TEKNIK MENGONSTRUKSI DISTRIBUSI BIVARIAT COPULA CLAYTON PADA DATA MARGINAL DISKRIT DENGAN IMPLIKASI KEBERGANTUNGAN

¹⁾ Andi Fitriawati, ²⁾ Werry Febrianti, ³⁾ Ariestha Widyastuty Bustan, ⁴⁾ Amris
^{1,2)} Institut Teknologi Sumatera, ³⁾ Universitas Pasifik Morotai, ⁴⁾ IAIN Fattahul Muluk Papua
¹⁾ andi.fitriawati@at.itera.ac.id, ²⁾ werry.febrianti@ma.itera.ac.id,
³⁾ ariesthawidyastutybustan@gmail.com, ⁴⁾ amrispoenya@gmail.com

Abstract

Received :
28/06/2020

Accepted :
11/08/2020

Published :
11/08/2020

If we are having two types of data, $\{x\}$ and $\{y\}$, which are come from two discrete marginal that are different, then the interesting things that are known such the value of $f_{X,Y}$ and/or $F_{X,Y}$. It means that their bivariate distribution needs to be constructed. In constructing the bivariate distribution, the information about dependency of data is needed. We can use Copula to construct the bivariate distribution when the data has a dependency. Copula is one of the popular tools that used to construct bivariate or multivariate distributions with the implications of dependency. However, based on the Sklar theorem that when the data comes from a discrete marginal distribution, constructing the Copula bivariate distribution directly will produce a Copula C that is not unique. So that, it will result in unclear interpretation, especially in the dependency. Therefore, we need certain techniques to construct Copula bivariat distributions on discrete marginal data. The idea is to make the discrete marginal distribution be continuous through jitters transformation. Let X^* dan Y^* are results of jitters, then F_{X^*} dan F_{Y^*} are continous. So that, we can construct Copula bivariate distribution by the results of jitters. We use Clayton Copula. Furthermore, we construct the bivariate distribution by Clayton Copula, such as $F_{X,Y} = C_{v^*,w^*}^* = \max\left(\left(v^{*- \theta} + w^{*- \theta} - 1\right)^{-\theta^{-1}}, 0\right)$ and $f_{X,Y} = c_{v^*,w^*}^* = (\theta + 1)(v^*w^*)^{-(\theta+1)}(v^{*- \theta} + w^{*- \theta} - 1)^{-(2\theta+1)\theta^{-1}}$, with $v^* = F_{X^*}^*$ dan $w^* = F_{Y^*}^*$.

Keywords: Dependency, discrete marginal, Clayton Copula, jitters.

Abstrak

Ketika memiliki dua jenis data $\{x\}$ dan $\{y\}$ yang berasal dari dua marginal diskrit yang berbeda F_X dan F_Y , maka hal menarik yang ingin diketahui adalah nilai dari $f_{X,Y}$ dan/atau $F_{X,Y}$. Berkaitan dengan hal ini, maka perlu dikonstruksi distribusi bivariatnya. Dalam mengkonstruksi distribusi bivariat, diperlukan informasi mengenai adanya kebergantungan yang dimiliki oleh data. Jika data memiliki kebergantungan, maka mengkonstruksi distribusi bivariatnya dapat menggunakan Copula. Copula merupakan salah satu alat populer yang digunakan untuk mengkonstruksi distribusi bivariat maupun multivariat dengan implikasi kebergantungan. Namun, perlu diperhatikan bahwa mengkonstruksi distribusi bivariat Copula dari marginal diskrit akan menghasilkan Copula C yang tidak unik. Sehingga akan menghasilkan interpretasi yang tidak jelas, terutama sifat kebergantungannya. Hal ini sebagaimana yang dinyatakan dalam teorema Sklar. Oleh sebab itu, perlu adanya teknik tertentu dalam mengkonstruksi distribusi bivariat Copula pada data marginal diskrit. Idennya, dengan mengkontinukan distribusi marginalnya melalui transformasi jitters. Misalkan diperoleh X^* dan Y^* yang merupakan hasil jitters maka $F_{X^*}^*$ dan $F_{Y^*}^*$ masing-masing kontinu. Hasil jitters inilah yang kemudian digunakan untuk mengkonstruksi distribusi bivariatnya menggunakan Copula. Copula yang digunakan adalah Copula Clayton. Menggunakan Copula Clayton maka dikonstruksi distribusi bivariatnya sehingga diperoleh $F_{X,Y} = C_{v^*,w^*}^* = \max\left(\left(v^{*- \theta} + w^{*- \theta} - 1\right)^{-\theta^{-1}}, 0\right)$ dan $f_{X,Y} = c_{v^*,w^*}^* = (\theta + 1)(v^*w^*)^{-(\theta+1)}(v^{*- \theta} + w^{*- \theta} - 1)^{-(2\theta+1)\theta^{-1}}$, dimana $v^* = F_{X^*}^*$ dan $w^* = F_{Y^*}^*$.

