

A Diagnostic Test Based on a 9-Component Mixture Gaussian Copula Model

A. Nanthakumar

Correspondence: State University of New York at Oswego, Oswego NY 13126

Received: March 18, 2021 Accepted: April 27, 2021 Online Published: May 16, 2021

doi:10.5539/ijsp.v10n4p1

URL: <https://doi.org/10.5539/ijsp.v10n4p1>

Abstract

Here in this paper, we investigate the performance of a diagnostic test based on a mixture Gaussian Copula which incorporates a Markov Chain. Suppose that in the context of an infectious disease, there are three states; Susceptible (S), Infected (I), or Recovered (R). We compare the performance of this approach with the ROC (Receiver Operating Characteristic) Curve which is usually used in diagnostic studies.

Keywords: diagnostic test, Markova chain, mixture-Gaussian copula, infectious disease

1. Introduction

In biometry especially in epidemiological modeling SIR (Susceptible, Infected, and Recovered) model has been in use for a long time. Each compartment S (Susceptible), I (Infected), and R (Recovered) have their rates such the rate of susceptibility $s(t)$, rate of infection $i(t)$, and the rate of recovery $r(t)$ are modeled by using differential equations. This requires a good understanding of differential equations and partial differential equations. In addition, this approach at times requires iterative numerical methods to obtain the solutions. On the other hand, Copula models do not require the differential equations or the iterative numerical methods. These are probabilistic models and are fairly easy to manage compared to other existing mathematical models such as SIR model. The Copula based methods originated as a result of the pioneering work done by Sklar (1959). The Copulas use the dependence structure and the direction of the association for modeling. There are several copulas and each differ based on their properties. The Copula models fall under one of two major categories; Archimedean Copulas and non-Archimedean family of Copulas. In addition, there are copulas for the discrete type variables and the continuous type variables. These Copulas have applications in Actuarial Science, Biometry, Economics, Finance, Engineering etc. In Actuarial Science, Bowers et al (1997), Cox and Oakes (1984), David and Moeschberger (1978), Carriere (1994) used the Copulas to construct the competing risk models. Marshall and Olkin (1988) used the Copulas to construct the machine frailty models. Zheng and Klein (1995) considered the use of Copula models in the context of survival models. In Epidemiology, especially in the case of malaria, Demongeot et al (2013) used an Archimedean Copula known as the Gumbel Copula to study the interaction among the SIR compartments by using the rate of susceptibility, rate of infection, and the rate of recovery which were based on the Ross-Macdonald model. These are mainly differential equations. They used the Gumbel Copula model to derive the conditional distribution for the interaction from one compartment to another compartment. Moreover, the Copulas have applications in Quantile Regression too. For the literature review on the copulas, the interested readers are referred to Nelson (2006).

Here in this paper, the focus is on studying the suitability of a diagnostic test in the context of an epidemic disease which uses the Copula model. Nanthakumar (2013) used a mixture Gaussian Copula model to study the suitability of diagnostic tests in the context of a two state Markov Chain. Before that, Pundir (2011), Krazanowski and Hand (2009), Gonen (2007), Pepe (2003), Zhou et al (2002), Shultz (1995) and others have investigated the use of Receiver Operating Characteristic Curve (ROC Curve) to study the suitability of diagnostic tests based on single and multiple variables.

This paper extends the earlier results obtained by Nanthakumar (2013) to a three compartment situation. The objective here is to evaluate the diagnostic ability of the pre-treatment measurement (V_1) taken at time $t=1$ and the post-treatment measurement (V_2) taken at time $t=2$. In this regard, we will use the probability that $P(V_2 < V_1)$ based on a mixture-Gaussian Copula to evaluate the diagnostic ability. We compare the performance of this approach with the ROC (Receiver Operating Characteristic) curve which is usually used in diagnostic studies. This paper is divided into many sections. We present the methodology in section 2, application and numerical results in section 3, and the

discussion and conclusion in section 4.

2. Methodology

We believe that these three compartments (states) follow a transition pattern according to the following transition probability matrix,

$$\begin{pmatrix} b & 1-b & 0 \\ 0 & c & 1-c \\ 0 & 0 & 1 \end{pmatrix}$$

Note that,

$$b = P(\text{state is } S \text{ at time } t = 2 | \text{state was } S \text{ at time } t = 1)$$

and

$$c = P(\text{state is } I \text{ at time } t = 2 | \text{state was } I \text{ at time } t = 1)$$

Moreover, note that at time $t = 1$, the state space (S, I, R) has the probability distribution; $P(S) = \alpha$, $P(I) = \beta$, and $P(R) = 1 - \alpha - \beta$.

