Metodološki zvezki, Vol. 3, No. 1, 2006, 109-120

Imputation Algorithm Using Copulas

Ene Käärik [1]

Abstract

In this paper the author demonstrates how the copulas approach can be used to find algorithms for imputing dropouts in repeated measurements studies. One problem with repeated measurements is the knowledge that the data is described by joint distribution. Copulas are used to create the joint distribution with given marginal distributions. Knowing the joint distribution we can find the conditional distribution of the measurement at a specific time point, conditioned by past measurements, and this will be essential for imputing missing values. Using Gaussian copulas, two simple methods for imputation are presented. Compound symmetry and the case of autoregressive dependencies are discussed. Effectiveness of the proposed approach is tested via series of simulations and results showing that the imputation algorithms based on copulas are appropriate for modelling dropouts.

1 Introduction

In repeated measurements study each unit is measured several times. In practice it is common that some of them terminate prematurely before the end of the study.

Therefore, it may be necessary to substitute (impute) *dropouts* in the data. Those dropouts themselves are sometimes of scientific interest, especially in case of small data sets. Imputation is commonly applied to compensate for nonresponse in sample surveys as well.

Rubin (1976), and Little and Rubin (1987) introduced a hierarchical classification of missing data mechanisms (see, for example Diggle and Kenward, 1994; Little, 1995). A dropout process is said to be (1) *completely random* (CRD), when dropout and measurement processes are independent, (2) *random* (RD), when dropout process depends on observed measurements, and (3) *informative* (ID), when dropout process additionally depends on unobserved measurements.

Modelling dropouts is difficult procedure, because typically there is little information in the data to identify dropouts, while modelling depends on the dropout process.

[1] Institute of Mathematical Statistics, University of Tartu, J. Liivi 2, Tartu, Estonia; Ene.Kaarik@ut.ee

We use the idea of imputing dropouts via conditional distributions. Therefore, we need to know the joint distribution of repeated measurements. Copulas provide a convenient way to present the joint distribution.

The term "copula" comes from Latin, and refers to connecting or joining together. A copula is a function that joins a multivariate probability distribution to a collection of univariate marginals, i.e. a copula is a multivariate probability distribution defined on the n-dimensional unit cube $[0,1]^n$ whose marginal distributions are uniform on the interval $[0,1]$.

There are many different copulas. One of the simplest is the independence (or *product*) copula. If the random variables are *independent*, then the copula function that links their marginals is the product copula. Hence, in the case of CRD, it is possible to use the product copula to model the joint distribution of measurement and dropout processes.

2 Notation and basic definitions

Let us introduce some notation.

Suppose that n units are sampled repeatedly over time. The aim is to measure each unit m times, but due to dropouts some of them are measured at $s \leq m$ time points. In general we have a sample of size n of measurements X_j (the time point is usually denoted by j), and the data form a matrix $\mathbf{X} = \{x_{ij}\}, i = 1,\ldots,n; j = 1,\ldots,m$, where due to dropouts some values are missing. For simplicity, we use notations without subscript for the subject's indicator i, usually the lowercase letter is used for subject who drops out, and subscript denotes the time point.

In consequence we are going to observe continuous or discrete outcome variables $X_1,\ldots,X_m$. Let k be the time point at which the dropping out process starts. We shall consider a sequence of measurements up to k observations $X_1, X_2, \ldots, X_k$ and we can assume that until the time point $k-1$ we have complete data $X_1, X_2, \ldots, X_{k-1}$. The vector $H = (X_1, X_2, \ldots, X_{k-1})$ is called the *history* of the measurements.

Suppose X_j has a marginal distribution F_j $(j = 1,\ldots,k)$. In our approach we assume that the marginal distribution F_k belongs to the same family as the distributions of the previous time points. In general, the distribution of the k-variate random vector $X = (X_1, X_2, \ldots, X_k)$ is unknown. Often it is possible to determine a family of marginal distributions, but there might not exist any known family of multivariate distributions suitable to describe the joint distribution of the vector X. If the joint distribution of the vector X is known, then the conditional distribution of X_k, conditioned by the history H, can be used to find the estimate of the missing value.