Kata Kunci: Copula Clayton, jitters, kebergantungan, marginal diskrit.

1. Pendahuluan

Di era industri 4.0 saat ini, data memiliki peranan yang sangat penting dalam berbagai aspek kehidupan. Secara etimologis, kata “data” berasal dari bahasa latin “datum” yang artinya sesuatu yang diberikan. Menurut Setiawan dan Munir (2006), data adalah nilai yang merepresentasikan deskripsi dari suatu objek atau kejadian (*event*). Hal ini juga sesuai dengan pernyataan Vercellis (2009), yaitu “*data represent a structured codification of single primary entities, as well as of transactions involving two or more primary entities.*” Dengan demikian, dapat dikatakan bahwa data mempresentasikan fakta-fakta, baik angka, gambar, simbol dan sebagainya yang terstruktur dari suatu percobaan atau kejadian.

Data yang dimiliki akan menjadi semakin bermakna ketika diolah. Hasil pengolahan data berupa informasi. Sebagaimana yang dikemukakan oleh Vercellis (2009) bahwa “*information is the outcome of extraction and processing activities carried out on data, and it appears meaningful for those who receive it in a specific domain.*” Data dan informasi ini memiliki peranan yang sangat penting, baik dalam membuat perencanaan, mengambil keputusan, melakukan evaluasi dan sebagainya. Misalnya, pemerintah Indonesia perlu memiliki data terkait penyebaran penderita positif COVID-19 dalam membuat kebijakan.

Lebih lanjut, ketika memiliki dua jenis data maka hal menarik yang ingin diketahui adalah peluang kedua jenis data tersebut dapat terjadi secara serentak/bersamaan. Hal ini berarti bahwa perlu dikonstruksi distribusi bivariatnya, baik fungsi peluang maupun fungsi peluang kumulatif (fungsi distribusi). Untuk mengkonstruksi distribusi bivariat, perlu diketahui distribusi marginal dari masing-masing data dan sifat kebergantungannya. Adanya informasi mengenai kebergantungan pada data akan mempengaruhi teknik yang digunakan dalam mengkonstruksi distribusi bivariatnya. Jika kedua jenis data tidak memiliki kebergantungan, maka mengkonstruksi distribusi bivariatnya dengan mengalikan masing-masing distribusi marginalnya. Sedangkan, jika kedua jenis data memiliki kebergantungan, maka mengkonstruksi distribusi bivariatnya dapat menggunakan Copula. Copula merupakan alat populer yang digunakan untuk mengkonstruksi distribusi bivariat maupun multivariat yang memiliki kebergantungan. Dalam mengkonstruksi distribusi bivariat Copula maka diperlukan distribusi marginal dari data. Misalkan $\{x\}$ dan $\{y\}$ merupakan dua jenis himpunan data yang berasal dari variabel acak X dan Y dengan fungsi distribusi berturut-turut adalah F_X dan F_Y . Dengan melakukan transformasi pada setiap variabel acaknya, melalui fungsi distribusinya, menjadi variabel acak baru yang berdistribusi uniform standar maka diperoleh distribusi bivariatnya adalah $C(v, w) = P(V \leq v, W \leq w)$; dimana $v = F_X$ dan $w = F_Y$ serta v, w dengan support $[0, 1]$.

