SAPIENZA
UNIVERSITÀ DI ROMA

La *Pair Copula Construction* nel contesto assicurativo

Dipartimento Di Scienze Statistiche – Scuola di Scienze Statistiche
Scienze Attuariali (XXXV ciclo)

Mariagrazia Rositano
Matricola 1532856

Relatore
Prof. Fabio Baione

Anno Accademico 2023/2024

Tesi discussa il 25 Gennaio 2024
di fronte a una commissione esaminatrice composta da:

Prof.ssa Anna Rita Bacinello (presidente)

Prof.ssa Mariarosaria Coppola

Prof. Massimiliano Menzietti

La *Pair Copula Construction* nel contesto assicurativo
Sapienza Università di Roma

Questa tesi è stata composta con LaTeX e la classe Sapthesis.

Email dell'autore: mariagrazia.rositano@uniroma1.it

Dicono che il vero amore spagini addirittura l'ordine precostituito dell'universo,
vediamo allora se riesce a spaginare persino me.
A mio marito, che mi ha insegnato la cosa più importante della mia carriera,
la cosa più importante della mia vita: qualunque forma assumerà l'amore sarà
topologicamente equivalente al nostro che vale doppio infinito.
Senza di te non sono niente.

Ringraziamenti

Un ringraziamento speciale al mio relatore il Professor Fabio Baione, persone "come te non ne fanno più, hanno buttato lo stampo". Ti ringrazio di avermi introdotto nel mondo della *pair-copula-construction*. Premetto che se avessi capito sin da subito quanto sarebbe stato impegnativo questo percorso, forse avrei fatto *ghosting*: come dice la mia migliore amica (che è una persona speciale e che ringrazio, grazie Viviana) non vedo l'ora di invocare il Comma 22 di Joseph Heller e scappare per salvarmi la vita in tutte le situazioni possibili.

Grazie Fabio, di tutto: grazie per quanto mi hai insegnato, professionalmente e umanamente. Grazie per pazienza e per la fiducia.

E guardandomi indietro prima del Professor Baione, ho imparato tanto da un altro Fabio, il Professor Grasso, il primo che ho incontrato al Dipartimento di Scienze Statistiche, e ancora ho imparato tanto dal Professor Luca Passalacqua e dalla Professoressa Susanna Levantesi che è sempre stata un riferimento durante questi anni. Ringrazio te Fabio e tutti loro. E forse "come voi non ne fanno più, hanno buttato lo stampo". A voi devo tanto, grazie, per la vostra guida e il vostro incoraggiamento durante tutti gli anni di studio in Sapienza, questo rimarrà sempre.

Inoltre ringrazio tutti coloro che durante questo percorso ho tediato con i miei dubbi e domande, con orgoglio ho collezionato le risposte della Professoressa Claudia Czado, del Professor Thomas Nagler e della Professoressa Johanna Nešlehová. Grazie anche a voi.

Unitamente a loro ringrazio il Professor Riccardo Cesari, anche lui ha creduto che potessi fare bene, grazie.

Ringrazio inoltre i miei genitori per aver reso possibile tutto questo. Senza di loro non sarei la persona che sono e ringrazio anche la mia antitesi mio fratello, grazie Francesco, senza di te certamente non sarei quella che sono.

Da ultimo vorrei ringraziare mio marito, Vittorio. Qualunque cosa io dica non sarà mai abbastanza quindi non dirò proprio nulla.

Indice

Introduzione

> Dependence relations between random variables is one of the most widely studied subjects in probability and statistics. The nature of the dependence can take a variety of forms and unless some specific assumptions are made about the dependence, no meaningful statistical model can be contemplated.
>
> Jogdeo (2014)

Overview e obiettivi

È difficile sopravvalutare l'importanza delle misure di dipendenza nella statistica e nella scienza: le interconnessioni e le dipendenze sono da sempre presenti in natura, nelle interazioni umane, nelle leggi che regolano l'economia e nei processi industriali. Tuttavia, si ricorda che correlazione nulla non implica indipendenza (vale però il viceversa) e, pertanto, se si misura la dipendenza di X e Y e si trova che il coefficente di correlazione lineare è nullo, allora si potrebbe sospettare che ci sia una relazione causale tra X e Y, che non è lineare. Questo è il tipico esempio in cui il legame tra variabili X e Y è non monotono. Un buon esempio di quanto detto lo si può trovare nell'articolo Volokh (2015) che dice:

> *Now of course this doesn't prove that gun laws have no effect on total homicide rates. Correlation, especially between just two variables, doesn't show causation.*

Comprendere la qualità tra le interconessioni e saperle rappresentare in modo adeguato non solo è d'interesse centrale nel campo della finanza e delle scienze attuariali (per i rischi sistemici e l'*upper tail analysis*) ma in generale assume un ruolo nella politica sociale: basti pensare alla scuola (alla qualità del livello d'istruzione offerto e il numero di studenti in una classe), la riforma dell'Istituto della pena, la sanità, e la previdenza nella forme di presenza "pubblica" e "privata". Attraverso lo studio della dipendenza tra le variabili si può rispondere a domande come: il degrado sociale appartiene alle periferie? Se sì, come può essere reindirizzato?

L'interesse e lo studio sistematico dei rischi congiunti si annoda alla fine del diciannovesimo secolo tra la scuola statistica anglosassone (Pearson e Fischer) e quella italiana (Gini e Pietra). L'impegno profuso congiuntamente da queste scuole, unitamente alla partecipazione della ricerca su tale argomento in tutta Europa, ha definito un *framework* teorico monumentale che culmina in due fondamentali risultati presentati nello stesso anno. Correva l'anno 1959, Rényi (1959) proponeva gli assiomi che avrebbero definito le proprietà per una misura di dipendenza (prima di allora non si era mai parlato di misura di dipendenza) e Abe Sklar (1959) enunciava l'ononimo teorema che sarebbe stato il più importante dell'intera teoria delle copule.

Anche dopo diversi anni che si studiano le copule, persino nella definizione può capitare di cogliere qualcosa di nuovo. E al netto della ricchezza dei significati che appartengono a questa forma funzionale, la domanda che ci si pone è: perchè scrivere oggi una tesi di dottorato sulle copule? Sicché su tale argomento ne sono state scritte diverse.

Si potrebbe parlare di copule come se ne parlava dieci anni fa, anche quindici anni, e comunque non si avrebbe torto: sarebbe comunque una tesi attuale, ancor più nel perimetro delle scienze attuariali. La variazione del livello prudenziale introdotta da *Solvency* 2, che si risolve in un nuovo *framework* di vigilanza, rappresentato dall'approccio *bottom-up* – in cui moduli di rischio si aggregano dal basso verso l'alto – attraverso una struttura pre-confezionata stabilita dal regolatore europeo, naturalmente conferisce alle copule il ruolo di *competitor* a quella pre-stabilita. E come annunciato, parlare di copule con questa chiave consentirebbe comunque di vincere e si vincebbe a mani basse.

Ma oggi, più di dieci anni fa, più di quindici anni fa, ha senso parlare di copule e parlarne ha un sapore diverso se si inserisce questa teoria in un contesto ancora più ampio, e proprio dentro questo perimetro acquista un lustro nuovo, diventando non solo qualcosa che "sfida" Solvency e le dipendenze "semplici" ma definisce solchi "nuovi" nel campo della ricerca. Infatti nel corso degli ultimi venti anni, la letteratura è stata concitata e i "limiti" di questa teoria sono stati riscritti grazie tre "ingredienti":

- il *coniugo* tra la teoria copule e la teoria dei grafi;
- lo sviluppo di soluzioni *software* che consentono l'implementazione dei modelli statistici;
- l'innovazione tecnologica che sancisce l'aumento progressivo della potenza di calcolo degli elaboratori.

Sotto questi tre ingredienti, tutti e tre da considerare in modo congiuto, si innesta una nuova metodologia che prende il nome di *pair-copula construction* (*PCC*) che nasce come "panacea" ai "limiti" osservati nella teoria delle copule. È risaputo che il concetto di copula, ovvero di distribuzione bivariata con marginali unidimensionali uniformi, può essere esteso al caso multivariato e in questo caso si parla di n-copula. Molte famiglie di copule multivariate sono state trovate negli anni (Dall'Aglio, Kotz e Salinetti (1991); Doruet Mari e Kotz (2001); Joe (1997); Nelsen (2006)). Inoltre la letteratura insegna che molte delle distribuzioni parametriche

bivariate con marginali continue, come la normale, hanno la loro corrispondente copula multivariata. Tuttavia, queste generalizzazioni sono spesso "limitate". Infatti Joe (1997) dimostra che più aumenta la dimensione della copula e più essa perde la capacità di cogliere la "vera" dipendenza tra variabili. Brechmann e Schepsmeier (2013) affermano:

> *The use of copulas is however challenging in higher dimensions where standard multivariate copulas suffer from rather inflexible structures.*

In tre lustri, circa quindici anni di ricerca, è stata ridata nuova linfa a questa teoria. Difatti, da una parte la ricerca è proseguita nella speranza di mantenere la facilità e la flessibilità che si trova nelle copule di tipo parametrico e in particolare nelle sue generalizzazioni, le copule archimedee innestate (*full-nested-archimedean copula*) coniugando questa impostazione alla teoria dei grafi, definendo il concetto di *R-vine-copula*. Cooke (1997) fù il primo nel 1997 a parlare di *vine-dependent* che semplicemente può essere tradotta come "dipendenza a vite".

La *pair-copula construction* (*PCC*) detta anche *R-Vine Copula*, in italiano "copula a *R-vite*", è un tecnica grafica di modellizzazione che consente di costruire strutture multivariate attraverso l'impiego esclusivo di copule bivariate semplici o condizionate. La "svolta statistica" per le *R-vine* è stata quando Aas et al. (2009) hanno descritto una tecnica di inferenza per *R-vine* che faceva uso della massima verosimiglianza come stima: la funzione di log-verosomiglianza può essere scomposta nella somma della verosomiglianza di ogni singolo arco. Da allora la teoria delle copule *R-vine* è stata costantemente migliorata e ampiamente studiata in letteratura compresa nell'analisi bayesiana.

Il risultato del *building-blocks approach* mediante *PCC* è letteralente una cascata di copule, di parametri e di modelli. Infatti:

- un modello avente n variabili è formato da $\binom{n}{2} = \frac{n(n-1)}{2}$ bi-copule;
- per ogni bi-copula nel modello, a meno che sia una bi-copula indipendente, corrisponde almeno un parametro;
- esistono, per dimensione n, $\frac{n!}{2} \times 2^{\binom{n-2}{2}}$ possibili alternative per la struttura grafica (Morales Napoles (2011)).

I problemi dell'enumerazione che negli anni sono stati affrontati attraverso l'utilizzo di modelli troncati (modelli in cui si presuppone l'indipendenza condizionata) e modelli sparsi (modelli in cui si presuppone l'indipendenza a priori per un certo numero di variabili): tuttavia questa strada non è stata risolutiva. Infatti, già in dimensione $n = 11$ si hanno circa 1.3×10^{18} di modelli selezionabili. In tal senso la ricerca si è spinta verso l'individuazione di strategie che portassero a soluzioni approssimate usando algoritmi *greedy*, ovvero algoritmi che facessero scelte locali (nel perimetro delle copule a *R-vine* significa "albero-per-albero"). In quanto la log-verosomiglianza di un arco dipende anche dai parametri determinati negli archi che si sono uniti per crearlo nei livelli superiori: l'algoritmo più utilizzato è quello di Dißmann et al. (2013). Congiuntamente sono stati creati nuovi indicatori per valutare la bontà di adattamento del modello (Nagler, Bumann e Czado (2019)) e

di comparazione tra modelli. D'altro canto, la definizione di tale impianto teorico ha necessariamente portato allo sviluppo di soluzioni *software* che consentissero la concreta implementazione. La *pair-copula construction* non è possibile se non si fa uso di *software* e librerie *ad-hoc* per l'implementazione di tale metodologia. Unitamente alla disponibilità di soluzioni *software* per lo studio della *vine-copula construction*, la potenza degli elaboratori ci consente di trovare soluzioni in tempi accettabili. E in questo senso, e per i motivi sopra citati, cinque anni fa non si poteva scrivere questo elaborato come è stato scritto oggi. Inoltre già cinque anni fa la teoria nel campo della *PCC* non era matura come lo è ora.

Premesso quanto sopra, la domanda di ricerca di questa tesi è: "*la metodologia* PCC *è appropriata per essere usata in ambito attuariale*?". L'utilizzo della *PCC* nelle scienze attuariali ha senso per diversi motivi:

- la centralità dello studio delle strutture di dipendenza nelle scienze attuariali;
- la necessità di modellare fenomeni multivariati nel campo delle assicurazioni;
- la dipendenza di tali fenomeni non sempre è rappresentabile attraverso strutture "semplici" e spesso si osserva la presenza di rischi che presentano asimmetrie e code leptocurtiche.

La presente ricerca mette in luce che tale metodologia è affetta da diverse criticità e ne analizza l'impatto sia in ambito teorico che nel contesto attuariale. In particolare, le principali problematiche affrontate sono:

- il problema dell'enumerazione ovvero il numero di modelli di tipo *PCC* esistenti definiti per una dimensione fissata;
- l'elevato numero di parametri dei modelli *PCC* e le strategie adottate per ridurre la complessità del modello;
- le difficoltà nel confronto tra modelli *PCC* concorrenti;
- gli algoritmi per la selezione del modello e le problematiche computazionali connesse;
- il problema della modellazione di variabili discrete utilizzando modelli continui e la non unicità della funzione di copula;
- il comportamento di questa metodologia facendo uso di *dataset* attuariali.

Va osservato che molte di queste problematiche non sono "distintive"della *PCC* ma sono condivise da metodologie concorrenti o addirittura connaturate al problema. Per esempio, il numero di parametri e la difficoltà di interpretare i risultati sono condivisi con tutti i modelli multivariati di dimensione sufficentemente alta e similmente vale per la selezione del modello.

Nelle scienze attuariali, il trattamento delle variabili discrete, sia esse variabili ordinali e/o categoriali nominali (il problema dell'inversione è un'eredità teorica delle copule che si propaga nella *PCC*), è fondamentale. Marshall (1996) afferma "*things are not so simple with discrete marginals*". In letteratura tale problema è

stato analizzato in maniera limitata nelle copule e per le *R-vine* la letteratura si sofferma solo su questioni di tipo computazionali (problema di integrazioni numeriche nei punti di discontinuità). Embrechts (2009) quando parla di copule nel contesto discreto dice "*everything that can go wrong, will go wrong*" ed effettivamente sia dal punto di vista teorico che nell'implementazioni pratiche, quando si ricorre all'uso di variabili discrete tutto quello che potrebbe andare male, va male. Prima di tutto "si inciampa" quando si fa uso di variabili discrete perchè la non continuità delle marginali fa si che la funzione di copula non sia unica, e dunque il problema dell'inversione definisce non solo aspetti problematici dal punto di vista teorico ma anche nell'implementazione della *PCC*. Genest e Nešlehová (2007) mostrano che il valore del τ_K di Kendall in *senso* probabilistico non è più uguale al valore del τ_C di Kandall nel *senso* della copula. La letteratura come detto è scarsa e in tal senso non offre soluzioni: alcune volte le variabili si rendono continue, altre si effettua il *jittering* (Denuit e Lambert (2005) e Nagler (2018)) delle variabili. In presenza di variabili categoriali si ricorre all'utilizzo di variabili *dummy* per recuperare il "senso" della concordanza.

Nello studio pratico della metodologia con i *dataset* attuariale si è cercato di risolvere le problematiche con approcci innovativi, in particolare si è definito un approccio *tailored-mixed*, ovvero una metodologia che utilizza l'algoritmo *greedy* di Dißmann solo per alcuni "tratti" e sfrutta la filtrazione sugli alberi precedenti. Inoltre l'elaborato analizza il problema della *PCC* nel contesto inferenziale, identificando le principali semplificazioni che ne consentono l'applicazione: selezionare la *R-vine* migliore per un determinato *dataset* è tutt'altro che un problema banale. L'algoritmo che viene utilizzato nella ricerca attuale non trova un ottimo globale e l'analisi della sua qualità si basa solo su risultati campionari. Seppur la selezione livello-per-livello e l'ottimo globale potrebbero non corrispondere, tale problema è stato dimostrato trascurabile poichè asintoticamente si raggiunge l'ottimo (Czado (2019)). Gruber e Czado (2015) hanno dimostrato che in scenari di simulazione in dimensione sei l'algoritmo *greedy* di Dißmann raggiunge almeno il 75% della vera log-verosimiglianza.

In particolare, la risposta "attuariale" è la chiave della "domanda di ricerca" in quanto le applicazioni fino ad oggi considerate della *PCC* nel perimetro assicurativo nè considerano e nè analizzano gli aspetti fondamentali di questa metodologia come ad esempio il delicato aspetto relativo alla dimensionalità e gli sviluppi in campo computazionale e il problema dell'inversione che viene affrontato sempre in modo marginale.

Infatti, in letteratura si osservano pochi esempi di applicazioni delle copule in campo assicurativo e spesso queste si limitano a modellazioni di tipo bivariato (Czado, Kastenmeier et al. (2012), Krämer et al. (2013) e Shi e Yang (2018)). In dimensione superiore si osserva solo il contributo di Ricotta (2019) il quale mostra un approccio con 4 variabili aleatorie. Si sottolinea che in tali trattazioni non si ci si interroga né rispetto al problema dell'inversione, non si fa uso nè di Dißmann e non si affronta in modo organico il problema dell'enumerazione che abbiamo visto risolversi durante la trattazione in podo "parziale" attraverso approccio computazionale.

La tesi è organizzata come segue. Nel capitolo 1 si definisce la funzione di copula e si presentano i principali risultati; il capitolo 2 espone gli aspetti fondamentali relativi alla dipendenza tra variabili aleatorie; nel capitolo 3 si definiscono le principali

famiglie di copule e le più recenti costruzioni moderne; nel capitolo 4 si presenta la *pair-copula construction*; nel capitolo 5 si affronta il problema inferenziale nella *pair-copula construction*; nel capitolo 6 sono trattate le variabili discrete nel contesto della *PCC*; il capitolo 7 illustra le librerie che vengono utilizzate per implementare la *PCC*; e i capitoli da 8 a 10 forniscono l'applicazione della *PCC* nel perimetro delle scienze attuariali. Infine, le conclusioni sono contenute nel capitolo 11.

Il presente lavoro di ricerca fornisce dei contributi e degli aspetti innovativi per i seguenti punti:

- Razionalizzazione sistematica delle fonti e l'applicazione della *PCC* nel contesto delle scienze attuariali;
- Rilevazione dei limiti d'applicazione di tale metodologia dal punto di vista teorico tenendo conto alla specificità del mercato assicurativo;
- L'analisi delle risultanze mediante l'uso dell'algoritmo di Dißmann utilizzando come *driver* di ricerca le diverse *tipologie* delle variabili trattate e la *dimensionalità*;
- L'implementazione della modellizzazione *PCC* facendo uso della libreria `rvinecopulib` definendo le frontiere dal punto di vista computazionale.

Capitolo 1

Le Copule

> Féron (1956), in studying three-dimensional distributions had introduced auxiliary functions, defined on the unit cube, that connected such distributions with their one-dimensional margins. [...] While writing this, I decided I needed a name for these functions. Knowing the word "copula" as a grammatical term for a word or expression that links a subject and predicate, I felt that this would make an appropriate name for a function that links a multidimensional distribution to its one-dimensional margins, and used it as such.
>
> A. Sklar (1996)

Lo studio della dipendenza tra variabili aleatorie può essere affrontato attraverso diversi strumenti metodologici, più o meno raffinati. Le copule consentono di separare le distribuzioni marginali dalla rispettiva struttura di dipendenza, la quale può essere rappresentata in modo univoco attraverso la funzione di copula. Questo capitolo si apre introducendo il concetto di copula. Prosegue quindi con alcune proprietà che caratterizzano tali funzioni fino ad arrivare al teorema più importante di tutta la teoria delle copule: il teorema di Sklar (teorema 1.2) che verrà presentato nelle sua due formulazioni. Nel seguito, si presenteranno altri risultati fondamentali di questa teoria. Da ultimo, si segnala che la seguente trattazione non ha finalità "enciclopedica": non ne aveva il libro scritto da Nelsen (2006)[1], figurasi questo che

[1] The book begins with the basic properties of copulas and then proceeds to present methods for constructing copulas and to discuss the role played by copulas in modeling and in the study of dependence. The focus is on bivariate copulas, although most chapters conclude with a discussion of the multivariate case. As an introduction to copulas, it is not an encyclopedic reference, and thus it is necessarily incomplete many topics that could have been included are omitted. (Nelsen (2006))

vuole "traghettare" il lettore verso il *topic* della tesi, che è quello della *pair-copula construction*.

1.1 Definizioni e Proprietà Base

Si consideri un vettore aleatorio $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ avente distribuzioni marginali:

$$F_1(x_1), F_2(x_2), \ldots, F_n(x_n) \tag{1.1}$$

e distribuzione congiunta:

$$H(x_1, x_2, \ldots, x_n). \tag{1.2}$$

Si ricorda inoltre che:

$$F_i(x_i) \coloneqq \mathbb{P}\{X_i \leq x_i\}, \tag{1.3}$$

$$H(x_1, x_2, \ldots, x_n) \coloneqq \mathbb{P}\{X_1 \leq x_1; X_2 \leq x_2; \ldots ; X_n \leq x_n\} \tag{1.4}$$

e che:

$$F_i \circ X_i \sim \mathrm{U}(0,1) = \mathrm{U}(\mathbb{I}) \tag{1.5}$$

per ogni $i \in \{1, \ldots, n\}$, dove $\mathbb{I} = [0,1]$.

Sia $\operatorname{Ran} \boldsymbol{X} \subseteq \mathbb{R}^n$ l'insieme dei valori che il vettore aleatorio $\boldsymbol{X}$ può assumere. A ciascuna n-upla di numeri reali $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \operatorname{Ran} \boldsymbol{X} \subseteq \mathbb{R}^n$ si può associare una $(n+1)$-upla

$$\big(F_1(x_1), F_2(x_2), \ldots, F_n(x_n), H(x_1, \ldots, x_n)\big) \in \mathbb{I}^{n+1} = [0,1]^{n+1} \subset \mathbb{R}^{n+1}. \tag{1.6}$$

Questo elemento di $\mathbb{I}^{n+1} = \mathbb{I}^n \times \mathbb{I}$ può essere visto come il grafico di una funzione $\mathbb{I}^n \to \mathbb{I}$ che associa ad ogni n-upla ordinata di valori di distribuzioni univariate $\big(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)\big)$ il valore della distribuzione congiunta $H(x_1, \ldots, x_n)$. Questo tipo di funzioni perde il nome di copula ed è l'argomento di questo capitolo.

Definizione 1.1 (Copula)**.** *Una* copula n-dimensionale *è una funzione di distribuzione multivariata su* $\mathbb{I}^n = [0,1]^n \subset \mathbb{R}^n$ *con distribuzioni marginali uniformi standard.*

Il concetto di copula può essere definito in modo alternativo. Si consideri una copula n-dimensionale $C(\boldsymbol{u}) = C(u_1, \ldots, u_n)$, è evidente che quest'ultima sia una funzione dall'ipercubo unitario n-dimensionale $\mathbb{I}^n$ nell'intervallo unitario $\mathbb{I}$ ovvero una mappa della forma $C : \mathbb{I}^n \to \mathbb{I}$. Inoltre, essendo una distribuzione multivariata soddisfa varie proprietà aggiuntive.

Teorema 1.1 (Definizione alternativa di copula)**.** *Ogni copula n-dimensionale C soddisfa le seguenti proprietà:*

1. *per ogni $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{I}^n$ si ha che $C(\boldsymbol{u}) = 0$ se almeno una coordinata u_i è nulla;*

2. *$C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ per tutti gli $i \in \{1, \ldots, n\}$ e $u_i \in \mathbb{I}$.*

3. *Vale la* disugualianza triangolare*, ovvero per tutti gli* $\boldsymbol{u}_1 = (u_{1,1}, \ldots, u_{1,n})$ *e* $\boldsymbol{u}_2 = (u_{2,1}, \ldots, u_{2,n})$ *in* $\mathbb{I}^n$ *con* $u_{1,i} \leq u_{2,i}$ *per ogni* $i \in \{1, \ldots, n\}$ *si ha che:*

$$\sum_{j_1=1}^{2} \cdots \sum_{j_n=1}^{2} (-1)^{j_1+\cdots+j_n} C(u_{j_1,1}, \ldots, u_{j_n,n}) \geq 0. \tag{1.7}$$

Vale anche il viceversa ovvero ogni funzione $C\colon \mathbb{I}^n \to \mathbb{I}$ *che soddisfa le proprietà sopraindicate è una copula.*

D'ora in poi si userà la notazione $\boldsymbol{u} = (u_1, \ldots, u_n)$ per i vettori aleatori con distribuzioni marginali uniformi in $\mathbb{I}$.

Poichè si tratteranno principalmente copule bivariate, risulta utile presentare il teorema 1.1 per la dimensione 2.

Corollario 1.1. *Una funzione* $C\colon \mathbb{I}^2 \to \mathbb{I}$ *è una copula bivariata se e solo se soddisfa le seguenti proprietà:*

1. $C(u, 0) = 0 = C(0, v)$*;*
2. $C(u, 1) = u$ *e* $C(1, v) = v$*;*
3. *per ogni* $(u_1, v_1), (u_2, v_2) \in \mathbb{I}^2$ *tali che* $u_1 \leq u_2$ *e* $u_2 \leq v_2$ *risulta che*

$$C(u_1, v_1) + C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) \geq 0. \tag{1.8}$$

Osservazione 1.1. *Se* C *è una copula* n*-dimensionale allora tutte le sue marginali di dimensione maggiore di uno sono a loro volta copule.*

Definizione 1.2 (Densità di Copula)**.** *Sia* C *una copula assolutamente continua di dimensione* n*, si definisce* densità di copula *per una funzione di copula* C *la seguente funzione:*

$$c(u_1, \ldots, u_n) \coloneqq \frac{\partial^n}{\partial u_1, \ldots, \partial u_n} C \tag{1.9}$$

1.2 Teorema di Sklar (1959)

L'importanza delle copule nello studio delle funzioni di distribuzione multivariate è riassunta dal seguente elegante teorema, che dimostra, in primo luogo, che tutte le distribuzioni multivariate contengono copule e, in secondo luogo, che le copule possono essere utilizzate insieme a funzioni di distribuzione univariate per costruire funzioni di distribuzione multivariate. Il teorema è stato proposto per la prima volta in Abe Sklar (1959).

Teorema 1.2 (Teorema di Sklar)**.** *Sia* H *una funzione di distribuzione congiunta con marginali* $F_1, \ldots, F_n$*, allora esiste una copula* $C\colon \mathbb{I}^n \to \mathbb{I}$ *tale che per tutti gli* $\boldsymbol{x} = (x_1, \ldots, x_n) \in \overline{\mathbb{R}} = [-\infty, +\infty]$*:*

$$H(x_1, \ldots, x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big) \tag{1.10}$$

Se le marginali sono continue allora C *è unica. Diversamente,* C *è univocamente determinata su* $\operatorname{Ran} F_1 \times \cdots \times \operatorname{Ran} F_n$ *dove* $\operatorname{Ran} F_i = F_i(\overline{\mathbb{R}})$ *rappresenta il codominio*

di F_i. Inoltre, se C è una n-copula e $F_1, \ldots, F_n$ sono funzioni di distribuzioni, allora la funzione H definita sopra è una funzione di distribuzione n-dimensionale di marginali $F_1, \ldots, F_n$.

Il teorema garantisce che per variabili aleatorie continue, le marginali univariate e la struttura di dipendenza multivariata possano essere separate univocamente e che la copula descriva completamente la struttura di dipendenza. Il teorema può essere invertito, cosicché la copula può esprimersi in termini di funzione di ripartizione congiunta e funzioni dei quantili delle marginali. Si osserva che non sempre le distribuzioni marginali sono strettamente crescenti e dunque non possiedono la funzione inversa; sia A. McNeil, Frey e Embrechts (2005) che Nelsen (2006) hanno affrontato questo tema, definendo il primo la funzione inversa generalizzata e l'altro la funzione quasi-inversa. Come si osserva le due definizioni sono del tutto equivalenti; questo capitolo le introduce entrambe per due motivi: il primo perché spesso in letteratura si fa confusione e si pensa che siano concetti distinti, e il secondo per rendere il lettore consapevole di entrambe le notazioni.

Definizione 1.3 (Funzione inversa generalizzata)**.** *Sia F una funzione di distribuzione, si denota con $F^{\leftarrow}$ la sua* funzione inversa generalizzata *definita nel modo seguente:*

$$F^{\leftarrow}(u) = \inf\{x|F(x) \geq u\} = \sup\{x|F(x) \leq u\} \tag{1.11}$$

Definizione 1.4 (Funzione quasi-inversa)**.** *Sia F la funzione di distribuzione, allora si definisce con la funzione quasi-inversa nel modo seguente:*

$$F^{(-1)}(u) = \inf\{x|F(x) \geq u\} = \sup\{x|F(x) \leq u\} \tag{1.12}$$

Ovviamente se F è strettamente crescente allora la funzione inversa generalizzata (e dunque la funzione quasi-inversa) concidono con la funzione inversa nella sua accezione ordinaria. Attraverso la funzione inversa generalizzata (o funzione quasi-inversa) si può introdurre il teorema di Sklar nella sua forma inversa.

Teorema 1.3 (Formulazione inversa del teorema di Sklar)**.** *Sia H una funzione di distribuzione congiunta con marginali $F_1, \ldots, F_n$. Se $F_i^{\leftarrow}$ è l'inversa generalizzata di F_i per ogni $i \in \{1, \ldots, n\}$, allora la funzione $C\colon \mathbb{I}^n \to \mathbb{I}$ definita come*

$$C(u_1, \ldots, u_n) \coloneqq H\big(F_1^{\leftarrow}(u_1), \ldots, F_n^{\leftarrow}(u_n)\big) \tag{1.13}$$

è una copula.

Se $h, f_1, \ldots, f_n$ sono le densità di $H, F_1, \ldots, F_n$, la formulazione dell'equazione (1.13) in termine di densità di copula è la seguente

$$c(u_1, \ldots, u_n) = \frac{h\big(F_1^{\leftarrow}(u_1), \ldots, F_n^{\leftarrow}(u_n)\big)}{\prod_{i=1}^{n} f_i(F_i^{\leftarrow}(u_i))}. \tag{1.14}$$

Le equazioni (1.10) e (1.13) sono fondamentali per tutta la teoria della copule:

- la forma diretta del teorema mostra come generare distribuzioni congiunte a partire dalle distribuzioni marginali attraverso la funzione la funzione C;

- la forma inversa del teorema ci consente di estrarre la funzione di copula partendo dalla funzione di ripartizione multivariata. Inoltre, attraverso questa riusciamo a creare famiglie di copule parametriche a partire da distribuzioni multivariate parametriche.

Si noti che l'artificio geometrico contenuto nella formulazione inversa (teorema 1.3) opera di fatto una normalizzazione delle componenti nello spazio di probabilità riportando la variabile congiunta a distribuzioni uniformi attraverso la densità di copula. Di fatto le v.a. uniformi assumono lo stesso ruolo dei versori della geometria euclidea. Inoltre, grazie alle modalità in cui opera questa trasformazione la funzione di copula non genera "distorsioni" tra i punti della distribuzione di partenza e di arrivo, senza "perdere" le informazioni contenute nella congiunta. La "speciale trasformazione di scala" è quella quantilica, e proprio grazie a questa trasformazione si riesce ad *estrarre* tutta la dipendenza contenuta nelle diverse marginali all'interno della funzione di copula C. Infatti nella equazione (1.13) si può vedere come le copule esprimano la dipendenza su una scala quantilica poiché il valore $C(u_1, \dots, u_n)$ è la probabilità congiunta che X_1 si trova "al di sotto" del suo u_1-quantile, X_2 si trova "al di sotto" del suo u_2-quantile e così via (A. McNeil, Frey e Embrechts (2005)).

Definizione 1.5 (Trasformazione su scala quantilica)**.** *Sia F una distribuzione con inversa generalizzata $F^{\leftarrow}$. Se $U \sim \mathrm{U}(0,1)$ ha una distribuzione uniforme standard allora:*

$$\mathbb{P}\{F^{\leftarrow}(U) \leq x\} = \mathbb{P}\{\inf\{z|F(z) \geq U\} \leq x\} = F(x). \tag{1.15}$$

Dove per completezza si ricorda la definizione di quantile.

Definizione 1.6 (Quantile di ordine α)**.** *Il quantile di ordine α per una variabile aleatoria Z è dato dalla seguente:*

$$q_\alpha(z) = \inf\{z \in F(z)|\, \mathbb{P}\{Z \geq \alpha\}\}$$

Di seguito due esempi di come si declina il teorema di Sklar nella sua forma inversa:

Esempio 1.1 (Copula di una distribuzione Gaussiana multivariata)**.** *Siano rispettivamente Φ e Φ^n_Σ le funzioni di distribuzione di una normale standard e di una normale standard multivariata con matrice aventi Σ come matrice di correlazione lineare, si definisce* copula Gaussiana *di dimensione n la seguente funzione:*

$$C(u_1, \dots, u_n) = \Phi^n_\Sigma(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)) \tag{1.16}$$

Esempio 1.2 (Copula di una distribuzione bivariata esponenziale di Gumbel)**.** *Sia H_θ la distribuzione bivariata della funzione esponenziale di Gumbel definita nel modo seguente:*

$$H_\theta(x,y) = \begin{cases} 1 - \exp(-x) - \exp(-y) + \exp(-x-y-\theta xy) & \text{per } x \geq 0, y \geq 0 \\ 0 & \text{altrimenti} \end{cases} \tag{1.17}$$

dove $\theta \in \mathbb{I}$. Allora le distribuzioni marginali sono esponenziali con funzione inversa generalizzata pari a:

$$F^{\leftarrow}(u) = -\ln(1-u) \tag{1.18}$$

$$G^{\leftarrow}(v) = -\ln(1-v) \tag{1.19}$$

per ogni $u, v \in \mathbb{I}$. *Quindi la copula corrispondente sarà:*

$$C_\theta(u,v) = u + v - 1 + (1-u)(1-v)\exp(-\theta \ln(1-u)\ln(1-v)). \tag{1.20}$$

Tra le diverse proprietà delle copule si ricorda che la funzione di copula è invariante per a trasformazioni delle marginali strettamente crescenti.

