin{aligned} \mathbf{C} = \frac{1}{6}[(c_3(\mathbf{Y}) - c_3(\mathbf{X})) + 3(D(\mathbf{Y}) - D(\mathbf{X})) \otimes (E(\mathbf{Y}) - E(\mathbf{X})) \\ + (E(\mathbf{Y}) - E(\mathbf{X}))^{\otimes 2} (E(\mathbf{Y}) - E(\mathbf{X}))']. \end{aligned} \quad (2.9)$$

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

Let us introduce the operators

$$\frac{d}{d\mathbf{x}} := \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_p} \right)',$$

$$\frac{d}{d\mathbf{x}}^{\otimes 2} := \left(\frac{\partial^2}{\partial x_1^2}, \frac{\partial^2}{\partial x_1 x_2}, \dots, \frac{\partial^2}{\partial x_p^2} \right)'$$

and

$$\frac{d}{d\mathbf{x}}^{\otimes 3} := \left(\frac{\partial^3}{\partial x_1^3}, \frac{\partial^3}{\partial x_1^2 x_2}, \dots, \frac{\partial^3}{\partial x_p^3} \right)'.$$

It is straightforward to see that $\frac{d}{d\mathbf{x}}$, $\frac{d}{d\mathbf{x}}^{\otimes 2}$ and $\frac{d}{d\mathbf{x}}^{\otimes 3}$ are p -, p^2 - and p^3 -vectors respectively. Applying the operators $\frac{d}{d\mathbf{x}}$, $\frac{d}{d\mathbf{x}}^{\otimes 2}$ and $\frac{d}{d\mathbf{x}}^{\otimes 3}$ to $f_{\mathbf{X}}(\mathbf{x})$ we get

$$\frac{d}{d\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) = \left(\frac{\partial f_{\mathbf{X}}(\mathbf{x})}{\partial x_1}, \frac{\partial f_{\mathbf{X}}(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f_{\mathbf{X}}(\mathbf{x})}{\partial x_p} \right)' = \text{vec} f_{\mathbf{X}}^{(1)}(\mathbf{x}),$$

$$\frac{d}{d\mathbf{x}}^{\otimes 2} = \left(\frac{\partial^2 f_{\mathbf{X}}(\mathbf{x})}{\partial x_1^2}, \frac{\partial^2 f_{\mathbf{X}}(\mathbf{x})}{\partial x_1 x_2}, \dots, \frac{\partial^2 f_{\mathbf{X}}(\mathbf{x})}{\partial x_1 x_p}, \dots, \frac{\partial^2 f_{\mathbf{X}}(\mathbf{x})}{\partial x_p^2} \right)' = \text{vec} f_{\mathbf{X}}^{(2)}(\mathbf{x})$$

and

$$\frac{d}{d\mathbf{x}}^{\otimes 3} = \left(\frac{\partial^3 f_{\mathbf{X}}(\mathbf{x})}{\partial x_1^3}, \frac{\partial^3 f_{\mathbf{X}}(\mathbf{x})}{\partial x_1^2 x_2}, \dots, \frac{\partial^3 f_{\mathbf{X}}(\mathbf{x})}{\partial x_1^2 x_p}, \dots, \frac{\partial^3 f_{\mathbf{X}}(\mathbf{x})}{\partial x_p^3} \right)' = \text{vec} f_{\mathbf{X}}^{(3)}(\mathbf{x}).$$

We rewrite the equality (2.6) in the form

$$f_{\mathbf{Y}}(\mathbf{x}) = \left(1 - \left(\mathbf{a}, \frac{d}{d\mathbf{x}}\right) + \left(\text{vec}\mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}\right) - \left(\text{vec}\mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}\right) + \dots\right) f_{\mathbf{X}}(\mathbf{x})$$

For distribution functions we formulate the following result.

Theorem 2.2.1. *Let \mathbf{X} and \mathbf{Y} be random p -vectors. Then the next relation holds between the probability distribution functions $F_{\mathbf{Y}}(\mathbf{x})$ and $F_{\mathbf{X}}(\mathbf{x})$:*

$$F_{\mathbf{Y}}(\mathbf{x}) = \left(1 - \left(\mathbf{a}, \frac{d}{d\mathbf{x}}\right) + \left(\text{vec}\mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}\right) - \left(\text{vec}\mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}\right) + \dots\right) F_{\mathbf{X}}(\mathbf{x}) \quad (2.10)$$

Proof. Between multivariate density function $f_{\mathbf{X}}(\mathbf{x})$ and distribution function $F_{\mathbf{X}}(\mathbf{x})$ the next relation holds:

$$\frac{\partial^p F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \dots \partial x_p} = f_{\mathbf{X}}(\mathbf{x}). \quad (2.11)$$

Using Propositions 1.4.8, 1.4.9 and equality (2.11) we get

$$\int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} f_{\mathbf{X}}(\mathbf{u}) du_1 \dots du_p = \frac{d}{s du} F_{\mathbf{X}}(\mathbf{u}) \circ (s d\mathbf{u}) = F_{\mathbf{X}}(\mathbf{x})$$

and

$$\begin{aligned} & \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} \left(\mathbf{a}, \frac{d}{d\mathbf{x}}\right) f_{\mathbf{X}}(\mathbf{u}) du_1 \dots du_p \\ &= \left(\mathbf{a}, \frac{d}{d\mathbf{x}}\right) \int_{(-\infty, x_1) \times \dots \times (-\infty, x_p)} \frac{d}{s d\mathbf{u}} F_{\mathbf{X}}(\mathbf{u}) \circ (s d\mathbf{u}) = \left(\mathbf{a}, \frac{d}{d\mathbf{x}}\right) F_{\mathbf{X}}(\mathbf{x}). \end{aligned}$$

Applying the same principles to $\left(\text{vec}\mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}\right) f_{\mathbf{X}}(\mathbf{x})$ and $\left(\text{vec}\mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}\right) f_{\mathbf{X}}(\mathbf{x})$ we get

$$\int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} \left(\text{vec}\mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}\right) f_{\mathbf{X}}(\mathbf{u}) du_1 \dots du_p = \left(\text{vec}\mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}\right) F_{\mathbf{X}}(\mathbf{x})$$

and

$$\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} (\text{vec} \mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}) f_{\mathbf{X}}(\mathbf{u}) du_1 \dots du_p = (\text{vec} \mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}) F_{\mathbf{X}}(\mathbf{x}).$$

So we have proved the equality (2.10). \square

We shall examine the equality (2.10) term by term. That means we study the terms $(\mathbf{a}, \frac{d}{d\mathbf{x}}) F_{\mathbf{X}}(\mathbf{x})$, $(\text{vec} \mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}) F_{\mathbf{X}}(\mathbf{x})$ and $(\text{vec} \mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}) F_{\mathbf{X}}(\mathbf{x})$ one by one. Results of this study are formulated as next three lemmas.

Lemma 2.2.1. *Let the vector functions $f(\mathbf{x})$ and $F(\mathbf{x})$ be such that (2.11) holds and let \mathbf{a} be a constant p -vector. Then*

$$\int_{\mathbb{R}^p} (\mathbf{a}, \frac{d}{d\mathbf{x}}) f(\mathbf{x}) \circ ({}_s d\mathbf{x}) = a_i f(\mathbf{x}) - \sum_{j=1}^p (a_i - a_j) \frac{\partial F(\mathbf{x})}{\partial x_j}, \quad i = 1, 2, \dots, p.$$

Proof. Let us take the i th component of vector \mathbf{a} . We can add and subtract to $(\mathbf{a}, \frac{d}{d\mathbf{x}}) f(\mathbf{x})$ the terms $a_i \frac{\partial f(\mathbf{x})}{\partial x_j}$, $j = 1, 2, \dots, p$, $j \neq i$. So we get

$$(\mathbf{a}, \frac{d}{d\mathbf{x}}) f(\mathbf{x}) = a_i \sum_{j=1}^p \frac{\partial f(\mathbf{x})}{\partial x_j} + \sum_{j=1}^p (a_j - a_i) \frac{\partial f(\mathbf{x})}{\partial x_j}. \quad (2.12)$$

Applying Proposition 1.4.6 we get

$$\int_{\mathbb{R}^p} a_i \sum_{j=1}^p \frac{\partial f(\mathbf{x})}{\partial x_j} \circ ({}_s d\mathbf{x}) = a_i f(\mathbf{x}). \quad (2.13)$$

To the second term on right hand side of (2.12) we apply Proposition 1.4.9:

$$\begin{aligned} \int_{\mathbb{R}^p} \sum_{j=1}^p (a_j - a_i) \frac{\partial f(\mathbf{x})}{\partial x_j} \circ ({}_s d\mathbf{x}) &= \sum_{j=1}^p (a_i - a_j) \frac{\partial}{\partial x_j} \int_{\mathbb{R}^p} f(\mathbf{x}) \circ ({}_s d\mathbf{x}) \\ &= \sum_{j=1}^p (a_i - a_j) \frac{\partial F(\mathbf{x})}{\partial x_j}. \end{aligned} \quad (2.14)$$

After subtracting the equality (2.14) from equality (2.13) we get the statement. \square

Lemma 2.2.2. *Let the vector functions $f(\mathbf{x})$ and $F(\mathbf{x})$ be such that (2.11) holds and let \mathbf{B} be a constant $p \times p$ -matrix. Then*

$$\int_{\mathbb{R}^p} \text{vec}' \mathbf{B} \frac{d^2 f(\mathbf{x})}{d\mathbf{x}^2} \circ ({}_s d\mathbf{x}) = \text{vec}' \mathbf{B} \frac{d^2 F(\mathbf{x})}{d\mathbf{x}^2}.$$

Proof. To prove the statement we apply Propositions 1.4.8 and 1.4.9. Using assumption (2.11) we get

$$\int_{\mathbb{R}^p} \text{vec}' \mathbf{B} \frac{d^2 f(\mathbf{x})}{d\mathbf{x}^2} \circ ({}_s d\mathbf{x}) = \left(\text{vec} \mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2} \right) \int_{\mathbb{R}^p} f(\mathbf{x}) \circ ({}_s d\mathbf{x}) = \text{vec}' \mathbf{B} \frac{d^2 F(\mathbf{x})}{d\mathbf{x}^2}.$$

So we have got the statement of lemma. \square

Lemma 2.2.3. *Let the vector functions $f(\mathbf{x})$ and $F(\mathbf{x})$ satisfy (2.11) and let \mathbf{C} be a constant $p^2 \times p$ -matrix. Then*

$$\int_{\mathbb{R}^p} \text{vec}' \mathbf{C} \frac{d^3 f(\mathbf{x})}{d\mathbf{x}^3} \circ ({}_s d\mathbf{x}) = \text{vec}' \mathbf{C} \frac{d^3 F(\mathbf{x})}{d\mathbf{x}^3}.$$

Proof. Applying Propositions 1.4.8 and 1.4.9 and assumption (2.11) we get

$$\int_{\mathbb{R}^p} \text{vec}' \mathbf{C} \frac{d^3 f(\mathbf{x})}{d\mathbf{x}^3} \circ ({}_s d\mathbf{x}) = \left(\text{vec} \mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3} \right) \int_{\mathbb{R}^p} f(\mathbf{x}) \circ ({}_s d\mathbf{x}) = \text{vec}' \mathbf{C} \frac{d^3 F(\mathbf{x})}{d\mathbf{x}^3}.$$

So we have got the statement of lemma. \square

Using equality (2.10) and Lemmas 2.2.1, 2.2.2 and 2.2.3 we can formulate the following result.

Theorem 2.2.2. *Let \mathbf{X} and \mathbf{Y} be random vectors. Then the next relation holds between the probability distribution functions $F_{\mathbf{Y}}(\mathbf{x})$ and $F_{\mathbf{X}}(\mathbf{x})$:*

$$\begin{aligned} F_{\mathbf{Y}}(\mathbf{x}) &= F_{\mathbf{X}}(\mathbf{x}) - a_i f_{\mathbf{X}}(\mathbf{x}) + \sum_{j=1}^p (a_i - a_j) \frac{\partial F_{\mathbf{X}}(\mathbf{x})}{\partial x_j} \\ &+ \text{vec}' \mathbf{B} \frac{d^2 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^2} - \text{vec}' \mathbf{C} \frac{d^3 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^3} + \dots \end{aligned} \quad (2.15)$$

Proof. Let us start from equality (2.10). Using Lemmas 2.2.1-2.2.3 we get

$$\left(\mathbf{a}, \frac{d}{d\mathbf{x}} \right) F_{\mathbf{X}}(\mathbf{x}) = a_i f(\mathbf{x}) - \sum_{j=1}^p (a_i - a_j) \frac{\partial F(\mathbf{x})}{\partial x_j}$$

$$= a_i f_{\mathbf{X}}(\mathbf{x}) + \sum_{j=1}^p (a_j - a_i) \frac{\partial F_{\mathbf{X}}(\mathbf{x})}{\partial x_j},$$

$$(\text{vec} \mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}) F_{\mathbf{X}}(\mathbf{x}) = \text{vec}' \mathbf{B} \frac{d^2 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^2}$$

and

$$(\text{vec} \mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}) F_{\mathbf{X}}(\mathbf{x}) = \text{vec}' \mathbf{C} \frac{d^3 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^3}.$$

□

In a special case we get the following corollary.