Namun, perlu diperhatikan bahwa jika data berasal dari marginal diskrit maka mengkonstruksi distribusi bivariat menggunakan Copula akan menimbulkan interpretasi yang tidak jelas, terutama sifat kebergantungannya. Berdasarkan teorema Sklar bahwa jika diberikan suatu fungsi distribusi bivariat H dengan marginal kontinu, F_X dan F_Y , maka akan terdapat suatu Copula C yang unik. Dengan kata lain, teorema ini hanya menjamin keunikan suatu Copula pada marginal kontinu tetapi tidak pada

marginal diskrit. Hal yang sama juga dinyatakan oleh Genest dan Néslehová (2007) bahwa Copula yang berasal dari marginal diskrit tidak unik.

Oleh sebab itu diperlukan suatu teknik untuk mengkonstruksi distribusi bivariat Copula dari marginal diskrit. Idenya, yaitu mengkontinukan distribusi marginal diskritnya melalui transformasi jitters. Hasil jitters inilah yang kemudian digunakan untuk mengkonstruksi distribusi bivariat Copula. Adapun jenis Copula yang akan digunakan adalah Copula Clayton.

2. Metode Penelitian

Penelitian ini dilakukan dengan menggunakan data simulasi dari distribusi binomial (100, 0.15) dan poisson (20) yang masing-masing berukuran 200. Ukuran data ini disesuaikan dengan pernyataan Roscoe (1975) yang dikutip Sekaran (2006) bahwa ukuran data sampel minimum adalah 30 data. Adapun beberapa variabel yang digunakan dalam penelitian ini, sebagai berikut:

X : Data berdistribusi binomial (100, 0.15)

Y : Data berdistribusi poisson (20)

X* : Data hasil jitters dari X

Y* : Data hasil jitters dari Y

Prosedur penelitian ini meliputi beberapa tahap diantaranya tahap mengeksplorasi karakteristik Copula Clayton, mengidentifikasi model jitters, melakukan uji nilai Kendall's tau dari data hasil transformasi terhadap data aslinya, membangkitkan data dari distribusi binomial (100, 0.15) dan poisson (20), melakukan transformasi jitters pada data asli agar data berdistribusi kontinu, mengamati perilaku penyebaran data hasil jitters terhadap data yang aslinya, menganalisis nilai ukuran kebergantungan data hasil jitters terhadap data aslinya, mengkonstruksi distribusi bivariat dari data hasil jitters menggunakan Copula Clayton.

3. Hasil dan Pembahasan

Dalam mengkonstruksi distribusi bivariat dari data marginal diskrit yang memiliki kebergantungan menggunakan Copula Clayton, diperlukan suatu teknik. Idenya dimulai dengan mengkontinukan data marginal diskrit melalui transformasi jitters. Hasil jitters harus mampu mempresentasikan data aslinya, baik dari perilaku penyebaran datanya maupun ukuran kebergantungannya. Ukuran kebergantungan yang akan digunakan adalah Kendall's tau. Selanjutnya, hasil jitters inilah yang akan digunakan untuk mengkonstruksi distribusi bivariat Copula Clayton. Semua proses mengkonstruksi distribusi bivariat Copula Clayton ini akan diilustrasikan melalui data simulasi yang dibangkitkan dari distribusi binomial dan poisson yang diasumsikan memiliki kebergantungan.

3.1. Copula Clayton

Copula Archimedean merupakan salah satu jenis Copula yang dikonstruksi dari fungsi generator $\varphi(t)$, dimana $t \in [0,1]$ (Nelsen, 2006). Adapun bentuk basis dari Copula Archimedean pada kasus bivariat, yaitu:

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

1)

Salah satu jenis Copula Archimedean yang akan digunakan untuk mengkonstruksi distribusi bivariat pada penelitian ini, yaitu Copula Clayton. Karakteristik dari Copula Clayton disajikan pada Tabel 1.

