Kettering University
Digital Commons @ Kettering University

Mathematics Presentations And Conference Materials | Mathematics

4-1-2017

Local Dependence Structure of the Bivariate Normal Distribution

Boyan N. Dimitrov
Kettering University, bdimitro@kettering.edu

Kreg Deachin
Kettering University, deac9618@kettering.edu

Follow this and additional works at: https://digitalcommons.kettering.edu/mathematics_conference

Part of the Mathematics Commons

Recommended Citation
Dimitrov, Boyan N. and Deachin, Kreg, "Local Dependence Structure of the Bivariate Normal Distribution" (2017). *Mathematics Presentations And Conference Materials*. 46.
https://digitalcommons.kettering.edu/mathematics_conference/46

This Presentation is brought to you for free and open access by the Mathematics at Digital Commons @ Kettering University. It has been accepted for inclusion in Mathematics Presentations And Conference Materials by an authorized administrator of Digital Commons @ Kettering University. For more information, please contact digitalcommons@kettering.edu.

Local Dependence Structure of the Bivariate Normal Distribution

Boyan Dimitrov and Kreg Deachin

**Kettering University,
Flint, Michigan 48504, USA**

Outline

- **Some previous results on measuring dependence**
- **From random events to random variables**
- **Correlated Bivariate Normal Distribution**
- **Local dependence in correlated Bivariate Normal distribution; Surface of dependence on squares**
- **[*x*, *x*+1]x[*y*,*y*+1], (x,y)∈[-3.3, 3.5]x[-3.5,3.5], and on squares [*x*, *x*+.5]x[*y*,*y*+.5]**
- **Conclusions**

Introduction

- **In several previous publications we developed an idea how probability tools can be used to measure strength of dependence between random events**

- **In the present talk we propose to use it for measuring magnitude of local dependences between random variables.**

- **As illustration, we demonstrate how it works in measuring dependence inside the normally distributed random variables, using the regression coefficients**

- **Short illustrations (graphics and tables) are showing the use of these measures in already known popular Bivariate Normal distribution with different correlation values**

How people INDICATE dependence?

The dependence in uncertainty is a complex concept.

- **In the classical approach conditional probability is used to determine if two events are dependent, or not:** *A* and *B* are independent when the probability for their joint occurrence equals to the product of the probabilities for their individual appearance, i.e. when

$$P(A \cap B) = P(A).P(B)$$

- **Otherwise, the two events are dependent.**

How to measure dependence?

- **To measure dependence between random events**

B. Dimitrov (2010, Some Obreshkov Measures of Dependence and Their Use, *Compte Rendus de l'Acad. Bulgare des Sci., v.* ***63, No.1, pp. 15-18)***

revived some measures of dependence for random events based on notion of probabilities of the events.

From that discussion and among the four proposed measures we selected the Regression coefficients as suitable measure of magnitude of dependence when the two events are dependent.

- **Some discussion and examples have been presented at the last year 2016 Kingsville meeting.**

Regression Coefficients as Measures of dependence between random events

- **Definition 1.** ***Regression coefficient*** of the event ***A*** with respect to the event ***B*** is called 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

$$R_B(A) = P(A \mid B) - P(A \mid \overline{B})$$

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

- The regression coefficient is always defined, for any pair of events ***A*** and ***B*** (zero, sure, arbitrary).

- The regression coefficient of ***B*** with respect to the event ***A is defined symmetrically***

Properties of Regression coefficients

(r1) The equality to zero $R_B(A) = R_A(B) = 0$ holds only if the events ***A*** and ***B*** are independent.

(r2) $R_A(A) = 1$; $R_{\overline{A}}(A) = -1.$

(r3) $$R_B(\sum A_j) = \sum R_B(A_j)$$

(r4) $$R_S(A) = R_{\text{ø}}(A) = 0$$

(r5) The regression coefficients are numbers with equal signs

To be valid $R_B(A) = R_A(B)$ it is necessary and sufficient to have

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

Regression coefficients measure the strength of dependence between random events.

- **The relations**

$$R_B(A) = \frac{P(A \cap B) - P(A)P(B)}{P(B)[1 - P(B)]}$$

- **and**

$$R_A(B) = \frac{P(A \cap B) - P(A)P(B)}{P(A)[1 - P(A)]}$$

- **explain when it will be $R_B(A) = R_A(B)$.**

Regression coefficients - properties

(r6) The regression coefficients and are numbers between –1 and 1, i.e. they satisfy the inequalities

$$-1 \le R_B(A) \le 1 \qquad -1 \le R_A(B) \le 1$$

(r6.1) The equality $R_B(A) = 1$ holds only when ***A* coincides** (is equivalent) **with the event *B***.

Then is also valid the equality $R_A(B) = 1$;

(r6.2) The equality $R_B(A) = -1$ holds only when event ***A* coincides** (or is equivalent) with the event $\overline{B}$ - the **complement of *B*.**

Then is also valid $R_A(B) = -1$, and respectively $\overline{A} = B$.

Regression coefficients – a proposition to classify the strength of dependence

- **In our opinion, it is possible one event to have stronger dependence on the other than the reverse.**

- **This measure suits for measuring the magnitude of dependence between events.**

- **The distance of the regression coefficient from the zero (where the independence is) could be used to classify the strength of dependence, e,g.**

  - almost independent (when $R_A(B) < .05$) ;
  - weakly dependent (when $.05 < |R_A(B)| < .2$) ;
  - moderately dependent (when $.2 < |R_A(B)| < .45$) ;
  - in average dependent (when $.45 < |R_A(B)| < .8$) ;
  - strongly dependent (when $|R_A(B)| > .8$) ;

Predictions using Regression coefficients

- One serious advantage of the *Regression coefficients is to use this* magnitude of dependence to evaluate the posterior distribution of one event when information of the other event occurred is available. We have

$$P(A \mid B) = P(A) + R_B(A)[1-P(B)].$$

- This formula competes with the BAYES RULE, that requires joint probability $P(A \cap B)$. We offer to use the strength of dependence $R_B(A)$ instead.

One interesting fact

If we introduce

- **Definition 3.** Correlation coefficient between two events *A* and *B* is defined by the number

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

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

- Then an equivalent representation will be

$$R_{A,B} = \frac{P(A\cap B) - P(A)P(B)}{\sqrt{P(A)P(\bar{A})}\sqrt{P(B)P(\bar{B})}}$$

Which corresponds to the Pierson-Brave Contingency coefficient $\boldsymbol{\varphi}$.

From Events to Random Variables

- The introduced measures allow **to see the interaction between** any pair of numeric **r.v.'s (*X,Y*) throughout the sample space**

- **Understand and use the local dependence.**

- **Let $F(x,y)=P(X \le x, Y \le y)$ - the joint c.d.f.**

- **Marginals $F(x) = P(X \le x)$, $G(y)=P(Y \le y)$.**

From Events to Random Variables

- Introduce the events

- $A_x = \{x \leq X \leq x + \Delta_1 x\}$; $B_y = \{y \leq Y \leq y + \Delta_2 y\}$,

for any $x, y \in (-\infty, \infty)$.

