

PEMODELAN KEBERGANTUNGAN DALAM MENGGONSTRUKSI DISTRIBUSI BIVARIAT COPULA FRANK PADA DATA MARGINAL DISKRIT MELALUI TRANSFORMASI NORMAL STANDAR DAN JITTERS

(Modelling Of Dependence to Construct a Frank Copula Bivariate Distribution on Discrete Marginal Data by Jitters and Standard Normal Transformation)

Andi Fitriawati, S.Pd., M.Si^{1*}, Dani Al Mahkya, S.Si., M.Si², Radot MH Siahaan, S.Si., M.Si³,
Dian Angraini, S.Si., M.Sc⁴

^{1,2,3,4} Program Studi Sains Aktuaria, Jurusan Sains, Institut Teknologi Sumatera
Jl. Terusan Ryacudu, Lampung Selatan, 35365, Lampung, Indonesia

e-mail: andi.fitriawati@at.itera.ac.id^{1*}, dani.almahkya@at.itera.ac.id², radot.siahaan@at.itera.ac.id³,
dian.angraini@at.itera.ac.id⁴

Abstrak: Data diskrit merupakan data empirik hasil realisasi variabel acak diskrit maupun kontinu. Ketika memiliki dua jenis data diskrit, seringkali ingin dikonstruksi distribusi bivariatnya untuk berbagai keperluan, baik fungsi peluang maupun fungsi distribusinya. Namun, saat data yang dimiliki terdapat kebergantungan, maka mengkonstruksi distribusi bivariatnya tidaklah mudah. Oleh sebab itu, digunakan Copula. Permasalahan lain timbul ketika data yang dimiliki tidak hanya memiliki kebergantungan tetapi juga berasal dari marginal diskrit. Berdasarkan teorema Sklar, penggunaan Copula dalam mengkonstruksi distribusi bivariat pada marginal diskrit akan menghasilkan suatu Copula C yang tidak unik. Akibatnya akan menimbulkan interpretasi yang tidak jelas, terutama untuk sifat kebergantungannya. Oleh sebab itu, diperlukan suatu teknik untuk mengkonstruksi distribusi bivariat dari data tersebut, yaitu dengan mengkontinukan distribusi marginalnya. Mengkontinukan distribusi marginalnya dilakukan melalui transformasi normal standar dan jitters. Hasil transformasi mampu mempresentasikan data aslinya. Hal ini terlihat dari perilaku penyebaran data dan ukuran kebergantungan dari data hasil transformasi dengan data aslinya adalah sama. Ukuran kebergantungan yang digunakan, yaitu Korelasi Pearson dan Kendall's tau. Selanjutnya, hasil transformasi ini kemudian digunakan untuk mengkonstruksi distribusi bivariat dari data yang dimiliki menggunakan Copula. Copula yang digunakan adalah Copula Frank dengan asumsi bahwa data tidak memiliki kebergantungan ekor atas maupun bawah. Jadi, fungsi peluang bivariat dan/atau fungsi distribusi bivariat dari data hasil transformasi mempresentasikan fungsi peluang bivariat dan/atau fungsi distribusi bivariat dari data aslinya. Seluruh prosesnya akan diilustrasikan melalui data simulasi.

Kata Kunci: Marginal diskrit, transformasi normal standar, jitters, Korelasi Pearson, Kendall's tau, Copula Frank.

Abstract: Discrete data is an empirical data as realizations of either discrete or continuous random variables. When having two types of discrete data, often want to be constructed their bivariate distribution for various purposes, either probability function or distribution function. However, when having data dependence, then to construct its bivariate distribution is not easy. Therefore, used Copula. Another problem arises when the data that is owned not only has dependencies but also comes from discrete marginal. Based on Sklar's theorem, the use of Copula in constructing bivariate distribution on discrete marginal will result in a non-unique Copula C. Consequently it will cause unclear interpretation, especially for the dependency. Therefore, it is necessary that a technique to construct the bivariate distribution from the data by continuing the marginal distribution. To continuing the marginal distribution is done through standard normal transformation and jitters. The results of transformations are able to present the original data. It is seen from their correlation is

same, both of Pearson correlation and Kendall's tau. Furthermore, the results of this transformation are then used to construct bivariate distribution of data owned using Copula. The copula used is Frank Copula with assuming that the data does not have the upper or lower tail dependent. Thus, the bivariate probability function and/or the bivariate distribution function from the transformation result present the bivariate probability function and/or the bivariate distribution function from the original data. We provide examples of simulated data to illustrate such construction.

Keywords: *Discrete marginal, standard normal transformation, jitters, Pearson correlation, Kendall's tau, Frank Copula.*

1. PENDAHULUAN

Data diskrit merupakan data empirik hasil realisasi variabel acak diskrit maupun kontinu. Hal menarik yang ingin diketahui ketika memiliki dua jenis data diskrit, yaitu peluang (dan/atau peluang kumulatif) kedua jenis data tersebut dapat terjadi secara bersamaan. Berkaitan dengan hal ini, maka perlu dikonstruksi distribusi bivariatnya, baik fungsi peluang maupun fungsi distribusi (peluang kumulatif).

Untuk memudahkan dalam mengkonstruksi distribusi bivariat, seringkali diasumsikan bahwa data berasal dari dua variabel yang saling bebas. Sehingga mengkonstruksi fungsi peluang bivariatnya hanya perlu melakukan perkalian dari kedua fungsi peluang marginalnya. Realitanya adalah tidak semua data bisa diasumsikan saling bebas satu sama lain. Apalagi jika data tersebut memiliki sifat kebergantungan yang cukup kuat. Sifat kebergantungan ini bahkan menjadi salah satu persoalan yang banyak dipelajari dalam probabilitas dan statistika [3].

Sifat kebergantungan pada data merupakan suatu komponen penting dalam proses penginterpretasian yang tidak boleh diabaikan. Oleh sebab itu, tidaklah benar jika dua jenis data yang memiliki kebergantungan selalu diasumsikan berasal dari dua variabel yang saling bebas ketika ingin mengkonstruksi distribusi bivariatnya. Mengkonstruksi distribusi bivariat dari data yang memiliki kebergantungan memang tidaklah mudah. Apalagi ketika data tersebut berasal dari dua distribusi yang berbeda. Meskipun demikian, hal tersebut tetap harus dilakukan. Oleh sebab itu, diperlukan suatu metode untuk memodelkan kebergantungan data dalam mengkonstruksi distribusi bivariatnya.

Copula adalah salah satu alat populer yang digunakan untuk mengkonstruksi distribusi bivariat, terutama pada data yang memiliki kebergantungan. Hal ini dikarenakan kelebihan dari Copula yang mampu memodelkan sifat kebergantungan data. Dalam mengkonstruksi distribusi bivariat dengan Copula diperlukan distribusi marginalnya. Misalkan $\{x\}$ dan $\{y\}$ merupakan dua jenis himpunan data diskrit yang berasal dari variabel acak X dan Y dengan fungsi distribusi berturut-turut adalah F_X dan F_Y . Selanjutnya, untuk mengkonstruksi distribusi bivariat Copula dari variabel acak X dan Y maka perlu dilakukan transformasi pada setiap variabel acaknya, melalui fungsi distribusinya, menjadi variabel acak baru yang berdistribusi uniform standar. Sehingga, fungsi distribusi bivariat Copula didefinisikan sebagai $C(v, w) = P(V \leq v, W \leq w)$; dimana $v = F_X$ dan $w = F_Y$ serta v, w dengan support $[0,1]$.