Note that since α and β are the state space probabilities for compartments 1 and 2 respectively in a three compartment model with $\alpha + \beta \leq 1$.

A random sample of n subjects who received treatment for this infectious disease were observed. Let V_1 be a health-related pre-treatment measurement (like temperature) taken at time =1 from a subject observed at random from this sample where the infectious disease is either prevalent or about to become prevalent. Note that, this subject could have come from any of the three compartments; S (Susceptible), I (Infected), or R (Recovered). Let us suppose at time = 1, for those who are susceptible to this disease (S), the distribution is $N(\mu_{11}, \sigma_{11}^2)$; for those who are infected by this disease (I), the distribution is $N(\mu_{12}, \sigma_{12}^2)$; for those who have already recovered from this disease (R), the distribution is $N(\mu_{13}, \sigma_{13}^2)$.

Similarly, let V_2 be a health-related post-treatment measurement V_2 taken at a different time (say time =2) for the second time from the same subject where the infectious disease is prevalent or about to become prevalent. Moreover, let us suppose at time = 2, for those who are susceptible, the distribution is $N(\mu_{21}, \sigma_{21}^2)$; for those who are infected, the distribution is $N(\mu_{22}, \sigma_{22}^2)$; for those who have recovered, the distribution is $N(\mu_{23}, \sigma_{23}^2)$.

We are using the probability $p = P(V_2 < V_1)$ to evaluate the diagnostic ability of the measurements. But computing this probability involves the joint distribution of (V_1, V_2) . This is where we need the Copula. Here, we use a mixture Gaussian Copula which captures the transition among the states and at the same time gives a fairly approximate estimate of this probability $p = P(V_2 < V_1)$.

As noted earlier, let V_1 be the measurement taken at time = 1 (say for example temperature) and V_2 be the measurement taken at time = 2.

$$P(V_1 \leq v_1) = P(V_1 \leq v_1 | \text{susceptible})P(\text{susceptible}) + P(V_1 \leq v_1 | \text{inf ected})P(\text{inf ected}) + P(V_1 \leq v_1 | \text{re coverd})P(\text{re covered}) \tag{1}$$

This leads to the result that the pre-treatment measurement V_1 taken at time $t = 1$ follow a mixture normal distribution as indicated below,

$$V_1 \sim \alpha N(\mu_{11}, \sigma_{11}^2) + \beta N(\mu_{12}, \sigma_{12}^2) + (1 - \alpha - \beta) N(\mu_{13}, \sigma_{13}^2) \tag{2}$$

Similarly, under the assumption that the transition takes place according to the first-order Markov Chain described earlier, the post-treatment measurement V_2 taken at time $t = 2$ follow a different mixture normal distribution

$$V_2 \sim b \alpha N(\mu_{21}, \sigma_{21}^2) + ((1-b)\alpha + c\beta) N(\mu_{22}, \sigma_{22}^2) + (1 - \alpha - c\beta) N(\mu_{23}, \sigma_{23}^2) \tag{3}$$

As we can see, the marginal distributions are each three component mixture of normal distributions. It appears reasonable to model the joint distribution of V_1, V_2 as a nine-component mixture of bivariate Gaussian Copulas.

2.1 Mixture of Bivariate Gaussian Copulas

Here, we define the nine-component mixture Gaussian Copulas.

Let $F(v_1, v_2) = P(V_1 \leq v_1, V_2 \leq v_2)$ be the joint cumulative distribution function.

