Le Vine Copule

Alessandro Ricotta, PhD
(alessandro.ricotta@univr.it)

Edoardo Luini, PhD
(edoardo.glaucoluini@unicatt.it)

Seminario del Comitato Scientifico dell'Ordine degli Attuari

27/01/2023

Agenda

Le Vine Copule

1. Aspetti Introduttivi
2. Pair-copula Construction (PCC)
3. PCC e Vine: le Vine Copule
4. Stima delle Vine Copule
5. I pacchetti "CDVine" e "VineCopula" nel linguaggio R
6. Riferimenti Bibliografici
7. Appendice

Agenda

Le Vine Copule

1. **Aspetti Introduttivi**
2. Pair-copula Construction (PCC)
3. PCC e Vine: le Vine Copule
4. Stima delle Vine Copule
5. I pacchetti "CDVine" e "VineCopula" nel linguaggio R
6. Riferimenti Bibliografici
7. Appendice

Vine copule

Motivazione

Le copule permettono di rappresentare un fenomeno multivariato separandone il comportamento marginale dalla struttura di dipendenza.

Esiste un'ampia gamma di famiglie di copule parametriche.

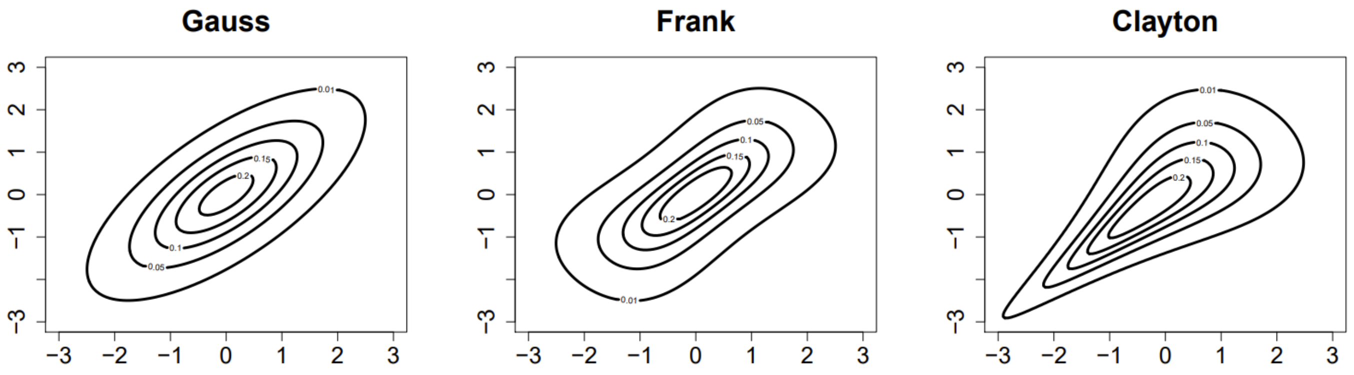


Immagine da: “Introduction to vine copulas”, Nicole Kraemer & Ulf Schepsmeier. NIPS Workshop, Granada, December 18, 2011.

Tuttavia le classiche copule multivariate:

- Possono essere poco flessibili al crescere della dimensionalità del contesto.
- Non consentono strutture di dipendenza diverse tra sottogruppi di variabili aleatorie.
- La disponibilità di dataset multivariati adeguati e di qualità è ancora limitata.

Le Vine copula permettono di superare queste problematiche.

Vine copule

Aspetti introduttivi:

Le vine copule sono basate su un costrutto metodologico che coniuga:

- **Copule bidimensionali**, quali elementi costitutivi.
- **Vine** (Bedford e Cooke (2001, 2002)), un insieme di alberi (grafi) annidati (*nested trees*).

Questi determinano la struttura di dipendenza multivariata.

Tale approccio è flessibile e permette di selezionare le copule bivariate tra il vasto insieme di famiglie parametriche disponibili.

Alla base vi è la nozione di **pair-copula construction** (PCC) (Joe (1996)), che permette di fattorizzare la densità di una variabile aleatoria multivariata n-dimensionale per mezzo di copule bivariate.

La costruzione di tali copule fa quindi riferimento a:

- Tecniche di statistica inferenziale (ad es. massima-verosimiglianza o approcci bayesiani) per calibrare le copule bivariate.
- Teoria dei grafi per detrerminare l’organizzazione e la struttura di dipendenza dei dati.

Vine copule

Premesse

Sia X una v.a. unidimensionale assolutamente continua, allora:

- $F_X(x) = P(X \leq x)$ è la funzione di ripartizione (FdR);
- $F^{-1}(p) = x \;\; con \;\; p \in [0,1]$, tale per cui, $F(x) = p$ è definita funzione quantilica
- $f_X(x) = \frac{\partial F_X(x)}{\partial x}$ è la funzione di densità;

Sia (X, Y) una v.a. bidimensionale assolutamente continua, allora:

- $F(x, y) = P(X \leq x, Y \; \leq y)$ è la funzione di ripartizione

Teorema di Sklar

Sia $F(x, y)$ una FdR bidimensionale con FdR marginali $F(x)$ e $F(y)$. Allora esiste una copula C tale per cui per ogni $(x, y) \in \mathbb{R}^2$

$$F(x, y) = C\big(F(x), F(y)\big),$$

$$f(x, y) = c\big(F(x), F(y)\big) f(x) f(y)$$

Si ha inoltre che

$$C(u, v) = F_{X,Y}\left(F_X^{-1}(u), F_Y^{-1}(v)\right) \quad con \; u, v \in [0,1]$$

Vine copule

Premesse

Sia $\boldsymbol{X}$ una v.a. n-dimensionale $(X_1, X_2, \dots, X_n)$; la FdR della v.a. bidimensionale (X_1, X_2) condizionata al vettore in $n-2$ dimensioni di v.a. $(X_3, \dots, X_n)$ può essere definita come[1]:

$$F_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n) = C_{1,2|3,\dots,n}\big(F_{1|3,\dots,n}(x_1|x_3, \dots, x_n), F_{2|3,\dots,n}(x_2|x_3, \dots, x_n)|x_3, \dots, x_n\big) \quad 1.1$$

La copula sottostante alla rispettiva funzione di ripartizione condizionata $F_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n)$ risulta:

$$C_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n) = F_{1,2|3,\dots,n}\big(F^{-1}_{1|3,\dots,n}(x_1|x_3, \dots, x_n), F^{-1}_{2|3,\dots,n}(x_2|x_3, \dots, x_n)|x_3, \dots, x_n\big) \quad 1.2$$

Si noti che le v.a. condizionanti impattano, non solo sugli argomenti della copula, ma anche sulla forma funzionale della copula, ossia i parametri caratteristici della stessa. Infine, la funzione di densità di $F_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n)$ può essere formulata come:

$$\begin{aligned} &f_{1,2|3,\dots,n}(x_1, x_2|x_3, \dots, x_n) \\ &= c_{1,2|3,\dots,n}\big(F_{1|3,\dots,n}(x_1|x_3, \dots, x_n), F_{2|3,\dots,n}(x_2|x_3, \dots, x_n)|x_3, \dots, x_n\big) \cdot f_{1|3,\dots,n}(x_1|x_3, \dots, x_n) \\ &\cdot f_{2|3,\dots,n}(x_1|x_3, \dots, x_n) \end{aligned} \quad 1.3$$

Tali risultati derivano dal teorema di Sklar.

[1] Al fine di uniformare la notazione con la letteratura inerente la pair-copula construction e le vine copule, per le funzioni di ripartizione e densità n-variate si utilizzano le seguenti formulazioni: $F_{X_1,\dots,X_n}(x_1, \dots, x_n) = F_{1,\dots,n}(x_1, \dots, x_n)$ e $f_{X_1,\dots,X_n}(x_1, \dots, x_n) = f_{1,\dots,n}(x_1, \dots, x_n)$.

Agenda

Le Vine Copule

1. Aspetti Introduttivi
2. **Pair-copula Construction (PCC)**
3. PCC e Vine: le Vine Copule
4. Stima delle Vine Copule
5. I pacchetti "CDVine" e "VineCopula" nel linguaggio R
6. Riferimenti Bibliografici
7. Appendice

Vine copule

Pair-copula construction (PCC)

Come noto, la funzione di densità della v.a. $\boldsymbol{X}$ n-dimensionale, dotata di v.a. marginali continue, può essere scritta esplicitandone la copula sottostante:

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = c_{1,2,\dots,n}\big(F_1(x_1), F_2(x_2), \dots, F_n(x_n)\big) \cdot f_1(x_1) \cdot f_2(x_2) \cdot \ \dots \cdot f_n(x_n) \qquad 2.1$$

Obiettivo della pair-copula construction è quello di scrivere la relazione precedente scomponendo la funzione di densità della copula n-dimensionale tramite copule bidimensionali.