Lebih jauh, teorema Sklar menyatakan bahwa jika diberikan suatu fungsi distribusi bivariat H dengan marginal F_X dan F_Y kontinu maka akan terdapat suatu Copula C yang unik. Artinya, teorema ini menjamin keunikan suatu Copula pada marginal kontinu tetapi tidak pada marginal diskrit. Hal ini berarti bahwa penggunaan Copula dalam mengkonstruksi distribusi bivariat pada marginal diskrit (F_X dan F_Y diskrit) akan menghasilkan suatu Copula C yang tidak unik. Ketidak-unikan ini akan menimbulkan interpretasi yang tidak jelas, terutama sifat kebergantungannya [1]. Dapat dikatakan bahwa mengkonstruksi distribusi bivariat dari marginal kontinu menggunakan Copula tidaklah masalah. Namun, tidak untuk marginal diskrit.

Oleh sebab itu, perlu diuraikan suatu teknik mengkonstruksi distribusi bivariat Copula dalam memodelkan kebergantungan pada marginal diskrit agar mampu menghasilkan suatu Copula C yang unik. Idenya, yaitu mengkontinukan distribusi marginal diskritnya terlebih dahulu. Hasil distribusi marginal kontinu inilah yang kemudian dikonstruksi menggunakan Copula. Sehingga, distribusi bivariat Copula yang dihasilkan bisa unik sesuai teorema Sklar.

Mengkontinukan suatu distribusi dapat dilakukan melalui proses transformasi normal standar. Hampir semua distribusi dari suatu data (terutama data berukuran besar) dapat dihampiri dengan distribusi normal standar. Ketika diperoleh bahwa data mengikuti suatu distribusi diskrit tertentu, maka data tersebut dapat ditransformasikan menjadi berdistribusi normal standar. Dengan kata lain, data berdistribusi diskrit tersebut telah dikontinukan menjadi distribusi kontinu normal standar. Selain transformasi normal standar, mengkontinukan suatu distribusi diskrit dapat juga dilakukan dengan transformasi jitters. Transformasi jitters telah digunakan oleh Madsen dan Fang [4] ketika mengkontinukan data diskritnya. Hal yang sama juga digunakan oleh Shi dan Valdez [8] untuk mengkontinukan data longitudinalnya.

Data hasil transformasi normal standar maupun jitters harus mampu mempresentasikan data yang aslinya. Sehingga, distribusi bivariat dari data hasil transformasi mampu mempresentasikan distribusi bivariat data aslinya. Oleh karena itu, data hasil transformasi harus memiliki karakteristik maupun sifat kebergantungan yang sama dengan data aslinya. Untuk membuktikan hal ini, maka akan dilakukan analisis terkait perilaku penyebaran data hasil transformasi terhadap data yang aslinya. Hal ini dapat juga terlihat dari ukuran kebergantungan antara data transformasi dan data aslinya adalah sama.

Adapun ukuran kebergantungan yang digunakan, yaitu korelasi Pearson dan Kendall's tau. Nilai korelasi Pearson dan Kendall's tau yang diperoleh dari data transformasi normal standar diharapkan sama dengan nilai korelasi Pearson dan Kendall's tau dari data yang aslinya. Sama halnya dengan nilai Kendall's tau dari data hasil jitters juga diharapkan sama dengan nilai Kendall's tau dari data aslinya.

Penelitian ini akan memaparkan suatu teknik memodelkan kebergantungan dalam mengkonstruksi distribusi bivariat Copula pada data marginal diskrit melalui transformasi normal standar dan jitters. Adapun Copula yang akan digunakan pada penelitian ini adalah Copula Frank. Copula ini dipilih dengan asumsi bahwa data tidak memiliki kebergantungan ekor atas maupun ekor bawah. Semua prosesnya akan diilustrasikan melalui data simulasi yang dibangun dari distribusi binomial dan poisson.

2. METODOLOGI

2.1. Sumber Data

Data yang digunakan adalah data simulasi dari distribusi binomial (100, 0.25) dan poisson (20) yang masing-masing berukuran 200. Hal ini sesuai dengan pernyataan Roscoe (1975) yang dikutip Sekaran [7] bahwa ukuran sampel minimum 30.

2.2. Variabel Penelitian

Beberapa variabel yang akan digunakan dalam penelitian ini sebagai berikut:

X : Data berdistribusi binomial (100, 0.25)

Y : Data berdistribusi poisson (20)

X^* : Data hasil transformasi normal standar dari X

Y^* : Data hasil transformasi normal standar dari Y

$X^@$: Data hasil transformasi jitters dari X

$Y^@$: Data hasil transformasi jitters dari Y

2.3. Tahapan Penelitian

Adapun tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mengidentifikasi model transformasi normal standar dan jitters.
2. Uji nilai Korelasi Pearson dan Kendall's tau dari hasil transformasi terhadap data aslinya.
3. Mengidentifikasi karakteristik Copula Frank.

4. Membangkitkan data dari distribusi binomial dan poisson.
5. Melakukan transformasi normal standar dan jitters pada data dengan tujuan agar data menjadi berdistribusi kontinu.
6. Mengamati perilaku penyebaran data hasil transformasi terhadap data yang aslinya.
7. Menganalisis nilai ukuran kebergantungan data hasil transformasi terhadap data yang aslinya
8. Mengkonstruksi distribusi bivariat dari data hasil transformasi menggunakan Copula Frank.
9. Membandingkan nilai peluang yang diperoleh dari data transformasi normal standar dan jitter terhadap data aslinya.

3. HASIL DAN PEMBAHASAN

3.1. Transformasi Normal Standar

Misalkan $\{x\}$ merupakan data diskrit berukuran n , $n \geq 30$. Data diskrit ini berasal dari marginal diskrit X . Untuk memperoleh X^* yang kontinu, maka dilakukan transformasi normal standar sebagai berikut:

$$x^* = \frac{x - \mu_X}{\sigma_X} \quad (1)$$

Adapun fungsi peluangnya, yaitu:

$$f^*(x^*) = \frac{e^{(-\frac{1}{2}x^{*2})}}{\sqrt{2\pi}} \quad (2)$$

dimana $X^* \sim N(0,1)$.

3.2. Jitters

Misalkan $\{x\}$ merupakan data diskrit yang berasal dari marginal diskrit X . Untuk memperoleh $X^{\textcircled{a}}$ yang kontinu, maka dilakukan transformasi jitters. Diperoleh:

$$X^{\textcircled{a}} = X - U \quad (3)$$

dimana $U \sim \text{Uniform}(0,1)$.