Corollary 2.2.1. *Let*

$$a_i = a_1, \quad i = 1, 2, \dots, p,$$

$$b_{ij} = B_1, \quad i, j = 1, 2, \dots, p$$

and

$$c_{(i,j)(k,1)} = C_1, \quad i, j, k = 1, 2, \dots, p.$$

Then

$$F_{\mathbf{Y}}(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) - a_1 f_{\mathbf{X}}(\mathbf{x}) + B_1(\mathbf{1}_p, \frac{df(\mathbf{x})}{d\mathbf{x}}) - C_1(\mathbf{1}_{p^2}, \frac{d^2 f(\mathbf{x})}{d\mathbf{x}^2}) + \dots$$

Proof. Let us start from equality (2.15). Using Lemmas 2.2.1-2.2.3 we get

$$a_i f_{\mathbf{X}}(\mathbf{x}) + \sum_{j=1}^p (a_i - a_j) \frac{\partial F_{\mathbf{X}}(\mathbf{x})}{\partial x_j} = a_1 f_{\mathbf{X}}(\mathbf{x}),$$

$$\text{vec}' B \frac{d^2 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^2} = B_1(\mathbf{1}_{p^2}, \frac{d^2 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^2}) = B_1(\mathbf{1}_p, \frac{df(\mathbf{x})}{d\mathbf{x}})$$

and

$$\text{vec}' C \frac{d^3 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^3} = B_1(\mathbf{1}_{p^3}, \frac{d^3 F_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^3}) = C_1(\mathbf{1}_{p^2}, \frac{d^2 f(\mathbf{x})}{d\mathbf{x}^2}).$$

Summarizing these results we get the statement of the corollary. □

2.2.2 General relation between two distribution functions

In this subsection we give a general relation between two distribution functions using Theorem 2.2.1 and Lemmas 2.2.1-2.2.3.

Let us have a random p -vector \mathbf{X} with the density function $f(\mathbf{x})$ and the distribution function $F(\mathbf{x})$. Next we introduce new notation:

$$\mathbf{u}_{-j} = (u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_p)$$

and

$$d\mathbf{u}_{-j} = du_1 \dots du_{j-1} du_{j+1} \dots du_p.$$

The notation is also extended to the case when several coordinates are excluded from their full set. For example,

$$\mathbf{u}_{-i-j} = (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_{j-1}, u_{j+1}, \dots, u_p).$$

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, X_j) and $f_{ijk}(x_i, x_j, x_k)$ be the joint density function of (X_i, X_j, X_k) . Let $f(x_j|\mathbf{x}_{-j})$ be the conditional density function of X_j . Then the conditional distribution function of X_j is defined by the equality

$$F(x_j|\mathbf{x}_{-j}) = \int_{-\infty}^{x_j} f(u_j|\mathbf{x}_{-j}) du_j.$$

The number of combinations from n elements by k elements is denoted by C_k^n .

In the notations given above we formulate the following theorem.

Theorem 2.2.3. *Let \mathbf{X} be a random p -vector with known distribution function $F(\mathbf{x})$ and density function $f(\mathbf{x})$. Let \mathbf{Y} be a random p -vector with unknown distribution function $F_{\mathbf{Y}}(\mathbf{x})$. Then $F_{\mathbf{Y}}(\mathbf{x})$ is expressed through the distribution of \mathbf{X} as follows:*

$$\begin{aligned} F_{\mathbf{Y}}(\mathbf{x}) &= F(\mathbf{x}) - a_i f(\mathbf{x}) + \sum_{j=1}^p (a_i - a_j) f_j(x_j) F(\mathbf{x}_{-j}|x_j) \\ &+ \sum_{i=1}^p b_{ii} \frac{df_i(x_i)}{dx_i} F(\mathbf{x}_{-i}|x_i) + f_i(x_i) \frac{\partial F(\mathbf{x}_{-i}|x_i)}{\partial x_i} \\ &+ 2 \sum_{i,j=1}^p b_{ij} f_{ij}(x_i, x_j) F(\mathbf{x}_{-i-j}|x_i, x_j) \end{aligned}$$

$$\begin{aligned}
& - \sum_{i=1}^p c_{(i,i)(1,i)} \left(\frac{d^2 f_i(x_i)}{dx_i^2} F(\mathbf{x}_{-i}|x_i) + 2 \frac{df_i(x_i)}{dx_i} \frac{\partial F(\mathbf{x}_{-i}|x_i)}{\partial x_i} \right. \\
& \quad \left. + f_i(x_i) \frac{\partial^2 F(\mathbf{x}_{-i}|x_i)}{\partial x_i^2} \right) \\
& - 3C_2^p \sum_{i,j=1}^p (c_{(j,j)(1,i)} + c_{(j,i)(1,j)} + c_{(i,j)(1,j)}) \\
& \times \left(\frac{\partial f_{ij}(x_i, x_j)}{\partial x_j} F(\mathbf{x}_{-i-j}|x_i, x_j) + f_{ij}(x_i, x_j) \frac{\partial F(\mathbf{x}_{-i-j}|x_i, x_j)}{\partial x_j} \right) \\
& - 3C_2^p \sum_{i,j=1}^p (c_{(i,i)(1,j)} + c_{(i,j)(1,i)} + c_{(j,i)(1,i)}) \\
& \times \left(\frac{\partial f_{ij}(x_i, x_j)}{\partial x_i} F(\mathbf{x}_{-i-j}|x_i, x_j) + f_{ij}(x_i, x_j) \frac{\partial F(\mathbf{x}_{-i-j}|x_i, x_j)}{\partial x_i} \right) \\
& - 6C_3^p \sum_{i,j,k=1}^p c_{(i,j)(1,k)} f_{ijk}(x_i, x_j, x_k) F(\mathbf{x}_{-i-j-k}|x_i, x_j, x_k) \\
& \quad + \dots
\end{aligned} \tag{2.16}$$

Proof. The statement is obtained by applying consequently Lemmas 2.2.1, 2.2.2 and 2.2.3. Let us start from equality (2.15). At first we find partial derivatives $\frac{\partial F(\mathbf{x})}{\partial x_j}$. We get

$$\begin{aligned}
\frac{\partial F(\mathbf{x})}{\partial x_j} &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_{j-1}} \int_{-\infty}^{x_{j+1}} \dots \int_{-\infty}^{x_p} f(x_j, \mathbf{u}_{-j}) d\mathbf{u}_{-j} \\
&= f_j(x_j) \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_{j-1}} \int_{-\infty}^{x_{j+1}} \dots \int_{-\infty}^{x_p} f(\mathbf{u}_{-j}|x_j) d\mathbf{u}_{-j} \\
&= f_j(x_j) F(\mathbf{x}_{-j}|x_j), \quad j = 1, 2, \dots, p.
\end{aligned}$$

It is to see that

$$\begin{aligned}
\text{vec}' B \frac{d^2 F(\mathbf{x})}{d\mathbf{x}^2} &= \sum_{i,j=1}^p b_{ij} \frac{\partial^2 F(\mathbf{x})}{\partial^2 x_i x_j}, \\
\text{vec}' C \frac{d^3 F(\mathbf{x})}{d\mathbf{x}^3} &= \sum_{i,j,k=1}^p c_{(i,j)(1,k)} \frac{\partial^3 F(\mathbf{x})}{\partial^2 x_i x_j x_k}.
\end{aligned}$$

For higher order partial derivatives of $F(\mathbf{x})$ we get

$$\frac{\partial^2 F(\mathbf{x})}{\partial x_i^2} = \frac{df_i(x_i)}{dx_i} F(\mathbf{x}_{-i}|x_i) + f_i(x_i) \frac{\partial F(\mathbf{x}_{-i}|x_i)}{\partial x_i}$$

and

$$\begin{aligned} \frac{\partial^3 F(\mathbf{x})}{\partial x_i^3} &= \frac{d^2 f_i(x_i)}{dx_i^2} F(\mathbf{x}_{-i}|x_i) + 2 \frac{df_i(x_i)}{dx_i} \frac{\partial F(\mathbf{x}_{-i}|x_i)}{\partial x_i} \\ &\quad + f_i(x_i) \frac{\partial^2 F(\mathbf{x}_{-i}|x_i)}{\partial x_i^2}, \quad i = 1, 2, \dots, p. \end{aligned}$$

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

$$\begin{aligned} \frac{\partial^2 F(\mathbf{x})}{\partial x_i \partial x_j} &= \frac{\partial f_i(x_i) F(\mathbf{x}_{-i}|x_i)}{\partial x_j} \\ &= \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_{i-1}} \int_{-\infty}^{x_{i+1}} \int_{-\infty}^{x_{j-1}} \int_{-\infty}^{x_{j+1}} \dots \int_{-\infty}^{x_p} f(\mathbf{u}_{-i-j}, x_j | x_i) d\mathbf{u}_{-i-j} \\ &= f_{ij}(x_i, x_j) F(\mathbf{x}_{-i-j} | x_i, x_j), \quad i \neq j, \quad i, j = 1, 2, \dots, p. \end{aligned}$$

For the third order mixed derivatives we get

$$\begin{aligned} \frac{\partial^3 F(\mathbf{x})}{\partial x_i \partial x_j^2} &= \frac{\partial f_{ij}(x_i, x_j) F(\mathbf{x}_{-i-j} | x_i, x_j)}{\partial x_j} \\ &= \frac{\partial f_{ij}(x_i, x_j)}{\partial x_j} F(\mathbf{x}_{-i-j} | x_i, x_j) + f_{ij}(x_i, x_j) \frac{\partial F(\mathbf{x}_{-i-j} | x_i, x_j)}{\partial x_j} \end{aligned}$$

and

$$\frac{\partial^3 F(\mathbf{x})}{\partial x_i \partial x_j \partial x_k} = f_{ijk}(x_i, x_j, x_k) F(\mathbf{x}_{-i-j-k} | x_i, x_j, x_k), \quad i \neq j \neq k,$$

$$i, j, k = 1, 2, \dots, p.$$

Replacing partial derivatives, the second and the third order derivatives and all the involved mixed derivatives into equality (2.15) we get the statement of the theorem. \square

Let us apply the statement (2.16) in the two-dimensional case.

Corollary 2.2.2. *Let $\mathbf{X} = (X_1, X_2)$ and $\mathbf{Y} = (Y_1, Y_2)$ be with known and unknown distribution functions respectively. Then*

$$\begin{aligned}
F_{\mathbf{Y}}(\mathbf{x}) &= F(\mathbf{x}) - a_1 f_X(\mathbf{x}) + (a_1 - a_2) f_2(x_2) F(x_1|x_2) \\
&+ 2b_{12} f(\mathbf{x}) + b_{11} \frac{\partial f_1(x_1) F(x_2|x_1)}{\partial x_1} + b_{22} \frac{\partial f_2(x_2) F(x_1|x_2)}{\partial x_2} \\
&\quad - (c_{((1,1)(1,2))} + c_{(2,1)(1,1)} + c_{(1,2)(1,1)}) \frac{\partial f(\mathbf{x})}{\partial x_1} \\
&\quad - (c_{(2,2)(1,1)} + c_{(2,1)(1,2)} + c_{(1,2)(1,2)}) \frac{\partial f(\mathbf{x})}{\partial x_2} \\
&- \left(c_{(1,1)(1,1)} \frac{\partial^2 f_1(x_1) F(x_2|x_1)}{\partial^2 x_1} + c_{(2,2)(1,2)} \frac{\partial^2 f_2(x_2) F(x_1|x_2)}{\partial^2 x_2} \right) \\
&\quad + \dots
\end{aligned}$$

2.3 Approximation with the normal distribution

For p -variate normal distribution with mean value $\mathbf{0}_p$ and covariance matrix Σ we introduce the notation $N_p(0, \Sigma)$. In this section we apply the results of the previous section in the case when we approximate the unknown bivariate distribution function through the normal distribution $N_2(0, \Sigma)$. We give also some examples of this kind of approximation. Finally we apply the approximation based on the normal distribution for modeling the distribution function of real data.

2.3.1 Representation of the unknown distribution function through the normal distribution

We introduce first the Hermite matrix-polynomials for a p -vector \mathbf{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 Fisher and Cornish (1960). We call this type of approximation as Edgeworth type expansion.

Definition 2.3.1. *Let \mathbf{x} be a p -vector. Then the matrix $H_k(\mathbf{x}, \Sigma)$ is called Hermite matrix-polynomial if it is defined by the equality*

$$\frac{d^k f_{\mathbf{X}}(\mathbf{x})}{d\mathbf{x}^k} = (-1)^k H_k(\mathbf{x}, \Sigma) f_{\mathbf{X}}(\mathbf{x}), \quad k = 1, 2, \dots$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the density function of the normal distribution $N_p(0, \Sigma)$.