We will generate the joint distribution using copulas. Theoretical basis of the multivariate modelling using copulas is provided by Sklar (1959), showing that the k-dimensional joint distribution function could be decomposed into its k marginal distributions as a copula, which completely describes the dependence between the k variables.

Using known marginal distributions $F_1(x_1),\ldots,F_k(x_k)$ and a copula C, the function

$$C(F_1(x_1),\ldots,F_k(x_k)) = F(x_1,\ldots,x_k)$$

defines a joint distribution function (see Nelsen, 1999).

If the marginal distributions are continuous, then the copula C is unique for every fixed F and equals to

$$C(u_1, \dots, u_k) = F(F_1^{-1}(u_1), \dots, F_k^{-1}(u_k)),$$

where $F_1^{-1}, \dots, F_k^{-1}$ are the quantile functions with given marginals, and $u_1, \dots, u_k$ are uniform $[0, 1]$ variables.

If C and $F_1, \dots, F_k$ are differentiable, then the joint density $f(x_1, \dots, x_k)$ corresponding to the joint distribution $F(x_1, \dots, x_k)$ can be written as a product of the marginal densities and the copula density, and can be expressed as

$$f(x_1, \dots, x_k) = f_1(x_1) \cdot \ldots \cdot f_k(x_k) \cdot c(F_1, \dots, F_k; R^*), \tag{2.1}$$

where R^* is the matrix of dependence measures, $f_j(x_j)$ is the density corresponding to F_j, $(j = 1, \dots, k)$, and the copula density c is defined as derivative of the copula.

In this paper we consider the Gaussian copula, which represents the dependencies between univariate marginals, allowing any positive-definite correlation matrix R. As the number of different dependence parameters in the case of the k-variate distribution is $k(k-1)/2$, some simple model structure describing the dependence matrix could be used.

According to (2.1), we get the normal joint density

$$\phi_N(x_1, \dots, x_k|R) = \phi_1(x_1) \cdot \ldots \cdot \phi_1(x_k) \cdot c_N[\Phi_1(x_1), \dots, \Phi_1(x_k); R^*], \tag{2.2}$$

where c_N is the multivariate normal copula density, Φ_1 is the univariate standard normal distribution function, and ϕ_1 is its density.

For constructing a multivariate density we use the marginals $F_1(x_1), \dots, F_k(x_k)$ and the copula density c_N as a dependence function. Defining $Y_j = \Phi_1^{-1}[F_j(X_j)]$, $j = 1, \dots, k$, we get the following formula for the copula density (Clemen and Reilly, 1999)

$$c_N[\Phi_1(y_1), \dots, \Phi_1(y_k); R^*] = \frac{\exp\{-Y^T(R^{-1} - I)Y/2\}}{|R|^{1/2}}, \tag{2.3}$$

where $Y = (Y_1, \dots, Y_k)$ and I is the $k \times k$ identity matrix.

Thus, we obtain the joint density as follows

$$\phi_N(x_1, \dots, x_k|R) = \phi_1(x_1) \cdot \ldots \cdot \phi_1(x_k) \cdot \frac{\exp\{-Q_k^T(R^{-1} - I)Q_k/2\}}{|R|^{1/2}}, \tag{2.4}$$

where $Q_k = (\Phi_1^{-1}[F_1(x_1)], \dots, \Phi_1^{-1}[F_k(x_k)])$.

3 Imputation

To present the formula for imputation we should find the conditional density for the variable X_k (which includes dropout) using the history H (complete data).

Taking into consideration the history, the correlation matrix of k measurements can be partitioned as

$$R = \begin{pmatrix} R_{k-1} & r \\ r^T & 1 \end{pmatrix},$$

where R_{k-1} is the correlation matrix of the history $H = (X_1, \ldots, X_{k-1})$, and $r = (r_{1k}, \ldots, r_{(k-1)k})^T$ is the vector of correlations between the history and the time point k (usually we mean Spearman's correlations, otherwise we can use the $\arcsin$-transformation to transform the Pearson correlations to the Spearman ones).