1.3 Bande di Frechèt-Hoeffding

La disuguaglianza nota come bande, o limiti, di Frechèt-Hoeffding è di fondamentale importanza per la teoria delle copule.

Si considerino le seguenti tre funzioni M^n, Π^n e W^n definite su $\mathbb{I}^n$. Sia $\boldsymbol{u} = (u_1, \ldots, u_n)$, allora:

$$M^n(\boldsymbol{u}) \coloneqq \min\{u_1, \ldots, u_n\} \tag{1.21}$$

$$\Pi^n(\boldsymbol{u}) \coloneqq \prod_{i=1}^{n} u_i \tag{1.22}$$

$$W^n(\boldsymbol{u}) \coloneqq \max\Big\{0, \Big(\sum_{i=1}^{n} u_1\Big) - 1\Big\} \tag{1.23}$$

Teorema 1.4 (Teorema di Fréchet-Hoeffding)**.** *Le funzioni* M^n *e* Π^n *sono copule per ogni* $n \geq 2$. *La funzione* W^n *è una copula solo per* $n = 2$ *(A. McNeil, Frey e Embrechts (2005)).*

Inoltre, valgono i seguenti limiti *o* bande di Frechèt-Hoeffding*: sia* C *una qualsiasi* n*-copula allora per ogni* $\boldsymbol{u} \in \mathbb{I}^n$ *vale la seguenza relazione:*

$$W^n(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M^n(\boldsymbol{u}). \tag{1.24}$$

Le funzioni W^n e M^n devono il proprio nome al teorema 1.4.

Definizione 1.7. W^n *prende il nome di* limite inferiore di Frechèt-Hoeffding*, mentre* M^n *viene chiamato* limite superiore di Frechèt-Hoeffding*.*

Definizione 1.8. Π^n *è detta* copula indipendente *o* copula prodotto*.*

Utilizzando il teorema 1.2 si può riscrievere il teorema 1.4 in termini di distribuzioni multidimensionali e le sue marginali.

Proposizione 1.1. *Sia* H *una funzione di distribuzione* n*-dimensionale con marginali* $F_1, \ldots, F_n$ *allora:*

$$\max\Big\{0, \Big(\sum_{i=1}^{n} F_i(x_i)\Big) - 1\Big\} \leq H(\boldsymbol{u}) \leq \min\{F_1(x_1), \ldots, F_n(x_n)\} \tag{1.25}$$

Sebbene W^n, come affermano A. McNeil, Frey e Embrechts (2005), non sia una copula per $n \geq 3$, il limite inferiore della disugualizanza è il migliore possibile come spiega Nelsen (2006), nel senso che per dimensioni strettamente maggiori di 2 e per ogni $\boldsymbol{x}$ nell'ipercubo n-dimensionale esiste una copula $C_{\boldsymbol{x}}$ tale che $C_{\boldsymbol{x}}(\boldsymbol{x}) = W^n(\boldsymbol{x})$.

Attraverso il teorema 1.4 si può introdurre i concetti di *co*-monotonia, *contro*-monotonia e indipendenza.

1.4 *Co*-monotonicità e *contro*-monotonicità

Per i concetti di *co*-monotonicità e *contro*-monotonicità si fa riferimento a quanto proposto da Schmeidler (1986) e Yaari (1987).

Definizione 1.9 (Variabili *co*-monotoniche)**.** *Due variabili aleatorie X_1 e X_2 sono* co-monotoniche *se la loro funzione di ripartizione congiunta $H(x_1, x_2)$ è:*

$$H(x_1, x_2) = \min\{F_1(x_1), F_2(x_2)\}. \tag{1.26}$$

Definizione 1.10 (Variabili *contro*-monotoniche)**.** *Due variabili aleatorie X_1 e X_2 la cui funzione di ripartizione congiunta $H(x_1, x_2)$ è:*

$$H(x_1, x_2) = \max\{0, F_1(x_1) + F_2(x_2) - 1\} \tag{1.27}$$

sono dette contro-monotoniche*.*

Usando le funzioni M^n e W^n tale definizione può essere riscritta come:

Proposizione 1.2. *Due variabili aleatorie X_1 e X_2 sono* co-*monotone se il vettore (X_1, X_2) ha come copula M^2. Sono invece* contro-*monotone se il vettore (X_1, X_2) ha come copula W^2.*

Vale inoltre il seguente teorema:

Teorema 1.5. *Due variabili aleatorie X_1 e X_2 sono* co-*monotone se e solo se esiste una variabile aleatoria Y e due funzioni f_1 e f_2 non decrescenti tali che:*

$$(X_1, X_2) = \big(f_1(Y), f_2(Y)\big) \tag{1.28}$$

Per la *contro*-monotonicità vale invece il seguente:

Teorema 1.6. *Due variabili aleatorie X_1 e X_2 sono* contro-*monotone se e solo se*

$$(X_1, X_2) = \big(g_1(Y), g_2(Y)\big) \tag{1.29}$$

per una qualche una variabile aleatoria Y e due funzioni g_1 e g_2 con la prima crescente e la seconda decrescente, o viceversa.

Poichè M^n è una copula anche per $n > 2$, il concetto di *co*-monotonicità può essere esteso a dimensioni superiori, la stessa cosa non vale invece per la *contro*-monotonicità, che vale solo per la dimensione 2.

Definizione 1.11 (n variabili *co*-monotoniche)**.** *Le variabili aleatorie $X_1, \ldots, X_n$ sono* co-*monotoniche se il vettore $\boldsymbol{X} = (X_1, \ldots, X_n)$ ha come copula M^n.*

Un analogo del teorema 1.5 vale anche per dimensioni superiori.

Teorema 1.7. *Le variabili aleatorie $X_1, \ldots, X_n$ sono* co-*monotone se e solo se esiste una variabile aleatoria Y e n funzioni $z_1, \ldots, z_n$ non decrescenti tali che:*

$$(X_1, \ldots, X_n) = \big(z_1(Y), \ldots, z_n(Y)\big). \tag{1.30}$$

La copula associata a due rischi *co*-monotonici deve quindi essere il limite superiore della disuguaglianza di Frechèt-Hoeffding; viceversa per variabili contro-monotoniche, la copula corrisponde al limite inferiore. Si nota che due rischi co-monotonici sono comunemente monotonici; a differenza dei rischi *contro*-monotonici conseguentemente non sono in grado di compensarsi l'un l'altro. In termini concreti la *co*-monotonicità è associabile alla nozione di *pooling risk* mentre la contro-monotonicità corrisponde alla nozione di *non-pooling risk*. Si ricorda che Puccetti e Scarsini (2010) hanno generalizzato la nozione di *co*-monotonicità e *contro*-monotonicità all'ambiente multivariato e che un'importante conseguenza della proprietà di *co*-monotonicità e *contro*-monotonicità risulta essere l'additività sui quantili. Infatti accade che se X e Y godono della proprietà *co*-monotonicità o *contro*-monotonicità, accade che, dato il quantile q di ordine α vale la seguente relazione:

$$q_\alpha(X_1, X_2) = q_\alpha(X_1) + q_\alpha(X_2) \tag{1.31}$$

Si ricorda che il $\mathbb{V}\text{ar}_\alpha$ è l'α-quantile della distribuzione, essa è una grandezza introdotta da JP Morgan nel 1994, e poiché rispetta gli assiomi di Artzner[2] è a pieno titolo *una misura a rischio*, inoltre poiché gode della proprietà di co-monotonia è anche una *spectral measure of risk* definizione data da Acerbi (2002)[3].

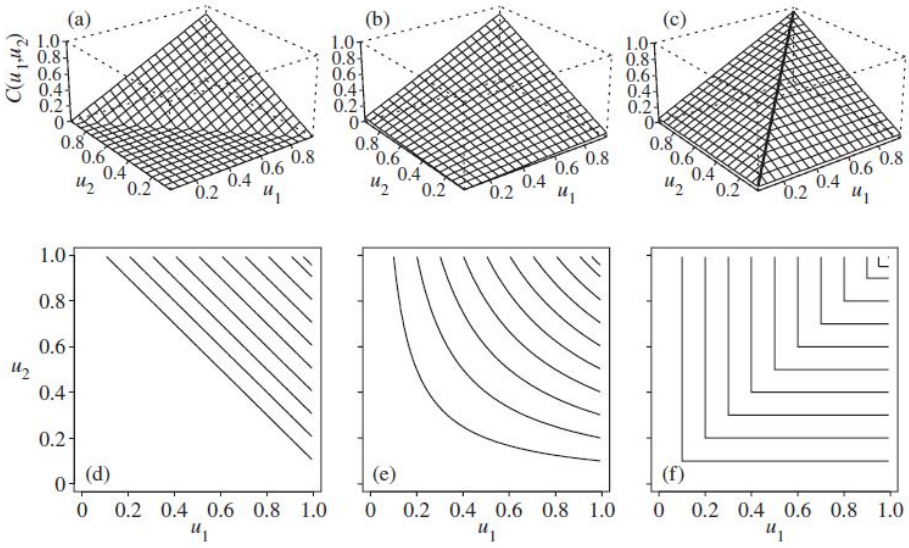


Figura 1.1. Curve di Livello delle copule fondamentali: *contro*-monotonicità (a), indipendenza (b), *co*-monotonicità (c).

1.5 Copula Prodotto

Anche il concetto di indipendenza tra variabili aleatorie trova rappresentazione in termini di copule e in questo caso la dimostrazione è una diretta conseguenza del

[2] Nel 1999 Artzner fissa gli assiomi omonimi che definiscono le proprietà di misura a rischio
[3] *Positivity, Normalization, Monotonicity.*

Teorema di Sklar (teorema 1.2) e del fatto che $X_1, \ldots, X_n$ sono indipendenti se e solo se la loro distribuzione congiunta si fattorizza nel prodotto delle marginali. In buona sostanza quello che accade è che se le marginali sono indipendenti inducono in modo "naturale" una copula che è prodotto di copule unidimensionali.

Si ricorda quindi la definizione di variabili indipendenti:

Definizione 1.12 (Indipendenza)**.** *Siano* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *un vettore aleatorio avente distribuzioni marginali* $F_1, \cdots F_n$ *e distribuzione congiunta* H*. Le variabili* $X_1, \ldots, X_n$ *sono v.a* indipendenti *se e solo se per qualunque* $\boldsymbol{x} = (x_1, \ldots, x_n) \in \operatorname{Ran} \boldsymbol{X}$*:*

$$\mathbb{P}\{X_1 \leq x_1, \ldots, X_n \leq x_n\} = \prod_{i=1}^{n} \mathbb{P}\{X_i \leq x_i\}, \tag{1.32}$$

che in termini di funzione di ripartizione equivale a:

$$H(\boldsymbol{x}) = \prod_{i=1}^{n} F_i(x_i) \tag{1.33}$$

Gli eventi sono indipendenti se il verificarsi dell'uno non influisce sulla probabilità di realizzarsi degli altri, ovvero la probabilità dell'evento congiunto corrisponde al prodotto delle probabilità dei singoli eventi. In termini di copule, se le variabili aleatorie sono indipendenti, il funzionale che descrive il legame tra le funzioni di ripartizione marginali e quella congiunta è la copula prodotto e se delle variabili aleatorie sono associate ad un copula detta appunto "copula prodotto" che di fatto è il prodotto di copule di dimensione inferiore tra loro indipendenti.

Teorema 1.8. *Siano* $X_1, \ldots, X_n$ *variabili aleatorie continue. Allora* $X_1, \ldots, X_n$ *sono indipendenti se e solo se:*

$$C_{X_1,\ldots,X_n}(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i = \Pi^n \tag{1.34}$$

La rappresentazione sopra esposta, come anticipato, è detta "copula prodotto". Si osserva che l'ipotesi di indipendenza ha il vantaggio di semplificare drasticamente le analisi poiché rende sufficiente la conoscenza delle distribuzioni marginali per fornire una descrizione esaustiva del fenomeno. Infatti, se le informazioni sulle distribuzioni marginali univariate sono facilmente ricavabili dall'analisi dei dati, inferire la struttura di dipendenza richiede, nei casi più favorevoli, una quantità di osservazioni molto più elevata e un notevole aggravio computazionale. Tuttavia assumere l'indipendenza tra i rischi è solitamente irrealistico. L'indipendenza tra variabili aleatorie, i cui vantaggi operativi sarebbero indubbi, dovrebbe essere adottata con cautela, limitatamente ai casi in cui l'evidenza sui dati la supporta. Si pensi, a titolo esemplificativo, al calcolo del premio annuo in una assicurazione danni di tipo R.C. Auto. Nel calcolo del premio puro l'ipotesi di lavoro classica è che il numero dei sinistri sia indipendente dall'ammontare dei danni e che questi siano tra loro indipendenti e identicamente distribuiti: questa caratterizzazione consente la fattorizzazione del premio equo in componenti moltiplicative. Mentre in un approccio collettivo di portafoglio le ipotesi appaiono plausibili, risultano irrealistiche nell'approccio individuale. L'utilizzo dell'ipotesi di indipendenza nel calcolo

del premio si tradurrebbe in una probabilità maggiore di quella stimata di carenza delle riserve qualora la sinistrosità e l'importo dei risarcimenti fossero positivamente correlati.

Risulta inoltre utile accennare al concetto di indipendenza condizionata.

Definizione 1.13. *Siano* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *e* $\boldsymbol{Y} = (Y_1, \ldots, Y_s)$ *due vettori aleatori,* $F_{\boldsymbol{X}|\boldsymbol{Y}}$ *la distribuzione congiunta condizionata di* $\boldsymbol{X}$ *dato* $\boldsymbol{Y}$ *e* $\{F_{x_1|\boldsymbol{Y}}\}_{1\leq i\leq n}$ *le distribuzioni marginali condizionate. I vettori* $X_1, \ldots, X_n$ *sono* indipendenti condizionatamente a $\boldsymbol{Y}$ *se per ogni* $\boldsymbol{x} \in \operatorname{Ran} \boldsymbol{X}$ *e* $\boldsymbol{y} \in \operatorname{Ran} \boldsymbol{Y}$

$$F_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{Y} = \boldsymbol{y}) = \prod_{i=1}^{n} F_{x_i|\boldsymbol{Y}}(x_i|\boldsymbol{Y} = \boldsymbol{y}) \tag{1.35}$$

Si noti che la copula associata a $F_{\boldsymbol{X}|\boldsymbol{Y}}$ è la copula prodotto e che quindi non dipende direttamente dalla variabile condizionante $\boldsymbol{Y}$.

1.6 Copula di sopravvivenza, copula duale e co-copula

Spesso, operando con variabili aleatorie è conveniente fare riferimento alla funzione di sopravvivenza piuttosto che alla funzione di ripartizione. La probabilità di sopravvivenza oltre una certa soglia x è data dalla funzione di sopravvivenza:

$$\bar{F}(x) = \mathbb{P}\{X > x\} = 1 - F(x) \tag{1.36}$$

La rappresentazione di una distribuzione di probabilità tramite la copula e le funzioni di ripartizioni marginali prescinde dalla specifica distribuzione di probabilità; poiché anche la funzione di sopravvivenza ha una distribuzione di probabilità, essa stessa può essere esplicitata univocamente attraverso le funzione di ripartizione marginali e una copula. Se viene considerato il caso più semplice possibile, ovvero il caso bivariato avente come v.a X e Y con distribuzione congiunta H, la funzione congiunta di sopravvivenza risulta essere:

$$\bar{H}(x, y) = \mathbb{P}\{X > x, Y > y\} \tag{1.37}$$

Risulta a questo punto interessante notare il legame che sussiste fra le distribuzioni marginali di sopravvivenza e quella congiunta può essere prolungato anche alle copule, ovvero esiste una copula detta appunto copula di "sopravvivenza" $\widehat{C}$ che si lega alla funzione di copula attraverso la seguente relazione:

$$\widehat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$$

$\widehat{C}$ è definita nel quadrato unitario $\mathbb{I}^2$ come anche u e v.

Esempio 1.3. *Nell'esempio 1.2 veniva introdotta la copula di Gambel bivariata come:*

$$C_\theta(u, v) = u + v - 1 + (1 - u)(1 - v)\exp\big(-\theta \ln(1 - u)\ln(1 - v)\big) \tag{1.38}$$

Allo stesso modo in cui accade per le variabili univariate, la funzione di sopravvivenza in questo caso ha una formulazione più immediata della funzione di distribuzione:

$$\widehat{C}_\theta(u, v) = uv\exp\big(-\theta \ln(u)\ln(v)\big). \tag{1.39}$$

Esempio 1.4. *In modo del tutto analogo accade con la distribuzione di Pareto bivariata Hutchinson e Lai (1990). Siano X e Y variabili casuali la cui funzione di sopravvivenza congiunta sia data da:*

$$\bar{H}_\theta(x,y) = \begin{cases} (1+x+y)^{-\theta} & \text{per } x,y \geq 0 \\ (1+x)^{-\theta} & \text{per } x \geq 0 \text{ e } y < 0 \\ (1+y)^{-\theta} & \text{per } x < 0 \text{ e } y \geq 0 \\ 1 & \text{per } x,y < 0 \end{cases} \tag{1.40}$$

Nelsen (2006) mostra che invertendo le funzioni di distribuzione delle marginali si ottiene:

$$\widehat{C}_\theta(u,v) = (u^{-\frac{1}{\theta}} + u^{-\frac{1}{\theta}} - 1)^{-\theta} \tag{1.41}$$

Schweizer e Abe Sklar (1983) mostrano che esistono altre due funzioni direttamente collegate alla funzione di copula e alla copula di sopravvivenza che sono il *duale* e la *co-copula*. Genest e Nešlehová (2007) mostrano un'altra funzione collegata alla copula di notevole importanza per il trattamento delle variabili discrete ovvero l'estensione bilineare della copula C^{Π} che verrà definita nel capitolo 6. Il duale di una copula C è denotato con $\widetilde{C}$ ed è una funzione definita:

$$\widetilde{C}(u,v) = u + v - C(u,v)\,. \tag{1.42}$$

Mentre la co-copula è denotata con C^* ed è una funzione definita nel seguente modo:

$$C^*(u,v) = 1 - C(1-u, 1-v)\,. \tag{1.43}$$

Nel caso bivariato si può esprimere la probabilità di un evento nel modo seguente:

$$\mathbb{P}\{X \leq x, Y \leq y\} = C\big(F(X), G(Y)\big) \tag{1.44}$$

$$\mathbb{P}\{X > x, Y > y\} = \widehat{C}\big(\bar{F}(x), \bar{G}(y)\big) \tag{1.45}$$

$$\mathbb{P}\{X \leq x \text{ oppure } Y \leq y\} = \tilde{C}\big(F(X), G(Y)\big) \tag{1.46}$$

$$\mathbb{P}\{X > \text{ oppure } Y > y\} = C^*\big(\bar{F}(x), \bar{G}(y)\big) \tag{1.47}$$

1.7 Proprietà di invarianza

Un problema di fondamentale interesse nello studio delle variabili aleatorie è il comportamento rispetto a trasformazioni monotone delle componenti; la tematica ha valenza teorica e pratica: sono molteplici le situazioni in cui è conveniente operare con trasformate delle variabili originali. Le copule si rivelano lo strumento per eccellenza poiché risultano invarianti rispetto a trasformazioni strettamente crescenti delle misure aleatorie. Nel caso bivariato valgono risultati più forti: è possibile esplicitare la trasformata con la copula iniziale, anche in condizioni più rilassate. Infatti risulta sufficente che le trasformazioni considerate siano strettamente monotone.

Teorema 1.9. *Siano X_1 e X_2 variabili aleatorie continue aventi come copula C_{X_1,X_2}, e siano α e β funzioni strettamente monotone sul rispettivo codominio, si ponga $\alpha(X_1) = Y_1$ e $\beta(X_2) = Y_2$:*

- *se α e β sono strettamente crescenti:*

$$C_{Y_1,Y_2}(u,v) = C_{X_1,X_2}(u,v) \tag{1.48}$$

  Ovvero C risulta inveriante rispetto a trasformazioni strettamente crescenti di X_1 e X_2.

- *se α è strettamente crescente e β strettamente decrescente:*

$$C_{Y_1,Y_2}(u,v) = u - C_{X_1,X_2}(u,1-v) \tag{1.49}$$

- *se α strettamente decrescente e β strettamente crescente:*

$$C_{Y_1,Y_2}(u,v) = v - C_{X_1,X_2}(1-u,v) \tag{1.50}$$

- *se α e β sono strettamente decrescenti:*

$$C_{Y_1,Y_2}(u,v) = u + v - 1 + C_{X_1,X_2}(1-u,1-v) = \widehat{C}_{X_1,X_2}(u,v) \tag{1.51}$$

La copula esprime in maniera esaustiva la dipendenza non parametrica tra le tra le variabili aleatorie e dunque lo studio della stessa è riconducibile alle proprietà delle copule.

Teorema 1.10. *Siano X un vettore aleatorio n-dimensionale, $(F_1, \ldots, F_n)$ le rispettive distribuzioni marginali continue aventi come copula C copula allora se $T_1, \ldots, T_n$ sono delle funzioni strettamente crescenti allora il vettore aleatorio:*

$$(T_1(X_1), \ldots, T_n(X_n)) \tag{1.52}$$

ha anch'esso copula C[4]*.*

1.8 Simmetria

Se X è una variabile aleatoria e a è un numero reale, si può affermare che X è simmetrica rispetto alla distribuzione di una variabile aleatoria se $X - a$ e $a - X$ hanno la stessa distribuzione, ovvero se per ogni $x \in \mathbb{R}$:

$$\mathbb{P}\{X - a \leq x\} = \mathbb{P}\{a - X \leq x\} \tag{1.53}$$

Quando X è continua con distribuzione F, questo equivale a:

$$F(a + x) = \bar{F}(a - x) \tag{1.54}$$

Ora se si considera il caso bivariato cosa significa per una coppia di v.a. (X, Y) dire che c'è "simmetria" rispetto a un punto (a, b)? C'è un numero di risposte a questa domanda che conducono a diverse tipi di "simmetria bivariata".

[4]Sarà diversa la risposta per dimensioni maggiori.

Definizione 1.14. *Siano X e Y v.a. e (a, b) un punto di $\mathbb{R}^2$:*

- (X, Y) *è* marginalmente simmetrica *rispetto a (a, b) se X e Y sono simmetriche rispettivamente rispetto ad a e b.*
- (X, Y) *è radialmente simmetrica rispetto a (a, b) se la distribuzione congiunta di $(X - a, Y - b)$ è la stessa della distribuzione congiunta di $(a - X, b - Y)$.*
- (X, Y) *è congiuntamente simmetrica rispetto a (a, b) se $(X - a, Y - b)$, $(X - a, b - Y)$, $(a - X, Y - b)$ e $(a - X, b - Y)$ hanno una comune distribuzione congiunta.*

Quando X e Y sono continue, la condizione di simmetria radiale può essere espressa in termini di distribuzione marginale e funzione di sopravvivenza delle v.a. X e Y in maniera del tutto analoga al caso univariato.

Teorema 1.11. *Siano X e Y v.a continue con distribuzione congiunta H e marginali F e G rispettivamente. Preso (a, b) un punto di $\mathbb{R}^2$. La distribuzione (X, Y) è radialmente simmetrica per (a, b) se e solo se:*

$$H(a + x, b + y) = H(a - x, b - y), \forall (x, y) \in \mathbb{R}^2 \tag{1.55}$$

Il termine "radiale" di fatto significa che i punti $(a + x, b + y)$ e $(a - x, b - y)$ giacciono su raggi con direzioni opposte rispetto a (a, b). Graficamente il teorema mostra che le regioni hanno sempre la stessa area[5] .

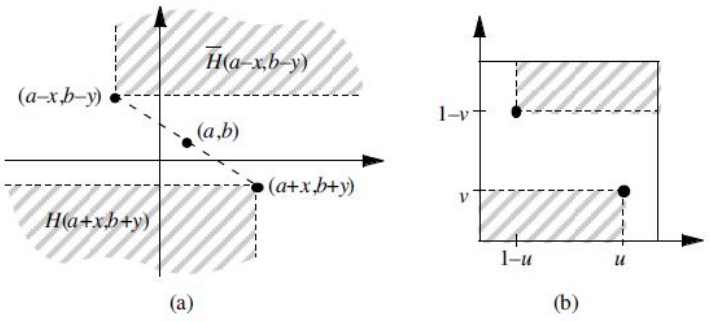


Figura 1.2. Regioni di equi-probabilità indotte da variabili aleatorie con simmetria radiale.

Esempio 1.5. *La distribuzione normale bivariata con parametri $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ e ρ è radialmente simmetrica rispetto al punto avente coordinate (μ_X, μ_Y). La dimostrazione è semplice (ma noiosa) da dimostrare a causa del calcolo degli integrali doppi.*

Teorema 1.12. *Siano X e Y v.a continue con distribuzione congiunta H, e distribuzioni marginali F e G, e aventi C come funzione di copula. Inoltre si supponga*

[5] Nelsen (2006) "always have equal H-volume".

che X e Y siano simmetriche rispettivamente per x e y. Allora (X, Y) è radialmente simmetrica rispetto ad (a, b) (ovvero H soddisfa la equazione (1.55)*) se e solo se $C = \widehat{C}$, ovvero se e solo se*

$$C(u, v) = u + v - 1 + C(1 - u, 1 - v) \tag{1.56}$$

per ogni $(u, v) \in \mathbb{I}^2$.

È utile osservare che Joe (1997) parla di *simmetria riflessiva* di una copula che soffisfa l'equazione (1.56). Si noti che il teorema 1.12 implica che se $(U, V) \sim C$ e C soddisfa l'equazione (1.56) allora $(1 - U, 1 - V) \sim C$.

Un'altra forma di simmetria consiste nella scambiabilità.

Definizione 1.15. *Due variabili aleatorie X e Y di dicono* scambiabili *se (X, Y) e (Y, X) sono identicamente distribuite.*

Quindi se H è la funzione di distribuzione congiunta di X e Y allora:

$$H(x, y) = H(y, x) \tag{1.57}$$

per ogni $(x, y) \in \mathbb{R}^2$.

Chiaramente due variabili scambiabili devono essere identicamente distribuite. Vale inoltre il seguente teorema:

Teorema 1.13. *Siano X e Y v.a continue con distribuzione congiunta H, e distribuzioni marginali F e G, e aventi C come funzione di copula. X e Y sono scambiabili se e solo se $F = G$ e $C(u, v) = C(v, u)$ per ogni $(u, v) \in \mathbb{I}^2$.*

Quando $C(u, v) = C(v, u)$ per ogni $(u, v) \in \mathbb{I}^2$, si dirà che la copula C è *simmetrica*. La proprietà di scambiabilità si vedrà è importante nel perimetro della *pair-copula construction*. Alcune copule simmetriche erano state introdotte: Π^2, M^n, W^n, come anche quelle presentate in esempio 1.1 e esempio 1.2.

1.9 Distribuzioni condizionate attraverso la funzione di copula

L'espressione che segue ci consente di esprimere la densità condizionata in termini di densità della copula e densità di una marginale.

Teorema 1.14. *La densità condizionale di due v.a X e Y può essere scritta nel modo seguente:*

$$f_{X|Y}(x|y) = c_{XY}(F_X(x), F_Y(y)) f_Y(Y) \tag{1.58}$$

In dimensione maggiore di 2 l'espressione si riscrive nel seguente modo. Siano $C_k(u_1, \ldots, u_k) = C(u_1, \ldots, U_k, 1, 1, 1)$ con $k = 2, \ldots, n-1$ le marginali k-dimensionali di $C = C_n(u_1, \ldots, u_n) = C(u_1, \ldots, u_n)$, dove C è la funzione di distribuzione congiunta delle variabili aleatorie $U_1, \ldots, U_n$. Allora la distribuzione condizionata di U_k dati i valori di $U_1, \ldots, U_{k-1}$ è data da:

$$C_k(u_k|u_1, \ldots, u_{k-1}) = \mathbb{P}\{U_k \leq u_k | U_1 = u_1, \ldots, U_{k-1} = u_{k-1}\} \tag{1.59}$$

$$= \frac{\partial^{k-1} C_k(u_1, \ldots, u_k)}{\partial u_1 \ldots u_{k-1}} \left\{ \frac{\partial^{k-1} C_{k-1}(u_1, \ldots, u_{k-1})}{\partial u_1 \ldots u_{k-1}} \right\}^{-1} \tag{1.60}$$

Ovviamente tutto funziona se la funzione di copula è assolutamente continua, questo consente la k-differenziabilità.

Capitolo 2

Misure di Dipendenza

> For non-normal random variables, Pearson's correlation and concepts based on linearity are not necessarily the best concepts to work with.
>
> Joe (1997)

L'atomo del carbonio possiede moltissime forme allotropiche[1]. La maggior parte di quest'ultime sono state prodotte nei laboratori, ma ne possiede diverse forme anche naturalmente, come la grafite, il diamante, il carbonio amorfo e i i fullereni. La grafite è uno dei materiali più morbidi in natura mentre il diamante è il più duro conosciuto. Il diamante e la grafite si differenziano principalmente nel tipo di legame covalente e nella disposizione di atomi all'interno della struttura cristallina. Nel diamante, il carbonio è disposto lungo i quattro vertici di un tetraedro e tutti gli elettroni sono impiegati nel formare dei legami, mentre nella grafite gli atomi di carbonio sono disposti secondo maglie esagonali e non tutti gli elettroni formano lo stesso tipo di legame, ma vi è una coppia di elettroni in grado di formare un debole legame in direzione perpendicolare agli strati reticolari.

Questo esempio relativo alle forme allotropiche è calzante anche nel contesto probabilistico, in quanto distribuzioni marginali unite con una struttura di dipendenza differente generano una distribuzione multivariata che può essere anche molto diversa.

Per molti anni, la correlazione nell'ambito dell'analisi della dipendenza tra variabili aleatorie è stata uno degli strumenti più diffusi nella finanza e nel mondo assicurativo; tuttavia essa è stata, e in parte ancora lo è, uno dei concetti più "fraintesi". Infatti, la storia ci dimostra che endemicamente i fenomeni multivariati nel mondo reale raramente possono essere rappresentati attraverso una correlazione lineare. A riguardo Embrechts, A. J. McNeil e Straumann (2002) scrivevano:

> *Although insurance has traditionally been built on the assumption of independence and the law of large numbers has governed the determination of premiums, the increasing complexity of insurance and reinsurance*

[1] Ovvero elementi chimici che esistono in diverse forme.

> *products has led recently to increased actuarial interest in the modelling of dependent risks (Wang (1998)); an example is the emergence of more intricate multi-line products. The current quest for a sound methodological basis for integrated risk management also raises the issue of correlation and dependence. Although contemporary financial risk management revolves around the use of correlation to describe dependence between risks, the inclusion of non-linear derivative products invalidates many of the distributional assumptions underlying the use of correlation. In insurance these assumptions are even more problematic because of the typical skewness and heavy-tailedness of insurance claims data.*

Il passo sopracitato, "vecchio" di circa 20 anni è ancora molto moderno, metteva in luce tutti i limiti sui modelli "semplici" che venivano utilizzati per il *pricing* dei prodotti "multilinea" che oggi vengono definiti "multiramo" o in generale *IBIPs* per il comparto vita e coperture "multirischio" per il comparto danni. Unitamente Embrechts, A. J. McNeil e Straumann (2002) facevano emergere la centralità delle code e in generale degli eventi estremi.

Evidentemente il ruolo di marginalità assunto da strutture di dipendenza più sofisticate è stato riconsiderato, dando ragione nuovamente alla storia, che ha mostrato, e continua a mostrare, che i fenomeni platicurtici si verificano e che tutto può succedere anche sulle code leggere di una distribuzione.

In questo capitolo, si ripercorreranno gli aspetti più essenziali per quanto riguarda le principali misure di dipendenza utilizzate nel mondo delle scienze attuariali, si osserveranno quindi le misure di dipendenza fondamentali: il coefficiente di correlazione lineare, la correlazione di rango nelle sue diverse forme e il teorema d'indipendenza asintotica usato per le dipendenze di coda.

In particolare, si osserverà che la correlazione è uno strumento utile ed estremamente "semplice" e si dimostra essere "appropriato" principalmente nel perimetro dei modelli gaussiani multivariati e nel contesto delle distribuzioni ellittiche.

Gli altri due tipi di misure di dipendenza dalle quali le copule prendono a piene mani sono il τ_K-di Kendall e ρ_S di Spearman. Si vedrà che queste ultime, diversamente dalla correlazione lineare, dipendono in modo diretto dalla forma della distribuzione bivariata e non dal comportamento delle marginali, ed è proprio questa caratteristica che rende tali misure particolarmente adatte alla parametrizzazione delle copule.

Si inizierà presentando lo sviluppo della teoria della dipendenza tra variabili aleatorie, riportando anche il contributo della Scuola di Statistica Italiana e come questi contributi si siano incastrati nell'alveo della teoria delle copule.

Da ultimo, si segnala che anche questo capitolo non ha la presunzione di una trattazione "enciclopedica" ma si ripercorreranno gli aspetti legati alle misure di dipendenza funzionali a questo elaborato.

2.1 Brevi Cenni Storici

In letteratura, l'introduzione alle misure di dipendenza avviene attraverso l'enumerazione di tre indici che prendono il nome dei loro padri fondatori:

- il ρ di Pearson da Karl Pearson che lo introdusse in Pearson (1895)[2];
- il ρ_S di Spearman da Charles Spearman che lo introdusse in Spearman (1904) e Spearman (1906);
- il τ_K di Kendall da Maurice G. Kendall che lo introdusse in Kendall (1938).

In modo un pò superficiale, si potrebbe giungere a conclusioni affrettate e pensare che l'analisi dei rischi congiunti sia stato oggetto d'interesse solo per la scuola anglosassone, che fa capo ai nomi di K. Pearson, R.A. Fisher, J. Neyman, E.S. Pearson e A. Wald. Tuttavia, iniziando a studiare un pò i carteggi, si nota come la ricerca in questo campo si annoda con le ricerche di statistica descrittiva sviluppate in Italia già dagli ultimi decenni del diciannovesimo secolo aventi come precursori studiosi come Corrado Gini e Gaetano Pietra. Tra le principali ricerche nel campo della correlazione portate avanti in Italia si ricorda:

- i lavori di Gili e Bettuzzi che trattano della correlazione nell'ambito degli studi sulla concordanza;
- i lavori di Rizzi e Vicari che forniscono l'espressione della matrice di correlazione parziale;
- i lavori di Crescimanni che ha elaborato alcune applicazioni non-*standard* sulla correlazione con esiti di particolare interesse[3] e sulla "correlazione assoluta";
- il lavoro di Buscemi che ha proposto un particolare indice di autocorrelazione spaziale, applicabile a uno spazio k-dimensionale definito in funzione di un indice di oscillazione k-dimensionale[4].