As follows from Definition 2.3.1, the Hermite matrix-polynomials are obtained by matrix differentiation.

The Hermite matrix polynomials up to the third order are given by equalities (Kollo, 1991 p. 141):

$$H_0(\mathbf{x}, \Sigma) = 1;$$

$$H_1(\mathbf{x}, \Sigma) = \mathbf{x}'\Sigma^{-1};$$

$$H_2(\mathbf{x}, \Sigma) = \Sigma^{-1}\mathbf{x}\mathbf{x}'\Sigma^{-1} - \Sigma^{-1};$$

$$H_3(\mathbf{x}, \Sigma) = (\Sigma^{-1}\mathbf{x})^{\otimes 2}\mathbf{x}'\Sigma^{-1} - \text{vec}\Sigma^{-1}(\mathbf{x}'\Sigma^{-1}) - (\Sigma^{-1}\otimes\Sigma^{-1}\mathbf{x}) - (\Sigma^{-1}\mathbf{x}\otimes\Sigma^{-1}).$$

In the univariate case when $X \sim N(0, \sigma^2)$ the Hermite polynomials $h_i(x)$, $i = 0, 1, 2$ take the following form:

$$h_0(x) = 1,$$

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

and

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

Now we apply Theorem 2.2.3 in the bivariate case, $\mathbf{X} \sim N_2(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

Let ρ be the Pearson correlation coefficient between X_1 and X_2 and let us use the following notation:

$$g(x_1) = \frac{\frac{x_2}{\sqrt{\sigma_{22}}} - \frac{x_1}{\sqrt{\sigma_{11}}}\rho}{\sqrt{1 - \rho^2}} \quad (2.17)$$

and

$$g(x_2) = \frac{\frac{x_1}{\sqrt{\sigma_{11}}} - \frac{x_2}{\sqrt{\sigma_{22}}}\rho}{\sqrt{1 - \rho^2}}. \quad (2.18)$$

Let $\Phi(x)$ be the distribution function of the standard normal distribution.

In the next theorem we present a formal expansion of the distribution function of the bivariate random vector through the normal distribution $N_2(0, \Sigma)$.

Theorem 2.3.1. Let $\mathbf{X} \sim N_2(0, \Sigma)$ and let $F_{\mathbf{X}}(\mathbf{x})$ be the distribution function of \mathbf{X} and $F_{\mathbf{Y}}(\mathbf{x})$ be the unknown distribution function of bivariate random vector \mathbf{Y} . Let $f_1(x_1)$ and $f_2(x_2)$ be the marginal density functions of X_1 and X_2 , respectively. Then

$$\begin{aligned}
F_{\mathbf{Y}}(\mathbf{x}) = & F_{\mathbf{X}}(\mathbf{x}) + \{a_2 + 2b_{12} + (\mathbf{C}_{12}, H_1(\mathbf{x}, \Sigma))\} f_{\mathbf{X}}(\mathbf{x}) \\
& + \{(a_1 - a_2)f_2(x_2)\} \Phi(g(x_2)) \\
& - b_{11}\{h_1(x_1) - g'(x_1)\} f_1(x_1) \Phi(g(x_1)) \\
& - b_{22}\{h_1(x_2) - g'(x_2)\} f_2(x_2) \Phi(g(x_2)) \\
& - c_{(1,1)(1,1)}\{h_2(x_1)f_1(x_1)\Phi(g(x_1)) \\
& \quad - 2h_1(x_1)f_1(x_1)f_1(g(x_1))g'(x_1) \\
& \quad - f_1(x_1)h_1(g(x_1))f_1(g(x_1))g'(x_1)^2\} \\
& - c_{(2,2)(1,2)}\{h_2(x_2)f_2(x_2)\Phi(g(x_2)) \\
& \quad - 2h_1(x_2)f_2(x_2)f_1(g(x_2))g'(x_2) \\
& \quad - f_2(x_2)h_1(g(x_2))f_2(g(x_2))g'(x_2)^2\} + \dots \tag{2.19}
\end{aligned}$$

where

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

and \mathbf{a} , \mathbf{B} and \mathbf{C} are defined by equalities (2.7), (2.8) and (2.9).

Proof. Let us start from Corollary 2.2.2. First we show that

$$F(x_2|x_1) = \Phi(g(x_1)). \tag{2.20}$$

We get by integration

$$\begin{aligned}
\int_{-\infty}^{x_2} f(x_1, u_2) du_2 &= f_1(x_1) \int_{-\infty}^{x_2} f(u_2|x_1) du_2 \\
&= f_1(x_1) F(x_2|x_1).
\end{aligned}$$

Let us find the integral $\int_{-\infty}^{x_2} f(x_1, u_2) du_2$:

$$\begin{aligned}
& \int_{-\infty}^{x_2} f(x_1, u_2) du_2 \\
&= \frac{1}{2\pi|\Sigma|^{1/2}} \int_{-\infty}^{x_2} \exp\left(\frac{1}{2}(x_1, u_2)\Sigma^{-1}(x_1, u_2)\right) du_2
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \int_{-\infty}^{x_2} \exp\left(\frac{\sigma_{22}x_2^2 - 2\sigma_{12}x_1u_2 + \sigma_{11}u_2^2}{2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}\right) du_2 \\
&= \frac{1}{\sqrt{2\pi(\sigma_{11})}} \frac{1}{\sqrt{\sigma_{22}(1-\rho^2)}} \exp\left(-\frac{x_1^2}{2\sigma_{11}}\right) \\
&\quad \times \int_{-\infty}^{x_2} \exp\left(-\frac{(\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\rho x_1 - u_2)^2}{2\sigma_{22}(1-\rho^2)}\right) du_2 \\
&= f_1(x_1) \int_{-\infty}^{x_2} \exp\left(-\frac{(\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\rho x_1 - u_2)^2}{2\sigma_{22}(1-\rho^2)}\right) du_2.
\end{aligned}$$

We have got

$$F(x_2|x_1) = \frac{1}{\sqrt{\sigma_{22}(1-\rho^2)}} \int_{-\infty}^{x_2} \exp\left(-\frac{(\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\rho x_1 - u_2)^2}{2\sigma_{22}(1-\rho^2)}\right) du_2.$$

It is easy to detect that

$$\begin{aligned}
&D\left(\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\rho X_1 - X_2\right) \\
&= \frac{\sigma_{22}}{\sigma_{11}}\rho^2 D(X_1) + D(X_2) + 2\text{cov}\left(\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\rho X_1, -X_2\right) \\
&= \rho^2\sigma_{22} + \sigma_{22} - 2\rho\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\text{cov}(X_1, X_2) \\
&= \rho^2\sigma_{22} + \sigma_{22} - 2\rho^2\sigma_{22} = \sigma_{22}(1-\rho^2).
\end{aligned}$$

Using equality (2.17) we get

$$g(x_1) = \frac{x_2 - \sqrt{\frac{\sigma_{22}}{\sigma_{11}}}\rho x_1}{\sqrt{\sigma_{22}(1-\rho^2)}}.$$

So the equality (2.20) is proven. Using equality (2.18) we can prove in the same way that

$$F(x_1|x_2) = \Phi(g(x_2)). \quad (2.21)$$

At the next step we find higher order partial derivatives of $F_{\mathbf{X}}(\mathbf{x})$. Differentiation gives us

$$\frac{\partial^2 F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 x_2} = f_{\mathbf{X}}(\mathbf{x}), \quad (2.22)$$

$$\frac{\partial^3 F_{\mathbf{X}}(\mathbf{x})}{\partial x_1^2 x_2} = -(H_1(\mathbf{x}, \boldsymbol{\Sigma}))_1 f_{\mathbf{X}}(\mathbf{x}), \quad (2.23)$$

$$\frac{\partial^3 F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 x_2^2} = -(H_1(\mathbf{x}, \boldsymbol{\Sigma}))_2 f_{\mathbf{X}}(\mathbf{x}). \quad (2.24)$$

Using univariate Hermite polynomials we get

$$\begin{aligned} \frac{\partial f_1(x_1)F(x_2|x_1)}{\partial x_1} &= -h_1(x_1)f_1(x_1)\Phi(g(x_1)) \\ &\quad + f_1(x_1)f_1(g(x_1))g'(x_1) \end{aligned} \quad (2.25)$$

and

$$\begin{aligned} \frac{\partial^2 f_1(x_1)F(x_2|x_1)}{\partial x_1^2} &= h_2(x_1)f_1(x_1)\Phi(g(x_1)) \\ &\quad - 2h_1(x_1)f_1(x_1)f_1(g(x_1))g'(x_1) \\ &\quad + f_1(x_1)h_1(g(x_1))f_1(g(x_1))g'(x_1)^2. \end{aligned} \quad (2.26)$$

In the same way

$$\begin{aligned} \frac{\partial f_2(x_2)F(x_1|x_2)}{\partial x_2} &= -h_1(x_2)f_2(x_2)\Phi(g(x_2)) \\ &\quad + f_2(x_2)f_2(g(x_2))g'(x_2) \end{aligned} \quad (2.27)$$

and

$$\begin{aligned} \frac{\partial^2 f_2(x_2)F(x_1|x_2)}{\partial x_2^2} &= h_2(x_2)f_2(x_2)\Phi(g(x_2)) \\ &\quad - 2h_1(x_2)f_2(x_2)f_2(g(x_2))g'(x_2) \\ &\quad + f_2(x_2)h_1(g(x_2))f_2(g(x_2))g'(x_2)^2. \end{aligned} \quad (2.28)$$

Using Definition 2.3.1 and replacing the expressions (2.20)-(2.28) into the equality of Corollary 2.2.2, we get the statement. \square

2.3.2 Simulation

In the simulation experiment we compare our approximating distribution function with the empirical distribution function for certain statistics. We have drawn samples of size n from i.i.d. (independent identically distributed) bivariate random vectors $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$, $\mathbf{X}_i \sim N_2(0, \Sigma)$, $i = 1, 2, \dots, n$ with

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where ρ is the Pearson correlation coefficient between X_{i1} and X_{i2} . The sampling values $\mathbf{x}_i = (x_{i1}, x_{i2})$, $i = 1, 2, \dots, n$, are presented in a matrix $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, thus being a realization of the random matrix

$$\mathcal{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n).$$

The number of replications $k = 200$. For each replicated sample, we compute the (unbiased) sample covariance matrix, what is a realization of the random matrix

$$\mathbf{S} = \frac{1}{n-1} \mathcal{X}(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n) \mathcal{X}'.$$

Define the matrix $\mathbf{L} = \sqrt{n}(\mathbf{S} - \boldsymbol{\Sigma})$. Let us form random vector $\mathbf{T} = (L_{11}, L_{22})'$ and let vector \mathbf{t} be a realization of \mathbf{T} .

To find an approximation of the distribution function of \mathbf{T} we need the first moments of it. The moments of $\text{vec}(\mathbf{S})$ are studied in Traat (1984) up to the order four. The second order moment of $\text{vec}(\mathbf{S})$ is

$$\mathbf{M}_2(\mathbf{S}) = \frac{1}{n-1} (\mathbf{I}_{p^2} + \mathbf{K}_{2,2}) (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})$$

and the third order moment of $\text{vec}(\mathbf{S})$ is

$$\mathbf{M}_3(\mathbf{S}) = \frac{1}{n^2} (\mathbf{M}_6^* - \text{vec}(\mathbf{M}_4) \otimes \text{vec}(\boldsymbol{\Sigma}) - \mathbf{W} - \mathbf{K}_{p^4, p^2} \mathbf{W}) + o(n^3)$$

where

$$\begin{aligned} \mathbf{W} &= (\mathbf{M}_4 \otimes \text{vec}(\boldsymbol{\Sigma}) + \text{vec}(\mathbf{M}_3) \otimes \mathbf{M}'_3 + \mathbf{M}_3 \otimes \mathbf{M}_3 + \mathbf{M}'_3 \otimes \text{vec}(\mathbf{M}_3) \\ &\quad - \text{vec}(\boldsymbol{\Sigma}) \otimes \text{vec}'(\boldsymbol{\Sigma}) \otimes \text{vec}(\boldsymbol{\Sigma})), \\ \mathbf{M}_3 &= E(\mathbf{X}_i \otimes \mathbf{X}'_i \otimes \mathbf{X}_i), \\ \mathbf{M}_4 &= E(\mathbf{X}_i \otimes \mathbf{X}'_i \otimes \mathbf{X}_i \otimes \mathbf{X}'_i) \end{aligned}$$

and $\mathbf{M}_6^* = E(\mathbf{X}_i \otimes \mathbf{X}'_i \otimes \mathbf{X}_i \otimes \mathbf{X}'_i \otimes \mathbf{X}_i \otimes \mathbf{X}_i)$ (Traat, 1984).

