Serdica J. Computing **8** (2014), No 3, 233–254

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

DEPENDENCE STRUCTURE OF SOME BIVARIATE DISTRIBUTIONS

Boyan Dimitrov

Abstract. Dependence in the world of uncertainty is a complex concept. However, it exists, is asymmetric, has magnitude and direction, and can be measured. We use some measures of dependence between random events to illustrate how to apply it in the study of dependence between non-numeric bivariate variables and numeric random variables. Graphics show what is the inner dependence structure in the Clayton Archimedean copula and the Bivariate Poisson distribution. We know this approach is valid for studying the local dependence structure for any pair of random variables determined by its empirical or theoretical distribution. And it can be used also to simulate dependent events and dependent r/v/'s, but some restrictions apply.

1. Introduction. Let A and B be two arbitrary random events and $P(B) > 0$ (i.e. B is a *possible* event). The conditional probability of A, given B, is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \tag{1}$$

ACM Computing Classification System (1998): G.3, J.2.
Key words: Bivariate Poisson, Clayton Copula, Local dependence, Measures of dependence, Regression coefficient.

and is used to introduce the *independence*: When the following is fulfilled

$$P(A|B) = P(A), \tag{2}$$

then A **is independent of** B. The independence is equivalent to the fulfillment of the equation

$$P(A \cap B) = P(A)P(B). \tag{3}$$

Most textbooks on probability (e.g. [2]) stop the use of conditional probability at the definition of independence and few rules, like Total Probability rule and Bayes rule. Independence is symmetric. It is **mutual.** Moreover, any pair A and $\overline{B}$ (the complement of B), $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$ are mutually independent too. The inconvenience of equation (2) in the definition of independence is that it requires $P(B) > 0$. When $P(B) = 0$, the conditional probability is not defined, since (1) will involve an improper division by zero. However, if $P(B) = 0$ (then B is called *impossible, or zero event*), equation (3) is fulfilled, whatever the event A is. Hence, if we define the conditional probability $P(A|B) = P(A)$ in cases $P(B) = 0$, both equations (2) and (3) equivalently express the independence no matter that one of the events is impossible. In addition when $P(B) = 1$ (sure event), the independence of B with any other event A is also a fact, confirmed by either (2) or (3).

Therefore, *a zero event and a sure event are independent with any other event.* ***The most important fact is that when equality in*** (2) ***or*** (3) ***does not hold, the events A and B are dependent.*** In this article we explore the situations when dependence between random events is a fact. We show that the strength of this dependence can be measured, has magnitude and direction, and is asymmetric. Further we explain how dependence between events can be practically used in the study of local dependence between random variables. And we illustrate this idea on some known bivariate distributions, like the empirical distribution between non-numeric variables, or like the components of the bivariate random vector with Clayton copula joint distribution. The graphical illustrations confirm the usefulness and the simplicity of this detailed analysis of inner dependence structures for pairs of dependent r.v. s.

2. Dependent events. Four measures of dependence.

2.1. Connection between random events. Dependence in the world of uncertainty is a complex concept. We use the concept of dependence proposed by Obreshkov [9] and discussed in Dimitrov [3]. References on the later publication will be used to avoid long explanations and omit proofs. We call the probabilities $P(A)$ and $P(B)$ marginal probabilities of the participating events, and $P(A \cap B)$ is called joint probability of these events.

Definition 1. *The number*

$$\delta(A, B) = P(A \cap B) - P(A)P(B) \tag{4}$$

is called connection between events A and B.

The following **properties of the connection** hold:

$\delta 1$) The connection $\delta(A, B)$ equals zero if and only if the events are independent.

$\delta 2$) The connection between events A and B is symmetric.

$$\delta(A, B) = \delta(B, A).$$

$\delta 3$) If $A_1, A_2, \ldots, A_j, \ldots$ are mutually exclusive events, then the following is fulfilled

$$\delta\left(\sum A_j, B\right) = \sum \delta(A_j, B).$$

The function $\delta(A, B)$ is additive with respect to either of its arguments. Therefore, it is also a continuous function as the probabilities in its construction are.

$\delta 4$) The connection $\delta(A, B)$ satisfies $\delta(A \cup C, B) = \delta(A, B) + \delta(C, B) - \delta(A \cap C, B)$, and this is equivalent to the rule for probability of the union for arbitrary two events. Therefore, most of the properties of the probability function for random events can be transferred as properties of the connection function including extensions for union of any finite number of events.

$\delta 5$) The connection between events A and $\overline{B}$ equals (by magnitude) the connection between events A and B, but has an opposite sign, i.e., it is true that $\delta(A, \overline{B}) = -\delta(B, A)$. The connection between the complementary events $\overline{A}$ and $\overline{B}$ is the same as between A and B, i.e., it is fulfilled that $\delta(\overline{A}, \overline{B}) = \delta(B, A)$.

$\delta 6$) If the occurrence of A implies the occurrence of B, i.e., when we have $A \subseteq B$, then it is fulfilled that $\delta(A, B) = P(A)P(\overline{B})$ and the connection between the events A and B is positive.

$\delta 7$) If A and $B$are mutually exclusive, i.e., when$A \cap B = \emptyset$, then $\delta(A, B) = -P(A)P(B)$ and the connection between the events A and B is negative.

$\delta 8$) When $\delta(A, B) > 0$, the occurrence of one of the two events increases the conditional probability for the occurrence of the other event. The following relation is true:

$$P(A|B) = P(A) + \frac{\delta(A, B)}{P(B)}. \tag{5}$$

If $\delta(A,B) < 0$ (then also $P(A)P(B) \neq 0$), the occurrence of one of the events decreases the chances for the other one to occur. *Equation* (5) indicates that the knowledge of the connection is very important, and *can be used for calculation of the posteriori probabilities, as when we apply the Bayes' rule*!

The numeric value of the connection $\delta(A,B)$ for two events is sufficient, together with their prior (marginal) probabilities, in order to exactly evaluate the posterior probability of either of the two events when the other one occurs. We anticipate most applications of the connection as a measure of dependence to be oriented towards such purposes. We call them *predictions*. The connection $\delta(A,B)$ also is sufficient to restore the joint probability $P(A \cap B)$ for two events with marginal probabilities $P(A)$ and $P(B)$.

