Using Edgeworth expansion approximating two- and three-dimensional probability distribution functions

Margus Pihlak
Talinn University of Technology
Department of Mathematics
Ehitajate tee 5 Tallinn 19086
Email: margusp@staff.ttu.ee

***Abstract*—In this article we study the approximation of multivariate distribution functions by means of Taylor series. We generalize two-dimensional Edgeworth expansion to three-dimensional case. We develop the results presented in the paper [11].**

I. Introduction

In this paper we present one method to approximate the unknown multivariate distribution function with known distribution function. The method is based on Taylor expansion. This method is in last 20 years relatively well studied. The idea of approximation by means of Taylor series in univarite case was suggested by R. A. Fisher and E. A. Cornish in [1]. The approximation distribution functions of random variables in multivariate case needs application of results of matrix algebra. 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: P. J. Dwyer and M. S. MacPhail [2], H. Neudecker ([9]), E. C. MacRae ([7]) and T. Kollo ([6]). In the papers T. Kollo and D. von Rosen ([4] and [5]), 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. 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 [10] where matrix integral is introduced and its basic properties studied.

The aim of this paper is to develop the results presented in [11].

The paper is organized in the following way. In Sections 2 and 3 we study results of matrix algebra applied on Edgeworth expansion. In Section 4 we present so called approximating operators in three-dimensional case. In Section 5 we realize Edgeworth expansion in two- and three-dimensional cases.

II. Preparation

In this section we present the results of matrix algebra applied on Edgeworth expansion.

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}. 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. A p-vector with zeros as coordinates is denoted as $\mathbf{0}_p$.

Now we describe main matrix operations from so-called newer matrix algebra ([8]). If we handle partitioned matrix $\mathbf{X}$ then its blocks in the i-th row and j-th column of block is denoted by $[\mathbf{X}]_{ij}$.

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

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

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:

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

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 $\mathbf{a}^{\otimes k}$,

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

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:

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

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

Let the elements of the matrix $\mathbf{Y} : r \times s$ be functions of matrix $\mathbf{X} : p \times q$. Assume that for all $i = 1, 2, ..., p,\ j =$

$1, 2, ..., q,\ k = 1, 2, ..., r$ and $l = 1, 2, ..., s$ 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.

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*

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

where

$$\frac{d}{d\mathrm{vec}'\mathbf{X}} = (\frac{\partial}{\partial x_{11}}, \ldots, \frac{\partial}{\partial x_{p1}}, \ldots, \frac{\partial}{\partial x_{1q}}, \ldots, \frac{\partial}{\partial x_{pq}}).$$

There exists also another widely used form of the matrix derivative. The matrix derivative defined by MacRae ([7]) keeps the structure of involved matrices.

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

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

In approximation of multivariate distribution functions the inverse operation of matrix derivative is needed. The notion "matrix integral" has been introduced in [10]. In the paper [10] beside basic properties of the matrix integral several examples are given to demonstrate practical usage of the notion.

The definition of matrix integrals based on the MacRae's matrix derivative is defined as follows.

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

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

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 **??** is used also to define the definite matrix integral.

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.

The matrix integral can find by means of star product of matrices. his operation is introduced in [7]. She has denoted this operation by $*$.

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.

The technic how to find the matrix integral by means of star product is presented in [10].

III. Matrix techniques on approximation of distributions

First we study cumulants of a random vector. Let us have a random p-vector $\mathbf{X}$ with coordinates $X_i,\ i = 1, 2, ..., p$. Let $\mathbf{x}$ be a realization of this vector. The characteristic function of the random vector $\mathbf{X}$ is defined as follows:

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

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

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

The k-th order cumulant $c_k(\mathbf{X})$ of $\mathbf{X}$ is the k-th matrix derivative of the cumulant function:

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

The technique of two density values integration is presented in [11]. In this paper the general equation between two distribution functions are presented. We can present this equation present formally as follows:

$$F_{\mathbf{Y}}(\mathbf{x}) = (1 - (\mathbf{a}, \frac{d}{d\mathbf{x}}) + (\mathrm{vec}\mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2}) - (\mathrm{vec}\mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3}) + \ldots) F_{\mathbf{X}}(\mathbf{x}) \tag{2}$$

where F_X in known distribution function, F_Y is unknown distribution function, p-vector

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

$p \times p$-matrix

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

and $p^2 \times p$-matrix

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

Matrix $\mathbf{C}$ can also be considered as a partitioned matrix consisting of p blocks where each block is a $p \times p$-matrix. For statistical meanings of vector $\mathbf{a}$ and matrices $\mathbf{B}$ and $\mathbf{C}$ interested reader can see in [4].

