ISSN: 0067-2904
GIF: 0.851

Using Linear Local Dependence Measure to Study the factors that are Leading to the Growth of Preferring the Application in Private Universities of Iraq

Fedaa N. Abdulahad*, Ranen Z. Ahmood
Department of Mathematics and Computer Applications, College of Science, Al-Nahrain University, Baghdad, Iraq

Abstract

In this paper a measure of linear local dependence has been used between two random variables and a study is conducted for the properties of this measure where two examples of bivariate probability distributions has been considered, which are bivariate Gumbel distribution and bivariate Beta-Stacy distribution, and applied on data collected by using a questionnaire conducted to study the reasons for the increase of application in private collages in Iraq. Five elements has been considered as random variables and the dependence has been measured between every two elements to estimate how correlated these elements are and their effect on the application in private collages of Iraq generally and Baghdad specifically.

Keywords: dependency, Gumbel Distribution, Beta-Stacy Distribution, Correlation Coefficient.

استخدام مقياس للاعتمادية الخطية لدراسة العوامل المؤدية الى الاقبال الشديد على الجامعات الاهلية في العراق

فداء نوئيل عبدالاحد*، رنين زيد حمود
قسم الرياضيات وتطبيقات الحاسوب، كلية العلوم، جامعة النهرين، بغداد، العراق

الخلاصة:

في هذا البحث تم استخدام مقياس جديد للاعتمادية الخطية بين متغيرين عشوائيين ودراسة خصائص هذا المقياس واخذ مثالين من التوزيعات الاحتمالية الثنائية وتطبيق المقياس عليها والتي كانت توزيع Gumbel الثنائي وتوزيع Beta-Stacy الثنائي كما تم تطبيق المقياس على بيانات مأخوذة من استمارات استبيان صممت لغرض دراسة اسباب زيادة الاقبال على الكليات الاهلية في العراق. تم اختيار خمسة اسباب كمتغيرات عشوائية وقياس الاعتمادية بين كل متغيرين لتقدير العلاقة بين هذه العناصر وتأثيرها في التقديم على هذه الكليات في العراق بصورة عامة وبغداد بصورة خاصة.

1. Introduction

It has been noticed lately that the number of students applying in the private collages is increasing all over Iraq and especially in Baghdad which led to noticeable increase in the number of these collages despite of the relatively high studying fees also it has been noticed that they accept applicants of different ages and diverse academic backgrounds this all led to considering a study that includes some of the reasons that encourages application to the private collages.

*Email:fn5382@yahoo.com

Researches have been conducted a new statistical measure several years ago to study the relation between the variables where Bjerve and Doksum [1], Doksum [2], Blyth [3,4] and Jones [5] introduce and discuss a "correlation curve", which is a generalization of Pearson correlation coefficient, and Bairamov [6] proposed a linear local dependence function between two variables X and Y based on regression concept. Nadarajah [7] provide details analysis (algebraic and numerical) of the linear local dependence function for the class of bivariate extreme value distributions. Tavangar and Asadi[8] showed that the measure of local dependence can be applied to measure the dependency between two residual lifetime random variables.

In this paper the considered measure proposed by Bairamov and kozuboski(2003) is used to compare between five variables chosen based on a questionnaire conducted to study the effects of the increase on the application for the private collages in Iraq in an attempt to find a relation between them, where the variables are X_1 is acceptance of low high school graduation average(overall degree), X_2 is for flexibility and tolerance with the students, X_3 is for the intention of the governmental employees to improve their income, X_4 is for accepting non high school graduates those who cannot apply to the governmental universities and X_5 is the absence of an age limit for applying which is considered one of the strongest reasons for the growing interest in applying to these colleges.

The linear dependence equation is driven between two variables for two probability distributions which are bivariate Gumbel distribution and bivariate Beta-Stacy distribution and a comparison is done between the results obtained from this measure and results obtained from the traditional linear dependence equation, the comparison showed that the results of the two methods are very close.