Then, $f(v_1, v_2) = \frac{\partial^2 F(v_1, v_2)}{\partial v_1 \partial v_2}$ is the joint density function. We will model this joint density function as follows.

$$\begin{aligned}
 f(v_1, v_2) = & \frac{p_1}{2\pi\sigma_{11}\sigma_{21}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{11}}{\sigma_{11}}\right)^2 + \left(\frac{v_2-\mu_{21}}{\sigma_{21}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{11}}{\sigma_{11}}\right)\left(\frac{v_2-\mu_{21}}{\sigma_{21}}\right)\right\}} \\
 & + \frac{p_2}{2\pi\sigma_{12}\sigma_{22}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{12}}{\sigma_{12}}\right)^2 + \left(\frac{v_2-\mu_{22}}{\sigma_{22}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{12}}{\sigma_{12}}\right)\left(\frac{v_2-\mu_{22}}{\sigma_{22}}\right)\right\}} \\
 & + \frac{p_3}{2\pi\sigma_{13}\sigma_{23}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{13}}{\sigma_{13}}\right)^2 + \left(\frac{v_2-\mu_{23}}{\sigma_{23}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{13}}{\sigma_{13}}\right)\left(\frac{v_2-\mu_{23}}{\sigma_{23}}\right)\right\}} \\
 & + \frac{p_4}{2\pi\sigma_{11}\sigma_{22}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{11}}{\sigma_{11}}\right)^2 + \left(\frac{v_2-\mu_{22}}{\sigma_{22}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{11}}{\sigma_{11}}\right)\left(\frac{v_2-\mu_{22}}{\sigma_{22}}\right)\right\}} \\
 & + \frac{p_5}{2\pi\sigma_{11}\sigma_{23}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{11}}{\sigma_{11}}\right)^2 + \left(\frac{v_2-\mu_{23}}{\sigma_{23}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{11}}{\sigma_{11}}\right)\left(\frac{v_2-\mu_{23}}{\sigma_{23}}\right)\right\}} \\
 & + \frac{p_6}{2\pi\sigma_{12}\sigma_{21}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{12}}{\sigma_{12}}\right)^2 + \left(\frac{v_2-\mu_{21}}{\sigma_{21}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{12}}{\sigma_{12}}\right)\left(\frac{v_2-\mu_{21}}{\sigma_{21}}\right)\right\}} \\
 & + \frac{p_7}{2\pi\sigma_{12}\sigma_{23}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{12}}{\sigma_{12}}\right)^2 + \left(\frac{v_2-\mu_{23}}{\sigma_{23}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{12}}{\sigma_{12}}\right)\left(\frac{v_2-\mu_{23}}{\sigma_{23}}\right)\right\}} \\
 & + \frac{p_8}{2\pi\sigma_{13}\sigma_{21}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{13}}{\sigma_{13}}\right)^2 + \left(\frac{v_2-\mu_{21}}{\sigma_{21}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{13}}{\sigma_{13}}\right)\left(\frac{v_2-\mu_{21}}{\sigma_{21}}\right)\right\}} \\
 & + \frac{p_9}{2\pi\sigma_{13}\sigma_{22}\sqrt{1-\rho^2}} e^{\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{v_1-\mu_{13}}{\sigma_{13}}\right)^2 + \left(\frac{v_2-\mu_{22}}{\sigma_{22}}\right)^2 - 2\rho\left(\frac{v_1-\mu_{13}}{\sigma_{13}}\right)\left(\frac{v_2-\mu_{22}}{\sigma_{22}}\right)\right\}} \tag{4}
 \end{aligned}$$

Remark: The Copulas are supposed to yield the marginal distributions when the data is collapsed. Therefore, collapsing the data and then equating the marginal distributions yield the following equations.

$$p_1 + p_4 + p_5 = \alpha \tag{5}$$

$$p_2 + p_6 + p_7 = \beta \tag{6}$$

$$p_3 + p_8 + p_9 = 1 - \alpha - \beta \tag{7}$$

$$p_1 + p_6 + p_8 = b\alpha \tag{8}$$

$$p_2 + p_4 + p_9 = (1 - b)\alpha + \beta \tag{9}$$

$$p_3 + p_5 + p_7 = 1 - \alpha - c\beta \tag{10}$$

Also note that

$$\sum_{i=1}^9 p_i = 1. \tag{11}$$

Moreover, the sum of equations (5), (6), (7) \Rightarrow (11).

Also, the sum of equations (8), (9), (10) \Rightarrow (11).

This means only four of these equations are independent although there are nine parameters. Since there are more number of parameters than the number of independent equations, we need to set five of these parameters at fixed values in such a way that we have a feasible solution to this system of equations.

