Approximation of multivariate distribution functions

MARGUS PIHLAK

June 29. 2007

Tartu University

Institute of Mathematical Statistics

Formulation of the problem

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 $y_1, y_2, ..., y_n$. Our aim is to present the unknown distribution function G by means of a known distribution function F. Let F be the distribution function of random variable X. It is assumed that function F is k times continuously differentiable. Then we can approximate the function G by means of the function F as

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

where $a_l, l = 1, 2, \ldots k,$ are some coefficients. The problem is how to determine the coefficients $a_l, l = 1, 2, \ldots, k$ in equality (1).

Solution in univariate case

1937-Cornish and Fisher. 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 function of X as a Taylor series. Then the inverse Fourier transform is used to get from the Taylor expansion an approximation of the probability density function of Y through the density function of X. Integrating this approximation we get an approximation of the unknown distribution function in the form (1) where the coefficients $a_l, l = 1, 2, \ldots, k$ are functions of the first k cumulants of random variables X and Y.

First way for approximation of multivariate distribution functions

Preparation

1) Vectorization

The vectorization operation is denoted by vec. For $\mathbf{X} : p \times q$ matrix vec$\mathbf{X} : pq \times 1$ is the following pq-vector:

$$\text{vec}\mathbf{X} = (x_{11}, \ldots, x_{p1}, x_{12}, \ldots, x_{p2},$$

$$\ldots x_{1q}, \ldots, x_{pq})'.$$

2) Kronecker product

This operation is denoted by $\otimes$. Let us have the 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:

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

where

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

3) Matrix derivative

Neudecker (1969) *The derivative of the matrix* $\mathbf{Y} : r \times s$ *by the matrix* $\mathbf{X} : p \times q$ *is the matrix* $\frac{d\mathbf{Y}}{d\mathbf{X}} : rs \times pq$ *expressed as*

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

where

$$\frac{d}{d\text{vec}'\mathbf{X}} = (\frac{\partial}{\partial x_{11}}, \ldots, \frac{\partial}{\partial x_{p1}}, \ldots,$$

$$\frac{\partial}{\partial x_{1q}}, \ldots, \frac{\partial}{\partial x_{pq}}).$$

MacRae (1974) *Derivative of the matrix* $\mathbf{Y}$: $r \times s$ *by the matrix* $\mathbf{X}$: $p \times q$ *is a* $pr \times qs$*-matrix* $\dfrac{\partial \mathbf{Y}}{\partial \mathbf{X}}$: *defined by equality*

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

where

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

Relation between two multivariate density functions.

Kollo and von Rosen (1995, 1998)

By means of matrix derivative we can define the characteristic function and the k-th order cumulants of random vector. Let $\mathbf{X}$ be a random p-vector and $\mathbf{t} \in \Re^p$. Then the k-th order

cumulant of $\mathbf{X}$

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

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

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. Assume that $f_{\mathbf{X}}(\mathbf{x})$ is uniformly continuous and continuously differentiable necessary number of times by argument $\mathbf{x}$. Let us denote the k-th order derivative of the function $f_{\mathbf{X}}(\mathbf{x})$ by $f^{(k)}_{\mathbf{X}}(\mathbf{x})$. Then in notations explained above the next equality holds (Kollo and von Rosen, 1995):

$$f_{\mathbf{Y}}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) - (E(\mathbf{Y}) - E(\mathbf{X}))'\text{vec}f_{\mathbf{X}}^{(1)}(\mathbf{x})$$

$$+\frac{1}{2}\text{vec}'\{D(\mathbf{Y}) - D(\mathbf{X}) + (E(\mathbf{Y}) - E(\mathbf{X}))$$

$$(E(\mathbf{Y}) - E(\mathbf{X}))'\}\text{vec}f_{\mathbf{X}}^{(2)}(\mathbf{x})$$

$$-\frac{1}{6}\text{vec}'\{(c_3(\mathbf{Y}) - c_3(\mathbf{X})) + 3\text{vec}'(D(\mathbf{Y}) - D(\mathbf{X}))$$

$$\otimes(E(\mathbf{Y}) - E(\mathbf{X}))$$

$$+(E(\mathbf{Y}) - E(\mathbf{X}))'^{\otimes 3}\}\text{vec}f_{\mathbf{X}}^{(3)} + ... \qquad (2)$$

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

Matrix integral

Matrix integral is defined as inverse operation of matrix derivative.