Da ultimo si ricorda che Nelsen (2006) utilizza le proprietà sulle misure di concordanza introdotte da Scarsini (1984). Proprio quel *set* di proprietà desiderabili introdotte da Scarsini (1984) vengono poi riprese in Rényi (1959). Quest'ultimo, attraverso gli omonimi assiomi[5], definisce per la prima volta il grado di dipenden-

[2] L'indice di Pearson è anche detto coefficiente di correlazione lineare di Bravais-Pearson perché frutto del lavoro di tre diversi statistici: August Bravais pubblicò nel 1846 un paper in cui viene presentata la formulazione matematica di "correlazione statistica"; nel 1885, Sir Francis Galton definì un indice che quantifica la forza della relazione tra le stature dei genitori e dei figli; e, nel 1895, infine, Karl Pearson ha ripreso il lavoro di Galton e Bravais ed ha sviluppato il coefficiente così come è conosciuto attualmente.

[3] Seguendo un suggerimento di Benedetti

[4] A quest'ultimo argomento avevano già dato contributi, in tempi recenti, la stessa Crescimanni 1986, oltre a Lionetti 1980 e Vinci 1987.

[5] *Assiomi di Rényi:* si assuma che $R(X, Y)$ misuri il grado di dipendenza di X e Y e soddisfi le seguenti condizioni:

- $R(X, Y)$ è definito per tutte le variabili casuali continue non costanti;
- $R(X, Y)$ può non essere uguale a $R(Y, X)$;
- $0 \leq R(X, Y) \leq 1$;
- $R(X, Y) = 0$ se e solo se X e Y sono indipendenti;
- $R(X, Y) = 1$ se e solo se $Y = f(X)$ quasi sicuramente per una funzione Borel-misurabile;
- se g è una biezione su un insieme misurabile secondo Borel su $\mathbb{R}$ allora $R(g(X, Y)) = R(X, Y)$.

za come se fosse una struttura algebrica. Infine, Schweizer e Wolff (1981) furono i primi che hanno messo esplicitamente in relazione le copule con lo studio della dipendenza tra variabili casuali. In Schweizer e Wolff (1981) vengono discussi e modificati gli assiomi di Rényi presentando le proprietà di invarianza delle copule sotto trasformazioni strettamente monotone per variabili aleatorie introducendo la misura di dipendenza oggi nota come σ di Schweizer e Wolff. Nelle loro parole:

> *[...] under a.s. strictly increasing transformations of X and Y, the copula is invariant while the margins may be changed at will, it follows that it is precisely the copula which captures those properties of the joint distribution which are invariant under a.s. strictly increasing transformations. Hence the study of rank statistics—insofar as it is the study of properties invariant under such transformations—may be characterized as the study of copulas and copula-invariant properties.*

2.2 ρ di Pearson

Il coefficente di correlazione ρ di Pearson rappresenta una misura di dipendenza "lineare" tra variabili aleatorie.

Definizione 2.1 (Coefficente di Correlazione ρ di Pearson)**.** *Siano $X_1, \dots, X_n$ variabili casuali non degeneri con varianza finita, allora il* coefficiente di correlazione lineare ρ *è definito come segue:*

$$\rho(X_1, \dots, X_n) = \frac{\mathbb{C}\text{ov}(X_1, \dots, X_n)}{\sqrt{\mathbb{V}\text{ar}(X_1) \cdots \mathbb{V}\text{ar}(X_n)}} \tag{2.1}$$

dove $\mathbb{C}\text{ov}(X_1, \dots, X_n) = \mathbb{E}[X_1 \cdots X_n] - \mathbb{E}[X_1] \cdots \mathbb{E}[X_n]$ è la covarianza delle variabili casuali. Il coefficente ρ assume valori in $[-1, 1]$[6]*.*

Se $X_1, \dots, X_n$ sono variabili indipendenti allora vale sempre la seguente espressione:

$$\mathbb{E}[X_1 \cdots X_n] = \prod_{i=1}^{n} \mathbb{E}[X_i] \tag{2.2}$$

e quindi $\rho(X_1, \dots, X_n) = \mathbb{C}\text{ov}(X_1, \dots, X_n) = 0$.

Inoltre, si dimostra che il coefficiente di correlazione ρ è invariante rispetto a trasformazioni monotone delle marginali ovvero:

$$\rho(\alpha_1 + \beta_1 X_1, \alpha_2 + \beta_2 X_2) = \alpha_1 + \beta_1 \rho(X_1) + \alpha_2 + \beta_2 \rho(X_2) \tag{2.3}$$

$$\rho(\alpha_1 + \beta_1 X_1, \dots, \alpha_n + \beta_n X_n) = \sum_{i=1}^{n} \alpha_i + \beta_i \sum_{i=1}^{n} \rho(X_i) \tag{2.4}$$

Tuttavia, ρ non verifica la proprietà di invarianza per trasformazioni non lineari. Un classico esempio per vedere cosa accade con funzioni non lineari è usare una

[6] Se l'indicatore assume il valore -1 si parla di perfetta correlazione negativa, se assume il valore $+1$ si parla di perfetta correlazione positiva.

distribuzione Normale standard $X \sim \mathcal{N}(0,1)$ e il suo quadrato ovvero $Y = X^2$:

$$\begin{aligned} \mathbb{C}\mathrm{ov}(X,Y) &= \mathbb{C}\mathrm{ov}(X,X^2) && (2.5)\\ &= \mathbb{E}[X \cdot X^2] - \mathbb{E}[X]\,\mathbb{E}[X^2] && (2.6)\\ &= \mathbb{E}[X^3] - \mathbb{E}[X]\,\mathbb{E}[X^2] && (2.7)\\ &= 0 && (2.8) \end{aligned}$$

poiché i momenti di una normale standard di ordine dispari sono tutti nulli:

$$\mathbb{E}[X^{2k+1}] = 0 \tag{2.9}$$

Esempio 2.1. *Marginali date e correlazione possono determinare distribuzioni congiunte differenti, infatti Embrechts, A. J. McNeil e Straumann (2002) mostrano che anche nel caso di distribuzioni ellittiche la distribuzione congiunta è non univoca. Se si considerano X e Y distribuzioni marginali normali aventi $\rho_K = 0$ ci sono molte possibili bivariate per X e Y. La distribuzione congiunta F_{XY} non è univocamente determinata: ne esistono molte, come ad esempio la copula di Farlie-Gumbel-Morgenstern come dimostrato Balakrishnan e Lai (2009) e Joe (1997).*

$$C(u,v) = uv + \theta uv(1-u)(1-v) \tag{2.10}$$

Si veda anche la figura 2.1.

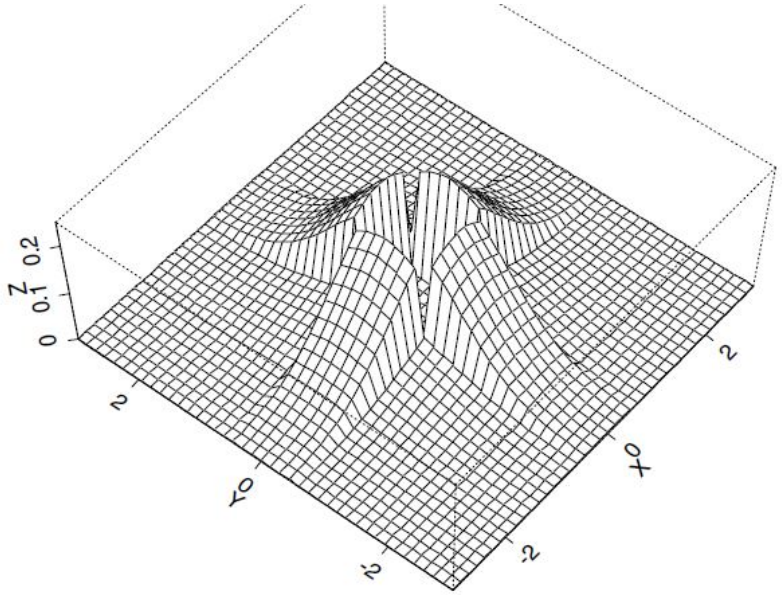


Figura 2.1. Densità di una distribuzione bivariata "non-normale" con marginali normali.

2.3 La Concordanza

In relazione alle variabili aleatorie bidimensionali e alle distribuzioni marginali bivariate di variabili aleatorie in dimensione generica è utile il concetto di concordanza.

Informalmente, una coppia di variabili casuali è concorde se valori "grandi" di una tendono ad essere associati a valori "grandi" dell'altra e a valori "piccoli" di una con valori "piccoli" dell'altra.

Definizione 2.2. *Siano (x_i, y_i) e (x_j, y_j) due osservazioni da un vettore (X, Y) di variabili casuali continue, si dirà che (x_i, y_i) e (x_j, y_j) sono concordi se:*

$$x_i < x_j \text{ e } y_i < y_j \text{ oppure se } x_i > x_j \text{ e } y_i > y_j \tag{2.11}$$

analogamente, si dice che (x_i, y_i) e (x_j, y_j) sono discordi se:

$$x_i < x_j \text{ e } y_i > y_j \text{ oppure se } x_i > x_j \text{ e } y_i < y_j \tag{2.12}$$

È possibile usare una formulazione alternativa, ovvero:

$$(x_i, y_i) \text{ e } (x_j, y_j) \text{ sono concordi se } (x_i - x_j)(y_i - y_j) > 0, \tag{2.13}$$

$$(x_i, y_i) \text{ e } (x_j, y_j) \text{ e discordi se } (x_i - x_j)(y_i - y_j) < 0. \tag{2.14}$$

Una misura di concordanza, affinché catturi adeguatamente il comportamento evidenziato, dovrà soddisfare una serie di criteri. Un possibile insieme di proprietà è quella introdotta in Scarsini (1984):

Definizione 2.3 (Assiomi di Scarsini)**.** *Una misura di associazione κ tra due variabili aleatorie X_1 e X_2 è una misura di concordanza se soddisfa le seguenti proprietà:*

- *κ è definita per ogni coppia di variabili aleatorie continue X_1 e X_2;*
- *κ è tale che $\kappa \in [-1, 1], \kappa_{X,X} = 1$ e $\kappa_{X,-X} = -1$;*
- *κ è simmetrica;*
- *Se X_1 e X_2 sono indipendenti, allora $\kappa_{X_1,X_2} = 0$*
- $\kappa_{-X_1,X_2} = \kappa_{X_1,-X_2} = -\kappa_{X_1,X_2}$

Una misura che soddisfa tali proprietà, soddisfa anche il seguente teorema:

Teorema 2.1. *Sia κ una misura di concordanza tra due variabili aleatorie continue X_1, X_2:*

- *Se X_2 è quasi certamente una funzione crescente di X_1, allora $\kappa_{X_1,X_2} = 1$*
- *Se X_2 è quasi certamente una funzione decrescente di X_1, allora $\kappa_{X_1,X_2} = -1$*
- *Se α e β sono funzioni quasi certamente strettamente monotone, rispettivamente sul loro codominio, allora $\kappa_{\alpha X_1,\beta X_2} = \kappa_{X_1,X_2}$*

Poiché la copula costituisce una descrizione esaustiva della struttura di dipendenza di una distribuzione multivariata e le misure di concordanza ne rappresentano solo una parte, è intuitivo supporre che queste venga espressa in qualche modo attraverso la copula.

Tra le misure di concordanza proposte in letteratura che soddisfano le proprietà di Scarsini rivestono una particolare rilevanza nello studio del comportamento delle marginali bivariate e delle distribuzioni multivariate: il τ_K di Kendall, la ρ_s di Spearman e i coefficenti di coda[7]. Inoltre si vedrà come queste misure siano collegate alla funzione di copula.

2.4 τ_K di Kendall

Definizione 2.4. *Siano (X_1, Y_1) e (X_2, Y_2) vettori casuali indipendenti e identicamente distribuiti, ciascuno con la funzione di distribuzione congiunta H, τ_K di Kendall*[8] *è definito come la probabilità di concordanza meno la probabilità di discordanza:*

$$\tau_K = \tau(X, Y) = \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\} \quad (2.15)$$

L'indice esprime il "disallineamento" delle distribuzioni X e Y: se gli $\{x_i\}$ sono ordinati in maniera crescente, la misura nella quale gli $\{y_i\}$ corrispondenti si discostano da tale ordinamento è sintomo della qualità del legame tra X e Y. Si noti che in caso di perfetta concordanza e perfetta discordanza, il τ_K assume valori rispettivamente 1 e -1.

La versione campionaria del τ_K è la seguente:

Definizione 2.5. *Sia $\{(x_1, y_1), \ldots, (x_d, y_d)\}$ un campione casuale di d realizzazioni da una variabile aleatoria (X, Y). Nel campione si avrà un numero di coppie $\{(x_i, y_i), (x_j, y_j)\}$ pari a $\binom{n}{2}$ e ognuna potrà essere discorde o concorde. Posti il numero delle coppie concordi e quello delle coppie discordi rispettivamente pari a c e d, il $\widehat{\tau}_K$ campionario corrisponde a:*

$$\widehat{\tau}_K = \frac{c - d}{c + d} = \frac{c - d}{\binom{n}{2}} \quad (2.16)$$

La misura τ_K è caratterizzabile con le copule. Per farlo abbiamo bisogno del seguente teorema:

Teorema 2.2. *Siano (X_1, Y_1) e (X_2, Y_2) vettori aleatori indipendenti di variabili aleatorie continue con funzione di ripartizione congiunta rispettivamente H_1 e H_2, aventi le stesse funzioni di ripartizione marginali F_X (per X_1 e X_2) e F_Y (per Y_1 e Y_2). Siano inoltre C_1 e C_2 le copule rispettivamente di (X_1, Y_1) e (X_2, Y_2). Ovvero si ha che:*

$$H_1(x, y) = C_1(F_X(x), F_Y(y)) \quad (2.17)$$

$$H_2(x, y) = C_2(F_X(x), F_Y(y)). \quad (2.18)$$

Sia Q la differenza tra la probabilità di concordanza e quella di discordanza di (X_1, Y_1) e (X_2, Y_2) ovvero sia:

$$Q \coloneqq \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_2) < 0\}. \quad (2.19)$$

[7] In questa trattazione non verranno presentate altre importanti misure di concordanza come la g di Gini, la β di Blomqvist e la γ di Goodman e Kruskal in quanto non invervengono utilizzate nel contesto della *pair-copula construction*.

[8] Detto anche rango di Kendall

Allora si ha che

$$Q = Q(C_1, C_2) = 4 \iint_{\mathbb{I}^2} C_2(u, v)\, \mathrm{d}C_1(u, v) - 1\,. \tag{2.20}$$

La funzione Q definita nel teorema 2.2 ha alcune proprietà interessanti:

Corollario 2.1. *Siano C_1, C_2 e Q come nel teorema 2.2, allora:*

1. *Q è simmetrica, ovvero $Q(C_1, C_2) = Q(C_2, C_1)$;*
2. *le copule C_1, C_2 possono essere sostituite dalle loro copule di sopravvivenza senza modificare il risultato, ovvero $Q(C_1, C_2) = Q(\widehat{C}_1, \widehat{C}_2)$.*

Utilizzando il teorema 2.2 è possibile esprimere il τ_K di Kendal per un vettore aleatorio in modo alternativo, ovvero usando la teoria delle copule.

Teorema 2.3. *Siano X e Y due variabili aleatorie continue di distribuzione congiunta F e la cui copula associata sia C, allora il τ_K di Kendall di X e Y coincide con:*

$$\tau_K(X, Y) = \tau_C = 4 \iint_{\mathbb{I}^2} C(u, v)\, \mathrm{d}C(u, v) - 1 = Q(C, C). \tag{2.21}$$

L'integrale in equazione (2.21) può essere interpretato come il valore atteso della funzione $C(U, V)$ con U e V uniformi *standard* e distribuzione congiunta $C(U, V)$, allora la formula si può riscrivere nel seguente modo:

$$\tau_K(X, Y) = 4\, \mathbb{E}[C(U, V)] - 1 \tag{2.22}$$

Nel caso in cui il prodotto delle derivate parziali $\frac{\partial C}{\partial u}$ e $\frac{\partial C}{\partial v}$ della copula sia integrabile su $\mathbb{I}^2$, allora si può ricavare la seguente caratterizzazione.

Teorema 2.4. *Sia C una copula tale che il prodotto $\frac{\partial C}{\partial u}\frac{\partial C}{\partial v}$ sia integrabile in $\mathbb{I}^2$, allora:*

$$\iint_{\mathbb{I}^2} C(u, v)\, \mathrm{d}C(u, v) = \frac{1}{2} - \iint_{\mathbb{I}^2} \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v)\, \mathrm{d}u\, \mathrm{d}v \tag{2.23}$$

Sotto le ipotesi, si ha che:

$$\tau_C = 1 - 4 \iint_{\mathbb{I}^2} \frac{\partial}{\partial u} C(u, v) \frac{\partial}{\partial v} C(u, v)\, \mathrm{d}u\, \mathrm{d}v\,. \tag{2.24}$$

2.5 ρ_S di Spearman

Definizione 2.6. *Siano (X_1, Y_1), (X_2, Y_2) e (X_3, Y_3) vettori casuali indipendenti e identicamente distribuiti provenienti dalla variabile (X, Y), allora ρ_S di Spearman è definito come:*

$$\rho_S(X, Y) = 3\Big(\mathbb{P}\{(X_1 - X_2)(Y_1 - Y_3) > 0\} - \mathbb{P}\{(X_1 - X_2)(Y_1 - Y_3) < 0\}\Big) \tag{2.25}$$

Dalla formula precedente risulta che la grandezza ρ_S individua, a meno di una costante moltiplicativa, la differenza tra la probabilità delle concordanze e quella delle discordanze tra le coppie (X_1, Y_1) e (X_2, Y_3).

Sia $\{(x_i, y_i)\}_{1\leq i\leq n}$ un campione casuale del vettore casuale (X, Y). Si ricorda che il *rango* di x_i è la sua posizione nel vettore ordinato delle $\{x_i\}_{1\leq i\leq n}$. Sia quindi R_i il rango di x_i e S_i il rango di y_i, allora il $\widehat{\rho}_S$ di Spearman empirico è uguale a:

$$\widehat{\rho}_S = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}} \tag{2.26}$$

Dove $\bar{R}$ e $\bar{S}$ sono i ranghi attesi per le variabili rango R e S. D'altra parte, è evidente che $\bar{R} = \bar{S}$ e che:

$$\bar{S} = \bar{R} = \frac{1}{n}\sum_{i=1}^n R_i = \frac{n(n+1)}{2n} = \frac{n+1}{2}. \tag{2.27}$$

Usando questo risultato si ricava:

$$\widehat{\rho}_S = \frac{1 - 6\sum_{i=1}^n (R_i - S_i)^2}{n^3 - n}. \tag{2.28}$$

Anche per il ρ_S di Spearman esiste un analogo del teorema 2.3. Ovvero vale il seguente:

Teorema 2.5. *Siano X e Y due variabili aleatorie continue di distribuzione congiunta F e la cui copula associata sia C, allora il ρ_S di Spearman di X e Y coincide con:*

$$\rho_S = \rho_C = 3Q(C, \Pi^2) \tag{2.29}$$

$$= 12 \iint_{\mathbb{I}^2} uv \, \mathrm{d}C(u, v) - 3 \tag{2.30}$$

$$= 12 \iint_{\mathbb{I}^2} C(u, v) \, \mathrm{d}u \, \mathrm{d}v - 3\,. \tag{2.31}$$

La constante 3 in equazione (2.31) è una costante di normalizzazione in quanto $Q(C, \Pi^2)$ ha valori in $\left[-\frac{1}{3}, \frac{1}{3}\right]$.

Siccome accade che:

$$3 = \frac{12}{4} = 12 \iint_{\mathbb{I}^2} uv \, \mathrm{d}u \, \mathrm{d}v, \tag{2.32}$$

allora è possibile riscrivere la equazione (2.31) nel seguente modo:

$$\rho_S = \rho_C = 12 \iint_{\mathbb{I}^2} (C(u, v) - uv) \, \mathrm{d}u \, \mathrm{d}v. \tag{2.33}$$

In altre parole, ρ_C è proporzionale al volume con segno compreso tra la copula C e la copula Π^2. Perciò, ρ_C è una misura della distanza media tra la distribuzione di X e Y e la distribuzione indipendente (rappresentata dalla copula prodotto).

Il ρ_S è noto anche come *grado* del coefficiente di correlazione. I gradi sono l'analogo dei ranghi in una popolazione, cioè date due realizzazioni x e y di due variabili aleatorie X e Y, con funzione di ripartizione rispettivamente F_X e F_Y,

allora i gradi di x e y sono $u = F_X(x)$ e $v = F_Y(y)$. Sia C la copula associata a (X, Y) e siano $U = F_X \circ X$ e $V = F_Y \circ Y$ le distribuzioni associate alle variabili aleatorie di grado. Dato che $U \sim V \sim U(0,1)$, allora $\mathbb{E}(U) = \mathbb{E}(V) = \frac{1}{2}$ e $\mathbb{V}\mathrm{ar}(U) = \mathbb{V}\mathrm{ar}(V) = \frac{1}{12}$. Risulta quindi che:

$$\rho_S = \rho_C = 12 \iint_{\mathbb{I}^2} C(u,v)\,\mathrm{d}u\,\mathrm{d}v - 3 \tag{2.34}$$

$$= 12\,\mathbb{E}(UV) - 3 \tag{2.35}$$

$$= \frac{\mathbb{E}(UV) - \frac{1}{4}}{\frac{1}{12}} \tag{2.36}$$

$$= \frac{\mathbb{E}(UV) - \mathbb{E}(U)\,\mathbb{E}(V)}{\sqrt{\mathbb{V}\mathrm{ar}(U)\,\mathbb{V}\mathrm{ar}(V)}} \tag{2.37}$$

Dunque, il ρ_S di Spearman per una coppia di variabili aleatorie X e Y è equivalente ad una forma funzionale del coefficiente di correlazione lineare di Pearson tra i gradi di X e Y (ovvero le variabili aleatorie U e V).

L'interesse per il τ_K e il ρ_S risiede anche nel fatto che esse siano delle misure di concordanza con le proprietà richieste dalla definizione 2.3.

Teorema 2.6. *Sia τ_K che ρ_S sono misure di concordanza, ovvero soddisfano le proprietà della definizione 2.3 e del teorema 2.1.*

Vi sono inoltre vincoli sui valori che le due misure di concordanza possono assumere congiuntamente. Ovvero vale il seguente teorema.

Teorema 2.7. *Siano X, Y due variabili aleatorie continue e siano $\tau = \tau_K(X,Y)$ e $\rho = \rho_S(X,Y)$, allora valgono le seguenti disequazioni:*

$$-1 \leq 3\tau - 2\rho \leq 1 \tag{2.38}$$

$$\left(\frac{1+\tau_K}{2}\right)^2 \leq \frac{1+\rho_S}{2} \tag{2.39}$$

$$\left(\frac{1-\tau_K}{2}\right)^2 \leq \frac{1-\rho_S}{2} \tag{2.40}$$

$$\tag{2.41}$$

Corollario 2.2. *Siano X, Y due variabili aleatorie continue e siano $\tau_K = \tau_K(X,Y)$ e $\rho_D = \rho_S(X,Y)$, allora:*

- *se $\tau_K \geq 0$ allora:*
$$\frac{3\tau_K - 1}{2} \leq \rho_S \leq \frac{1 + 2\tau_K - \tau_K^2}{2} \tag{2.42}$$
- *se $\tau_K \leq 0$ allora:*
$$\frac{\tau_K^2 + 2\tau - 1}{2} \leq \rho_S \leq \frac{1 + 3\tau_K}{2} \tag{2.43}$$

Carriére (1994) mostra che la relazione tra ρ_S e τ_K nelle famiglie di copule in cui il funzionale dipende da un solo parametro, τ_K può essere sfruttato per costruire un test di ipotesi per verificare se un dato campione può provenire da una particolare famiglia.

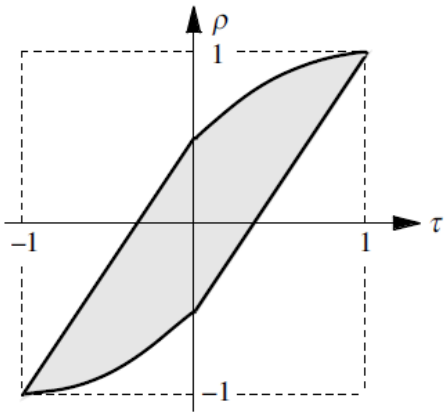


Figura 2.2. Regione del piano dei valori assunti da τ_K e ρ_S.

2.6 Dipendenza di Coda

Nel corso degli anni sono state presentate diverse altre misure di dipendenza, come ad esempio la dipendenza sui quadranti di Lehmann (1966) (e la relativa generalizzazione sugli ortanti *Orthant Dependence*) oppure le misure di dipendenza basate sul coefficente di Gini. Un'importante misura di dipendenza che ha fortissimo interesse pratico nella teoria delle copule è la dipendenza di coda (*tail-dependence*). La valutazione delle dinamiche sulle code è fondamentale nel *risk-management* sia in campo bancario che assicurativo e trova naturale collocazione nel perimetro della teoria dei valori estemi.

Essa è stata introdotta da Sibuya (1960) e Ledford e Tawn (1996) e rappresenta la probabilità che la variabile Y ecceda per un dato quantile $q_Y(\alpha)$ condizionatamente al fatto che i valori della variabile X eccedano il quantile $q_X(\alpha)$ per lo stesso α. I coefficienti di dipendenza di coda, proprio come quelli di correlazione di rango, sono delle metriche che dipendono esclusivamente dalla copula e per tanto ne ereditano la proprietà di invarianza rispetto a trasformazioni strettamente crescenti.

Teorema 2.8 (Teorema d'indipendenza asintotica)**.** *Siano X_1 e X_2 variabili aleatorie con funzione di ripartizione F_1 e F_2, si definisce* coefficente di coda superiore *λ_U di una distribuzione bivariata avente copula C:*

$$\lambda_U = \lim_{\alpha \to 1^-} \mathbb{P}\{X_2 > F_2^{-1}(\alpha) | X_1 > F_1^{-1}(\alpha)\} = \lim_{\alpha \to 1^-} \frac{1 - 2\alpha + C(\alpha, \alpha)}{1 - \alpha}, \qquad (2.44)$$

mentre si definisce il coefficente di coda inferiore *λ_U come:*

$$\lambda_L = \lim_{\alpha \to 0^+} \mathbb{P}\{X_2 < F_2^{-1}(\alpha) | X_1 < F_1^{-1}(\alpha)\} = \lim_{\alpha \to 0^+} \frac{C(\alpha, \alpha)}{\alpha} \qquad (2.45)$$

Nel caso in cui:

- *I limiti esistano e i valori di λ_U e di λ_L siano compresi tra $(0, 1]$ ci si trova nel caso di* dipendenza asintotica *(rispettivamente* sulla coda superiore *e/o* inferiore*);*

- *I limiti esistano e i valori di* λ_U *e di* λ_L *assumano valori pari a zero, ci si trova nel caso di* indipendenza asintotica sulla coda superiore e inferiore.

Esempio 2.2. *Se si campionano due distribuzioni aventi distribuzioni marginali uguali, lo stesso valore del coefficiente di correlazione di rango* ρ_K *ma diverse copule: una con struttura gaussiana e una con copula di Gumbel si osserverà che nel primo caso si è in presenza di indipendenza asintotica* $\lambda_U = 0$ *mentre nel secondo caso* $\lambda_U = 0.768$ *(si veda figura 2.3).*

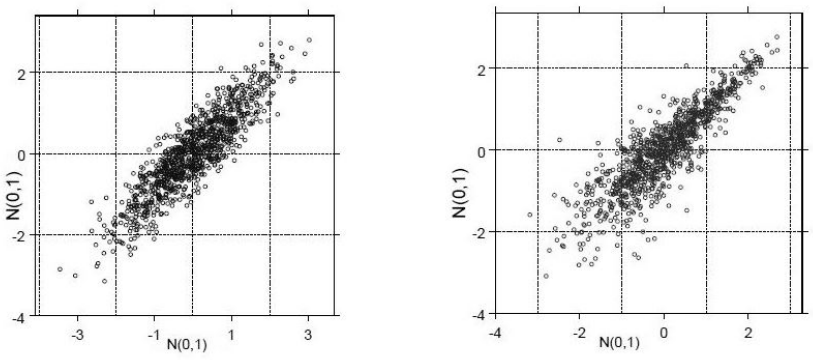


Figura 2.3. Dipendenza di coda: a destra copula gaussiana a sinistra copula di Gumbel.

Esiste inoltre un'altra formulazione dell'indipendenza asintotica nell'ipotesi che il limite esista:

$$\lambda_U = 2 \lim_{\alpha \to 1^-} \mathbb{P}\{X_2 > VaR_\alpha(X_2) | X_1 > VaR_\alpha(X_1)\} \tag{2.46}$$

Analoghe definizioni per il coefficente di dipendenza sulla coda inferiore.

2.7 Limiti delle tradizionali strutture di dipendenza

Riassumento quanto osservato in questo capitolo, le strutture di dipendenza semplici, seppur siano degli strumenti flessibili e di facile applicazione, sono ormai in disuso. In particolare la correlazione lineare non rappresenta più la misura canonica utilizzata per studiare la dipendenza tra fenomeni perchè essa è affetta dalle seguenti "patologie":

- $\rho = 0$ definisce variabili aleatorie indipendenti solo in casi particolari;
- le varianze delle variabili aleatorie devono essere finite in quanto sorgono problematiche in presenza di fenomeni leptocurtici[9];
- la correlazione non è invariante sotto trasformazioni non lineari strettamente crescenti.

Inoltre, nella fase di costruzione emergono altri due aspetti di rilievo:

[9]Si rimanda all'appendice per questa nozione.

- non è biunivoca la corrispondenza tra *congiunta-marginali* assegnata una matrice di correlazione Σ;
- non sempre esiste la congiunta date le marginali e una assegnata la matrice di correlazione.

All'esigenza di misure di dipendenza che riuscissero a cogliere delle relazioni più rappresentative si trova risposta con le copule e tale risposta si rivelerà esaustiva per circa quarant'anni.

Capitolo 3

Famiglie di Copule

> Whereas the bivariate case was central in most of the publications and seems to be well explored at present, there is still an ongoing and active debate on the construction of multivariate copula models.
>
> Fischer (2010)

3.1 The "*family*" concept

Ormai da diversi anni il dibattito in tema di copule si concentra sui metodi di costruzione di modelli multivariati e in particolare modelli multivariati di grandi dimensioni. Kurowicka e Joe (2010) affermano in tal senso:

> *With regard to dependence in higher dimensions, much is not known. For example, we do not know whether an arbitrary correlation matrix is also a rank correlation matrix and the negative dependence for dimensions greater than four.*

Esistono diversi approcci per la costruzione di distribuzioni multivariate: metodi che si basano su misture e rappresentazioni stocastiche, oppure metodi in cui le copule sono ottenute come limite dei valori estremi. Inoltre, non si potrebbe parlare di copule e non parlare delle famiglie di Laplace[1], le quali giocano un ruolo fondamentale nella costruzione di alcune famiglie di copule. Infine per parlare di *pair-copula construction* non si può non parlare dei metodi grafici per la costruzione di copule multivariate e del problema dell'inferenza legata a questa tipologia di modelli.

La letteratura, sia in campo statistico che probabilistico, ha offerto, nel caso bivariato, uno spazio più ampio alle famiglie di distribuzioni continue (Hutchinson e Lai (1990)) rispetto a quelle discrete (S. Kocherlakota e K. Kocherlakota (1992)).

[1] Che si vedrà essere alcuni dei generatori delle famiglie parametriche ad uno parametro e due parametri.

Infatti, S. Kocherlakota e K. Kocherlakota (1992) osservano che molte famiglie nel caso bivariato discreto non godono di "ovvie e piacevoli" generalizzazioni di cui godono le variabili continue.

Esiste un catalogo talmente ricco non solo di copule ma di "famiglie" di copule che solo elencarle è un'attività sfidante, e da cui questo capitolo si asterrà. Verranno fatti dei "brevi" cenni, come è stato ripetuto più volte in questi primi capitoli, cercando di rilevare i concetti fondamentali che si ritengono propedeuci ai fini della *pair-copula construction* unitamente alla loro rilevanza nel perimetro delle scienze attuariali.

L'appoccio da manuale è distiguere le copule ellittiche da quelle archimedee, e poi generalmente si conclude la trattazione parlando delle copule parametriche. In tal senso, quando si parla di copule parametriche, non è immediato cogliere la loro giusta collocazione: le copule archimedee sono delle particolari copule parametriche oppure il contrario? Un'altro grande dilemma è sull'aggettivazione "archimedea". Perchè? Da dove deriva? Se si è tanto ostinati e si continua lo studio in tal senso si "cede il passo" quando si inizia a leggere delle famiglie di copule `B` e `BB`.

I manuali principali in questo senso sono Joe (1997) e Nelsen (2006), e per questioni più avanzate si rimanda a Kurowicka e Joe (2010).

Si forniranno prima, nella sezione 3.2, i concetti di distribuzioni sferiche ed ellittiche, da cui deriveranno le rispettive copule sferiche ed ellittiche, poi verranno introdotte le copule archimedee nella sezione 3.3. Dopodichè, nella sezione 3.4 si vedranno le principali famiglie parametriche, distinguendole tra famiglie di copule a un parametro (*Bivariate-One-Parameter Families*) e famiglie a due parametri (*Bivariate-Two-Parameter Families*) che sono fondamentali nel contesto della *PCC* in quanto sono quelle che ricorrono negli *output* dei capitoli da 8 a 10.

Infine si osserverà, nel sezione 3.5, che nel'alveo di una visione più ampia delle famiglie archimedee, ovvero nelle generalizzazione di copule archimedee, si ritrova il tipico approccio alla modellizzazione che è proprio della *pair-copula construction*. Si ritroverà infatti, che l'accoppiamento di tipo "*pair-copula*" ovvero quello che accoppia "coppie di copule" è presentato per la prima volta nelle copule archimedee annidate (*nested archimedean copulae (FNA)*).