In particular, we focus on the following correlation structures:

(1) *compound symmetry*, where the correlations between all time points are equal, $r_{ij} = \rho, \ \ i, j = 1, \ldots, k$; and

(2) *first order autoregressive*, where the dependence between observations decreases as the measurements get further in time, $r_{ij} = \rho^{|j-i|}, \ \ i, j = 1, \ldots, k$.

The correlation structure and ρ can be estimated by R_{k-1}.

However, there are many other possible correlation structures. We started our analysis with these two structures because they are the most commonly used, the first corresponds to the situation where the observations do not change over time, and the second one corresponds to the situation, where the observations change over time according to the wide-spread autoregressive model.

Hence, from to the definition of the joint density and partition of the correlation matrix, we get the conditional density as follows (see Clemen and Reilly, 1999; Song, 2000)

$$f(y_k|H;R^*) = \phi_1(y_k) \cdot \exp\{-\frac{1}{2}[\frac{(y_k - r^T R_{k-1}^{-1}(y_1, \ldots, y_{k-1})^T)^2}{(1 - r^T R_{k-1}^{-1} r)} - y_k^2]\}$$
$$\cdot (1 - r^T R_{k-1}^{-1} r)^{-1/2}. \tag{3.1}$$

To impute the dropouts we should find the maximum likelihood estimate. It is easy to see (Käärik, 2005) that when maximizing (3.1) with respect to y_k, we get a general form of imputation algorithm

$$\hat{y}_k = r^T \cdot R_{k-1}^{-1} \cdot Y_{k-1}^*, \tag{3.2}$$

where $Y_{k-1}^* = (y_1, \ldots, y_{k-1})^T$ is the vector of observations for the subject who drops out at the time point k.

Formula (3.2) is the starting point for consequent algorithms, where we consider the particular correlation structures.

3.1 The case of compound symmetry

Assume that R has the constant correlation structure or so-called compound symmetry structure. Then the vector of correlations between the history and time point k is $r = (\rho, \ldots, \rho)^T$ and the $(k-1) \times (k-1)$ correlation matrix for history has the following structure

$$R_{k-1} = \begin{pmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \vdots & & \ddots & \\ \rho & \rho & \ldots & 1 \end{pmatrix}.$$

The inverse of this correlation matrix R_{k-1}^{-1} has a well-known simple form with equal elements in the main diagonal and equal elements in the off-diagonal (see Rao, 1965).

Taking into account the whole history consisting of the $(k-1)$ measurements and (3.2) in the case of compound symmetry, the following *formula for imputation* is valid (Käärik, 2005):

$$\hat{y}_k^{CS} = \frac{\rho}{1+(k-2)\rho} \sum_{j=1}^{k-1} y_j, \tag{3.3}$$

where $y_1, \ldots, y_{k-1}$ are the observed values for the subject who dropped out at time point k.

Hence, in the case of compound symmetry we use a weighted sum of past values for an imputed value.

3.2 The case of autoregressive dependencies

Consider now the first order autoregressive correlation structure. Then the vector of correlations between the history and time point k is $r = (\rho^{k-1}, \rho^{k-2} \ldots, \rho)^T$. The $(k-1) \times (k-1)$ correlation matrix for history is the following

$$R_{k-1} = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{k-2} \\ \rho & 1 & \rho & \ldots & \rho^{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{k-2} & \rho^{k-3} & \rho^{k-4} & \ldots & 1 \end{pmatrix}.$$

The inverse of the correlation matrix R_{k-1}^{-1} is a three-diagonal matrix. The main properties of this type of three-diagonal matrices are well-known (see Kendall and Stuart, 1976; Raveh, 1985).