Definition 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 matrix $\mathbf{Z}$ is denoted as

$$\int_{\Re^{pq}} \mathbf{Z} \circ d\mathbf{X} = \mathbf{Y}. \tag{3}$$

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 is used also to define the definite matrix integral.

Definition 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.

For decreasing the dimensions of matrices MacRae (1974) has introduced the star product of matrices. She has denoted this operation as *.

Definition 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.*

The next statement gives us the relation between star product and matrix integral.

Theorem 1 *Let* $\mathbf{Z} = \dfrac{\partial \mathbf{Y}}{\partial \mathbf{X}}$. *Then*

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

Example

Let us have functions $g(\mathbf{x}) = x_1^2 + x_2^2$ and $G(\mathbf{x}) = \frac{1}{3}(x_1^3 x_2 + x_2^3 x_1)$. In the next example we take $\mathbf{C} = \mathbf{0}_2$.

We get

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

and

$$g^{(1)}(\mathbf{x}) = \begin{pmatrix} 2x_1 & 2x_2 \end{pmatrix}.$$

Applying the star product we get

$$\int_{\Re^p} g^{(2)}(\mathbf{x}) \circ d\mathbf{x} = \int_{\Re^p} dx * g^{(2)}(\mathbf{x})$$

$$= \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = (g^{(1)}(\mathbf{x}))'.$$

Pihlak, M. (2004) Matrix integral. *Linear Algebra and Its Applications* **388**, 315-325

Relation between two multivariate distribution functions

We have to integrate the expression (2). Let us write the equality (2) in the form

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

$$-\text{vec}'\mathbf{C}\text{vec} f_{\mathbf{X}}^{(3)}(\mathbf{x}) + \ldots \qquad (4)$$

where

$$\mathbf{a} = (E(\mathbf{X}) - E(\mathbf{Y})), \qquad (5)$$

$$\mathbf{B} = \frac{1}{2}[D(\mathbf{Y}) - D(\mathbf{X}) + (E(\mathbf{Y}) - E(\mathbf{X}))(E(\mathbf{Y}) - E(\mathbf{X}))'] \tag{6}$$

and

$$\mathbf{C} = \frac{1}{6}[(c_3(\mathbf{Y}) - c_3(\mathbf{X})) + 3(D(\mathbf{Y}) - D(\mathbf{X})) \otimes (E(\mathbf{Y}) - E(\mathbf{X})) + (E(\mathbf{Y}) - E(\mathbf{X}))^{\otimes 2}(E(\mathbf{Y}) - E(\mathbf{X}))'] . \tag{7}$$

Applying properties of matrix integral we get from equality (4) the expansion of the distribution function $F_{\mathbf{Y}}(\mathbf{x})$ formulated in

Approximation of multivariate distribution functions. (English) Dissertationes Mathematicae Universitatis Tartuensis 50. Tartu: Tartu University Press; Tartu: Univ. Tartu, Faculty of Mathematics and Computer Science

In bivariate case we get between known distribution function $F(\mathbf{x})$ and unknown distribution function $F_{\mathbf{Y}}(\mathbf{x})$ the following expression:

$$F_{\mathbf{Y}}(\mathbf{x}) = F(\mathbf{x}) - a_1 f(\mathbf{x}) + (a_1 - a_2) f_2(x_2) F(x_1|x_2)$$

$$+ 2b_{12} f(\mathbf{x}) + b_{11} \frac{\partial f_1(x_1) F(x_2|x_1)}{\partial x_1}$$

$$+ b_{22} \frac{\partial f_2(x_2) F(x_1|x_2)}{\partial x_2}$$

$$- (c_{((1,1)(1,2)} + c_{(2,1)(1,1)} + c_{(1,2)(1,1)}) \frac{\partial f(\mathbf{x})}{\partial x_1}$$

$$- (c_{(2,2)(1,1)} + c_{(2,1)(1,2)} + c_{(1,2)(1,2)}) \frac{\partial f(\mathbf{x})}{\partial x_2}$$

$$- (c_{(1,1)(1,1)} \frac{\partial^2 f_1(x_1) F(x_2|x_1)}{\partial^2 x_1}$$

$$+ c_{(2,2)(1,2)} \frac{\partial^2 f_2(x_2) F(x_1|x_2)}{\partial^2 x_2}) + \ldots.$$

Edgeworth type expansion for distribution functions

Let us introduce Hermite matrix-polynomials for a p-vector $\mathbf{x}$.