- **Then the measures of dependence between events *A* and *B* turn into *a measure of local dependence between the pair of r.v.'s X and Y on the rectangle***

$$D=[x, x + \Delta_1 x]\times[y, y + \Delta_2 y].$$

From Events to Random Variables

- Naturally, they can be named and calculated as follows:
- **Regression coefficient** of *X* with respect to *Y*, and of *Y* with respect to *X* on the rectangle D= [x, $x+\Delta_1 x$]×[y, $y+\Delta_2 y$]. By Definition 1 we get

$R_Y((X,Y) \in D)$=

$$\frac{\Delta_D F(x,y) - [F(x+\Delta_1 x) - F(x)][G(y+\Delta_2 y) - G(y)]}{[F(x+\Delta_1 x) - F(x)]\{1 - [F(x+\Delta_1 x) - F(x)]\}}$$

- Here by **$\Delta_D F(x,y)$** is denoted the **two dimensional finite difference** for the function $F(x,y)$ on rectangle ***D=[x, x+Δ_1x]×[y, y+Δ_2y].***

From Events to Random Variables

- Namely

$$\Delta_D F(x,y) = F(x+\Delta_1 x, y+\Delta_2 y) - F(x+\Delta_1 x, y) -$$
$$- F(x, y+\Delta_2 y) + F(x, y).$$

- In an analogous way is defined $R_X((X,Y)$ **в** $D)$**.** Just denominator in the expression is changed respectively.

- **Correlation coefficient** $\rho_Y((X,Y)$ **в** $D)$ between the r.v.'s X **and** Y **on rectangle** $D=[x, x+\Delta_1 x]\times[y, y+\Delta_2 y]$ can be presented in similar way by the **use of Definition 2**. We omit detailed expressions as something obvious.

Regression coefficients

- **The biggest advantage of the *Regression coefficients* as measures of the magnitude of dependence is their easy interpretation, described above, and the fact that they come available from the knowledge of the probabilities of the respective events, or proportional number of individuals in the sets of subpopulations of interests.**

- **In Probability modeling which use multivariate distribution we see GREAT Advantages when knowing one component within an interval, to predict everything that may happen with the other component.**

- **Next we illustrate specific rules in calculation of Regression Coefficients as measures of dependence to analyze the local dependence structure in Bivariate Normal distribution.**

Correlated Bivariate Normal Distribution

- The pair (***X,Y***) has pdf

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\nu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\nu_1}{\sigma_1}\right)\left(\frac{y-\nu_2}{\sigma_2}\right)+\left(\frac{y-\nu_2}{\sigma_2}\right)^2\right]}$$

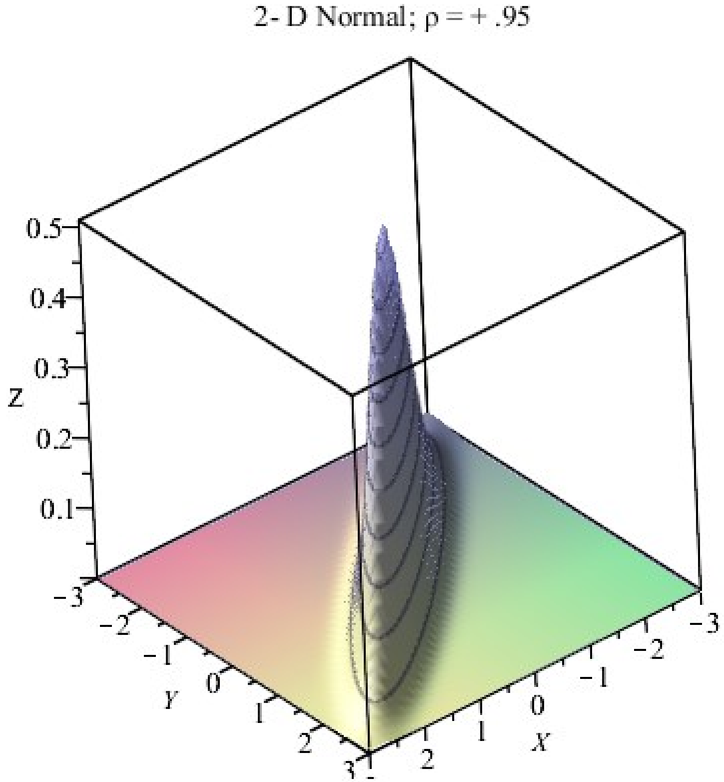
- **Here $\rho_{X,Y}$ is the Correlation coefficient**

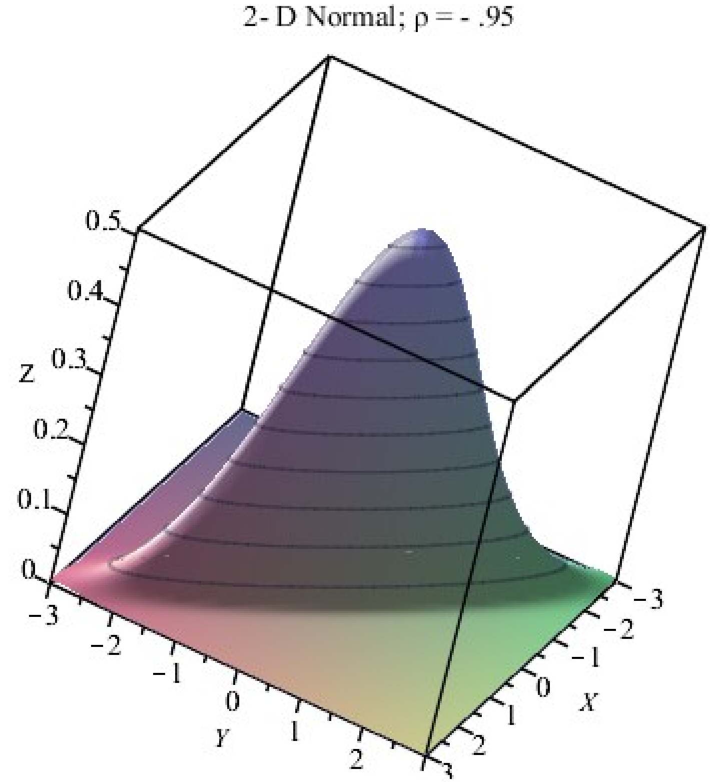
- **The marginal $F_X(x)$ and $G_Y(y)$ are normal distributions.**

- ***We use standard normal marginals, and correlated components with different numeric values of the correlation coefficient*** **$\rho_{X,Y}$** in our illustrations.

-

Correlated Bivariate Normal Distribution Graphs: ρ = + - .95





Correlated Bivariate Normal –Local Dependence Functions

- The functions

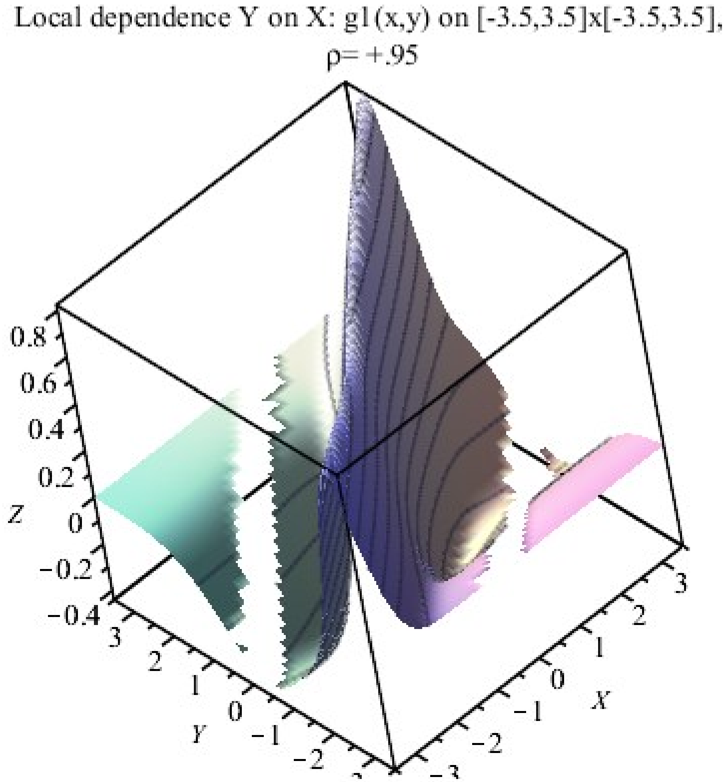
$$g1(x,y)=\frac{\int_x^{x+1}\int_y^{y+1} e^{-\frac{1}{2(1-\rho^2)}\left[u^2-2\rho uv+v^2\right]}du\ dv-\sqrt{1-\rho^2}\int_x^{x+1}e^{-u^2/2}du.\int_y^{y+1}e^{-v^2/2}dv}{\sqrt{2\pi}\sqrt{1-\rho^2}\int_x^{x+1}e^{-u^2/2}du\left(1-(1/\sqrt{2\pi})\int_x^{x+1}e^{-u^2/2}du\right)}$$