La PCC permette di giungere alla seguente formulazione[1]:

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = \prod_{k=1}^{n} f_k(x_k) \cdot$$

$$\prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\dots,i+j-1}\left(F_{i|i+1,\dots,i+j-1}\big(x_i|x_{i+1,\dots,}x_{i+j-1}\big), F_{i+j|i+1,\dots,i+j-1}\big(x_{i+j}|x_{i+1,\dots,}x_{i+j-1}\big)|(x_{i+1,\dots,}x_{i+j-1})\right)$$

2.2

[1] La fattorizzazione in esame individua la densità di una particolare vine copula definita D-vine.

Vine copule

Pair-copula construction (PCC)

In particolare, riscrivendo la densità di una v.a. multivariata come prodotto delle funzioni di densità delle variabili unidimensionali condizionate[1,2],

$$f_{1,2,\dots,n}(x_1,x_2,\dots,x_n)=f_1(x_1)\cdot f_{2|1}(x_2|x_1)\cdot\ \dots\cdot f_{n|1,2,\dots,n-1}(x_n|x_1,x_2,\dots,x_{n-1}) \qquad 2.3$$

Ogni elemento della relazione precedente può essere espresso come il prodotto della relativa pair-copula e la funzione di densità marginale condizionata (Aas et al. (2009)):

$$f_{x_i|\boldsymbol{v}}(x_i|\boldsymbol{v})=c_{x_i,v_j|\boldsymbol{v}_{-j}}\left(F_{x_i|\boldsymbol{v}_{-j}}(x_i|\boldsymbol{v}_{-j}),F_{v_j|\boldsymbol{v}_{-j}}(v_j|\boldsymbol{v}_{-j})|\boldsymbol{v}_{-j}\right)f_{x_i|\boldsymbol{v}_{-j}}(x_i|\boldsymbol{v}_{-j}) \qquad 2.4$$

dove,

- $\boldsymbol{v}$ rappresenta il vettore $(x_1,\dots,x_{i-1})$;
- v_j individua il generico elemento del vettore $\boldsymbol{v}$;
- $\boldsymbol{v}_{-j}$ è il vettore $\boldsymbol{v}$ al netto dell'elemento v_j.

e iterando il processo si giunge alla relazione (2.2).

[1] La decomposizione è unica a meno di una permutazione delle variabili.

[2] $f_{1,2,\dots,n}(x_1,x_2,\dots,x_n)=f_1(x_1)\cdot\frac{f_{1,2}(x_1,x_2)}{f_1(x_1)}\cdot\ \dots\cdot\frac{f_{1,2,\dots,n}(x_1,x_2,\dots,x_n)}{f_{1,2,\dots,n-1}(x_1,x_2,\dots,x_{n-1})}$

Vine copule

Pair-copula construction (PCC)

La fattorizzazione della densità multivariata tramite pair-copula construction è, quindi, per sua natura un **processo ricorsivo**. Le funzioni di densità condizionate della relazione (2.3),

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot \ \dots \cdot f_{n|1,2,\dots,n-1}(x_n|x_1, x_2, \dots, x_{n-1})$$

possono essere scomposte tramite l'opportuna pair-copula moltiplicata per una funzione di densità condizionata; quest'ultima, a sua volta, può essere scomposta tramite pair-copula. Iterando in modo ricorsivo tale procedimento, è possibile scrivere la densità della variabile multidimensionale di partenza come il prodotto tra le funzioni di densità marginali e le rispettive copule bivariate; data una variabile in n dimensioni, le copule bivariate coinvolte sono pari a $\binom{n}{2}$.

La scomposizione della variabile n -dimensionale, inoltre, può essere ottenuta tramite molteplici fattorizzazioni tra loro equivalenti da un punto di vista teorico.

Appendice 1: Pair-copula construction per $n = 3$

Vine copule

Pair-copula construction (PCC)

La PCC coinvolge anche copule bivariate condizionate; la variabile condizionante (o le variabili condizionanti se $n \geq 4$) compare non solo negli argomenti della pair-copula, ma impatta anche la forma funzionale, ossia il parametro (o i parametri) da cui dipende la pair-copula.

Al fine di rendere trattabile il problema inferenziale legato alla stima dei parametri delle copule condizionate, soprattutto al crescere della dimensione del problema, nella pratica si è soliti formulare la cosiddetta «**ipotesi semplificata**» della pair-copula construction. Tale assunzione prevede che le copule condizionate dipendano dalle variabili condizionanti *solo* per mezzo degli argomenti stessi delle copule, definiti dalle funzioni di ripartizione condizionate.

A titolo esemplificativo,

$$
\begin{aligned}
f_{1,2,3,4}(x_1, x_2, x_3, x_4) = {} & f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\
& \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \cdot c_{3,4}(F_3(x_3), F_4(x_4)) \\
& \cdot c_{1,3|2}(F_1(x_1|x_2), F_3(x_3|x_2)\boxed{|x_2)} \cdot c_{2,4|3}(F_2(x_2|x_3), F_4(x_4|x_3)\boxed{|x_3)} \\
& \cdot c_{1,4|2,3}(F_1(x_1|x_2, x_3), F_4(x_4|x_2, x_3)\boxed{|x_2, x_3)}
\end{aligned}
$$

$$
\begin{aligned}
f_{1,2,3,4}(x_1, x_2, x_3, x_4) \approx {} & f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot f_4(x_4) \\
& \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{2,3}(F_2(x_2), F_3(x_3)) \cdot c_{3,4}(F_3(x_3), F_4(x_4)) \\
& \cdot c_{1,3|2}(F_1(x_1|x_2), F_3(x_3|x_2)\ ) \cdot c_{2,4|3}(F_2(x_2|x_3), F_4(x_4|x_3)) \\
& \cdot c_{1,4|2,3}(F_1(x_1|x_2, x_3), F_4(x_4|x_2, x_3))
\end{aligned}
$$

Agenda

Le Vine Copule

Vine copule

PCC e Vine: le vine copule

I vine sono particolari strutture grafiche costituite da un insieme annidato di alberi connessi tra loro.

***Definizione* (Albero)**: un albero $T = \{N, E\}$ è un grafo aciclico dove N rappresenta l'insieme dei nodi ed E l'insieme degli archi, definiti come coppie di nodi non orientati.

Un vine $\mathcal{V}$ di n elementi è definito come un insieme di alberi contenuti l'uno nell'altro e collegati tra loro, $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$, dove gli archi dell'albero j sono i nodi dell'albero $j+1$, con $j = 1, \dots, n-2$. Un *regular* vine di n variabili, invece, prevede che due archi nell'albero j siano connessi tra loro nell'albero $j+1$ se e solo se tali archi possiedono un nodo in comune.

***Definizione* (Vine, regular vine)**: $\mathcal{V}$ è un vine di n elementi se:

- $\mathcal{V} = \{T_1, \dots, T_{n-1}\}$;
- T_1 è un albero con $N_1 = \{1, \dots, n\}$ nodi e un insieme di archi E_1; per $i = 2, \dots, n-1$, T_i è un albero con $N_i = E_{i-1}$ nodi.

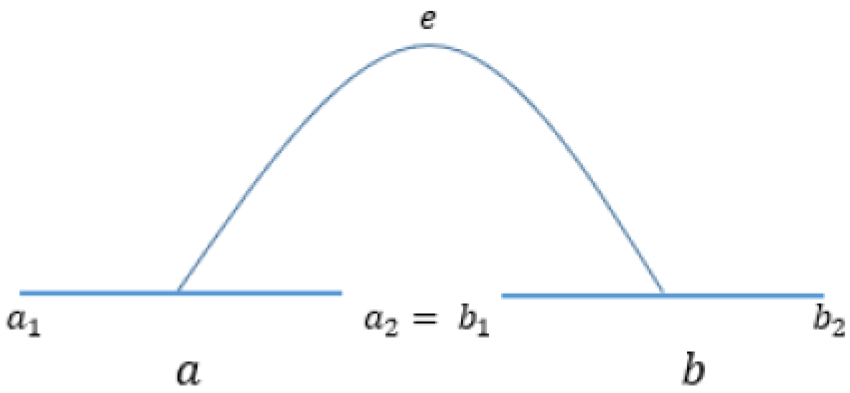
$\mathcal{V}$ è un regular vine (R-vine) di n elementi se oltre alle condizioni precedenti vale la condizione di prossimità:

- per $i = 2, \dots, n-1$, se $\{a, b\} \in E_i$, dove $a = \{a_1, a_2\}$ e $b = \{b_1, b_2\}$, allora $\#a \cap b = 1$, dove # individua il simbolo di cardinalità di un insieme.

Vine copule

PCC e Vine: le vine copule

Un R-vine di n elementi, quindi, è un insieme ordinato di $n-1$ alberi, dove gli archi dell'albero j diventano i nodi dell'albero $j+1$. La condizione di prossimità, inoltre, garantisce che due archi nell'albero $j+1$ siano connessi tra loro solo se nell'albero j hanno un nodo in comune. In un regular vine $\mathcal{V}$ di n elementi il numero totale di archi è pari a $\binom{n}{2}$.