Let $\mathbf{M}_1(\mathbf{T})$, $\mathbf{M}_2(\mathbf{T})$ and $\mathbf{M}_3(\mathbf{T})$ be the first, second and third moments of \mathbf{T} respectively. We get these moments from the corresponding moment matrices of $\text{vec}(\mathbf{S})$ by indicating necessary elements:

$$\mathbf{M}_1(\mathbf{T}) = (0, 0),$$

$$\mathbf{M}_2(\mathbf{T}) = \frac{n}{n-1} \begin{pmatrix} (\mathbf{M}_2(\mathbf{S}))_{11} & (\mathbf{M}_2(\mathbf{S}))_{14} \\ (\mathbf{M}_2(\mathbf{S}))_{41} & (\mathbf{M}_2(\mathbf{S}))_{44} \end{pmatrix}$$

and

$$\mathbf{M}_3(\mathbf{T}) = n^{-\frac{1}{2}} \begin{pmatrix} (\mathbf{M}_3(\mathbf{S}))_{11} & (\mathbf{M}_3(\mathbf{S}))_{14} \\ (\mathbf{M}_3(\mathbf{S}))_{41} & (\mathbf{M}_3(\mathbf{S}))_{44} \\ (\mathbf{M}_3(\mathbf{S}))_{13,1} & (\mathbf{M}_3(\mathbf{S}))_{13,4} \\ (\mathbf{M}_3(\mathbf{S}))_{16,1} & (\mathbf{M}_3(\mathbf{S}))_{16,4} \end{pmatrix}. \quad (2.29)$$

Our aim is to approximate the distribution function of \mathbf{T} through the distribution function of $\mathbf{Y} \sim N_2(0, \mathbf{M}_2(\mathbf{T}))$. In this approximation we use the

results of Theorem 2.3.1. It is easy to detect that $c_1(\mathbf{Y}) = c_1(\mathbf{T}) = \mathbf{0}_2$, $c_2(\mathbf{Y}) = c_2(\mathbf{T}) = \mathbf{M}_2(\mathbf{T})$ and $c_3(\mathbf{Y}) = \mathbf{0}_{4 \times 2}$. Let $F_{\mathbf{Y}}(\mathbf{t})$ be the distribution function of \mathbf{Y} and $F_{\mathbf{T}}(\mathbf{t})$ be the distribution function of \mathbf{T} . Then, using the equalities (2.4), (2.19) and (2.29), we get

$$\begin{aligned}
F_{\mathbf{T}}(\mathbf{t}) &= F_{\mathbf{Y}}(\mathbf{t}) + \frac{1}{6}[(\mathbf{M}_3, H_1(\mathbf{t}, \mathbf{M}_2(\mathbf{T})))f_{\mathbf{Y}}(\mathbf{t}) \\
&\quad - (\mathbf{M}_3(\mathbf{T}))_{11}\{h_2(t_1)f_1(t_1)\Phi(g(t_1)) - 2h_1(t_1)f_1(t_1)f_1(g(t_1))g'(t_1) \\
&\quad \quad - g(t_1)f_1(g(t_1))f_1(t_1)g'(t_1)^2\} \\
&\quad - (\mathbf{M}_3(\mathbf{T}))_{42}\{h_2(t_2)f_2(t_2)\Phi(g(t_2)) - 2h_1(t_2)f_2(t_2)f_2(g(t_2))g'(t_2) \\
&\quad \quad - g(t_2)f_2(g(t_2))f_2(t_2)g'(t_2)^2\}] + \dots \tag{2.30}
\end{aligned}$$

where $g(t_1)$ and $g(t_2)$ are defined by equalities (2.17)-(2.18),

$$\mathbf{M}_3 = \begin{pmatrix} (\mathbf{M}_3(\mathbf{T}))_{12} + (\mathbf{M}_3(\mathbf{T}))_{21} + (\mathbf{M}_3(\mathbf{T}))_{31} \\ (\mathbf{M}_3(\mathbf{T}))_{22} + (\mathbf{M}_3(\mathbf{T}))_{32} + (\mathbf{M}_3(\mathbf{T}))_{41} \end{pmatrix},$$

$f_{\mathbf{Y}}(\mathbf{t})$ is the density function of \mathbf{Y} , $f_1(t_1)$ and $f_2(t_2)$ are marginal density functions of Y_1 and Y_2 respectively.

We consider two sample sizes, $n = 20$ and $n = 50$. Matrix $(n-1)\mathbf{S}$ is Wishart distributed (Wishart distribution was introduced in Wishart (1928)). We estimate the distribution function of the diagonal elements of \mathbf{S} (vector \mathbf{T}) by constructing empirical distribution function $F_k(\mathbf{t})$ for both sample sizes. To measure the goodness of approximation given by (2.30) we construct the following quantity:

$$d = \frac{\sum_{i=1}^k (F_k(\mathbf{t}_i) - F_{\mathbf{T}}(\mathbf{t}_i))^2}{k} \tag{2.31}$$

where \mathbf{t}_i is a realization of the random vector \mathbf{T} in the i -th replication.

We repeated the simulation experiment 10 times and computed the mean of d over these 10 replications. The results of simulation are presented in Table 2.1. The table presents the mean values of d (denoted by \bar{d}) for the two sample sizes, for the normal approximation and theoretical approximation (2.30). In this table "Normal" means that instead of $F_{\mathbf{T}}(\mathbf{t})$ the distribution function of the normal distribution $N_2(0, \mathbf{M}_2(\mathbf{T}))$ is used and "Theoretic" means that $F_{\mathbf{T}}(\mathbf{t})$ is found by means of equality (2.30).

Table 2.1 Goodness of distribution function approximation depending on sample size.

Sample size	Distribution function	Value of $\bar{d} \times 10^6$
20	Normal	1720
	Theoretic	1030
50	Normal	438
	Theoretic	343

Goodness-of-fit can be estimated visually by the $Q-Q$ plot. In the Figure 2.1 we have 3 lines. The straight line with slope 1 corresponds to the empirical distribution function $F_k(\mathbf{t})$ which is estimated from simulation experiment. Line "Normal" shows the behavior of the normal approximation. Line "Theor" shows the behavior of the approximation given by (2.30).

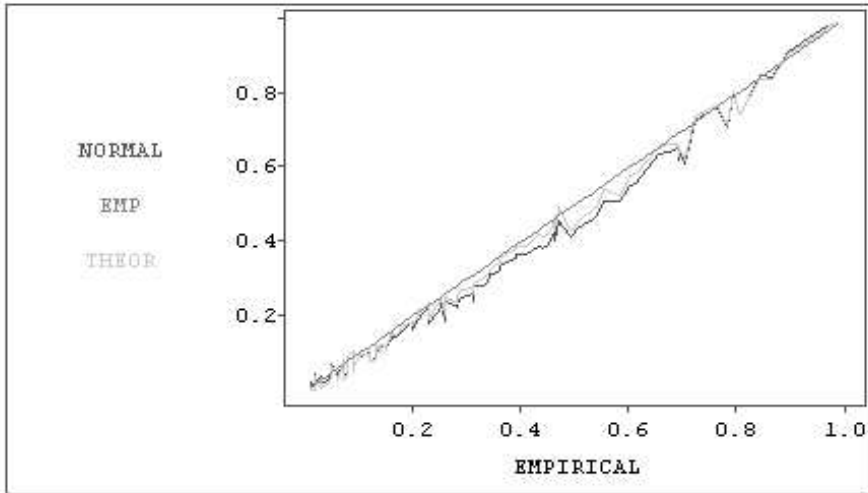


Figure 2.1 The $Q-Q$ plot describing goodness of approximation of the distribution function of \mathbf{T} , $n = 20$.

From Table 2.1 we can conclude that approximation is getting better when the sample size is growing and gives quite good fit even when $n = 20$. Comparing relative gains of the theoretical over normal approximation for the two sample sizes we get: $\frac{172 - 103}{103} = 0.67$ when $n = 20$, while $\frac{438 - 343}{343} = 0.28$ when $n = 50$. From equality (2.29) we can see that the third moment of \mathbf{T} is proportional to $n^{-\frac{1}{2}}$. The role of $\mathbf{M}_3(\mathbf{T})$ decreases in equality (2.30), if the sample size n increases.

2.3.3 Application of the Edgeworth type approximation

In this subsection we apply the results of Theorem 2.3.1 to real data (Adermann and Pihlak, 2005). The data is formed by tree measurement of Estonian National Forest Inventory from the period of 1999-2003. Measurements of 3257 spruces and 4853 pines were used. The sample is representative covering all the forest types throughout the country. Sample trees are chosen using the Bitterlich basal area gauge on volume plots allocated on the basis of systematic random choice method. The chance that a tree is included into the sample is proportional to its basal area. Four standard density classes (< 0.6 , $0.6 \dots 0.7$, $0.8 \dots 0.9$ and > 0.9) of the widespread Estonian tree species: pine (*Pinus sylvestris*) and spruce (*Picea abies*) were analysed in the course of data processing. The density class shows relative density of the surrounding stand. The relative density is defined as the ratio of the basal area of stand and basal area of normal stand. Normal stand is defined as even-aged one-storied stand with relative density 1.0 which uses maximally the environment of growth. The maximum and minimum of tree height H , diameter at the breast height (DBH) and the sample sizes of both species are presented in Table 2.2. The breast height of the tree is 1.3 meter.

Table 2.2. The maximum and minimum of tree height (H) (in m) and DBH (in cm) and sample sizes (Adermann and Pihlak, 2005).

Species	$H_{\min} \dots H_{\max}$	$DBH_{\min} \dots DBH_{\max}$	Sample size
Pine	1.4...38	0.6...59.7	4052
Spruce	1.4...36	0.9...65	3257

Our aim is to approximate the joint distribution function of H and DBH with the distribution function of normal distribution. The functional dependence between DBH and height has been intensively studied in the literature. One of the most widely used 2-parameter family of functions is elaborated by M. Näslund (Näslund, 1941). Following Näslund in the classical deterministic model variables H and DBH are connected by the following equality

$$\sqrt{\frac{1}{H - 1.3}} = \frac{b_0}{DBH} + b_1 \quad (2.32)$$

where b_0 and b_1 are the parameters. In Chapter 3 we shall see that H can be modeled by normal distribution and DBH follows a log-normal distribution. In the following we denote DBH by X_1 and H by X_2 .

Next we generate data

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

where $\ln(X_1)$ has the standard normal distribution and the dependence between X_1 and X_2 is described by equality (2.32). Let us call data generated from \mathbf{X} the forestry model data. We center and normalize the components X_1 and X_2 . Let $F_{\mathbf{X}}(\mathbf{x})$ be the distribution function of \mathbf{X} and \mathbf{S} be the sample covariance matrix calculated from the forestry model data. Let $F_{\mathbf{Y}}(\mathbf{x})$ and $f_{\mathbf{Y}}(\mathbf{x})$ be the distribution and density functions of normal distribution $N_2(0, \mathbf{S})$ respectively. We approximate the function $F_{\mathbf{X}}(\mathbf{x})$ with the function $F_{\mathbf{Y}}(\mathbf{x})$ using the equality (2.30). Let $c_3(\mathbf{X}) = E(\mathbf{X} \otimes \mathbf{X}' \otimes \mathbf{X})$. Then we get

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= F_{\mathbf{Y}}(\mathbf{x}) + \frac{1}{6}[(\mathbf{C1}, H_1(\mathbf{x}, \mathbf{S}))f_{\mathbf{Y}}(\mathbf{x}) \\ &\quad - (c_3(\mathbf{X}))_{11}\{h_2(x_1)f_1(x_1)\Phi(g(x_1)) - 2h_1(x_1)f_1(x_1)f_1(g(x_1))g'(x_1) \\ &\quad \quad - g(x_1)f_1(g(x_1))f_1(x_1)g'(x_1)^2\} \\ &\quad - (c_3(\mathbf{X}))_{42}\{h_2(x_2)f_2(x_2)\Phi(g(x_2)) - 2h_1(x_2)f_2(x_2)f_2(g(x_2))g'(x_2) \\ &\quad \quad - g(x_2)f_2(g(x_2))f_2(x_2)g'(x_2)^2\}] + \dots \end{aligned} \quad (2.33)$$

where

$$\mathbf{C1} = \begin{pmatrix} (c_3(\mathbf{X}))_{12} + (c_3(\mathbf{X}))_{21} + (c_3(\mathbf{X}))_{31} \\ (c_3(\mathbf{X}))_{22} + (c_3(\mathbf{X}))_{32} + (c_3(\mathbf{X}))_{41} \end{pmatrix}$$

and $f_1(x_1)$ and $f_2(x_2)$ are marginal density functions of Y_1 and Y_2 , respectively.

We apply equality (2.31) to the distribution functions $F_{\mathbf{X}}(\mathbf{x})$ and $F_{\mathbf{Y}}(\mathbf{x})$.