Adapun fungsi distribusinya, yaitu:

$$\begin{aligned} P(X^{\textcircled{a}} \leq x^{\textcircled{a}}) &= P(X - U \leq x^{\textcircled{a}}) \\ &= \sum_{k=0}^{\infty} P(X \leq x^{\textcircled{a}} + U | X = k) P(X = k) \\ &= \sum_{k=0}^{\infty} P(U > k - x^{\textcircled{a}}) P(X = k) \\ &= \sum_{k=0}^{\infty} \{1 - P(U \leq k - x^{\textcircled{a}})\} P(X = k) \\ &= \sum_{k=0}^{\infty} \{1 - P(U \leq k - x^{\textcircled{a}})\} P(X = k) \\ &= \sum_{k=0}^{\lfloor x^{\textcircled{a}} \rfloor} \{1 - P(U \leq k - x^{\textcircled{a}})\} P(X = k) + \sum_{k=\lfloor x^{\textcircled{a}} \rfloor + 1}^{\infty} \{1 - P(U \leq k - x^{\textcircled{a}})\} P(X = k) \\ &= \{P(X \leq \lfloor x^{\textcircled{a}} \rfloor) - (0 + 0 + \dots + 0)\} + (x^{\textcircled{a}} - \lfloor x^{\textcircled{a}} \rfloor) P(X = \lfloor x^{\textcircled{a}} \rfloor + 1) \end{aligned}$$

$$= P(X \leq [x^@]) + (x^@ - [x^@])P(X = [x^@ + 1])$$

Sehingga diperoleh:

$$\begin{aligned} F^@(x^@) &= P(X \leq [x^@]) + (x^@ - [x^@])P(X = [x^@ + 1]) \\ &= F([x^@]) + (x^@ - [x^@])f([x^@ + 1]) \end{aligned} \quad (4)$$

Fungsi peluangnya yang bersesuaian, yaitu:

$$f^@(x^@) = P(X = [x^@ + 1]) = f([x^@ + 1]) \quad (5)$$

3.3. Nilai Korelasi Pearson dan Kendall's Tau Data Transformasi Normal Standar

Hasil transformasi normal standar mampu mempresentasikan data aslinya. Hal ini dapat terlihat dari ukuran kebergantungannya. Ukuran kebergantungan yang digunakan, yaitu Korelasi Pearson ρ dan Kendall's Tau τ . Nilai korelasi Pearson ρ_{X^*,Y^*} sama dengan $\rho_{X,Y}$. Berikut pemaparannya:

$$\begin{aligned} \rho_{X^*,Y^*} &= \frac{\text{Cov}(X^*, Y^*)}{\sigma_{X^*}\sigma_{Y^*}} = \frac{\text{Cov}(X - \mu_X, Y - \mu_Y)}{\sigma_X\sigma_Y} \\ &= \frac{[\text{Cov}(X, Y) - \text{Cov}(X, \mu_Y) - \text{Cov}(Y, \mu_X) + \text{Cov}(\mu_X, \mu_Y)]}{\sigma_X\sigma_Y} = \frac{\text{Cov}(X, Y)}{\sigma_X\sigma_Y} \\ &= \rho_{X,Y} \end{aligned} \quad (6)$$

Demikian juga untuk nilai dari τ_{X^*,Y^*} sama dengan $\tau_{X,Y}$, seperti pemaparan berikut:

$$\begin{aligned} P_{X^*,Y^*}(\text{concordant}) &= P[(X^*_1 - X^*_2)(Y^*_1 - Y^*_2) > 0] \\ &= P\left[\left(\frac{X_1 - \mu_X}{\sigma_X} - \frac{X_2 - \mu_X}{\sigma_X}\right)\left(\frac{Y_1 - \mu_Y}{\sigma_Y} - \frac{Y_2 - \mu_Y}{\sigma_Y}\right) > 0\right] \\ &= P\left[\left(\frac{X_1 - X_2}{\sigma_X}\right)\left(\frac{Y_1 - Y_2}{\sigma_Y}\right) > 0\right] = P[(X_1 - X_2)(Y_1 - Y_2) > 0] \\ &= P_{X,Y}(\text{concordant}) \end{aligned} \quad (7)$$

Sehingga diperoleh:

$$\begin{aligned} \tau_{X^*,Y^*} &= P_{X^*,Y^*}(\text{concordant}) - P_{X^*,Y^*}(\text{discordant}) \\ &= P_{X^*,Y^*}(\text{concordant}) - 1 + P_{X^*,Y^*}(\text{discordant}) \\ &= 2P_{X^*,Y^*}(\text{concordant}) - 1 = 2P_{X,Y}(\text{concordant}) - 1 \\ &= \tau_{X,Y} \end{aligned} \quad (8)$$

Persamaan (6) dan (8) menunjukkan bahwa nilai Korelasi Pearson maupun Kendall's tau untuk data hasil transformasi dengan data aslinya adalah sama. Dengan kata lain, kebergantungan yang terjadi antar data yang hasil transformasi sama dengan kebergantungan yang terjadi antar data aslinya. Dapat dikatakan bahwa data hasil transformasi normal standar mampu mempresentasikan data aslinya.

3.4. Nilai Kendall's Tau Data Hasil Jitters

Data hasil jitters juga mampu mempresentasikan data aslinya. Hal ini dapat dibuktikan melalui dua cara. Cara pertama, dibuktikan dengan menunjukkan bahwa orde *concordant* antar data hasil jitters sama dengan orde *concordant* antar data slinya. Cara kedua, dibuktikan dengan membandingkan nilai dari $\tau_{X^@,Y^@}$ dan $\tau_{X,Y}$.

Untuk cara pertama, misalkan terdapat dua variabel acak X dan Y . Dapat dikatakan bahwa (X_2, Y_2) lebih *concordant* daripada (X_1, Y_1) , dinotasikan $(X_1, Y_1) \prec_c (X_2, Y_2)$, jika:

$$P(X_1 \leq s, Y_1 \leq t) \leq P(X_2 \leq s, Y_2 \leq t); \quad \forall s, t \in R \quad (9)$$

Misalkan X dan Y masing-masing dikontinukan menjadi $X^@$ dan $Y^@$ oleh U_1 dan U_2 melalui proses jitters. Jika $(X_1, Y_1) \prec_c (X_2, Y_2)$, maka diperoleh:

$$\begin{aligned}
P(X_1^{\otimes} \leq s, Y_1^{\otimes} \leq t) &= P(X_1 - U_1 \leq s, Y_1 - U_2 \leq t) \\
&= \iint_{u_1, u_2 \in [0,1]} P(X_1 \leq s + u_1, Y_1 \leq t + u_2) f_{U_1}(u_1) f_{U_2}(u_2) du_1 du_2 \\
&\leq \iint_{u_1, u_2 \in [0,1]} P(X_2 \leq s + u_1, Y_2 \leq t + u_2) f_{U_1}(u_1) f_{U_2}(u_2) du_1 du_2 \\
&= P(X_2^{\otimes} \leq s, Y_2^{\otimes} \leq t)
\end{aligned} \tag{10}$$

Sehingga,

$$(X_1, Y_1) <_c (X_2, Y_2) \Rightarrow (X_1^{\otimes}, Y_1^{\otimes}) <_c (X_2^{\otimes}, Y_2^{\otimes}) \tag{11}$$

Dari persamaan (11) terbukti bahwa jika (X_1, Y_1) lebih *concordant* daripada (X_2, Y_2) maka $(X_1^{\otimes}, Y_1^{\otimes})$ juga lebih *concordant* daripada $(X_2^{\otimes}, Y_2^{\otimes})$. Jadi, terbukti bahwa orde *concordant* antar data hasil jitters sama dengan orde *concordant* antar data yang aslinya.