Condizione di prossimità. Gli archi a e b condividono il nodo $a_2 = b_1$;
risultano, quindi, collegati tra loro dall'arco e che sarà un nodo nell'albero successivo.

Vine copule

PCC e Vine: le vine copule

Per ogni arco dell'albero è possibile definire l'insieme dei vincoli, l'insieme condizionante e l'insieme condizionato. Gli elementi ottenibili da un dato arco sono definiti come l'insieme dei vincoli dell'arco. Inoltre, quando due archi sono uniti tra loro da un arco nell'albero successivo, l'intersezione tra gli insiemi dei vincoli dei due archi definisce l'insieme condizionante; infine, la differenza simmetrica degli insiemi dei vincoli definisce l'insieme condizionato.

***Definizione* (Insiemi dei vincoli, condizionante, condizionato)**: Per ogni generico arco e del vine si ha:

- l'insieme dei vincoli associato ad $e \in E_i$ è U_e^*;
- per $i = 1, \dots, n-1$, $e \in E_i$ con $e = \{j, k\}$, l'insieme condizionante associato ad e è $U_j^* \cap U_k^*$;
- per $i = 1, \dots, n-1$, $e \in E_i$ con $e = \{j, k\}$, l'insieme condizionato associato ad e è $U_j^* \triangle U_k^*$.

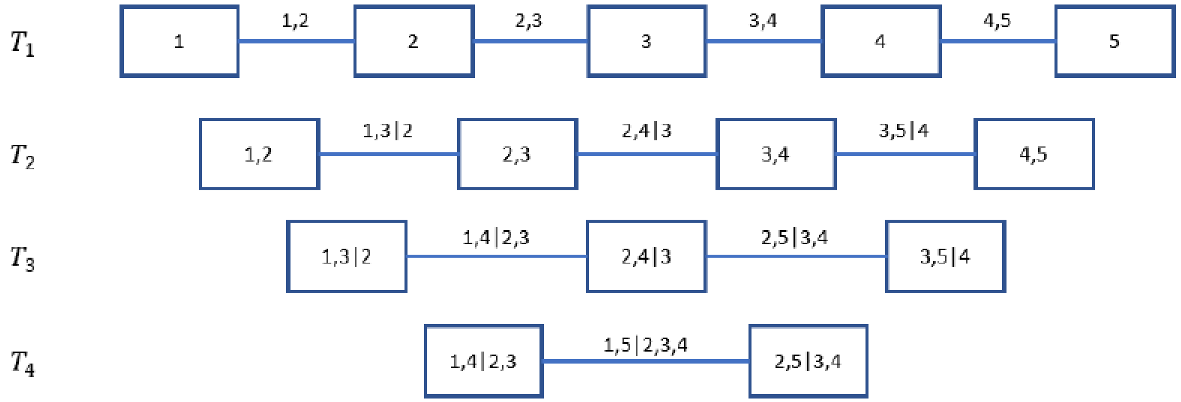
Si noti che per $e \in E_1$, l'insieme condizionante è vuoto. La condizione di prossimità, che definisce i regular vine, garantisce che l'insieme condizionato sia sempre composto da due elementi. Inoltre, ogni coppia di elementi costituenti l'insieme condizionato compare, all'interno dei nodi, una sola volta nella struttura vine.

Vine copule

PCC e Vine: le vine copule

Esempio *(segue)*

Si consideri il seguente regular vine (D-vine) costituito da 5 elementi e, quindi, definito da 4 alberi, $\mathcal{V} = \{T_1, T_2, T_3, T_4\}$:

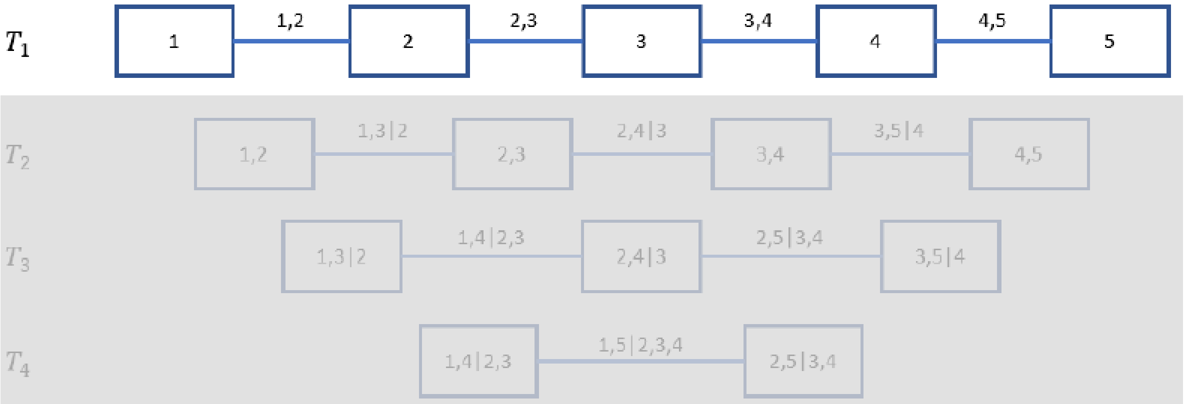


Vine copule

PCC e Vine: le vine copule

Esempio *(segue)*

Si consideri il seguente regular vine (D-vine) costituito da 5 elementi e, quindi, definito da 4 alberi, $\mathcal{V} = \{T_1, T_2, T_3, T_4\}$*:*



*Il **primo albero**,* T_1*, è costituito dai nodi* N_1 *e dagli archi* E_1 *seguenti:*

-$N_1 = \{1,2,3,4,5\}$

-$E_1 = \{\{1,2\},\{2,3\},\{3,4\},\{4,5\}\}$

(i) <u>insieme dei vincoli</u> per i 4 archi è $\{1,2\},\{2,3\},\{3,4\},\{4,5\}$.

(ii) <u>insiemi condizionanti</u> sono vuoti.

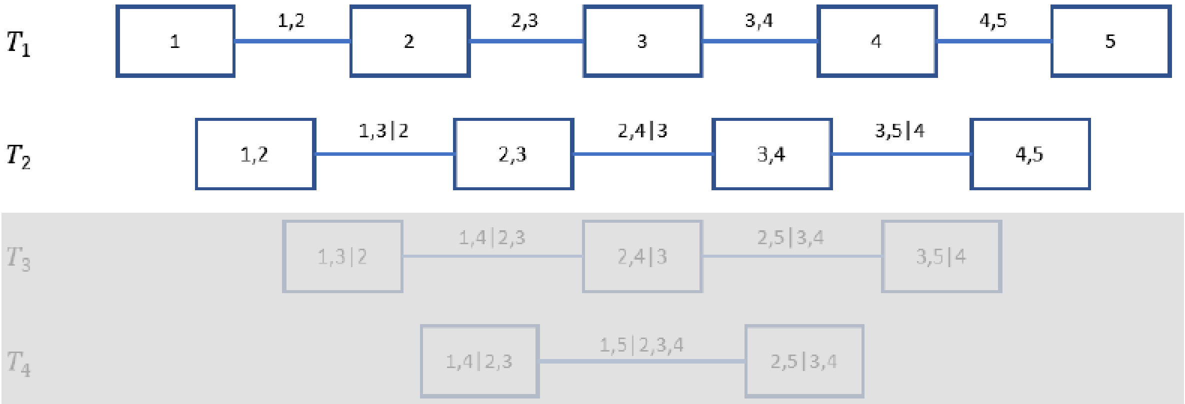
(iii) <u>insiemi condizionati</u> coincidono con quelli dei vincoli.

Vine copule

PCC e Vine: le vine copule

Esempio *(segue)*

Si consideri il seguente regular vine (D-vine) costituito da 5 elementi e, quindi, definito da 4 alberi, $\mathcal{V} = \{T_1, T_2, T_3, T_4\}$*:*



Albero $\boldsymbol{T_2}$, *possiede nodi* N_2 *coincidenti con gli archi dell'albero precedente e archi* E_2 *pari a:*

-$N_2 = E_1$

-$E_2 = \{(\{1,2\},\{2,3\}); (\{2,3\},\{3,4\}); (\{3,4\},\{4,5\})\}$

(i) *<u>insieme dei vincoli</u>* *per i 3 archi:* $(\{1,2\},\{2,3\}); (\{2,3\},\{3,4\}); (\{3,4\},\{4,5\})$.

(ii) *<u>insiemi condizionanti</u>* *sono definiti dall'intersezione degli insiemi dei vincoli degli archi collegati tra loro; per l'arco* $(\{1,2\},\{2,3\})$ *è pari a* $\{1,2\} \cap \{2,3\} = \{2\}$, *per l'arco* $(\{2,3\},\{3,4\})$ *è pari a* $\{2,3\} \cap \{3,4\} = \{3\}$, *mentre per l'arco* $(\{3,4\},\{4,5\})$ *è* $\{3,4\} \cap \{4,5\} = \{4\}$.