Tabel 1. Karakteristik Copula Clayton

$C(v, w)$	$\max\left((v^{-\theta} + w^{-\theta} - 1)^{-\theta^{-1}}, 0\right)$
θ	$\frac{2\tau}{1-\tau} \quad (\theta > 0; 0 < \tau \leq 1)$
$\varphi(t)$	$\frac{1}{\theta}(t^{-\theta} - 1)$
$\varphi^{-1}(s)$	$(1 + \theta s)^{-\theta^{-1}}$
λ_{upper}	0
λ_{lower}	$2^{-\theta^{-1}}$

Tabel 1 menunjukkan bahwa Copula Clayton terdefinisi untuk $\tau \in (0, 1]$. Artinya, Copula ini hanya mampu memodelkan kebergantungan positif. Informasi lainnya yang dapat diperoleh dari tabel tersebut, yaitu adanya hubungan antara Kendall's tau (τ) dengan parameter Copula (θ). Semakin besar nilai τ maka semakin besar pula nilai θ , begitupun sebaliknya. Hal ini diilustrasikan pada Gambar 1.

Gambar 1. Ilustrasi hubungan antara Kendall's tau (τ) dan parameter Copula (θ)

Informasi sebelumnya menyatakan bahwa semakin besar nilai τ maka semakin besar pula nilai θ . Dari tabel terlihat bahwa semakin besar nilai θ maka semakin besar pula nilai fungsi peluang maupun fungsi distribusinya. Sehingga, dapat dikatakan juga bahwa semakin besar nilai τ maka semakin besar pula nilai fungsi peluang maupun fungsi distribusinya. Berikut ilustrasi dari fungsi peluang dan fungsi distribusi dengan nilai $\tau = 0.15; 0.4; 0.65; 0.9$ secara berturut-turut sebagai berikut:

Gambar 2. Visualisasi fungsi peluang (atas) dan fungsi distribusi (bawah) Copula Clayton dengan $\tau = 0.15; 0.4; 0.65; 0.9$ (kiri ke kanan)

Dari gambar 2 terlihat bahwa Copula Clayton memberikan peluang yang hampir besar pada nilai-nilai yang rendah. Dengan kata lain, adanya kebergantungan ekor bawah. Kebergantungan ekor ini juga dapat dilihat dari nilai λ_{upper} dan λ_{lower} . Copula memiliki kebergantungan ekor bawah jika $\lambda_{\text{lower}} > 0$ dan memiliki kebergantungan ekor atas $\lambda_{\text{upper}} > 0$. Dari tabel 1 terlihat bahwa Copula Clayton memiliki nilai $\lambda_{\text{upper}} = 0$ dan $\lambda_{\text{lower}} = 2^{-\theta^{-1}}$, artinya Copula ini memiliki kebergantungan ekor bawah.

Informasi lainnya yang dapat diperoleh dari tabel 1, yaitu pada fungsi generatornya, semakin besar nilai t menuju 1 maka semakin kecil nilai $\phi(t)$ menuju 0 untuk suatu nilai τ atau θ tertentu. Hal ini menunjukkan bahwa $\phi(t)$ merupakan fungsi monoton turun yang terdefinisi untuk $t \in [0,1]$, dengan $\phi(1) = 0$. Berikut ilustrasinya untuk $\tau = 0.25$:

Gambar 3. Ilustrasi hubungan antara t dengan $\phi(t)$ untuk $\tau = 0.25$

Lebih lanjut, semakin besar nilai τ maka semakin besar pula nilai $\phi(t)$. Berikut ilustrasinya untuk $t = 0.5$:

Gambar 4. Ilustrasi hubungan τ dengan $\phi(0.5)$

3.2. Pengaruh Teorema Sklar pada Data Marginal Diskrit

Teorema Sklar merupakan suatu teori dasar dari Copula, yang menjelaskan peran Copula dalam membentuk suatu distribusi bivariat. Teorema ini menyatakan bahwa jika diberikan suatu fungsi distribusi bivariat dari dua fungsi distribusi marginal yang kontinu, maka terdapat suatu Copula C yang unik. Artinya, teorema ini hanya menjamin keunikan suatu Copula pada marginal kontinu tetapi tidak pada marginal diskrit.