Untuk cara kedua, misalkan X^{\otimes} dan Y^{\otimes} merupakan variabel acak kontinu hasil jitters oleh U_1 dan U_2 . Diperoleh bahwa:

$$\begin{aligned}
P^{\otimes}(\text{concordant}) &= P[(X_1^{\otimes} - X_2^{\otimes})(Y_1^{\otimes} - Y_2^{\otimes}) > 0] \\
&= P[(X_1 - U_{11} - X_2 + U_{12})(Y_1 - U_{21} - Y_2 + U_{22}) > 0] \\
&= P[X_1 = X_2, Y_1 = Y_2]P[(U_{12} - U_{11})(U_{22} - U_{21}) > 0] + \\
&\quad P[X_1 = X_2, Y_1 > Y_2]P[(U_{12} - U_{11}) > 0] + \\
&\quad P[X_1 = X_2, Y_1 < Y_2]P[(U_{12} - U_{11}) < 0] + \\
&\quad P[X_1 > X_2, Y_1 = Y_2]P[(U_{22} - U_{21}) > 0] + \\
&\quad P[X_1 < X_2, Y_1 = Y_2]P[(U_{22} - U_{21}) < 0] + \\
&\quad P[(X_1 - X_2)(Y_1 - Y_2) > 0]
\end{aligned} \tag{12}$$

Karena $U_{12} - U_{11}$ dan $U_{22} - U_{21}$ adalah variabel acak kontinu dengan peluang simetri sekitar nol, maka:

$$\begin{aligned}
P^{\otimes}(\text{concordant}) &= \frac{1}{2} P(\text{tie}) + P[(X_1 - X_2)(Y_1 - Y_2) > 0] \\
&= \frac{1}{2} P(\text{tie}) + P(\text{concordant})
\end{aligned} \tag{13}$$

Dengan demikian,

$$\begin{aligned}
\tau(X^{\otimes}, Y^{\otimes}) &= P^{\otimes}(\text{concordant}) - P^{\otimes}(\text{discordant}) = 2P^{\otimes}(\text{concordant}) - 1 \\
&= 2P(\text{concordant}) + P(\text{tie}) - 1 \\
&= P(\text{concordant}) + P(\text{tie}) - P(\text{discordant})
\end{aligned} \tag{14}$$

Jika $P(\text{tie}) = 0$ maka $\tau_{X^{\otimes}, Y^{\otimes}} = \tau_{X, Y}$. Dapat dikatakan bahwa hasil jitters mampu mempresentasikan nilai aslinya.

3.5. Copula Frank

Copula Archimedean merupakan salah satu jenis Copula yang dikonstruksi dari fungsi generator $\varphi(t)$, dimana t dengan support $[0,1]$. Adapun bentuk basis dari Copula Archimedean pada kasus bivariat, yaitu:

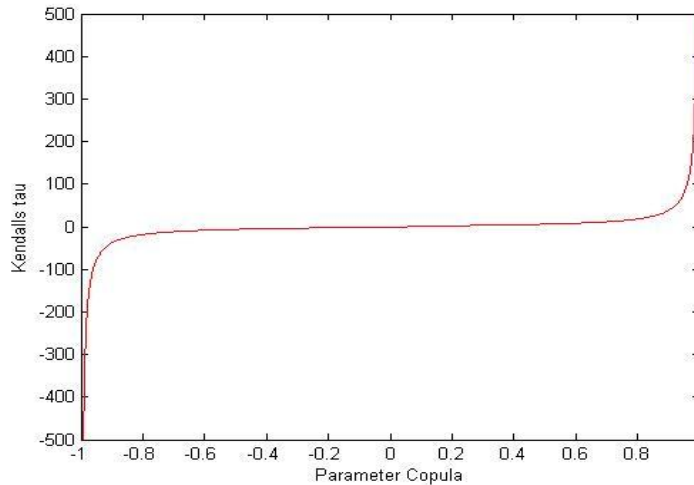
$$C(v, w) = \varphi^{-1}(\varphi(v) + \varphi(w)) \tag{15}$$

Salah satu jenis Copula Archimedean yang akan digunakan untuk mengkonstruksi distribusi bivariat pada penelitian ini, yaitu Copula Frank. Karakteristik dari Copula Frank disajikan pada tabel berikut.

Tabel 1. Karakteristik Copula Frank

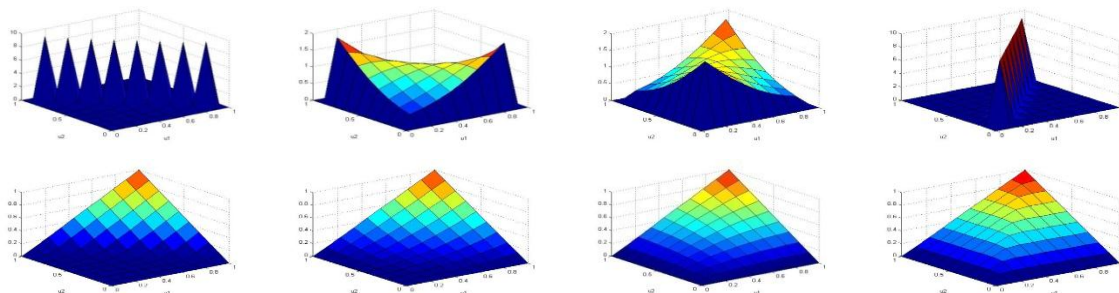
$C(v, w)$	$-\theta^{-1} \ln(1 + \frac{(e^{-\theta v} - 1)(e^{-\theta w} - 1)}{e^{-\theta} - 1})$
θ	$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt \quad (\theta \neq 0; \tau \neq 0)$
$\varphi(t)$	$-\ln(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1})$
$\varphi^{-1}(s)$	$-\theta^{-1} \ln(1 + (e^{-\theta} - 1)e^{-s})$
λ_{upper}	0
λ_{lower}	0

Tabel 1 menunjukkan bahwa Copula Frank terdefinisi untuk $\tau \in [-1, 1] \setminus 0$. Artinya, Copula ini mampu memodelkan ketergantungan positif maupun negatif. Informasi lainnya yang dapat diperoleh dari tabel tersebut, yaitu mengenai hubungan antara Kendall's tau (τ) dengan parameter Copula (θ). Nilai τ berbanding lurus dengan nilai θ . Artinya, semakin besar nilai θ maka semakin besar pula nilai τ dan sebaliknya. Hal ini diilustrasikan sebagai berikut:



Gambar 1. Ilustrasi Hubungan Antara Parameter Copula (θ) dan Kendall's Tau (τ)