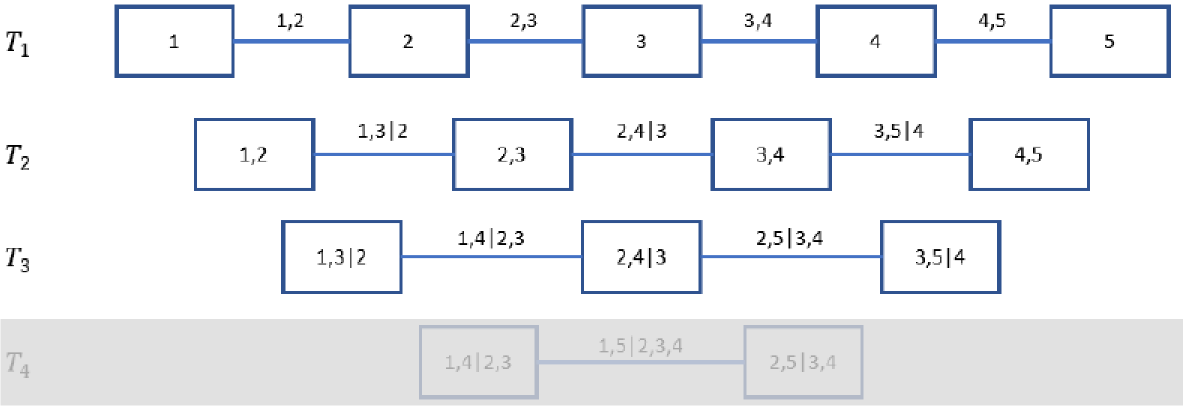
(iii) *<u>insiemi condizionati</u>, invece, sono definiti come la differenza simmetrica tra gli insiemi dei vincoli; per ogni arco, quindi, gli insiemi condizionati sono* $\{1,2\} \triangle \{2,3\} = \{1,3\}$, $\{2,3\} \triangle \{3,4\} = \{2,4\}$, $\{3,4\} \triangle \{4,5\} = \{3,5\}$.

Vine copule

PCC e Vine: le vine copule

Esempio *(segue)*

Si consideri il seguente regular vine (D-vine) costituito da 5 elementi e, quindi, definito da 4 alberi, $\mathcal{V} = \{T_1, T_2, T_3, T_4\}$*:*



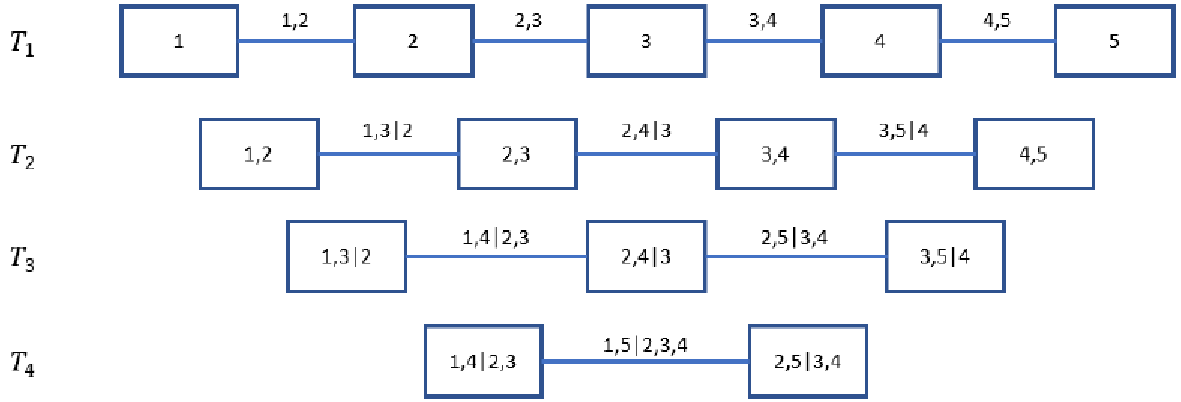
Il ***terzo albero***, T_3, *è caratterizzato dai seguenti nodi e archi:*
-$N_3 = E_2$
-$E_3 = \{(\{1,3|2\},\{2,4|3\}); (\{2,4|3\},\{3,5|4\})\}$
(i) *<u>insiemi dei vincoli</u> per i due archi sono* $(\{1,3,2\},\{2,4,3\})$ *e* $(\{2,4,3\},\{3,5,4\})$.
(ii) *<u>insieme condizionante</u> per l'arco* $(\{1,3|2\},\{2,4|3\})$ *è* $\{1,3,2\} \cap \{2,4,3\} = \{2,3\}$, *mentre per l'arco* $(\{2,4|3\},\{3,5|4\})$ *è* $\{2,4|3\} \cap \{3,5|4\} = \{3,4\}$.
(iii) *<u>insiemi condizionati</u> sono dati da,* $\{1,3,2\} \triangle \{2,4,3\} = \{1,4\}$ *e* $\{2,4,3\} \triangle \{3,5,4\} = \{2,5\}$.

Vine copule

PCC e Vine: le vine copule

Esempio *(fine)*

Si consideri il seguente regular vine (D-vine) costituito da 5 elementi e, quindi, definito da 4 alberi, $\mathcal{V} = \{T_1, T_2, T_3, T_4\}$*:*



Il ***quarto albero****,* T_4*, infine, è caratterizzato dai seguenti nodi e archi:*

-$N_4 = E_3$

-$E_4 = \{\{1,4|2,3\}, \{2,5|3,4\}\}$

L'<u>*insieme dei vincoli*</u> *dell'unico arco è* $(\{1,4,2,3\}, \{2,5,3,4\})$.

L'<u>*insieme condizionante*</u> *è* $\{1,4,2,3\} \cap \{2,5,3,4\} = \{2,3,4\}$.

L'<u>*insieme condizionato*</u> *è* $\{1,4,2,3\} \triangle \{2,5,3,4\} = \{1,5\}$.

Vine copule

PCC e Vine: le vine copule

In letteratura è stato mostrato come la densità di una variabile aleatoria multivariata in n dimensioni possa essere rappresentata tramite vine. In particolare, le vine copule sono definite associando ad ogni arco del vine una copula condizionata in cui le v.a. sono definite dall'insieme condizionato, mentre le variabili condizionanti sono definite dall'insieme condizionante dell'arco.

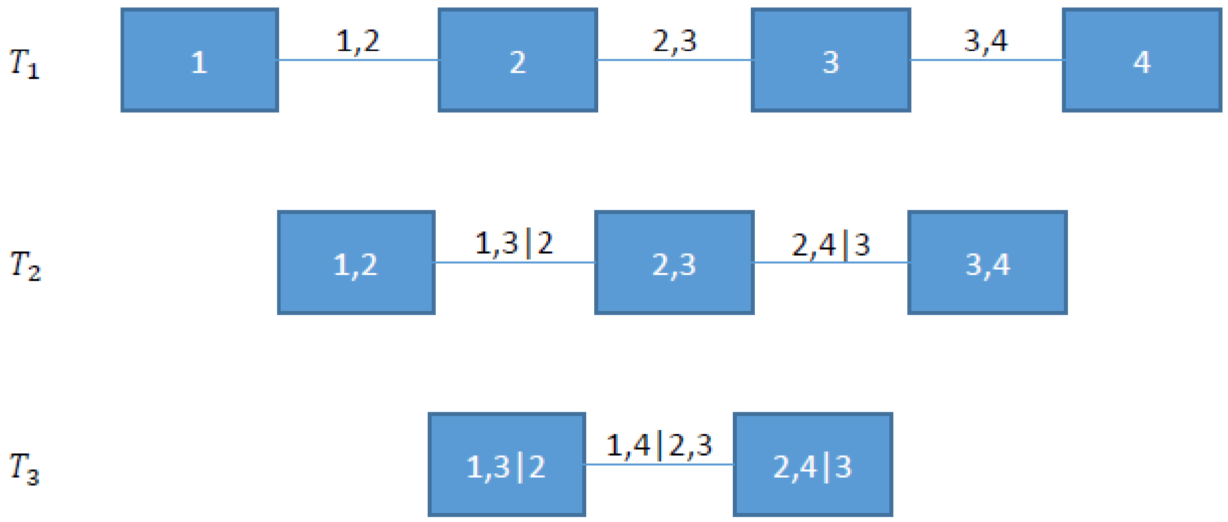
Si noti che la struttura dei regular vine garantisce che tutte le copule coinvolte siano bivariate; In tale situazione, quindi, le copule bivariate coinvolte sono pari a $\binom{n}{2}$ di cui $n-1$ non condizionate e le restanti $(n-1)(n-2)/2$ condizionate.

Le vine copule, quindi, sono copule multivariate costruite tramite la fattorizzazione determinata dalla pair-copula construction, le cui molteplici scomposizioni sono organizzate e definite ricorrendo alla struttura grafica definita dai regular vine.

