

A Copula Based Test for a Two Component Bivariate Mixture Distribution

Ampalavanar Nanthakumar

Department of Mathematics, State University of New York at Oswego, Oswego NY, USA

Email address:

ampala.nanthakumar@oswego.edu

To cite this article:

Ampalavanar Nanthakumar. A Copula Based Test for a Two Component Bivariate Mixture Distribution. *International Journal of Statistical Distributions and Applications*. Vol. 3, No. 4, 2017, pp. 61-66. doi: 10.11648/j.ijstd.20170304.11

Received: August 28, 2017; **Accepted:** September 13, 2017; **Published:** November 1, 2017

Abstract: The paper presents a Copula based approach to test for a two component bivariate mixture distribution. The regular joint density is modeled by using the Copula and then the Locally Most Powerful test (LMP) test is derived by using this Copula based regular density. This is a fairly simple test compared to the dip / depth test developed by Hartigan. Our simulation results show that this Copula based (LMP) test is very powerful in detecting a mixture.

Keywords: Bivariate, Mixture, Copula

1. Introduction

Mixtures of multivariate distributions are a recurring phenomenon in many fields and the need for a reasonably good hypothesis test has been felt by the scientific community for a long time. Wolfe (1970) used the likelihood ratio to test for a two-component normal mixture against a normal null hypothesis. The test is very sensitive to the normality assumptions. Another test developed by Engelman and Hartigan (1969) will not work when the bimodal alternative is not a normal mixture. The test developed by Giacomelli et al (1971) needs the modes specified in advance. Another test, 'Dip Test of Unimodality' developed by Hartigan and Hartigan (1985) requires a fairly large sample for this test to work. In a way, the presence of multiple modes is an indication that a data follows a mixture of distributions. But, the irony is that a mixture of distributions may not possess multiple modes. Moreover, the maximum likelihood estimate for the density may not exist in the neighborhood of the mode. The situation becomes trickier in the case of multivariate distributions. So, when the "Dip" test fails to detect multiple modes that does not necessarily mean that the underlying distribution is not a mixture of distributions. On the other hand, the copulas are used as a general way of formulating a multivariate distribution in such a way that the dependence can be infused in a reasonable manner. This is based on a simple idea that the joint distribution can be represented as a transformation of the

underlying marginal distributions (see Sklar 1959). There are several copulas and each of these differ according to the strength of the dependence and the direction of the association. For the literature review, the interested readers are referred to Nelson (2006). Here in this paper, we consider two particular copulas; the Gaussian Copula and the Clayton Copula. The Clayton Copula belongs to the family of Archimedean Copulas whereas the Gaussian Copula does not belong to the Archimedean family. The Copulas are used in the construction of the regular joint density and then the Locally Most Powerful test (see Rao (1973)) is derived by using this copula joint density. Based on these Copula densities, a locally most powerful (LMP) test is developed in place of the "Dip" test to test for a possible mixture of multivariate (in this case, bivariate) distributions. Here each component of the mixture is a bivariate distribution. In this paper, two different types of copulas are studied namely Gaussian and Clayton copulas. This LMP test is seen to be very powerful in detecting whether a distribution is a mixture or not in this bivariate situation.

We divide the paper into several sections. The first two sections are devoted for the introduction and the third section contains the methodology. The fourth section is for the numerical results and the last section has the conclusion.

2. Methodology

2.1. Copulas

In this section, we present the results pertaining to the copulas.

2.1.1. Definition

A copula is a multivariate joint distribution defined on the k dimensional unit cube $[0,1]^k$ such that every marginal distribution is uniform on the interval $[0,1]$.

In other words, $C: [0,1]^k \rightarrow [0,1]$ is a k -dimensional copula if

- (a). $C(u) = 0$ whenever $u \in [0,1]^k$ has at least one component equal to 0.
- (b). $C(u) = u_i$ whenever $u \in [0,1]^k$ has all the components equal to 1 except the i^{th} one which is equal to u_i .
- (c). $C(u)$ is k -increasing.