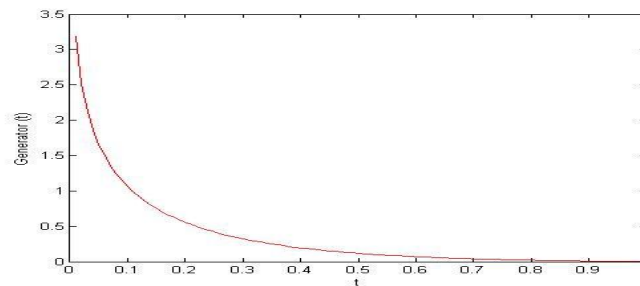
Selain itu, semakin besar nilai τ maka semakin besar pula nilai fungsi peluang dan fungsi distribusinya. Hal yang sama juga berlaku untuk nilai θ -nya karena θ berbanding lurus dengan nilai τ . Berikut ilustrasi dari fungsi peluang dan fungsi distribusi dengan nilai $\tau = -90; -30; 30; 90$ secara berturut-turut sebagai berikut:



Gambar 2. Visualisasi Fungsi Peluang dan Fungsi Distribusi Copula Frank dengan $\tau = -90; -30; 30; 90$

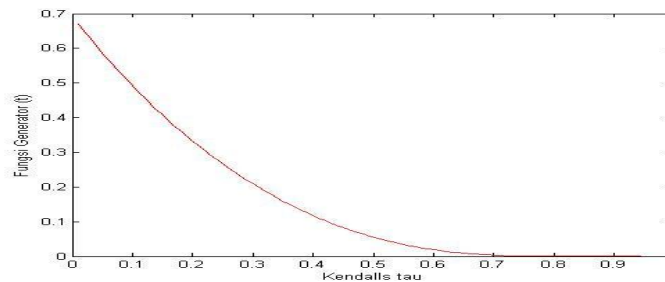
Dari gambar 2 terlihat bahwa Copula Frank memberikan peluang yang hampir sama pada nilai-nilai yang rendah maupun tinggi, dengan kata lain tidak ada kebergantungan ekor atas maupun ekor bawah. Tidak adanya kebergantungan ekor atas juga ditunjukkan oleh nilai $\lambda_{upper} = 0$. Sedangkan tidak adanya kebergantungan ekor bawah juga ditunjukkan oleh nilai $\lambda_{lower} = 0$. Sehingga Copula jenis ini cocok digunakan untuk memodelkan data yang tidak memiliki kebergantungan ekor atas maupun ekor bawah.

Lebih lanjut, adanya kaitan antara nilai t dengan fungsi generator $\varphi(t)$. Semakin besar nilai t menuju 1 maka semakin kecil nilai $\varphi(t)$ menuju 0 untuk suatu nilai τ atau θ tertentu. Hal ini berarti bahwa $\varphi(t)$ merupakan fungsi monoton turun yang terdefinisi untuk $t \in [0,1]$, dengan $\varphi(1) = 0$. Berikut ilustrasinya untuk $\tau = 0.4$:



Gambar 3. Ilustrasi Hubungan Antara t dengan Fungsi Generator untuk $\tau = 0.4$

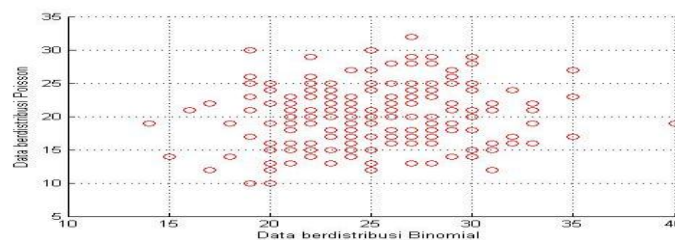
Lebih jauh, nilai τ berbanding terbalik dengan nilai fungsi generator $\varphi(t)$. Artinya, semakin besar nilai τ maka semakin kecil nilai $\varphi(t)$. Hal yang sama juga berlaku untuk nilai θ , karena θ berbanding lurus dengan τ . Berikut ilustrasinya:



Gambar 4. Ilustrasi Hubungan τ dengan Fungsi Generator

3.6. Simulasi

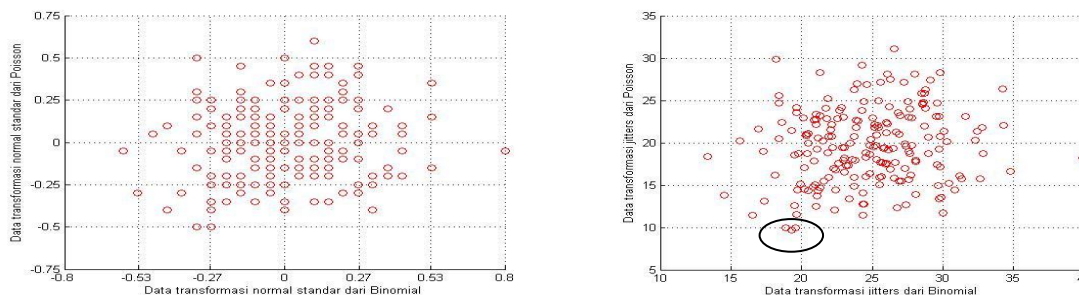
Data simulasi $\{x\}$ dan $\{y\}$ berasal dari dua distribusi diskrit yang berbeda. Distribusi dari $\{x\}$ dan $\{y\}$ berturut-turut adalah binomial (100,0.25) dan poisson (20). Masing-masing data berjumlah 200. Adapun penyebaran datanya dapat dilihat pada gambar berikut.



Gambar 5. Scatter Plot Data

Pada gambar 5 terlihat bahwa data menyebar di sekitar nilai (25,20). Sedangkan, ukuran kebergantungannya, yaitu $\rho = 0.146$ dan $\tau = 0.094$. Nilai kebergantungan yang diperoleh menunjukkan bahwa kedua jenis data memiliki kebergantungan positif yang relatif cukup kuat.

Kedua jenis data yang dimiliki berdistribusi marginal diskrit. Sehingga perlu dilakukan transformasi agar diperoleh distribusi marginal kontinu. Transformasi dilakukan dengan menggunakan pendekatan normal standar dan jitters. Adapun penyebaran data hasil transformasinya sebagai berikut



Gambar 6. Scatter Plot Data Transformasi Normal Standar dan Jitter (Kiri ke Kanan)

Gambar 6 (sebelah kiri) menunjukkan bahwa perilaku penyebaran data transformasi normal standar terlihat sama dengan perilaku penyebaran data aslinya pada Gambar 5. Ukuran kebergantungannya juga sama, yaitu $\rho^* = 0.146$ dan $\tau^* = 0.094$. Pada data transformasi normal standar, data menyebar di sekitar nilai $(0, 0)$. Nilai $(0, 0)$ ini mempresentasikan nilai $(25, 20)$ pada data aslinya. Hal yang sama juga berlaku untuk nilai lainnya, (x^*, y^*) mempresentasikan (x, y) . Dengan demikian, data transformasi normal standar mampu mempresentasikan data aslinya.