We generate data \mathbf{X} 10 times. For each data we compute d according to (2.31). The value \bar{d} is the calculated average. In the case of normal approximation $\bar{d} = 0.0168$ and in the case when applying the equality (2.33) we have $\bar{d} = 0.00202$. We can conclude that the expression (2.33) corrects essentially the fit of the distribution function.

Next we apply expression (2.33) to the forestry data to model the joint distribution function of DBH and H . The goodness-of-fit of approximations is tested by means of equality (2.31). The results of approximation are presented in Table 2.3.

Table 2.3 Goodness of distribution function approximation on different tree species and density classes.

Species	Density class	Distribution function	Value of $d \times 10^6$
spruce	low	Normal	1190
		Theoretic	514
spruce	high	Normal	1370
		Theoretic	217
pine	low	Normal	496
		Theoretic	183
pine	high	Normal	152
		Theoretic	95.1

From Table 2.3 we can conclude that the term of asymmetry in expression (2.33) has significant role when approximating the joint distribution function of DBH and H .

The results of approximation can be visually estimated in Figures 2.2 and 2.3. In these figures we can see the $Q-Q$ plots for empirical distribution, normal approximation and approximation of the distribution function obtained by means of equality (2.33). The line "Emp" is for empirical distribution, "Normal" denotes the normal approximation and line "Theoretic" describes the behavior of approximation obtained by equality (2.33).

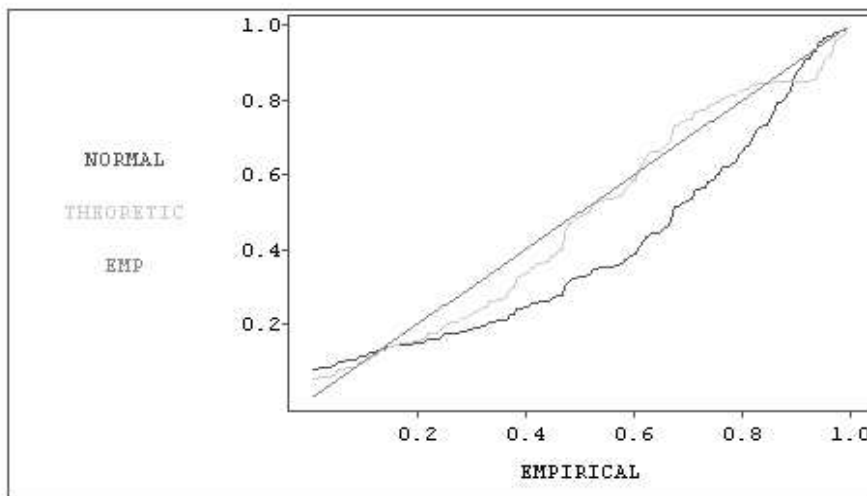


Figure 2.2 The $Q-Q$ plot for the forestry model data.

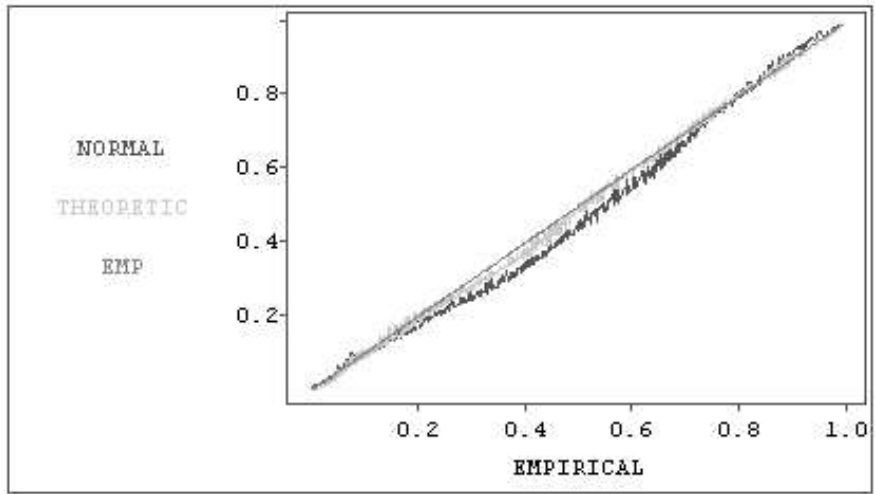


Figure 2.3 The $Q-Q$ plot for spruce data, in the case of high density class.

Chapter 3

Approximation by copulas

In this chapter we study another method for approximation of multivariate probability distributions. This method is based on copula theory. Initiator of the copula theory is A. Sklar who laid down foundations of the theory (Sklar, 1959). By Sklar's approach a copula is a joint distribution of uniformly distributed random variables which are connected through a probability transform with the original random variables with given marginal distributions. First we give some basic results of the copula theory. Then we turn to the data, examined in subsection 2.3.3. There we modeled the joint distribution of tree's height H and tree's diameter at the breast height (DBH) by our new approximation method. In this chapter we follow Adermann and Pihlak (2005) and present a copula approximation. In the last section we compare the two models obtained by copulas and by Edgeworth type expansion.

3.1 Some basic results of copula theory

Let us denote $\mathbf{I} = [0, 1]$ and let \mathbf{I}^2 be the unit square. Then a copula is defined as follows.

Definition 3.1.1. *The function $C : \mathbf{I}^2 \rightarrow \mathbf{I}$ is called copula if 1) for every $u, v \in \mathbf{I}$,*

$$C(u, 0) = C(0, v) = 0$$

and

$$C(u, 1) = u, \quad C(1, v) = v;$$

2) for every u_1, u_2, v_1 and $v_2 \in \mathbf{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

We shall explain Definition 3.1.1 on an example.

Example 3.1.1 Let $u, v \in \mathbf{I}$. Let we have a function $\Pi(u, v) = uv$. We show that $\Pi(u, v)$ is a copula. It is easy detect that $uv \in \mathbf{I}$. It follows also straightforwardly that

$$\Pi(0, v) = \Pi(u, 0) = 0$$

and

$$\Pi(1, v) = 1v = v, \quad \Pi(u, 1) = u1 = u.$$

Let u_1, u_2, v_1 and $v_2 \in \mathbf{I}$. Assume that $u_1 \leq u_2$ and $v_1 \leq v_2$. We have to show that $\Pi(u_2, v_2) - \Pi(u_2, v_1) - \Pi(u_1, v_2) + \Pi(u_1, v_1) \geq 0$. We get

$$\begin{aligned} \Pi(u_2, v_2) - \Pi(u_2, v_1) - \Pi(u_1, v_2) + \Pi(u_1, v_1) &= u_2v_2 - u_2v_1 - u_1v_2 + u_1v_1 \\ &= (u_2 - u_1)(v_2 - v_1) \geq 0, \end{aligned}$$

if $u_1 \leq u_2$ and $v_1 \leq v_2$.

We have shown that $\Pi(u, v)$ satisfies the conditions of Definition 3.1.1.

The copula $\Pi(u, v)$ is called *product copula* (Nelsen, 1999, p. 9).

3.1.1 Copulas and random variables

In this subsection we study the relation between copulas and bivariate distributions. Let us have continuous random variable X with the distribution function F . If random variable U is uniformly distributed on $[0, 1]$ then the following equality holds:

$$U = F(X). \tag{3.1}$$

The equality (3.1) is the basic relation for the copula theory. The next theorem gives the main result in the bivariate case.

Theorem 3.1.1. *Let H be a joint bivariate distribution function with marginal distribution functions F and G . Then there exists a copula C such that for all $x, y \in \mathfrak{R}$*

$$H(x, y) = C(F(x), G(y)).$$

Theorem 3.1.1 is called the Sklar's theorem (Nelsen, 1999, p. 18). Sklar's theorem elucidates the role that copulas play in the relationship between the multivariate distribution function and its univariate margins. Using Theorem 3.1.1 we get

$$\begin{aligned} C(F(x), G(y)) &= P(U \leq F(x), V \leq G(y)) = P(F^{-1}(U) \leq x, G^{-1}(V) \leq y) \\ &= P(X \leq x, Y \leq y) = H(x, y) \end{aligned}$$

where random variables U and V are uniformly distributed on $[0, 1]$, i. e. the joint distribution function of X and Y can be presented as copula.

3.1.2 Dependence in copula theory

Let us examine different forms of dependencies in the copula theory. Let (x_i, y_i) and (x_j, y_j) be two observations from (X, Y) . We say that (x_i, y_i) and (x_j, y_j) are *concordant* if $x_i < x_j$ and $y_i < y_j$, or if $x_i > x_j$ and $y_i > y_j$. Similarly, we say that (x_i, y_i) and (x_j, y_j) are *discordant* if $x_i < x_j$ and $y_i > y_j$, or if $x_i > x_j$ and $y_i < y_j$.

Definition 3.1.2. *Let c denote the number of concordant pairs, and d the number of discordant pairs. Then Kendall's tau for the sample is defined as*

$$\tau = \frac{c - d}{c + d}.$$

Different measures of association between X and Y such as Kendall's τ or Spearman's ρ and "concordance function" q are used. The concordance function q is defined as follows (Nelsen, 1999):

$$q = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

where $(X_1, Y_1)'$ and $(X_2, Y_2)'$ are independent vectors of continuous random variables. In these notations the next theorem is valid (Nelsen, 1999, p. 127-128).

Theorem 3.1.2. *Let H_1 and H_2 be the joint distribution functions of (X_1, Y_1) and (X_2, Y_2) , respectively. Let C_1 and C_2 denote the copulas of (X_1, Y_1) and (X_2, Y_2) , respectively, so that $H_1(x, y) = C_1(F(x), F(y))$ and $H_2(x, y) = C_2(F(x), F(y))$. Then*

$$q(C_1, C_2) = 4 \int \int_{\mathbf{I}^2} C_2(u, v) dC_1(u, v) - 1. \quad (3.2)$$

We can see that q is a nondecreasing function in both arguments. Let $(X, Y)'$ be a vector of continuous random variables. Let $\tau_{X,Y}$ denote the Kendall's τ between X and Y . Then

$$\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] = q.$$

It means that Kendall's τ is defined as the probability of concordance minus the probability of discordance. The relation between a copula and Kendall's rank correlation coefficient is given in the following theorem (Nelsen 1999, p. 129):

Theorem 3.1.3. *Let X and Y be continuous random variables whose copula is C . Then the Kendall's tau between X and Y*

$$\tau_{X,Y} = q(C, C). \quad (3.3)$$

The Spearman's ρ is also based on concordance and discordance.

Definition 3.1.3. *Let $(X_1, Y_1)'$, $(X_2, Y_2)'$ and $(X_3, Y_3)'$ be independent random vectors with a common joint distribution. Then Spearman's ρ is defined as*

$$\rho = \rho_{X,Y} = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]).$$

Next theorem gives the relation between Spearman's ρ and copula C .

Theorem 3.1.4. *Let X and Y be continuous random variables whose copula is C . Then the Spearman's ρ (denoted by $\rho_{X,Y}$) is given by*

$$\rho_{X,Y} = 12 \int \int_{\mathbf{I}^2} C(u, v) dudv - 3.$$

Theorem 3.1.4 is presented in Nelsen (1999, p. 135). The next theorem, due to Daniels (1950), gives universal inequalities for Kendall's τ and Spearman's ρ .

Theorem 3.1.5. *Let X and Y be continuous random variables, and let $\tau_{X,Y}$ and $\rho_{X,Y}$ be Kendall's τ and Spearman's ρ , respectively. Then*

$$-1 \leq 3\tau_{X,Y} - 2\rho_{X,Y} \leq 1.$$

Theorem 3.1.5 is firstly proven by Kruskal (1958).

3.2 Archimedean and Gaussian copulas

3.2.1 Archimedean copulas

One very important wide class of copulas is known as Archimedean copulas. This class consists of families of one-parameter distributions. Let us introduce the function $\varphi_\theta(z)$ where z is a real argument and $\theta \in \mathfrak{R}$ is a parameter. Assume that φ_θ is a convex, decreasing function from $(0, 1]$ to $[0, \infty)$ such that $\varphi_\theta(1) = 0$.

Definition 3.2.1. *Copula $C(u, v)$ is called Archimedean copula, if*

$$C(u, v) = \varphi_\theta^{-1}(\varphi_\theta(u) + \varphi_\theta(v)), \quad u, v \in (0, 1].$$

Hence the representation of an Archimedean copula allows to reduce the study of a bivariate copula to investigation of a univariate function. The function φ_θ is called the generating function of an Archimedean copula (Nelsen, 1999, p. 92).

Using the equalities (3.2)-(3.3) and Definition 3.2.1 one can present Kendall's correlation coefficient $\tau_{X,Y}$ through the generating function φ_θ (Nelsen, 1999, p.130):

$$\tau_{X,Y} = 1 + 4 \int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} dt. \quad (3.4)$$

Archimedean copulas are divided by type of the function φ_θ into families. Long list of different families of Archimedean copulas is presented in Nelsen (1999, p. 94-97).