Vine copule

PCC e Vine: le vine copule

Considerando a titolo esemplificativo un regular vine (D-vine) in 4 dimensioni, la PCC e lo strumento grafico del vine posso essere coniugati come segue:



vine distribution →

$$f_{1,2,3,4} = \underbrace{\prod_{i=1}^{4} f_i}_{\text{nodi in } T_1} \cdot \underbrace{c_{1,2} \cdot c_{2,3} \cdot c_{3,4}}_{\substack{\text{archi in } T_1 \\ \text{nodi in } T_2}} \cdot \underbrace{c_{1,3|2} \cdot c_{2,4|3}}_{\substack{\text{archi in } T_2 \\ \text{nodi in } T_3}} \cdot \underbrace{c_{1,4|2,3}}_{\text{arco in } T_3}$$

vine copula →

$$c_{1,2,3,4} = c_{1,2} \cdot c_{2,3} \cdot c_{3,4} \cdot c_{1,3|2} \cdot c_{2,4|3} \cdot c_{1,4|2,3}$$

Vine copule

PCC e Vine: le vine copule

In letteratura le vine copule maggiormente studiate e utilizzate sono quelle associate alle strutture dei regular vine definite C-vine (*canonical vine*) e D-vine (*drawable vine*).

***Definizione* (C-vine, D-vine)**: Un regular vine $\mathcal{V} = \{T_1, \ldots, T_{n-1}\}$ è chiamato,

- C-vine se ogni albero T_i ha un unico nodo di grado $n - i$;
- D-vine se ogni nodo in T_i ha al più grado pari a 2, dove il grado è definito come il numero di archi connessi al nodo.

Considerando $n > 3$ v.a. le configurazioni differenti di regular vine sono in numero pari a,

$$n!/2 \times 2^{\binom{n-2}{2}} \qquad 3.1$$

Circoscrivendo l'analisi ai soli C-vine e D-vine e considerando solo copule che godono della proprietà dell'interscambiabilità il numero di strutture vine risulta essere considerevole, pari a:

$$n!/2 \qquad 3.2$$

Si noti che per $n = 3$ tutti i regular vine appartengono alla stessa classe, per $n = 4$, invece, possono dare luogo sia a C-vine che a D-vine, infine, per $n = 5$ compaiono regular vine che non sono né C-vine né D-vine.

Vine copule

PCC e Vine: le vine copule

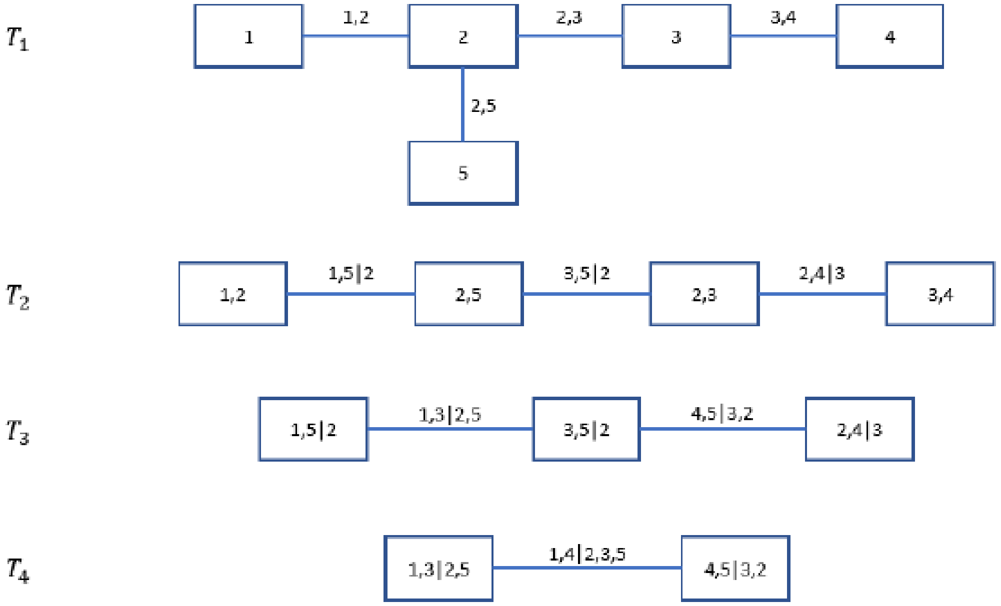
n	R-vine	C-vine, D-vine
3	3	3
4	24	12
5	480	60
6	23.040	360
7	2.580.480	2.520
8	$6,606\,\mathrm{e}{+}08$	20.160
9	$3,805\,\mathrm{e}{+}11$	181.440
10	$4,871\,\mathrm{e}{+}14$	1.814.400

Vine copule

PCC e Vine: le vine copule

R-vine

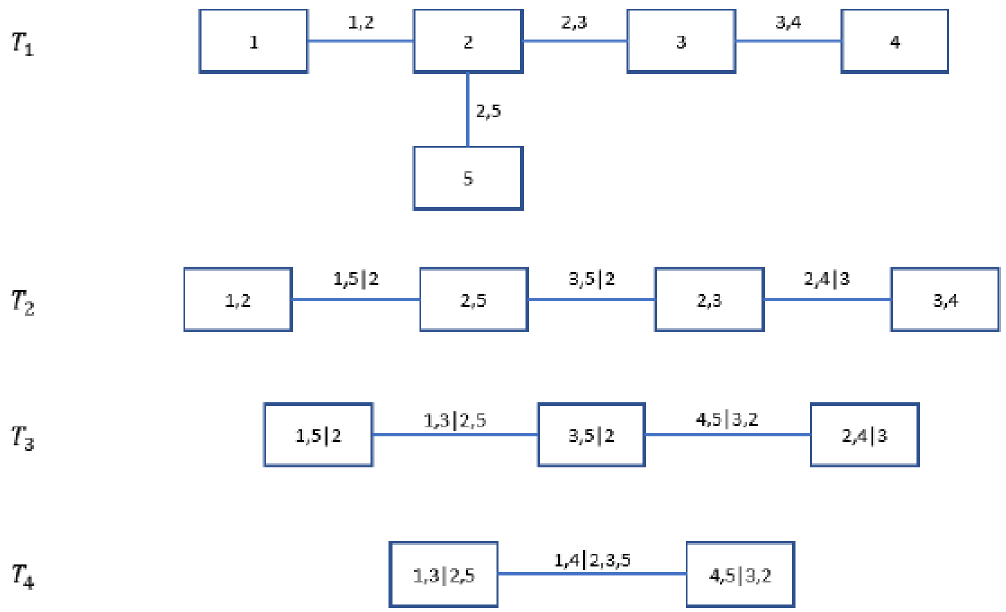
n	R-vine	C-vine, D-vine
3	3	3
4	24	12
5	480	60
6	23.040	360
7	2.580.480	2.520
8	6, 606 e+08	20.160
9	3, 805 e+11	181.440
10	4, 871 e+14	1.814.400



Vine copule

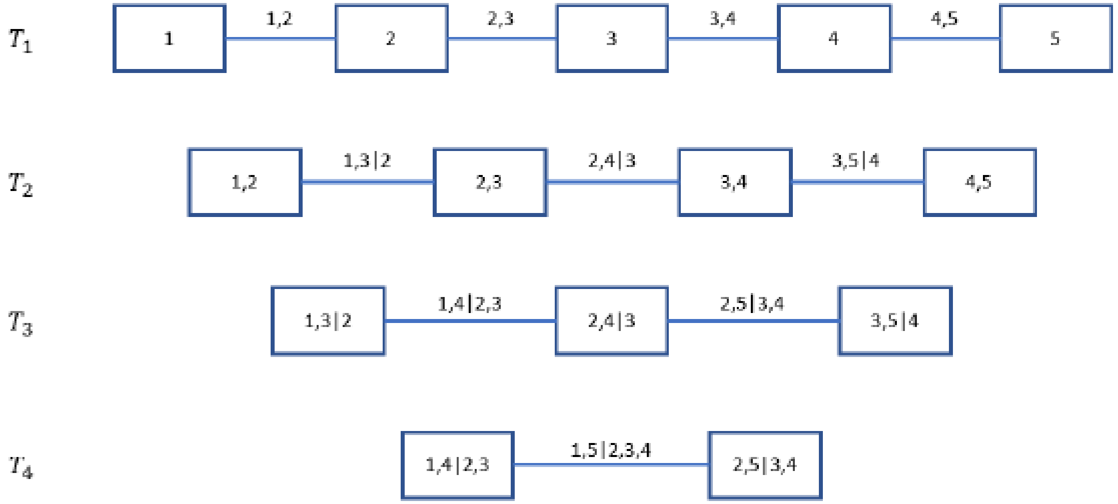
PCC e Vine: le vine copule

R-vine



n	R-vine	C-vine, D-vine
3	3	3
4	24	12
5	480	60
6	23.040	360
7	2.580.480	2.520
8	6, 606 e+08	20.160
9	3, 805 e+11	181.440
10	4, 871 e+14	1.814.400