Luckily in this case, the transition from state I (Infected) at time $t = 1$ to state S (Susceptible) at time $t = 2$ is not theoretically feasible. Similarly, the transition from state R (Recovery) at time $t = 1$ to state S (Susceptible) at time $t = 2$ is assumed to be not theoretically feasible. Likewise, the transition from state R (Recovery) at time $t = 1$ to state I (Infected) at time $t = 2$ is not theoretically possible.

So, we will set the corresponding parameters $p_6 = 0, p_8 = 0, p_9 = 0$ or to values closer to 0 in addition to fixing two other parameters.

Note that $V_2 < V_1$ is an indication of improvement as a result of this treatment in this diagnostic study. So, the interest is in computing the probability $P(V_2 < V_1)$.

Note that based on this copula density,

$$P(V_2 < V_1) = \sum_{i=1}^9 p_i P(V_2 - V_1 < 0 | group = i)$$

3. Application

3.1 Numerical Example

As we know there are infectious diseases like Cholera, Malaria, SARS, COVID-19 that have the potential to affect the human population from time to time. There are mainly three possible states when it comes to these infectious diseases; susceptible (S), infected (I), and recovered (R). Suppose that an infectious disease is prevalent in a region and due to that people are advised to take a preventive medicine (treatment). In order to ensure the effectiveness of this treatment, two measurements (a pre-treatment measurement V_1 was taken at time $t = 1$. Then again, a post-treatment measurement V_2 was taken at time $t = 2$ from the same individual. Here, we assume that the state to state transition is taking place according to a Markov Chain as described earlier.

Here we assume the estimates of $P(V_2 < V_1)$ or analogously ($P(V_2 > V_1)$) can provide a true measure of the health status if the measurements (V_1, V_2) are taken carefully. Let us suppose that the transition is taking place according to the transition probability matrix given by

$$\begin{pmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}$$

Also, we are making the assumption that all the component distributions are normally distributed with the means and standard deviations given as follows.

Table 1.

	Susceptible (S)	Infected (I)	Recovery (R)
Pre-treatment (V_1)	Mean= $\mu_{11} = 98.2$ Std = $\sigma_{11} = 0.8$	Mean = $\mu_{12} = 103.5$ Std = $\sigma_{12} = 0.9$	Mean = $\mu_{13} = 98.6$ Std = $\sigma_{13} = 0.4$
Post-treatment (V_2)	Mean = $\mu_{21} = 98.4$ Std = $\sigma_{21} = 0.5$	Mean = $\mu_{22} = 99.9$ Std = $\sigma_{22} = 0.7$	Mean = $\mu_{23} = 98.5$ Std = $\sigma_{23} = 0.3$

In addition, we are making one additional assumption that the correlation that could exist between V_1 and V_2 is a constant within the categories where it is believed to exist. Let us say that $\rho = 0.3$.

It is to be noted that at time $t = 1$, 70% of the population was susceptible, 20% infected, and the remaining 10% were already immune to the disease. At time $t = 2$, the percentages are as follows due to the transition explained by the Markov Chain,

$$(0.7 \quad 0.2 \quad 0.1) \begin{pmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{pmatrix} = (0.525 \quad 0.355 \quad 0.12)$$

This means that at time $t = 2$, 52.5% of the population was still susceptible, 35.5% were infected, and 12% were either immune or fully recovered.

Based on the above percentages, the mixing proportions can be estimated as follows.

$$p_1 + p_4 + p_5 = 0.7 \tag{12}$$

$$p_2 + p_6 + p_7 = 0.2 \tag{13}$$

$$p_3 + p_8 + p_9 = 0.1 \tag{14}$$

$$p_1 + p_6 + p_8 = 0.525 \tag{15}$$

$$p_2 + p_4 + p_9 = 0.355 \tag{16}$$

$$p_3 + p_5 + p_7 = 0.12 \tag{17}$$

Also note that $\sum_{i=1}^9 p_i = 1$

This means that there is redundancy and that there are only four independent equations in nine variables. This creates a situation that are infinitely many solutions. We will accept the solutions that are between 0 and 1 as these are probabilities (or mixing proportions) for the mixture Gaussian Copula.