Let A be a matrix of order m with the autoregressive correlation structure, and B be a three-diagonal matrix of order m. Then the inverse matrix $A^{-1} = cB$ is a three-diagonal matrix, where $c = \frac{1}{\rho^2-1}$ and

- $b_{ij} = 0$, if $|i-j| > 1$;
- $b_{11} = b_{mm} = -1$ and $b_{ii} = -(1+\rho^2), i = 2, \ldots, m-1$;
- $b_{ij} = \rho$, if $|i-j| = 1$.

So, the inverse matrix of the correlation matrix R_{k-1} has following structure

$$R_{k-1}^{-1} = \frac{1}{\rho^2-1} \cdot \begin{pmatrix} -1 & \rho & 0 & \ldots & 0 & 0 \\ \rho & -(1+\rho^2) & \rho & \ldots & 0 & 0 \\ 0 & \rho & -(1+\rho^2) & \ldots & 0 & 0 \\ 0 & 0 & \rho & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -(1+\rho^2) & \rho \\ 0 & 0 & 0 & \ldots & \rho & -1 \end{pmatrix}.$$

Using the matrix R_{k-1}^{-1} and (3.2) we get

$$y_k = (\rho^{k-1}, \rho^{k-2}, \ldots, \rho) \cdot R_{k-1}^{-1} \cdot (y_1, \ldots, y_{k-1})^T$$

$$= (\rho^{k-1}, \rho^{k-2}, \ldots, \rho) \cdot \frac{1}{\rho^2 - 1} \cdot \begin{pmatrix} -1 & \rho & \ldots & 0 & 0 \\ \rho & -(1+\rho^2) & \ldots & 0 & 0 \\ 0 & \rho & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & -(1+\rho^2) & \rho \\ 0 & 0 & \ldots & \rho & -1 \end{pmatrix} \cdot (y_1, \ldots, y_{k-1})^T.$$

Simplifying the last equation we obtain

$$\hat{y}_k = \rho \cdot y_{k-1}. \tag{3.4}$$

Taking into account the general trend in data, we get the modified formula for imputation

$$\hat{y}_k^{AR} = \rho \frac{S_k}{S_{k-1}} (y_{k-1} - \bar{Y}_{k-1}) + \bar{Y}_k, \tag{3.5}$$

where y_{k-1} is the last observed value for the subject, $\bar{Y}_{k-1}$ and $\bar{Y}_k$ are the mean values of the time points k and $k-1$, respectively, while S_k and S_{k-1} are the corresponding standard deviations. Basically, we use here a standardizing procedure and some kind of simple regression predictor for imputed value.

4 Simulation study

The goal of the simulation study was to test the effectiveness of the imputation methods (3.3) and (3.5) by comparison them with some well-known imputation methods in the case of different missing data mechanisms and sample sizes.

The idea of the comparisons was to start with normally distributed data and then check the robustness of the imputation methods by moving away from the normal distribution. We performed two simulation studies: (1) using the standard normal distribution and (2) using skewed distribution.

We have used the standardized absolute difference between the observed value and the imputed value as a quality measure. First we generated a complete dataset and then a dataset with dropouts using a fixed missing data mechanism from the complete set.

4.1 Generation of the complete data

In the first simulation study we generated a complete data matrix from a multivariate normal distribution using (1) a constant correlation structure, (2) an autoregressive correlation structure with the correlation coefficients $\rho = 0.5$ and $\rho = 0.7$.

We generated data from 3-, 6- and 12-dimensional normal distributions with sample sizes $n = 10$ and $n = 20$, assuming that the data represent repeated measurements. Due to small sample sizes every value is important, hence we have to impute the missing values.