In applications two very important families of Archimedean copulas are Clayton (Clayton, 1978) and Gumbel (Gumbel, 1960) families. The generating function φ_θ is defined for a Clayton copula as

$$\varphi_\theta(t) = \frac{t^{-\theta} - 1}{\theta} \quad (3.5)$$

and for a Gumbel copula as

$$\varphi_\theta(t) = (-\ln t)^\theta. \quad (3.6)$$

Using equality (3.4) we can represent parameter θ of these copula families through the Kendall's $\tau_{X,Y}$. For Clayton family of copulas we have

$$\theta = \frac{2\tau_{X,Y}}{1 - \tau_{X,Y}}.$$

For Gumbel copulas we have

$$\theta = \frac{1}{1 - \tau_{X,Y}}.$$

It follows straightforwardly that for Clayton family of copulas Kendall's τ is:

$$\tau_{X,Y} = \frac{\theta}{\theta + 2}.$$

For Gumbel copulas we get

$$\tau_{X,Y} = \frac{\theta - 1}{\theta}.$$

Not always the parameter of Archimedean copula can be presented through Kendall's τ . As an example we refer to Plackett family (Plackett, 1965). Let parameter $\theta > 0$ and $\theta \neq 1$. Then the Plackett's family of copulas is defined as follows:

$$C_\theta(u, v) = \frac{[1 + (\theta - 1)(u + v)] - \sqrt{[1 + (\theta - 1)(u + v)]^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)}.$$

Applying Theorem 3.1.4 to $C_\theta(u, v)$ we get

$$\rho_{X,Y} = \frac{\theta + 1}{\theta - 1} - \frac{2\theta}{(\theta - 1)^2} \ln \theta.$$

So in this case Spearman's ρ can be presented as a function of θ .

3.2.2 Copulas and density functions

Copula theory traditionally characterizes random vector through its distribution function. Common approach to multivariate distributions is based on density functions. Kotz and Seeger (1991) focused to construction of multivariate probability density functions for copulas. In this approach the object of modeling is the density weighting function which gives a joint density, when multiplied by specified marginal densities. In Krzysztofowicz and Kelly (1994) this idea is realized via the normal distribution through so-called normal quantile transform.

For random variable Z with the distribution function $F_Z(x)$ the function $F_Z^{-1}(x)$ is called the quantile function of Z if the inverse function exists. Let us denote the quantile function for Z by Q .

Let $(X, Y)'$ denote a vector of continuous random variables with specified continuous marginal distributions: the density function f and the distribution function F of X and the density function g and the distribution function G of Y . Let Q be the normal quantile function. Then we define standard normal variables V and W through the quantile functions:

$$V = Q(F(X))$$

and

$$W = Q(G(Y)).$$

Let γ be the Pearson correlation coefficient between V and W . The coefficient γ is not invariant under monotone transformations. Fortunately the linear correlation coefficient γ between normally distributed random variables can be uniquely expressed through nonparametric measures of dependence which are invariant under strictly monotone transformation of variables. In particular, Spearman's rank correlation coefficient ρ between V and W , as well as between X and Y , can be expressed through γ :

$$\rho = \frac{6}{\pi} \arcsin\left(\frac{\gamma}{2}\right).$$

Also Kendall's correlation coefficient τ between V and W , as well as between X and Y , has the similar representation:

$$\tau = \frac{2}{\pi} \arcsin(\gamma).$$

Both equalities are due to Kruskal (1958).

The density function $h(x, y)$ of (X, Y) can be expressed as a product

$$h(x, y) = \phi(Q(F(X)), Q(G(Y)))f(x)g(y) \quad (3.7)$$

where

$$\phi(v, w) = \frac{1}{\sqrt{1 - \gamma^2}} \exp\left(\frac{-\gamma(\gamma v^2 - 2vw + \gamma w^2)}{2(1 - \gamma^2)}\right) \quad (3.8)$$

is the density weighting function (Krzysztofowicz and Kelly, 1994). The function $h(x, y)$ in (3.7) is called the density function of the Gaussian copula.

3.2.3 Two-dimensional Gaussian copula

Let us study a Gaussian copula with log-normal and normal marginal distributions. Let us have a random vector $(X, Y)'$ where X is log-normally

distributed with the density function

$$f(x) = \frac{1}{xs\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x)-m}{s}\right)^2}. \quad (3.9)$$

In equality (3.9) m and s denote the mean and the standard deviation of $\ln(X)$, and Y is normally distributed with mean μ and standard deviation σ .

Our aim is to find the joint density function of (X, Y) . We perform the transformation

$$Z = \ln(X) \sim N(m, s^2)$$

and use correspondingly $z = \ln(x)$. Let us denote the Jacobian matrix of the transformation $(x, y) \rightarrow (z, y)$ by J . So we get

$$J = \begin{pmatrix} \frac{1}{x} & 0 \\ 0 & 1 \end{pmatrix}.$$

The Jacobian of our transformation $|J| = \frac{1}{x}$. Let $f_{Z,Y}(z, y)$ and $f_{X,Y}(x, y)$ be the joint density functions of (Z, Y) and (X, Y) respectively. Then

$$f_{X,Y}(x, y) = \frac{1}{x} f_{Z,Y}(z(x, y), y).$$

Using equality (3.7) we get

$$f_{X,Y}(x, y) = \phi(Q(F(\ln(x))), Q(G(y))) f(x)g(y). \quad (3.10)$$

Now we study equality (3.10). After using (3.8) and density expression of the bivariate normal distribution we get

$$\begin{aligned} f_{Z,Y}(z, y) &= \frac{1}{2\pi x s \sigma} \frac{1}{\sqrt{1-\gamma^2}} \exp\left(\frac{-\gamma(2\gamma \ln(v) - 2 \ln(v)w + \gamma w^2)}{2(1-\gamma^2)}\right) \\ &\times \exp\left(-\frac{1}{2}\left(\frac{(\ln x - m)^2}{s^2} + \frac{(y - \mu)^2}{\sigma^2}\right)\right). \end{aligned} \quad (3.11)$$

In (3.11) γ is the linear correlation coefficient between $\ln(V)$ and W .

3.3 Application of copulas for distribution approximation

In this section we apply copula models to the forestry data described in subsection 2.3.3. Our aim is to model the relation between tree's height

H and diameter at the breast height (DBH) by means of copulas. Several Archimedean copulas and the Gaussian copula were used in modeling of the bivariate distribution of interest.

3.3.1 Goodness-of-fit test for Archimedean copulas

We are going to show how to construct a multivariate Kolmogorov-Smirnov goodness-of-fit test for Archimedean copulas.

Approximation of the bivariate empirical distribution function with Archimedean copulas is one widely used way of modeling. A proper $\varphi_\theta(z)$ has to be found for this kind of approximation.

Testing concordance between an Archimedean copula and real data is based on the following construction (Genest and Rivest, 1993). Let us have a random sample:

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n).$$

Define the random variables Z_i ,

$$Z_i = \frac{\#\{(X_j, Y_j) : X_j < X_i \& Y_j < Y_i\}}{n-1}, \quad i = 1, 2, \dots, n.$$

From a realization $(x_i, y_i), i = 1, 2, \dots, n$ now we get z_1, z_2, \dots, z_n as values of the i.i.d. random variables Z_1, Z_2, \dots, Z_n . Then we construct the empirical distribution function of the artificial data:

$$K_n(z) = \frac{\#\{z_i : z_i < z\}}{n}.$$

It is shown (Frees and Valdez, 1998) that the distribution function of Z can be presented through the generating function $\varphi_\theta(z)$:

$$K(z) = z - \frac{\varphi_\theta(z)}{\varphi'_\theta(z)}.$$

Fitting of the copula $C(u, v)$ with the data $(x_i, y_i), i = 1, 2, \dots, n$, can be estimated by means of the Kolmogorov-Smirnov statistic

$$D = \sup_z |K_n(z) - K(z)| \tag{3.12}$$

where n is the sample size.

3.3.2 Multivariate Kolmogorov-Smirnov test

The method of Genest and Rivest (1993), described in previous subsection, is generalized to the multivariate case by Justel, Peña and Zamar (1997).

Let \mathbf{X} be a random p -vector with joint distribution function F . For \mathbf{X} the next theorem holds due to Rosenblatt (1952).

Theorem 3.3.1. *Let $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ be a random vector with joint density function*

$$f(x_1, x_2, \dots, x_p) = f_1(x_1)f_2(x_2|x_1) \dots f_p(x_p|x_1, \dots, x_{p-1}),$$

and define the transformation $\mathbf{Y} = T(\mathbf{X})$ by

$$Y_1 = F(X_1),$$

$$Y_i = F(X_i|X_1, \dots, X_{i-1}), \quad i = 2, \dots, p.$$

Then Y_i , $i = 1, 2, \dots, p$ are i.i.d. random variables with $Y_i \sim U(0, 1)$.

The construction of multivariate Kolmogorov-Smirnov test is based on Theorem 3.3.1. The straightforward extension of the Kolmogorov-Smirnov test statistic (3.12) to the multivariate case is given by equality

$$d_n = \sup_{\mathbf{y}} |G_n(\mathbf{y}) - y_1 y_2 \dots y_p| \quad (3.13)$$

where n is the sample size and $G_n(\mathbf{y})$ is the empirical distribution function for transformed data y_i . The statistic defined by (3.13) is not invariant under permutations. That means, relabeling of the components of \mathbf{X} would give a different transformation $\mathbf{y} = T(\mathbf{x})$ and, therefore, different value of (3.13). To remove this disadvantage the following sequence of transformations is defined (Justel et. al, 1997):

$$y_1^j = F(z_1^j),$$

$$y_2^j = F(z_2^j|z_1^j),$$

...

$$y_p^j = F(z_p^j|z_1^j, z_2^j, \dots, z_{p-1}^j)$$

where $z_1^j, z_2^j, \dots, z_p^j$, for $j = 1, 2, \dots, p!$, is the j th permutation of the values x_1, x_2, \dots, x_p . For every sequence the following value is calculated:

$$d_n^j = \max_{\mathbf{y}^j} |G_n(\mathbf{y}^j) - y_1^j y_2^j \dots y_p^j|. \quad (3.14)$$

The multivariate Kolmogorov-Smirnov test statistic is given by the next equality

$$D_n = \max_{j=1,2,\dots,p!} d_n^j. \quad (3.15)$$

So the multivariate Kolmogorov-Smirnov test consists of the following steps:

1) for every permutation of coordinates of the vector $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ we perform transformation $\mathbf{Y} = T(\mathbf{X})$ where T is defined in Theorem 3.3.1;

2) for every permutation $j = 1, 2, \dots, p!$ we find d_n^j by equality (3.14);

3) finally we find test statistic D_n by equality (3.15).

3.3.3 Modeling the joint distribution of H and DBH

In this subsection we present results of modeling of the joint distribution of tree's height Y and its diameter at the breast height (DBH). Let us denote the tree's DBH by X . Let $F(x)$ and $G(y)$ be distribution functions and $f(x)$ and $g(y)$ density functions, of X and Y respectively. According to the Sklar's theorem (Theorem 3.1.1) we can find a copula $C(F(x), G(y))$ to model the bivariate distribution of interest. We estimate first the probability density functions f and g . It came out that Y can be modeled by normal distribution and X follows a log-normal distribution. So the density function of X is described by equality (3.9). The estimated values of the parameters of the model distributions are presented in the following two tables as well as the results of goodness-of-fit test by the Kolmogorov-Smirnov criteria.

Table 3.1 (Adermann and Pihlak, 2005) Estimation of marginal distribution of X

Species	Density class	Parameters	Test statistic	P-value
Pine	low	$m = 2.99$ $s = 0.532$	0.014	0.061
Pine	high	$m = 2.97$ $s = 0.633$	0.0135	0.074
Spruce	low	$m = 2.96$ $s = 0.542$	0.016	0.059
Spruce	high	$m = 2.75$ $s = 0.599$	0.027	0.083

Table 3.2 (Adermann and Pihlak, 2005) Estimation of marginal distribution of Y

Species	Density class	Parameters	Test statistic	P-value
Pine	low	$\mu = 16.0 \sigma = 6.31$	0.013	0.094
Pine	high	$\mu = 17.9 \sigma = 6.70$	0.013	0.094
Spruce	low	$\mu = 17.4 \sigma = 6.74$	0.015	0.067
Spruce	high	$\mu = 15.7 \sigma = 6.30$	0.021	> 0.15

The fourth column in Tables 3.1-3.2 consists of values of the Kolmogorov-Smirnov test statistic, which measures the maximal difference between empirical and theoretical distribution functions. The fifth column of these tables gives P-values of the Kolmogorov-Smirnov test.

Now we find the joint density function of X and Y using the equalities (3.7), (3.8) and (3.11). The results of approximation with Gaussian copulas are presented in Figures 3.1-3.2. The parameter γ is the Pearson's correlation coefficient in equality (3.8). In these figures the asymmetry of the joint distribution is caused by the log-normality of X .