Gambar 6 (sebelah kanan) menunjukkan bahwa perilaku penyebaran data transformasi jitter terlihat agak sedikit berbeda dengan perilaku penyebaran data aslinya pada Gambar 5. Setelah dilakukan analisis lebih lanjut, ditemukan beberapa pasangan data yang bernilai sama, yaitu $x_i = x_j$ dan $y_i = y_j, \exists i \neq j$. Sedangkan, pada data hasil jitter tidak ditemukan lagi pasangan data yang sama, artinya $x_i^@ \neq x_j^@$ dan $y_i^@ \neq y_j^@, \forall i \neq j$.

Pada Gambar 5, semua pasangan data yang bernilai sama diplot oleh *scatter* yang saling tumpang tindih. Sehingga terlihat seolah-olah hanya diplot oleh satu *scatter*. Sedangkan pada Gambar 6, *scatter* pasangan data yang awalnya bernilai sama tersebut menjadi berbeda karena nilainya juga menjadi berbeda. Oleh sebab itu, *scatter* yang awalnya tumpang tindih jadi sedikit bergeser. Salah satu contohnya adalah data yang dilingkari pada masing-masing Gambar 5 dan Gambar 6. Pada Gambar 5, pasangan data yang dilingkari adalah pasangan data ke-31 yang bernilai $(20, 10)$ dan pasangan data ke-141 yang bernilai sama $(20, 10)$. Kedua pasangan data ini seolah-olah hanya diplot oleh satu *scatter*, tetapi sebenarnya *scatter*-nya saling tumpang tindih. Selanjutnya, pada Gambar 6, pasangan data yang dilingkari adalah pasangan data ke-31 yang menjadi bernilai $(19.561, 9.937)$ dan pasangan data ke-141 yang menjadi bernilai $(19.332, 9.703)$. Sehingga, *scatter* dari dua pasangan data ini tidak lagi sama atau tumpang tindih, tetapi saling bergeser (menjauh satu sama lain). Perbedaan ini terjadi karena nilai $u_{1i} \neq u_{1j}$ dan $u_{2i} \neq u_{2j}, \forall i \neq j$. Oleh sebab itu, pada Gambar 5 untuk daerah yang dilingkari hanya terlihat dua *scatter*. Sedangkan pada Gambar 6 untuk daerah yang dilingkari menjadi tiga *scatter* karena dua *scatter* yang awalnya saling tumpang tindih menjadi saling menjauh. Selain dua pasangan data ke-31 dan ke-141, yaitu $(20, 10)$, masih beberapa pasangan data lainnya yang juga bernilai sama. Banyaknya pasangan data yang awalnya bernilai sama, ketika ditransformasikan oleh proses jitter menjadi sedikit berbeda inilah yang menyebabkan penyebaran data hasil jitter pada Gambar 6 menjadi sedikit berbeda dari penyebaran data aslinya pada Gambar 5. Seandainya $x_i \neq x_j$ dan $y_i \neq y_j, \forall i \neq j$, maka penyebaran data hasil jitters akan mirip dengan penyebaran data aslinya.

Lebih lanjut, data transformasi jitter memiliki nilai $\tau^@ = 0.092$. Nilai antara $\tau^@$ dan τ sedikit berbeda. Perbedaan tersebut disebabkan oleh adanya data $x_i = x_{i+1}$ (atau $y_j = y_{j+1}$). Berdasarkan persamaan (8), menentukan nilai Kendall's tau berbasis pada konsep *concordance*, yaitu dengan menentukan banyaknya pasangan yang *concordant* dan *discordant*. Dapat dikatakan bahwa (x_i, x_{i+1}) dan (y_j, y_{j+1}) adalah *concordant* jika $(x_i - x_{i+1})(y_j - y_{j+1}) > 0$. Sedangkan (x_i, x_{i+1}) dan (y_j, y_{j+1}) adalah *discordant* jika $(x_i - x_{i+1})(y_j - y_{j+1}) < 0$. Artinya, untuk data $x_i = x_{i+1}$ (atau $y_j = y_{j+1}$) tidak

terhitung sebagai pasangan yang *concordant* maupun *discordant*. Namun, setelah proses jitters, data $x_i = x_{i+1}$ (atau $y_j = y_{j+1}$) menjadi $x_i^{\textcircled{a}} \neq x_{i+1}^{\textcircled{a}}$ (atau $y_j^{\textcircled{a}} \neq y_{j+1}^{\textcircled{a}}$). Data $x_i^{\textcircled{a}} \neq x_{i+1}^{\textcircled{a}}$ (atau $y_j^{\textcircled{a}} \neq y_{j+1}^{\textcircled{a}}$) kemudian akan terhitung sebagai pasangan *concordant* atau *discordant*. Hal ini tentu menyebabkan terjadinya perubahan banyaknya pasangan yang *concordant* maupun *discordant* pada data hasil jitter terhadap data aslinya. Akibatnya, nilai $\tau^{\textcircled{a}}$ -nya menjadi berubah.

Pada penelitian ini, data $x_i = x_{i+1}$ (atau $y_j = y_{j+1}$) sebanyak 21. Setelah melalui transformasi jitter, dari 21 data tersebut 12 data menjadi data *concordant* dan 9 data menjadi *discordant*. Banyaknya data yang *concordant* dan *discordant* tersebut saling meniadakan sehingga tersisa 2 data yang *concordant*. Artinya, data *concordant* pada hasil jitters lebih banyak 2 dibandingkan data aslinya. Perubahan banyaknya data yang *concordant* pada data jitters inilah yang menyebabkan nilai $\tau^{\textcircled{a}}$ dan τ berbeda. Meskipun demikian, perbedaannya tidaklah signifikan. Sehingga, masih dapat dikatakan bahwa data transformasi jitter mampu mempresentasikan data aslinya.

Langkah selanjutnya, yaitu menentukan fungsi distribusi dan fungsi peluang dari data transformasi. Fungsi peluang dari data transformasi normal standar didefinisikan oleh Persamaan (2), sedangkan fungsi distribusinya merupakan kumulatif dari fungsi peluangnya. Adapun fungsi peluang dari data transformasi jitter didefinisikan oleh Persamaan (5), sedangkan fungsi distribusinya didefinisikan oleh Persamaan (4).

3.7. Mengkonstruksi Distribusi Bivariat Copula Frank Hasil Transformasi

Berdasarkan Tabel 1, fungsi distribusi bivariat Copula Frank didefinisikan sebagai berikut:

$$C(v, w) = -\theta^{-1} \ln\left(1 + \frac{(e^{-\theta v} - 1)(e^{-\theta w} - 1)}{e^{-\theta} - 1}\right) \quad (16)$$

Fungsi peluangnya yang bersesuaian, yaitu:

$$c(v, w) = \frac{-\theta e^{-\theta(v+w)}(e^{-\theta} - 1)}{(e^{-\theta} - 1 + (e^{-\theta v} - 1)(e^{-\theta w} - 1))^2} \quad (17)$$