The second simulation study was performed with skewed distributed marginals. Suppose $X = (X_1, \ldots, X_k)$ has the k-variate normal distribution. To get skewed marginals $Z_1, \ldots, Z_k$ the data were transformed using the following rule

$$z_{ij} = \begin{cases} b_1 v_j & \text{for maximum value } v_j = \max_i x_{ij}, \\ b_2 x_{ij}, & \text{for every other positive value } x_{ij}, \\ x_{ij}, & \text{otherwise.} \end{cases}$$

The constants are equal $b_1 = 10$ and $b_2 = 5$. Thus, we get the transformed data $Z_j = (z_{1j}, \ldots, z_{nj}), (j = 1, \ldots, k)$, which are extended to the positive direction.

4.2 Generation and imputation of the dropouts

The dropouts occur at the last time points $X_k, k = 3, 6, 12$, and we examine three cases of missing mechanism: CRD, RD and ID.

We delete the observation according to the definitions of the CRD, RD and ID denoted before. Two methods of imputation based on formulas (3.3) and (3.5) were used.

In both simulation studies the methods of imputation were compared with two well-known methods:

1. Imputation by formula (3.3) versus imputation by linear prediction, where the observation at the last time point was modelled using previous time points $X_k = \beta_0 + \beta_1 X_1 + \ldots + \beta_{k-1} X_{k-1}$.

2. Imputation by (3.5) versus imputation using the *LOCF*-method (*Last Observation Carried Forward*) [2].

4.3 Experimental design

In both simulation studies we generated $3 \times 2 \times 2 \times 3 = 36$ different data sets: $k = 3, 6, 12$ (data from 3-, 6-, 12-dimensional normal distribution), $n = 10, 20$ (small sample sizes), $\rho = 0.5,\ \rho = 0.7$ and 3 missingness mechanisms (CRD, RD and ID). For each combination formed by the above simulation factors, 500 runs were performed.

4.4 Calculations

To analyze the obtained results, the average absolute bias was calculated as average difference between observed values and imputed values. Results were presented in units of standard deviation of given marginals.

Let w_k be the observed value for the subject who drops out at time point k (i. e. $w_k = x_k$ or $w_k = z_k$ according to the simulation study), w_{kv} be the corresponding imputed value using (3.3) or (3.5) (i.e. $w_{kv} = \hat{y}_k^{CS}$ or $w_{kv} = \hat{y}_k^{AR}$, respectively) and w_{kp} be the

[2] When the main interest is the outcome at endpoint of the study (for example in clinical trials), the *LOCF* is the most frequently used approach for dealing with missing values in continuous variables.

corresponding imputed value using well-known rules (linear prediction or *LOCF*). The standardized biases are calculated as follows

$$SB_1 = \frac{w_k - w_{kv}}{S_k}, \quad SB_2 = \frac{w_k - w_{kp}}{S_k},$$

where S_k is the standard deviation of observed values at last time point k.

Mean biases B_1 as the average bias for (3.3) or (3.5), and B_2 as the average bias for linear prediction or *LOCF* rules, were calculated by averaging absolute values of standardized biases SB_1 and SB_2 over 500 runs.

The average standard deviations of biases were calculated over 500 runs and denoted S_1 and S_2, respectively.

4.5 Results

To estimate the effectiveness of new imputation rules, we compare the mean biases B_1 versus B_2, and standard deviations S_1 versus S_2.

In the case of compound symmetry in both simulation studies, the results show the advantage of (3.3) compared to the linear regression (see Table 1).

Table 1: Results of two simulation studies in the case of compound symmetry.

I	CRD	RD	ID
B_1	0.0247	0.0414	1.5109
B_2	0.0485	0.0961	1.7835
S_1	0.6895	0.7897	1.0236
S_2	1.0945	1.5627	2.0957
II	CRD	RD	ID
B_1	0.0245	0.1173	1.8994
B_2	0.0870	0.3035	2.0685
S_1	0.6918	0.8216	1.0243
S_2	1.4107	2.0112	2.0647

We can see that in all cases the new formula (3.3) gives better results: it has smaller bias and is more stable compared to the imputation using the linear regression ($B_1 < B_2,\ S_1 < S_2$). Of course in the case of ID both methods perform not well; nevertheless, the new one gives smaller bias here too. In the case of informative dropouts, the bias is greater than in the case of random or completely random dropouts, as is usual.