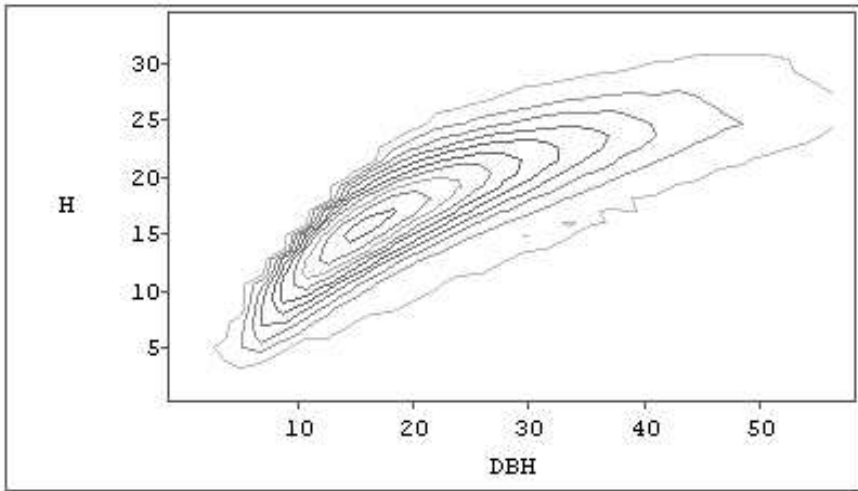


Figure 3.1 (Pihlak and Adermann, 2005) Joint distribution of H (in m) and DBH (in cm) by Gaussian copula for spruce in high density class, $\gamma = 0.889$

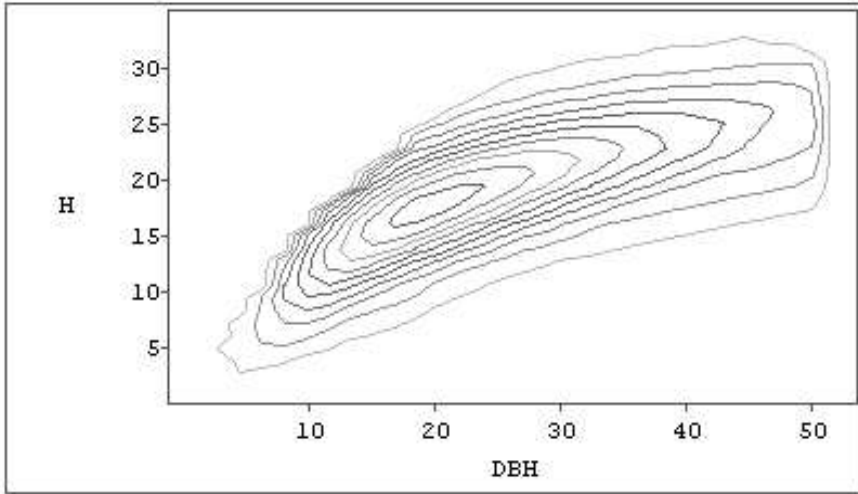


Figure 3.2 (Pihlak and Adermann, 2005) Joint distribution of H (in m) and DBH (in cm) for pine by Gaussian copula in high density class, $\gamma = 0.838$

The results of modeling with Clayton and Gumbel copulas are presented in the following table.

Table 3.3 (Adermann and Pihlak, 2005). The results of modeling with Archimedean copulas

Species	Density class	Model	θ	D
spruce	low	Clayton	4.43	0.0368
spruce	high	Clayton	4.60	0.0636
spruce	low	Gumbel	3.22	0.0582
spruce	high	Gumbel	3.30	0.0531
pine	low	Clayton	2.66	0.0287
pine	high	Clayton	3.44	0.0329
pine	low	Gumbel	2.33	0.0694
pine	high	Gumbel	2.72	0.0671

The third column in Table 3.3 indicates the type of Archimedean copula. The value 'Clayton' means the Clayton type copula (3.5) and 'Gumbel' the Gumbel type copula (3.6). The values in the fifth column of this table are calculated by equality (3.12).

From Table 3.3 one can see that the Clayton type copula gives for pine better fit than the Gumbel type copula. For spruce the Clayton type copula

is the best for the low density class. The high density class for spruce is better modeled by the Gumbel type copula.

3.4 Comparing different models of distribution

In this section we compare the models obtained by approximation based on normal distribution and the models obtained by Gaussian copulas. For this comparison we use the bivariate Kolmogorov-Smirnov test.

3.4.1 Bivariate Kolmogorov-Smirnov test

For comparison of two different models we use the equality (3.15) in the bivariate case. Let us denote as before the tree's DBH by X and the tree's height by Y . Let F be the joint distribution function of (X, Y) and n the sample size. Let us standardize random variables X and Y . Then we define random variables $Z_1^1 = F(X)$, $Z_2^1 = F(Y|X)$ and $Z_1^2 = F(Y)$, $Z_2^2 = F(X|Y)$. From these random variables we form random vectors $\mathbf{Z}^1 = (Z_1^1, Z_2^1)'$ and $\mathbf{Z}^2 = (Z_1^2, Z_2^2)'$. Following (3.13) we construct the values of test statistics

$$d_n^1 = \max_{1 \leq i \leq n} |G_{\text{emp}}(z_i^1) - z_{1i}^1 z_{2i}^1|$$

and

$$d_n^2 = \max_{1 \leq i \leq n} |G_{\text{emp}}(z_i^2) - z_{1i}^2 z_{2i}^2|$$

where $z_i^1 = (z_{1i}^1, z_{2i}^1)$, $z_i^2 = (z_{1i}^2, z_{2i}^2)$, $z_{1i}^1, z_{2i}^1, z_{1i}^2$ and z_{2i}^2 are the i th realizations of random variables Z_1^1, Z_2^1, Z_1^2 and Z_2^2 , respectively, $i = 1, \dots, n$. The function G_{emp} denotes the empirical distribution function of \mathbf{Z}^1 or \mathbf{Z}^2 . Using (3.15) we get the value of the bivariate statistic for goodness-of-fit test

$$D_n = \max\{d_n^1, d_n^2\}. \quad (3.16)$$

Let us consider conditional distribution functions $F(y|x)$ and $F(x|y)$. It follows straightforwardly that

$$F(x|y) = \int_{-\infty}^x f(u|y) du = \frac{1}{f(y)} \int_{-\infty}^x f(u, y) du$$

where $f(y)$, $f(u|y)$ and $f(u, y)$ are marginal, conditional and joint density functions, respectively. Using the equalities (3.7)-(3.8) we get

$$F(x|y) = \int_{-\infty}^x \frac{1}{\sqrt{1-\gamma^2}}$$

$$\times \exp\left(\frac{-\gamma(\gamma Q(F(u))^2 - 2Q(F(u))Q(G(y)) + \gamma Q(G(y))^2)}{2(1 - \gamma^2)}\right) f(u) du \quad (3.17)$$

and

$$F(y|x) = \int_{-\infty}^x \frac{1}{\sqrt{1 - \gamma^2}} \times \exp\left(\frac{-\gamma(\gamma Q(F(x))^2 - 2Q(F(x))Q(G(u)) + \gamma Q(G(u))^2)}{2(1 - \gamma^2)}\right) g(u) du \quad (3.18)$$

where $F(x)$ and $G(y)$ are the marginal distribution functions of X and Y , $f(x)$ and $g(y)$ are the marginal density functions of X and Y , Q is the quantile function and γ is the Pearson's correlation coefficient between $Q(F(X))$ and $Q(G(Y))$. The integrals (3.17)-(3.18) can be found numerically.

In the case of Edgeworth type expansion we use equalities (2.10) and (2.33). We get

$$F(x|y) = \left(1 - \frac{1}{6}(\text{vec}'\mathbf{C3}, \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)^{\otimes 3}) + \dots\right) \Phi(x|y) \quad (3.19)$$

and

$$F(y|x) = \left(1 - \frac{1}{6}(\text{vec}'\mathbf{C3}, \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)^{\otimes 3}) + \dots\right) \Phi(y|x) \quad (3.20)$$

where $\mathbf{C3}$ in the third order cumulant of (X, Y) and $\Phi(x|y)$ is the conditional distribution function of the standard normal distribution. Using the equality (2.20) we get

$$\Phi(x|y) = \Phi\left(\frac{x - y\rho}{\sqrt{1 - \rho^2}}\right)$$

and

$$\Phi(y|x) = \Phi\left(\frac{y - x\rho}{\sqrt{1 - \rho^2}}\right)$$

where ρ is the Pearson's correlation coefficient between X and Y . It follows from equalities (3.19)-(3.20) that we have to find the third order derivatives and mixed derivatives of $\Phi\left(\frac{x - y\rho}{\sqrt{1 - \rho^2}}\right)$ and $\Phi\left(\frac{y - x\rho}{\sqrt{1 - \rho^2}}\right)$ by x and y to get expressions for the conditional distribution functions $F(x|y)$ and $F(y|x)$. Let $g(x)$ and $g(y)$ be defined by equalities (2.17) and (2.18). Let ρ be the Pearson's correlation coefficient between X and Y and

$$\mathbf{G12} = \begin{pmatrix} (\mathbf{C3})_{12} + (\mathbf{C3})_{21} + (\mathbf{C3})_{31} \\ (\mathbf{C3})_{22} + (\mathbf{C3})_{32} + (\mathbf{C3})_{41} \end{pmatrix}.$$

Then we get

$$\begin{aligned}
F(x|y) = \Phi(g(y)) - \frac{1}{6} \left\{ \left((\mathbf{C3})_{11} \left(\frac{1}{\sqrt{1-\rho^2}} \right)^3 - (\mathbf{G12})_1 \left(\frac{1}{\sqrt{1-\rho^2}} \right)^2 \frac{r}{\sqrt{1-\rho^2}} \right. \right. \\
+ (\mathbf{G12})_2 \frac{1}{\sqrt{1-\rho^2}} \left(\frac{r}{\sqrt{1-\rho^2}} \right)^2 - (\mathbf{C3})_{42} \left(\frac{r}{\sqrt{1-\rho^2}} \right)^3 \Big) \\
\left. \left. (g(y)^2 - 1) \frac{1}{\sqrt{2\pi}} \exp(0.5g(y)^2) \right\} \quad (3.21)
\end{aligned}$$

and

$$\begin{aligned}
F(y|x) = \Phi(g(x)) - \frac{1}{6} \left\{ \left((\mathbf{C3})_{11} \left(\frac{1}{\sqrt{1-\rho^2}} \right)^3 - (\mathbf{G12})_1 \left(\frac{1}{\sqrt{1-\rho^2}} \right)^2 \frac{r}{\sqrt{1-\rho^2}} \right. \right. \\
+ (\mathbf{G12})_2 \frac{1}{\sqrt{1-\rho^2}} \left(\frac{r}{\sqrt{1-\rho^2}} \right)^2 - (\mathbf{C3})_{42} \left(\frac{r}{\sqrt{1-\rho^2}} \right)^3 \Big) \\
\left. \left. (g(x)^2 - 1) \frac{1}{\sqrt{2\pi}} \exp(0.5g(x)^2) \right\}. \quad (3.22)
\end{aligned}$$

3.4.2 Results

In this subsection we compare different approximations on our example of forestry data and summarize the results of approximation of multivariate distribution functions. We compare the models of approximation obtained by Edgeworth expansion and by copulas. For comparison we use the bivariate Kolmogorov-Smirnov test. To test goodness-of-fit of copula approximation we integrate numerically equalities (3.17) and (3.18) to find the conditional distributions. For testing goodness-of-fit of Edgeworth type approximation we find the conditional distributions by means of equalities (3.21) and (3.22). The results of goodness-of-fit test are presented in Table 3.4.

Table 3.4 The results of the Kolmogorov-Smirnov test for Edgeworth type approximation and for Gaussian copula

Species	Density	Dn_{edg}	Dn_{cop}
spruce	low	0.0859	0.1279
spruce	high	0.0839	0.1164
pine	low	0.1302	0.1555
pine	high	0.1276	0.1726

In Table 3.4 the value Dn_{edg} equals to the value of the Kolmogorov-Smirnov statistic for the Edgeworth type approximation (equality (2.33)), and the value Dn_{cop} is the corresponding value for the Gaussian copula ((3.7)-(3.8)). Both values are found following equality (3.16).

As can be seen from Table 3.4 both models fit for spruce better than for pine. The density approximation by the copula model is not as good as the Edgeworth type approximation. From the third and fourth columns of Table 3.4 we can see that the model obtained by Edgeworth type expansion fits better than the model obtained by Gaussian copula. So we can conclude that the equality (2.33) describes better the asymmetry of DBH than the log-normal distribution in the equality (3.7).