2. A Local Dependence Function

Let X and Y be random variables (r.v's) with marginal distribution functions (d.f.'s) and densities (p.d.f.'s). Consider the following function of two variables

$$H(x,y)=\frac{E\{(X-E(X|Y=y))(Y-E(Y|X=x))\}}{\sqrt{E\{(X-E(X|Y=y))^2\}}\sqrt{E\{(Y-E(Y|X=x))^2\}}} \qquad \text{... (1)}$$

which is obtained from the expression of the Pearson correlation coefficient by replacing mathematical expectation E(X) and E(Y) by conditional expectation $E(X|Y=y)$ and $E(Y|X=x)$, respectively. By construction, H(x,y) can be interpreted as a local dependence function characterizing the dependence between X and Y at the point (x,y). After some simple algebra, equation (1) can be written as

$$H(x,y)=\frac{\rho+\phi_X(y)\phi_Y(x)}{\sqrt{1+\phi_X^2(y)}\sqrt{1+\phi_Y^2(x)}} \qquad \text{... (2)}$$

Where

$$\rho=\frac{Cov(X,Y)}{\sigma_X\sigma_Y} \qquad \text{... (3)}$$

Is the Pearson correlation coefficient of X and Y,

$$\phi_X(y)=\frac{E(X)-E(X|Y=y)}{\sigma_X},\phi_Y(x)=\frac{E(Y)-E(Y|X=x)}{\sigma_Y}, \qquad \text{... (4)}$$

σ_X and σ_Y are the standard deviations of X and Y respectively. The function H(x,y) will be referred to as a local dependence function[6].

Now in general assume that A and B are two sets such that $A\times B \subseteq N_{X,Y}$. Motivated by the definition of linear local dependence function in equation (1), we propose the following set function for measuring the local association between X and Y

$$H(A,B)=\frac{E\{(X-E(X|Y\in B))(Y-E(Y|X\in A))\}}{\sqrt{E\{(X-E(X|Y\in B))^2\}}\sqrt{E\{(Y-E(Y|X\in A))^2\}}} \qquad \text{... (5)}$$

H(A,B) measures the dependency between two random variables X and Y under the condition that X and Y belong to subsets A and B of their supports, respectively. Similar to equation (2), it can be shown that an alternative expression for equation (5) is

$$H(A,B)=\frac{\rho+\phi_X(B)\phi_Y(A)}{\sqrt{1+\phi_X^2(B)}\sqrt{1+\phi_Y^2(A)}} \qquad \text{... (6)}$$

Where

$$\phi_X(B) = \frac{E(X) - E(X|Y \in B)}{\sigma_X}, \phi_Y(A) = \frac{E(Y) - E(Y|X \in A)}{\sigma_Y}, \quad \ldots (7)$$

Note that if we take A={x} and B={y}, then H(A,B) reduces to H(x,y) in (1)[8]. Following are, some properties of the local dependence function H:

Let (X,Y) have a bivariate distribution with finite second moments, Pearson correlation coefficient ρ, support $N_{X,Y}$, and local dependence function $H=H_{X,Y}$. Then:

1. If X and Y are independent, then H(x,y)=0 for all (x,y) $\in N_{X,Y}$.
2. $|H(x,y)| \leq 1$ for all (x,y) $\in N_{X,Y}$
3. If $H(x,y) = \pm 1$ for some (x,y) $\in N_{X,Y}$, then $\rho \neq 0$.
4. If Y=aX+b, a.s., then H(X,Y)=sign(a), a.s.
5. If $\rho = \pm 1$ then $H(x,y) = \pm 1$, a.s.
6. If $\tilde{X} = aX + b;\ \tilde{Y} = cY + d, then\ H_{\tilde{X},\tilde{Y}}(\tilde{x},\tilde{y}) = sign(ac)H_{X,Y(x,y)}$. Where $\tilde{x} = ax + b\ and\ \tilde{y} = cy + d$.
7. If H(x,y) = 0 for all (x,y) $\in N_{X,Y}$, then either $EX = E(X|Y = y)\ or\ EY = E(Y|X = x)$ for all (x,y) $\in N_{X,Y}$, and $\rho = 0$.
8. The point (x^*,y^*) satisfying $\phi_X(y^*) = \phi_Y(x^*) = 0$ is a saddle point of H and $H(x^*,y^*) = \rho$.[6,8]

3. Examples:

In this section we illustrate the concept of local linear dependence function by means of two examples, chosen to demonstrate the special features of the function at hand.