In Table 2 we see the results of the simulation studies in the case of the first order autoregressive correlation structure.

Table 2: Results of two simulation studies in the case of autoregressive dependencies.

I	CRD	RD	ID
B_1	0.0199	0.0629	2.1261
B_2	0.0213	0.1787	1.0929
S_1	0.8296	0.8528	0.9599
S_2	0.8776	0.8959	1.4408
II	CRD	RD	ID
B_1	0.0426	0.0597	2.6449
B_2	0.0870	0.3035	2.0685
S_1	0.6918	0.8216	1.0243
S_2	0.8776	0.8959	1.4408

Again, the new method (3.5) is more stable ($S_1 < S_2$ in all cases). In case of CRD and RD, the new method gives smaller bias compared to the *LOCF*-method ($B_1 < B_2$ in the first two columns).

Formula (3.5) did not work well when we had informative dropout. In this case the bias was larger if compared to the *LOCF*-method, but standard deviations were smaller ($B_1 > B_2, S_1 < S_2$ in the last column).

5 Illustration

The data from PWC170[3] study were carried out as an example for modelling dropouts. Fifteen athletes (Estonian skiing team) performed seven consecutive workloads on a bicycle ergometer and the average heart rate was recorded in every step. So we had repeated measurements at seven time points $X_1, \ldots, X_7$, and we supposed one observation was missing for some reason at the last time point. Heart rate data for the subject who dropped out were 91, 94, 107, 118, 129, 137 beats per minute and the last value was missing (the observed value was 148 bpm).

For imputing the dropout, we had to follow certain rules due to the above derived Gaussian copula based approach.

1. Estimation of marginal distributions.
   We used the Kolmogorov-Smirnov and Anderson-Darling tests for normality, and, as usual, small samples almost passed the normality test, thus we did not reject the normality assumption. The means and standard deviations were calculated for every time point and summarized in Table 3.

Table 3: Means and standard deviations of heart rate in PWC170 test.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Mean	92.9	99.3	109.5	120	131.7	144.4	152.5
Std Deviation	11.41	12.20	12.27	12.46	13.95	13.55	14.07
n	15	15	15	15	15	15	14

[3]Physical Working Capacity (PWC170) test: the workload at a heart rate of 170 beats per minute (bpm) is used to estimate aerobic fitness in athletes.

2. Estimation of the correlation structure of data.
   Of course in solving practical tasks it is difficult to specify the correct correlation structure. Many methods allow the specification of a 'working' correlation matrix that is intended to approximate the true correlation matrix. In our correlation matrix of the data, the correlations decreased monotonically over time, so the natural choice was autoregressive correlation structure. Calculation of the 'working' correlation matrix gave us the Pearson correlation coefficient $r = 0.88$ and the Spearman's $\rho = 0.87$.

3. Estimation of the missing value using formula (3.5)

$$\hat{y}_k^{AR} = \rho \frac{S_k}{S_{k-1}}(y_{k-1} - \bar{Y}_{k-1}) + \bar{Y}_k = 0.87 \frac{14.07}{13.55}(137 - 144.4) + 152.5 = 145.8$$

If we use the *LOCF*-method we get the imputed value 137 bpm, which is much more biased.

As illustrated by this example, our imputation strategy allows us to get presumptive results for practical use.

6 Discussion

In this paper we have introduced the analysis of incomplete repeated measurements using Gaussian copula, and derived two expressions for imputing dropouts.

These algorithms require determination of the correlation structure, which can be estimated from history of measurements.

In general, in both simulation studies the results showed that the imputation algorithms based on the copula approach are quite appropriate for modelling dropouts.

- The bias is smaller in the case of CRD and RD.
- Standard deviations are more stable.
- The formula (3.3) could be used for small data sets with several repeated measurements $(k > n)$, when linear prediction does not work.
- The formula (3.5) contains more information about data than the LOCF-method.