Berdasarkan teorema Sklar tersebut, mengkonstruksi distribusi bivariat Copula dari data marginal diskrit akan menghasilkan Copula yang tidak unik. Akibatnya, akan menghasilkan interpretasi yang tidak jelas, terutama sifat kebergantungannya. Hal ini sesuai dengan pernyataan Denuit dan Lambert (2005) bahwa penggunaan Copula secara langsung pada marginal diskrit akan menghasilkan ketidakunikan pada distribusi bersamanya sebagai sebuah fungsi dari distribusi marginalnya. Hal yang sama juga dinyatakan oleh Genest dan Nešlehová (2007). Oleh sebab itu, dalam mengkonstruksi distribusi bivariat Copula dari data marginal diskrit, maka datanya dikontinukan terlebih dahulu agar keunikan Copula yang dihasilkan dapat terjamin oleh teorema Sklar.

3.3. Jitters

Jitters merupakan salah satu cara yang digunakan untuk mengkontinukan data marginal diskrit. Madsen dan Fang (2010) menggunakan jitters saat memodelkan data longitudinal diskritnya. Selain itu, Shi dan Valdez (2012) juga menggunakan jitters pada data longitudinalnya. Selama proses transformasi jitters, tidak ada informasi yang hilang (Madsen dan Fang, 2010). Lebih lanjut, hubungan kebergantungan antar data asli dan data hasil jitters adalah sama (Denuit dan Lambert, 2005).

Misalkan $\{x\}$ adalah data dari variabel acak diskrit X . Untuk memperoleh X^* yang kontinu, maka dilakukan transformasi jitters. Diperoleh:

$$X^* = X - U \quad 2)$$

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

Adapun fungsi distribusinya, yaitu:

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

$$\begin{aligned}
&= \sum_{k=0}^{[x^*]} \{1 - P(U \leq k - x^*)\} P(X = k) \\
&\quad + \sum_{k=[x^*+1]}^{\infty} \{1 - P(U \leq k - x^*)\} P(X = k) \\
&= \{P(X \leq [x^*]) - (0 + 0 + \dots + 0)\} + (x^* - [x^*]) P(X = [x^* + 1]) \\
&= P(X \leq [x^*]) + (x^* - [x^*]) P(X = [x^* + 1])
\end{aligned}$$

Sehingga diperoleh:

$$\begin{aligned}
F^*(x^*) &= P(X \leq [x^*]) + (x^* - [x^*]) P(X = [x^* + 1]) \\
&= F([x^*]) + (x^* - [x^*]) f([x^* + 1])
\end{aligned} \tag{3}$$

Fungsi peluangnya yang bersesuaian, yaitu:

$$f^*(x^*) = P(X = [x^* + 1]) = f([x^* + 1]) \tag{4}$$

3.4. Data Jitters Mempresentasikan Data Aslinya

Hasil jitters 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 aslinya. Cara kedua, dibuktikan dengan membandingkan nilai Kendall's tau hasil jitters terhadap data aslinya.

Cara pertama, misalkan terdapat dua data dari variabel acak diskrit X dan Y. Dapat dikatakan bahwa (X_2, Y_2) lebih concordant daripada (X_1, Y_1) , dinotasikan $(X_1, Y_1) <_c (X_2, Y_2)$, jika:

$$P(X_1 \leq s, Y_1 \leq t) \leq P(X_2 \leq s, Y_2 \leq t); \quad \forall s, t \in R \tag{5}$$

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

$$\begin{aligned}
P(X_1^* \leq s, Y_1^* \leq t) &= P(X_1 - U_1 \leq s, Y_1 - U_2 \leq t) \\
&= \iint_{u_1, u_2 \in [0,1]} P(X_1 \leq s + u_1, Y_1 \leq t + u_2) f_{U_1}(u_1) f_{U_2}(u_2) du_1 du_2 \\
&\leq \iint_{u_1, u_2 \in [0,1]} P(X_2 \leq s + u_1, Y_2 \leq t + u_2) f_{U_1}(u_1) f_{U_2}(u_2) du_1 du_2 \\
&= P(X_2^* \leq s, Y_2^* \leq t)
\end{aligned} \tag{6}$$

sehingga,

$$(X_1, Y_1) <_c (X_2, Y_2) \Rightarrow (X_1^*, Y_1^*) <_c (X_2^*, Y_2^*) \tag{7}$$

Dari persamaan (7) terlihat bahwa jika (X_1, Y_1) lebih *concordant* daripada (X_2, Y_2) maka (X_1^*, Y_1^*) juga lebih *concordant* daripada (X_2^*, Y_2^*) . Sehingga terbukti bahwa orde *concordant* antar data hasil jitters sama dengan orde *concordant* antar data yang aslinya. Jadi, data jitters mampu mempresentasikan data aslinya.