Adapun pemodelan kebergantungan pada fungsi distribusi bivariat Copula Frank dari data transformasi normal standar maupun jitters, sebagai berikut:

$$C^*(v^*, w^*) = (-0.8541)^{-1} \ln\left(1 + \frac{(e^{-0.8541v^*} - 1)(e^{-0.8541w^*} - 1)}{e^{-0.8541} - 1}\right) \quad (18)$$

$$C^{\textcircled{a}}(v^{\textcircled{a}}, w^{\textcircled{a}}) = (-0.8287)^{-1} \ln\left(1 + \frac{(e^{-0.8287v^{\textcircled{a}}} - 1)(e^{-0.8287w^{\textcircled{a}}} - 1)}{e^{-0.8287} - 1}\right) \quad (19)$$

Fungsi peluangnya yang bersesuaian, yaitu:

$$c^*(v^*, w^*) = \frac{-0.8541 e^{-0.8541(v^*+w^*)}(e^{-0.8541} - 1)}{(e^{-0.8541} - 1 + (e^{-0.8541v^*} - 1)(e^{-0.8541w^*} - 1))^2} \quad (20)$$

$$c^{\textcircled{a}}(v^{\textcircled{a}}, w^{\textcircled{a}}) = \frac{-0.8287 e^{-0.8287(v^{\textcircled{a}}+w^{\textcircled{a}})}(e^{-0.8287} - 1)}{(e^{-0.8287} - 1 + (e^{-0.8287v^{\textcircled{a}}} - 1)(e^{-0.8287w^{\textcircled{a}}} - 1))^2} \quad (21)$$

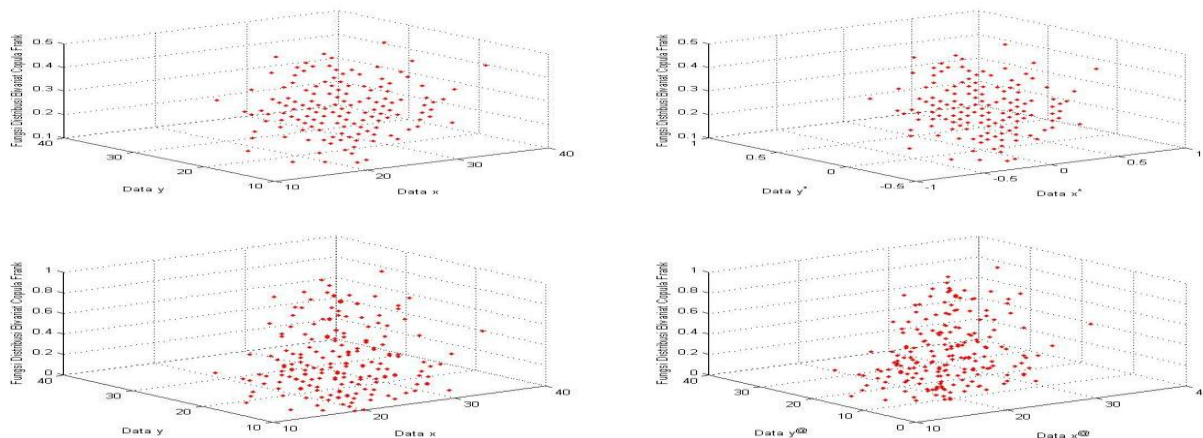
dimana: $v^* = F^*_{x^*}$ dan $w^* = F^*_{y^*}$; $v^{\textcircled{a}} = F^{\textcircled{a}}_{x^{\textcircled{a}}}$ dan $w^{\textcircled{a}} = F^{\textcircled{a}}_{y^{\textcircled{a}}}$

Dapat dikatakan bahwa $F(x, y) = C^*(v^*, w^*)$ atau $F(x, y) = C^{\textcircled{a}}(v^{\textcircled{a}}, w^{\textcircled{a}})$. Artinya, nilai fungsi distribusi bivariat Copula Frank dari hasil transformasi normal standar dapat mempresentasikan fungsi distribusi bivariat dari data aslinya. Demikian juga nilai fungsi distribusi bivariat Copula Frank dari hasil jitters dapat mempresentasikan fungsi distribusi bivariat dari data aslinya. Hal ini diilustrasikan melalui 17 data simulasi pada tabel berikut.

Tabel 2. Data Asli, Data Hasil Transformasi Normal Standar, Data Hasil Jitters, Fungsi Distribusi Bivariat Copula Frank Hasil Transformasi Normal Standar dan Fungsi Distribusi Bivariat Copula Frank Hasil Jitters

x	y	x^*	y^*	$x^{@}$	$y^{@}$	$C^*(v^*, w^*)$	$C^{@}(v^{@}, w^{@})$
29	26	0.213333	0.3	28.71313	25.86202	0.385577	0.769576
29	18	0.213333	-0.1	28.10464	17.77236	0.294472	0.304143
30	24	0.266667	0.2	29.55941	23.10503	0.375321	0.702766
29	26	0.213333	0.3	28.4568	25.86445	0.385577	0.75631
30	21	0.266667	0.05	29.81198	20.7636	0.339936	0.563449
29	25	0.213333	0.25	28.55535	24.63553	0.374762	0.725841
30	25	0.266667	0.25	29.59641	24.692	0.386763	0.772453
29	27	0.213333	0.35	28.76573	26.27283	0.396154	0.781483
20	19	-0.26667	-0.05	19.77843	18.7111	0.21486	0.07348
26	16	0.053333	-0.2	25.49124	15.07945	0.245067	0.109606
28	20	0.16	0	27.37321	19.62963	0.30784	0.413216
26	18	0.053333	-0.1	25.57648	17.65583	0.266134	0.235201
31	12	0.32	-0.4	30.02779	11.67359	0.23769	0.030937
22	23	-0.16	0.15	21.00788	22.12007	0.269871	0.167557
24	19	-0.05333	-0.05	23.02156	18.16999	0.256229	0.171094

Dari Tabel 2, diperoleh informasi bahwa $F(29,18) = 0.294472$ jika menggunakan transformasi normal standar. Namun, jika menggunakan jitters maka $F(29,18) = 0.304143$. Selain itu, dari tabel ini juga terlihat bahwa untuk $i < j$, jika $x_i > x_j$ maka $x_i^* > x_j^*$ dan $x_i^{@} > x_j^{@}$. Hal ini menunjukkan bahwa perilaku data asli, data hasil transformasi normal standar dan data hasil jitters adalah sama. Selain itu, untuk $i < j$, jika $C^*(v^*_i, w^*_i) > C^*(v^*_j, w^*_j)$ maka $C^{@}(v^{@}_i, w^{@}_i) > C^{@}(v^{@}_j, w^{@}_j)$. Hal ini juga menunjukkan bahwa nilai fungsi distribusi dari data hasil transformasi normal standar dan jitters memiliki perilaku yang sama. Hal yang sama juga ditunjukkan oleh Gambar 7.