It is clear that in the case of informative dropouts we do not get good results because the dropout process is not random, and without supplementary information we cannot expect good results.

Thus, the new approach has essential advantages and therefore could have widespread implementation in practice.

1. Normality of marginals is not necessary. Furthermore, the marginals may have different distributions. The normalizing transformation will be used.
2. The simplicity of formulas (3.3) and (3.5) for calculation.

3. Effectiveness, especially in the case of of small sample size n relative to the number of measurements (time points) k.

Certainly the Gaussian copula is not the only use of this approach. Multivariate normal distribution and linear correlation form the basis for most models used to model dependence. Even though this distribution has a wide range of dependence structures it is quite seldom suitable for modeling real data. Linear correlation is a natural measure of dependence in the context of normal distribution.

The parametric class of copulas named *Archimedean* copulas (see Nelsen,1998; Genest, 1987; Genest and Rivest, 1993) has attracted particular interest, since its elements have a number of properties which make them simple to use. Members of the Archimedean copula class are constructed by a continuous, strictly decreasing and convex function, which is called the *generator* of the copula. Different choices of the generator yield several important families of copulas. For example, the *Frank's copula* with one parameter, which statistical properties are given in (Genest, 1987). Vandenhende and Lambert (2002) used *Frank's copula* to model the dependence between dropout and responses. We plan to explore these ideas more thoroughly later.

Copulas provide a natural approach to handle dependencies between repeated measurements. They are not difficult to apply, while being reliable in many situations where the correlation structure is known.

Acknowledgment

This work is supported by Estonian Science Foundation grants No 5521 and No 5203.

References

[1] Clemen, R.T. and Reilly, T. (1999): Correlations and copulas for decision and risk analysis. Fuqua School of Business, Duke University. *Management Science*, **45**, 208–224.

[2] Diggle, P.J. and Kenward, M. G. (1994): Informative dropout in longitudinal data analysis. *Applied Statistics*, **43**, 1, 49–93.

[3] Genest, C. (1987): Frank's family of bivariate distributions. *Biometrika*, **74**, 549–555.

[4] Genest, C. and Rivest, L.P. (1993): Statistical inference procedure for bivariate Archimedean copulas. *JASA*, **88**, 423, 1034–1043.

[5] Kendall, M. and Stuart, A. (1976): Design and analysis, and time-series. *The Advanced Theory of Statistics*, **3**, 646, Moscow: Nauka (Russian).

[6] Käärik, E. (2005): Handling dropouts by copulas. *WSEAS Transactions on Biology and Biomedicine,* **1**, 2, 93–97.

[7] Little, R.J.A. (1995): Modeling the dropout mechanism in repeated-measures studies. *JASA*, **90**, 431, 1112–1121.

[8] Little, J. A. and Rubin, D.B. (1987): *Statistical Analysis with Missing Data*. New York: Wiley.

[9] Nelsen R.B. (1999): An Introduction to Copulas. *Lecture Notes in Statistics*, 139, New York: Springer Verlag.

[10] Rao, C. R. (1965): *Linear Statistical Inference and its Applications*. New York: Wiley.

[11] Raveh, A. (1985): On the use of the inverse of the correlation matrix in multivariate data analysis. *The American Statistician*, **39**, 1, 39–42.

[12] Reilly, T. (1999): Modelling correlated data using the multivariate normal copula. *Proceedings of the Workshop on Correlated Data*, Trieste, Italy.

[13] Rubin, D.B. (1976): Inference and Missing Data. *Biometrika*, **63**, 581–592.

[14] 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.

[15] Song, P.X.K. (2000): Multivariate dispersion models generated from Gaussian Copula. *Scandinavian Journal of Statistics*, **27**, 305–320.

[16] Vandenhende, F. and Lambert, P. (2002): On the joint analysis of longitudinal responses and early discontinuation in randomized trials. *Journal of Biopharmaceutical Statistics*, **12** , 425–440.