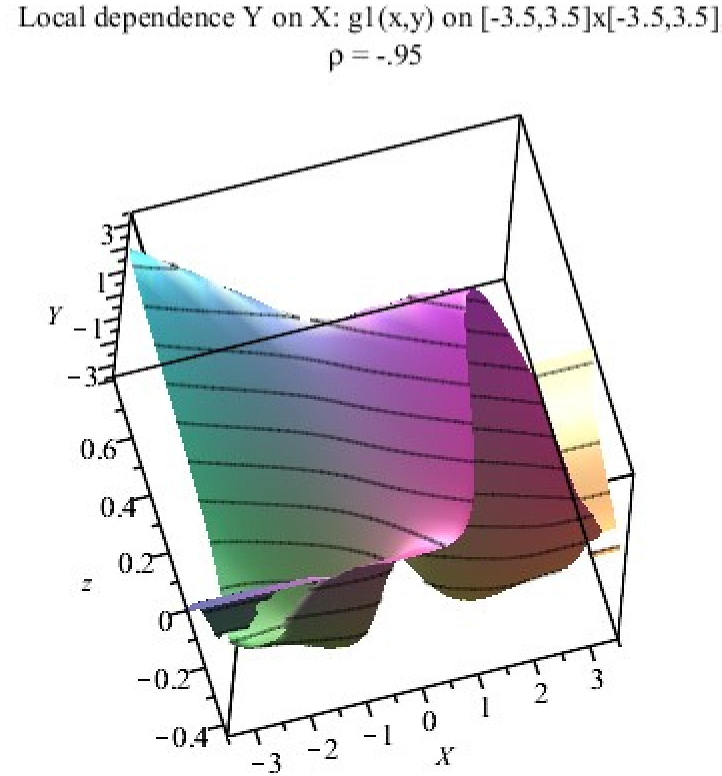
- and

$$g2(x,y)=\frac{\int_x^{x+1}\int_y^{y+1} e^{-\frac{1}{2(1-\rho^2)}\left[u^2-2\rho uv+v^2\right]}du\ dv-\sqrt{1-\rho^2}\int_x^{x+1}e^{-u^2/2}du.\int_y^{y+1}e^{-v^2/2}dv}{\sqrt{2\pi}\sqrt{1-\rho^2}\int_y^{y+1}e^{-u^2/2}du\left(1-(1/\sqrt{2\pi})\int_y^{y+1}e^{-u^2/2}du\right)}$$

represent the local dependence between correlated components Y with respect to X in g1(x,y) , and of X with respect to Y in g2(x,y) on the unit square [x, x+1]x[y, y+1] located at the point (x,y).

Correlated Bivariate Normal –Local Dependence Function g1(x,y): ρ = + - .95





Correlated Bivariate Normal –Local Dependence Function g1(x,y): Numeric values on integer points in the square [-3,3]x[-3,3]

X \ Y	-3	-2	-1	0	1	2	3
-3	.6531	.1818	-.3884		-.1389	-.0219	-.0013
-2	.0324	.6289	-.1354	-.3948		-.0248	-.0015
-1	-.0325	-0707	.6396	-.2937	-.2062		-0020
0		-.2062	-.2937	.6396	-.0707	-.0325	
1	-.0248		-.3948	-.1354	‘6829	.0324	-.0015
2	-.0219	-.1389		-.3484	.1818	.6531	.0244
3	-.0214	-.1361	-.3418		-.1353	.4884	.5773

Correlated Bivariate Normal –Local Dependence Function g1(x,y): Numeric values on integer points in the square [-3,3]x[-3,3]

The calculations in the missed cells (all made with MAPLE) were non-numeric:

> $evalf(g1(-1,-1));$

$$0.6396420599$$

> $evalf(g1(1,1));$

$$0.6829419756$$

> $evalf(g1(0,-3))$

$$2.267077886\left(\int_{0.}^{1.}\left(0.3913472140\,\mathrm{erf}(6.793662205+2.151326365\,u)\,\mathrm{e}^{-0.5000000002\,u^2}-0.3913472140\,\mathrm{erf}(4.529108136+2.151326365\,u)\,\mathrm{e}^{-0.5000000002\,u^2}\right)\mathrm{d}u\right)-0.03249079578$$