We call the events A and B *positively associated when* $\delta(A,B) > 0$, and *negatively associated when* $\delta(A,B) < 0$. The reason for this is the relationship (5) which shows the increase or decrease of the conditional probability for the occurrence of one of the events when the other one occurs.

Remark 1. Let us introduce the indicator of the random event Aas $I_A = 1$, when A occurs, and $I_A = 0$ when the complement $\overline{A}$ occurs. Then it is true that $E(I_A) = P(A)$ and

$$Cov(I_A, I_B) = \delta(A,B).$$

Therefore, the connection between two random events equals the covariance between their indicators.

Comment. Similarly to the covariance between r.v. s, the numerical value of the connection $\delta(A,B)$ does not show the magnitude of the dependence between A and B. It is intuitively clear that the strongest connection should be between two coinciding events, i.e., the strongest connection must hold when $A = B$. In such cases we have $P(A) = P(B)$, and also $\delta(A,B) = P(A) - P^2(A) = P(A)[1 - P(A)]$. Therefore, $\delta(A,A)$ depends on the value of $P(A)$ and is close to zero when A is close to the sure, or to the impossible event.

Due to this reason other measures are used to establish the strength of the dependence between two random events. We focus on them in the next section.

2.2. Regression coefficients as measure of dependence. The reason for such a name will be given at the end of this section. First we introduce this measure as it is suggested by Obreshkov [9].

Definition 2. *We call a regression coefficient $r_B(A)$ of the event A with respect to the event B the difference between the conditional probability for the event A given the event B, and the conditional probability for the event A given* *the complementary event $\overline{B}$, namely*

$$r_B(A) = P(A|B) - P(A|\overline{B}). \tag{6}$$

This measure of the dependence of the event A on the event B *is* directed dependence.

The regression coefficient $r_A(B)$ of the event B with respect to the event A, namely

$$r_A(B) = P(B|A) - P(B|\overline{A}). \tag{7}$$

The following statements hold:

(**r1**) Equality to zero $r_B(A) = r_A(B) = 0$ holds if and only if the two events are independent.

(**r2**) The regression coefficients $r_B(A)$ and $r_A(B)$ are numbers with equal signs and this is the sign of their connection $\delta(A, B)$. It holds the relationship

$$\delta(A, B) = r_B(A)P(B)[1 - P(B)] = r_A(B)P(A)[1 - P(A)]. \tag{8}$$

The numerical values of $r_B(A)$ and $r_A(B)$ may not always be equal. There **exists an asymmetry in the dependence between random events, and this reflects the nature of real life.**

For $r_B(A) = r_A(B)$ to be valid, it is necessary and sufficient for the following to be fulfilled

$$P(A)[1 - P(A)] = P(B)[1 - P(B)].$$

(**r3**) When $r_B(A) > 0$, the occurrence of the event B increases the conditional probability of the occurrence of the event A. It is true:

$$P(A|B) = P(A) + r_B(A)[1 - P(B)] \tag{9}$$

Knowledge of the regression coefficients is important, and can be used for calculation of the posterior probabilities as when applying Bayes' rule.

(**r4**) The regression coefficients $r_B(A)$ and $r_A(B)$ are numbers between -1 and 1, i.e., they satisfy the inequalities

$$-1 \leq r_B(A) \leq 1; \quad -1 \leq r_A(B) \leq 1.$$

(**r4.1**) The equality $r_B(A) = 1$ holds only when the random event A coincides with (or is equivalent to) the event B. Then the equality $r_A(B) = 1$ is also valid;

(**r4.2**) The equality $r_B(A) = -1$ holds only when the random event A coincides with (or is equivalent to) the event $\overline{B}$—the complement of the event B. Then $r_A(B) = -1$, and respectively$\overline{A} = B$, is also valid.

We interpret the properties (r4) of the regression coefficients in the following way: The closer the numerical value of $r_B(A)$ is to 1, "the denser the events A and B are inside within each other, considered as sets of outcomes of the experiment". In a similar way we interpret also the negative values of the regression coefficient: "The closer the numerical value of $r_B(A)$ is to -1, the denser the events A and $\overline{B}$ are within each other, considered as sets of outcomes of the experiment".

(**r5**) It is fulfilled that $r_{\overline{B}}(A) = -r_B(A)$, and $r_B(\overline{A}) = -r_B(A)$. Also the identities $r_{\overline{A}}(B) = r_A(\overline{B}) = -r_B(A)$ hold.

(**r6**) It is true that for any mutually exclusive sequence of events

$$r_B\left(\sum\nolimits_j A_j\right) = \sum\nolimits_j r_B(A_j).$$

(**r7**) The regression function possesses the property

$$r_B(A \cup C) = r_B(A) + r_B(C) - r_B(A \cap C).$$

Remark 2. The fact that $r_B(A) = r_A(B) = \delta(A, B) = 0$, indicates that $A \cap B \neq \emptyset$.

(**r8**) Freshet-Hoefding inequalities for the Regression Coefficients between two random events:

$$\max\left\{-\frac{P(A)}{1-P(B)}, -\frac{1-P(A)}{P(B)}\right\} \leq r_B(A) \leq \min\left\{\frac{P(A)}{P(B)}, \frac{1-P(A)}{1-P(B)}\right\};$$

$$\max\left\{-\frac{P(B)}{1-P(A)}, -\frac{1-P(B)}{P(A)}\right\} \leq r_A(B) \leq \min\left\{\frac{P(B)}{P(A)}, \frac{1-P(B)}{1-P(A)}\right\}.$$

These properties are anticipated to be used in simulation of dependent random events with desired values of the regression coefficients, and with given marginal probabilities $P(A)$ and $P(B)$. The restrictions must be satisfied when modelling dependent events by use of regression coefficients.

For example if $P(A) = 0.3$, $P(B) = 0.6$, then the regression coefficient $r_B(A)$ must be kept within the interval $\max[-0.3/0.4, -0.7/0.6] = -0.75$, and $\min[0.3/0.6, 0.7/0.4] = 0.5$, i.e., only the values $r_B(A) \in [-0.75, 0.5]$ are legal for the assumed regression coefficient in simulations under the given marginal probabilities. For $r_A(B)$ it must be fulfilled that $r_A(B)$ takes values within the interval $\max[-0.6/0.7, -0.4/0.3] = -6/7$, and $\min[0.6/0.3, 0.4/0.7] = 4/7$, i.e., only the values $r_A(B) \in [-8/7, 4/7]$ are legal for the assumed regression coefficient $r_A(B)$ in simulations in these conditions, given by the marginal probabilities of the two events.