Definition 4 *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*

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

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

As follows from Definition 4, 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

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

For approximation of the unknown distribution function with distribution function of normal distribution the next statement is valid.

Theorem 2 *Let* $\mathbf{X} \sim N_2(0, \mathbf{\Sigma})$ *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* $\mathbf{Y}$. *Then*

$$F_{\mathbf{Y}}(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) + \{a_2 + 2b_{12} + (\mathbf{C}_{12}, H_1(\mathbf{x}, \mathbf{\Sigma}))\} f_{\mathbf{X}}(\mathbf{x})$$

$$+ \{(a_1 - a_2) f_2(x_2)\} \Phi(g(x_2))$$

$$- b_{11} \{h_1(x_1) - g'(x_1)) f_1(x_1)\} \Phi(g(x_1))$$

$$- b_{22} \{h_1(x_2) - g'(x_2)) f_2(x_2)\} \Phi(g(x_2))$$

$$- c_{(1,1)(1,1)} \{h_2(x_1) f_1(x_1) \Phi(g(x_1))$$

$$- 2h_1(x_1) f_1(x_1) f_1(g(x_1)) g'(x_1)$$

$$- f_1(x_1)) h_1(g(x_1)) f_1(g(x_1)) g'(x_1)^2\}$$

$$- c_{(2,2)(1,2)} \{h_2(x_2) f_2(x_2) \Phi(g(x_2))$$

$$- 2h_1(x_2) f_2(x_2) f_1(g(x_2)) g'(x_2)$$

$$- f_2(x_2) h_1(g(x_2)) f_2(g(x_2)) g'(x_2)^2\} + \ldots \quad (8)$$

where

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

$\mathbf{a}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are defined by equalities (5), (6) and (7) and*

$$g(x_1) = \frac{\frac{x_2}{\sqrt{\sigma_{22}}} - \frac{x_1}{\sqrt{\sigma_{11}}}\rho}{\sqrt{1-\rho^2}} \tag{9}$$

and

$$g(x_2) = \frac{\frac{x_1}{\sqrt{\sigma_{11}}} - \frac{x_2}{\sqrt{\sigma_{22}}}\rho}{\sqrt{1-\rho^2}}. \tag{10}$$

.

Relation between functions $g(x_1)$, $g(x_2)$ and function Φ (distribution function of standard normal distribution):

$$F(x_2|x_1) = \Phi(g(x_1))$$

and

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

Second way way for approximation of multivariate distribution functions. Copulas

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.

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

Definition 5 *The function* $C : \mathbf{I}^2 \to \mathbf{I}$ *is called copula if*

1) for every $u, v \in \mathbf{I}$,

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

and

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

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

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

The key result of the copula theory is formulated in the following theorem.

Theorem 3 *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* $x, y \in \mathbf{R}$

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

Theorem 3 is called the Sklar's theorem (Nelsen, 1999).

A. Sklar is initiator of the copula theory. He laid down foundations of the theory.

Archimedean copula

One very important wide class of copulas is known as Archimedean copulas. This class consists of families of one-parameter distributions. Let $\theta \in \Re$ and $\varphi_\theta = \varphi_\theta(z)$. Assume that $\varphi_\theta(z)$ is a convex, decreasing function from $(0, 1]$ to $[0, \infty)$ such that $\varphi_\theta(1) = 0$.

Definition 6 *The function*

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

is said to be an Archimedean copula.

Between Kendall's correlation coefficient τ and generating function φ_θ the next relation is valid (Nelson, 1999):

$$\tau = 1 + 4 \int_0^1 \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} dt.$$

Archimedean copulas can be divide by type of $\varphi_\theta(z)$ into families.