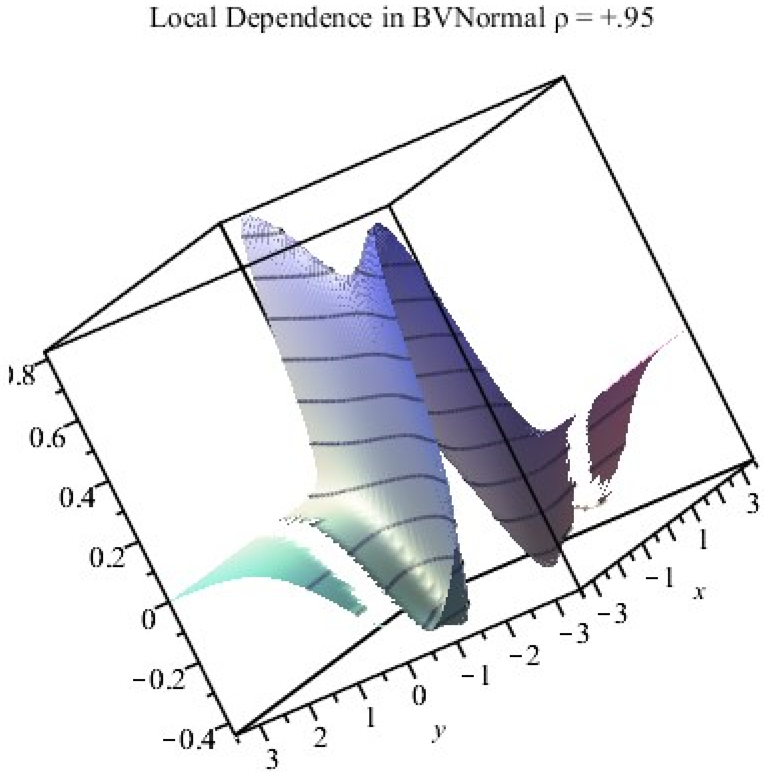
> $evalf(g1(1,-2))$

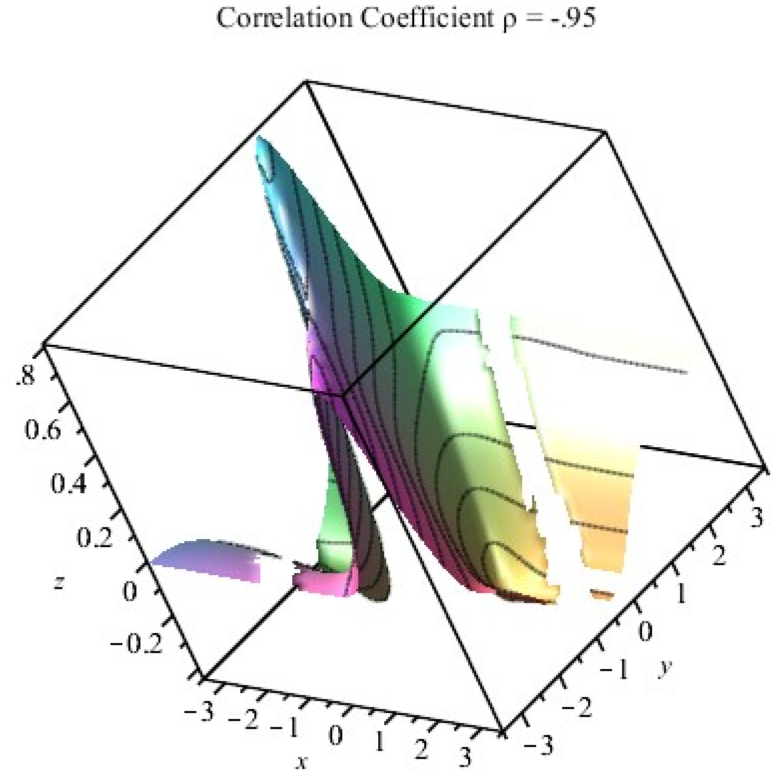
$$4.340308176\left(\int_{1.}^{2.}\left(0.3913472140\,\mathrm{erf}(4.529108136+2.151326365\,u)\,\mathrm{e}^{-0.5000000002\,u^2}-0.3913472140\,\mathrm{erf}(2.264554068+2.151326365\,u)\,\mathrm{e}^{-0.5000000002\,u^2}\right)\mathrm{d}u\right)-0.1572803240$$