Moreover, as we can see from the following equations, the solutions to the mixing proportions impact the estimates of the overall covariance, overall variance, and the overall probability for effectiveness of the treatment.

$$Cov(V_1, V_2) = \sum_{i=1}^9 p_i Cov((V_1, V_2) | Group = i) \tag{18}$$

$$Var(V_1) = \sum_{i=1}^9 p_i Var((V_1) | Group = i) \tag{19}$$

$$Var(V_2) = \sum_{i=1}^9 p_i Var((V_2) | Group = i) \tag{20}$$

$$\begin{aligned} P(V_2 < V_1) &= P(V_2 - V_1 < 0) \\ &= \sum_{i=1}^9 p_i P((V_2 - V_1 < 0) | Group = i) \end{aligned} \tag{21}$$

We will proceed with the calculations as follows,

$$\begin{aligned} P(V_2 - V_1 < 0) &= p_1 P\left(\frac{V_2 - V_1 - (\mu_{21} - \mu_{11})}{\sqrt{\sigma_{21}^2 + \sigma_{11}^2 - 2\rho\sigma_{21}\sigma_{11}}} < \frac{0 - (\mu_{21} - \mu_{11})}{\sqrt{\sigma_{21}^2 + \sigma_{11}^2 - 2\rho\sigma_{21}\sigma_{11}}}\right) \\ &+ p_2 P\left(\frac{V_2 - V_1 - (\mu_{22} - \mu_{12})}{\sqrt{\sigma_{22}^2 + \sigma_{12}^2 - 2\rho\sigma_{22}\sigma_{12}}} < \frac{0 - (\mu_{22} - \mu_{12})}{\sqrt{\sigma_{22}^2 + \sigma_{12}^2 - 2\rho\sigma_{22}\sigma_{12}}}\right) \\ &+ p_3 P\left(\frac{V_2 - V_1 - (\mu_{23} - \mu_{13})}{\sqrt{\sigma_{23}^2 + \sigma_{13}^2 - 2\rho\sigma_{23}\sigma_{13}}} < \frac{0 - (\mu_{23} - \mu_{13})}{\sqrt{\sigma_{23}^2 + \sigma_{13}^2 - 2\rho\sigma_{23}\sigma_{13}}}\right) \\ &+ p_4 P\left(\frac{V_2 - V_1 - (\mu_{22} - \mu_{11})}{\sqrt{\sigma_{22}^2 + \sigma_{11}^2 - 2\rho\sigma_{22}\sigma_{11}}} < \frac{0 - (\mu_{22} - \mu_{11})}{\sqrt{\sigma_{22}^2 + \sigma_{11}^2 - 2\rho\sigma_{22}\sigma_{11}}}\right) \\ &+ p_5 P\left(\frac{V_2 - V_1 - (\mu_{23} - \mu_{11})}{\sqrt{\sigma_{23}^2 + \sigma_{11}^2 - 2\rho\sigma_{23}\sigma_{11}}} < \frac{0 - (\mu_{23} - \mu_{11})}{\sqrt{\sigma_{23}^2 + \sigma_{11}^2 - 2\rho\sigma_{23}\sigma_{11}}}\right) \\ &+ p_6 P\left(\frac{V_2 - V_1 - (\mu_{21} - \mu_{12})}{\sqrt{\sigma_{21}^2 + \sigma_{12}^2 - 2\rho\sigma_{21}\sigma_{12}}} < \frac{0 - (\mu_{21} - \mu_{12})}{\sqrt{\sigma_{21}^2 + \sigma_{12}^2 - 2\rho\sigma_{21}\sigma_{12}}}\right) \\ &+ p_7 P\left(\frac{V_2 - V_1 - (\mu_{23} - \mu_{12})}{\sqrt{\sigma_{23}^2 + \sigma_{12}^2 - 2\rho\sigma_{23}\sigma_{12}}} < \frac{0 - (\mu_{23} - \mu_{12})}{\sqrt{\sigma_{23}^2 + \sigma_{12}^2 - 2\rho\sigma_{23}\sigma_{12}}}\right) \\ &+ p_8 P\left(\frac{V_2 - V_1 - (\mu_{21} - \mu_{13})}{\sqrt{\sigma_{21}^2 + \sigma_{13}^2 - 2\rho\sigma_{21}\sigma_{13}}} < \frac{0 - (\mu_{21} - \mu_{13})}{\sqrt{\sigma_{21}^2 + \sigma_{13}^2 - 2\rho\sigma_{21}\sigma_{13}}}\right) \\ &+ p_9 P\left(\frac{V_2 - V_1 - (\mu_{22} - \mu_{13})}{\sqrt{\sigma_{22}^2 + \sigma_{13}^2 - 2\rho\sigma_{22}\sigma_{13}}} < \frac{0 - (\mu_{22} - \mu_{13})}{\sqrt{\sigma_{22}^2 + \sigma_{13}^2 - 2\rho\sigma_{22}\sigma_{13}}}\right) \end{aligned} \tag{22}$$

Next, we will evaluate the probability $P(V_2 < V_1)$ by using the solutions for the mixing proportions as given in the next table.