Clayton family

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

$$\theta = \frac{2\tau}{1 - \tau}$$

Gumbel family

$$\varphi_\theta(t) = (-\ln t)^\theta$$

$$\theta = \frac{1}{1 - \tau}$$

Two-dimensional Gaussian copula

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

Let $(X, Y)'$ denote a vector of continuous variables with arbitrary 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 U be the normal quantile function. Then we define variables V and W having the standard normal distribution as

$$V = U(F(X))$$

and

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

Let γ be the Pearson linear correlation coefficient between V and W. Then the density function h of (X, Y) can be expressed as

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

where

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

is the density weighting function (Krzysztofowicz and Kelly, 1994). The function h is called the density function of the Gaussian copula.

Application of both methods

Let us approximate the joint distribution of tree height H and diameter at the breast height (DBH) by different models.

The functional dependence between DBH and height has been intensively studied in the literature. The most widely used 2-parameter family of functions is elaborated by M. Näslund (Näslund, 1941). Following Näslund, the dependence between variables H and DBH can be expressed by the following equality

$$\sqrt{\frac{1}{H-1.3}} = \frac{b_0}{\text{DBH}} + b_1,$$

where b_0 and b_1 are the parameters.

Let us denote DBH by X and H by Y.

The data is formed by tree measurement of Estonian National Forest Inventory from the period of 1999-2003. Widespread Estonian tree species: pine (*Pinus sylvestris*) and spruce (*Picea abies*) were analysed in the course of data processing.

Approximating using Edgeworth expansion

We approximate the distribution of random vector $\mathbf{X} = (X, Y)'$ with distribution function of $\mathbf{Y} \sim N_2(0, \mathbf{S})$ where $\mathbf{S}$ is the sample covariance matrix of $\mathbf{X}$.

Applying equality (8) we get for distribution function of $\mathbf{X}$ the following approximation:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{Y}}(\mathbf{x}) + \frac{1}{6}[(\mathbf{C1}, H_1(\mathbf{x}, \mathbf{S}))f_{\mathbf{Y}}(\mathbf{x})$$

$$-(c_3(\mathbf{X}))_{11}\{h_2(x_1))f_1(x_1)\Phi(g(x_1))$$

$$-2h_1(x_1)f_1(x_1)f_1(g(x_1))g'(x_1)$$

$$-g(x_1)f_1(g(x_1))f_1(x_1)g'(x_1)^2\}$$

$$-(c_3(\mathbf{X}))_{42}\{h_2(x_2)f_2(x_2)\Phi(g(x_2))$$

$$-2h_1(x_2)f_2(x_2)f_2(g(x_2)g'(x_2))$$

$$-g(x_2)f_2(g(x_2))f_2(x_2)g'(x_2)^2\}] + \ldots \quad (11)$$

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.

The goodness of approximation is estimated by the measure

$$d = \frac{\sum_{i=1}^{k} (F_k(\mathbf{x}_i) - F_{\mathbf{X}}(\mathbf{x}_i))^2}{k}$$

where k denotes the sample size.

Table 1. Goodness of distribution function approximation on different tree species and density classes.

Species	Density	Distribution	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

Approximating using copulas

Goodness of fit test for Archimedean copula

Let us have a random sample of observations, $(X_1, Y_1), ..., (X_n, Y_n)$. Define the random variables Z_i,

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

From a sample $(x_i, y_i), i = 1, ..., n$ we get $z_1, ..., z_n$ as values of the i.i.d. random variables $Z_1, ..., Z_n$. Then we construct

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

Finally we define the theoretical distribution function of the Archimedean copula $C(F(X), G(Y))$:

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

(Frees and Valdez, 1998). The concordance of the copula $C(u, v)$ with empirical distribution function of (X, Y) can be estimated by means of Kolmogorov-Smirnov statistic

$$D = \max_z |K_n(z) - K(z)|,$$

where n is the sample size.

Results of approximation with copula

The results of approximation with Gaussian copulas are presented in Figures 1-2. The parameter γ is the Pearson's correlation coefficient. In these figures the asymmetry of the joint distribution is caused by the log-normality of DBH.