Remark 3. Let $I_A(\omega)$ and $I_B(\omega)$ be the indicator r.v. associated with the random events A and B as in Remark 1. The argument ω symbolizes an arbitrary outcome from the experiment. Formally, construct the following "regression model" which represents a possible linear relationship

$$I_A(\omega) = \alpha + \beta \quad I_B(\omega) + \varepsilon(\omega), \tag{10}$$

where $\varepsilon(\omega)$ is a r.v. which has a zero expectation and minimum variance. It allows one to "predict" the value of the indicator $I_A(\omega)$ if one knows the value of indicator $I_B(\omega)$, and admits an error $\varepsilon(\omega) = I_A(\omega) - [\alpha + \beta I_B(\omega)]$. In this prediction the values of the coefficients α^* and β^* are such numbers that the following is fulfilled

$$E[I_A(\omega) - \alpha - \beta I_B(\omega)] = 0, \qquad \text{and}$$

$$Var[I_B(\omega) - \alpha^* - \beta^* I_A(\omega)] = \min_{\alpha,\beta}\{Var[I_A(\omega) - \alpha - \beta I_B(\omega)]\}.$$

The "optimal" coefficient have the values

$$\alpha^* = P(A) + \delta(A,B)/P(\overline{B}) = P(A|\overline{B}). \tag{11}$$

and

$$\beta^* = P(A|B) - P(A|\overline{B}) = r_B(A). \tag{12}$$

The asymmetry in this form of dependence of one event on the other can be explained by the different capacity of the events. Events with less capacity (smaller amounts of favorable outcomes) will have less influence on events with larger capacity. Therefore, when $r_B(A)$ is less than $r_A(B)$, the event A is weaker in its influence on B. We accept it as reflecting what actually exists in the real life. By catching the asymmetry with the proposed measures we are convinced about their flexibility and utility features.

Remark 4. It is possible to **use** $r_B(A)$ **for ranking** the events by the magnitude of dependence on a certain event. Such gradation will use the distance of the regression coefficient from zero. For instance, if the values $|r_B(A)|$ are within distance 0.05 from zero, the event could be classified as "*almost independent* on the other", for distances between 0.05 to 0.2 from zero, the event may be classified as *weakly dependent* on the other; if the distance is between 0.2 and 0.45, the event could be classified as *moderately dependent*; from 0.45 to 0.8 it could be called as *dependent on the average*, and above 0.8 it could be classified as *strongly dependent*. Users will understand that this classification is very relative.

Actually, if we fix the event A, and consider any finite sequence B_1, B_2, $\ldots$, B_n of random events, then these events can be ordered according to their "magnitude of influence on event A" which corresponds to the absolute value of their regression coefficients $r_{B_k}(A)$ with respect to the event A.The higher the absolute value of $r_{B_k}(A)$, the stronger the influence of B_k on A is.

2.3. Correlation between two random events.

Definition 3. *We call a correlation coefficient between two events A and B the number*

$$R_{A,B} = \pm\sqrt{r_B(A)\cdot r_A(B)}, \tag{13}$$

where the sign, plus or minus, is the sign of either of the two regression coefficients.

An equivalent representation of the correlation coefficient $R_{A,B}$ in terms of the connection $\delta(A,B)$ holds, namely

$$R_{A,B} = \frac{\delta(A,B)}{\sqrt{P(A)P(\overline{A})P(B)P(\overline{B})}} = \frac{P(A\cap B) - P(A)P(B)}{\sqrt{P(A)P(\overline{A})}\sqrt{P(B)P(\overline{B})}}. \tag{14}$$

If it happens that some of the events A or B is a zero or sure event, then $\delta(A,B) = r_B(A) = r_A(B) = R_{A,B} = 0$ despite the formality in (14) remains the undefined quantity $0/0$.

Remark 5. The correlation coefficient $R_{A,B}$ between the events A and B equals the formal correlation coefficient ρ_{I_A,I_B} between the random variables I_A and I_B, the indicators of the two random events $A$and B,as defined in Remark 1. This explains the terminology proposed by Obreshkov [9].

The **correlation coefficient** $R_{A,B}$ between two random events is symmetric, is located between the numbers $r_B(A)$ and $r_A(B)$, and **possesses the following properties:**

R1. $R_{A,B} = 0$ holds if and only if the two events $A$and B are independent. The use of the numerical values of the correlation coefficient is similar to the use of the two regression coefficients. The closer $R_{A,B}$ is located to zero, the "closer" the two events A and B are to independence.

For random variables a similar statement is not true. The equality to zero of their mutual correlation coefficient does not mean independence, but only registers an absence of correlation. The two random variables are called then *non-correlated* and may not be independent.

R2. The correlation coefficient $R_{A,B}$ is always a number between -1 and $+1$, i.e., it is fulfilled that $-1 \le R_{A,B} \le 1$.

R2.1. The equality $R_{A,B} = 1$ holds if and only if the events A and B are equivalent, i.e., when $A = B$.

R2.2. The equality $R_{A,B} = -1$ holds if and only if the events A and $\overline{B}$ are equivalent, i.e., when $A = \overline{B}$ (then of course it holds also that $\overline{A} = B$).

The closer $R_{A,B}$ is to the number 1, the " denser within one another" the events A and B are, and when $R_{A,B} = 1$, the two events coincide (are equivalent).

The closer $R_{A,B}$ is to the number -1, the "denser one within the other" the events A and $\overline{B}$ are, and when $R_{A,B} = -1$, the two events coincide (are equivalent). Then the events $\overline{A}$ and B are denser within one another.

R3. The correlation coefficient $R_{A,B}$ has the same sign as the other measures of the dependence between two random events $A$and B (and this is the sign of the connection $\delta(A,B)$, as it is the sign of the two regression coefficients $r_B(A)$ and $r_A(B)$). Knowledge of $R_{A,B}$ allows calculating the posterior probability of one of the events under the condition that the other one occurred. For instance, $P(A| \quad B)$ will be determined by the rule

$$P(A|B) = P(A) + R_{A,B}\sqrt{\frac{P(\overline{B})P(A)P(\overline{A})}{P(B)}}. \tag{15}$$

This rule reminds again of Bayes' rule for posterior probabilities. The net increase or decrease in the posterior probability compared to the prior probability equals the quantity $R_{A,B}\sqrt{\dfrac{P(\overline{B})P(A)P(\bar{A})}{P(B)}}$, and depends only on the value of the mutual correlation $R_{A,B}$ (positive or negative).