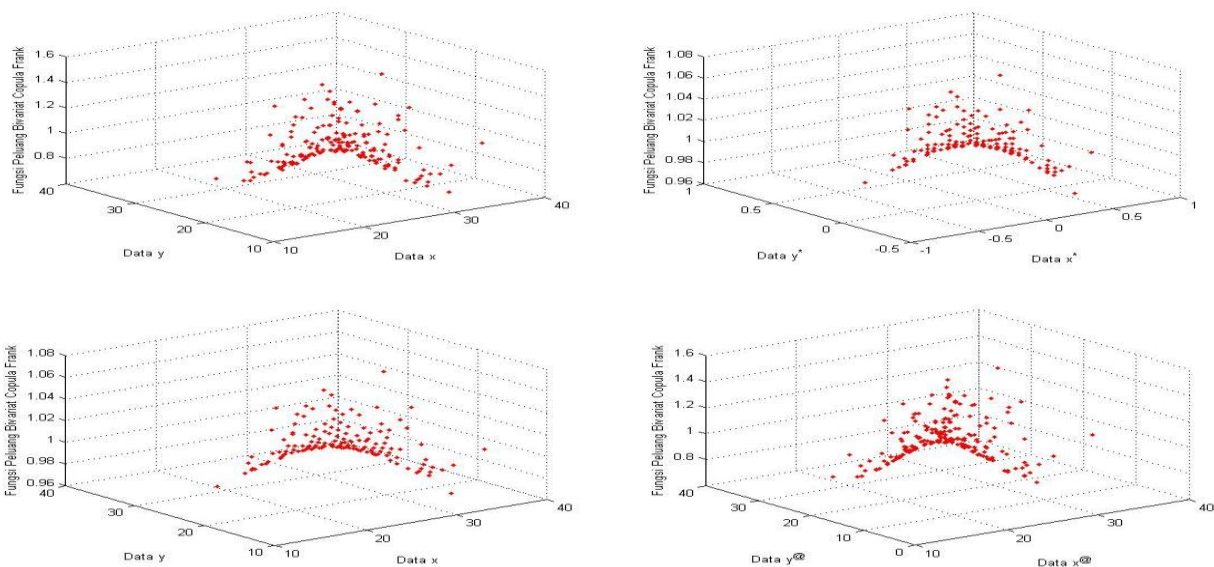
Gambar 7. Plot (x, y) Terhadap $C^*(v^*, w^*)$, (x^*, y^*) Terhadap $C^*(v^*, w^*)$, (x, y) Terhadap $C^{@}(v^{@}, w^{@})$, $(x^{@}, y^{@})$ Terhadap $C^{@}(v^{@}, w^{@})$ (Kiri ke Kanan, atas ke Bawah)

Selain itu, dapat juga dikatakan bahwa $f(x, y) = c^*(v^*, w^*)$ atau $f(x, y) = c^{@}(v^{@}, w^{@})$. Artinya, nilai fungsi peluang bivariat Copula Frank dari hasil transformasi normal standar dapat mempresentasikan fungsi peluang bivariat dari data aslinya. Demikian juga nilai fungsi peluang bivariat Copula Frank dari hasil jitters dapat mempresentasikan fungsi peluang bivariat dari data aslinya. Hal ini diilustrasikan melalui 17 data simulasi pada tabel berikut.

Tabel 3. Data Asli, Data Hasil Transformasi Normal Standar, Data Hasil Jitters, Fungsi Peluang Bivariat Copula Frank Hasil Transformasi Normal Standar dan Fungsi Peluang Bivariat Copula Frank Hasil Jitters

x	y	x^*	y^*	$x^{@}$	$y^{@}$	$c^*(v^*, w^*)$	$c^{@}(v^{@}, w^{@})$
29	26	0.213333	0.3	28.71313	25.86202	1.028939	1.233902
29	18	0.213333	-0.1	28.10464	17.77236	1.007656	0.930599
30	24	0.266667	0.2	29.55941	23.10503	1.026718	1.182034
29	26	0.213333	0.3	28.4568	25.86445	1.028939	1.221946
30	21	0.266667	0.05	29.81198	20.7636	1.016726	1.069659
29	25	0.213333	0.25	28.55535	24.63553	1.026811	1.201108
30	25	0.266667	0.25	29.59641	24.692	1.029736	1.239156
29	27	0.213333	0.35	28.76573	26.27283	1.030906	1.243241
20	19	-0.26667	-0.05	19.77843	18.7111	1.016726	1.02528
26	16	0.053333	-0.2	25.49124	15.07945	1.010951	0.940828
28	20	0.16	0	27.37321	19.62963	1.014406	1.0145
26	18	0.053333	-0.1	25.57648	17.65583	1.013287	0.983019
31	12	0.32	-0.4	30.02779	11.67359	0.974626	0.716131
22	23	-0.16	0.15	21.00788	22.12007	1.007122	0.890537
24	19	-0.05333	-0.05	23.02156	18.16999	1.015743	1.032247

Dari Tabel 3, diperoleh informasi bahwa $f(29,18) = 1.007656$ jika menggunakan transformasi normal standar. Namun, jika menggunakan jitters maka $f(29,18) = 0.930599$. Selain itu, dari Tabel 3 juga terlihat bahwa untuk $i < j$, jika $c^*(v^*_i, w^*_i) > c^*(v^*_j, w^*_j)$ maka $c^{@}(v^{@}_i, w^{@}_i) > c^{@}(v^{@}_j, w^{@}_j)$. Hal ini berarti bahwa nilai peluang dari data hasil transformasi normal standar dan jitters memiliki perilaku yang sama. Hal yang sama juga ditunjukkan oleh Gambar 8.



Gambar 8. Plot (x, y) Terhadap $c^*(v^*, w^*)$, (x^*, y^*) Terhadap $c^*(v^*, w^*)$, (x, y) Terhadap $c^{@}(v^{@}, w^{@})$, $(x^{@}, y^{@})$ Terhadap $c^{@}(v^{@}, w^{@})$ (Kiri ke Kanan, Atas ke Bawah)

4. KESIMPULAN

Berdasarkan hasil analisis diperoleh bahwa model fungsi peluang bivariat Copula Frank dan fungsi distribusi bivariat Copula Frank yang diperoleh dari hasil transformasi normal standar dan jitters mampu mempresentasikan fungsi peluang bivariat dan fungsi distribusi bivariat dari data aslinya, terutama dalam memodelkan sifat kebergantungannya.

DAFTAR PUSTAKA

- [1] Genest, C., & Neslehova, J., *A Primer on Copulas for Count Data*. ASTIN Bulletin 37(2), 475-515: 2007
- [2] Guegan, D., & Ladoucette, S.A., *Dependence Modelling of The Joint Extremes in a Portfolio Using Archimedean Copulas: Application to MSCI indices*. Cahiers de la Maison des Science Economiques: 2004
- [3] Jogde, K., *Concepts of Dependence, in Encyclopedia of Statistical Sciences, Vol.1, S. Kotz dan N.L, Johnson*. John Wiley & Sons, New York: 1982
- [4] Madsen, L., & Fang, Y., *Joint Regression Analysis for Discrete Longitudinal Data*. Biometrics, 67(3), 1175-1176: 2010ASTIN
- [5] Nelsen, B., *An Introduction to Copulas*. Springer: 2006
- [6] Rice, J.A., *Mathematical Statistic and Data Analysis*. Duxbury Press: 1995
- [7] Sekaran, Uma., *Metode Penelitian Bisnis*. Salemba Empat, Jakarta: 2006
- [8] Shi, P., & Valdez, E.A., *Longitudinal Modeling of Insurance Claim Counts Using Jitters*. Scandinavian Actuarial Journal: 2012