3.1 Bivariate Beta-Stacy Distribution:

Consider the following bivariate distribution, referred to as a bivariate Beta-Stacy distribution with p.d.f.

$$f(x,y) = \frac{1}{\Gamma\alpha\, \beta^{\alpha}\ \beta(\theta_1,\theta_2)} x^{\alpha-\theta_1-\theta_2}\, y^{\theta_1-1}(x-y)^{\theta_2-1} e^{-\frac{x}{\beta}} \quad \ldots (8)$$

Where X follows Gama distribution with two parameters and with mean $=EX = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$ [9,10]. From the equation (4)

$$\phi_x(y) = \frac{\alpha\beta - x}{\sqrt{\alpha\beta^2}} \quad \ldots (9)$$

$$\phi_y(x) = \frac{\frac{\theta_1\alpha\beta}{(\theta_1+\theta_2)} - \frac{\theta_1}{(\theta_1+\theta_2)}\, x}{\sqrt{\frac{(\theta_1+1)\theta_1\alpha\beta^2(1+\alpha)}{(\theta_1+\theta_2+1)(\theta_1+\theta_2)} - \frac{\theta_1^2\alpha^2\beta^2}{(\theta_1+\theta_2)^2}}} \quad \ldots (10)$$

The variables are independent with $\rho = 0$. For this distribution the local dependence function

$$H(x,y) = \frac{\rho + \dfrac{\alpha\beta - {}^{x\theta_1}\!/_{(\theta_1+\theta_2)}\,[\alpha\beta - x]}{\sqrt{\alpha\beta^2}\sqrt{\alpha\beta^2\frac{\theta_1}{(\theta_1+\theta_2)}\left[\frac{(\theta_1+1)(1+\alpha)}{(\theta_1+\theta_2+1)} - \frac{\theta_1}{(\theta_1+\theta_2)}\alpha\right]}}}{\sqrt{1+\frac{(\alpha\beta-x)^2}{\alpha\beta^2}}\sqrt{1+\dfrac{\frac{\theta_1^2}{(\theta_1+\theta_2)^2}[\alpha\beta - x]^2}{\frac{\theta_1}{(\theta_1+\theta_2)}\alpha\beta^2\left[\frac{(\theta_1+1)(1+\alpha)}{(\theta_1+\theta_2+1)} - \frac{\theta_1}{(\theta_1+\theta_2)}\alpha\right]}}} \quad \ldots (11)$$

3.2 Bivariate Gumbel distribution:

Consider the distribution of random vector (x,y)with p.d.f.[8]

$$f(x,y) = e^{-(x+y)}\{1 + \alpha(2e^{-x} - 1)(2e^{-y} - 1)\}\,; x,y > 0 \quad \ldots (12)$$

As the marginal distribution of X and Y are standard exponential, the correlation of X and Y is

$$\rho = \frac{1}{4}\alpha \quad \ldots (13)$$

And

$$\phi_x(y) = \alpha e^{-y} - \frac{1}{2}\alpha \quad , \quad \phi_y(x) = \alpha e^{-x} - \frac{1}{2}\alpha \quad \ldots (14)$$

For $\alpha = 0$ the variables are independent with$\rho = 0$. The local dependence function takes the form

$$H(x,y) = \frac{1/4\,\alpha + (\alpha e^{-y} - \frac{1}{2}\alpha)(\alpha e^{-x} - \frac{1}{2}\alpha)}{\sqrt{1 + (\alpha e^{-y} - \frac{1}{2}\alpha)^2}\sqrt{1 + (\alpha e^{-x} - \frac{1}{2}\alpha)^2}} \quad \ldots (15)$$

4. The numerical example:

In this section a study is conducted on the elements that lead to increase the application to the private collages by distributing a questionnaire sheet that is set specifically for this study. The distribution included three different universities, the first one is university of Baghdad which is governmental university, a second and a third are Al-Maamoon and A-Mansour private universities. Three departments are considered, two of them are scientific departments which are Computer and Communication engineering departments respectively and the third one is a humanitarian department which is Accounting department. The questionnaire was distributed to all students in the mentioned departments and on both genders. The sheet included questions about (5) reasons which are considered (5) the random variables as the measure of local dependence is applied on them and the variable as follows.

X_1: is acceptance of low high school graduation average.

X_2: is for flexibility and tolerance with the students.

X_3: is for the intention of the governmental employees to improve their income.

X_4: is for accepting non high school graduates those who cannot apply to the governmental universities.