Important Note. The proposed rules (5), (9) and (15) for evaluating the posterior probability $P(A|B)$ via any of the proposed measures of dependence can be turned into a powerful tool in calculating posterior probabilities, with a brilliant use of the statistical information for practical purposes. Note that the definitions of $\delta(A,B)$, $r_B(A)$, $r_A(B)$, and $R_{A,B}$ involve only probabilities. These probabilities have natural frequency statistical estimations. Imagine that A is an event in the future (tomorrow), and B is an event from the past (yesterday, or today), then we can immediately see what a tremendous tool for forecasting such dependence measures could be.

R4. It is fulfilled that $R_{\overline{A},B} = R_{A,\overline{B}} = -R_{A,B}$; $R_{\overline{A},\overline{B}} = R_{A,B}$.

R5. The Freshet-Hoefding inequalities for the Correlation Coefficient are as follows

$$\max\left\{-\sqrt{\frac{P(A)P(B)}{P(\overline{A})P(\overline{B})}}, -\sqrt{\frac{P(\overline{A})P(\overline{B})}{P(A)P(B)}}\right\} \leq R(A,B)$$
$$\leq \min\left\{\sqrt{\frac{P(A)P(\overline{B})}{P(\overline{A})P(B)}}, \sqrt{\frac{P(\overline{A})P(B)}{P(A)P(\overline{B})}}\right\}$$

Their use is similar to the ones for the regression coefficient. Notice their importance in construction (e.g., for simulation, or modeling purposes) of events with given individual probability, and desired mutual correlation.

For example if $P(A) = 0.3$, $P(B) = 0.6$, then the correlation coefficient $R_{A,B}$ must be kept within the interval $\max\left[-\sqrt{\frac{0.18}{0.28}}, -\sqrt{\frac{0.28}{0.18}}\right] = -\sqrt{\frac{9}{14}}$ as a left end, and $\min\left[\sqrt{\frac{0.12}{0.42}}, \sqrt{\frac{0.42}{0.12}}\right] = \sqrt{\frac{2}{7}}$ as a right end. Only values $R_{A,B} \in \left[-\sqrt{\frac{9}{14}}, \frac{2}{7}\right]$ are legitimate values for any assumed regression coefficient in simulations under the given marginal probabilities.

Some warnings. First of all, we notice that the introduced measures of dependence between random events are not transitive. It is possible that the random event A is positively associated with a random event B, and B is positively associated with a third random event , but the event A may be negatively associated with . The association for mutually exclusive events is negative, while for the non-exclusive pairs (A, B) and $(B,)$ every kind of dependence is possible. The dependence between compound events is not an integral feature, and is composed from a number of particular details.

3. Empirical estimation of the measures of dependence between two random events. The fact that the considered measures of dependence between random events are constructed from their marginal, joint, or conditional probabilities makes them attractive, easy for statistical estimation and practical use.

Let in N independent experiments (or observations) the random event A occur k_A times, the random event B occur k_B times, and the event $A \cap B$ occur

$k_{A\cap B}$ times. It is well known that the statistical estimator of the probability $P(A)$ is the ratio k_A/N, and the estimators of other probabilities are similar. In this way in the definitions of the introduced measures of dependence all the probabilities can be statistically estimated.

The estimator of the connection between the two events is given by the formula

$$\hat{\delta}(A,B) = \frac{k_{A\cap B}}{N} - \frac{k_A}{N}\cdot\frac{k_B}{N};$$

The estimators of the two regression coefficients are

$$\hat{r}_A(B) = \frac{\frac{k_{A\cap B}}{N} - \frac{k_A}{N}\cdot\frac{k_B}{N}}{\frac{k_A}{N}(1-\frac{k_A}{N})}; \quad \text{and} \quad \hat{r}_B(A) = \frac{\frac{k_{A\cap B}}{N} - \frac{k_A}{N}\cdot\frac{k_B}{N}}{\frac{k_B}{N}(1-\frac{k_B}{N})};$$

The estimator of the correlation coefficient is given by the rule

$$\hat{R}(A,B) = \frac{\frac{k_{A\cap B}}{N} - \frac{k_A}{N}\cdot\frac{k_B}{N}}{\sqrt{\frac{k_A}{N}(1-\frac{k_A}{N})\frac{k_B}{N}(1-\frac{k_B}{N})}}.$$

According to the rules of statistical estimation, all these estimators are consistent. Moreover, the estimator of the connection $\hat{\delta}(A,B)$ is unbiased. The estimators obtained in this way are also *maximum likelihood estimators* and have the respective MLE properties.

We notice that the use of the conditional probabilities in the estimations of the regression coefficients is not needed. We personally are excited by the opportunities offered by these measures.

4. Applications.

4.1. Categorical variables. As an illustration of the proposed measures of dependence between random events we analyze here an example from the book by Alan Agresti Categorical Data Analysis [1]. The following table represents the observed data about the yearly income of people and their job satisfaction.

The probabilities in each category

$$P_{i,j} = \frac{n_{i,j}}{n}, P_{i,.} = \frac{n_{.i}}{n}, P_{.,j} = \frac{n_{.j}}{n},$$

in the above table produce the join empirical distribution of the two categories. $P_{i,j}$ is the probability that a new observation will fall in the respective subcategory

Table 1. Observed Frequencies of Income and Job Satisfaction

Job Satisfaction					
Categories by the income in US $$	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied	Total Marginally
Less than 6000	20	24	80	82	206
6000–15 000	22	38	104	125	289
15 000–25 000	13	28	81	113	235
Above 25 000	7	18	54	92	171
Total Marginally	62	108	319	412	901

Table 2. Joint and marginal distributions between the observed categories and subcategories.