Cara kedua, misalkan X^* dan Y^* masing-masing merupakan variabel acak kontinu hasil jitters oleh U_1 dan U_2 . Diperoleh:

$$\begin{aligned}
 P^*(\text{concordant}) &= P[(X_1^* - X_2^*)(Y_1^* - Y_2^*) > 0] \\
 &= P[(X_1 - U_{11} - X_2 + U_{12})(Y_1 - U_{21} - Y_2 + U_{22}) > 0] \\
 &= P[X_1 = X_2, Y_1 = Y_2]P[(U_{12} - U_{11})(U_{22} - U_{21}) > 0] \\
 &\quad + \\
 &\quad P[X_1 = X_2, Y_1 > Y_2]P[(U_{12} - U_{11}) > 0] + \\
 &\quad P[X_1 = X_2, Y_1 < Y_2]P[(U_{12} - U_{11}) < 0] + \\
 &\quad P[X_1 > X_2, Y_1 = Y_2]P[(U_{22} - U_{21}) > 0] + \\
 &\quad P[X_1 < X_2, Y_1 = Y_2]P[(U_{22} - U_{21}) < 0] + \\
 &\quad P[(X_1 - X_2)(Y_1 - Y_2) > 0]
 \end{aligned} \tag{8}$$

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

$$\begin{aligned}
 P^*(\text{concordant}) &= \frac{1}{2}P(\text{tie}) + P[(X_1 - X_2)(Y_1 - Y_2) > 0] \\
 &= \frac{1}{2}P(\text{tie}) + P(\text{concordant})
 \end{aligned} \tag{9}$$

Sehingga,

$$\begin{aligned}
 \tau^*(X^*, Y^*) &= P^*(\text{concordant}) - P^*(\text{discordant}) \\
 &= P(\text{concordant}) + P(\text{tie}) - P(\text{discordant})
 \end{aligned} \tag{10}$$

Jika $P(\text{tie}) = 0$ maka $\tau^*_{X^*, Y^*} = \tau_{X, Y}$. Nilai kendalls tau data asli dan data hasil jitters adalah sama. Hal ini menunjukkan bahwa data asli dan data hasil jitters memiliki sifat kebergantungan yang sama. Sehingga, dapat dikatakan bahwa data jitters mampu mempresentasikan data aslinya.

3.5. Simulasi

Data simulasi $\{x\}$ dan $\{y\}$ berasal dari dua marginal diskrit yang berbeda. Distribusi dari $\{x\}$ dan $\{y\}$ berturut-turut adalah binomial (100, 0.15) dan poisson (20). Masing-masing data dibangkitkan sebanyak 200. Selanjutnya, kedua jenis data tersebut dikontinukan melalui transformasi jitters. Adapun penyebaran data asli dan data hasil jitters dapat dilihat pada gambar berikut.

Gambar 5. Scatter plot data asli dan data hasil jitters (kiri ke kanan)

Pada gambar 5 di sebelah kiri terlihat bahwa data asli menyebar di sekitar nilai (15, 20). Adapun nilai Kendall's tau yang dimiliki sebesar $\tau = 0.1142$. Nilai kebergantungan yang diperoleh menunjukkan bahwa kedua jenis data memiliki kebergantungan positif yang relatif cukup kuat. Sedangkan pada gambar 5 sebelah kanan terlihat bahwa data hasil transformasi jitters juga menyebar di sekitar (15, 20) dengan $\tau^* = 0.1147$. Namun, perilaku penyebaran data hasil jitters terlihat sedikit berbeda dengan data aslinya. Setelah dianalisis lebih lanjut, pada data asli ditemukan beberapa pasangan data yang bernilai sama atau $(x_i, y_i) = (x_j, y_j), \exists i \neq j$. Sementara pada data jitters tidak ditemukan pasangan data yang bernilai sama atau $(x_i^*, y_i^*) = (x_j^*, y_j^*), \forall i \neq j$.