X_5: is the absence of an age limit.

And after collecting the sheets and categorizing the results according to the variables and universities, the results came as shown in the Table-1:

Table 1-The results according to the variables and universities

Variables		X_1	X_2	X_3	X_4	X_5	No. of Students
departments	Collages						
Computers eng.	Baghdad	13	10	15	13	16	19
Communication eng.		19	18	24	10	17	27
Accounting dep.		10	10	13	10	10	13
Computers eng.	Al-Mansour	18	6	21	13	4	21
Communication eng.		10	2	14	12	8	17
Accounting dep.		9	4	12	10	9	16
Computers eng.	Al-Maamoon	14	12	18	16	18	21
Communication eng.		14	5	10	13	10	16
Accounting dep.		11	14	18	15	15	19

After collecting the data every two random variables were taken together and applying the form of H(x,y) on them to measure the correlation between them by using Matlab software and the results are shown in Table-2.

Table 2- The values of the measure H and the values of ρ

$H(x_1,x_j)$ j=2,3,4,5	0.4630	0.7371	0.8416	0.1198
ρ	0.4471	0.7480	0.7007	0.1333
$H(x_2,x_j)$ j=3,4,5	0.6523	0.8633	0.7968	
ρ	0.6771	0.7547	0.7799	
$H(x_3,x_j)$ j=4,5	0.7463	0.3019		
ρ	0.7608	0.3019		
$H(x_4,x_5)$	0.8171			
ρ	0.6752			

By looking at the data in the above table one can see that the results of correlations are very close to the proposed measure (H) due to the fact that the two methods are a measure of the relation between the variables. By looking at the values in Table-2 it can be said that the variable X_2 represents the strongest reasons as the employees, non-high school graduates and elderly applicants all rely on the factor of tolerance with the student applying to these universities. As the measure H showed high value of dependence between X_2 and X_3, X_4, X_5. Also it has been noted that the employees have the will to improve their income along with non-high school graduate whom usually academic grades are low dependent on factor X_2.

5. Conclusions

1. The results shows closeness between the values of correlation coefficients calculated [ρ] with the values produces by the new measure which indicates the success of this measure of finding the correlation between the variables.
2. Tolerance with the students in private universities greatly increased applying in these universities as the employees, non-high school graduates and elderly students dependent on this factor to earn the graduation certificates with least possible amount of effort.
3. The factor of the absence of age limit in applications considered the least effective among the other reasons in applying to private universities. This is based on the results of the measure used in the paper, this is due to the fact that the majority of students of these collage are young as they applied after finishing high school or the other institutions.

References

1. Bjerve, S. and Doksum, K. **1993**. Correlation curves: Measures of association as functions of covariate values. *Annuals of Statistics*,21(2), pp: 890-902.
2. Doksum, K., Blyth, S., Bradlow, E., Meng, X. L., and Zhao, H.**1994**. Correlation curves as local measures of variance explainedby regression. *Journal of the American Statistical Association*,89(426),pp: 571-582.
3. Blyth, S. **1994**. Karl Pearson and the correlation curve. *Internat.Statist. Rev*., 62(3),pp 393-403.
4. Blyth, S. **1994**. Measuring local association: an introduction to the correlation curve. *Sociol. Meth*., 24,pp 171-197.
5. Jones, M. C. **1996**. The local dependence function. *Biometrika*, 83(4), pp:899-904.
6. Bairamor, I., Kotz, S. & Kozubowski, T.J. **2003**. A New Measure of Linear Local Dependence, *Statistics*, Vol.37(3), pp: 243-258.
7. Nadarajah, S., Mitov, K., and Kotz, S. **2003**. Local dependence function for extreme value distributions. *Journal of Applied Statistics*, 30(10), pp:1081-1100.
8. Tavangar, M. & Asadi,M. **2008**. On a New Measure of Linear Local Dependence, *JIRSS*, Vol. 7, Nos. 1-2, pp: 35-56
9. Johnson, N.L. & Kotz, S. **1972**. Comparision of Scale Parameter of Beta- stacy Distribution, *Journal of the American Statistical Association*, 67(3387)
10. Johnson, N.L. & Kotz, S. **2000**. *Continuous multivariate, distributions*,2nd Edition, John Wiley & Sons, Canada.