Table 2.

p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	Copula Estimate of $P(V_2 < V_1)$
0.525	0.200	0.100	0.155	0.020	0.000	0.000	0.000	0.000	0.48174
0.435	0.120	0.020	0.205	0.060	0.040	0.040	0.050	0.030	0.44655
0.435	0.120	0.040	0.215	0.050	0.050	0.030	0.040	0.020	0.44847
0.520	0.095	0.000	0.160	0.020	0.005	0.100	0.000	0.100	0.42357
0.445	0.095	0.045	0.230	0.025	0.055	0.050	0.025	0.030	0.43783
0.470	0.110	0.065	0.225	0.005	0.040	0.050	0.015	0.020	0.44591
0.485	0.130	0.070	0.205	0.010	0.030	0.040	0.010	0.020	0.45285
0.495	0.140	0.070	0.195	0.010	0.020	0.040	0.010	0.020	0.45659
0.490	0.130	0.075	0.205	0.005	0.030	0.040	0.005	0.020	0.45287
0.480	0.140	0.065	0.195	0.025	0.030	0.030	0.015	0.020	0.45603
0.525	0.200	0.000	0.055	0.120	0.000	0.000	0.000	0.100	0.45737
0.335	0.065	0.030	0.290	0.075	0.120	0.015	0.070	0.000	0.43187
0.415	0.090	0.040	0.215	0.070	0.100	0.010	0.010	0.050	0.42896
0.415	0.085	0.065	0.245	0.040	0.100	0.015	0.010	0.025	0.43345
0.415	0.070	0.060	0.255	0.030	0.100	0.030	0.010	0.030	0.42745
0.425	0.100	0.100	0.255	0.020	0.100	0.000	0.000	0.000	0.44437
0.425	0.090	0.100	0.265	0.010	0.100	0.010	0.000	0.000	0.44118
0.345	0.100	0.000	0.255	0.100	0.080	0.020	0.100	0.000	0.44513
0.445	0.095	0.045	0.230	0.025	0.055	0.050	0.025	0.030	0.43783
0.420	0.095	0.015	0.180	0.100	0.100	0.005	0.005	0.080	0.42326

Note that the empirical estimate of $P(V_2 < V_1)$ is 0.477. As you can see from the above table (Table 2), the Copula based estimates are fairly close to the empirical estimate. Not only that, these estimates are also robust across the feasible solutions for the mixing proportions. It is clear that the solutions for the mixing proportions do not affect the Copula based estimate for $P(V_2 < V_1)$. This was the case in the previous study conducted by this author in the context of two component mixture.

3.2 ROC Curve

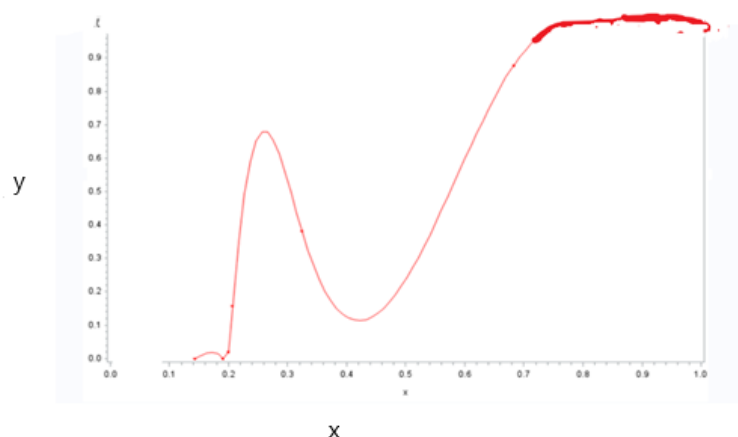
We can study the effectiveness of the diagnostic by using the ROC curve too. For this, we will use the area under the ROC (AUROC) for the evaluation.

$$\text{Let } x = P(V_1 > \tau)$$

$$y = P(V_2 > \tau)$$

where τ is a threshold value chosen within the range of the pre-treatment and post-treatment measurements.