IV. APPROXIMATION IN THREE-DIMENSIONAL CASE

In this section we apply the operator described by equality (2) in three-dimensional case. Let $f_{\mathbf{X}}(\mathbf{x})$ be the probability density function of random vector $\mathbf{X}$. Let the marginal density functions are denoted as $f_i(x_i)$ and $f_ij(x_i, x_j)$ where $i, j = 1, 2, ..., p$. Then for part of Mean value we get

$$(\mathbf{a}, \frac{d}{d\mathbf{x}})F_{\mathbf{X}}(\mathbf{x}) = -a_1 f_{\mathbf{X}}(\mathbf{x})+$$
$$+(a_1 - a_2)f_2(x_2)F(x_1, x_3|x_2)+$$
$$+(a_1 - a_3)f_3(x_3)F(x_1, x_2|x_3). \quad (3)$$

Part of variance can present as follows:

$$(\text{vec}\mathbf{B}, \frac{d}{d\mathbf{x}}^{\otimes 2})F_{\mathbf{X}}(\mathbf{x}) = \sum_{i=1}^{3} \frac{\partial f_i(x_i)F(\mathbf{x}_{-i}|x_i)}{\partial x_i}+$$
$$+2b_{12}F(x_3|x_1, x_2)f(x_1, x_2)+$$
$$+2b_{13}F(x_2|x_1, x_3)f(x_1, x_3)+$$
$$+2b_{23}F(x_1|x_2, x_3)f(x_2, x_3). \quad (4)$$

Part of skewness is presented by the following form:

$$(\text{vec}\mathbf{C}, \frac{d}{d\mathbf{x}}^{\otimes 3})F_{\mathbf{X}}(\mathbf{x}) =$$
$$-9\sum_{i,j=1}^{3}(c_{(j,j)(1,i)} + c_{(j,i)(1,j)} + c_{(i,j)(1,j)})\times$$
$$\times\Big(\frac{\partial f_{ij}(x_i, x_j)}{\partial x_j}F(\mathbf{x}_{-i-j}|x_i, x_j)+$$
$$+f_{ij}(x_i, x_j)\frac{\partial F(\mathbf{x}_{-i-j}|x_i, x_j)}{\partial x_j}\Big)-$$
$$-9\sum_{i,j=1}^{3}(c_{(i,i)(1,j)} + c_{(i,j)(1,i)} + c_{(j,i)(1,i)})\times$$
$$\times\Big(\frac{\partial f_{ij}(x_i, x_j)}{\partial x_i}F(\mathbf{x}_{-i-j}|x_i, x_j)+$$
$$+f_{ij}(x_i, x_j)\frac{\partial F(\mathbf{x}_{-i-j}|x_i, x_j)}{\partial x_i}\Big) - 6f_{\mathbf{X}}(\mathbf{x}). \quad (5)$$

The operators (3)-(5) in p-dimensional case are presented in [11].

V. EDGEWORTH EXPANSION

Let us introduce for p-variate normal distribution with mean value $\mathbf{0}_p$ and covariance matrix $\mathbf{\Sigma}$ the notation $N_p(0, \mathbf{\Sigma})$. Let σ_{ij} be the element of ith row on jth of matrix $\mathbf{Sigma}$. In [11] the unknown bivariate distribution function is approximated through the normal distribution $N_2(0, \mathbf{\Sigma})$.

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 [3]. We call this type of approximation as Edgeworth type expansion.

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

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

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

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

The Hermite matrix polynomials up to the third order are given by equalities in [6].

Let $F_{\mathbf{X}}(\mathbf{x})$ be two-dimensional normal distribution function. Then applying operators (3)-(5) we get for unknown distribution function $F_{\mathbf{Y}}(\mathbf{x})$ the following equation:

$$\begin{aligned}
F_{\mathbf{Y}}(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) + \{a_2 + 2b_{12} + (\mathbf{C}_{12}, H_1(\mathbf{x}, \mathbf{\Sigma}))\}f_{\mathbf{X}}(\mathbf{x}) \\
+\{(a_1 - a_2)f_2(x_2)\}\Phi(g(x_2)) \\
-b_{11}\{h_1(x_1) - g'(x_1))f_1(x_1)\}\Phi(g(x_1)) \\
-b_{22}\{h_1(x_2) - g'(x_2))f_2(x_2)\}\Phi(g(x_2)) \\
-c_{(1,1)(1,1)}\{h_2(x_1)f_1(x_1)\Phi(g(x_1)) \\
-2h_1(x_1)f_1(x_1)f_1(g(x_1))g'(x_1) \\
-f_1(x_1))h_1(g(x_1))f_1(g(x_1))g'(x_1)^2\} \\
-c_{(2,2)(1,2)}\{h_2(x_2)f_2(x_2)\Phi(g(x_2)) \\
-2h_1(x_2)f_2(x_2)f_1(g(x_2))g'(x_2) \\
-f_2(x_2)h_1(g(x_2))f_2(g(x_2))g'(x_2)^2\} + \ldots \quad (6)
\end{aligned}$$

where

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

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

and

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

where ρ is linear correlation coefficient between random variables X_1 and X_2 and $\Phi(x)$ denotes the standard normal distribution functuion.