Joe (1996) è stato il primo a introdurre il concetto di "*nested models*". L'unica differenza che si trova tra una *FNA* e *PCC* è nell'accoppiamento: nella *pair-copula construction* l'accoppiamento avviene non per coppie generiche ma per "coppie di copule bivariate".

Si osserva che la trattazione ha da un lato una finalità di tipo teorico in quanto mostra i principali elementi costituenti alla base dalla *PCC* dall'altro essa non è una "mera rassegna" delle diverse famiglie di copule ma è la chiave di lettura per rendere leggibile l'*output* dell'algortimo di Dißmann (che viene introdotto nella sezione 5.6) e apprezzare la sensibilità delle diverse strutture di dipendenza che possono essere rappresentate con la metodologia di copula.

3.2 Distribuzioni Sferiche

Le distribuzioni sferiche sono interpretabili come misture di distribuzioni uniformi su sfere di differente raggio in $\mathbb{R}$, in particolare sono distribuzioni simmetriche di

vettori aleatori incorrelati a media nulla.

Le distribuzioni sferiche si definiscono attraverso matrici ortonormali ovvero matrici tale che il prodotto di se stessa per la sua trasposta è la matrice identità:

$$Z^T Z = Z Z^T = I_n \,. \tag{3.1}$$

Le matrici ortogonali formano un gruppo in algebra lineare detto il gruppo delle isometrie $Z \in \mathcal{O}(n)$, non è un caso che una conseguenza dell'equazione (3.1) è che la sua trasposta e l'inversa coincidono e che esse sono matrici simmetriche. In generale un vettore $X = (X_1, X_2, \ldots, X_n)$ ha una distribuzione sferica se:

$$ZX =_d X \tag{3.2}$$

per ogni $Z \in \mathcal{O}(n)$ ortogonale.

Definizione 3.1 (Distribuzioni Sferiche)**.** *Un vettore aleatorio $X = (X_1, \ldots, X_n)$ ha una distribuzione sferica, $X \sim S_n(\Phi)$, se ha la rappresentazione stocastica $X =_d RU$, con R variabile aleatoria positiva indipendente dal vettore aleatorio U uniformemente distribuito nell'ipersfera unitaria $S_{n-1} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{x}\boldsymbol{x}\}$ inoltre se X è una distribuzione sferica per ogni $Z \in \mathbb{R}^{n\times n}$ ortogonale. $ZX =_d X$. La funzione caratteristica di X è $\Phi(y) = \mathbb{E}[\exp(iy^t X)] = \Phi(y^t y) = \Phi(\sum_n^i y_i^2)$.*

Oltre alla distribuzione normale multivariata *standard*, appartengono alla classe di distribuzioni sferiche la distribuzione multivariata *standard* di tipo T-Student e la distribuzione logistica *standard*. Si segnala che sebbene si operi sempre su vettori incorrelati solo nel caso gaussiano l'incorrelazione equivale all'indipendenza delle componenti.

Definizione 3.2 (Distribuzioni Ellittiche)**.** *Un vettore aleatorio $\boldsymbol{X}$ di dimensione n si dice appartenere alla* famiglia di distribuzione ellittiche *se e solo se la funzione di densità $f(x)$ possiede la seguente rappresentazione:*

$$f(\boldsymbol{X}; \boldsymbol{\mu}) = k_n \sqrt{|\Sigma|} g((\boldsymbol{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu})) \tag{3.3}$$

Dove:

- $k_n \in \mathbb{R}$ *è una costante dipendente solo dalla dimensione n;*
- $\boldsymbol{\mu} \in \mathbb{R}^n$ *è il vettore della medie;*
- $\Sigma \in \mathbb{R}^{n\times n}$ *è una matrice simmetrica definita positiva con determinante $|\Sigma|$;*
- $g \colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ *è una classe di funzioni.*

La distribuzione Normale multivariata e la distribuzione T-Student multivariata appartengono alla classe di distribuzioni ellittiche perché rispettano la forma funzionale sopracitata. In particolare, si osserva che la normale multivariata nel caso in cui $\Sigma = I_n$, appartiene alla classe di distribuzioni sferiche (normale multivariata sferica). Le distribuzioni ellittiche godono di diverse proprietà, si ricordano quelle più importanti nel perimetro della modellistica:

- Le marginali di una distribuzione ellittica sono esse stesse ellittiche in quanto hanno lo stesso generatore;
- Combinazioni lineari di distribuzioni ellittiche indipendenti aventi la stessa matrice Σ sono ancora distribuzioni ellittiche indipendenti[2];
- Il coeffciente di correlazione lineare è la naturale misura di dipendenza tra variabili aleatorie con distribuzione congiunta ellittica.

H.-B. Fang, K.-T. Fang e Kotz (2002) forniscono un'espressione generale in forma chiusa per le copule ellittiche.

Le distribuzioni ellittiche hanno la virtù di estendersi facilmente in grandi dimensioni: in dimensione n possiedono almeno $\frac{n(n-1)}{2}$ parametri. Tuttavia, le asimmetrie radiali e il comportamento di code asimmetriche non possono essere catturate all'interno di questa famiglia di copule. Le applicazioni e i limiti delle copule ellittiche sono discusse in modo più dettagliato da Frahm, Junker e Szimayer (2003), mentre Hult e Lindskog (2002) trattano la dipendenza agli estemi della coda della copula.

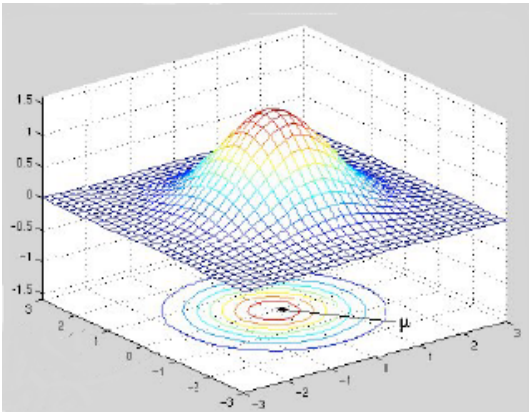


Figura 3.1. Distribuzione normale bivariata sferica.

Nella figura 3.2 sono presentati gli *scatter plot* e le curve di livello di una copula gaussiana (a sinistra) e una copula T-Student (a destra) aventi stessi parametri[3].

Una generalizzazione delle copule ellittiche sono rappresentate dalle copule ellittiche generalizzate di tipo iperbolico (*elliptical-generalized-hyperbolic-copulae (GH)*), che sono state introdotte da Barndorff-Nielsen sia nella forma univariata che nella forma multivariata e sono diventate molto famose nel contesto finanziario in quanto possiedono code *semi-heavy tails*, ovvero code più pesanti rispetto alla distribuzione gaussiana ma più leggere rispetto alla distribuzione T-Student (non a caso esse sono un caso limite). Esistono tutti i momenti della distribuzione *GH* e la funzione generatrice dei momenti è disponibile in forma chiusa.

[2]Questa è analoga alla proprietà riproduttiva della normale.
[3]Sono stati considerati i casi $\tau_K \in \{0.25, 0.75\}$.

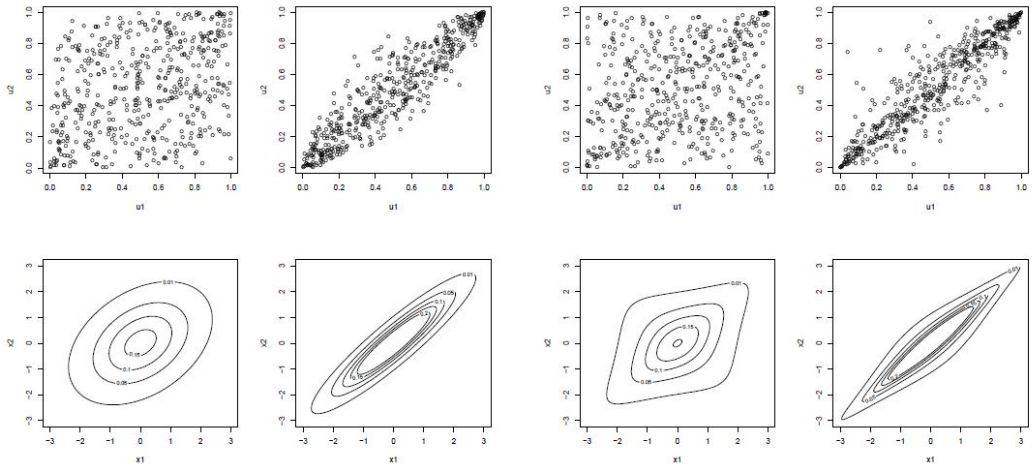


Figura 3.2. Copule ellittiche bivariate: dipendenza simmetrica.

3.3 Copule Archimedee

L'aggettivazione "archimedee" trova un senso dalla seguente proprietà *archimedea*:

> *Dati comunque due segmenti di lunghezza rispettivamente s_1 e s_2 tali che $s_1 \leq s_2$, esiste sempre un multiplo intero di s_1 maggiore o uguale a s_2.*

Questa proprietà geometrica, si vedrà, acquista un senso anche nel perimetro della famiglia di copule che ora si andranno a definire.

Definizione 3.3 (Trasformata di Laplace)**.** *Sia M la funzione di ripartizione di distribuzione univariata positiva tale che $M(0) = 0$ si definisce ϕ la* trasformata di Laplace *di M la funzione:*

$$\phi(s) = \int_0^\infty \exp\{-s\omega\}\,\mathrm{d}M(\omega),\ s \geq 0 \tag{3.4}$$

Si noti che con questa definizione, $\phi(-t)$ è la funzione generatrice dei momenti di M.

Per ogni arbitraria funzione di ripartizione univariata F si dimostra che esiste *un'unica* funzione di ripartizione G tale che:

$$F(x) = \int_0^\infty G^\alpha(x)\,\mathrm{d}M(\alpha) = -\phi\big(-\log G(x)\big)\,. \tag{3.5}$$

L'equazione (3.5) può essere invertita ricavando:

$$G(x) = \exp\Big\{-\phi^{-1}\big(F(x)\big)\Big\}\,. \tag{3.6}$$

Nel caso bivariato, si consideri una distribuzione F di distribuzioni marginali F_1 e F_2. Dati $G_1 = \exp\{-\phi^{-1}(F_1)\}$ e $G_2 = \exp\{-\phi^{-1}(F_2)\}$, la funzione di ripartizione della congiunta è data da:

$$\int_0^\infty G_1^\alpha G_2^\alpha\,\mathrm{d}M(\alpha) = \phi\big(-\log(G_1) - \log(G_2)\big) = \phi\big(\phi^{-1}(F_1) + \phi^{-1}(F_2)\big)\,. \tag{3.7}$$

La copula che si ottiene prendendo F_1 e F_2 distribuzioni uniformi è la seguente:

$$C(u_1, u_2) = \phi(\phi^{-1}(u_1) + \phi^{-1}(u_2)) \,. \tag{3.8}$$

Tutte le copule che rispettano questa forma funzionale, piuttosto semplice, sono dette copule archimedee. E proprio dalla equazione sopracitata deriva la famosa aggettivazione "archimedee". Affinché esse abbiano una forma chiusa è necessario che sia ϕ che ϕ^{-1} abbiano una forma chiusa (si veda appendice A.7 per degli esempi).

Si potrebbero costruire infinite distribuzioni bivariate a partire da misture della forma:

$$\int G_1(x_1, \alpha) G_1(x_2, \alpha) dM(\alpha) \,, \tag{3.9}$$

ma in genere queste costruzioni non producono copule con una forma chiusa.

Inoltre, in dimensione m, siano $F_1, \ldots, F_m$ le funzioni di ripartizione di funzioni univariate, una "naturale" e semplice estensione di copule archimedee in dimensione m aventi F della seguente forma:

$$F = \phi\Big(\sum_{j=1}^{m} \phi^{-1}(F_j)\Big) \,, \tag{3.10}$$

e la copula archimedea multivariata è data dalla seguente espressione:

$$C(u) = \phi\Big(\sum_{j=1}^{m} \phi^{-1}(u_j)\Big) \,. \tag{3.11}$$

Essa è una permutazione simmetrica delle m uniformi scambiabili.

Questa famiglia di copule sono state esplorate in modo approfondito in Joe (1996), Nelsen (2006) e diversi altri lavori come A. McNeil, Frey e Embrechts (2005) e A. McNeil e Nešlehová (2009). Nella figura 3.3, si può vedere lo *scatter plot* e le curve di livello di una copula di Gumbel[4] e una copula di Clayton[5] con i medesimi parametri[6].

3.4 Famiglie di copule parametriche

Un obiettivo della teoria della modellazione della dipendenza e delle copule multivariate è quello di sviluppare famiglie parametriche che siano appropriate, in quanto si dimostra che sebbene le copule multivariate archimedee siano modelli estremamente flessibili all'aumentare della dimensione colgono sempre meno la dipendenza negativa. Si introdurranno la `B` *family* e la `BB` *family*. La notazione `B` e `BB` è stata introdotta da Joe (1997) e da Nikoloulopoulos, Joe e Li (2012), e rappresentano rispettivamente le famiglie di copule bivariate parametriche a un parametro e a due parametri.

[4] Che si vedrà essere anche detta `B6`

[5] Che si vedrà essere anche detta `B4`

[6] Viene considerato il caso $\tau_K \in \{0.25, 0.75\}$.

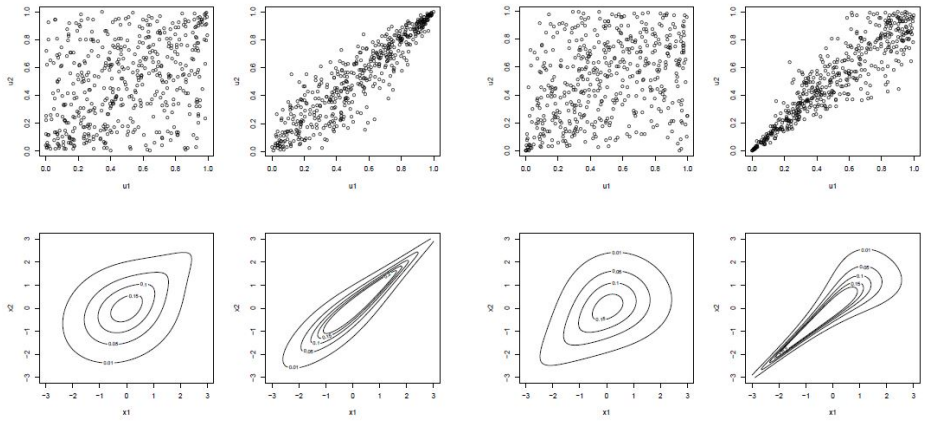


Figura 3.3. Copule Archimedee bivariate: dipendenza di coda "upper" e "lower".

3.4.1 Famiglie di copule bivariate a un parametro

Segue il catalogo delle famiglie di copule a un parametro. Le prime 8 possiedono le seguenti caratteristiche:

- il parametro può essere intepretato come un fattore di interpolazione tra la copula indipendente e il limite superiore di Frechet;
- sono assolutamente continue;
- hanno supporto su tutto $\mathbb{I}^2$;
- hanno simmetria radiale.

Se queste condizioni vengono rilassate esistono infinite famiglie a un parametro oltre a quelle elencate qui o in qualsiasi altra fonte. Seguendo Joe (1997), abbiamo presentato alcune di queste famiglie, anche se queste non sono supportate dai *software* per le *PCC* e, conseguentemente, i capitoli da 8 a 10 non le usano.

Il fatto che le prime 8 famiglie di tipo `B` godano della proprietà di "parsimonia" è un buon *punto d'inizio* per la modellizzazione. Esse sono utili sia nel contesto bivariato, che multivariato, che per le serie storiche di Markov del primo ordine. La notazione $C(u, v; \delta)$ è usata per le famiglie con dipendenza dal parametro δ: all'aumentare del valore in modulo del parametro aumenta il livello di dipendenza. Oltre alle proprietà di dipendenza, verranno annoverate altre proprietà come ad esempio la riflessione e l'esistenza della forma in trasformata di Laplace del tipo:

$$C(u_1, u_2) = \phi(\phi^{-1}(u_1) + \phi^{-1}(u_2)) \tag{3.12}$$

L'elenco delle famiglie parametriche di trasformate di Laplace che si useranno è consultabile nella appendice A.7.

Nel seguito si useranno le seguenti notazioni:

$$\overline{u} = 1 - u\,; \tag{3.13}$$
$$\overline{v} = 1 - v\,; \tag{3.14}$$
$$\tilde{u} = -\log u\,; \tag{3.15}$$
$$\tilde{v} = -\log v\,. \tag{3.16}$$

Famiglia B1. Normale Bivariata

Per $0 \leq \delta \leq 1$ si ha:

$$C(u, v; \delta) = \Phi_\delta(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{3.17}$$

dove Φ è la funzione di ripartizione di una normale *standard* $\mathcal{N}(0, 1)$, Φ^{-1} è la sua inversa e Φ_δ è la funzione di ripartizione di una normale bivariata *standard* con correlazione δ. In altre parole, per $x = \Phi^{-1}(u)$, $y = \Phi^{-1}(v)$, la sua densità è:

$$c(u, v; \delta) = (1 - \delta^2)^{-\frac{1}{2}} \exp\Big\{-\frac{1}{2}(1 - \delta^2)^{-1}[x^2 + y^2 - 2\delta xy]\Big\} \exp\Big\{\frac{1}{2}[x^2 + y^2]\Big\} \tag{3.18}$$

Proprietà: simmetria radiale, estensione multivariata, estensione alla dipendenza negativa.

Famiglia B2. Plackett (1965)

Per $0 \leq \delta < \infty$ si ha:

$$C(u, v; \delta) = \frac{1}{2}\eta^{-1}\Big\{1 + \eta(u + v) - \Big[\big(1 + \eta(u + v)\big)^2 - 4\delta\eta uv\Big]^{\frac{1}{2}}\Big\} \tag{3.19}$$

dove $\eta = \delta - 1$.

Proprietà: simmetria radiale, estensione alla dipendenza negativa.

Famiglia B3. Frank (1979)

Per $0 \leq \delta < \infty$ si ha:

$$C(u, v; \delta) = -\delta^{-1} \log\left(\frac{\eta - (1 - \exp\{-\delta u\})(1 - \exp\{-\delta v\})}{\eta}\right) \tag{3.20}$$

dove $\eta = 1 - \exp(-\delta)$.

Proprietà: simmetria radiale, estensione multivariata parziale, estensione a dipendenza negativa, mistura di potenze con la trasformata di Laplace LTD.

Famiglia B4. Kimeldori e Sampson (1975)

Questa copula è spesso conosciuta con il nome di *copula di Clayton*. Per $0 \leq \delta < \infty$ si ha:

$$C(u, v; \delta) = (u^{-\delta} + v^{-\delta} - 1)^{-\frac{1}{\delta}} \tag{3.21}$$

Proprietà: dipendenza di coda inferiore, estensione multivariata parziale, estensione a dipendenza negativa, mistura di potenze con la trasformata di Laplace LTB.

Famiglia `B5`. Joe (1993)

Per $0 \leq \delta < \infty$ si ha:

$$C(u, v; \delta) = 1 - (\overline{u}^{\delta} + \overline{v}^{\delta} - \overline{u}^{\delta}\overline{v}^{\delta})^{\frac{1}{\delta}} \tag{3.22}$$

Proprietà: dipendenza di coda superiore, estensione multivariata parziale, estensione a dipendenza negativa, mistura di potenze con la trasformata di Laplace LTC.

Famiglia `B6`. Gumbel (1935)

Per $0 \leq \delta < \infty$ si ha:

$$C(u, v; \delta) = \exp\{-(\tilde{u}^{\delta} + \tilde{v}^{\delta})^{-\frac{1}{\delta}}\} \tag{3.23}$$

Proprietà: dipendenza di coda superiore, copula ai valori estremi, estensione multivariata parziale, estensione a dipendenza negativa, mistura di potenze con la trasformata di Laplace LTA.

Famiglia `B7`. Galambos (1975)

Per $0 \leq \delta < \infty$ si ha:

$$C(u, v; \delta) = uv \exp\{(\tilde{u}^{-\delta} + \tilde{v}^{-\delta})^{-\frac{1}{\delta}}\} \tag{3.24}$$

Proprietà: dipendenza di coda superiore, copula ai valori estremi, estensione multivariata parziale.

Famiglia `B8`. Hüsler e Reiss (1989)

Per $0 \leq \delta < \infty$ si ha:

$$C(u, v; \delta) = \exp\left\{-\tilde{u}\Phi\left[\delta^{-1} + \frac{1}{2}\delta \log\left(\frac{\tilde{u}}{\tilde{v}}\right)\right] - \tilde{v}\Phi\left[\delta^{-1} + \frac{1}{2}\delta \log\left(\frac{\tilde{v}}{\tilde{u}}\right)\right]\right\} \tag{3.25}$$

dove Φ è la funzione di ripartizione di una normale *standard* $\mathcal{N}(0, 1)$.

Proprietà: dipendenza di coda superiore, copula ai valori estremi, estensione multivariata.

Si riporta nella tabella 3.1 il valore del parametro al variare di ρ_S per le prime otto famiglie di bicopule `B`.

ρ_S	B1	B2	B3	B4	B5	B6	B7	B8
0	0	1	0	0	1	1	0	0
0.1	0.105	1.35	0.60	0.14	1.12	1.07	0.28	0.58
0.2	0.209	1.84	1.22	0.31	1.27	1.16	0.40	0.73
0.3	0.313	2.52	1.88	0.51	1.46	1.26	0.51	0.88
0.4	0.416	3.54	2.61	0.76	1.69	1.38	0.65	1.05
0.5	0.518	5.12	3.45	1.06	1.99	1.54	0.81	1.24
0.6	0.618	7.76	4.47	1.51	2.39	1.75	1.03	1.50
0.7	0.717	12.7	5.82	2.14	3.00	2.07	1.34	1.86
0.8	0.813	24.2	7.90	3.19	4.03	2.58	1.86	2.45
0.9	0.908	66.1	12.2	5.56	6.37	3.73	3.01	3.73
1	1	∞	∞	∞	∞	∞	∞	∞

Tabella 3.1. Valori dei parametri corrispondenti per un ρ_S di Spearman fissato.

Alcune famiglie di copule bivariate a un solo parametro che non sono così semplici o che non hanno proprietà altrettanto gradevoli di quelle già elencate sono riportate di seguito.

Famiglia B9. Raftery (1984)

Per $-1 \leq \delta \leq 1$ si ha:

$$C(u, v; \delta) = B(\max\{u, v\}, \min\{u, v\}; \delta)\,, \tag{3.26}$$

dove:

$$B(x, y; \delta) = x - \frac{1-\delta}{1+\delta} x^{\frac{1}{1-\delta}} \left[y^{\frac{-\delta}{1-\delta}} - y^{\frac{1}{1-\delta}} \right] \tag{3.27}$$

Proprietà: dipendenza di coda inferiore.

Famiglia B10. Morgenstern (1956)

Per $-1 \leq \delta \leq 1$ si ha:

$$C(u, v; \delta) = uv\big[1 + \delta(1-u)(1-v)\big] \tag{3.28}$$

Proprietà: simmetria radiale, dipendenza positiva per $\delta > 0$ e negativa per $\delta < 0$, estensione multivariata.

Famiglia B11

Per $0 \leq \delta \leq 1$ si ha:

$$C(u, v; \delta) = \delta \min\{u, v\} + (1-\delta)uv \tag{3.29}$$

Proprietà: simmetria radiale, estensione multivariata, componenti singolari di massa δ.

Famiglia `B12`

Per $0 \leq \delta \leq 1$ si ha:

$$C(u, v; \delta) = \delta\big[\min\{u, v\}\big]^{\delta} + [uv]^{1-\delta} \tag{3.30}$$

Proprietà: simmetria radiale, estensione multivariata, componenti singolari di massa $\frac{\delta}{2-\delta}$.

3.4.2 Famiglie di copule bivariate a due parametri

Le famiglie a due parametri possono essere utilizzate per catturare più di una tipo di dipendenza, ad esempio un parametro può essere adottato per cogliere la dipendenza dalla coda superiore mentre il secondo per rappresentare la concordanza, oppure un parametro per la dipendenza dalla coda superiore e uno per la dipendenza dalla coda inferiore. Di seguito sono riportate le principali famiglie bivariate a due parametri. Esse hanno tutte la forma funzionale seguente:

$$C(u, v, \theta; \delta) = \psi\Big(-\log K\Big(\exp\{-\psi^{-1}(u)\}, \exp\{-\psi^{-1}(v)\}\Big)\Big) \tag{3.31}$$

Dove K è una famiglia tra le `B` e ψ è una famiglia delle trasformate di Laplace.

Famiglia `BB1`

Considerando l'equazione (3.31) se si pone K alla famiglia `B6` e ψ la famiglia delle trasformate di Laplace LTB si ricava:

$$C(u, v, \theta; \delta) = \left\{1 + [u^{-\theta} - 1]^{\delta} + [u^{-\theta} - 1]^{\frac{1}{\delta}}\right\}^{\frac{1}{\delta}}, \tag{3.32}$$

per $\theta > 0$ e $\delta \geq 1$.

Proprietà: indipendenza, dipendenza di coda inferiore o inferiore al variare dei parametri e copula ai valori estremi.

Famiglia `BB2`

Considerando l'equazione (3.31) se si pone K alla famiglia `B4` e ψ la famiglia delle trasformate di Laplace LTB si ricava:

$$C(u, v, \theta; \delta) = \left\{1 + \delta^{-1} \log\Big(\exp\{\delta(u - \theta - 1)\} + \exp\{\delta(v^{-\theta} - 1)\} - 1\Big)\right\}^{\frac{1}{\theta}}, \tag{3.33}$$

per $\theta > 0$ e $\delta > 0$.

Proprietà: per $\theta, \delta \to 0$ la famiglia diventa una `B4`, dipendenza di coda inferiore o inferiore al variare dei parametri, concordanza al variare dei parametri.

Famiglia BB3

Considerando l'equazione (3.31) se si pone K alla famiglia B4 e ψ la famiglia delle trasformate di Laplace LTA si ricava:

$$C(u,v,\theta;\delta) = \exp\left\{-\left[\delta^{-1}\log\Big(\exp\{\delta\tilde{u}^{\theta}\}+\exp\{\delta\tilde{v}^{\theta}\}-1\Big)\right]^{\frac{1}{\theta}}\right\}, \tag{3.34}$$

per $\theta \geq 1$ e $\delta > 0$.

Proprietà: per $\theta = 1$ la famiglia diventa una B4 mentre per $\theta \to 0$ la famiglia diventa una B6, al variare dei parametri si ha indipendenza oppure dipendenza di coda inferiore o superiore, copula ai valori estremi e concordanza.

Famiglia BB4

Considerando l'equazione (3.31) se si pone K alla famiglia B7 e ψ la famiglia delle trasformate di Laplace LTB si ricava:

$$C(u,v,\theta;\delta) = \left\{u^{-\theta}+v^{-\theta}-1-\left[(u^{-\theta}-1)^{-\delta}+(v^{-\theta}-1)^{-\delta}\right]^{-\frac{1}{\delta}}\right\}^{-\frac{1}{\theta}}, \tag{3.35}$$

per $\theta \geq 0$ e $\delta > 0$.

Proprietà: per $\delta \to 0$ la famiglia diventa una B4 mentre per $\theta \to 0$ la famiglia diventa una B7, al variare dei parametri si ha indipendenza oppure dipendenza di coda inferiore o superiore, copula ai valori estremi e concordanza.

Famiglia BB5

Considerando l'equazione (3.31) se si pone K alla famiglia B7 e ψ la famiglia delle trasformate di Laplace LTA si ricava:

$$C(u,v,\theta;\delta) = \exp\left\{-\left[\tilde{u}^{\theta}+\tilde{v}^{\theta}-(\tilde{u}^{-\theta\delta}+\tilde{v}^{-\theta\delta})^{-\frac{1}{\delta}}\right]^{\frac{1}{\theta}}\right\} \tag{3.36}$$

per $\theta \geq 1$ e $\delta > 0$.

Proprietà: per $\delta \to 0$ la famiglia diventa una B6 mentre per $\theta = 1$ la famiglia diventa una B7, al variare dei parametri si ha dipendenza di coda inferiore o superiore, copula ai valori estremi e concordanza.

Famiglia BB6

Considerando l'equazione (3.31) se si pone K alla famiglia B6 e ψ la famiglia delle trasformate di Laplace LTC si ricava:

$$C(u,v,\theta;\delta) = 1-\left(1-\exp\left\{-\left[\left(-\log(1-\overline{u}^{\theta})\right)^{\delta}+\left(-\log(1-\overline{v}^{\theta})\right)^{\delta}\right]^{\frac{1}{\delta}}\right\}\right)^{-\frac{1}{\theta}} \tag{3.37}$$

per $\theta \geq 1$ e $\delta \geq 1$.

Proprietà: per $\theta = 1$ la famiglia diventa una `B6` mentre per $\delta = 1$ la famiglia diventa una `B5`, al variare dei parametri si ha dipendenza di coda inferiore o superiore, la concordanza aumenta quando θ aumenta.

Famiglia `BB7`

Considerando l'equazione (3.31) se si pone K alla famiglia `B4` e ψ la famiglia delle trasformate di Laplace LTC si ricava:

$$C(u,v,\theta;\delta) = 1 - \Big(1 - \big[(1-\overline{u}^{\theta})^{-\delta} + (1-\overline{v}^{\theta})^{-\delta} - 1\big]^{-\frac{1}{\delta}}\Big)^{-\frac{1}{\theta}} \tag{3.38}$$

per $\theta \geq 1$ e $\delta > 0$.

Proprietà: per $\theta = 1$ la famiglia diventa una `B4`, al variare dei parametri si ha indipendenza o dipendenza di coda inferiore o superiore, la concordanza aumenta quando θ aumenta.

Famiglia `BB8`

Questa copula è creata usando la famiglia delle trasformate di Laplace a due parametri LTJ. La sua distribuzione è:

$$C(u,v,\theta;\delta) = \delta^{-1}\left[1 - \Big\{1 - \big[1-(1-\delta)^{\theta}\big]^{-1}\big[1-(1-\delta u)^{\theta}\big]\big[1-(1-\delta v)^{\theta}\big]\Big\}^{\frac{1}{\theta}}\right] \tag{3.39}$$

per $\theta \geq 1$ e $0 \leq \delta \leq 1$.

Proprietà: per $\delta = 1$ la famiglia diventa una `B5` mentre per $\theta \to \infty$ la famiglia diventa una `B3`, la famiglia non possiede dipendenze di coda, la concordanza aumenta quando θ o δ aumentano.

Famiglia `BB9`

Questa copula è creata usando la famiglia delle trasformate di Laplace a due parametri LTL. La sua distribuzione è:

$$C(u,v,\theta;\alpha) = \exp\Big\{-\big[(\alpha - \log u)^{\theta} + (\alpha - \log v)^{\theta} - \alpha^{\theta}\big]^{\frac{1}{\theta}} + \alpha\Big\} \tag{3.40}$$

per $\theta \geq 1$ e $\alpha > 0$.

Proprietà: La concordanza aumenta all'aumentare di θ o al diminuire di α.

Famiglia `BB10`

Questa copula è creata usando la famiglia delle trasformate di Laplace a due parametri LTM. La sua distribuzione è:

$$C(u,v,\theta;\alpha) = uv\Big\{1 - \theta\big(1-u^{\frac{1}{\alpha}}\big)\big(1-v^{\frac{1}{\alpha}}\big)\Big\}^{-\alpha} \tag{3.41}$$

per $0 \leq \theta \leq 1$ e $\alpha > 0$.

Proprietà: la concordanza cresce con θ per α fissato, descresce in α per $\theta = 1$, ma non vi è alcuna relazione tra concordanza e il variare di α per $0 \leq \theta < 1$.

3.5 Metodi di costruzione gerarchici

Nell'esigenza sempre più marcata di strutture multivariate di grandi dimensioni e unitamente alle proprietà di composizione di cui godono le forme funzionali delle copule archimedee Joe (1997), Whelan (2004) e Savu e Trede (2010) introducono il concetto di struttura gerarchica nell'alveo della teoria delle copule. L'idea alla base di questo approccio è costruire una struttura gerarchica di copule archimedee su differenti livelli.

Kurowicka e Joe (2010) distinguono le *classiche* copule archimedee dalle queste forme che hanno suscitato particolare interesse grazie a numerose proprietà che le rendono semplici da implementare e analizzare: il modo più comune di definire una copula archimedea multivariata è la copula archimedea scambiabile (*exchangeable archimedean copula (EAC)*). La *partially nested archimedean construction (PNAC)* è una costruzione di scambiabilità parziale che deve soddisfare alcune particolari condizioni affinchè il risultato di tale costruzione sia esso stesso una copula.

Le condizioni sono diverse, ad esempio sul generatore ϕ deve essere una funzione decrescente, invertibile e tale che $\phi(1) = 0$.

Le principali famiglie di tipo annidato sono le seguenti:

- *Full Nested Archimedean*;
- *Partially Nested Archimedean*;
- *Hierarchical Nested Archimedean.*

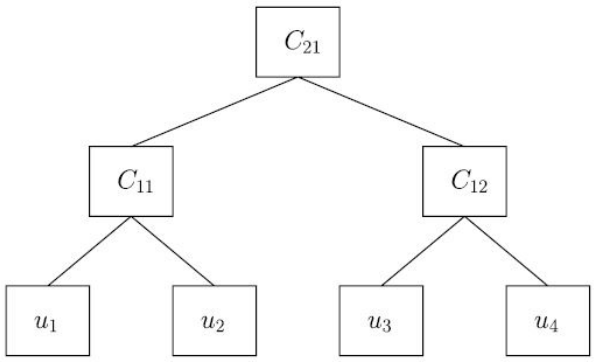


Figura 3.4. Costruzione di copule archimedee annidate parzialmente.

Come si vede nelle figure da 3.5 a 3.7, le copule vengono accoppiate in una copula risultante, utilizzando un generatore ϕ (che esso stesso può essere combinato con un'altro generatore ψ questo grazie alle proprietà di cui godono le copule archimedee). Le copule *fully nested archimedean* (FNA) sono state introdotte da

Joe (1997), Whelan (2004) e Savu e Trede (2010) e come si può vedere nella figura 3.5 si costruisce una gerarchia di copule archimedee a diversi livelli. Nel caso delle *fully nested* ovvero delle copule annidatate completamente la bi-copula viene generata non da un'altra bi-copula ma da una bi-copula e una singola copula (si aggiunge quindi una copula alla volta) ad ogni accoppiamento si aumenta il livello della struttura gerarchica.