> $evalf(g1(3,2))$

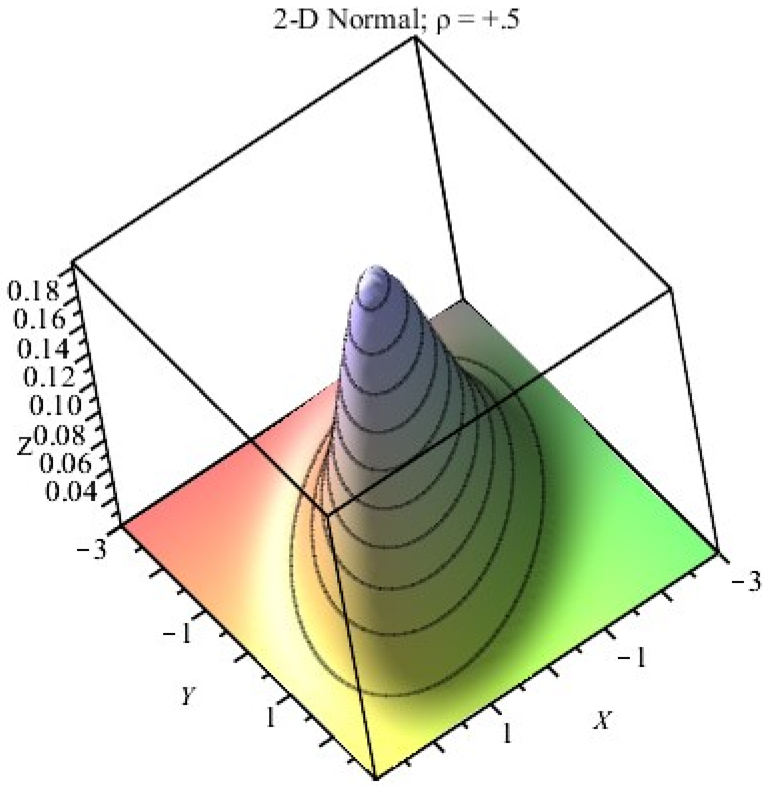
$$0.3884267371$$

Correlated Bivariate Normal –Local Dependence Function g2(x,y): ρ = + - .95

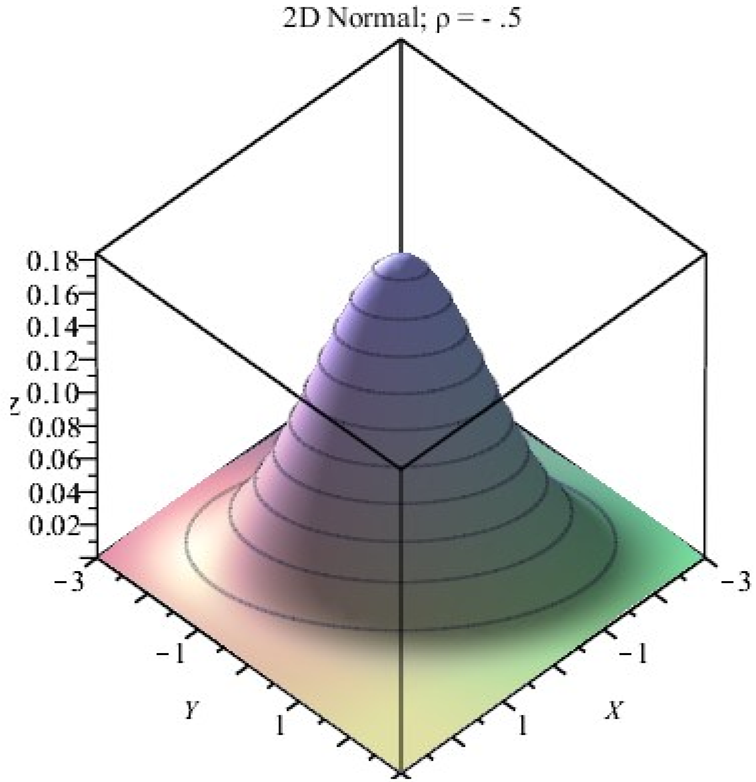




Correlated Bivariate Normal density: ρ = + - .5

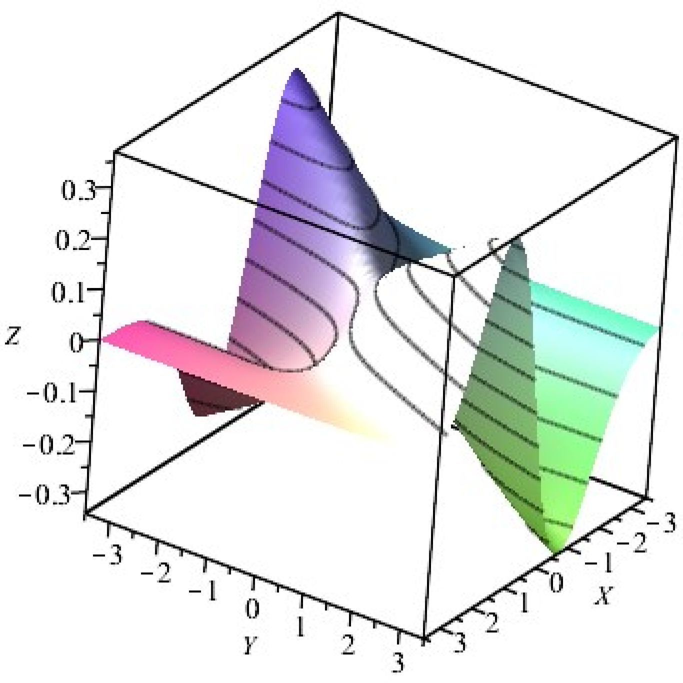


±

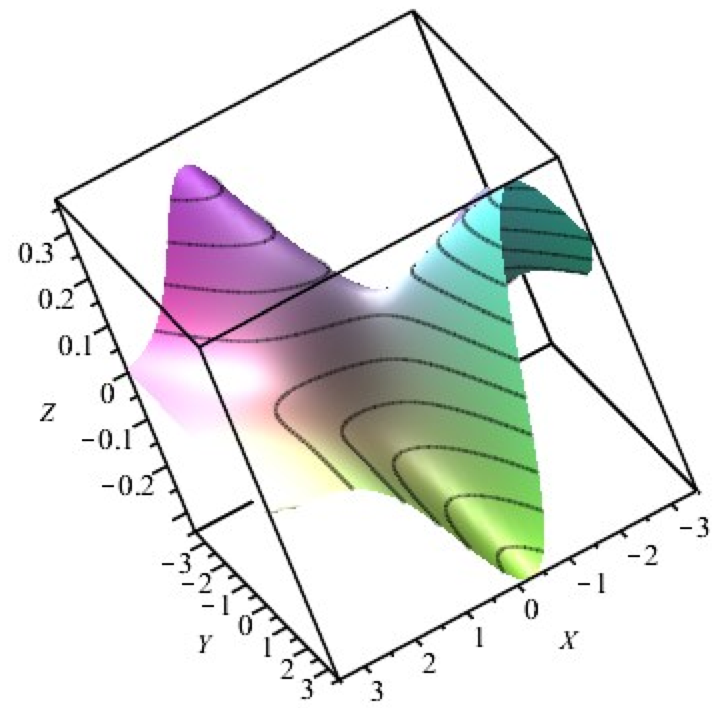


Correlated Bivariate Normal –Local Dependence Function g2(x,y): ρ = + - .5

Local dependence X on Y: g2(x,y) on [-3.5,3.5]x[-3.5,3.5],
ρ= +.5

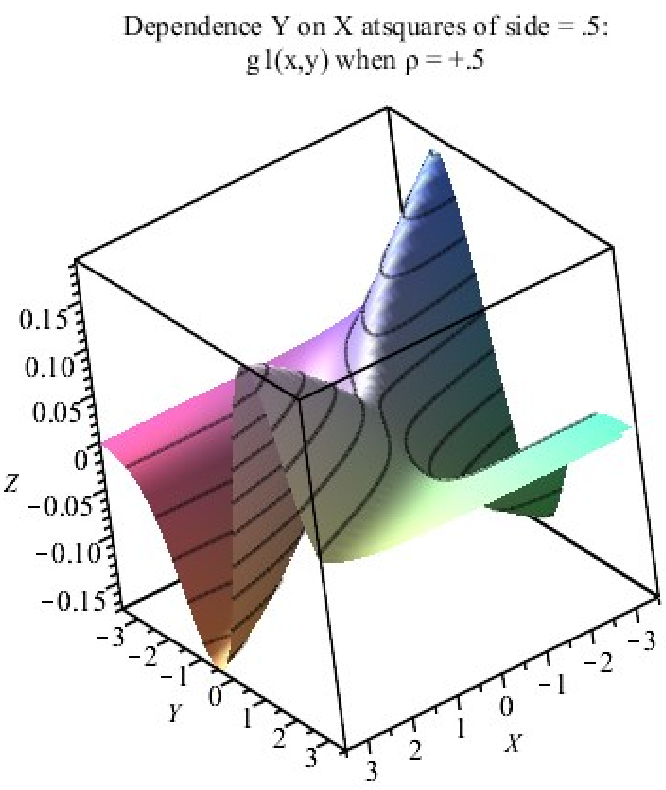


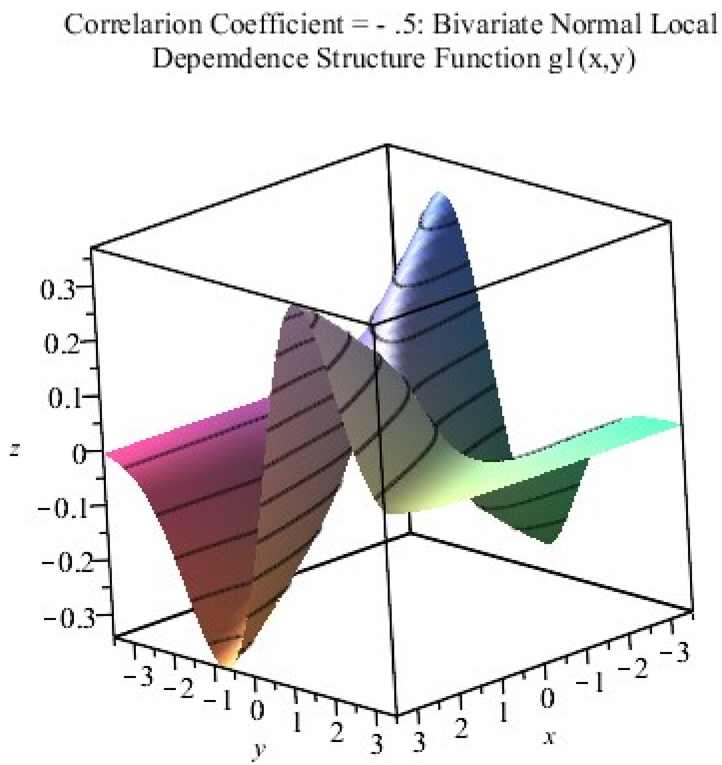
Local dependence X on Y: g2(x,y) on [-3.5,3.5]x[-3.5,3.5],
ρ= -.5



Correlated Bivariate Normal –Local Dependence Function g1(x,y): ρ = + .5

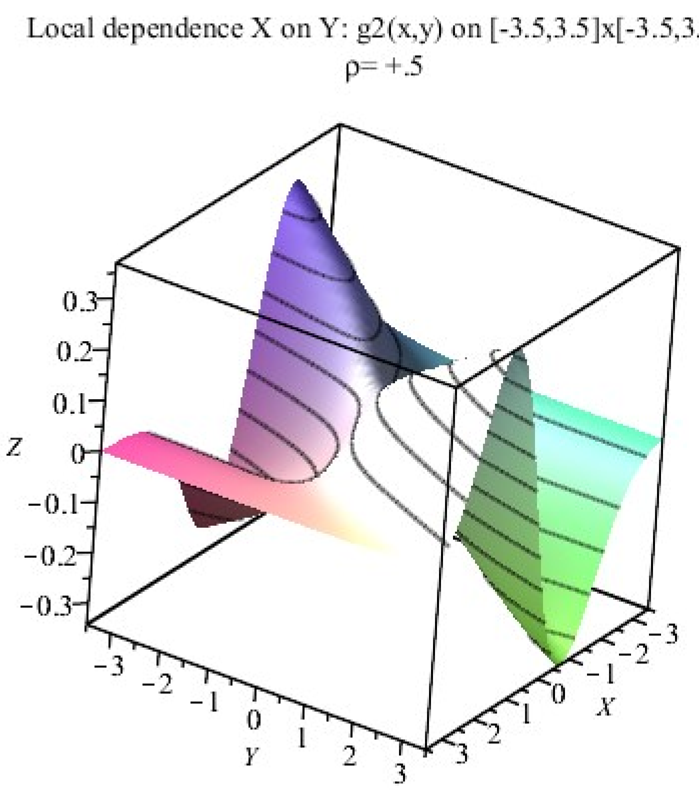
- Dependence as function of size of square size [.5 vs 1].