2.1.2. Copula Construction

Here, we discuss the construction of the Copulas. There are primarily two major types of Copulas; Archimedean Copula and the non-Archimedean Copulas.

Archimedean Copula:

This is a family of copulas and the k -dimensional Archimedean Copula is defined as follows.

$$C(u_1, u_2, \dots, u_k) = \Psi^{-1} \left(\sum_{i=1}^k \Psi(u_i) \right) \quad (1)$$

where Ψ is known as the generator function and u_i is the marginal distribution of the i^{th} component.

Any generator function which satisfies the following properties is the basis for a copula.

$\Psi(1) = 0$, limit $\Psi(x) = \infty$ as $x \rightarrow 0$, and $\Psi'(x) < 0$, $\Psi''(x) > 0$

2.2. Special Cases

2.2.1. Clayton Copula

Let the generator function $\Psi(u) = u^{-\alpha} - 1$ where $\alpha > 1$. One can show that the functional inverse, $\Psi^{-1}(u) = (1+u)^{-1/\alpha}$.

In the bivariate case, the Archimedean formulation yields the Clayton Copula as

$$C(u_1, u_2) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha} \quad (2)$$

2.2.2. Gaussian Copula

$$C(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \quad (3)$$

where $\Phi_2(x, y)$ represents the bivariate normal cdf and

$\Phi^{-1}(x)$ represents the inverse of the normal cdf.

Theorem #1 (Sklar 1959):

The copula $C(u_1, u_2, \dots, u_k) = F(x_1, x_2, \dots, x_k)$ where $F(x_1, x_2, \dots, x_k)$ is the joint distribution function and $u_i = F_i(x_i)$ is the marginal distribution function.

Note: The copula density $c(u_1, u_2, \dots, u_k)$ is defined as

$$c(u_1, u_2, \dots, u_k) = \frac{\partial^k C(u_1, u_2, \dots, u_k)}{\partial u_1 \partial u_2 \dots \partial u_k} \quad (4)$$

Result:

$$c(u_1, u_2, \dots, u_k) = \frac{f(x_1, x_2, \dots, x_k)}{f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_k(x_k)} \quad (5)$$

where $f(x_1, x_2, \dots, x_k)$ is the joint density function and $f_i(x_i)$ is the marginal density function.

Note: In this paper, the hypothesis test is derived by using the regular density function $f(x_i, y_i)$ for the two component mixture situation.

3. Methodology

As we know, the two component situation arises in the context of many situations. In educational data, in econometrical data, in astrophysical data, the two component situation is very common.

Suppose that we have a bivariate two component mixture with the component densities given by f_1 and f_2 respectively for the components with p as the mixing proportion. Then, the mixture density function is given by

$$f(x_i, y_i) = (1-p) \cdot f_1(x_i, y_i) + p \cdot f_2(x_i, y_i) \quad (6)$$

and the likelihood function is

$$L = \prod_{i=1}^n f(x_i, y_i) \quad (7)$$

By taking the natural logarithm, we have

$$\ln L = \sum_{i=1}^n \ln f(x_i, y_i) \quad (8)$$

Next, taking the partial derivative with respect to p and then evaluating the derivative at $p = 0$ yields,

$$\text{At } p = 0, \quad \frac{\partial \ln L}{\partial p} = \sum_{i=1}^n \left(\frac{f_2(x_i, y_i)}{f_1(x_i, y_i)} - 1 \right) \quad (9)$$

This means that in order to test the hypothesis,

$H_0: p = 0$ versus $H_1: p > 0$

The LMP test should be: Reject H_0 when $\overline{W}_n > C$, where C is a constant and

$$\overline{W}_n = \frac{\sum_{i=1}^n \frac{f_2(x_i, y_i)}{f_1(x_i, y_i)}}{n} \quad (10)$$

As we can see from (3.5), the test-statistic involves the ratio of the joint density functions. Here, we will use the Clayton Copula to construct the joint density function from the marginal distributions.