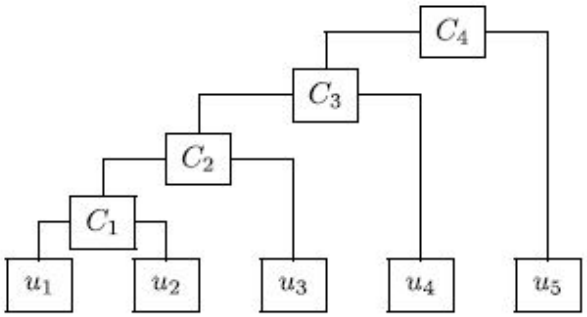


Figura 3.5. Copule *full nested archimedean* nel caso $n = 5$.

In alternativa vengono mescolate le ordinarie copule archimedee con le *FNA*, generando le copule *partially nested Archimedean (PNA)*.

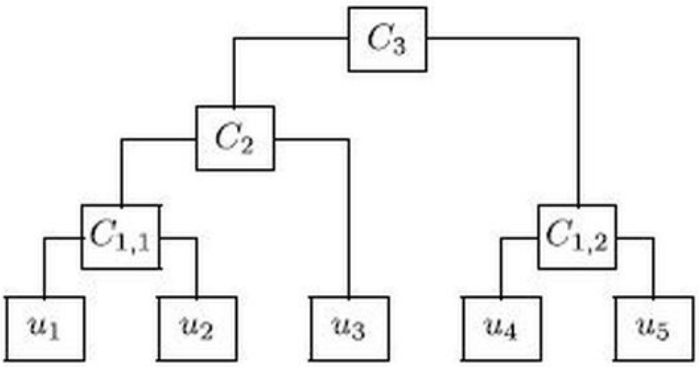


Figura 3.6. copule *partially nested archimedean* nel caso $n = 5$.

Nelle copule di tipo *hierarchical nested archimedean* si costruisce una gerarchia di copule archimedee a diversi livelli, indicizzati da $l = 1, \ldots, L$.

Morillas (2005) e Liebscher (2008) forniscono una definizione di copule *generalized multiplicative archimedean* attraverso la definizione di un generatore moltiplicativo definito nel modo seguente:

$$C(u_1, \ldots, u_n) = \theta^{[-1]}(\theta(u_1) \cdot \ldots \cdot \theta(u_n)) \tag{3.42}$$

Infine, utilizzando un approccio analogo alla *nested archimedean construction* è quello della *pair-copula construction* che si osserva, conserva la logica di tipo

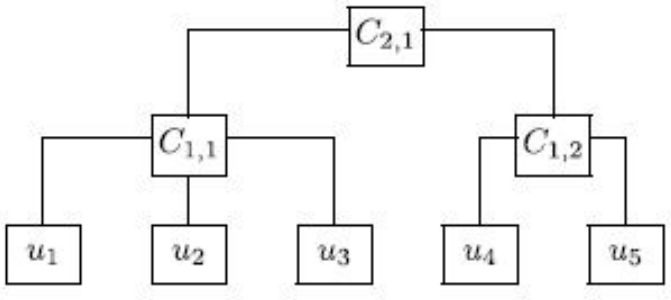


Figura 3.7. Copule *hierarchical nested archimedean* nel caso $n = 5$.

gerarchico. Lo schema di modellazione si basa sulla decomposizione di una densità multivariata in $\frac{n(n-1)}{2}$ densità di copula bivariata, di cui le prime $n-1$ sono strutture di dipendenza incondizionate. Nella *PPC* a differenza della *PNAC* le copule non devono nemmeno appartenere alla stessa classe.

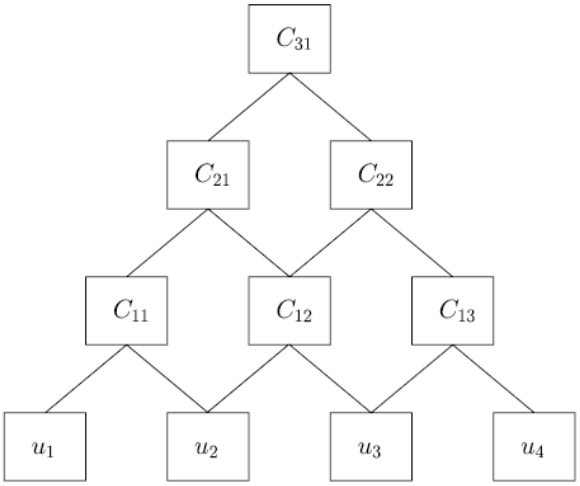


Figura 3.8. Costruzione di tipo *pair-copula*.

3.6 Limiti delle tradizionali copule

Riassumendo quanto osservato in questo capitolo, le copule, e in particolare le copule in dimensione 2, rappresentano uno strumento flessibile ed esaustivo per lo studio della dipendenza nel caso bidimensionale. Tuttavia, all'aumentare del numero delle variabili, Joe (1996) afferma che esse non sono più la misura *canonica* per studiare la dipendenza tra fenomeni in quanto, la n-copula per n sufficentemente grande non coglie più la "reale" dipendenza. Inoltre al netto delle copule gaussiane e di Student multivariate di cui esiste l'estensione multivariata, la selezione di copule parametriche di grandi dimensioni è piuttosto limitata. I recenti sviluppi

in quest'area tendono verso strutture gerarchiche, ovvero basate su copule aventi strutture gerarchiche. La più promettente di queste è la costruzione a *pair-copule*, ovvero la *pair copula construction (*PCC*)*, che sarà l'oggetto dei prossimi capitoli.

Capitolo 4

La *Pair-Copula Construction*

> Though dating back to 1959 when the term "copulae" was coined, copula models only started their triumphal procession in the mid-1990s.
>
> Fischer (2010)

Nel capitolo 3 è stato osservato che seppur esistano generalizzazioni a n dimensioni delle copule bidimensionali, queste ultime hanno forti limitazioni sulle strutture di dipendenza che riescono a soddisfare. In Joe (1997) vengono proposte varie modalità per costruire copule multivariate, ma non tutte hanno una formulazione analitica per le funzioni di distribuzioni ed in generale sono difficili da interpretare.

Per ovviare al problema, è stato introdotto il metodo delle *pair-copula constructions* (o *PCC*). Nella *PCC*, la distribuzione multivariata viene rappresentata attraverso un metodo grafico (ovvero basato sui grafi) in cui la componente stocastica del modello viene modellata usando delle copule bivariate e i cui argomenti sono funzioni di distribuzione condizionata. Il primo ad utilizzare questo tipo di decomposizione è stato Joe (1996), anche se è stato il lavoro di Bedford e Cooke (2002) ad avergli dato la forma attuale. In particolare, questi autori hanno introdotto lo strumento delle *regular vine* che si vedrà nella sezione 4.2. La teoria ha però preso piede solo dopo il lavoro di Aas et al. (2009), che ne ha anche fissato il nome.

Per facilitare la comprensione di questa metodologia, nella sezione 4.1 si fornirà un esempio esplicativo di come una funzione di densità multivariata possa essere decomposta come prodotto di copule bivariate, condizionate e non, e delle densità marginali. La sezione si concluderà dunque con un accenno all'ipotesi semplificante per le *PCC*, che sarà enucleata nella sezione 4.4. La componente grafica delle *PCC* e i prerequisiti di teoria dei grafi necessari per comprenderla, saranno introdotti nella sezione 4.2. La sezione si concluderà con un'analisi delle rappresentazioni matriciali proposte per le *regular vine* (sezione 4.2.5). Si accennerà inoltre al numero di *regular vine* su un certo n (sezione 4.2.4). Infine, nella sezione 4.3 si fornirà una definizione formale alle *PCC* e saranno enunciati i principali risultati teorici.

4.1 Decomposizione di una Copula in *Pair-Copulas*

Per comprendere il principio alla base della decomposizione *PCC*, si mostrerà come sia possibile decomporre una distribuzione multivariata in prodotti di copule bivariate condizionate.

Prima di procedere con l'esempio, è utile richiamare qualche risultato elementare di teoria della probabilità. Una delle decomposizioni più sfruttate dai modelli grafici è la *regola del concatenamento di probabilità.*

Proposizione 4.1 (Regola del Concatenamento di Probabilità)**.** *Siano* $\{A_i\}_{1\leq i\leq n}$ *n eventi casuali, allora* $\mathbb{P}(\bigcap_1^n A_i) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\cdots\mathbb{P}(A_n|\bigcap_1^{n-1} A_i)$

Nel caso di n variabili aleatorie $\{X_i\}_{1\leq i\leq n}$ si può riscrivere la proposizione 4.1 usando le funzioni di ripartizione o le loro densità (quando esistono).

Proposizione 4.2 (Regola del Concatenamento per Densità di Probabilità)**.** *Siano* $\{X_i\}_{1\leq i\leq n}$ *n variabili aleatorie con funzione di ripartizione congiunta* $F_{1,\dots,n}$ *assolutamente continua con marginali* $\{F_i\}_{1\leq i\leq n}$ *continue e strettamente crescenti, allora valgono le seguenti equazioni:*

$$F_{1,\dots,n}(x_1,\dots,x_n) = F_1(x_1)F_{2|1}(x_2|x_1)\cdots F_{n|1,\dots,n-1}(x_n|x_1,\dots,x_n) \tag{4.1}$$

$$f_{1,\dots,n}(x_1,\dots,x_n) = f_1(x_1)f_{2|1}(x_2|x_1)\dots f_{n|1,\dots,n-1}(x_n|x_1,\dots,x_{n-1}) \tag{4.2}$$

dove $f_{1,\dots,n}$ *è la funzione di densità congiunta e le* $\{f_i\}_{1\leq i\leq n}$ *sono le funzioni di densità marginali.*

Si noti che la decomposizione dipende dall'ordine scelto dalle variabili e che quindi esistono $n!$ fattorizzazioni di questa forma, uno per ogni permutazione degli indici. Per esempio, se $n = 3$ allora ognuna di queste decomposizione è possibile:

$$f_{1,2,3}(x_1,x_2,x_3) = f_1(x_1)f_{2|1}(x_2|x_1)f_{3|1,2}(x_3|x_1,x_2), \tag{4.3}$$

$$f_{1,2,3}(x_1,x_2,x_3) = f_1(x_1)f_{3|1}(x_3|x_1)f_{2|1,3}(x_2|x_1,x_3), \tag{4.4}$$

$$f_{1,2,3}(x_1,x_2,x_3) = f_2(x_2)f_{1|2}(x_1|x_2)f_{3|1,2}(x_3|x_1,x_2), \tag{4.5}$$

$$f_{1,2,3}(x_1,x_2,x_3) = f_2(x_2)f_{3|2}(x_3|x_2)f_{1|2,3}(x_1|x_2,x_3), \tag{4.6}$$

$$f_{1,2,3}(x_1,x_2,x_3) = f_3(x_3)f_{1|3}(x_1|x_3)f_{2|1,3}(x_2|x_1,x_3), \tag{4.7}$$

$$\text{e } f_{1,2,3}(x_1,x_2,x_3) = f_3(x_3)f_{2|3}(x_2|x_3)f_{1|2,3}(x_1|x_2,x_3). \tag{4.8}$$

È altresì importante sottolineare che la funzione densità di una variabile aleatoria $\boldsymbol{X}$ condizionata a $\boldsymbol{Y} = \boldsymbol{y}$, ovvero la funzione

$$f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{f_{\boldsymbol{X},\boldsymbol{y}}(\boldsymbol{x},\boldsymbol{y})}{f_{\boldsymbol{Y}}(\boldsymbol{y})}\,, \tag{4.9}$$

è una funzione nella sola $\boldsymbol{X}$ in quanto è stato posto $\boldsymbol{Y} = \boldsymbol{y}$, mentre diventa una funzione in $\boldsymbol{X}$ e $\boldsymbol{Y}$ quando il valore di $\boldsymbol{y}$ non è fissato. Quindi le funzioni $f_{2|1}$ e $f_{3|1,2}$ nell'equazione (4.3) sono funzioni rispettivamente a 2 e 3 variabili e non in una sola variabile come è usuale vederle.

Si consideri quindi l'equazione (1.10) del teorema 1.2 (ovvero il teorema di Sklar). Questa equazione può essere scritta anche in termini delle funzioni di densità ovvero:

$$\begin{aligned}\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{1,\dots,n}(x_1,\dots,x_n) &= \frac{\partial^n}{\partial x_1 \cdots \partial x_n} C(F_1(x_1),\dots,F_n(x_n)) \\ f_{1,\dots,n}(x_1,\dots,x_n) &= \left(\frac{\partial^n}{\partial u_1 \cdots \partial u_n} C(u_1,\dots,u_n)\right) \frac{\mathrm{d}u_1}{\mathrm{d}x_1} \cdots \frac{\mathrm{d}u_n}{\mathrm{d}x_n} \\ f_{1,\dots,n}(x_1,\dots,x_n) &= c_{1,\dots,n}(F_1(x_1),\dots,F_n(x_n)) f_1(x_1) \cdots f_n(x_n)\end{aligned} \tag{4.10}$$

dove $c_{1,\dots,n}$ è la densità di copula associata a $F_{1,\dots,n}$ e $u_i = F_i(x_i)$.

Nel caso di due variabili X_i e X_j, usando l'equazione (4.10) nella definizione della densità congiunta $f_{i|j}$ si ricava:

$$f_{i|j}(x_i|x_j) = c_{ij}(F(x_i), F(x_j)) f_i(x_i) \tag{4.11}$$

ovvero si può scrivere la densità condizionata in funzione della densità di copula.

Si consideri quindi il caso $n = 3$ e l'equazione (4.3). Usando la equazione (4.11), si può sostituire $f_{2|1}$ con $c_{1,2}(F(x_1), F(x_2)) f_2(x_2)$. Mentre, per quanto riguarda $f_{3|1,2}$, si osserva che:

$$f_{3|1,2}(x_3|x_1,x_2) = \frac{f_{1,3|2}(x_1,x_3|x_2)}{f_{1|2}(x_1|x_2)} \tag{4.12}$$

e al contempo, per la equazione (4.10), si ha che:

$$f_{1,3|2}(x_1,x_3|x_2) = c_{1,3;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) f_{1|2}(x_1|x_2) f_{3|2}(x_3|x_2)\,, \tag{4.13}$$

dove, per ogni x_2, la funzione $c_{1,3;2}(u,v;x_2)$ è la funzione di densità della copula $C_{1,3;x_2}$ associata alla funzione di ripartizione condizionata $F_{1,3|2}(x_1,x_3|x_2)$. In altre parole, $c_{1,3;2}$ è una funzione in $\mathbb{I}^2 \times \mathrm{Ran}(Y)$ o equivalentemente è una famiglia di funzioni indicizzata da $\mathrm{Ran}(Y)$. Inoltre, $c_{1,3;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2)$ dipende da x_2 anche attraverso i suoi argomenti $F_{1|2}(x_1|x_2)$ e $F_{3|2}(x_3|x_2)$. Questo aspetto verrà ripreso meglio nelle sezioni 4.3 e 4.4 dove si vedrà come viene ovviato il problema nella teoria della *PCC* e le conseguenze di tale ipotesi semplificante.

Ritornando al nostro esempio, si ricava che:

$$\begin{aligned} f_{3|1,2}(x_3|x_1,x_2) &= c_{1,3;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) f_{3|2}(x_3|x_2) && (4.14)\\ &= c_{1,3;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{2,3}(F_2(x_2), F_3(x_3)) f_3(x_3) && (4.15)\end{aligned}$$

Dove usando equazione (4.11) in $f_{3|2}(x_3|x_2)$ per ricavare la scomposizione finale.

Si osserva inoltre che usando un approccio simile si sarebbe potuto arrivare alla decomposizione:

$$f_{3|1,2}(x_3|x_1,x_2) = c_{2,3;1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1); x_1) c_{1,3}(F_1(x_1), F_3(x_3)) f_3(x_3) \tag{4.16}$$

Sarebbe infatti stato sufficiente usare $f_{2,3|1}(x_2,x_3|x_1)$ invece che $f_{1,3|2}(x_1,x_3|x_2)$.

Riunendo i risultati ottenuti si può riscrivere la equazione (4.3) come:

$$\begin{aligned} f_{1,2,3}(x_1,x_2,x_3) &= f_1(x_1)f_{2|1}(x_2|x_1)f_{3|1,2}(x_3|x_1,x_2) \\ &= f_1(x_1)f_2(x_2)f_3(x_3) \\ &\quad \cdot c_{1,2}(F(x_1),F(x_2))c_{2,3}(F_2(x_2),F_3(x_3)) \\ &\quad \cdot c_{1,3;2}\big(F_{1|2}(x_1|x_2),F_{3|2}(x_3|x_2);x_2\big) \end{aligned} \tag{4.17}$$

Usando un procedimento simile è possibile costruire una simile decomposizione per un n qualsiasi. In particolare, vale la seguente proposizione.

Proposizione 4.3. *Sia $f_{1,\dots,n}$ la funzione di densità congiunta delle n variabili aleatorie $\{X_i\}_{1\leq i\leq n}$. Allora, omettendo gli argomenti delle funzioni per semplificare le formule, risulta che $f_{1|2} = c_{1,2}\cdot f_1$ e che, per ogni $k\geq 3$:*

$$f_{k|1,\dots,k-1} = f_k \cdot c_{k-1,k} \cdot \left[\prod_{l=1}^{k-2} c_{l,k;l+1,\dots,k-1}\right] \tag{4.18}$$

Si è già visto la formula per $f_{3|1,2}$ e come quella formula non sia unica.

Per comprendere meglio a cosa corrisponde il prodotto è utile scrivere la equazione (4.18) per $k=4$ e $k=5$:

$$f_{4|1,2,3} = f_4 \cdot c_{3,4} \cdot c_{2,4;3} \cdot c_{1,4;2,3} \tag{4.19}$$

$$f_{5|1,2,3,4} = f_5 \cdot c_{4,5} \cdot c_{3,5;4} \cdot c_{2,5;3,4} \cdot c_{1,5;2,3,4} \tag{4.20}$$

$$\tag{4.21}$$

Il *pattern* continua per k crescenti, infatti è possibile esprimere la densità $f_{1,\dots,n}$ facendo uso dell'equazione (4.18) e la proposizione 4.1 nella seguente forma:

$$f_{1,\dots,n} = \left[\prod_{k=1}^{n} f_k\right] \cdot \left[\prod_{k=2}^{n} c_{k-1,k}\right] \cdot \left[\prod_{k=3}^{n}\prod_{l=1}^{k-2} c_{l,k;l+1,\dots,k-1}\right] \tag{4.22}$$

Questa fattorizzazione mostra come sia sempre possibile esprimere una distribuzione multivariata come prodotto di *pair-copule* eventualmente condizionate. D'altra parte, come si ha avuto modo di osservare, questa decomposizione non è unica. Rispetto alla equazione (4.2), in cui è facile osservare che permutazioni diverse producono decomposizioni diverse, per le *PCC* non è così semplice. Nella sezioni 4.2 e 4.3 si osserverà come sia possibile descrivere ed enumerare la *PCC* usando la teoria dei grafi.

4.2 Le *regular vine*

Al fine di comprendere e analizzare le decomposizioni in *pair-copule*, risulta utile rappresentare queste strutture attraverso *grafi*[1]. A questo scopo, Bedford e Cooke (2002) hanno introdotto i concetti di *regular vine*, *regular vine distribution* e *regular vine copula*. Questa sezione si concentrerà sulle *regular vine*, mentre si analizzaranno altri due concetti nella sezione 4.3.

[1] Ovvero attraverso un approccio grafico.

Le *regular vine* sono un concetto puramente di teoria dei grafi, senza alcuna componente stocastica, e per comprenderle al meglio si dedicherà la sezione 4.2.1 all'introduzione dei suoi prerequisiti teorici. Al fine di utilizzare una notazione il più possibile *standard*, si è deciso di utilizzare quella di Diestel (2017). Si rimanda a tale trattazione per ulteriori approfondimenti. Si noti inoltre che Bedford e Cooke (2002) usano una notazione leggermente diversa, ma le differenze non sono sostanziali.

4.2.1 Nozioni di Teoria dei Grafi

Seguendo l'impostazione di Diestel (2017), tutti i grafi in questa tesi saranno semplici, ovvero senza cappi e archi multipli tra i vertici.

Dato un insieme V, si userà la notazione $[V]^k \subset \wp(V)$ per indicare i sottoinsiemi di k elementi di V. Gli insiemi di k elementi saranno anche detti k-insiemi e gli elementi di $[V]^k \subset \wp(V)$ sono anche detti k-sottoinsiemi. In particolare, $[V]^2 \subset \wp(V)$ è l'insieme coppie (non ordinate) di elementi di V.

Definizione 4.1. *Un* grafo *è una coppia di insiemi* (V, E) *tale che* $E \subseteq [V]^2$*. Gli elementi di* V *sono detti* vertici *e gli elementi di* E *sono detti* archi*.*

Per evitare problemi, si assume che $V \cap E = \emptyset$. I vertici di un grafo sono anche chiamati *nodi* o *punti*, e i suoi archi sono alle volte chiamati *lati*.

Dato un grafo G, è usuale segnare con $V(G)$ l'insieme dei suoi vertici e con $E(G)$ l'insieme dei suoi archi, indipendentemente da come questi insiemi siano stati definiti in precedenza. Se $V = V(G)$ si dirà che G è un grafo *su* V.

Definizione 4.2. *L'*ordine *di un grafo* G *è il numero dei suoi vertici e si denota con* $|G|$*. Il numero dei suoi archi prende il nome di* dimensione *del grafo* G *e viene denotato con* $\|G\|$*. Un grafo* G *è* finito *se* $|G|$ *è finito.*

Nella presente tesi ogni grafo è finito quindi, nel seguito, quando si parlerà di grafi si intenderanno sempre grafi finiti.

Il grafo $(\emptyset, \emptyset)$ prende il nome di *grafo nullo* ed è segnato con $\emptyset$. I grafi di ordine 0 e 1 sono detti *banali*. Un grafo G in cui $E(G) = [V(G)]^2$ viene detto *grafo completo*.

Si forniscono ora alcune definizioni su archi e vertici:

Definizione 4.3. *Un arco* $e \in E(G)$ *di dice* incidente *ad un vertice* $v \in V(G)$ *se* $v \in e$*. L'insieme degli archi incidenti ad un vertice* $v \in V(G)$ *si denota con* $E(v)$*.*

Definizione 4.4. *Si chiama* grado $d(v)$ *di un vertice* $v \in V(G)$ *il numero degli archi incidenti a* v*, ovvero* $d(v) = |E(v)|$*. Un vertice di grado* 0 *viene detto* isolato*.*

Definizione 4.5. *Dato un arco* $e \in E(G)$*, i due vertici* $u, v \in V(G)$ *tali che* $e = \{u, v\}$ *si dicono* estremi *di* e*. Si dice inoltre che* e giunge u *e* v*.*

Definizione 4.6. *Due archi* $e, f \in E(G)$ *si dicono* adiacenti*, o* vicini*, se sono incidenti ad uno stesso vertice. Due archi non adiacenti di dicono* indipendenti*. Similmente si dice che un sottoinsieme* $E' \subseteq E(G)$ *è* indipendente *se non contiene coppie di archi adiacenti.*

Definizione 4.7. *Due vertici $u, v \in V(G)$ si dicono* adiacenti *se sono gli estremi di un qualche arco di G. Due vertici non adiacenti di dicono* indipendenti*. Similmente si dice che un sottoinsieme $V' \subseteq V(G)$ è* indipendente *se non contiene coppie di vertici adiacenti.*

Per non creare confusione con la notazione delle *regular vine*, si userà la notazione $\mathrm{Adj}_G(v)$ per indicare l'insieme dei vertici adiacenti ad $v \in V(G)$ invece che la notazione usata da Diestel (2017). Si ometterà il riferimento al grafo G quando è chiaro dal contesto.

Per l'ipotesi di semplicità che è stata fatta all'inizio, $d(v) = |E(v)| = |\mathrm{Adj}(v)|$.

Proposizione 4.4. *Sia G un grafo, allora $\|G\| = \frac{1}{2}\sum_{v \in V(G)} d(v)$.*

Unione e intersezione di grafi si fa facendo l'unione e l'intersezione degli insiemi dei vertici e degli archi. Ovvero:

$$V(G \cup G') = V(G) \cup V(G') \tag{4.23}$$
$$V(G \cap G') = V(G) \cap V(G') \tag{4.24}$$
$$E(G \cup G') = E(G) \cup E(G') \tag{4.25}$$
$$E(G \cap G') = E(G) \cap E(G') \tag{4.26}$$

Se $G \cap G' = \emptyset$, allora si dirà che i due grafi sono *disgiunti*.

Definizione 4.8. *Un grafo G' si dice* sottografo *di un grafo G se $V(G') \subseteq V(G)$ e $E(G') \subseteq E(G)$. Il sottografo G' di G si dice* proprio *se $G \neq G'$.*

Se G' è un sottografo di G, si dice anche che G *contiene* G' e si scrive $G' \subseteq G$. Si noti che $G' \subseteq G$ se e solo se $G' \cap G = G'$.

Definizione 4.9. *Un sottografo $G' \subset G$ si dice* sottografo indotto *se per ogni $v, f \in V(G')$, $\{u, v\} \in E(G')$ se e solo se $\{u, v\} \in E(G)$.*

Un sottografo indotto G' è quindi completamente definito da G e dall'insieme $V(G')$. Se $U \subseteq V(G)$, si userà la notazione $G[U]$ per indicare il sottografo indotto generato da U. Se $G' \subset G$ è comune usare la notazione $G[G']$ al posto di $G[V(G')]$.

Per ogni $U \subset V(G)$ e $G' \subset G$ si usa inoltre le notazioni $G - U$ e $G - G'$ per indicare i sottografi $G[V(G) - U]$ e $G[V(G) - V(G')]$. Quando $U = \{v\}$ per un qualche vertice $v \in V(G)$, si ometteranno le parentesi graffe, ovvero si scriverà $G - v = G - \{v\}$.

Se $F \subseteq [V(G)]^2$, si usano le notazione $G - F$ e $G + F$ per indicare i grafi $(V(G), E(G) - F)$ e $(V(G), E(G) \cup F)$. Quando $F = \{e\}$, è comune omettere le parentesi.

Definizione 4.10. *Un grafo $G = (V, E)$ è detto* arco-massimale *rispetto ad una particolare proprietà se G soddisfa quella proprietà ma nessun grafo (V, F) con $F \supsetneq E$ soddisfa quella proprietà.*

Un sottografo $G' \subseteq G$ è detto massimale/minimale *rispetto ad una particolare proprietà se lo è rispetto alla relazione di sottografo.*

Definizione 4.11. *Sia $G = (V, E)$ un grafo, il* grafo di linea $L(G)$ *di G è il grafo su E tale che $\{e, f\} \in E(L(G))$ se $e \cap f = \{v\}$ per un qualche $v \in V$, ovvero se sono adiacenti in G.*

Definizione 4.12. *Sia $U = (v_0, \ldots, v_n)$ una sequenza ordinata di $n + 1$ vertici di un grafo G tali che $e_i = \{v_i, v_{i+1}\} \in E(G)$ per ogni $0 \leq i < n$. Il sottografo $P = (\{v_i\}_{1 \leq i \leq n}, \{e_i\}_{1 \leq i < n}) \subset G$ prende il nome di* percorso *di lunghezza n su G. I vertici v_0 e v_n sono detti* estremi *del percorso P. I vertici $v_1, \ldots, v_{n-1}$ sono detti vertici* interni *del percorso.*

Un cammino *è un percorso in cui tutti i vertici sono distinti. Un percorso in cui gli unici vertici ripetuti sono gli estremi viene chiamato* ciclo.

Dato un percorso tra due vertici distinti, è sempre possibile costruire un cammino tra i suoi due estremi. Similmente, dato un percorso da un vertice in sé stesso, è sempre possibile estrarne un ciclo.

Il concetto di cammino su un grafo G permette di definire la relazione di connessione di un grafo.

Definizione 4.13. *Sia G un grafo, due vertici $v, w \in V(G)$ si dicono* connessi *se esiste un cammino tra di loro. Due vertici non connessi si dicono* sconnessi. *La relazione di connessione è una relazione di equivalenza; le sue classi di equivalenza si dicono* componenti connesse *del grafo.*

Un grafo G in cui tutti i vertici sono connessi si dice connesso. *Un sottografo è connesso se lo è come grafo. Un grafo non connesso si dice* sconnesso.

Si è quindi pronti a definire il concetto di albero.

Definizione 4.14. *Una* foresta *o* grafo aciclico *è un grafo G che non possiede cicli. Una foresta si dice* albero *se è connesso. Un sottografo connesso di un albero prende il nome di* sottoalbero.

I vertici di un albero si dicono foglie *se hanno un solo arco incidente, ovvero se hanno grado* 1.

Il seguente teorema mostra le varie definizioni equivalenti per il concetto di albero.

Teorema 4.1. *Le seguenti proprietà sono equivalenti per un grafo (finito) T:*

1. *T è un albero, ovvero è aciclico e connesso;*
2. *T è connesso e $\|T\| = |T| - 1$;*
3. *T è aciclico e $\|T\| = |T| - 1$;*
4. *T è connesso e $T - e$ è sconnesso per ogni $e \in E(T)$;*
5. *T è connesso e $T - v$ è sconnesso per ogni $v \in V(T)$;*
6. *ogni due vertici $v, w \in V(T)$ sono connessi da un unico cammino;*
7. *T è aciclico e $T + e$ contiene cicli per ogni $e \in [V(T)]^2 - E(T)$.*

Le seguenti proposizioni discendono direttamente dalla definizione di albero:

Proposizione 4.5. *Sia T un albero e $U \subset V(T)$, allora $T[U]$ è un sottoalbero se e solo se è connesso.*

Proposizione 4.6. *Ogni sottoalbero di un albero T è indotto dai suoi vertici.*

Proposizione 4.7. *Due sottoalberi di un albero T sugli stessi vertici coincidono.*

Si useranno inoltre queste due tipologie di alberi:

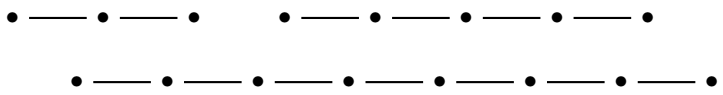


Figura 4.1. Qualche esempio di albero a cammino.

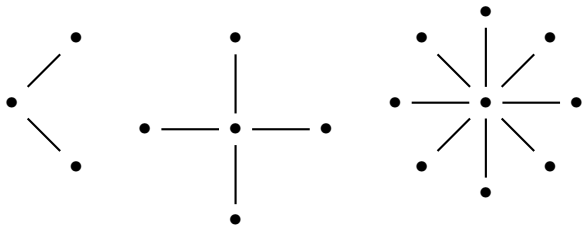


Figura 4.2. Qualche esempio di albero a stella.

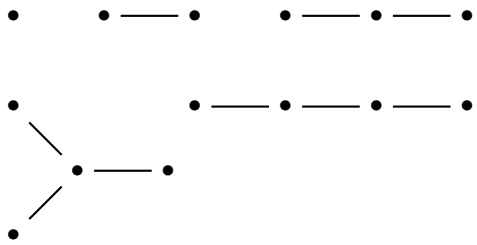


Figura 4.3. Le tipologie di albero fino a quattro vertici.

Definizione 4.15. *Un albero T viene detto* a cammino *se $d(v) \leq 2$ per ogni $v \in V(T)$.*

Definizione 4.16. *Un albero T viene detto* a stella *se possiede un vertice $v \in V(T)$ con $d(v) = \|T\|$.*

Proposizione 4.8. *Un albero T a cammino con $|T| \geq 2$, possiede due foglie distinte $u, v \in V(T)$ e il cammino tra di loro coincide con T.*

Proposizione 4.9. *In un albero T a stella tutti i vertici tranne uno sono foglie.*

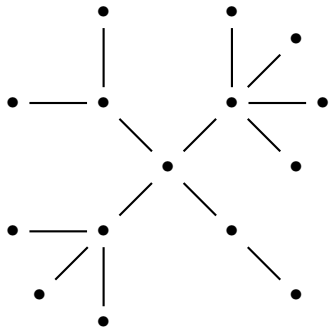


Figura 4.4. Un esempio di albero che non è né a stella né a cammino.

La figura 4.1 contiene qualche esempio di alberi a cammino, e nella figura 4.2 si propongono alcuni esempi di alberi a stella. I nomi dei nodi sono stati omessi per meglio evidenziare in cosa consistano le due tipologie di alberi.

Osservazione 4.1. *Come si può vedere dalla figura 4.3, ogni albero con meno di quattro vertici è sia a cammino che a stella. Gli alberi con* 4 *vertici sono o a stella o a cammino, mentre dai* 5 *vertici in su, possono esistere grafi che non appartengono a nessuna delle due categorie. Un esempio con* 15 *vertici è fornito nella figura 4.4.*

Un risultato sugli alberi a cammino e a stella che risulterà utile nelle prossime sezioni è il seguente:

Proposizione 4.10. *Sia T_c un albero a cammino e T_s un albero a stella. Allora $L(T_c)$ è un albero a cammino e $L(T_s)$ è un grafo completo.*

Non è difficile osservare che ogni albero possa essere "composto" come unione di alberi a stella ed alberi a cammino. Non dovrebbe quindi stupire troppo il seguente:

Proposizione 4.11. *Sia G un grafo tale che $L(G)$ è un albero, allora sia G che $L(G)$ sono alberi a cammino.*

4.2.2 La definizione di *regular vine*

Ora che sono state introdotte le necessarie nozioni sui grafi si è pronti a introdurre le *regular vine*. Nella trattazione di Bedford e Cooke (2002) le *regular vine* vengono presentate come una sottoclasse di una classe di oggetti più ampia chiamate *vine*. Siccome le *vine* non sono utilizzate nella pratica, similmente a quanto fatto in Czado (2019), è stato deciso di semplificare la trattazione e introdurre direttamente le *regular vine* o *R-vine*.