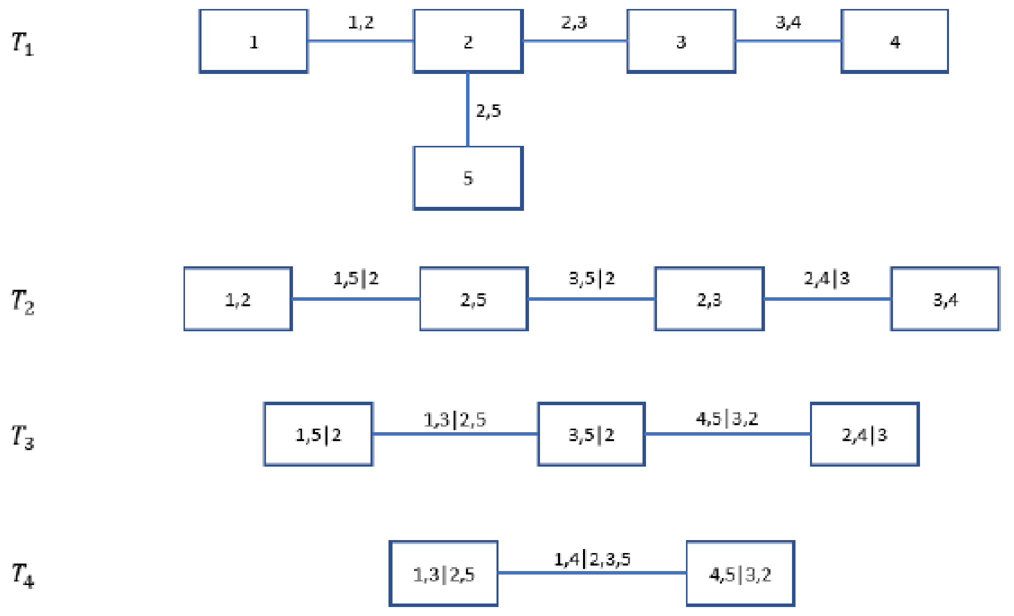
D-vine



Vine copule

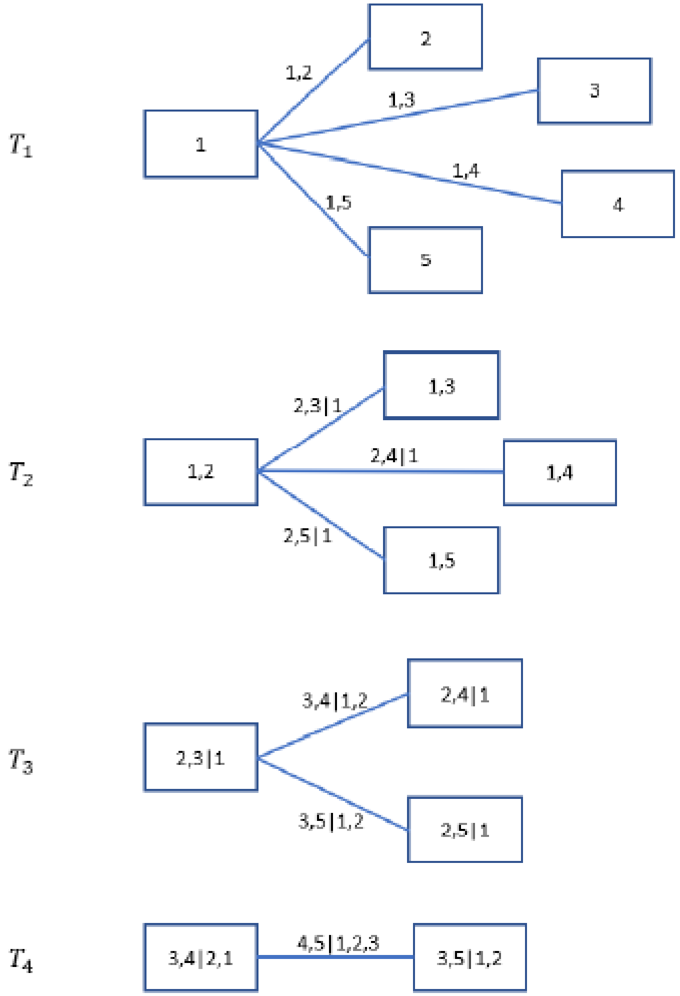
PCC e Vine: le vine copule

R-vine

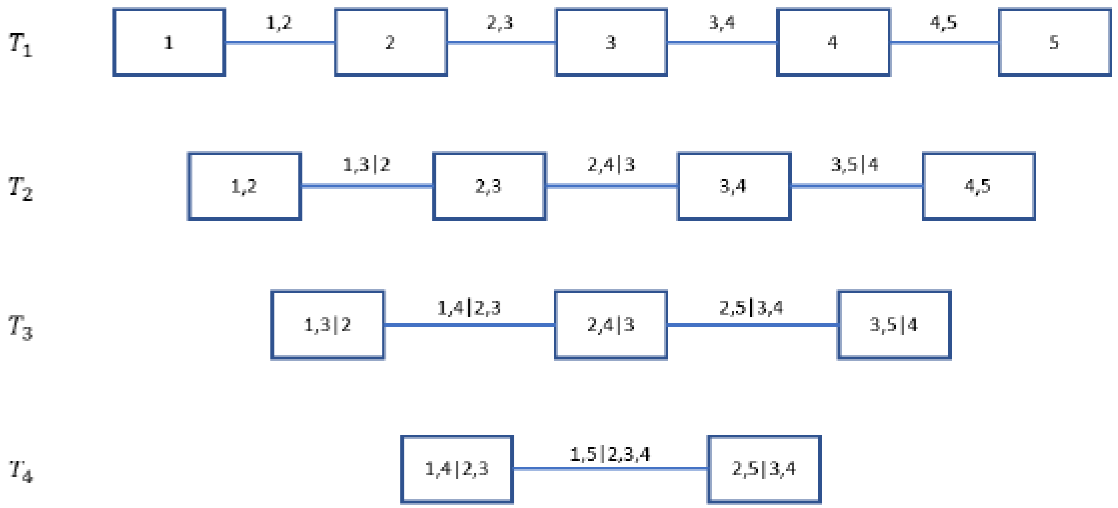


n	R-vine	C-vine, D-vine
3	3	3
4	24	12
5	480	60
6	23.040	360
7	2.580.480	2.520
8	6, 606 e+08	20.160
9	3, 805 e+11	181.440
10	4, 871 e+14	1.814.400

C-vine



D-vine



Vine copule

PCC e Vine: le vine copule

Per simulare da una vine copula è possibile utilizzare un algoritmo generale e analogo al *metodo della distribuzione condizionata* usato per le copule bidimensionali. Si consideri il caso n-dimensionale con variabili marginali uniformi, i passaggi da seguire sono i seguenti:

1) Generare due v.a. w_i con $i = 1, \dots, n$ v.a. indipendenti e uniformi in [0,1];

2) Porre

$$x_1 = w_1$$

$$x_2 = F_{2|1}^{-1}(w_2|x_1)$$
$$x_3 = F_{3|1,2}^{-1}(w_3|x_1, x_2)$$
$$\dots$$
$$x_n = F_{n|1,\dots,n-1}^{-1}(w_n|x_1, \dots, x_{n-1})$$

dove le funzioni $F_{j|1,\dots,j-1}^{-1}\left(w_j\middle|x_1, \dots, x_{j-1}\right)$, per ogni j, sono ottenute ricorsivamente con la relazione $F_{x_i|\boldsymbol{v}}(x_i|\boldsymbol{v})$ mostrata in Joe (1996).

3) Le realizzazioni distribuite secondo la vine copula saranno $(x_1, \dots, x_n)$.

Agenda

Le Vine Copule

Vine copule

Stima delle vine copule

Dato un campione di osservazioni generate da una v.a. n-dimensionale, nel contesto in esame il focus della stima risulta essere la parte del modello vine associato alle sole copule, lasciando ad altra sede il problema della modellizzazione delle v.a. marginali.

Da un punto di vista teorico, le molteplici configurazioni ottenibili tramite PCC con cui è possibile fattorizzare la funzione di densità di una variabile multidimensionale sono tra loro equivalenti.

Da un punto di vista inferenziale, di contro, l'utilizzo dell'ipotesi semplificata associata alla PCC, l'uso solamente di alcune famiglie bivariate di copule per stimare la struttura di dipendenza tra le variabili coinvolte, nonché la procedura di stima dei parametri associati alle copule creano l'esigenza di selezionare la vine copula più adeguata a descrivere la struttura di dipendenza insita nei dati oggetto di studio.

Il problema inferenziale legato alle vine copule, basato sull'ipotesi semplificata della PCC, si compone di tre fasi (Dißmann et al. (2013)):

1) scelta del modello regular vine (C-vine, D-vine) e dell’ordine delle variabili;
2) selezione delle famiglie di copule bivariate da adottare per ogni arco del vine;
3) stima dei parametri caratteristici per ogni famiglia di copula selezionata.