Job Satisfaction					
Income US $$	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied	Total (marginal) distribution
Less than 6000	0.02220	0.02664	0.08879	0.09101	0.22864
6000–15 000	0.02442	0.04217	0.11543	0.13873	0.32075
15 000–25 000	0.01443	0.03108	0.08990	0.12542	0.26083
Above 25 000	0.00776	0.01998	0.05993	0.10211	0.18978
Total (marginal) distribution	0.06881	0.11987	0.35405	0.45727	1.00000

i by income, and subcategory j by job satisfaction. Table 2 presents the join distribution for the two categorical variables.

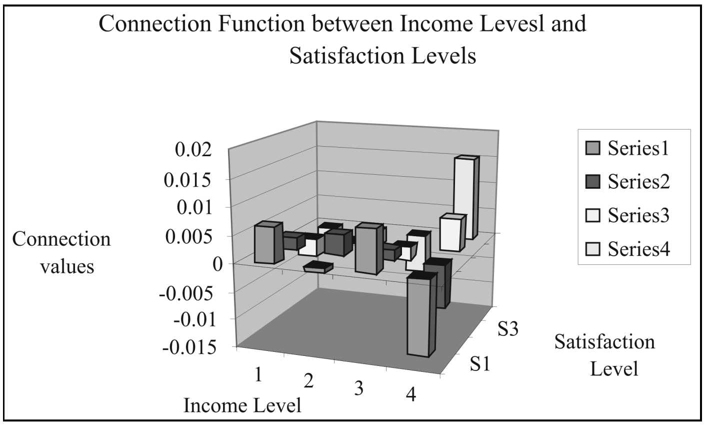
Applying the rules for the proposed measures of dependence between random events, and using the empirical probabilities in Table 2, we obtain these measures as shown in the tables from 3 to 6. A positive sign indicates a positive local dependence between the two sub-categories, and a negative sign indicates the opposite in this locality. The negative association is marked in bold, and the positive areas of association in normal.

Since numbers speak less than graphs, we immediately give graphic presentation of these two-argument functions giving the dependence structure between the observed categories (cross sections of any two sub-categories). As the ancient Greeks used to say, just "look and decide".

For instance, in Table 4 the number $r_{VeryDissatisfied}(< 6000) = 0.100932704$ indicates positive dependence of the category of the lowest income "< 6000" on the category: "Very Dissatisfied" for the Job Satisfaction. The same number with negative sign $r_{VeryDissatisfied}(\overline{< 6000}) = -0.100932704$ indicates

Table 3. Empirical Estimations of the connection function for each particular category of Income and Job Satisfaction $\delta(IncomeGroup_i, Satisfaction_j)$

	Job Satisfaction			
Income US $$	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied
Less than 6000	0.006467282	**−0.00077**	0.00784	**−0.01354**
6000–15 000	0.002349193	0.003722	0.001868	**−0.00794**
15 000-25 000	**−0.00351771**	**−0.00019**	**−0.00245**	0.00615
Above 25 000	**−0.00529876**	**−0.00277**	**−0.00726**	0.015329
Total sum in a column	0	0	0	0

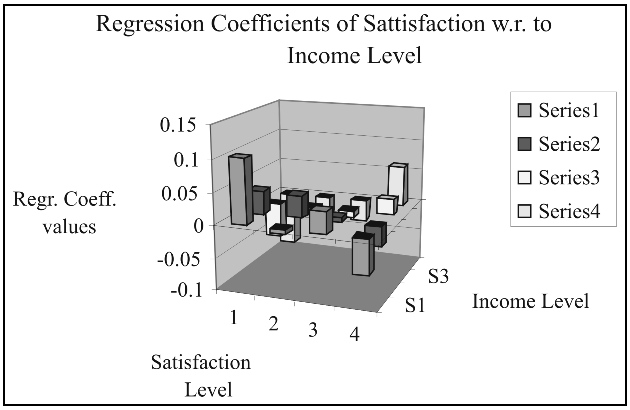


the negative strength of dependence of all the other income categories, opposite to "< 6000" on the category: "Very Dissatisfied" for the Job Satisfaction. Similarly to the connection function, the sums of numbers from several cells in a column of Table 4 (or in a row of Table 5) will indicate the strength of dependence of the union of the sub-categories of the respective factor "Income" on the sub-category of "Job, Satisfaction" corresponding to the column (with analogous switch of factor's interpretation).

The two matrices for regression coefficients allow calculating the correlation coefficients between every pair of sub-categories of the two factors. Table 6 summarizes these calculations. The numbers actually represent the numerical estimations of the respective mutual local correlation coefficients. Obviously, each of these numbers gives the **local average measure of dependence** between the two factors. Unfortunately, the sum of the numbers in a vertical or horizontal line does not have the same or similar meaning as in the cases of regression in regression matrices. Also, the sums of the numbers in a row or in a column do

Table 4. Empirical Estimations of the regression coefficient between each particular level of income with respect to the job satisfaction $r_{Satisfaction_j}(IncomeGroup_i)$

Job Satisfaction				
Income US $$	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied
Less than 6000	0.100932704	**−0.00727**	0.034281	**−0.05456**
6000–15 000	0.036663063	0.035276	0.00817	**−0.03199**
15 000–25 000	**−0.05489976**	**−0.00176**	**−0.0107**	0.024782
Above 25 000	**−0.08269601**	**−0.02625**	**−0.03175**	0.061768



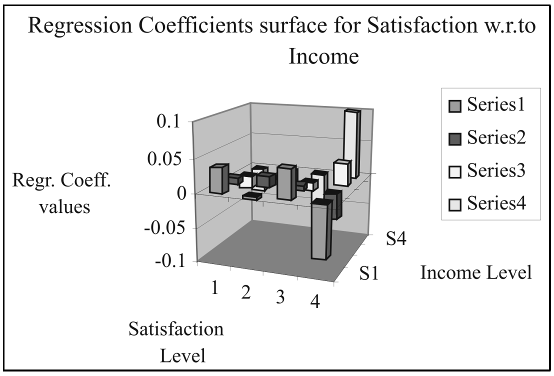
not equal zero as above.

Some predictions of the Income sub-group are now possible. After we know the job satisfaction subcategory, the marginal probabilities and the values of local connections (or local regression coefficients, or correlation) we can use any of the equations (5), (9), or (15). Table 7 presents these posterior probabilities. For comparison, the "prior" probabilities for each subcategory are given in the margin (the last column).