Definizione 4.17. *Una* regular vine *(*R-vine *nel seguito) $\mathcal{V}$ su n elementi è una successione di n alberi $\mathcal{V} = (T_1, \ldots, T_n)$, tali che:*

1. *per ogni $1 \leq i \leq n$, $T_i = (V_i, E_i)$;*
2. *$V_1 = \{1, \ldots, n\}$ è l'insieme degli interi da* 1 *a n compresi;*

3. *per ogni* $2 \leq i \leq n$, $V_i = E_{i-1}$, *ovvero i vertici di un albero* T_i *sono gli archi dell'albero precedente* T_{i-1};

4. *(*condizione di prossimità*) per ogni* $2 \leq i \leq n$, *se* $\{e_1, e_2\} \in E_i$ *allora* $e_1 \cap e_2 \neq \emptyset$, *ovvero* e_1 *e* e_2 *sono adiacenti come archi di* T_{i-1}.

Prima di analizzare meglio questo oggetto, risulta utile definire due sottoclassi delle *R-vine*:

Definizione 4.18. *Una* D-vine, *o* drawable vine, *è una* R-vine *in cui* T_i *è a cammino per ogni* $1 \leq i \leq n$.

Definizione 4.19. *Una* C-vine, *o* canonical vine, *è una* R-vine *in cui* T_i *è a stella per ogni* $1 \leq i \leq n$.

Osservazione 4.2. *La condizione 4 di prossimità è equivalente ad affermare che, per ogni* $2 \leq i \leq n$, $T_i \subset L(T_{i-1})$. *Come si è visto nella proposizione 4.11, il grafo di linea* $L(T_{i-1})$ *è raramente un albero e quindi l'inclusione è generalmente propria. Nella proposizione 4.10 si è osservato che il grafo di linea di un albero a cammino è ancora un albero a cammino mentre il grafo di linea di un albero a stella è il grafo completo. In genere, il grafo di linea di un albero è una via di mezzo tra questi due estremi.*

Il fatto che il grafo di linea di un albero a cammino è ancora un albero a cammino ha conseguenze nella costruzione e nella enumerazione delle D-vine. *Infatti, la condizione 4 di prossimità e la proposizione 4.10 forzano la scelta di tutti gli alberi dopo il primo. Si analizzerà questo aspetto più un dettaglio nella sezione 4.2.4.*

È importante osservare che ogni *R-vine* $\mathcal{V}$ è completamente determinata dai suoi archi. Infatti V_1 è fissato e $V_i = E_{i-i}$ per gli altri alberi. Per semplificare alcune definizioni in questa sezione, si si pone $E_0 = V_1 = \{1, \ldots, n\}$.

Dato che ogni T_i è un albero e $V_i = E_{i-1}$, si ha che $|V_1| = n$ e che $|V_i| = |V_{i-1}| - 1$. Pertanto, $|T_i| = |V_i| = n + 1 - i$ e $\|T_i\| = |E_i| = n - i$. Risulta pertanto evidente che T_n è un albero di un singolo vertice e nessun arco[2].

La condizione 4 di prossimità fa si che ogni cammino in T_i sia associato ad un cammino in T_{i-1} ed in generale ogni sottoalbero di T_i è associato ad un sottoalbero di T_{i-1}. Non è invece vera la relazione opposta, infatti il sottografo di T_i indotto dagli archi di un sottoalbero di T_{i-1} potrebbe non essere connesso.

Sia $e \in E_i$ per un qualche $2 \leq i \leq n$, allora $T_i[e] = (e, \{e\})$ è banalmente un sottoalbero di T_i. Per le considerazioni appena fatte, a questo sottoalbero ne sono connessi altri in $T_{i-1}, \ldots, T_1$. Al fine di studiare questi alberi, si definiscono i seguenti insiemi.

Definizione 4.20. *L'*unione j-esima U_e^j *di* e, *con* $0 \leq j \leq i$, *è l'insieme definito come:*

- *se* $j = 0$, *allora* $U_e^0 = \{e\}$;

[2]Per questa ragione non tutte le fonti che sono state consultate contengono quest'ultimo albero. Per esempio, Czado (2019) considerano solo $n - 1$ alberi, mentre Bedford e Cooke (2002) ne considerano n.

- *altrimenti* $U_e^j = \{a \in E_{i-j} : \exists b \in U_e^{j-1},\, a \in b\}$.

Come si è notato, per ogni $e \in E_i$ e $0 \leq j \leq i$, $(U_e^{j+1}, U_e^j) \subset T_{i-j}$ è un sottoalbero. L'insieme di questi alberi ha una struttura molto simile ad una *R-vine*. Infatti l'unica condizione che non soddisfa è la condizione sui vertici in T_1. Il fatto che siano sottoalberi, implica la seguente proposizione:

Proposizione 4.12. *Sia $\mathcal{V}$ una R-*vine *ed $e \in E_i$ per un qualche $1 \leq i < n$, allora $|U_e^j| = j + 1$ per ogni $0 \leq j \leq i$.*

Nella definizione ricorsiva di U_e^j, l'insieme dipende solo dagli elementi di U_e^{j-1} e non direttamente da $e \in E_i$. Pertanto, è facile mostrare che se $x \in U_e^j$ per qualche $e \in E_i$ e $0 \leq j < i$, allora $U_x^k \subset U_e^{j+k}$ per ogni $0 \leq k \leq i - j$. Inoltre, $U_e^{j+k} = \bigcup_{x \in U_e^j} U_x^k$. In particolare, se $e = \{x, y\} \in E_i$ per $1 \leq i < n$ allora $U_e^j = U_x^{j-1} \cup U_y^{j-1}$.

Quest'ultima caratteristica consente di definire $U_e^i \subset E_0$ in maniera più semplice ed intuitiva:

Definizione 4.21. *L'*unione completa *A_e di e è l'insieme definito nel seguente modo:*

- *se $i \leq 1$, allora $A_e = e$;*
- *se $1 < i \leq n$, allora $A_e = A_x \cup A_y$ dove x e y sono gli estremi di e.*

Proposizione 4.13. *Sia $\mathcal{V}$ una R-*vine *ed $e \in E_i$ per un qualche $1 \leq i \leq n - 1$, allora $U_e^i = A_e$.*

Questo insieme inoltre caratterizza completamente e. Vale infatti la seguente proposizione:

Proposizione 4.14. *Sia $\mathcal{V}$ una R-*vine *ed $e, f \in E_i$ per un qualche $1 \leq i < n$, allora $A_f = A_e$ se e solo se $e = f$.*

L'unione completa è alla base dei seguenti insiemi:

Definizione 4.22. *Sia $e = \{x, y\} \in E_i$ per $1 \leq i < n$, l'*insieme condizionante *di e è l'insieme:*

$$\mathcal{D}_e = A_x \cap A_y$$

*L'*insieme condizionato *di e è l'insieme:*

$$\mathcal{C}_e = A_x \triangle A_y = (A_x - \mathcal{D}_e) \cup (A_y - \mathcal{D}_e)$$

Si userà inoltre la notazione $\mathcal{C}_{e,x} = A_x - \mathcal{D}_e$ e $\mathcal{C}_{e,y} = A_y - \mathcal{D}_e$.

Proposizione 4.15. *Sia $e = \{x, y\} \in E_i$ per $1 \leq i < n$, allora $A_e = \mathcal{C}_{e,x} \cup \mathcal{C}_{e,y} \cup \mathcal{D}_e$.*

Si noti che la decomposizione espressa nella proposizione 4.15 è in insiemi disgiunti, ovvero $\{\mathcal{C}_{e,x}, \mathcal{C}_{e,y}, \mathcal{D}_e\}$ è una partizione di A_e.

Proposizione 4.16. *Sia $\mathcal{V} = (T_1, \ldots, T_{n-1})$ una R-*vine *ed $e = \{x, y\} \in E_i$ per un qualche $1 \leq i < n$, allora:*

1. $|A_e| = i + 1$;
2. $|\mathcal{C}_{e,x}| = |\mathcal{C}_{e,y}| = 1$;
3. $|\mathcal{D}_e| = i - 1$.

Usando la decomposizione definita dalla proposizione 4.15 e la proposizione 4.14 è possibile rappresentare un arco $e \in E_i$ con la tripletta di insiemi $(\mathcal{C}_{e,x}, \mathcal{C}_{e,y}, \mathcal{D}_e)$.

Definizione 4.23. *Sia $\mathcal{V}$ una* R-vine*, l'insieme:*

$$\mathcal{CV} = \{(\mathcal{C}_{e,x}, \mathcal{C}_{e,y}, \mathcal{D}_e) : e = \{x, y\} \in E_i,\ 1 \leq i < n\}$$

prende il nome di insieme dei vincoli *di $\mathcal{V}$.*

Siccome l'insieme dei vincoli $\mathcal{CV}$ di $\mathcal{V}$ permette di derivare ogni arco di $\mathcal{V}$ e una *R-vine* è completamente determinata dai suoi archi, allora risulta banalmente la seguente proposizione.

Proposizione 4.17. *Se due* R-vine *$\mathcal{V}_1$ e $\mathcal{V}_2$ possiedono lo stesso insieme di vincoli $\mathcal{CV} = \mathcal{CV}_1 = \mathcal{CV}_2$ allora $\mathcal{V}_1 = \mathcal{V}_2$.*

Per semplificare la notazione, è usuale scrivere il vincolo generico $(\mathcal{C}_{e,x}, \mathcal{C}_{e,y}, \mathcal{D}_e)$ come $(a_e, b_e; \mathcal{D}_e)$ (si veda per esempio Czado (2019)). Ovvero, quando si vorrà rappresentare un arco $e \in E_i$ attraverso il suo vincolo, si userà la notazione $e = (a_e, b_e; \mathcal{D}_e)$. Quando a_e, b_e e $\mathcal{D}_e$ saranno espressi con i loro elementi si ometteranno le parentesi graffe. Per esempio, se $a_e = \{3\}$, $b_e = \{7\}$ e $\mathcal{D}_e = \{1, 4, 8\}$, si scriverà che $e = (3, 7; 1, 4, 8)$.

Si noti che fissato una *R-vine* $\mathcal{V}$, l'insieme $\{a_e, b_e\}$ identifica univocamente l'arco $e \in E(\mathcal{V})$, ovvero non esistono due archi distinti e_1 e e_2 tali che $\{a_{e_1}, b_{e_1}\} = \{a_{e_2}, b_{e_2}\}$.

4.2.3 Un esempio

Si consideri una generica *R-vine* $\mathcal{V} = (T_1, T_2, T_3, T_4, T_5)$ su 5 oggetti. Per definizione, $V_1 = E_0 = \{1, 2, 3, 4, 5\}$ mentre l'unica condizione su E_1 è che T_1 sia un albero. Siccome lo si può scegliere in maniera arbitraria, si consideri l'albero definito da: $E_1 = \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{4, 5\}\}$ ovvero l'albero rappresentato in figura 4.5. Quest'albero non è né a cammino né a stella.

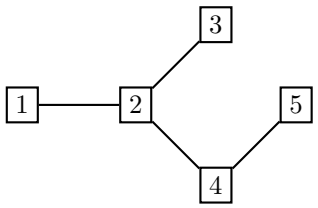


Figura 4.5. L'albero T_1 dell'esempio della sezione 4.2.3.

Si può passare quindi a T_2. È noto che $V_2 = E_1$ e che $T_2 \subset L(T_1)$. In altre parole, gli archi in E_2 sono coppie di elementi di E_1. Pertanto, avranno la forma

$\{\{u,v\},\{v,w\}\} = (u,w;v)$ per qualche $u,v,w \in V_1$, con $\{u,v\},\{v,w\} \in E_1$. Si sa inoltre che T_2 è un albero connesso. Il grafo di linea $L(T_1)$ è rappresentato in figura 4.6.

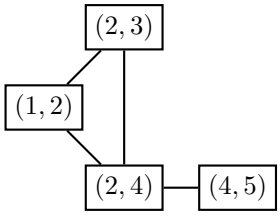


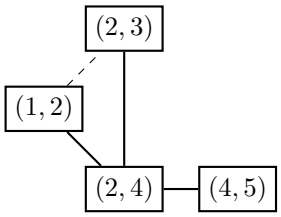
Figura 4.6. Il grafo di linea $L(T_1)$ dell'albero T_1.

Si procederà elemento per elemento. Si deve avere necessariamente $(2,5;4) = \{\{2,4\},\{4,5\}\} \in E(T_2)$ perché $\{2,4\}$ è l'unico arco adiacente a $\{4,5\}$. Siccome il sottografo generato da $\{\{1,2\},\{2,3\},\{2,4\}\}$ è un grafo completo bisognerà fare una scelta. In questo caso sarà sufficiente scartare un singolo arco di $L(T_1)$ per avere un albero.

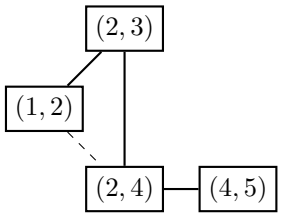
Si hanno dunque 3 possibili scelte per E_2:

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

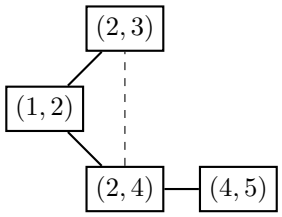
Nella figura 4.7 sono stati rappresentati i tre possibili T_2. L'arco di $L(T_1)$ che è stato "rimosso" è stato rappresentato con una linea tratteggiata.



(a) L'albero T_2 con E_2 definito come nella opzione 1.



(b) L'albero T_2 con E_2 definito come nella opzione 2.



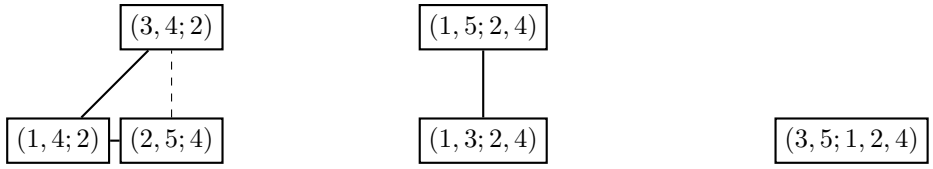
(c) L'albero T_2 con E_2 definito come nella opzione 3.

Figura 4.7. I possibili alberi per T_2.

Com'è semplice osservare, figura 4.7a è un albero a stella, mentre figure 4.7b e 4.7c sono a cammino. Come è stato già accennato nella osservazione 4.2, la struttura a cammino di un albero determina univocamente ogni albero successivo. Pertanto, se si scegliesse l'opzione 3 per E_2 (figura 4.7c), allora si avrà necessariamente $E_3 = \{(3,4;1,2),(1,5;2,4)\}$ ed $E_4 = \{(3,5;1,2,4)\}$. L'opzione 2, rappresentata in figura 4.7b, è simile.

Al contrario, l'opzione 1 offre più scelta. Infatti, ci troveremmo esattamente nella stessa situazione di E_2, con tre possibili E_3: uno per ogni arco che "si rimuove". La

figura 4.8 mostra una delle tre possibili scelte per i restanti alberi. Nella figura 4.8a è stato segnato con un linea tratteggiata l'arco rimosso da $L(T_2)$.



(a) T_3 **(b)** T_4 **(c)** T_5

Figura 4.8. Un esempio dei restanti 3 alberi di $\mathcal{V}$.

4.2.4 Enumerazione delle *R-vine*

n	# *C-vine*/*D-vine*	# *R-vine*
2	1	1
3	3	3
4	12	24
5	60	480
6	360	23040
7	2520	2580480
8	20160	660602880
9	181440	380507258880
10	1814400	487049291366400
11	19958400	1371530804487782400
12	239500800	842668526272935065600
13	3113510400	11217603421803331159326720 0
14	43589145600	3216311253099451110002157158400
15	653837184000	197610163390430276198532535812096000
16	10461394944000	$2.4758800785707605 \times 10^{27}$
17	177843714048000	$4.056481920730334 \times 10^{31}$
18	3201186852864000	$1.3292279957849159 \times 10^{36}$
19	60822550204416000	$8.711228593176025 \times 10^{40}$
20	1216451004088320000	$1.141798154164768 \times 10^{46}$
21	25545471085854720000	$2.9931553532536893 \times 10^{51}$
22	562000363888803840000	$1.5692754338466702 \times 10^{57}$
23	12926008369442488320000	$1.645504557321206 \times 10^{63}$
24	310224200866619719680000	$3.450873173395282 \times 10^{69}$
25	7755605021665492992000000	$1.4474011154664525 \times 10^{76}$
26	201645730563302817792000000	$1.2141680576410807 \times 10^{83}$
27	5444434725209176080384000000	$2.037035976334486 \times 10^{90}$
28	152444172305856930250752000000	$6.835158514946912 \times 10^{97}$
29	4420880996869850977271808000000	$4.586997231980143 \times 10^{105}$
30	132626429906095529318154240000000	$6.156563468186638 \times 10^{113}$

Tabella 4.1. Enumerazione delle *C-vine*, *D-vine* e *R-vine*.

Nella sezione 4.3 si introdurrà il concetto di *pair-copula construction* e si vedrà come le *R-vine* siano usate per rappresentarne la struttura. Siccome *R-vine* diverse porteranno a modelli differenti, è utile sapere quante siano le *R-vine* differenti.

Per n piccoli è facile generare ogni possibile *R-vine*. Per $n = 2$ vi è una sola *vine*, per $n = 3$ ogniuna delle 6 *R-vine* è sia una *C-vine* che una *D-vine*. Per $n = 4$, tutte le *R-vine* sono o *C-vine* o *D-vine*: vi sono 12 *C-vine* e 12 *D-vine* per un totale di 24 *R-vine*. Oltre questo valore cominciano ad esserci *R-vine* che non sono né *C-vine* che *D-vine* e diventa molto complicato elencarle tutte senza un *computer*.

Una *R-vine* su n elementi è in buona sostanza una successione di n alberi in cui l'ordine degli alberi parte da n e decresce fino a 1 all'ultimo albero. Cayley nel 1889[3] ha dimostrato che il numero di alberi di ordine n è n^{n-2}. Siccome la scelta del primo albero è arbitraria, è evidente che il numero di *R-vine* è almeno n^{n-2} e può essere al massimo $\prod_{i=2}^{n} i^{i-2}$ (ovvero il numero di successioni di n alberi di ordine descente[4]). Fortunatamente la condizione di prossimità assicura che il numero non sia così grande. Morales Napoles (2010)[5] ha dimostrato che vi sono:

$$\frac{n!}{2} \times 2^{\binom{n-2}{2}} \tag{4.27}$$

R-vine su n elementi.

Il numero di *R-vine* è una delle ragioni per cui molta teoria si è concentrata su *D-vine* e *C-vine*. Il loro numero è infatti molto più maneggevole e per calcolarlo non è necessario usare complessi strumenti matematici.

Siccome il grafo di linea di un cammino è ancora un cammino, una *D-vine* è completamente determinata dal suo primo albero. In altre parole, una *D-vine* è univocamente determinata da una permutazione degli elementi da 1 a n in cui il primo elemento sia minore dell'ultimo (questa condizione è necessaria perché il grafo non è orientato e quindi il cammino inverso definisce la stessa *D-vine*). Siccome il numero di permutazioni di n elementi è $n!$, ci sono $\frac{n!}{2}$ *D-vine*.

Nelle *C-vine*, ogni albero ha un vertice centrale e l'elenco di questi vertici definisce completamente la *C-vine*. Non è difficile calcolare che anche le *C-vine* su n elementi sono $\frac{n!}{2}$, come le *D-vine*. Un pò più difficile è invece notare che ogni albero "elimina" un elemento tra 1 e n e che quindi sia possibile definire anche le *C-vine* usando una permutazione degli elementi tra 1 e n (in questo caso è l'ordine degli ultimi due elementi a non contare).

La tabella 4.1 contiene il numero di *R-vine*, *D-vine* e *C-vine* al variare di n. Al fine di comprendere quanto grandi siano quei valori è utile compararli con il numero di atomi nell'universo, che è stato stimato essere tra i 10^{79} e i 10^{82}. Dalla tabella si vede che è sufficente $n = 26$ per superare quel valore (la riga evidenziata in giallo). Per le *D-vine* e *C-vine* ci vuole un po' di più, ma ci riescono per $n = 62$.

Questi valori fanno sì che ogni algoritmo per la ricerca della migliore *vine* dovrà ridurre al minimo il numero di *vine* che prende in considerazione. Si vedrà questo aspetto in dettaglio nella sezione 5.6.

[3] L'articolo è stato ristampato in Cayley (2009).

[4] Nella trattazione, è stato introdotto il concetto di *R-vine* senza definire quello di *vine*. Morales Napoles (2011) mostra che il numero delle *vine* (*regular* e non) è effettivamente pari a $\prod_{i=2}^{n} i^{i-2}$.

[5] Va osservato che Morales Napoles (2010) è una tesi di dottorato e seppur contenga la dimostrazione del numero di *R-vine*, essa verte su un altro argomento. La dimostrazione del numero di *R-vine* è stata successivamente pubblicata e rifinita in Morales Napoles (2011). Una dimostrazione alternativa e più semplice è fornita in Cooke, Kurowicka e Wilson (2015).

4.2.5 Rappresentazioni matriciali

Allo scopo di semplificare la generazione e la gestione delle *R-vine* nei *software*, sono state introdotte molteplici rappresentazioni matriciali di queste ultime. La prima rappresentazione matriciale è stata introdotta nella presentazione Kurowicka (2009). Quest'ultima è stata estesa per il calcolo della verosimiglianza da Dißmann (2010) e Dißmann et al. (2013).

Le matrici utilizzate dalle librerie `R` per questa tesi sono tutte matrici triangolari (o convertibili in matrici triangolari). Stöber e Czado (2016) usano matrici triangolari superiori, mentre in Dißmann et al. (2013) e Dißmann (2010) si vedrà un approccio che fa uso matrici triangolari inferiori.

La differenza tra le due rappresentazioni non risiede nella logica costruttiva ma bensí in una pura questione estetica. Infatti, è possibile passare da una rappresentazione ad altra con una rotazione di 180° della matrice. Per esempio, le due matrici seguenti:

$$T_L = \begin{pmatrix} 2 & & & & \\ 5 & 3 & & & \\ 3 & 5 & 4 & & \\ 1 & 1 & 5 & 5 & \\ 4 & 4 & 1 & 1 & 1 \end{pmatrix}, \text{ e } \qquad T_U = \begin{pmatrix} 1 & 1 & 1 & 4 & 4 \\ & 5 & 5 & 1 & 1 \\ & & 4 & 5 & 3 \\ & & & 3 & 5 \\ & & & & 2 \end{pmatrix} \tag{4.28}$$

rappresentano la stessa *R-vine*. La libreria di `R` `VineCopula` (si veda Nagler, Schepsmeier et al. (2023)) supporta entrambe le versioni; al contrario, `rvinecopulib` (si veda Nagler e Vatter (2023a)) usano una notazione diversa, simile ad una triangolare ma che si azzera al di sotto dell'antidiagonale. Seppur la documentazione della libreria (si veda Nagler e Vatter (2023a)) non faccia riferimento a questo aspetto, limitandosi a elencare definizioni e proprietà, la matrice usata da `rvinecopulib` non è altro che una notazione alternativa della stessa matrice. Può essere infatti vista o come una matrice triangolare superiore in cui ci sia stata una riflessione sulle colonne, oppure come una matrice triangolare inferiore in cui ci sia stata una riflessione sulle righe. Ad esempio, l'equivalente delle matrici T_L e T_U dell'equazione (4.28) è la matrice:

$$T_R = \begin{pmatrix} 4 & 4 & 1 & 1 & 1 \\ 1 & 1 & 5 & 5 & \\ 3 & 5 & 4 & & \\ 5 & 3 & & & \\ 2 & & & & \end{pmatrix} \tag{4.29}$$

Come si vedrà nella sezione 4.3 e nel capitolo 5, l'insieme dei vincoli $\mathcal{CV}$ è fondamentale nell'associazione tra componente grafica e componente stocastica di una *PCC*, non stupisce quindi che le rappresentazioni matriciali siano basate su di loro. Siccome queste rappresentazioni matriciali sono sostanzialmente uguali, similmente a Czado (2019), si descriverà soltanto la rappresentazione in matrici triangolari superiori di `VineCopula`. Al contrario di Czado (2019) si è ritenuto importante però proporre gli algoritmi anche per la versione di `rvinecopulib` in quanto è la libreria che è stata utilizzata nei capitoli da 8 a 10.

Definizione 4.24. *Sia M una matrice triangolare superiore con entrate (m_{ij}). La matrice M è una* R-vine matrix *(*superiore*) se soddisfa le seguenti proprietà:*

1. $1 \leq m_{ij} \leq n$ *per ogni* $i \leq j$ *e* $m_{ij} = 0$ *per* $i > j$*;*

2. $\{m_{1,i}, \ldots, m_{i,i}\} \subset \{m_{1,j}, \ldots, m_{j,j}\}$ *per* $1 \leq i < j \leq n$*, ovvero i valori di una specifica colonna sono contenuti in ogni colonna alla sua destra;*

3. $m_{i,i} \notin \{m_{1,i-1}, \ldots, m_{i-1,i-1}\}$*, ovvero il valore della diagonale di una colonna non è presente nella colonna alla sua sinistra.*

4. *per* $i = 3, \ldots, n$ *e* $k = 1, \ldots, i-1$ *esistono* (j, l) *con* $l < j < i$ *tali che*

$$\begin{aligned} &\{m_{k,i}, \{m_{1,i}, \{m_{1,i}, \ldots, m_{k-1,i}\}\}\} = \{m_{j,j}, \{m_{1,j}, \ldots, m_{l,j}\}\} \text{ oppure} \\ &\{m_{k,i}, \{m_{1,i}, \{m_{1,i}, \ldots, m_{k-1,i}\}\}\} = \{m_{l,j}, \{m_{1,j}, \ldots, m_{l-1,j}, m_{j,j}\}\}. \end{aligned} \tag{4.30}$$

Si chiamerà il rispettivo con matrici triangolari inferiori con R-vine matrix inferiore *mentre la versione usata da* `rvinecopulib` *si chiamerà con R-vine matrix* riflessa*.*

Per la *R-vine matrix* riflessa, gli autori di `rvinecopulib` hanno dato la seguente definizione:

Definizione 4.25. *Una* R-vine matrix *riflessa* $M = (m_{i,j})$ *viene detta in* ordine naturale *se* $m_{n-i+1,i} = i$ *per ogni* $1 \leq i \leq n$*.*

L'equazione (4.30) della definizione 4.24 è l'analogo della condizione 4 di prossimità per le *R-vine*. Gli algoritmi algoritmi 4.1 e 4.2 per convertire una *R-vine* in una *R-vine matrix* e viceversa sono presi da Stöber e Czado (2016) (nella versione riportata in Czado (2019)). L'equivalente dell'algoritmo 4.1 per la versione riflessa è presentata nell'algoritmo 4.3.

Con il proposito di comprendere meglio l'algoritmo 4.1, si userà per convertire l'esempio di *R-vine* presentata nella figura 4.8 della sezione 4.2.3, ovvero la *R-vine* con il seguente insieme dei vincoli:

$$\begin{aligned} E_1 &= \{(1,2;\emptyset), (2,3;\emptyset), (2,4;\emptyset), (4,5;\emptyset)\} \\ E_2 &= \{(1,4;2), (3,4;2), (2,5;4)\} \\ E_3 &= \{(1,5;2,4), (1,3;2,4)\} \\ E_4 &= \{(3,5;1,2,4)\} \end{aligned} \tag{4.31}$$

Al passo $i = 5$, si deve prendere un elemento di E_4. $\mathcal{X}$ è vuoto e E_4 ha un solo elemento: $(3,5;1,2,4)$. Pertanto si ha $e_5 = (3,5;1,2,4)$. L'unica scelta che va fatta è se usare $m_{5,5} = 3$ oppure $m_{5,5} = 5$. Scelto $m_{5,5} = 3$ si avrà $m_{4,5} = 5$. Siccome 3 compare una sola volta in ogni insieme degli archi E_i, anche le altre scelte sono ovvie, conducendo la matrice alla forma:

$$\begin{pmatrix} ? & ? & ? & ? & 2 \\ & ? & ? & ? & 4 \\ & & ? & ? & 1 \\ & & & ? & 5 \\ & & & & 3 \end{pmatrix}. \tag{4.32}$$

Algoritmo 4.1: Calcolo di una *R-vine matrix* superiore a partire da una *R-vine* $\mathcal{V}$.

input : Una *R-vine* $\mathcal{V}$ su n elementi
output: Una *R-vine matrix* superiore

$\mathcal{X} \leftarrow \emptyset$;
per $i \leftarrow n-1$ **a** 3 **fai**
  Selezionare $e_i \in E(T_i)$ tale che $\mathcal{C}_{e_i} = \{a_{e_i}, b_{e_i}\} \cap \mathcal{X} = \emptyset$;
  $m_{i,i} \leftarrow a_{e_i}$;
  $m_{i-1,i} \leftarrow b_{e_i}$;
  per $k \leftarrow i-1$ **a** 1 **fai**
    Selezionare $e_k \in E(T_k)$ con $\mathcal{C}_{e_k} = \{m_{i,i}, b_{e_k}\}$ e $b_{e_k} \notin \mathcal{X}$;
    $m_{k,i} \leftarrow b_{e_k}$;
  fine
  $\mathcal{X} \leftarrow \mathcal{X} \cup \{m_{i,i}\}$;
fine
Selezionare $a, b \notin \mathcal{X}$;
$m_{2,2} \leftarrow a$;
$m_{1,2} \leftarrow b$;
$m_{1,1} \leftarrow b$;
ritorna $M = (m_{i,j})$

Algoritmo 4.2: Calcolo di una *R-vine* $\mathcal{V}$ da una *R-vine matrix*.

input : Una *R-vine matrix* superiore $M = (m_{i,j})$ di dimensione $n \times n$
output: Una *R-vine* $\mathcal{V}$

$E_0 \leftarrow \{1, \ldots, n\}$;
$E_1 \leftarrow \{\{m_{1,2}, m_{2,2}\}\}$;
per $k \leftarrow 2$ **a** n **fai**
  $E_k = \emptyset$;
fine
per $i \leftarrow 3$ **a** n **fai**
  $e_1^i \leftarrow \{m_{1,i}, m_{i,i}\}$;
  $E_1 \leftarrow E_1 \cup \{e_1^i\}$;
  per $k \leftarrow 1$ **a** $i-2$ **fai**
    Selezionare $a_k \in E_k$ con $A_{a_k} = \{m_{1,i}, \ldots, m_{k+1,i}\}$;
    $e_{k+1}^i \leftarrow \{e_k^i, a_k\}$;
    $E_{k+1} \leftarrow E_{k+1} \cup \{e_{k+1}^i\}$;
  fine
fine
per $k \leftarrow 1$ **a** $n-1$ **fai**
  $T_k \leftarrow (E_{k-1}, E_k)$;
fine
$T_n \leftarrow (E_{n-1}, \emptyset)$;
$\mathcal{V} \leftarrow (T_1, \ldots, T_n)$;
ritorna $\mathcal{V}$

Algoritmo 4.3: Calcolo di una *R-vine matrix* riflessa a partire da una *R-vine* $\mathcal{V}$.

input : Una *R-vine* $\mathcal{V}$ su n elementi
output: Una *R-vine matrix* riflessa

$\mathcal{X} \leftarrow \emptyset$;
per $j \leftarrow 1$ **a** $n-2$ **fai**
  Selezionare $e_j \in E(T_{n-j})$ tale che $\mathcal{C}_{e_j} = \{a_{e_j}, b_{e_j}\} \cap \mathcal{X} = \emptyset$;
  $m_{n+1-j,j} \leftarrow a_{e_j}$;
  $m_{n-j,j} \leftarrow b_{e_j}$;
  per $k \leftarrow n-j-1$ **a** 1 **fai**
    Selezionare $e_k \in E(T_k)$ con $\mathcal{C}_{e_k} = \{m_{n+1-j,j}, b_{e_k}\}$ e $b_{e_k} \notin \mathcal{X}$;
    $m_{k,i} \leftarrow b_{e_k}$;
  fine
  $\mathcal{X} \leftarrow \mathcal{X} \cup \{m_{n+1-j,j}\}$;
fine
Selezionare $a, b \notin \mathcal{X}$;
$m_{2,n-1} \leftarrow a$;
$m_{1,n-1} \leftarrow b$;
$m_{1,n} \leftarrow b$;
ritorna $M = (m_{i,j})$

Definita l'ultima colonna, si può passare alla penultima, ovvero al passo $i = 4$. Si deve scegliere un arco e che non contiene 3 nell'insieme condizionato. Anche in questo caso si ha una sola possibilità, ovvero $(1, 5; 2, 4)$. Anche in questo caso, si può scegliere $m_{4,4} = 1$ oppure $m_{4,4} = 5$. In questo caso, si sceglierà $m_{4,4} = 5$ e quindi $m_{3,4} = 1$. Similmente al passo precedente, non c'è un solo arco con 5 per ogni E_i. Si ricava pertanto la matrice:

$$\begin{pmatrix} ? & ? & ? & 4 & 2 \\ & ? & ? & 2 & 4 \\ & & ? & 1 & 1 \\ & & & 5 & 5 \\ & & & & 3 \end{pmatrix}. \tag{4.33}$$

Nel passo $i = 3$, si ha $e_3 = (1, 4; 2)$. Se ne ricava $m_{3,3} = 1$, $m_{2,3} = 4$ e $m_{1,3} = 2$. Per gli ultimi 3 elementi si può considerare $a = 2$ e $b = 4$. Sicché, una *R-vine matrix* superiore per il nostro esempio è:

$$\begin{pmatrix} 4 & 4 & 2 & 4 & 2 \\ & 2 & 4 & 2 & 4 \\ & & 1 & 1 & 1 \\ & & & 5 & 5 \\ & & & & 3 \end{pmatrix}. \tag{4.34}$$

È evidente dai passi che sono stati fatti che esistono molteplici *R-vine matrices* superiore associate ad una certa *R-vine*. Per esempio, si poteva scegliere $m_{5,5} = 5$

e quindi avremmo potuto trovarci con la matrice:

$$\begin{pmatrix} 4 & 4 & 2 & 2 & 4 \\ & 2 & 4 & 4 & 2 \\ & & 1 & 1 & 1 \\ & & & 3 & 3 \\ & & & & 5 \end{pmatrix}, \tag{4.35}$$

o anche si poteva evitare di scegliere uno tra 3 e 5 fino all'ultimo e trovarci con:

$$\begin{pmatrix} 5 & 5 & 4 & 2 & 2 \\ & 4 & 5 & 4 & 4 \\ & & 2 & 5 & 1 \\ & & & 1 & 5 \\ & & & & 3 \end{pmatrix}. \tag{4.36}$$

Rispetto agli algoritmi in Stöber e Czado (2016) è stato aggiunta la condizione $b_{e_k} \notin \mathcal{X}$ perché quando è stata calcolata la matrice (4.36) è emerso che con $m_{3,3} = 2$ vi erano ben tre possibili e_1.