Lebih jauh lagi, semua pasangan data asli yang bernilai sama diplot oleh scatter yang sama dan saling tumpang tindih. Akibatnya, pasangan data yang sama hanya akan terlihat menjadi satu pasangan data. Setelah dikontinukan melalui transformasi jitters, maka semua pasangan data jitters tidak ada yang sama. Sehingga, scatter dari setiap pasangan data jadi berbeda dan terlihat lebih banyak.

Misalnya, pada pasangan data ke-128 yang bernilai (23, 15) dan pasangan data ke-190 yang juga bernilai (23, 15). Pasangan data tersebut berada dalam lingkaran besar pada gambar 5. Pada gambar 5 sebelah kiri, pasangan data tersebut terlihat diplot oleh satu scatter. Padahal data tersebut diplot oleh dua scatter yang sama dan saling tumpang tindih. Akibatnya, dua pasangan data tersebut terlihat menjadi satu pasangan satu saja. Jika dihitung secara visual, banyaknya data asli dalam lingkaran besar adalah 6. Pasangan data jitters ke-128 menjadi bernilai (22.726, 14.236) dan ke-190 bernilai (22.549, 14.714). Masing-masing pasangan data jitters ini diplot oleh scatter yang berbeda. Oleh sebab itu, banyaknya data jitters dalam lingkaran besar lebih banyak, yaitu 7. Lebih lanjut, banyaknya data asli di lingkaran kecil pada gambar 5 adalah 2 dan data jittersnya juga 2. Hal ini berarti bahwa memang hanya terdapat dua pasangan data asli yang berbeda di lingkaran kecil tersebut.

Banyaknya pasangan data asli yang awalnya bernilai sama, ketika melalui transformasi jitters menjadi berbeda, menyebabkan penyebaran data jitter menjadi sedikit berbeda dari penyebaran data aslinya. Namun, data jitters masih mampu mempresentasikan data aslinya. Hal ini dapat terlihat dari nilai Kendall's tau-nya yang sama, yaitu $\tau = \tau^* = 0.114$.

3.6. Mengkonstruksi Distribusi Bivariat Copula Clayton Data Jitters

Langkah selanjutnya, yaitu mengkonstruksi distribusi bivariat Copula Clayton, baik fungsi distribusi maupun fungsi peluang menggunakan data jitters. Fungsi distribusi dari data jitters didefinisikan oleh persamaan (3), sedangkan fungsi peluangnya didefinisikan oleh persamaan (4). Berdasarkan tabel 1, fungsi distribusi bivariat Copula Clayton data jitters didefinisikan sebagai berikut:

$$C^*(v^*, w^*) = \max\left((v^{*- \theta} + w^{*- \theta} - 1)^{-\theta^{-1}}, 0\right) \quad (11)$$

Fungsi peluangnya yang bersesuaian, yaitu:

$$c^*(v^*, w^*) = (\theta + 1)(v^* w^*)^{-(\theta+1)}(v^{*-\theta} + w^{*-\theta} - 1)^{-(2\theta+1)\theta^{-1}} \quad (12)$$

dimana: $v^* = F_{X^*}^*$ dan $w^* = F_{Y^*}^*$.

Persamaan (11) merupakan fungsi distribusi bivariat data asli yang berdistribusi marginal diskrit, dapat dikatakan bahwa $F(x, y) = C^*(v^*, w^*)$. Sementara persamaan (12) merupakan fungsi peluang bivariatnya, $f(x, y) = c^*(v^*, w^*)$. Jadi, distribusi bivariat Copula Clayton ini mempresentasikan distribusi bivariat dari data aslinya. Hal ini diilustrasikan melalui 20 data simulasi pada tabel berikut.