The greatest numbers inside the table show the "hot local positions", where the conditional probability increases compared to the prior (unconditional) probability. The blue colored numbers show the places of local decrease in the posterior probability. If someone answers "very dissatisfied", then the highest chance is that this is a person whose income is in the range of 6000–15 000. The chances that such answer comes from person of income "Less than 6000" increase by approximately .10. If someone answers "Very Satisfied", then the lowest chances are that this is a person of income "$<$ \$6000". This is totally different (as the entire order of income sub-classes) from the prior distribution of the income groups. Also, the

Table 5. Empirical Estimations of the regression coefficient between each particular level of the job satisfaction with respect to the income

	Job Satisfaction			
Income US $$	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied
Less than 6000	0.03667013	**−0.00435**	0.044454	**−0.07677**
6000–15 000	0.01078257	0.017082	0.008576	**−0.03644**
15 000-25 000	**−0.01824561**	**−0.00096**	**−0.01269**	0.0319
Above 25 000	**−0.03446045**	**−0.01801**	**−0.04723**	0.099694



sum of the numbers in a column gives 1, since there are all the possible parts of the sure event ($S = \sum_i A_i$).

Similarly, if one knows the income group A_i and has either of the measures of dependence and $P(B_j)$ for particular group j, then the conditional (posterior) probabilities $P(B_j|A_i)$ of the job satisfaction groups can be re-evaluated by the same rules. Table 8 presents these probabilities. For comparison, prior $P(B_j)$ are given on the margin (the last row).

Here we would like to observe that similar "categorizations" can be made for any two numeric random variables, and what we see and read in the above tables can be used for studies of the local structure of dependence between random variables.

4.2. Numeric random variables. We illustrate now the idea of transferring these measures of dependence between random events into measures of local dependence between random variables (r.v. s). These measures allow studying the behavior of interaction between any pair of numeric r.v. s (X, Y) throughout the sample space, and better understanding and use of dependence. Some examples of popular distributions will be provided further.

Table 6. Empirical Estimations of the correlation coefficient between each particular income group and the categories of the job satisfaction $R(IncomeGroup_i, Satisfaction_j)$

Job Satisfaction					
Very Dissatisfied	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied	Total $\sum_k P(A_i\|B_k) = 1$
< 6000	0.09709587	**−0.11651505**	0.388339748	**−0.398049335**	1
6000–15 000	0.07613406	0.13147311	0.359875292	**−0.432517537**	1
15 000–25 000	**−0.05532339**	**−0.11915807**	**−0.34466894**	0.480849596	1
$> 25\ 000$	**−0.04088945**	**−0.1052798**	**−0.3157867**	0.538044051	1
Unconditional Probabilities $P(B)$	0.06881	0.11987	0.35405	0.45727	1.00000

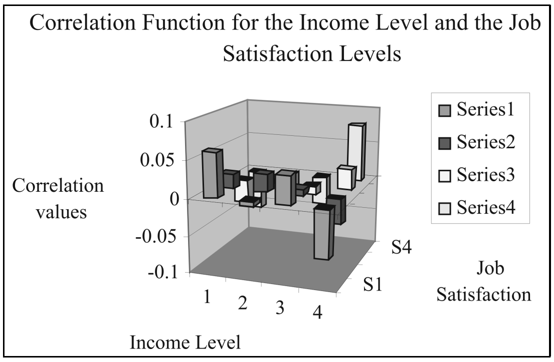


Table 7. Forecasted probabilities $P(A_i|B_j) = P(A_i) + \delta(A_i, B_j)/P(B_j)$ of particular income group given the categories of the job satisfaction

Income US $$	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied	Unconditional Probabilities $P(A)$
< 6000	0.32262753	**0.22224076**	0.250783788	**0.19902902**	0.22864
6000–15 000	0.35489028	0.35179778	0.326027397	**0.303387495**	0.32075
15 000–25 000	**0.20970789**	**0.25928089**	**0.253918938**	0.274279966	0.26083
$> 25\ 000$	**0.11277431**	**0.16668057**	**0.169269877**	0.223303519	0.18978
Total $\sum_i P(B_k\|A_i) = 1$	1.00000	1.00000	1.00000	1.00000	1.00000

Let the joint cumulative distribution function (c.d.f.) be $F(x, y) = P(X \leq x, Y \leq y)$, and marginals $F(x) = P(X \leq x$, $G(y) = P(Y \leq y)$.

Table 8. Forecast of the probabilities $p(B_j|A_i) = P(B_j) + \delta((A_i, B_j)/PA_i)$ of particular income group given the categories of the job satisfaction

Job Satisfaction					
Very Dissatisfied	Very Dissatisfied	Little Satisfied	Moderately Satisfied	Very Satisfied	Total $\sum_k P(A_i|B_k) = 1$
< 6000	0.09709587	**0.11651505**	0.388339748	**0.398049335**	1
6000–15 000	0.07613406	0.13147311	0.359875292	**0.432517537**	1
15 000–25 000	**0.05532339**	**0.11915807**	**0.344668941**	0.480849596	1
$> 25\ 000$	**0.04088945**	**0.1052798**	**0.3157867**	0.538044051	1
Unconditional Probabilities $P(B)$	0.06881	0.11987	0.35405	0.45727	1.00000

Let us introduce the events

$$A = \{X \le x\}; \quad B = \{Y \le y\}, \quad \text{for any} \ \ x, y \in (-\infty, \infty).$$

Then the measures of dependence between events A and B turn into a cumulative measure of dependence between the pair of r.v.s X and Y at the point (x, y). Naturally, they can be named and calculated as follows:

Connection at the point (x, y)

$$\delta(x, y) = F(x, y) - F(x)G(y);$$

Regression coefficient of X with respect to Y , and of Y with respect to X at the point (x, y), by the use of equation (8)

$$R_Y(X; x, y) = \frac{F(x, y) - F(x)G(y)}{F(x)[1 - F(x)]}, \quad \text{and} \quad R_X(Y; x, y) = \frac{F(x, y) - F(x)G(y)}{G(y)[1 - G(y)]}.$$

Correlation coefficient between the r.v.s X and Y at the point (x, y), by the use of equation (14)

$$R_{X,Y}(x, y) = \frac{F(x, y) - F(x)G(y)}{\sqrt{F(x)[1 - F(x)}\sqrt{G(y)[1 - G(y)]}}.$$

In Dimitrov et al. [4], [5] and [6], these measures have been studied and called cumulative measures of dependence between components of a multivariate random vector. As an important geometric illustration of the dependence structure between two r.v. s it is suggested to draw curves of constant dependence. These are the sets of all possible points of coordinates (x, y) in the plane where the respective measure is constant. If we denote by $m_{X,Y}(x,y)$ the value of any particular measure of dependence listed above, then these curves of constant dependence at level m_0 are defined by the equations

$$C(m_0) = \{(x,y);\ m_{X,Y}(x,y) = m_0,\ (x,y) \in R^2\}.$$

Moreover, given an interpretation of the equation

$$z = m_{X,Y}(x,y)$$

as an equation of a 3-d surface, such surfaces will tell us everything about the cumulative local dependence (in regard of the considered measure) between the pair (X,Y) on the plane R^2.