L'algoritmo 4.2 costruisce $\mathcal{V}$ usando direttamente la definizione 4.17 invece di costruire l'insieme dei vincoli $\mathcal{CV}$. Come esempio, si consideri la matrice dell'equazione (4.34). All'inizio dell'algoritmo 4.2 si ha:

$$\begin{aligned} E_1 &= \{(2,4;\emptyset)\}, \\ E_2 &= \emptyset, \\ E_3 &= \emptyset, \text{ e} \\ E_4 &= \emptyset. \end{aligned}$$

Al passo $i = 3$, l'algoritmo pone $e_1^3 = (1,4;\emptyset)$ e lo aggiunge a E_1. Quindi, si ha $a_1 = (2,4;\emptyset)$ e $e_2^3 = \{(1,4;\emptyset),(2,4;\emptyset)\} = (1,4;2)$. In altre parole, dopo il passo $i = 3$, ci si trova con:

$$\begin{aligned} E_1 &= \{(1,4;\emptyset),(2,4;\emptyset)\}, \\ E_2 &= \{(1,4;2)\}, \\ E_3 &= \emptyset, \text{ e} \\ E_4 &= \emptyset. \end{aligned}$$

Il passo $i = 4$ parte in modo simile: $e_1^4 = (4,5;\emptyset)$. Si avrà quindi $a_1 = (2,4;\emptyset)$, $a_2 = (1,4;2)$, $e_2^4 = \{(4,5;\emptyset),(2,4;\emptyset)\} = (2,5;2)$ e $e_3^4 = \{(2,5;2),(1,4;2)\} = (1,5;2,4)$. In buona sostanza, ci si ritrova con:

$$\begin{aligned} E_1 &= \{(1,4;\emptyset),(2,4;\emptyset),(4,5;\emptyset)\}, \\ E_2 &= \{(1,4;2),(2,5;2)\}, \\ E_3 &= \{(1,5;2,4)\}, \text{ e} \\ E_4 &= \emptyset. \end{aligned}$$

Algoritmo 4.4: Calcolo di una *R-vine* $\mathcal{V}$ da una *R-vine matrix* riflessa.

input : Una matrice *R-vine matrix* superiore $M = (m_{i,j})$ di dimensione $n \times n$
output: Una *R-vine* $\mathcal{V}$

$E_0 \leftarrow \{1, \ldots, n\}$;
per $t \leftarrow 1$ **a** $n-1$ **fai**
  $E_t \leftarrow \emptyset$;
  per $s \leftarrow t+1$ **a** n **fai**
    $E_t \leftarrow E_t \cup \Big\{\big(m_{s,s}, m_{t,s}; \{m_{t-1,s}, \ldots, m_{1,s}\}\big)\Big\}$;
  fine
  $T_t \leftarrow (E_{t-1}, E_t)$;
fine
$T_n \leftarrow (E_{n-1}, \emptyset)$;
$\mathcal{V} \leftarrow (T_1, \ldots, T_n)$;
ritorna $\mathcal{V}$

Algoritmo 4.5: Calcolo di una *R-vine* $\mathcal{V}$ da una *R-vine matrix* riflessa.

input : Una matrice *R-vine matrix* riflessa $M = (m_{i,j})$ di dimensione $n \times n$
output: Una *R-vine* $\mathcal{V}$

$E_0 \leftarrow \{1, \ldots, n\}$;
per $t \leftarrow 1$ **a** $n-1$ **fai**
  $E_t \leftarrow \emptyset$;
  per $s \leftarrow 1$ **a** $n-t$ **fai**
    $E_t \leftarrow E_t \cup \Big\{\big(m_{n+1-s,s}, m_{t,s}; \{m_{t-1,s}, \ldots, m_{1,s}\}\big)\Big\}$;
  fine
  $T_t \leftarrow (E_{t-1}, E_t)$;
fine
$T_n \leftarrow (E_{n-1}, \emptyset)$;
$\mathcal{V} \leftarrow (T_1, \ldots, T_n)$;
ritorna $\mathcal{V}$

Il passo $i = 5$ concluderà quindi la costruzione della *R-vine*.

La documentazione di `rvinecopulib` (Nagler e Vatter (2023a)) mostra che esite un metodo migliore per convertire una *R-vine matrix* riflessa direttamente nell'insieme dei vincoli $\mathcal{CV}$. Vale infatti la seguente relazione:

$$(m_{n+1-s,s}, m_{t,s}; \{m_{t-1,s}, \dots, m_{1,s}\}) \in \mathcal{CV} \tag{4.37}$$

per ogni $1 \leq t < n$ ed $1 \leq s \leq n - t$. Si noti che t rappresenta l'albero associata al vincolo, mentre s è l'indice dell'arco in un particolare ordinamento di T_t.

Siccome le varie matrici sono equivalenti, esiste un relazione simile anche per le matrici a *R-vine* superiori e inferiori. Per quelle superiori vale:

$$(m_{s,s}, m_{t,s}; \{m_{t-1,s}, \dots, m_{1,s}\}) \in \mathcal{CV} \tag{4.38}$$

per ogni $1 \leq t < n$ ed $t < s \leq n$.

Gli algoritmi relativi all'equazione (4.37) e equazione (4.38) sono gli algoritmi 4.4 e 4.5, verranno usati sulla matrice T_U dell'equazione (4.28) presentate all'inizio di questa trattazione delle matrici a *R-vine*.

Per l'albero T_1 si hanno gli elementi: $(m_{2,2}, m_{1,2}; \emptyset)$, $(m_{3,3}, m_{1,3}; \emptyset)$, $(m_{4,4}, m_{1,4}; \emptyset)$ e $(m_{5,5}, m_{1,5}; \emptyset)$, ovvero gli elementi $(5, 1; \emptyset)$, $(4, 1; \emptyset)$, $(3, 4; \emptyset)$ e $(2, 4; \emptyset)$.

Per l'albero T_2 si hanno gli elementi: $(m_{3,3}, m_{2,3}; m_{1,3})$, $(m_{4,4}, m_{2,4}; m_{1,4})$ e $(m_{5,5}, m_{2,5}; m_{1,5})$, ovvero gli elementi $(4, 5; 1)$, $(3, 1; 4)$ e $(2, 1; 4)$.

Per l'albero T_2 si hanno gli elementi: $(m_{4,4}, m_{3,4}; m_{2,4}, m_{1,4})$ e $(m_{5,5}, m_{3,5}; m_{2,5}, m_{1,5})$, ovvero gli elementi $(3, 5; 1, 4)$ e $(2, 3; 1, 4)$.

L'arco dell'ultimo albero è descritto da $(m_{5,5}, m_{4,5}; m_{4,5}, m_{2,5}, m_{1,5})$ ovvero è uguale a $(2, 5; 3, 1, 4)$.

4.3 *R-vine distributions*

La teoria finora trattata si è limitata alla teoria dei grafi. Si vuole quindi associare una struttura di tipo probabilistico ad una *R-vine* in modo di fornire una decomposizione in copule bivariate, condizionate e non, di una distribuzione multivariate. Si noti che nel seguito si userà d invece che n per la dimensione dello spazio associato alla *R-vine*. Questa scelta è basata sul fatto che nel capitolo 5 si userà n per la dimensione del campione.

Definizione 4.26. *La distribuzione congiunta F di un vettore d-dimensionale $\boldsymbol{X} = (X_1, \dots, X_d)$ possiede una R-*vine distribution *se è possibile definire una tripletta $(\boldsymbol{F}, \mathcal{V}, \mathcal{B})$ tale che:*

1. *$\boldsymbol{F} = (F_1, \dots, F_d)$ è un vettore di funzioni di distribuzione marginali continue ed invertibili tali che $X_i \sim F_i$ per ogni $1 \leq i \leq d$.*

2. *$\mathcal{V} = (T_1, \dots, T_d)$ è una R-*vine *su d elementi.*

3. *$\mathcal{B}$ è un insieme di copule bivariate simmetriche indicizzate dagli elementi di $\mathcal{CV}$.*

4. *Per ogni* $e = (a_e, b_e; \mathcal{D}_e) \in \mathcal{CV}$*, la copula* C_e *è la copula associata alla distribuzione condizionale di* X_{a_e} *e* X_{b_e} *dato* $\boldsymbol{X}_{\mathcal{D}_e} = \boldsymbol{x}_{\mathcal{D}_e}$*. Inoltre* C_e *non dipende dalla scelta di* $\boldsymbol{x}_{\mathcal{D}_e}$*.*

Si noti che la condizione che C_e non dipende dalla scelta di $\boldsymbol{x}_{\mathcal{D}_e}$ è una *assunzione semplificante*. La sezione 4.4 analizzerà come questa ipotesi sia appropriata o meno.

In futuro, si denoterà la copula C_e con $C_{a_e,b_e;\mathcal{D}_e}$ e la sua densità corrispondente con $c_{a_e,b_e;\mathcal{D}_e}$. Questa copula prende il nome di *pair-copula*. In generale è usuale ordinare gli indici delle copule in ordine crescente. Ovvero, quando si scriverà $C_{i,j;i_1\ldots,i_r}$ si avrà che $i < j$ e $i_1 < \cdots < i_r$.

Osservazione 4.3. *Siccome gli alberi di una* R-vine *non sono diretti, ci si deve necessariamente restringere a copule con simmetria radiale. Questa restrizione è più che altro formale dato che la teoria di questo capitolo rimane vera anche se si scegliesse di usare alberi diretti. D'altra parte, l'ipotesi di simmetria radiale è comune nella letteratura sulle* vine *allo scopo di usare la notazione insiemistica. In simulazioni e applicazioni è comunque possibile lavorare con copule non simmetriche se queste ultime si mostrassero necessarie. Nelle applicazioni dei capitoli da 8 a 10, tutte le copule parametriche usate sono con simmetrie radiale, ma sono state usate varie copule non-parametriche che non possiedono tale caratteristica.*

Bedford e Cooke (2002) hanno dimostrato che ad una tipletta $(\boldsymbol{F}, \mathcal{V}, \mathcal{B})$ che soffisfa i primi 3 punti della definizione 4.26 corrisponde un'unica distribuzione d-dimensionale F. In particolare, vale il seguente teorema.

Teorema 4.2. *Sia* $(\boldsymbol{F}, \mathcal{V}, \mathcal{B})$ *una tripletta di insiemi che soddisfa i primi tre punti della definizione 4.26, allora esiste un'unica distribuzione* d*-dimensionale* F *con densità*

$$f_{1,\ldots,d}(\boldsymbol{x}) = \Big(\prod_{i=1}^{d} f_i(x_i)\Big) \cdot \Big(\prod_{e\in\mathcal{CV}} c_{a_e,b_e;\mathcal{D}_e}\Big(F_{a_e|\mathcal{D}_e}(x_{a_e}|\boldsymbol{x}_{\mathcal{D}_e}), F_{b_e|\mathcal{D}_e}(x_{b_e}|\boldsymbol{x}_{\mathcal{D}_e})\Big)\Big) \quad (4.39)$$

tale che

$$F_{a_e,b_e|\mathcal{D}_e} = C_e\Big(F_{a_e|\mathcal{D}_e}(x_{a_e}|\boldsymbol{x}_{\mathcal{D}_e}), F_{b_e|\mathcal{D}_e}(x_{b_e}|\boldsymbol{x}_{\mathcal{D}_e})\Big). \quad (4.40)$$

Inoltre le marginali unidimensionali di F *sono date dalle* F_i *per ogni* $1 \le i \le d$*.*

Il teorema 4.2 permette quindi di specificare una *R-vine distribution* semplicemente usando la tripletta $(\boldsymbol{F}, \mathcal{V}, \mathcal{B})$.

Osservazione 4.4. *Il teorema 4.2 rimane valido nel caso in cui si considerassero archi diretti invece che indiretti. In questo caso la equazione* (4.39) *va modificata specificando che* a *è la testa e* b *è la coda dell'arco diretto* $a \to b$*. Questo aspetto compare già in Bedford e Cooke (2002), anche se gli autori non l'hanno esplicitato apertamente.*

Si è quindi pronti a dare una definizione di *PCC*.

Definizione 4.27. *Una* R-vine copula*, o* pair-copula construction *(abbreviata in* PCC*), è una* R-vine distribution $(\boldsymbol{F}, \mathcal{V}, \mathcal{B})$ *tale che* $F_i \sim \mathrm{U}(0,1)$ *per ogni* $1 \le i \le d$*.*

Si adotterà l'aggettivo "semplificata" per le *PCC* quando si vorrà distinguere tra decomposizioni che usano l'ipotesi semplificante e decomposizioni che non la usano. Si rimanda alla sezione 4.4 per approfondimenti su questo aspetto.

4.4 L'ipotesi semplificante per le *PCC*

Come è stato già detto nella trattazione di questo capitolo, l'ipotesi semplificante consiste nell'associare ad ogni arco di una *PCC* una specifica copula bivariata ignorando la dipendenza diretta dalle variabili condizionanti.

Va però osservato che la presenza delle funzioni di ripartizione condizionate per le variabili condizionate fa sì che la copula potrebbe non dipendere direttamente dalle variabili condizionanti. In altre parole, esistono distribuzioni che possono essere completamente rappresentate attraverso *PCC* semplificate. Czado (2019) fornisce i seguenti esempi:

Teorema 4.3. *Le seguenti classi di copule multivariate possono essere rappresentate da* PCC *semplificate:*

1. *La classe delle copule Gaussiane multivariate. In questo caso le pair-copule sono tutte copule Gaussiane bivariate con i parametri di dipendenza forniti dalle corrispondenti correlazioni parziali;*

2. *La classe delle copule multivariate t-Student con μ gradi di libertà e matrice di correlazione R. In questo caso, le pair copule dell'albero T_j sono t-Student con $\mu+j-1$ gradi di libertà e i parametri di dipendenza forniti dalle corrispondenti correlazioni parziali;*

3. *La classe delle copule multivariate di Clayton.*

Seppur nella pratica i fenomeni analizzati siano raramente rappresentabili fedelmente con le tre famiglie del teorema 4.3, il fatto che queste famiglie siano decomponibili in *PCC* semplificate assicura che la modellazione attraverso *PCC* sia necessariamente più accurata di una modellazione che si limiti ad usare quelle tre famiglie.

Hobæk Haff, Aas e Frigessi (2010), nella loro analisi sull'ipotesi semplificante, forniscono esempi non banali sia "positivi" che "negativi". Inoltre, dimostrano alcune proposizioni che possono essere usate per determinare se una distribuzione possa essere rappresentata con *PCC* semplificate. Queste proposizioni, similmente al teorema 4.3 sono limitate al caso di distribuzioni continue e richiedono la conoscenza della distribuzione multivariata che si vuole decomporre. Seppur siano interessanti dal punto di vista teorico, la loro utilità pratica è dubbia.

A titolo di esempio, Hobæk Haff, Aas e Frigessi (2010) dimostrano che se il τ_K di Kendall (o un'altra misura di concordanza simile) della copula condizionata dipende dalle variabili condizionanti allora la distribuzione non può essere decomposta in forma semplificata.

Hobæk Haff, Aas e Frigessi (2010) hanno inoltre fatto alcuni test con tre variabili aleatorie e hanno mostrato che una *PCC* semplificata fornisce spesso una buona approssimazione anche quando l'ipotesi semplificante non sia soddisfatta.

Czado (2019) osserva che l'elevato numero di *R-vine* (si veda sezione 4.2.4) e la possibilità di usare ogni tipo di copula bivariata, parametrica o non paramatrica, rende la classe delle *PCC* semplificate molto flessibile. Inoltre, Nagler e Czado (2016) hanno mostrato che l'uso di bicopule non parametriche nelle *PCC* semplificate evita la *maledizione della dimensionalità* nella stima di densità multivariate.

Un'analisi dell'uso delle copule non parametriche nel contesto di *PCC* semplificate è stata fatta in Nagler, Schellhase e Czado (2017).

Stöber, Joe e Czado (2013) hanno analizzato l'applicabilità dell'ipotesi semplificante a copule archimedee ricavando che, per $n \geq 3$, solo quelle basate su una trasformazione di Laplace Gamma o una sua estensione possiedono questa proprietà. Inoltre, hanno mostrato che le copule t-Student sono le uniche a possedere questa proprietà tra le copule che derivano da una mistura di normali. Infine, Stöber, Joe e Czado (2013) hanno mostrato come le *PCC* possano essere adattate a situazioni in cui l'ipotesi semplificante non vale e hanno presentato una metodologia per analizzare la distanza tra una distribuzione multivariata conosciuta e una distribuzione vicina che soddisfi l'ipotesi semplificante.

Acar, Genest e Nešlehová (2012) hanno invece presentato una metodologia di *smoothing* polinomiale che permette di rilassare l'ipotesi semplificante nel caso di *PCC* trivariate. Tuttavia, questo approccio è limitato a tre dimensioni e prevede uno *smoothing* multidimensionale, che potrebbe essere difficile da estendere a dimensioni superiori. Un approccio più fattibile in dimensioni superiori, è stato sviluppato da Vatter e Nagler (2018).

Nella modellazione inferenziale una analisi dell'applicabilità dell'ipotesi semplificante è certamente difficile, specialmente al crescere del numero di variabili. Va inoltre detto che la scelta delle funzioni marginali ha anch'essa un peso.

Capitolo 5

L'approccio Inferenziale nella *Pair-Copula Construction*

> The use of copulas is however challenging in higher dimensions where standard multivariate copulas suffer from rather inflexible structures.
>
> Brechmann e Schepsmeier (2013)

Uno degli usi più diffusi in materia di copule è la sua applicazione nel contesto inferenziale ovvero l'utilizzo delle copule per la determinazione della distribuzione congiunta di una popolazione partendo da un *dataset*. Tuttavia, si segnala che nel caso di distribuzioni discrete, questo approccio risente di notevoli limitazioni (si veda il capitolo 6).

Il settore attuariale, seppur nella pratica non abbia sfruttato questa metodologia con lo stesso successo di altri settori, ha mostrato più volte interesse nel suo uso per studiare la dipendenza tra variabili. Per esempio, l'International Actuarial Association (2004) propone di usare le copule per aggregare i rischi per la loro capacità di analizzare le dipendenze di coda.

Prima di intraprendere il percorso della modellazione attraverso *PCC*, è opportuno interrogarsi sulle modalità e sull'appropriatezza dell'utilizzo delle copule nel perimetro dell'inferenza statistica.

Sia $\boldsymbol{X} = (X_1, \ldots, X_n)$ un vettore aleatorio di dimensione n con funzione di distribuzione congiunta F e distribuzioni marginali $F_1, \ldots, F_n$. Per il teorema 1.2 di Sklar, allora esiste almeno una copula C[1] che soddisfa la seguente:

$$F(x_1, \ldots, x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big) \tag{5.1}$$

Un modello parametrico basato sulle copule è un modello che presuppone che $C \in C_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathcal{O}\}$ dove $\mathcal{O}$ è un insieme aperto di $\mathbb{R}^n$ per qualche $n \geq 1$.

A latere del problema della determinazione dei parametri è necessario studiarne la bontà di adattamento, ovvero quando l'ipotesi $C \in C_0$ sia appropriata.

[1] C è unica se F è continua.

Altresì, quando esiste un modello alternativo, si dovrebbe analizzare quale dei due sia migliore o sé invece siano entrambi da considerarsi equivalenti. Genest, Rémillard e Beaudoin (2009) comparano varie metodologie per analizzare la bontà di adattamento di questo tipo di modelli per $n = 2$.

Nella sezione 5.1 si analizzerà come sia caratterizzato un modello *PCC* parametrico. Nella sezione 5.2 si continuerà ad analizzare la modellazione inferenziale usando copule, in particolare ci si soffermerà sul perché sia conveniente spezzare la modellazione delle marginali da quello della struttura di dipendenza. Nella sezione 5.3 si vedrà come sia fatta la funzione di verosimiglianza di un modello *PCC*. Le sezioni da 5.4 a 5.6 saranno invece dedicate alle metodologie per la stima dei parametri del modello. In particolare, nelle sezione 5.6 si analizzeranno gli algoritmi per la determinazione della struttura *R-vine* del modello. Si finirà il capitolo con la sezione 5.7 in cui ci si soffermerà sui maggiori test e indicatori per comparare i modelli *PCC* parametrici. Questo capitolo seguirà principalmente Hobæk Haff (2013) e Czado (2019).

La sezione 5.8 abbandonerà la letteratura per tornare alla domanda di ricerca e fare il punto della situazione e riassumere i principali aspetti teorici del capitolo.

5.1 Modello parametrico di una *PCC*

Questa sezione si soffermerà su come sia costruito un modello inferenziale basato sulla decomposizione *PCC*. Per la definizione 4.26, una *R-vine distribution* su n elementi è una tripletta $(\boldsymbol{F}, \mathcal{V}, \mathcal{B})$, dove $\boldsymbol{F} = (F_1, \dots, F_n)$ sono le distribuzioni marginali, le $\mathcal{V}$ è una *R-vine* su n elementi e $\mathcal{B}$ è un insieme di copule indicizzato dagli archi di $\mathcal{V}$. Secondo la definizione 4.27, in una *PCC* $\boldsymbol{F}$ è formato da distribuzioni uniformi, ma nella pratica si ha raramente a che fare con distribuzioni marginali uniformi. Perciò, si parlerà di modellazione *PCC*, anche quando il modello è nei fatti una *R-vine distribution*[2].

Siccome l'insieme $\mathcal{B}$ è indicizzato da $\mathcal{V}$, è evidente che il primo elemento di un modello inferenziale *PCC* sia la *R-vine* stessa. Infatti a viti diverse corrispondono modelli differenti. Per quanto riguarda invece $\mathcal{B}$, è possibile scegliere un diverso modello per ogni arco.

Definizione 5.1. *Un* modello *PCC* parametrico *è una tripletta* $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$ *dove*

- $\mathcal{V}$ *è una* R-vine *su* n *elementi;*
- $\mathcal{B}(\mathcal{V})$ *è un insieme, indicizzato dagli archi di* $\mathcal{V}$*, e i cui elementi sono famiglie parametriche di copule bivariate;*
- $\Theta(\mathcal{B}(\mathcal{V}))$ *è l'insieme dei parametri per le famiglie* $\mathcal{B}(\mathcal{V})$*.*

[2]Questa scelta non crea di solito problemi perché è comune spezzare la stima dei parametri delle marginali da quelli delle copula. A seconda dei casi si può usare la distribuzione empirica costruita usando la funzione rango oppure un tipo di modellazione più rifinito. È fondamentale osservare che una scelta diversa per il modello delle marginali porta a copule differenti e pertanto a modelli *PCC* distinti. Questo aspetto verrà approfondito nelle sezioni da 5.2 a 5.4

Com'è evidente dalla teoria nel capitolo 4, la *R-vine* $\mathcal{V}$ di un modello *PCC* parametrico $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$ non possiede alcun significato stocastico intrinseco. D'altra parte, l'incertezza creata dai singoli modelli e l'errore potenziale insito nell'ipotesi simplificante (si veda sezione 4.4) dà alla scelta della *R-vine* una significatività che altrimenti gli sarebbe sconosciuta. Ma questa significatività è fattuale solo nel suo realizzarsi in un modello *PCC* parametrico completo. Non è possibile dunque trovare a priori una *R-vine* migliore senza dare attualità agli insiemi $\mathcal{B}(\mathcal{V})$ e $\Theta(\mathcal{B}(\mathcal{V}))$ ed aver trovato attraverso opportuni metodi inferenziali i parametri $\hat{\boldsymbol{\theta}} \in \Theta(\mathcal{B}(\mathcal{V}))$ che meglio descrivono i dati.

Perciò ogni metodologia che determina la migliore *R-vine* $\mathcal{V}$ senza al contempo costruire il resto del modello non può garantire di trovare una soluzione ottimale né globale né locale e il crescere del numero di osservazioni o di variabili non migliora le cose.

In buona sostanza, se la dipendenza insiemistica va dalla *R-vine* $\mathcal{V}$ ai parametri dei modelli, il problema inferenziale deve seguire un percorso a ritroso. Ovvero si possono definire i seguenti tre problemi inferenziali:

1. stima dei parametri di copula $\hat{\boldsymbol{\theta}} \in \Theta(\mathcal{B}(\mathcal{V}))$ per un modello $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$ fissato;

2. selezione dei modelli $(\mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$ e dei relativi parametri per una *R-vine* fissata;

3. selezione e stima dell'intero modello $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$.

Il primo problema sarà trattato nella sezione 5.4, il secondo nella sezione 5.5 e il terzo nella sezione 5.6. Prima di passare alle metodologie per risolvere questi problemi, è utile soffermarsi sulla funzione di verosimiglianza associata ad una *PCC* e alle sue conseguenze.

5.2 Stimatori di copula multivariati

In questa sezione si presenteranno alcune varianti dello stimatore di massima verosimiglianza per problemi inferenziali con modelli multivariati basati sulle copule. Similmente a Hobæk Haff (2013), ci si soffermerà sullo *stimatore* (*classico*) *di massima verosimiglianza* (*ML*) e l'*inference functions of margins* (*IFM*). Lo *stepwise semiparametric estimator* (*SSP*), ovvero la metodologia analizzata in Hobæk Haff (2013), sarà invece presentata nella sezione 5.4.

Si consideri dunque la distribuzione congiunta F, sconosciuta, di un vettore aleatorio n-dimensionale $\boldsymbol{X} = (X_1, \ldots, X_n)$ con distribuzioni marginali $F_1, \ldots, F_n$, anch'esse sconosciute.

Un campione di dimensione d estratto da $\boldsymbol{X}$ può essere rappresentato con la seguente matrice di dimensione $d \times n$:

$$\boldsymbol{x} = (\boldsymbol{x}_1{}^\intercal, \ldots, \boldsymbol{x}_d{}^\intercal) \tag{5.2}$$

dove:

$$\boldsymbol{x}_k = (x_{k,1}, \ldots, x_{k,n})^\intercal \tag{5.3}$$

per ogni $1 \leq i \leq d$.

Il problema inferenziale consiste nel determinare i parametri di:

$$F(\boldsymbol{x}, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n, \boldsymbol{\theta}) = C\big(F_1(x_1; \boldsymbol{\lambda}_1), \ldots, F_n(x_n; \boldsymbol{\lambda}_n); \boldsymbol{\theta}\big) \tag{5.4}$$

dove $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n$ sono i parametri dei modelli per le distribuzioni marginali e $\boldsymbol{\theta}$ sono i parametri di copula. Per semplicità, si supporrà nel resto del capitolo che la distribuzione congiunta F sia assolutamente continua e le F_i strettamente crescenti.

In questo caso, il problema diventa determinare i parametri in:

$$f(\boldsymbol{x}; \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n, \boldsymbol{\theta}) = c\big(F_1(x_1; \boldsymbol{\lambda}_1), \ldots, F_n(x_n; \boldsymbol{\lambda}_n); \boldsymbol{\theta}\big) \prod_{j=1}^{n} f_j(x_j; \boldsymbol{\lambda}_j) \tag{5.5}$$

dove f è la densità di F e $f_1, \ldots, f_n$ sono le densità di $F_1, \ldots, F_n$.

Perciò, le n funzioni di log-verosimiglianza associate alle distribuzioni marginali di $\boldsymbol{X}$ associata al campione $\boldsymbol{x}$ sono:

$$\ell_j(\alpha_j; \boldsymbol{x}) = \sum_{i=1}^{d} \log f_j(x_{ij}; \boldsymbol{\lambda}_j)\,, \tag{5.6}$$

per ogni $1 \leq j \leq n$. Mentre la funzione di log-verosimiglianza per la distribuzione congiunta è:

$$\begin{aligned} \ell(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n, \boldsymbol{\theta}; \boldsymbol{x}) &= \sum_{i=1}^{d} \log c\big(F_1(x_{i,1}; \boldsymbol{\lambda}_1), \ldots, F_n(x_{i,n}; \boldsymbol{\lambda}_n); \boldsymbol{\theta}\big) + \sum_{i=1}^{d} \ell_j(\alpha_j; \boldsymbol{x}) \\ &= \ell_C(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n, \boldsymbol{\theta}; \boldsymbol{x}) + \ell_M(\alpha_j; \boldsymbol{x})\,, \end{aligned} \tag{5.7}$$

dove con ℓ_C e ℓ_M si indicano le componenti della funzione di log-verosimiglianza associate rispettivamente alla funzione di copula e alle funzioni marginali. Si noti che ℓ_C dipende dalla marginali scelte.

Lo stimatore classico di massima verosimiglianza (ML) consiste nell'ottimizzare congiuntamente $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_n$ e $\boldsymbol{\theta}$ e ottenere le stime $\hat{\boldsymbol{\lambda}}_1^{ML}, \ldots, \hat{\boldsymbol{\lambda}}_n^{ML}$ e $\hat{\boldsymbol{\theta}}^{ML}$. Si rimanda a Hobæk Haff (2013) o ad un manuale classico sull'argomento (per esempio Lehmann e Casella (1998) o Lehmann (2010)) per comprendere come questo problema possa essere trattato.

Ciò che è invece importante osservare è che, nel caso delle *PCC*, la stima dei parametri usando lo stimatore ML richiede spesso ottimizzazioni di tipo numerico. Persino in basse dimensioni, come 4 o 5, il numero di parametri può essere molto alto, in particolare se si ammettono modelli con più parametri per ogni copula. Da qui la necessità di metodi più rapidi e computazionalmente più trattabili.

Un altro aspetto non trascurabile è che la correttezza del calcolo richiede che il modello sia corretto o per lo meno vicino al modello corretto[3] (si veda per esempio Claeskens e Hjort (2008)). In generale, però, lo stimatore non è robusto quando il modello diverge in maniera più significativa dal modello reale.

[3]La vicinanza si intende nel senso di Kullback-Leiber, si analizzerà cosa questo significhi nella sezione 5.7.3

Con l'obiettivo di rispondere alle limitazioni computazionali degli stimatori ML, sono stati introdotti due varianti in più passaggi.

Lo stimatore *inference functions of margins* (*IFM*) è stato introdotto e analizzato da McLeish e Small (1988), Xu (1996), Joe (1997) e Joe (2005) e non sfrutta in alcun modo le specifiche caratteristiche della decomposizione *PCC*. Questa metodologia consiste nel fare n ottimizzazioni separate ed indipendenti per le verosimiglianze univariate delle marginali seguite dall'ottimizzazione della verosimiglianza multivariata. Nello specifico:

1. le log-verosimiglianze $\ell_j(\boldsymbol{\lambda}_j; \boldsymbol{x})$ delle n marginali sono massimizzate separatamente per ottenere le stime $\hat{\boldsymbol{\lambda}}_1^{IFM}, \ldots, \hat{\boldsymbol{\lambda}}_n^{IFM}$;
2. la funzione $\ell(\hat{\boldsymbol{\lambda}}_1^{IFM}, \ldots, \hat{\boldsymbol{\lambda}}_n^{IFM}, \boldsymbol{\theta}; \boldsymbol{x})$ viene massimizzata rispetto alla variabile $\boldsymbol{\theta}$ per ottenere $\hat{\boldsymbol{\theta}}$.

In generale, l'ML e l'IFM producono risultati differenti in quando la semplificazione fornita dall'approccio IFM gli impedisce di cogliere quegli elementi della dipendenza tra le variabili che si nascondono nei modelli delle variabili marginali. Numerosi studi, tra cui Joe (2005) e Kim, M. Silvapulle e P. Silvapulle (2007), hanno mostrato che a meno che la dipendenza tra le variabili non sia estrema questa perdita di informazione è relativamente trascurabile. Entrambi gli stimatori sono consistenti e asintoticamente normali. Questo metodo è inoltre generalmente molto più rapido dello stimatore ML e può essere eventualmente usato come primo passo per una ottimizzazione ML completa. Ovviamente, quando il numero di parametri di $\boldsymbol{\theta}$ è elevato (come nel caso dei modelli *PCC*), il vantaggio computazionale dello stimatore IMF si assottiglia e rimane ancora troppo lento in pratica.

Un altro metodo a due passaggi consiste nel *semiparametric estimator for copula parameters* (*SP*). Questo metodo è simile all'IFM ed è stato introdotto da Genest, Ghoudi e Rivest (1995) e Shih e Louis (1995) e generalizzato da Tsukahara (2005). Aas et al. (2009) lo suggeriva per i modelli *PCC*. In questo metodo le distribuzioni marginali sono calcolate usando funzioni rango normalizzate. Questo metodo è più robusto di ML e IFM rispetto ad errori nella modellazione delle marginali (Kim, M. Silvapulle e P. Silvapulle (2007)). Dal punto di vista computazionale, questo metodo è paragonabile all'IFM.

5.3 Verosimiglianza di un modello *PCC* parametrico

Prima di trattare le metodologie specifiche per la *PCC*, è utile osservare come sia fatta la funzione di log-verosimiglianza per i modelli *PCC*.

A seguito della modellazione delle marginali, si definiscono le *pseudo*-osservazioni $\boldsymbol{u}$:

$$\boldsymbol{u} = (\boldsymbol{u}_1{}^{\intercal}, \ldots, \boldsymbol{u}_d{}^{\intercal}) \tag{5.8}$$

dove:

$$\begin{aligned} \boldsymbol{u}_k &= (F_1(x_{k,1}), \ldots, F_n(x_{k,n}))^{\intercal} \\ &= (u_{k,1}, \ldots, u_{k,n})^{\intercal} \end{aligned} \tag{5.9}$$

per $k = 1, \ldots, d$.

Al fine di semplificare la notazione, data una *R-vine* $\mathcal{V} = (T_1, \dots, T_n)$ si pone $E = E_1 \cup \cdots \cup E_{n-1}$. Ovvero quando si scriverà $e \in E$, si intenderà che $e \in E_i$ per un qualche $1 \leq i < n$.