3.1. Joint Density Construction

Clayton Copula:

According to (3.1), the mixture distribution is given by

$$F(x, y) = (1-p) \cdot F_1(x, y) + p \cdot F_2(x, y) \quad (11)$$

and by using the Clayton Copula, one can write

$$F(x, y) = (1-p) \cdot \left\{ [F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right\}^{\frac{-1}{\alpha_1}} + p \cdot \left\{ [F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right\}^{\frac{-1}{\alpha_2}} \quad (12)$$

So, the mixture joint density, $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$ is given by

$$\begin{aligned} f(x, y) = & (1-p) \cdot (1+\alpha_1) \cdot \left\{ [F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right\}^{\frac{-1}{\alpha_1}-2} \cdot [F_1(x)]^{-\alpha_1-1} \cdot [G_1(y)]^{-\alpha_1-1} \cdot f_1(x) \cdot g_1(y) \\ & + p \cdot (1+\alpha_2) \cdot \left\{ [F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right\}^{\frac{-1}{\alpha_2}-2} \cdot [F_2(x)]^{-\alpha_2-1} \cdot [G_2(y)]^{-\alpha_2-1} \cdot f_2(x) \cdot g_2(y) \end{aligned} \quad (13)$$

So, we have,

$$\frac{f_2(x_i, y_i)}{f_1(x_i, y_i)} = \frac{(1+\alpha_2)}{(1+\alpha_1)} \cdot \frac{\left\{ [F_1(x_i)]^{-\alpha_1} + [G_1(y_i)]^{-\alpha_1} - 1 \right\}^{\frac{1}{\alpha_1}+2}}{\left\{ [F_2(x_i)]^{-\alpha_2} + [G_2(y_i)]^{-\alpha_2} - 1 \right\}^{\frac{1}{\alpha_2}+2}} \cdot \frac{\{F_1(x_i) \cdot G_1(y_i)\}^{\alpha_1+1}}{\{F_2(x_i) \cdot G_2(y_i)\}^{\alpha_2+1}} \cdot \frac{f_2(x_i)}{f_1(x_i)} \cdot \frac{g_2(y_i)}{g_1(y_i)} \quad (14)$$

$$\text{Let } h(x, y) = \frac{f_2(x, y)}{f_1(x, y)} \quad (15)$$

Note that the partial derivative with respect to x ,

$$\begin{aligned} \frac{\partial h}{\partial x} = & \frac{(1+\alpha_2)}{(1+\alpha_1)} \cdot \frac{\left\{ [F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right\}^{\frac{1}{\alpha_1}+2}}{\left\{ [F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right\}^{\frac{1}{\alpha_2}+2}} \cdot \frac{[F_1(x)]^{\alpha_1+1}}{[F_2(x)]^{\alpha_2+1}} \cdot \frac{[G_1(y)]^{\alpha_1+1}}{[G_2(y)]^{\alpha_2+1}} \\ & \left\{ f_2'(x) \cdot [f_1(x)]^{-1} \cdot g_2(y) \cdot [g_1(y)]^{-1} - f_2(x) \cdot [f_1(x)]^{-2} \cdot f_1'(x) \cdot g_2(y) \cdot [g_1(y)]^{-1} \right\} \\ & + \frac{(1+\alpha_2)}{(1+\alpha_1)} \cdot \frac{\left\{ [F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right\}^{\frac{1}{\alpha_1}+2}}{\left\{ [F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right\}^{\frac{1}{\alpha_2}+2}} \cdot \frac{f_2(x)}{f_1(x)} \cdot \frac{g_2(y)}{g_1(y)} \\ & \left\{ (1+\alpha_1) \cdot [F_1(x)]^{\alpha_1} \cdot f_1(x) \cdot [F_2(x)]^{-\alpha_2-1} \cdot [G_1(y)]^{\alpha_1+1} \cdot [G_2(y)]^{-\alpha_2-1} \right. \\ & \left. - (1+\alpha_2) \cdot [F_2(x)]^{-\alpha_2-2} \cdot f_2(x) \cdot [F_1(x)]^{\alpha_1+1} \cdot [G_1(y)]^{\alpha_1+1} \cdot [G_2(y)]^{-\alpha_2-1} \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{[F_1(x)]^{\alpha_1+1}}{[F_2(x)]^{\alpha_2+1}} \cdot \frac{[G_1(y)]^{\alpha_1+1}}{[G_2(y)]^{\alpha_2+1}} \cdot \frac{f_2(x)}{f_1(x)} \cdot \frac{g_2(y)}{g_1(y)} \\
& \cdot \left\{ (1+2\alpha_2) \cdot \left[[F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right]^{\frac{-1}{\alpha_2}-3} \cdot \left[[F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right]^{\frac{1}{\alpha_2}+2} \cdot [F_2(x)]^{-\alpha_2-1} \cdot f_2(x) \right. \\
& \left. - (1+2\alpha_1) \cdot \left[[F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right]^{\frac{-1}{\alpha_2}-2} \cdot \left[[F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right]^{\frac{1}{\alpha_1}+1} \cdot [F_1(x)]^{-\alpha_1-1} \cdot f_1(x) \right\} \quad (16)
\end{aligned}$$