Vine copule

Stima delle vine copule

FASE 1: SCELTA DEL MODELLO E ORDINE DELLE VARIABILI

In linea teorica, al fine di selezionare la struttura vine più adeguata è necessario stimare tutte le possibili strutture esistenti in n dimensioni. L'esigenza pratica di selezionare la struttura vine più adeguata senza enumerarne tutte le possibili configurazioni esistenti ha trovato risposta in letteratura nello sviluppo di specifici algoritmi, quali ad esempio:

- (Dißmann et al. (2013)) - algoritmo euristico che consiste nel selezione gli alberi di un regular vine massimizzando la dipendenza tra le coppie di variabili aleatorie. L'obiettivo dell'algoritmo consiste, nello specifico, nel cogliere il maggior grado di dipendenza (τ di Kendall) nei primi alberi i quali hanno la maggiore influenza sul livello di adattamento complessivo del vine. Si noti che, considerando un D-vine, tale algoritmo consiste nel scegliere l'ordine delle variabili aleatorie del solo primo albero; tutti quelli successivi, infatti, sono univocamente determinati dalla configurazione dell'albero T_1.
- (Czado, Jeske e Hofmann (2013)) - algoritmo che prevede di selezionare gli alberi massimizzando per ogni arco dei pesi, definiti dalle grandezze AIC o dai valori del p-value di opportuni test statistici di adattamento relativi alle copule bivariate stimate. Si noti che tali algoritmi non velocizzano il problema di selezione del vine in quanto richiedono di stimare in ogni albero tutte le pair-copule coinvolte al fine di valutare l'albero che massimizza il criterio di scelta.

Vine copule

Stima delle vine copule

FASE 1: SCELTA DEL MODELLO E ORDINE DELLE VARIABILI

- (Kraus e Czado (2017)) - algoritmo che si prefigge un duplice obiettivo, costruire una vine copula che, analogamente all'algoritmo di Dißmann, colga il più possibile la dipendenza tra le v.a. negli alberi iniziali e, inoltre, che sia aderente il più possibile all'ipotesi semplificata sottostante alla PCC. Quest'ultima è valutata ricorrendo al test statistico definito *constant conditional correlations test* (CCC). In particolare, l'ipotesi semplificata della PCC prevede che, data una copula bivariata, le v.a. condizionate $(F_{x_i|\boldsymbol{v}}(x_i|\boldsymbol{v}))$ siano indipendenti dalle v.a. condizionanti. Il test è costruito considerando il coefficiente di correlazione lineare condizionato e verificando che quest'ultimo sia costante rispetto alle variabili condizionanti.

  Si noti che nell'albero T_1 non sono presenti copule condizionate, quindi tale algoritmo coincide con quello di Dißmann. Nei livelli successivi, invece, vengono scelti gli alberi che massimizzano i pesi associati agli archi, dove i pesi dipendono sia dal τ di Kendall tra le variabili che dal p-value associato al test di ipotesi CCC.

Per scegliere tra due o più modelli stimati di vine copule è possibile ricorre ai criteri di informazione di Akaike (AIC) e Bayesiano (BIC).

Anche i test statistici sviluppati da Vuong (1989) e Clarke (2007), inoltre, permettono di scegliere tra due modelli concorrenti tra loro quello più adeguato; in tali test la statistica si basa sul rapporto delle funzioni di verosimiglianza e risulta rispettivamente distribuita come una v.a. $N(0,1)$ e una v.a. $Bin(H, 0{,}5)$.

Vine copule

Stima delle vine copule

FASE 2: SELEZIONE DELLE FAMIGLIE DI COPULE BIVARIATE

La selezione di un'adeguata famiglia di copule bivariate per ogni arco degli alberi associati al vine può essere perseguita ricorrendo sia a strumenti grafici che a test statistici.

Strumenti grafici:

(i) Plot pseudo-osservazioni;

(ii) grafico *Chi* (Fisher, Switzer (1985));

(iii) cd funzione *lambda* (Wang, Wells (2000)).

Test statistici:

(i) test statistici basati sulla funzione $K(t)$ (Genest, Quessy e Rémillard (2006));

(ii) test basato sulla cosiddetta matrice di informazione di White (White (1982)).

Vine copule

Stima delle vine copule

Fase 3: stima dei parametri per ogni famiglia di copula selezionata

Nel contesto delle vine copule, in letteratura, la stima dei parametri delle copule bidimensionali coinvolte è perseguita tramite l'approccio della verosimiglianza.

Data la non conoscenza delle variabili marginali, le quali vengono modellizzate separatamente dal modello di dipendenza, la stima del vettore dei parametri delle copule è ottenuta seguendo un approccio **semiparametrico (SP)**, in cui vengono utilizzate le funzioni di ripartizione empiriche, o pseudo-osservazioni, delle variabili marginali. Come noto, le funzioni di ripartizione della v.a. marginali permettono, iterativamente, di specificare gli argomenti di tutte le copule condizionate coinvolte nella vine copula.

Il metodo della massima verosimiglianza prevede di stimare i parametri massimizzando la funzione di log-verosimiglianza della vine copula rispetto al vettore $\boldsymbol{\theta}$ di parametri.

Si noti, che le copule che compaiono negli alberi successivi al primo, $(T_2, \dots, T_{n-1})$, dipendono dai parametri delle copule precedenti; gli argomenti della funzione di log-verosimiglianza della vine copula, quindi, devono essere considerati congiuntamente al fine di ottenere il vettore di parametri $\widehat{\boldsymbol{\theta}}^{SP}$ che massimizza la funzione di verosimiglianza della vine copula in esame. Lo stimatore $\widehat{\boldsymbol{\theta}}^{SP}$, sotto opportune condizioni di regolarità, risulta essere consistente e asintoticamente Normale (Haff (2013), Tsukahara (2005)).

Vine copule

Stima delle vine copule

FASE 3: STIMA DEI PARAMETRI PER OGNI FAMIGLIA DI COPULA SELEZIONATA

La massimizzazione congiunta della funzione di verosimiglianza della vine copula risulta essere, al crescere della dimensione e del numero dei parametri del problema, un'operazione computazionalmente onerosa che necessita di un vettore adeguato di valori di partenza al fine di stimare i parametri delle copule coinvolte nel vine. Per gestire tale aspetto, in letteratura l'approccio SP è stato modificato al fine di adattarlo alla logica della PCC, dando luogo ad un approccio di stima **semiparametrico *stepwise* (SSP)**. La procedura stepwise prevede di stimare i parametri delle copule bivariate della PCC separatamente, considerando una copula alla volta, iniziando dall'albero al livello superiore per poi, in seguito, stimare i parametri delle copule site ai livelli inferiori.

Lo stimatore $\boldsymbol{\theta}^{SSP}$ risulta anch'esso uno stimatore consistente e asintoticamente Normale; $\boldsymbol{\theta}^{SSP}$ tuttavia è asintoticamente meno efficiente rispetto a $\boldsymbol{\theta}^{SP}$ in quanto ad ogni livello non tiene conto delle informazioni negli alberi successivi (Haff (2012, 2013)).

Lo stimatore stepwise, tuttavia, riduce l'onere computazione del processo di stima, risultando particolarmente adeguato in contesti inferenziali di dimensione elevata in cui il numero di parametri diviene numeroso. L'approccio stepwise, inoltre, può essere utilizzato per determinare i valori di partenza dell'approccio SP; stime molto differenti tra i metodi SP e SSP sono sintomo, inoltre, dell'errata scelta delle famiglie di copule utilizzate nel vine.

Agenda

Le Vine Copule

1. Aspetti Introduttivi
2. Pair-copula Construction (PCC)
3. PCC e Vine: le Vine Copule
4. Stima delle Vine Copule
5. **I pacchetti "CDVine" e "VineCopula" nel linguaggio R**
6. Riferimenti Bibliografici
7. Appendice

Vine copule

I pacchetti "CDVine" e "VineCopula" nel linguaggio R

L'ambiente di calcolo statistico **R** presenta due pacchetti dedicati alle vine copule, denominati «CDVine» (2015) e «VineCopula» (2018).

Di seguito alcune funzioni utili per sviluppare analisi multivariate con le vine copule:

- `pobs`, funzione che calcola le pseudo-osservazioni di una matrice di osservazioni,

```
pobs(x1,x2)
```

- `BiCopSelect`, funzione che stima una copula bivariata tra un set di famiglie definite,

```
BiCopSelect(u1, u2, family,…)
```

Si noti che le famiglie tra cui è possibile scegliere sono svariate.