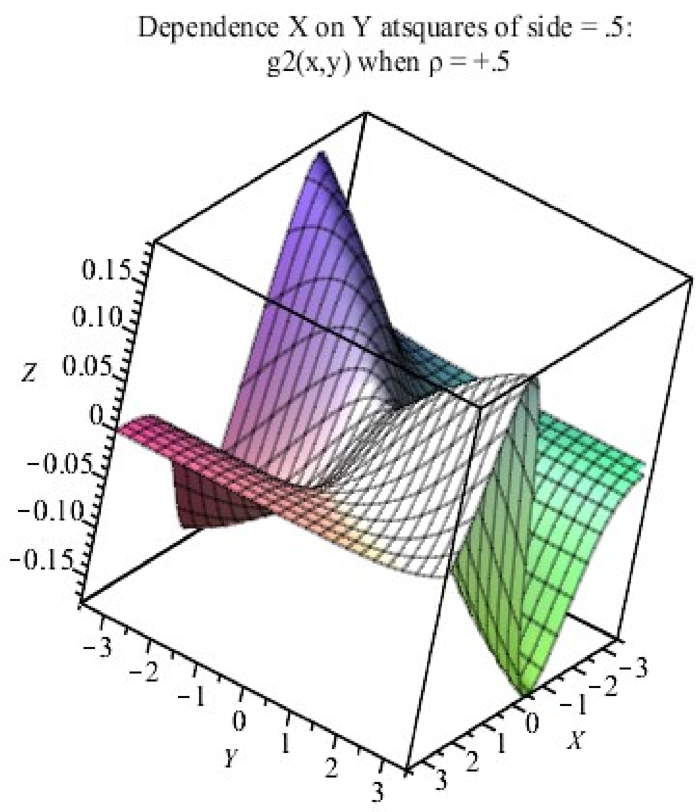




Correlated Bivariate Normal –Local Dependence Function g2(x,y): ρ = - .5

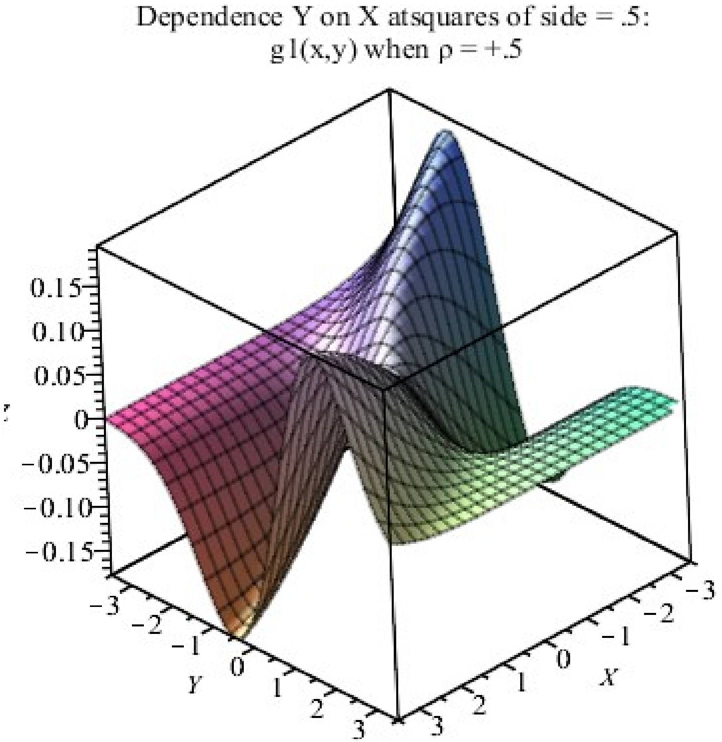
Dependence as function of size of square size [1 vs .5].

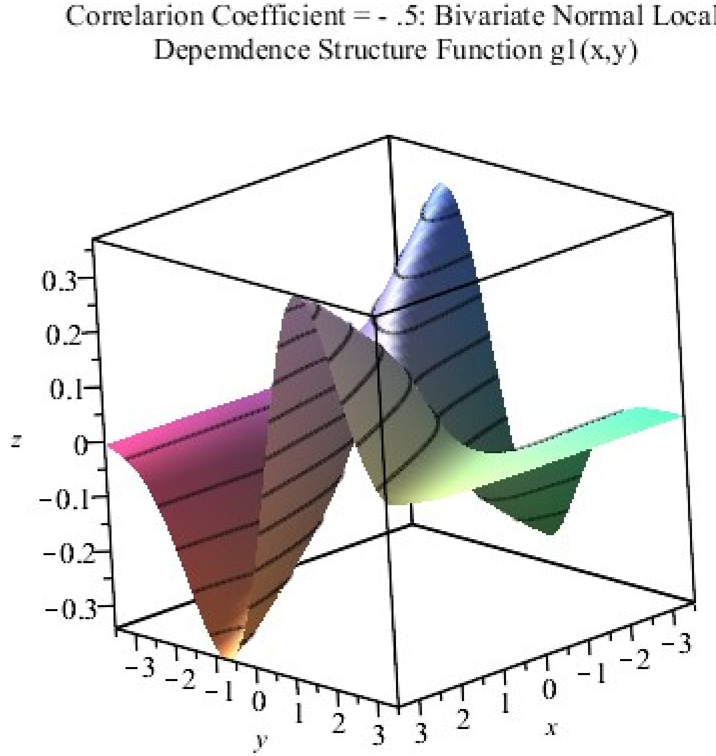




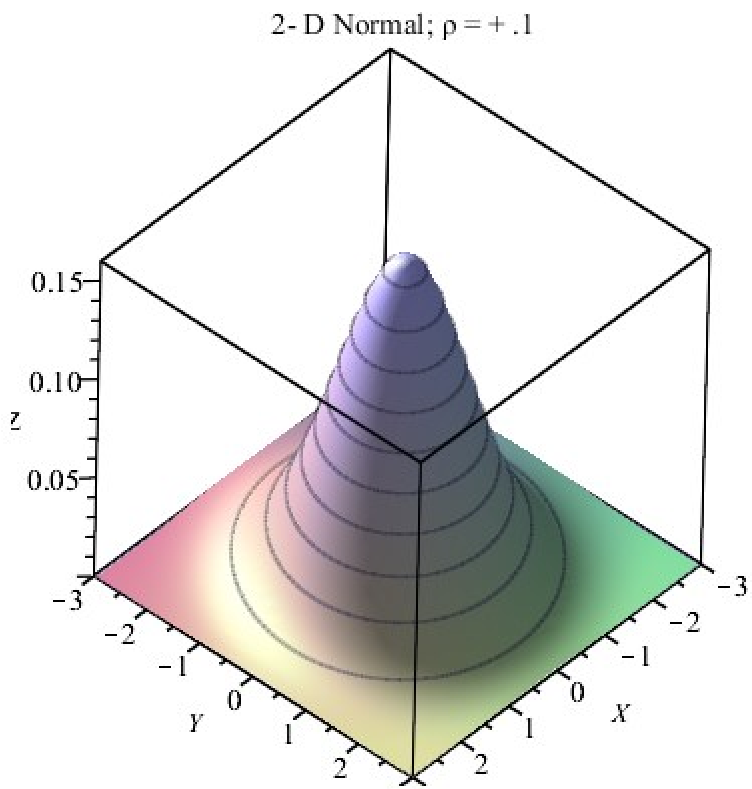
Correlated Bivariate Normal –Local Dependence Function g1(x,y): ρ = + - .5