Similarly, the partial derivative with respect to y ,

$$\begin{aligned}
\frac{\partial h}{\partial y} &= \frac{(1+\alpha_2)}{(1+\alpha_1)} \cdot \frac{\left\{ [F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right\}^{\frac{1}{\alpha_1}+2}}{\left\{ [F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right\}^{\frac{1}{\alpha_2}+2}} \cdot \frac{[F_1(x)]^{\alpha_1+1}}{[F_2(x)]^{\alpha_2+1}} \cdot \frac{[G_1(y)]^{\alpha_1+1}}{[G_2(y)]^{\alpha_2+1}} \\
& \cdot \left\{ g_2'(y) \cdot [g_1(y)]^{-1} \cdot f_2(x) \cdot [f_1(x)]^{-1} - g_2(y) \cdot [g_1(y)]^{-2} \cdot g_1'(y) \cdot f_2(x) \cdot [f_1(x)]^{-1} \right\} \\
& + \frac{(1+\alpha_2)}{(1+\alpha_1)} \cdot \frac{\left\{ [F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_2} - 1 \right\}^{\frac{1}{\alpha_1}+2}}{\left\{ [F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right\}^{\frac{1}{\alpha_2}+2}} \cdot \frac{f_2(x)}{f_1(x)} \cdot \frac{g_2(y)}{g_1(y)} \\
& \cdot \left\{ (1+\alpha_1) \cdot [G_1(y)]^{\alpha_1} \cdot g_1(y) \cdot [G_2(y)]^{-\alpha_2-1} \cdot [F_1(x)]^{\alpha_1+1} \cdot [F_2(x)]^{-\alpha_2-1} \right. \\
& \left. - (1+\alpha_2) \cdot [G_2(y)]^{-\alpha_2-2} \cdot g_2(y) \cdot [G_1(y)]^{\alpha_1+1} \cdot [F_1(x)]^{\alpha_1+1} \cdot [F_2(x)]^{-\alpha_2-1} \right\} \\
& + \frac{[F_1(x)]^{\alpha_1+1}}{[F_2(x)]^{\alpha_2+1}} \cdot \frac{[G_1(y)]^{\alpha_1+1}}{[G_2(y)]^{\alpha_2+1}} \cdot \frac{f_2(x)}{f_1(x)} \cdot \frac{g_2(y)}{g_1(y)} \\
& \cdot \left\{ (1+2\alpha_2) \cdot \left[[F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right]^{\frac{-1}{\alpha_2}-3} \cdot \left[[F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right]^{\frac{1}{\alpha_2}+2} \cdot [G_2(y)]^{-\alpha_2-1} \cdot g_2(y) \right. \\
& \left. - (1+2\alpha_1) \cdot \left[[F_2(x)]^{-\alpha_2} + [G_2(y)]^{-\alpha_2} - 1 \right]^{\frac{-1}{\alpha_2}-2} \cdot \left[[F_1(x)]^{-\alpha_1} + [G_1(y)]^{-\alpha_1} - 1 \right]^{\frac{1}{\alpha_1}+1} \cdot [G_1(y)]^{-\alpha_1-1} \cdot g_1(y) \right\} \quad (17)
\end{aligned}$$