`BiCopGofTest`, test di adattamento di White (1982),

```
BiCopGofTest(u_1, u_2, family, par, par2, method = "white",…)
```

```
BiCopChiPlot(u1, u2, PLOT = TRUE,…)
```

Vine copule

I pacchetti "CDVine" e "VineCopula" nel linguaggio R

- `BiCopLambda`, funzione che calcola la funzione lambda di una copula bivariata,

```
BiCopLambda(u1, u2, family, par, par2,…)
```

- `BiCopHfunc`, funzione che stima la FdR condizionata data una copula bivariata,

```
BiCopHfunc(u1, u2, family, par, par2,…)
```

- `pacotest`, funzione del pacchetto «pacotest» che implementa il test statistico definito *constant conditional correlations test* (CCC),

```
Pacotest(U, W, pacotestOptions)
```

dove `U` rappresenta la matrice dei dati associati alle v.a. condizionate, `W` la matrice dei dati relativi alle v.a. condizionanti e `pacotestOptions` le opzioni relative alla definizione del test statistico.

- `CDVineMLE`, funzione che stima i parametri di massima verosimiglianza di un C/D-vine, partendo da un set iniziale di valori di partenza per i parametri,

```
CDVineMLE(data, family, start=NULL, start2=NULL, type,…)
```

dove `data` rappresenta la matrice delle pseudo-osservazioni, `family` il vettore $n(n-1)/2$ delle famiglie selezionate, mentre `start` e `start2` sono i vettori $n(n-1)/2$ dei valori iniziali dei parametri oggetto di stima.

Vine copule

I pacchetti "CDVine" e "VineCopula" nel linguaggio R

- `CDVineAIC CDVineBIC`, funzioni che calcolano le grandezze AIC e BIC per una C/D-vine copula

```
CDVineAIC(data, family, par, par2, type,…)
```

- `CDVineVuongTest CDVineClarkeTest`, test di Vuong e Clarke tra due C/D-vine n-dimensionali

```
CDVineVuongTest(data, Model1.order, Model2.order, Model1.family,
Model2.family, Model1.par, Model2.par, Model1.par2, Model2.par2,
Model1.type,  Model2.type)
```

```
CDVineClarkeTest(data, Model1.order, Model2.order, Model1.family,
Model2.family, Model1.par, Model2.par, Model1.par2, Model2.par2,
Model1.type,  Model2.type)
```

Agenda

Le Vine Copule

Riferimenti bibliografici

- Aas, K. et al. (2009). « Pair-Copula Construction of Multiple dependence». In: Insurance: Mathematics and Economics 44.2, pp. 182-198.
- Bedford, T. e R.M. Cooke (2001). «Probability Density Decomposition for Conditionally Dependent Random Variables Modeled by Vines». In: Annals of Mathematics and Artificial Intelligence 32, pp. 245-268.
- Bedford, T. e R.M. Cooke (2002). «Vines - a new graphical model for dependent random variables». In: Ann. Statist. 30.4, pp. 1031-1068.
- Clarke, K.A. (2007). «A Simple Distribution-Free test for Nonnested Model Selection». In: Political analysis Advance Acces.
- Czado, C., S. Jeske e M. Hofmann (2013). «Selection strategies for regular vine copulae». In: Journal de la Société Française de Statistique 154, pp. 11-129.
- Dißmann, J. et al. (2013). «Selecting and estimating regular vine copulae and application to financial returns». In: Computational Statistics & Data Analysis 59, pp. 52-69.
- Fisher, N. e P. Switzer (1985). «Chi-plots for assessing dependence». In: Biometrika 72.2, pp. 253-265.
- Genest, C., J.F. Quessy e B. Remillard (2006). «Goodness-of-fit procedures for copula models based on the probability integral transformation». In: Scandinavian Journal of Statistics 33.2, pp. 337-366.

Riferimenti bibliografici

- Haff, I.H. (2012). «Comparison of estimators for pair-copula constructions». In: Journal of Multivariate Analysis 110, pp. 91-105.
- Haff, I.H. (2013). «Parameter estimation for pair-copula constructions». In: Bernoulli 19.2, pp. 462-491.
- Joe, H. (1996). «Families of m-Variate Distributions with Given Margins and $m(m-1)/2$ Bivariate Dependence Parameters». In: Distributions with Fixed Marginals and related Topics, pp. 120-141.
- Kraus, D. e C. Czado (2016). «D-vine copula based quantile regression». In: arXiv:1510.04161v4.
- Kraus, D. e C. Czado (2017). «Growing simplified vine copula trees: improving Dißmann's algorithm». In: arXiv:1703.05203v1.
- Stöber, J., H. Joe e C. Czado (2013). «Simplied pair copula construction - Limitations and extensions». In: Journal of Multivariate Analysis 119, pp. 101-118.
- Tsukahara, H. (2005). «Semiparametric estimation in copula models». In: The Canadian Journal of Statistics 33.2, pp. 357-375.
- Vuong, Q.H. (1989). «Likelihood ratio tests for model selection and non-nested hypotheses». In: Econometrica 57.2, pp. 307-333.
- Wang, W. e M.T. Wells (2000). «Model selection and semiparametric inference for bivariate failure-time data». In: Journal of the American Statistical Association 95.449, pp. 62-72.

Riferimenti bibliografici

- White, H. (1982). «Maximum likelihood estimation of misspecified models». In: Econometrica 50.1.

Pacchetti R:

- Schepsmeier, U. e Brechmann, E.C. (2015). «CDVine». R package version 1.4, https://cran.r-project.org/web/packages/CDVine/index.html
- Nagler, T., Schepsmeier, U. et al. (2021). «VineCopula». R package version 2.4.3, https://cran.r-project.org/web/packages/VineCopula/index.html

Agenda

Le Vine Copule

1. Aspetti Introduttivi
2. Pair-copula Construction (PCC)
3. PCC e Vine: le Vine Copule
4. Stima delle Vine Copule
5. I pacchetti "CDVine" e "VineCopula" nel linguaggio R
6. Riferimenti Bibliografici
7. **Appendice**

Appendice 1: Pair-copula construction per $n = 3$

Variabili aleatorie unidimensionali

Torna alla presentazione

Esempio *(segue)*

La pair-copula construction applicata nel caso in cui $n = 3$*, opera come segue.*

$$f_{1,2,3}(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1,2}\,(x_3|x_1, x_2)$$

Le funzioni di densità univariate condizionate possono essere espresse in termini di copule bivariate,

$$f_{2|1}(x_2|x_1) = \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)} = \frac{c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2)}{f_1(x_1)} = c_{1,2}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2)$$

Mentre $f_{3|1,2}\,(x_3|x_1, x_2)$ *può essere scritta osservando quanto segue,*

$$f_{2,3|1}\,(x_2, x_3|x_1) = \frac{f_{1,2,3}(x_1, x_2, x_3)}{f_1(x_1)} = \frac{f_1(x_1) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1,2}\,(x_3|x_1, x_2)}{f_1(x_1)}$$

$$= f_{2|1}(x_2|x_1) \cdot f_{3|1,2}\,(x_3|x_1, x_2)$$

Da cui

$$f_{3|1,2}\,(x_3|x_1, x_2) = \frac{f_{2,3|1}\,(x_2, x_3|x_1)}{f_{2|1}(x_2|x_1)}$$

Appendice 1: Pair-copula construction per $n = 3$

Variabili aleatorie unidimensionali

Torna alla presentazione

Inoltre, è possibile riformulare $f_{2,3|1}(x_2, x_3|x_1)$ *come segue,*

$$f_{2,3|1}(x_2, x_3|x_1) = c_{2,3|1}\big(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1\big) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1}(x_3|x_1)$$

Dunque,

$$f_{3|1,2}(x_3|x_1, x_2) = \frac{c_{2,3|1}\big(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1\big) \cdot f_{2|1}(x_2|x_1) \cdot f_{3|1}(x_3|x_1)}{f_{2|1}(x_2|x_1)}$$

$$= c_{2,3|1}\big(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1\big) \cdot f_{3|1}(x_3|x_1)$$

e osservando che,

$$f_{3|1}(x_2|x_1) = \frac{f_{3,1}(x_3, x_1)}{f_1(x_1)} = \frac{c_{3,1}(F_3(x_3), F_1(x_1)) \cdot f_3(x_3) \cdot f_1(x_1)}{f_1(x_1)} = c_{3,1}(F_3(x_3), F_1(x_1)) \cdot f_3(x_3)$$

si ottiene,

$$= c_{2,3|1}\big(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1\big) \cdot c_{3,1}(F(x_3), F(x_1)) \cdot f_3(x_3)$$

Appendice 1: Pair-copula construction per $n = 3$

Variabili aleatorie unidimensionali

Torna alla presentazione

Infine, è possibile scrivere la densità $f_{1,2,3}(x_1, x_2, x_3)$ *come*[1]*:*

$$f_{1,2,3}(x_1, x_2, x_3) =$$

$$f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \cdot c_{1,2}(F_1(x_1), F_2(x_2)) \cdot c_{3,1}(F(x_3), F(x_1)) \cdot c_{2,3|1}\left(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1\right)$$

Le copule bivariate presenti nella relazione precedente prendono il nome di pair-copule.

Esempio *(fine)*

[1] *La decomposizione precedente è solo una delle tre possibili fattorizzazioni in* $n = 3$ *dimensioni.*