3.4.3 Discussion and further development

Both methods, Edgeworth expansion and copulas, have their advantages and disadvantages. The disadvantage of Edgeworth type expansions is the complexity of formulas of approximation (equalities (2.16), (2.19) and (2.33)). The formulas of Gaussian copulas are much simpler. But for approximation by means of Gaussian copulas we need more information. The marginal density functions have to be estimated and the correlation between two random variables has to be found. The approximation by means of Taylor series is based on general theory. Another disadvantage of approximation by means of copulas is uncertainty. No rules exist for finding the best class of copulas. This kind of approximation is based mainly on empirical approach and intuition. Despite of complexity of equalities (2.19) and (2.33) they can be easily calculated by means of software SAS IML (Statistical Analyse System Interactive Matrix Language).

Finally we make some comments for further developments. One idea for further studies is to generalize equality (2.19) to higher-dimensional cases, firstly to the three-dimensional case. Applying equality (2.16) to approximation of an unknown distribution with Wishart distribution would also be one possible idea for further development. For example, in Kollo and von Rosen (1995) approximation of skewed multivariate density with density function of the Wishart distribution is explored. Also other non-symmetric distributions could be used in such a situation - skew normal and skew-elliptical, for example. On skew-normal density approximation Gupta, Kollo (2003) and Kollo, Selart (2004) can be mentioned. Second direction for further development is deeper study of the copula theory. Especially the theory of the choice of generating function $\phi_\theta(t)$ for Archimedean copulas

needs development and generalization to the multivariate case. It would be interesting to apply our method in data analysis. Application of our technique to different practical problems would most probably create new ideas for further theoretical development.

In future we plan apply our technique in approximation of distribution functions for different problems in life sciences.

Bibliography

- [1] Adermann, V., Pihlak, M. (2005) Using copulas for modeling the dependence between tree height and diameter at breast height. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, **9**, 77-85.
- [2] Anderson, T. W. (2003) *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.
- [3] Clayton D. G. (1978) A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141-151.
- [4] Cornish, E. A., Fisher, R. A. (1937) Moments and cumulants in the specification of distribution. *Revue de l'Institut International de Statistique*, **5**, 307-322.
- [5] Daniels, H. E. (1950) Rank correlation and population models *Journal of Royal Statistical Society, Series B*, **12**, 171-181.
- [6] Dwyer, P. J., MacPhail, M. S. (1948) Symbolic matrix derivatives. *The Annals of Mathematical Statistics*, **19**, 517-534.
- [7] Fisher, R. A., Cornish, E. A. (1960) The percentile points of distributions having known cumulants. *Technometrics*, **2**, No. 2, 209-225.
- [8] Frees, E. W., Valdez, E. A. (1998) Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 143-168.
- [9] Genest, C., Rivest L.-P. (1993) Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association*, **88**, No. 423, 1034-1043.
- [10] Gumbel, E. J. (1960) Distributions de valeurs extrêmes en plusieurs dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, **9**, 171-173.

- [11] Gupta, A. K., Kollo, T. (2003) Density expansions based on the multivariate skew normal distribution. *Sankhyā. The Indian Journal of Statistics*, **65**, 821-835.
- [12] Harville, A. (1997) *Matrix Algebra from a Statistician's Perspective*. Springer, New York.
- [13] Iital, A., Stålnacke, P., Deelstra, J., Loigu, E., Pihlak, M. (2005) Effects of large-scale changes in emissions on nutrient concentrations in Estonian rivers in the Lake Peipsi drainage basin. *Journal of Hydrology*, **304**, 261-273.
- [14] Justel, A., Peña, D., Zamar, R. (1997) A multivariate Kolmogorov-Smirnov test of goodness-of-fit. *Statistics and Probability Letters*, **35**, 251-259.
- [15] Kollo, T. (1991) *Matrix Derivative in Multivariate Statistics*. Tartu University Press, Tartu (in Russian).
- [16] Kollo, T., von Rosen, D. (1995) Approximating by the Wishart distribution. *Annals of the Institute of Statistical Mathematics*, **47**, 767-783.
- [17] Kollo, T., von Rosen, D. (1998) A unified approach to the approximation of multivariate densities. *Scandinavian Journal of Statistics*, **25**, 93-109.
- [18] Kollo, T., Selart, A. (2004) Density expansion for correlations and eigenvalues of the covariance matrix. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, **8**, 155-168.
- [19] Kollo, T., von Rosen, D. (2005) *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht.
- [20] Kruskal, W. H. (1958) Ordinary measures of association. *Journal of the American Statistical Association*, **53**, 814-861.
- [21] MacRae, E. C. (1974) Matrix derivatives with an applications to an adaptive linear decision problem. *The Annals of Statistics*, **7**, 381-394
- [22] Magnus, J. R., Neudecker, H. (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics.
- [23] Näslund, M. (1941) Funktioner och tabeller för kubering avstående träd. tall, gran och björk i norra Sverige. *Meddelanden från Statens Skogsförsöksanstalt*, **32**, 87-142.

- [24] Nelsen, R. B. (1999) *An Introduction to Copulas*. Springer-Verlag, New-York.
- [25] Neudecker, H. (1969) Some theorems on matrix differentiations with special reference to Kronecker matrix products. *Journal of the American Statistical Association*, **64**, 953-963.
- [26] Pihlak, M. (2004) Matrix integral. *Linear Algebra and Its Applications*, **388**, 315-325
- [27] Pihlak, M. (2006a) Approximation of multivariate distribution functions. *Acta Applicandae Mathematicae*, 16 pages (submitted).
- [28] Pihlak, M. (2006b) Approximation of multivariate distribution functions. In: *9th International Vilnius Conference on Probability Theory and Mathematical Statistics*, Abstracts of Communications, Vilnius, Lithuania, June 26 to June 30, 2006, p. 261.
- [29] Plackett, R. L. (1965) A class of bivariate distributions. *Journal of the American Statistical Association*, **60**, 516-522.
- [30] Rao, C. R., Rao, M. B. (1998) *Matrix Algebra and Its Applications to Statistics and Econometrics*. World Scientific, Singapore.
- [31] Rosenblatt, M. (1952) Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, **23**, 470-472.
- [32] Schott, R. (1997) *Matrix Analysis for Statistics*. Wiley, New York.
- [33] Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229-231.
- [34] Traat, I. (1984) Moments of the sample covariance matrix. *Proceedings of the Computer Centre of Tartu State University*, **51**, 108-126 (in Russian).
- [35] Traat, I. (1986) Matrix calculus for multivariate distributions. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 64-83.
- [36] Wishart, J. (1928) The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, **20A**, 32-52.

Mitmemõõtmeliste jaotusfunktsioonide lähendamine

Kokkuvõte

Käesoleva dissertatsiooni probleemiks on juhusliku vektori jaotuse lähendamine. Tavaliselt esitatakse jaotuste lähendamisel seos, mis esitab uuri-tava keeruka tihedusfunktsiooni teise lihtsama tihedusfunktsiooni kaudu. Selles töös on meie eesmärk töötada välja meetod jaotusfunktsioonide lähendamiseks. Kaasaegne mitmemõõtmeline statistika baseerub maatriksalgebral. Seetõttu on märgatav osa dissertatsioonist pühendatud vajaliku maatriksalgebra uuematele tulemustele.

Esimese peatüki põhitulemused on esitatud artiklis Pihlak (2004). Peatüki alguses tuuakse sisse maatriksalgebra vajalikud põhilised tulemused ning tähistused, mida töös kasutatakse. Kõige olulisemad kasutatavad mõisted on vec-operaator, kommutatsioonimaatriks, otsekorrutis ja maatrikstuletis, mis on esitatud koos omadustega. Seejärel tuuakse sisse uus maatriksoperatsioon, mida nimetatakse maatriksintegraaliks. Definiitsioonis 1.4.1 defineeritakse maatriksintegraal kui maatrikstuletise pöördoperatsioon. Osutub, et maatriksintegraali leidmisel saab kasutada maatriksite tähtkorrutist, maatriksoperatsiooni, mille defineeris E. MacRae 1974. aastal (MacRae, 1974).

Teine peatükk on pühendatud jaotusfunktsioonide lähendamismeetodi väljatöötamisele. Olulised autori tulemused selles osas on esitatud artiklis Pihlak (2006a) ja ette kantud Vilniuse rahvusvahelisel tõenäosusteooria ja matemaatilise statistika konverentsil (Pihlak, 2006b). Peatüki esimene paragrahv on referatiivne. Selles paragrahvis käsitletakse tihedusfunk-

sioonide lähendamist Tayloriga reaalsetel andmetel. Materjali esitus baseerub artiklil Kollo ja von Rosen (1995). Peatüki teises paragrahvis rakendatakse maatriksintegraali leidmaks tihedusfunktsioonide vahelisest seosest vastavate jaotusfunktsioonide vahelist seost. Jaotusfunktsioonide vaheline seos on esitatud Teoreemis 2.2.3. Kolmandas paragrahvis kasutatakse saadud tulemusi suvalise kahemõõtmelise jaotusfunktsiooni esitamiseks kahemõõtmelise normaaljaotuse kaudu. Lõpuks rakendatakse leitud jaotusfunktsiooni lähendit reaalsel andmel. Lähendatakse puu kõrguse ja rinnasdiameetri ühisjaotust kahemõõtmelise normaaljaotuse baasil leitud rittaarendusega. Vastav andmestik on kirjeldatud artiklis Adermann ja Pihlak (2005).

Kolmas peatükk käsitleb tundmatu jaotuse lähendamist koopulate abil. Koopula kujutab endast ühtlase jaotusega marginaalide ühisjaotust. Formaalselt on koopula defineeritud Definitsiooniga 3.1.1. Peatükk sisaldab artikli Adermann ja Pihlak (2005) tulemusi. Artikli esimeselt autorilt on pärit andmed ja metsandusuuringutes kasutatava klassikalise analüüsimeetodika kirjeldus. Meetodite võrdlus ja koopulate valik ning statistiline analüüs on teostatud teise autori poolt. Esimeses paragrahvis antakse ülevaade koopulate teooria põhilistest tulemustest. Nendest üks olulisemaid on Sklar'i teoreem (Sklar, 1959), mis annab koopula ja kahemõõtmelise jaotusfunktsiooni vahelise seose. Teises paragrahvis antakse ülevaade kahest koopulate klassist - arhimeedilistest koopulatest ja Gaussi koopulatest. Kolmandas paragrahvis lähendatakse puu kõrguse ja rinnasdiameetri ühisjaotust arhimeediliste ja Gaussi koopulaga. Arhimeediliste koopulate erinevatel juhtudel kasutati lähendamisel Claytoni (Clayton, 1978) ja Gumbeli (Gumbel, 1960) koopulaid. Samuti rakendatakse Kolmogorov-Smirnovi testi mitmemõõtmelisel juhul. Peatüki neljandas paragrahvis võrreldakse Kolmogorov-Smirnovi testiga Tayloriga rittaarendamisel saadud uut meetodit ja koopulate abil saadud lähendeid. Selgub, et Tayloriga reaalsetel andmetel saadud lähendite klassi valikuks, mis on vajalik koopulatega lähendamise korral.

Curriculum Vitae

Margus Pihlak

Citizenship: Estonian Republic

Born: October 27, 1970, Tallinn, Estonia

Marital Status: single

Address: Uus 58-43, Tartu 50606, Estonia

Contacts: e-mail: margus.pihlak@ut.ee

Education

1989-1994 Faculty of Biology and Geography, University of Tartu, BSc in Biology

1995-1999 Faculty of Mathematics, University of Tartu, BSc in Mathematical Statistics

1999-2001 Faculty of Mathematics, University of Tartu, MSc in Mathematical Statistics

2001-2006: Faculty of Mathematics and Computer Science, University of Tartu, PhD studies in Mathematical Statistics

Professional employment

Since 2000 Engineer, Estonian Marine Institute, University of Tartu

2001-2002 Specialist, Estonian Statistical Office

Since 2005 Lecturer, Institute of Mathematics, Tallinn Technical University

Curriculum Vitae

Margus Pihlak

Kodakondsus: Eesti Vabariik

Sünniaeg ja -koht: 27. oktoober 1970, Tallinn, Eesti

Perekonnaseis: vallaline

Aadress: Uus 58-43, Tartu 50606, Eesti

Kontaktandmed: e-mail: margus.pihlak@ut.ee

Hariduskäik

1989-1994 Tartu Ülikooli bioloogia-geograafiateaduskond, bakalaureusekraad bioloogia

1995-1999 Tartu Ülikooli matemaatikateaduskond, bakalaureusekraad matemaatilise statistika erialal

1999-2001 Tartu Ülikooli matemaatikateaduskond, magistrikraad matemaatilise statistika erialal

2001-2006 Tartu Ülikooli matemaatika-informaatikateaduskond, doktoriõpingud matemaatilise statistika erialal

Erialane teenistuskäik

Alates 2000 insener, Eesti Mereinstituut, Tartu Ülikool

2001-2002 peaspetsialist, Eesti Statistikaamet

Alates 2005 lektor, matemaatikainstituut, Tallinna Tehnikaülikool