Theorem:

As $n \rightarrow \infty$, under H_0 , \bar{W}_n follows a normal distribution with mean = $h(a, b)$ and

$$\text{Variance} = \frac{1}{n} \cdot \left\{ \left(\frac{\partial h}{\partial x} \Big|_{x=a, y=b} \right)^2 \cdot \text{var}(X) + \left(\frac{\partial h}{\partial y} \Big|_{x=a, y=b} \right)^2 \cdot \text{var}(Y) + 2 \cdot \left(\frac{\partial h}{\partial x} \Big|_{x=a, y=b} \right) \cdot \left(\frac{\partial h}{\partial y} \Big|_{x=a, y=b} \right) \cdot \text{cov}(X, Y) \right\}$$

Where a, b are the respective means of X, Y under the null distribution.

Example:

Let us suppose that in a given situation, only the marginal distributions are known to us and that

$F_1(x) = 1 - e^{-\lambda_1 x}$, $G_1(y) = 1 - e^{-\lambda_2 y}$ are the marginal distributions of the first component and

$F_2(x) = \Phi\left(\frac{(x-\mu_1)}{\sigma_1}\right)$, $G_2(y) = \Phi\left(\frac{(y-\mu_2)}{\sigma_2}\right)$ are the

marginal distributions of the second component.

Here, we assume that the joint density can be modeled by using the Clayton Copula.

So, according to (3.8), the density ratio is as follows,

$$\frac{f_2(x_i, y_i)}{f_1(x_i, y_i)} = \frac{(1 + \alpha_2)}{(1 + \alpha_1)} \cdot \frac{\left\{ \left(1 - e^{-\lambda_1 \cdot x_i}\right)^{-\alpha_1} + \left(1 - e^{-\lambda_2 \cdot y_i}\right)^{-\alpha_1} - 1 \right\}^{\frac{1}{\alpha_1} + 2}}{\left\{ \left(\Phi\left(\frac{x_i - \mu_1}{\sigma_1}\right) \right)^{-\alpha_2} + \left(\Phi\left(\frac{y_i - \mu_2}{\sigma_2}\right) \right)^{-\alpha_2} - 1 \right\}^{\frac{1}{\alpha_2} + 2}} \cdot \frac{(1 - e^{-\lambda_1 \cdot x_i})^{1 + \alpha_1} \cdot (1 - e^{-\lambda_2 \cdot y_i})^{1 + \alpha_1}}{\left(\Phi\left(\frac{x_i - \mu_1}{\sigma_1}\right) \right)^{1 + \alpha_2} \cdot \left(\Phi\left(\frac{y_i - \mu_2}{\sigma_2}\right) \right)^{1 + \alpha_2}} \cdot \frac{\varphi\left(\frac{x_i - \mu_1}{\sigma_1}\right) \cdot \varphi\left(\frac{y_i - \mu_2}{\sigma_2}\right) \cdot e^{\lambda_1 \cdot x_i} \cdot e^{\lambda_2 \cdot y_i}}{\lambda_1 \cdot \lambda_2 \cdot \sigma_1 \cdot \sigma_2} \quad (18)$$

3.2. Numerical Results

We did a lot of simulations by generating samples of sizes $n=100$. In order to generate the bivariate exponential sample, we set the parameters $\lambda_1 = 0.7$, $\lambda_2 = 0.8$ for the respective marginal exponential distributions. Here is how we generated the bivariate exponential sample with the dependence parameter λ . Set $\lambda < \min\{\lambda_1, \lambda_2\}$. For the simulation, we chose $\lambda = 0.6$.