We take the following examples from that study to illustrate how these relationships can be used to study the cumulative dependencies between random variables.

Example 1. The **Clayton Archimedean copula** is a two-dimensional distribution on the square $[0,1]\times[0,1]$, defined by Nelsen (see [8]) the joint pdf

$$C(x,y) = \max(x^{-\theta} + y^{-\theta} - 1,^{0})^{-\frac{1}{\theta}}, \quad (x,y) \in [0,1]\times[0,1].$$

Its connection function at a point (x,y) is then

$$\delta_{X,Y}(x,y) = \max(x^{-\theta} + y^{-\theta} - 1,^{0})^{-\frac{1}{\theta}} - x\cdot y, \quad (x,y) \in [0,1]\times[0,1].$$

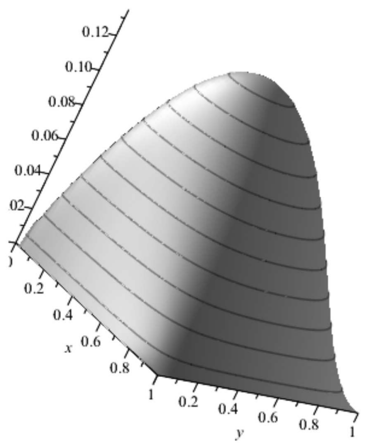


Fig. 1a. The connection function $\delta_{X,Y}(x,y)$

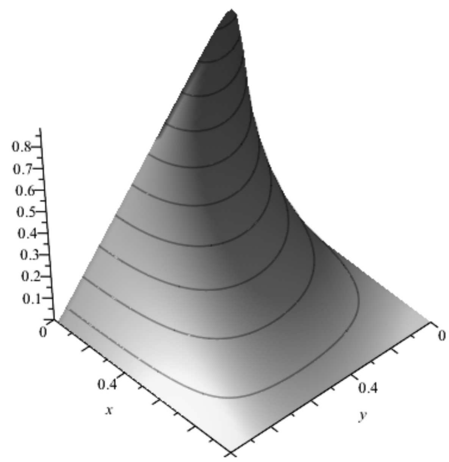


Fig. 1b. The regression coefficient function $R_X(Y;x,y)$

For the Clayton copula for the case with value of $\theta = 2$

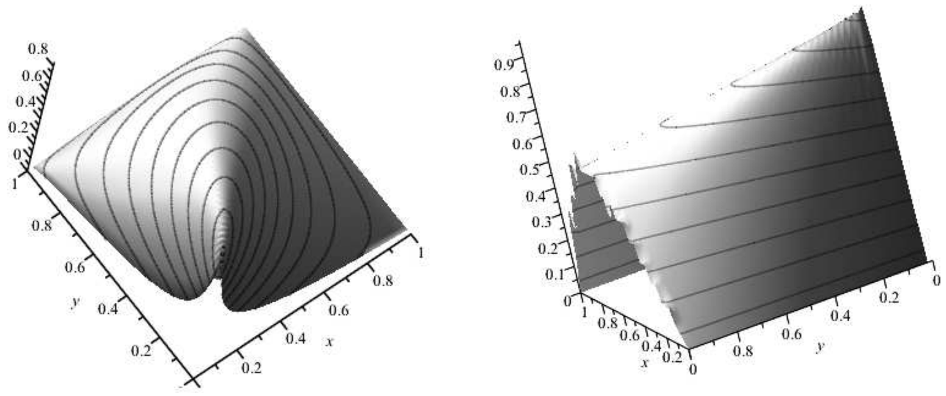


Fig. 2. Correlation coefficient function $\rho_{X.Y}(x,y)$ and Regression function $R_Y(X;x,y)$ structures for the Clayton copula in the case $\theta = 4$

The regression coefficient function $R_X(Y;x,y)$ at a point (x,y) is

$$R_X(Y;x,y) = \frac{\max(x^{-\theta} + y^{-\theta} - 1, 0)^{-\frac{1}{\theta}} - xy}{x(1-x)}, \quad (x,y) \in [0,1] \times [0,1].$$

Below are the surface graphs of these functions with the level curves for the case with value of $\theta = 2$.

The correlation coefficient function at a point (x,y) is

$$\rho_{X,Y}(x,y) = \frac{\max(x^{-\theta} + y^{-\theta} - 1, 0)^{-\frac{1}{\theta}} - xy}{\sqrt{x(1-x)y(1-y)}}, \quad (x,y) \in [0,1] \times [0,1].$$

Its surface for the case with value of $\theta = 4$, graph with level curves on it, is shown on Fig. 2. Also on the right-hand side of Fig. 2 is presented the Regression coefficient function $R_Y(X;x,y)$ for the same case.