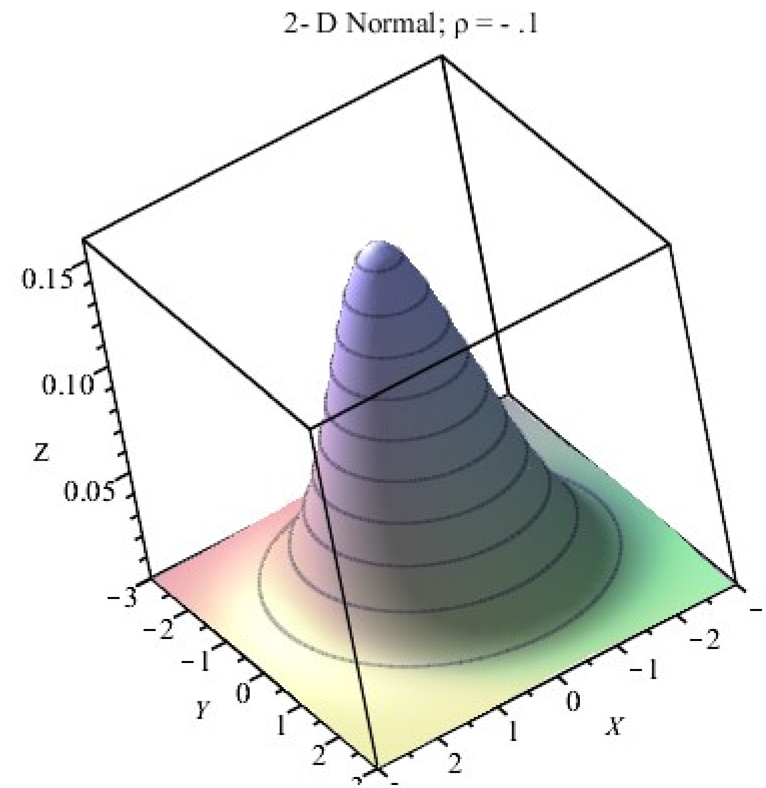
- Dependence as function of size of square size [.5 vs 1].



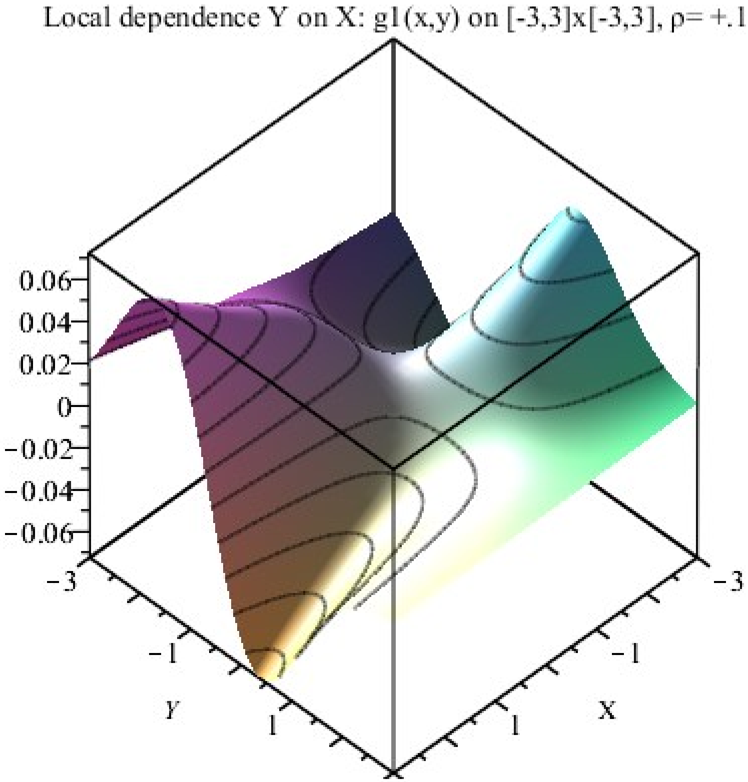


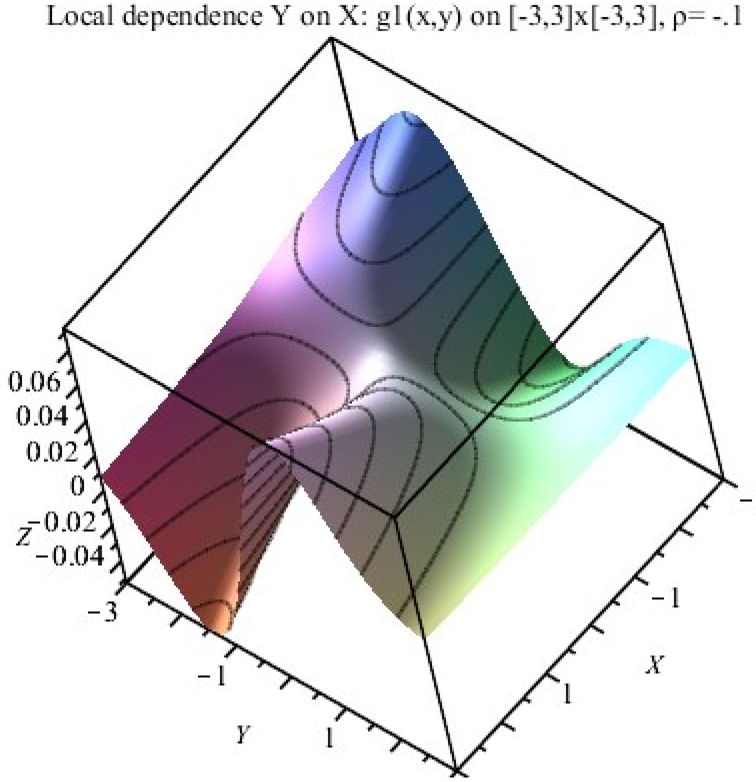
Correlated Bivariate Normal density: ρ = + - .1