Next, generate a sample of U with size $= n$ according to an exponential distribution with mean $= (\lambda_1 - \lambda)^{-1}$.

Again, generate a sample of size of V with size $= n$ according to an exponential distribution with mean $= (\lambda_2 - \lambda)^{-1}$.

Once again, generate a sample of size of W with size $= n$ according to an exponential distribution with mean $= \lambda^{-1}$.

$$\text{Let } X = \min\{U, W\} \text{ and } Y = \min\{V, W\} \quad (19)$$

We can easily show that (X, Y) follows a bivariate exponential with dependence parameter $= \lambda$ while having marginal distributions with means λ_1^{-1} and λ_2^{-1} respectively.

In order to generate the bivariate normal sample, we used the following equations.

$$X = \mu_1 + \sigma_1 \cdot Z_1 \quad (20)$$

and

$$Y = \mu_2 + \frac{\sigma_{12}}{\sigma_1} \cdot Z_1 + \sqrt{\sigma_2^2 - \left(\frac{\sigma_{12}}{\sigma_1}\right)^2} \cdot Z_2 \quad (21)$$

where Z_1 and Z_2 are independent standard normal variables and μ_1, μ_2 are the means and σ_1^2, σ_2^2 are the variances of X and Y respectively. Note that $\text{Cov}(X, Y) = \sigma_{12}$.

In order to generate the bivariate normal sample, we chose

$$\mu_1 = 5, \mu_2 = 6, \sigma_1 = 1.11, \sigma_2 = 1.12, \sigma_{12} = 0.5675.$$

Next, to construct the Clayton Copula, we chose $\alpha_1 = 4.67$ and $\alpha_2 = 1.67$.

3.3. Power Computation (Clayton Copula based)

Here, we present the estimated power of the hypothesis test for various choices of the mixing proportion based on 1000 simulation runs from the Clayton Copula model. We used the same set of values that were used in the previous example.

For the hypothesis test, $H_0: p = 0$ (no mixture) versus $H_1: p > 0$ (mixture)

Table 1. Mixing Proportion Power.

0.00	0.049
0.10	0.728
0.20	0.924
0.30	0.966
0.40	0.987
0.50	0.991
0.60	1.000
0.70	1.000
0.80	1.000
0.90	1.000
1.00	1.000

Note that in Table 1, p represents the mixing proportion.

4. Gaussian Mixture Copula Model

Here, we will investigate a mixture of Gaussian Copulas. Consider the following two component Gaussian mixture.

$$F(x, y) = (1 - p) \cdot F_1(x, y) + p \cdot F_2(x, y) \\ = (1 - p) \cdot \Phi_2(\Phi^{-1}(F_1(x)), \Phi^{-1}(G_1(y))) + p \cdot \Phi_2(\Phi^{-1}(F_2(x)), \Phi^{-1}(G_2(y))) \quad (22)$$

Where, $\Phi_2(s, t)$ is the bivariate normal cumulative distribution function.

The density function for this bivariate normal mixture is given below.