Now we generalize Edgeworth expansion presented in [11] to three-dimensional case. Approximating the unknown distribution function with normal distribution $N_3(0, \mathbf{\Sigma})$ we get

$$\frac{\partial F_{\mathbf{X}}(x_1, x_2, x_3)}{\partial x_1} = \frac{\partial}{\partial x_i} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} f_{\mathbf{X}}(u_1, u_2, u_3) du_1 du_2 du_3 =$$

$$= \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} f_{\mathbf{X}}(x_1, u_2, u_3) du_2 du_3 =$$

$$= \int_{-\infty}^{x_2} \int_{-\infty}^{x_3} f_{\mathbf{X}}(u_2, u_3|x_1) f_1(x_1) du_2 du_3 =$$

$$= f_1(x_1) F_{\mathbf{X}}(x_2, x_3|x_1).$$

In the same way we get that

$$\frac{\partial F_{\mathbf{X}}(\mathbf{x})}{\partial x_2} = f_2(x_2) F_{\mathbf{X}}(x_1, x_3|x_2)$$

and

$$\frac{\partial F_{\mathbf{X}}(\mathbf{x})}{\partial x_3} = f_3(x_3) F_{\mathbf{X}}(x_1, x_2|x_3).$$

For the term of second order derivative we get

$$\frac{\partial^2 F_{\mathbf{X}}(x_1, x_2, x_3)}{\partial x_1 \partial x_2} = f_{12}(x_1, x_2) F_{\mathbf{X}}(x_3|x_1, x_2),$$

$$\frac{\partial^2 F_{\mathbf{X}}(x_1, x_2, x_3)}{\partial x_1 \partial x_3} = f_{13}(x_1, x_3) F_{\mathbf{X}}(x_2|x_1, x_3)$$

and

$$\frac{\partial^2 F_{\mathbf{X}}(x_1, x_2, x_3)}{\partial x_2 \partial x_3} = f_{23}(x_2, x_3) F_{\mathbf{X}}(x_1|x_2, x_3).$$

For the term of third order derivative we get

$$\frac{\partial^3 F_{\mathbf{X}}(x_1, x_2, x_3)}{\partial x_1 \partial x_2 \partial x_3} = f_{\mathbf{X}}(\mathbf{x}).$$

VI. Conclusion

The Edgeworth expansion can use when the components of unknown random vector are strongly dependent but this dependance is not linear. In [11] the Edgeworth expansion is applied to on forestry data where the joint distribution function of trees height and trees diameter at brest height is approximated. For further developments is planned to apply equation (6) on three-dimensional case.

Acknowledgment

The author is thankful for financial support from the Estonian Science Foundation Grant 7656.

References

[1] Cornish, E. A. Fisher, R. A, Moments and cumulants in the specification of distribution. *Rev. de l'Inst int. de stat.*, **5**, 307-322, 1937.

[2] Dwyer, P. J., MacPhail, M. S, Symbolic matrix derivatives. *The Annals of Mathematical Statistics*, **19**, 517-534, 1948.

[3] Fisher, R. A., Cornish, E. A, The percentile points of distributions having known cumulants. *Technometrics*, **2**, No. 2, 209-225, 1960.

[4] Kollo, T., von Rosen, D, Approximating by the Wishart distribution. *Annals of the Institute of Statistical Mathematics*, **47**, 767-783, 1995.

[5] Kollo, T., von Rosen, D, A unified approach to the approximation of multivariate densities. *Scandinavian Journal of Statistics*, **25**, 93-109, 1998.

[6] Kollo, T., von Rosen, D, *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht, 2005.

[7] MacRae, E. C, Matrix derivatives with an applications to an adaptive linear decision problem. *The Annals of Statistics*, **7**, 381-394, 1974.

[8] Magnus, J. R., Neudecker, H, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics, 1999.

[9] Neudecker, H, Some theorems on matrix differentiations with special reference to Kronecker matrix products. *Journal of the American Statistical Association*, **64**, 953-963, 1969.

[10] Pihlak, M, Matrix integral *Linear Algebra and its Applications* **388**, 315-325, 2004.

[11] Pihlak, M, Approximation of Multivariate Distribution Functions. *Mathematica Slovaca*, **58**(5), 635 - 652, 2008.