Tabel 2. Data Asli, Data Jitters, Fungsi Distribusi Bivariat Copula Clayton dan Fungsi Peluang Bivariat Copula Clayton

x	y	x^*	y^*	$C^*(v^*, w^*)$	$c^*(v^*, w^*)$
16	24	15.82954	23.92738	0.559317	1.107737
20	26	19.08469	25.05151	0.800394	1.193294
13	18	12.67037	17.43202	0.135423	1.035112
15	19	14.12671	18.55783	0.232776	1.024629
10	17	9.174524	16.40111	0.030118	1.144279
13	28	12.22224	27.31143	0.260487	0.904107
22	25	21.43816	24.34589	0.832406	1.202506
16	24	15.671	23.83308	0.542707	1.101447
13	21	12.70928	20.91531	0.226399	0.985376
11	21	10.80326	20.01513	0.104744	0.918957
18	24	17.69438	23.65603	0.677864	1.152517
16	21	15.71666	20.64244	0.414618	1.067346
8	18	7.400075	17.22858	0.01168	0.962778
16	11	15.26737	10.0219	0.009331	0.581866
15	22	14.57935	21.50522	0.376832	1.050323
27	32	26.96813	31.10577	0.991633	1.256318
12	18	11.5607	17.23214	0.094131	1.054435
18	20	17.98734	19.56348	0.447152	1.057851
12	13	11.86943	12.72961	0.027511	1.172766
13	24	12.55802	23.57112	0.261655	0.949813

Dari tabel 2, diperoleh informasi bahwa untuk $i < j$, jika $x_i > x_j$ maka $x_i^* > x_j^*$. Hal ini menunjukkan bahwa perilaku data jitters sama dengan data aslinya. Selain itu, dari tabel 2 diketahui bahwa $F(16, 24) = 0.559317$, $F(20, 26) = 0.800394$, dan seterusnya. Sementara $f(16, 24) = 1.107737$, $f(20, 26) = 1.193294$, dan seterusnya. Adapun fungsi peluang dan fungsi distribusi bivariat dari semua data diilustrasikan pada sebagai berikut:

Gambar 6. 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 pemaparan sebelumnya, untuk mengkonstruksi distribusi bivariat Copula Clayton pada data marginal diskrit dengan implikasi kebergantungan maka distribusi marginalnya harus dikontinukan terlebih dahulu agar menghasilkan Copula C yang unik. Untuk mengkontinukan marginal diskrit maka dilakukan transformasi jitters. Data jitters inilah yang digunakan untuk mengkonstruksi distribusi bivariat Copula Clayton, baik fungsi peluang maupun fungsi distribusinya (peluang kumulatif). Hasil distribusi bivariat Copula Clayton dari data jitters mempresentasikan distribusi bivariat dari data aslinya karena data jitters mampu mempresentasikan data aslinya. Diperoleh distribusi bivariatnya, yaitu $F_{X,Y} = C^*_{v^*,w^*} = \max\left((v^{*-\theta} + w^{*-\theta} - 1)^{-\theta^{-1}}, 0\right)$ dan $f_{X,Y} = c^*_{v^*,w^*} = (\theta + 1)(v^*w^*)^{-(\theta+1)}(v^{*-\theta} + w^{*-\theta} - 1)^{-(2\theta+1)\theta^{-1}}$, dimana $v^* = F^*_{X^*}$ dan $w^* = F^*_{Y^*}$.

5. Ucapan Terimakasih

Terimakasih kepada semua pihak yang telah membantu proses penelitian dan penulisan ini.

Pustaka

Denuit, M., & Lambert, P. (2005). Constraints on Concordance Measures in Bivariate Discrete Data. *Journal of Multivariate Analysis*, 93, 40-57.

- Genest, C., & Nešlehová, J. (2007). A primer on Copulas for Count Data. *ASTIN Bulletin*, 37(2), 475-515.
- Madsen, L., & Fang, Y. (2010). Joint Regression Analysis for Discrete Longitudinal Data. *Biometrics*, 67(3), 1175-1176.
- Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd Edition. New York: Springer.
- Rice, J. A. (1995). *Mathematical Statistic and Data Analysis*. Duxbury Press.
- Sekaran, Uma. (2006). *Metode Penelitian Bisnis*. Jakarta: Salemba Empat.
- Setiawan, Wawan., & Munir. (2006). *Pengantar Teknologi Informasi: Sistem Informasi*. Bandung: Universitas Pendidikan Indonesia.
- Shi, P., & Valdez, E. A. (2012). Longitudinal Modeling of Insurance Claim Counts Using Jitters. *Scandinavian Actuarial Journal*.
- Song, P. X. K. (2007). *Correlated Data Analysis*. New York: Springer.
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: John Wiley & Sons Ltd.