$$f(x, y) = (1 - p) \cdot f_1(x, y) + p \cdot f_2(x, y) \quad (23)$$

Note that

$$f_1(x, y) = \frac{1}{\sqrt{1 - \rho_1^2}} \cdot e^{\frac{-\rho_1^2}{2(1 - \rho_1^2)} (\Phi^{-1}(F_1(x)) - \Phi^{-1}(G_1(y)))^2} \cdot f_1(x) \cdot g_1(y) \quad (24)$$

$$f_2(x, y) = \frac{1}{\sqrt{1 - \rho_2^2}} \cdot e^{\frac{-\rho_2^2}{2(1 - \rho_2^2)} (\Phi^{-1}(F_2(x)) - \Phi^{-1}(G_2(y)))^2} \cdot f_2(x) \cdot g_2(y) \quad (25)$$

and

$$h(x, y) = \frac{f_2(x, y)}{f_1(x, y)}$$

$$= \sqrt{\frac{(1-\rho_2^2)}{(1-\rho_1^2)}} \cdot e^{\left\{ \frac{\rho_2^2}{2(1-\rho_2^2)} (\Phi^{-1}(F_2(x)) - \Phi^{-1}(G_2(y)))^2 - \frac{\rho_1^2}{2(1-\rho_1^2)} (\Phi^{-1}(F_1(x)) - \Phi^{-1}(G_1(y)))^2 \right\}} \cdot \frac{f_2(x) \cdot g_2(y)}{f_1(x) \cdot g_1(y)} \quad (26)$$

4.1. Power Computation (Gaussian Copula Based)

Here, we present the estimated power of the hypothesis test for various choices of the mixing proportion based on 1000 simulation runs from the Gaussian Copula model. We used the same set of values that were used in the previous example. In order to describe the dependence, we found the corresponding correlation coefficients.

For this hypothesis test, $H_0 : p = 0$ (no mixture) versus $H_1 : p > 0$ (mixture)

Table 2. Mixing Proportion Power.

0.00	0.049
0.10	1.00
0.20	1.00
0.30	1.00
0.40	1.00
0.50	1.00
0.60	1.00
0.70	1.00
0.80	1.00
0.90	1.00
1.00	1.00

Note that in Table 2, p represents the mixing proportion.

4.2 Power Comparison of Clayton and Gaussian Copulas

The power computations in Tables 1 and 2, show that the Gaussian Copula based method is slightly better in this case in detecting a mixture compared to the Clayton Copula based method. Overall, both methods are very good.

5. Conclusion and Discussion

As we can see from the power computations, our method

is not only computationally very easy but also is very powerful. This test easily detects the bivariate mixture distributions. In this paper, only the Clayton and the Gaussian Copulas were considered. The author proposes to consider the other types of Copulas in the future work.

References

- [1] Engelman, L., and Hartigan, J. A. Percentage of a test for clusters. *Journal of American Statistical Association* (1969), 64, 1647-1648.
- [2] Frees, E. W., and Valdez, E. A. Understanding Relationships Using Copulas. *North American Actuarial Journal* (1998), 2(1), 1-25.
- [3] Giacomelli, F., Wiener, J., Kruskal, J. B., Pomeran, J. W., and Loud, A. V. Sub-Populations of blood lymphocytes as demonstrated by quantitative cytochemistry. *Journal of Histochemistry and Cytochemistry* (1971), 19, 426-433.
- [4] Hartigan, J. A., and Hartigan, P. M. The Dip Test of Unimodality. *Annals of Statistics* (1985), 13, 1, 70-84.
- [5] Nanthakumar, A. On the mixture Gaussian Copula to study the suitability of diagnostic tests. *Sri Lankan Journal of Applied Statistics* (2013), 14(2), 121-132.
- [6] Nanthakumar, A., Ganesalingam, S., and Ganesh, S. On Copula based discriminant rule. *Journal of Statistics & Management Systems* (2013), 16(6), 401-417.
- [7] Nanthakumar, A. Multivariate Kurtosis as a tool for comparing Copula Models", *International Journal of Statistics and Probability* (2016), 5(4), 67-78.
- [8] Nelson, R. B. An Introduction to Copulas (2006), Springer.
- [9] Rao, C. R. Linear Statistical Inference and its Applications (1973), 2nd Edition, Wiley, New York.
- [10] Sklar, A. Fonctions de repartition a n dimensions et leurs marges. *Publications de L' Institut de Statistique de L' Universite de Paris* (1959), 8, 229-231.