Example 2. The simplest **Bivariate Poisson distribution** with dependent components. This is a two-dimensional discrete distribution, for which the study of the local dependence structure using our approach is one of the best illustrations of how one can see all at once. We use the bivariate discrete distribution presented as in [7], by the three positive parameters (λ, μ, ν) family

$$P(X = x, Y = y) = e^{-\lambda-\mu-\nu} \frac{\lambda^x}{x!} \cdot \frac{\mu^y}{y!} \sum_{k=0}^{\min(x,y)} \binom{x}{k} \binom{y}{k} k! \left(\frac{\nu}{\lambda\mu}\right)^k.$$

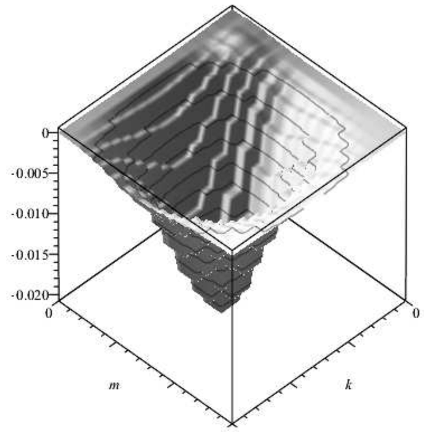


Fig. 3a. The connection function $\delta_{X,Y}(x,y)$

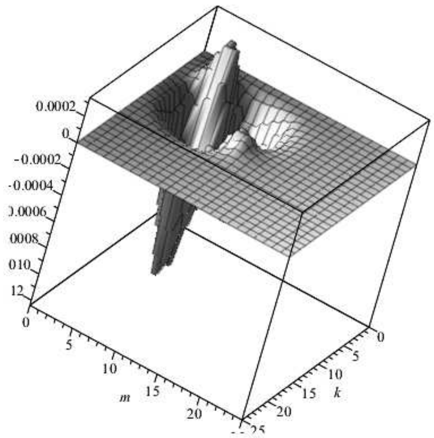


Fig. 3b. The correlation function $\rho_{X.Y}(x,y)$

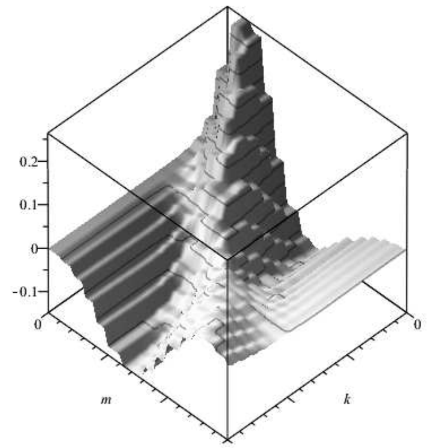


Fig. 3c. The regression coefficient function $R_Y(X;x,y)$

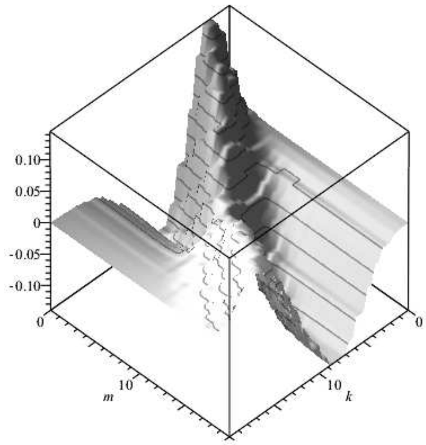


Fig. 3d. The regression coefficient function $R_X(Y;x,y)$

Here $x, y = 0, 1, 2, \ldots$ are the possible values of the variables X and Y. If M_1, M_2, and M_3 are three independent Poisson distributed r.v. s with parameters λ, μ, and ν respectively, then the dependence between X and Y comes from the fact that X is distributed as the sum $X = M_1 + M_3$, and $Y = M_2 + M_3$. Inclusion of M_3 in both sums makes them dependent. The marginal distributions of X and Y are Poisson with parameters $\lambda + \mu$ and $\lambda + \nu$ respectively. We say these facts to avoid explicating easy but cumbersome expressions for the connection function

$\delta_{X,Y}(x, y)$, the two regression coefficient functions $R_X(Y; x, y)$ and $R_X(Y; x, y)$, and for the correlation function $\rho_{X.Y}(x, y)$ at each point (x, y) with integer coordinates. When written and programmed, graphs of these functions for the case of $\lambda = 3$, $\mu = 2$ and $\nu = 5$ are shown on Fig. 3 below. As the ancient Greek geometers used to say, just watch and conclude what kind of dependence works at what point, and what is its strength.

5. Conclusion. The measures of dependence once well established and studied for the case of random events, can be easily and naturally turned into a powerful tool to study the local dependence structure between random variables of either non-numeric or numeric nature. This simple approach makes it convenient for inclusion in education within traditional university courses in probability and statistics, where dependence is presented vaguely and frequently is misunderstood.

The proposed approach can be used to study the local dependence structure in numerous multivariate distributions discussed in the studies and modeling various classic or new situations of applied probability and statistics.

There are multiple opportunities to use the local dependence between random variables to simulate such dependences on different areas of the plane. But careful warning is to comply with the Freshet-Hoefding restrictions in selecting dependence strengths.

Knowledge of the dependence structure can play an important role in decision making based on information about one component in risk models, investments, politics, financial mathematics, insurance and many other fields of applications.

REFERENCES

[1] Agresti A. Categorical Data Analysis. John Wiley & Sons, Hew York, 2006.

[2] Dimitrov B., N. Yanev. Probability and Statistics. A textbook, Sofia University “Kliment Ohridski”, Sofia (Second Edition 1998, Third Edition 2007), 1991.

[3] Dimitrov B. Some Obreshkov Measures of Dependence and Their Use. *Compte Rendus de l'Academie Bulgare des Sciences,* **63** (2010), No 1, 15–18.

[4] Dimitrov B. Measures of dependence in reliability. In: Proceedings of the 8-th International Conf. on Mathematical Methods in Reliability MMR2014, Stellenbosh, South Africa, July 1–4, 2013, 65–69.

[5] Esa S., B. Dimitrov Dependencies in the world of politics, In: Proceedings of the 8-th International Conf. on Mathematical Methods in Reliability MMR2014, Stellenbosh, South Africa, July 1–4, 2013, 70–73.

[6] Dimitrov B., S. Esa, N. Kolev, G. Pitselis. Transfer a Global Measures of Dependence into Cumulative Local, *Applied Mathematics,* **5** (2014), 315–327.

[7] Kocherlakota S. On the compounded bivariate Poisson Distribution: A unified treatment, *Ann. Inst. Statist.\Math.*, **49** (1988), No 1, 61–70.

[8] Nelsen R. An Introduction to Copulas. Springer, New York, 2006.

[9] Obreshkov N. Probability Theory. Nauka i Izkustvo, Sofia, 1963 (In Bulgarian).

Boyan Dimitrov
Kettering University
Dept. of Mathematics, Flint
Michigan 48504-4898, USA
e-mail: bdimitro@kettering.edu

Received December 4, 2014
Final Accepted February 23, 2015