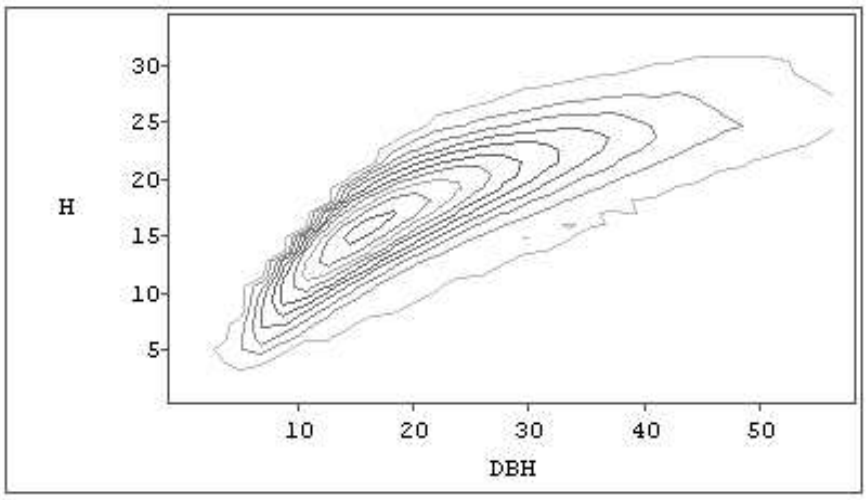

Figure 1 (Pihlak and Adermann, 2005) Joint distribution of H(in m) and DBH(in cm) by Gaussian copula for spruce in high density class, $\gamma = 0.889$

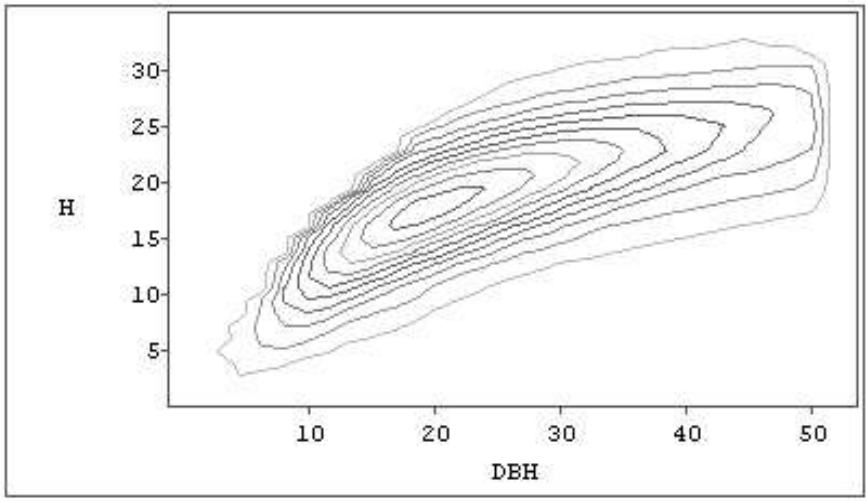

Figure 2 (Pihlak and Adermann, 2005) Joint distribution of H(in m) and DBH(in cm) for pine by Gaussian copula in high density class, $\gamma = 0.838$

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

Table 2. (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

Multivariate Kolmogorov-Smirnov test

Multivariate Kolmogorov-Smirnov test is the generalization of univariate Kolmogorov-Smirnov test (Justel, Peña and Zamar, 1997).

Bivariate Kolmogorov-Smirnov test

Let us 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 $\mathbf{Z}^2 = (Z_1^2, Z_2^2)$. 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 ith realizations of random variables Z_1^1, Z_2^1, Z_1^2 and Z_2^2, respectively, $i = 1, \ldots, n$.

The function G_{emp} is the empirical distribution function of $\mathbf{Z}^1$ or $\mathbf{Z}^2$. We get the value of the bivariate statistic for goodness-of-fit test

$$D_n = \max\{d_n^1, d_n^2\}. \qquad (12)$$

Applying bivariate Kolmogorov-Smirnov test

The results of goodness-of-fit test are presented in Table 3.

Table 3. The results of the Kolmogorov-Smirnov test for Edgeworth type expansion 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

Advantages and disadvantages

	Adv.	**Disadv.**
Edgeworth	General theory	Complexity
Gaussian	Simplicity	uncertainty

Some ideas for further development

• To generalize Theorem (8) to higher-dimensional cases, firstly to the three-dimensional case.

• The theory of the choice of generating function $\phi_\theta(t)$ for Archimedean copulas needs development and generalization to the multivariate case.

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

Thank You for Attention!