In this analysis, we will draw the ROC curve (y versus x) and evaluate the area under the curve. The area under the curve is a measure of the performance of the diagnosis test.



As you can see, the area under the ROC curve is about 0.5 and it is in agreement with the estimate given by the mixture-Gaussian Copula model.

4. Conclusion and Discussion

The purpose of this study was to see whether the mixture-Gaussian Copula model can be used in evaluating the diagnostic ability of the pre-treatment and post-treatment measurements V_1, V_2 when the joint distribution of these measurements is unknown. Here, we use a 9-component mixture Gaussian Copula to model the joint distribution of V_1, V_2 . As seen from this study, this mixture-Gaussian Copula model is doing fairly well in evaluating the diagnostic ability of the pre-treatment and post-treatment measurements. This is supported by the empirical as well as ROC-curve based estimates of $P(V_2 < V_1)$ which measures the diagnostic ability of this test.

Acknowledgement

I want to thank my former student Sean Chung for his help in checking the numerical computations.

References

- Bowers, N., Gerber, H., Hickman, J., Jones, D., & Nesbitt, C. (1997). *Actuarial Mathematics* (2nd ed.). Schaumburg, Ill, Society of Actuaries.
- Carriere, J. (1994). A Large Sample Test for One-Parameter Families of Copulas. *Communications in Statistics – Theory and Methods*, 23(5), 1311-1317. <https://doi.org/10.1080/03610929408831323>
- Cox, R. D., & Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, New York.
- David, H. A., & Moeschberger, M. L. (1978). *The Theory of Competing Risks*. New York, Mac Millan.
- Demongeot, J., Ghassani, M., Rachdi, M., Ouassou, I., & Taramasco, C. (2013). *Archimedean Copula and Contagion Modeling in Epidemiology. Networks & Heterogeneous Media*, 8(1), 149-170. <https://doi.org/10.3934/nhm.2013.8.149>
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141–151. <https://doi.org/10.1093/biomet/65.1.141>
- Gonen, M. (2007). Receiver Operating Characteristic (ROC) curves. *SUGI*, 31(210), 1-18.
- Krzanowski, W. J., & Hand, D. J. (2009). *ROC curves for continuous data*, *Monographs on Statistics and Applied Probability*. CRC Press, Taylor and Francis Group, New York.
- Marshall, A. W., & Olkin, I. (1988). *Families of Multivariate Distributions*. *Journal of the American Statistical Association*, 83, 834-841. <https://doi.org/10.1080/01621459.1988.10478671>
- Nanthakumar, A. (2013). On the Mixture Gaussian Copula to study the suitability of Diagnostic Tests. *Sri Lankan Journal of Applied Statistics*, 14(2), 121-132. <https://doi.org/10.4038/sljastats.v14i2.6217>
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York. <https://doi.org/10.1007/978-1-4757-3076-0>

- Pepe, M. S. (2003). *The Statistical Evaluation of medical tests for classification and prediction*. Oxford Statistical Science Series, Oxford University Press.
- Pundir, S. (2011). *Receiver Operating Characteristic Curve for Bi-Weibull distribution and its properties*. Proceedings of the Institute of Applied Statistics, Sri Lanka.
- Shultz, E. K. (1995). Multivariate Receiver Operating Characteristic Curve Analysis: Prostate Cancer Screening as an example. *Clinical Chemistry*, 41/8(B), 1248 – 1255. <https://doi.org/10.1093/clinchem/41.8.1248>
- Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de L' Institut de Staistique de Universite de Paris*, 8, 229-231.
- Zheng, M., & Klein, J. P. (1995). Estimates of Marginal Survival for Dependent Competing Risks based on an assumed Copula. *Biometrika*, 82, 127 – 138. <https://doi.org/10.1093/biomet/82.1.127>
- Zhou, X. H., Obuchowski, N. A., & McClish, D. K. (2002). *Statistical Methods in diagnostic medicine*. Wiley, New York. <https://doi.org/10.1002/9780470317082>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).