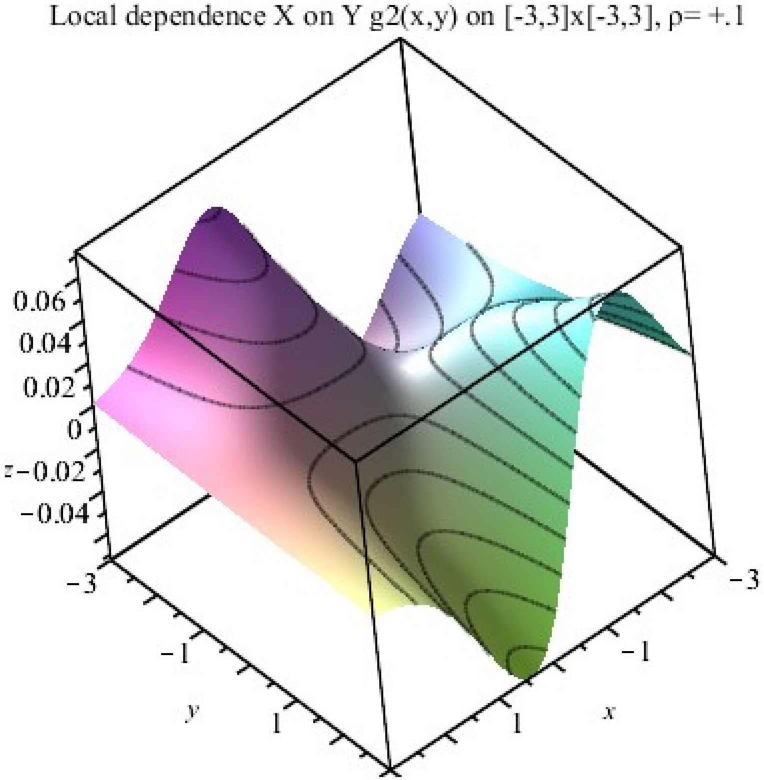


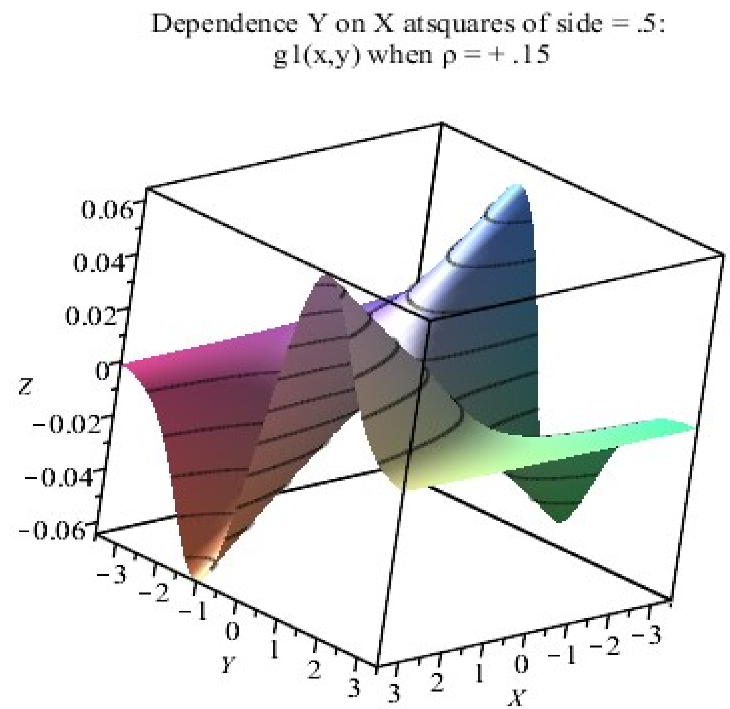
Correlated Bivariate Normal –Local Dependence Function g1(x,y): ρ = + - .1



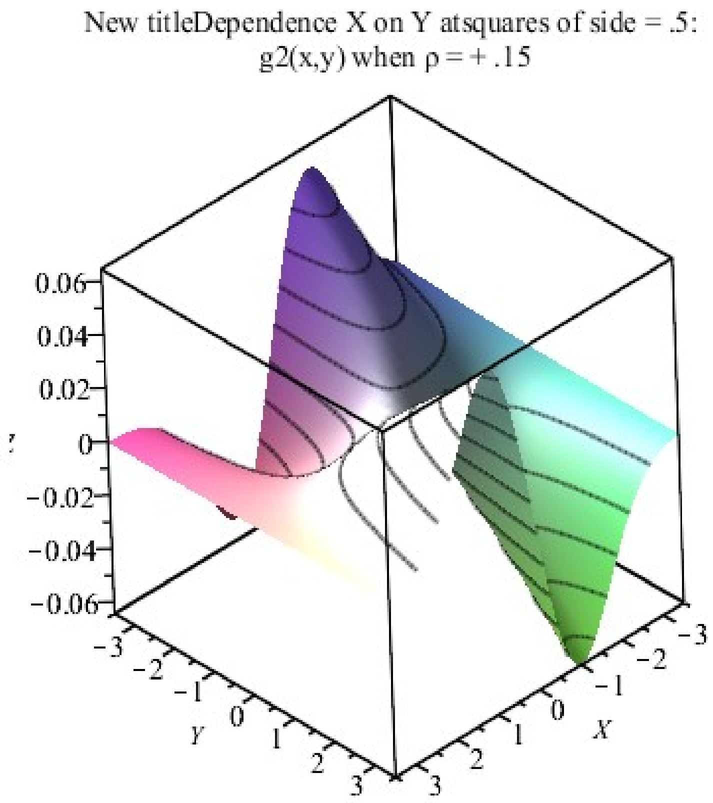


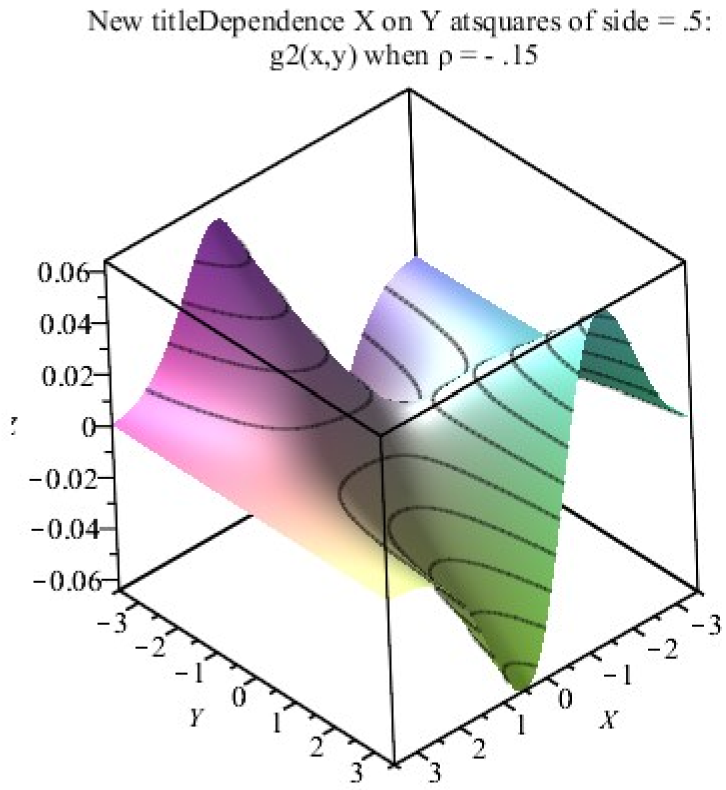
Correlated Bivariate Normal –Local Dependence Function g2(x,y): ρ = + - .1





Correlated Bivariate Normal Distribution





CONCLUSIONS

- We discussed Regression coefficients as measures of dependence between two random events.
- These measures are asymmetric, and exhibit natural properties.
- Their numerical values serve as indication for the magnitude of dependence between random events.
- These measures provide simple ways to detect independence, coincidence, degree of dependence.
- If either measure of dependence is known, it allows better prediction of the chance for occurrence of one event, given that the other one occurs.

CONCLUSIONS

- We discussed Regression coefficients as measures of dependence between the two components of BVND.
- These measures are examined by the 3d surface of dependence on squares [*x*, *x*+*a*]x[*y*,*y*+*a*] with *a*=.5; 1.0 and $(x,y) \in [-3.5, 3.5] \times [-3.5, 3.5]$
- We observe high positive local dependence close to the line *y*=*x*, and negative local dependences, also of relatively high magnitude, about the opposite signs *y*= - *x*. This magnitude vanishes as long the points become far from the origin (0,0).
- Notice reduction of magnitude in half on smaller square.
- **The ancient Greeks used to say: Just seat, watch, and make your own conclusions.**

CONCLUSIONS

- We extend the REGRESSION COEFFICIENTS measures from events to local dependence between random variables
- Our study of the local dependence IN THE BIVARIATE NORMAL distribution WITH CORRELATED COMPONENTS on squares inside [-3.5, 3.5]x[-3.5, 3.5] finds different behavior completely different than the global dependence.
- Graphical illustrations show things outside our expectations.
- **Local dependence can be essentially different on different regions in the field.**

References

- Agresti A. ***Categorical Data Analysis***. John Wiley & Sons, Hew York, 2006
- B. Dimitrov, Let start measure uncertain dependence in the classroom, MASAL Conference March 2016, Saginaw
- B. Dimitrov. **Dependence Structure of some bivariate Distributions,** Serdica J. Computing, **8**, No. 2, pp 233-254 (2014).
- B. Dimitrov. **Measures of dependence in reliability,** Proceedings of the **8-th MMR'2013**, Stellenbosh, South Africa, July 1- 4 , pp 65-69 (2013).
- B. Dimitrov. **Some Obreshkov Measures of Dependence and Their Use,** Compte Rendus de l'Academie Bulgare des Sciences, V. 63, No.1, pp. 15-18 (2010).
- S. Esa and B. Dimitrov. **Dependencies in the world of politics,** Proceedings of the **8-th MMR'2013**, Stellenbosh, South Africa, July 1-4 , pp 70 – 73 (2013).
- Obreshkov N. ***Probability Theory.*** Nauka i Izkustvo, Sofia, 1963 (In Bulgarian).