Definizione 5.2. *La* verosimiglianza *di modello* PCC *parametrico* $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$ *con parametri di copula* $\boldsymbol{\theta} = \{\boldsymbol{\theta}_e, e \in E\}$ *e dati osservati* $\boldsymbol{u}$ *è definita come:*

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{u}) &= \prod_{k=1}^{d} \prod_{j=1}^{n-1} \prod_{e \in E_j} c_{a_e, b_e; \mathcal{D}_e}(C_{a_e|\mathcal{D}_e}(u_{k,a_e}|\boldsymbol{u}_{k,\mathcal{D}_e}), C_{b_e|\mathcal{D}_e}(u_{k,b_e}|\boldsymbol{u}_{k,\mathcal{D}_e})) \\ &= \prod_{k=1}^{d} \prod_{e \in E} c_{a_e, b_e; \mathcal{D}_e}(C_{a_e|\mathcal{D}_e}(u_{k,a_e}|\boldsymbol{u}_{k,\mathcal{D}_e}), C_{b_e|\mathcal{D}_e}(u_{k,b_e}|\boldsymbol{u}_{k,\mathcal{D}_e}))\end{aligned} \tag{5.10}$$

La log-verosimiglianza *di modello* PCC *parametrico* $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$ *con parametri di copula* $\boldsymbol{\theta} = \{\boldsymbol{\theta}_e, e \in E\}$ *e dati osservati* $\boldsymbol{u}$ *è definita come:*

$$\ell(\boldsymbol{\theta}; \boldsymbol{u}) = \log \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{u}) \tag{5.11}$$

Per semplificare la notazione, nell'equazione (5.10) è stata eliminata la dipendenza della verosimiglianza $\mathcal{L}_e$ di uno specifico arco $e \in E_j$ dai parametri di copula calcolati negli alberi precedenti, ovvero dai parametri degli archi in $E_1, \dots, E_{j-1}$. Nello specifico, se $e = (a_e, b_e; \mathcal{D}_e) = \{x, y\} \in E_j$ per $x, y \in E_{j-1}$, le funzioni $C_{a_e|\mathcal{D}_e}$ e $C_{b_e|\mathcal{D}_e}$ sono definite usando $c_{a_x,b_x|\mathcal{D}_x}$ e $c_{a_y,b_y|\mathcal{D}_y}$ e pertanto dipendono dai loro parametri. Attraverso di loro, la pair copula $c_{a_e,b_e;\mathcal{D}_e}$, oltre a dipendere dai parametri $\boldsymbol{\theta}_{a_e,b_e;\mathcal{D}_e}$ specifici del suo modello, dipenderà anche dai parametri $\boldsymbol{\theta}_{a_x,b_x;\mathcal{D}_x}$, $\boldsymbol{\theta}_{a_y,b_y;\mathcal{D}_y}$ e, ricorsivamente, dai parametri degli archi usati per generare x e y.

La natura di questa dipendenza deriva dal fatto che le pseudo-osservazioni usate per modellare la *pair-copula* per l'arco $e \in E_j$ non sia $\boldsymbol{u}$, ma il risultato di varie trasformazioni di $\boldsymbol{u}$ ad opera degli archi precedenti nella struttura. Perciò ogni deformazione ed errore nella modellazione negli archi precedenti si ripercuote a cascata sugli alberi successivi.

Ad esempio, se si considera la *R-vine* dell'esempio della figura 4.8 della sezione 4.2.3 e ripresa nella equazione (4.31), allora $C_{5|1,2,4}$ dipenderà da $c_{1,5;2,4}$, $c_{1,4;2}$, $c_{2,5;4}$, $c_{1,2}$, $c_{2,4}$ e $c_{4,5}$ e perciò anche dai loro parametri.

5.4 Stime dei parametri in un modello *PCC* parametrico

L'equazione (5.10) e la positività delle funzioni densità suggeriscono che sia possibile massimizzare la log-verosimiglianza ℓ massimizzando ogni arco tenendo solo conto della loro dipendenza reciproca.

La semplicità della forma della log-verosimiglianza dei modelli *PCC* parametrici non è condiviso da alcune delle metodologie alternative per generare copule multivariate. Per esempio, questo non è possibile per le copule archimedee gerarchiche (si veda per esempio Savu e Trede (2010) e O. Okhrin, Y. Okhrin e Schmid (2013)).

Nel caso di copule a *R-vine* semplificate in n-dimensioni con *pair-copule* a parametro singolo, il numero di parametri da essere massimizzati è $\binom{n}{2} = \frac{n(n-1)}{2}$,

che cresce quadraticamente. Similmente a quanto fatto per le funzioni marginali, ottimizzare i parametri sequenzialmente invece che tutti insieme è senz'altro computazionalmente più efficiente. Il risultato potrà quindi essere preso così com'è oppure come punto di partenza per un'ottimizzazione globale.

Si presenta ora lo *stepwise semiparametric estimator* (*SSP*). Questo stimatore è stato introdotto da Hobæk Haff (2013) ed ampliato e implementato in Stöber e Schepsmeier (2013). Inoltre Stöber e Schepsmeier (2013) hanno costruito una stima per l'errore standard dello stimatore di massima verosimiglianza per la copula. Gli autori hanno inoltre aggiunto la funzione per calcolare questo errore nella libreria `VineCopula`.

Nel seguito si denoteranno con $\boldsymbol{\theta}_e$ i parametri di copula corrispondenti all'arco $e = (a_e, b_e; \mathcal{D}_e) \in T_i$ e con $\boldsymbol{\theta}(T_i)$ l'insieme dei parametri corrispondenti a tutti gli archi di T_i e con $\widehat{\boldsymbol{\theta}}(T_i)$ le loro stime. Si userà inoltre la notazione $\widehat{\boldsymbol{\theta}}(T_1, \dots, T_{i-1})$ per le stime dei parametri degli alberi da T_1 a T_{i-1} ovvero:

$$\widehat{\boldsymbol{\theta}}(T_1, \dots, T_{i-1}) = \bigcup_{j=1}^{i-1} \widehat{\boldsymbol{\theta}}(T_j) \tag{5.12}$$

La stima sequenziale di $\boldsymbol{\theta}_e$ per un arco $e = (a_e, b_e; \mathcal{D}_e) \in T_i$ basata sulle *pseudo-osservazioni*

$$u_{k,a_e|\mathcal{D}_e,\widehat{\boldsymbol{\theta}}(T_1,\dots,T_{i-1})} = C_{a_e|D_e}\big(u_{k,a_e}|\boldsymbol{u}_{k,\mathcal{D}_e}, \widehat{\boldsymbol{\theta}}(T_1, \dots, T_{i-1})\big) \tag{5.13}$$

$$u_{k,b_e|\mathcal{D}_e,\widehat{\boldsymbol{\theta}}(T_1,\dots,T_{i-1})} = C_{b_e|D_e}\big(u_{k,b_e}|\boldsymbol{u}_{k,\mathcal{D}_e}, \widehat{\boldsymbol{\theta}}(T_1, \dots, T_{i-1})\big) \tag{5.14}$$

per $1 \leq k \leq d$, è effettuata massimizzando:

$$\prod_{k=1}^{d} c_{a_e,b_e;\mathcal{D}_e}\big(u_{k,a_e|\mathcal{D}_e,\widehat{\boldsymbol{\theta}}(T_1,\dots,T_{i-1})}, u_{k,b_e|\mathcal{D}_e,\widehat{\boldsymbol{\theta}}(T_1,\dots,T_{i-1})}; \boldsymbol{\theta}_e\big) \tag{5.15}$$

oppure invertendo il τ di Kendall empirica basata sulle *pseudo*-osservazioni.

La componente della funzione di log-verosimiglianza basata sulle marginali può essere calcolata usando metodi parametrici o non-parametrici in modo simile a quanto fatto nell'IFM oppure usando funzioni rango normalizzate come nel SP. Nelle applicazioni di questo elaborato sono state utilizzate principalmente le funzioni rango normalizzate oppure le funzionalità incluse nella libreria `rvinecopulib`.

5.5 Selezione delle famiglie in un modello *PCC* parametrico

Si tratterà ora il problema di identificare l'insieme $\mathcal{B}(\mathcal{V})$. Quest'ultimo è formato da $\binom{n}{2} = \frac{n(n-1)}{2}$ famiglie di copule bivariate e i loro rispettivi insiemi di parametri $\Theta(\mathcal{B}(\mathcal{V}))$.

Se nel caso delle stime dei parametri ci si trova con un problema di ottimizzazione numerico, la selezione delle famiglie è una questione di comparazione di modelli bivariati. Va inoltre osservato che anche se si parla di selezione di famiglie, in realtà si potrebbe scegliere ogni tipologia di modello per quella particolare copula. Per

esempio, la libreria `rvinecopulib` permette di usare modelli non-parametrici nei casi in cui nessuna delle famiglie supportate sia ritenuta all'altezza.

Perciò, assumendo che si sia scelta una struttura a *R-vine* $\mathcal{V}$, si vuole capire come individuare le famiglie di copule $C_{a_e,b_e;\mathcal{D}_e}$ e i rispettivi parametri per un dato campione $\boldsymbol{x}$.

Czado (2019) suggerisce di utilizzare il *criterio di informazione di Akaike* (`AIC`) introdotto in Akaike (1973)[4]. Si noti che la scelta del modello avviene contestualmente a quella dei suoi parametri conformemente allo stimatore SSP introdotto nella sezione 5.4. Questo criterio è stato seguito da Manner (2007) e Brechmann (2010). In particolare, Brechmann (2010) compara questa strategia contro altre tre opzioni:

- la selezione della famiglia con il più grande *p-value* del test di bontà di adattamento basato sulla statistica di Cramér-von Mises[5];
- con la minima distanza tra la caratteristica di dipendenza (τ_K di Kendall oppure dipendenza di code λ) empirica e del modello;
- con il maggior numero di "vittorie" nelle comparazioni delle famiglie utilizzando il test di Vuong (1989).

Brechmann (2010), nel suo studio Monte Carlo su grande scala, ha identificato nell'`AIC` il metodo più affidabile di selezione della famiglia di copule bivariate. Pertanto, si approfondirà nel seguito questa metodologia. Risulta comunque semplice adattare il discorso nel caso si faccia uso di un'altra metrica per comparare i modelli.

Sia $\mathcal{V} = (T_1, \dots, T_n)$ la struttura a *R-vine* su cui si vuole costruire la copula. Per prima cosa, si considerino le famiglie di copule e i relativi parametri per l'albero T_1. Per un arco arbitrario $e = (a_e, b_e) \in E_1$, si usino i dati di copula $u_{k,a_e} = F_{a_e}(x_{k,a_e})$ e $u_{k,b_e} = F_{b_e}(x_{k,b_e})$ per $1 \leq k \leq d$. In particolare, sia $\mathcal{B}_e$ l'insieme delle possibili famiglie di copule parametriche bivariate per l'arco e.

Per ogni elemento $C^B \in \mathcal{B}_e$ di densità di copula c^B, si usano i dati di copula $\boldsymbol{u}_e = \{u_{k,a_e}, u_{k,b_e}; k = 1, \dots, d\}$ per trovare la stima $\widehat{\boldsymbol{\theta}}^B$ di massima verosimiglianza per la copula C^B come descritto nella sezione 5.4. Quindi si calcola l'`AIC` corrispondente, ovvero si calcola:

$$AIC(C^B, \widehat{\boldsymbol{\theta}}^B; \boldsymbol{u}_e) = -2 \sum_{k=1}^{d} \ln\big(c^B(u_{k,a_e}, u_{k,b_e}; \widehat{\boldsymbol{\theta}}^B)\big) + 2k^B \tag{5.16}$$

dove $2k^B$ è la dimensione del vettore dei parametri $\boldsymbol{\theta}^B$. Successivamente, si seleziona la copula C^B di parametri $\boldsymbol{\theta}^B$ che minimizza $\texttt{AIC}(C^B, \widehat{\boldsymbol{\theta}}^B; \boldsymbol{u}_e)$.

Va osservato che questa minimizzazione richiede di stimare tutti i parametri per tutte le famiglie di copule contenute in $\mathcal{B}_e$.

Siccome le tipologie di copule più comunemente usate sono a uno o due parametri, un metodo di selezione alternativo come il `BIC` (Schwarz (1978)) non risulta necessario per ridurre il numero totale di parametri.

[4]Si analizzeranno più in dettaglio i criteri per la comparazione di modelli nella sezione 5.7. Rispetto alla comparazione dei modelli *PCC*, la comparazione di famiglie di copule bi-variate è molto più semplice perché il rapporto tra i vari parametri delle famiglie è conosciuto.

[5]Proposto da Genest, Quessy e Rémillard (2006) e Genest, Rémillard e Beaudoin (2009)

Per la selezione degli archi degli alberi T_i con $i \geq 2$, è necessario utilizzare il metodo di stima sequenziale introdotto nella sezione 5.4. Per un arco $e = (a_e, b_e; \mathcal{D}_e) \in E_i = E(T_i)$, si usano le pseudo-osservazioni per la distribuzione di copula bivariata di (U_{a_e}, U_{b_e}) dato $\boldsymbol{U}_{\mathcal{D}_e}$. Queste *pseudo*-osservazioni sono definite nella equazione (5.13) e che si abbrevia come:

$$\widehat{u}_{k,a_e} = u_{k,a_e|\mathcal{D}_e;\widehat{\boldsymbol{\theta}}(T_1,\ldots,T_{i-1})} \tag{5.17}$$

$$\widehat{u}_{k,b_e} = u_{k,b_e|\mathcal{D}_e;\widehat{\boldsymbol{\theta}}(T_1,\ldots,T_{i-1})} \tag{5.18}$$

per $1 \leq k \leq d$. Si considera quindi l'insieme $\mathcal{B}_e$ delle possibili famiglie di copule parametriche per l'arco $e \in E_i$. Gli pseudo-dati $\boldsymbol{u}_e = \{\widehat{u}_{k,a_e}, \widehat{u}_{k,b_e}\}$ sono usati per stimare i parametri di copula $\boldsymbol{\theta}^B$ usando la massima verosimiglianza per ogni elemento $C^B \in \mathcal{B}_e$. Successivamente, similmente a quanto fatto per T_1, si determina $\texttt{AIC}(C^B, \widehat{\boldsymbol{\theta}}^B; \boldsymbol{u}_e)$ e si selezione la copula che minimizza questo valore. Questa metodologia permette di selezionare contemporaneamente le famiglie di copule e stimarne i loro parametri.

Chiaramente, questo metodo sequenziale "monta" l'incertezza nella selezione delle famiglie di *pair-copule* e i loro parametri con il crescere dei livelli. Pertanto, il modello finale andrà controllato in modo attento e comparato con modelli alternativi. A questo scopo si rimanda a sezione 5.7.

5.6 Selezione della vite in un modello *PCC* parametrico

In questo paragrafo si analizza il caso in cui la struttura $\mathcal{V}$ sia da determinare. Come è evidente non è possibile semplicemente provare tutte le possibili definizioni $(\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$ di copule a *R-vine* e poi scegliere la migliore. Si ricorda infatti che il numero di *R*-viti su n variabili è $n!2^{\binom{n-2}{2}-1}$ (si veda Morales Napoles (2011) e sezione 4.2.4). Questo significa che se anche non dovessimo identificare le famiglie di *pair-copule* e i loro parametri, il numero di possibili strutture renderebbe comunque il calcolo computazionalmente non trattabile. Questa problematica permane persino se ci si restringe alle *C-vine* o *D-vine* dato che il loro numero è $n!2^{-1}$ (Aas et al. (2009) e sezione 4.2.4). Inoltre confrontare modelli alternativi è tutt'altro che semplice, come si vedrà nella sezione 5.7.

Un aspetto che va sottolineato è che, come è stato già osservato nella sezione 5.1, la *R-vine* non possiede alcuna componente statistica intrinseca. L'aspetto decisionale degli algoritmi viene dunque spostato a qualcosa di esterno alla struttura. È dunque possibile separare la componente dell'algoritmo che genera gli elementi della *R-vine* e la parte decisionale che decide quale elemento scegliere tra quelli possibili.

Dato l'enorme numero di *R-vine* è normale che si cerchi rifugio nella classe degli algoritmi *greedy*, anche al costo di perdere l'ottimalità del risultato. Prima di analizzarli in dettaglio è utile richiamare il significato di algoritmo *greedy*.

Chiunque abbia studiato un pò di algoritmi ha una idea intuitiva di cosa siano gli algoritmi *greedy*, spesso attorniata da una serie di esempi, ma come ci ricorda Kleinberg e Tardos (2006)

> *It is hard, if not impossible, to define precisely what is meant by a greedy algorithm. An algorithm is greedy if it builds up a solution in small steps,*

> *choosing a decision at each step myopically to optimize some underlying criterion. One can often design many different greedy algorithms for the same problem, each one locally, incrementally optimizing some different measure on its way to a solution.*

Altri manuali, come Cormen et al. (2022), fanno un tentativo che si distribuisce su varie pagine e che esula senz'altro dallo scopo di questa tesi.

Si adotterà la seguente definizione[6] (ispirata al passaggio appena citato da Kleinberg e Tardos (2006)):

Definizione 5.3. *Un algoritmo* `A` *per risolvere un problema* `P` *si dice* greedy *se:*

1. *Il problema* `P` *può essere rappresentato come una sequenze di scelte* $S_1, S_2, \ldots, S_m$*;*
2. *Le scelta nel sottoproblema* S_j *dipende solo dalle scelte fatte precendentemente e l'algoritmo non torna mai sui suoi passi.*

Perciò, nel caso della generazione di *R-vine*, usare un algoritmo *greedy* vuol dire costruire la *vine* incrementalmente (per esempio un albero alla volta) usando nella scelta dei nuovi elementi solo gli alberi o archi già generati ed, eventualmente, gli elementi stocastici ad essi collegati.

Czado (2019) separa gli algoritmi *greedy* per la generazione delle *R-vine* in algoritmi *bottom-up* e algoritmi *top-down* in base all'ordine in cui gli alberi sono generati. Questa separazione possiede due difetti sostanziali:

1. ignora la possibilità che la *vine* sia generata con un ordine differente;
2. dipende da quale albero si considera in cima, se T_1 oppure T_{n-1}.

Infatti, Kurowicka (2010) presenta il suo algoritmo come *top-down* mentre Czado (2019) lo annovera tra quelli *bottom-up*.

Czado (2019) chiama *top-down* gli algoritmi che generano prima T_1 e poi "scendono" verso T_{n-1}. Gli algoritmi *bottom-up* sono invece quelli che fanno l'opposto.

Czado (2019) cita due algoritmi per generare le R-viti, anche se altri algoritmi sono stati proposti nel tempo:

1. l'algoritmo *top-down* di Dißmann et al. (2013);
2. l'algoritmo *bottom-up* di Kurowicka (2010).

Il primo, l'algoritmo di Dißmann et al. (2013), è quello più diffuso e utilizzato, ed attualmente è quello implementato per la costruzione *building blocks* della *pair-copula construction* sia la libreria `VineCopula` che `rvinecopulib`.

Nella sezione 4.2.3 si è costruita una *R-vine* usando un approccio *top-down*. L'approccio che viene adottato in questa trattazione ha molto in comune con l'algoritmo di Dißmann et al. (2013) e la vera distinzione tra i due risiede solamente nella modalità di scelta.

Un'aspetto fondamentale su cui porre attenzione è che la generazione dei parametri e delle famiglie nelle sezioni 5.4 e 5.5 è stata anch'essa *top-down*. Pertanto, si

[6] In alcuni manuali si suppone che la funzione dell'algoritmo *greedy* sia deterministica, ma nel questa trattazione caso questa ipotesi ci sembra eccessivamente restrittiva.

Algoritmo 5.1: Algoritmo di Dißmann et al. (2013).

input : Il vettore $\boldsymbol{u} = \{\boldsymbol{u}_i | 1 \leq i \leq d\}$ delle osservazioni
output: Una copula *R-vine* che modelizza il vettore $\boldsymbol{u}$

$\mathcal{E}_1 \leftarrow [\{1, \ldots, n\}]^2$;
Calcola i pesi $\omega_{i,j}$ per ogni $\{i, j\} \in \mathcal{E}_1$;
$T_1 \leftarrow$ `trova_albero_ricomprente_ottimo`$(\mathcal{E}_1, \{\omega_{i,j}\})$;
per ciascun $\{a_e, b_e; \emptyset\} \in E(T_1)$ **fai**
  $C_{a_e,b_e}, \widehat{\boldsymbol{\theta}}_{a_e,b_e} \leftarrow$ `seleziona_copula_e_parametro`$(a_e, b_e, \emptyset)$;
  Calcola pseudo-osservazioni dalla copula $C_{a_e,b_e}(\widehat{\boldsymbol{\theta}}_{a_e,b_e})$;
fine
per $m \leftarrow 2$ **a** $n - 1$ **fai**
  $\mathcal{E}_m \leftarrow$ `trova_archi_con_condizione_prossimità`$(E(T_{m-1}))$;
  Calcola i pesi $\omega_{a_e,b_e;\mathcal{D}_e}$ per ogni $\{a_e, b_e; \mathcal{D}_e\} \in \mathcal{E}_m$;
  $T_m \leftarrow$ `trova_albero_ricomprente_ottimo`$(\mathcal{E}_1, \{\omega_{i,j}\})$;
  per ciascun $\{a_e, b_e; \mathcal{D}_e\} \in E(T_m)$ **fai**
    $C_{a_e,b_e;\mathcal{D}_e}, \widehat{\boldsymbol{\theta}}_{a_e,b_e;\mathcal{D}_e} \leftarrow$ `seleziona_copula_e_parametro`$(a_e, b_e, \mathcal{D}_e)$;
    Calcola pseudo-osservazioni dalla copula $C_{a_e,b_e;\mathcal{D}_e}(\widehat{\boldsymbol{\theta}}_{a_e,b_e;\mathcal{D}_e})$;
  fine
fine
$\widehat{\mathcal{V}} \leftarrow (T_1, \ldots, T_{n-1})$;
$\widehat{\mathcal{B}} \leftarrow \{C_{a_e,b_e;\mathcal{D}_e} | \{a_e, b_e; \mathcal{D}_e\} \in E(\widehat{\mathcal{V}})\}$;
$\widehat{\boldsymbol{\theta}} \leftarrow \{\widehat{\boldsymbol{\theta}}_{a_e,b_e;\mathcal{D}_e} | \{a_e, b_e; \mathcal{D}_e\} \in E(\widehat{\mathcal{V}})\}$;
ritorna $(\widehat{\mathcal{V}}, \widehat{\mathcal{B}}, \widehat{\boldsymbol{\theta}})$

può usare quando detto in quelle sezioni anche se non si ha ancora generato l'intera struttura, perché la parte della struttura che conta è già stata generata.

Per comprendere l'algoritmo si deve quindi esplicitare in cosa consistano i sottoproblemi e come faccia la scelta. Si supponga che l'algoritmo abbia generato fino all'albero T_j, con $j \geq 1$, e si voglia generare l'albero T_{j+1}. Dalla teoria del capitolo 4, si sa che $T_{j+1} \subset L(T_j)$ e per ogni $e = (a_e, b_e; \mathcal{D}_e) \in L(T_j)$ è possibile trovare la copula corrispondente $C_{a_e,b_e;\mathcal{D}_e}$ usando i metodi descritti nella sezione 5.5.

Quello che fa l'algoritmo di Dißmann et al. (2013) è associare a $e \in L(T_j)$ un peso ω_e e trovare l'albero $T_{j+1} \subset L(T_j)$ che massimizzi la funzione $\omega(T_j) = \sum_{e \in T_j} \omega_e$. Per il primo albero si può fare la stessa cosa, semplicemente si sostituisce $L(T_j)$ con il grafo completo degli elementi $\{1, \ldots, n\}$. Si noti che la funzione peso può essere o meno dipendente dalla copula $C_{a_e,b_e;\mathcal{D}_e}$.

L'algoritmo appena descritto è esplicitato nell'algoritmo 5.1 dove:

- `seleziona_copula_e_parametro` seleziona le famiglie e i parametri usando la teoria presentata nella sezione 5.5;
- `trova_archi_con_condizione_prossimità` calcola il grafo di linea $L(T_j)$ dall'albero T_j;
- `trova_albero_ricomprente_ottimo` trova l'albero che massimizza $\omega(T_j)$.

L'albero ricoprente massimale può essere trovato usando algoritmi classici come quello di Prim o Kruskal (si guardi per esempio Cormen et al. (2022)).

Chiaramente questo algoritmo fa scelte localmente ottimali, siccome l'impatto degli alberi precedenti e successivi è ignorato. La strategia, però, è ragionevole riguardo la modellazione statistica con copule a *R-vine* perché le dipendenze più importanti misurate dalla funzione peso ω sono modellate per prime.

La letteratura ha presentato varie opzioni per la funzione peso ω_e. Una possibilità è quella di usare la log-verosimiglianza ℓ_e delle pair-copule $C_{a_e,b_e;\mathcal{D}_e}$. Ovvero si selezionerebbe l'albero che massimizza la log-verosimiglianza delle copule che lo compongono. Questa soluzione però, porta a modelli con troppo parametri perché un numero maggiore di parametri tende ad aumentare la verosimiglianza. Per esempio, una t-Student avrà sempre una maggiore verosimiglianza di una Gaussiana dato che quest'ultima è un caso speciale della prima. Tra le possibili scelte per la funzione ω si citano le seguenti:

- il τ di Kendall assoluto empirico (proposto da Dißmann et al. (2013) e Czado, Schepsmeier e Min (2012));
- l'`AIC` di ogni *pair-copula*;
- il grado di libertà stimato (negativo) di una copula t di Student come proposto in Mendes, Semeraro e Leal (2010);
- il *p-value* di un test di adattamento di copula e le sue varianti come proposto in Czado, Jeske e Hofmann (2013).

Gruber e Czado (2015) hanno mostrato che in scenari a 6-dimensioni, l'algoritmo di Dißmann et al. (2013), usando il τ di Kendall assoluto empirico come funzione ω,

è stato capace di produrre copule con un log-verosimiglianza almeno del 75% della reale log-verosimiglianza.

L'algoritmo di Dißmann et al. (2013) può essere adattato per considerare solo C-viti o D-viti invece che tutte le R-viti. Per C-viti il nodo centrale può essere scelto come il nodo con la massima somma dei pesi, comparato con gli altri (Czado, Schepsmeier e Min (2012)). Nel caso delle D-viti, solo l'ordine del primo albero necessita di essere scelto. Siccome gli alberi a D-viti sono percorsi Hamiltoniani, per la scelta di T_1 è necessario idenficare il massimo percorso hamiltoniamo. Questo problema è equivalente al problema del venditore, come discusso in Brechmann (2010). Siccome quest'ultimo è un problem `NP-hard`, non si conoscono algoritmi efficienti per risolverlo. Esistono però alcuni algoritmi approssimati che possono essere sfruttati.

Siccome l'algoritmo di Kurowicka (2010) genera la *R-vine* a partire da T_{n-1}, è evidente che non possa utilizzare alcuna informazione sulla struttura stocastica del modello. Questo rende l'algoritmo di Kurowicka (2010) molto meno utilizzabile nella pratica.

Kurowicka (2010) propone comunque un metodo per spingere il modello verso una struttura con il massimo numero di copule indipendenti negli ultimi alberi (ovvero i primi che sono generati dall'algoritmo). Il metodo proposto da Kurowicka (2010) consiste nell'usare le correlazioni parziali calcolate usando la matrice di correlazione ottenuta dai dati. Le correlazioni parziali sono calcolate usando la seguente formula:

$$\rho_{i,j;I\setminus\{i,j\}} = \frac{C_{i,j}}{\det(I)} \tag{5.19}$$

dove $\det(I)$ denota il determinante della matrice di correlazione delle variabili in I e $C_{i,j}$ è l'(i,j)-esima matrice dei cofattori. Questa scelta è legata al fatto che per le distribuzioni ellittiche, queste correlazioni parziali corrispondono alle correlazioni condizionate. Seppur una correlazione parziale uguale a zero non corrisponda ad indipendenza condizionata, il suggerimento di Kurowicka (2010) è di minimizzare le correlazioni parziali.

5.7 Comparazione di modelli *PCC* parametrici

Per quanto la variabilità insita in un modello *PCC* possa far pensare che sia capace di rappresentare ogni popolazione, non si deve mai dimenticare che ogni modello è approssimato. A questo scopo diventa essenziale avere a disposizione dei buoni indicatori sulla bontà del modello.

Un'altro aspetto che non va sottovalutato è il *trade-off* tra numero di parametri e capacità espressiva del modello. Se infatti aumentare il numero di parametri porta spesso ad una maggiore capacità espressiva del modello, dall'altro aumenta il rischio di *overfitting* e aumenta la complessità del modello rendendo più difficile una sua analisi.

L'applicabilità di queste metodologie spesso dipende se i modelli sono annidati o meno.

Definizione 5.4. *Siano* $\mathcal{M}_1$ *e* $\mathcal{M}_2$ *due modelli con insiemi di parametri* Θ_1 *e* Θ_2*. Si dice che i due modelli sono* annidati *se* $\Theta_1 \subset \Theta_2$ *oppure* $\Theta_2 \subset \Theta_1$*.*

I due modelli sono sovrapposti *se i due modelli non sono annidati e* $\Theta_1 \cap \Theta_2 \neq \emptyset$. *Due modelli non annidati o sovrapposti si dicono* strettamente non annidati.

Nel caso dei modelli *PCC*, i modelli sono strettamente non annidati solo quando le due viti non possiedono alcun arco in comune (o alternativamente si usino modelli distinti per lo stesso arco).

Proposizione 5.1. *Siano* $\mathcal{M}_1$ *e* $\mathcal{M}_2$ *due modelli* PCC *sulla stessa popolazione. Siano inoltre* $\mathcal{V}_1 = (T_{1,1}, \dots, T_{1,n})$ *e* $\mathcal{V}_2 = (T_{2,1}, \dots, T_{2,n})$ *le loro strutture a* R-vine. *Allora,* $\mathcal{M}_1$ *e* $\mathcal{M}_2$ *sono strettamente non annidati se e solo* $E_{1,1} \cap E_{2,1} = \emptyset$.

I modelli sono invece annidati quando uno dei due modelli può essere creato dall'altro sostituendo delle copule non indipendenti con copule indipendenti. Un caso particolare in cui questo avviene sono i modelli prodotti per troncamento di altri, dove un modello troncato è definito come segue:

Definizione 5.5. *Un* modello *PCC* troncato $\mathcal{M}$ *è un modello che usa solo copule indipendenti da un certo albero* T_m *in poi.*

5.7.1 Criteri di informazione di Akaike e Bayesiano

Un *criterio di informazione* di un modello è un qualsiasi meccanismo che usa i dati per dare un voto numerico a ciascun modello. Questo tipo di criteri sono alla base di quasi tutti i metodi di selezione dei modelli (si veda per esempio Claeskens e Hjort (2008)). Tra questi criteri, il *criterio di informazione di Akaike* (`AIC` in seguito) e il *criterio di informazione Bayesiano* (`BIC` in seguito) sono senz'altro i più usati e conosciuti. Il primo è stato introdotto in Akaike (1973) mentre il secondo in Schwarz (1978).

Definizione 5.6. *Data un campione* $\boldsymbol{u}$ *e un modello* PCC $\mathcal{M} = (\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$, *il* criterio di informazione di Akaike *(`AIC`) è definito come:*

$$\begin{aligned} AIC(\mathcal{M}) &= -2 \max_{\boldsymbol{\theta} \in \Theta(\mathcal{B}(\mathcal{V}))} \ell(\boldsymbol{\theta}; \boldsymbol{u}) + 2K \\ &= -2\ell(\widehat{\boldsymbol{\theta}}; \boldsymbol{u}) + 2K \end{aligned} \tag{5.20}$$

dove K *è il numero di parametri del modello, ovvero la lunghezza del vettore* $\widehat{\boldsymbol{\theta}}$.

Definizione 5.7. *Data un campione* $\boldsymbol{u}$ *e un modello* PCC $\mathcal{M} = (\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$, *il* criterio di informazione di Bayesiano *(`BIC`) è definito come:*

$$\begin{aligned} BIC(\mathcal{M}) &= -2 \max_{\boldsymbol{\theta} \in \Theta(\mathcal{B}(\mathcal{V}))} \ell(\boldsymbol{\theta}; \boldsymbol{u}) + K \ln d \\ &= -2\ell(\widehat{\boldsymbol{\theta}}; \boldsymbol{u}) + K \ln d \end{aligned} \tag{5.21}$$

dove K *è il numero di parametri del modello, ovvero la lunghezza del vettore* $\widehat{\boldsymbol{\theta}}$.

Czado (2019) osserva che questi criteri trovano solitamente uso nella comparazione tra modelli annidati ma che, secondo Ripley (1996), il criterio è comunque utilizzabile negli altri casi ma al costo di una maggiore variabilità della stima `AIC`.

È utile osservare che, similmente alla log-verosimiglianza, i due criteri di informazione si possono decomporre nella somma dei criteri di informazione relativi ad ogni singolo arco.

5.7.2 Modified Bayesian Information Criterion `mBIC`

Poichè il numero di parametri in un modello *PCC* cresce quadraticamente all'aumentare della dimensione d del modello, i criteri `BIC` e `AIC` si sono rivelati insoddisfacenti quando la creazione di modelli sparsi sia auspicata.

Definizione 5.8. *Un* modello *PCC* sparso $\mathcal{M}$ *è un modello in cui il numero di copule indipendenti è molto alto in rapporto al numero totale di copule del modello.*

Si osservi che nella definizione non vi è un limite numerico preciso per cosa significhi alto. Nella pratica si sceglie un certo rapporto oltre il quale si cerca di non andare.

L'utilizzo di modelli sparsi potrebbe apparire strano, ma è essenziale in settori come quello finanziario. Infatti, come osservano Nagler, Bumann e Czado (2019), un modello con 50 asset modellato con copule a due parametri possiede 2450 parametri possibili, un modello con 200 asset ne ha 20000, mentre la quantità di eventi disponibili in un determinato intervallo spaziato è finito.

Nagler, Bumann e Czado (2019) propongono un criterio alternativo al `BIC` che possiede un elemento di correzione per rendere i modelli più sparsi più appetibili.

Per farlo, Nagler, Bumann e Czado (2019) introducono una probabilità a priori ψ_e ad ogni arco e nella *vine* del modello. Più precisamente, gli autori assumono che le funzioni indicatrici $I_e = \mathbb{1}_{\boldsymbol{\theta}_e=0}$ sono variabili bernoulliane indipendenti con media ψ_e e propongono di scegliere $\psi_e = \psi_0^m$ per ogni $e \in T_m$. La risultante probabilità a priori è

$$\psi(\boldsymbol{\theta}) = \prod_{m=1}^{n-1} \psi_0^{mq_m} (1-\psi_0^m)^{n-m-q_m} \tag{5.22}$$

dove q_m è il numero di copule non indipendenti nell'albero T_m.

Definizione 5.9. *Dato un campione* $\boldsymbol{u}$ *e un modello* PCC $\mathcal{M} = (\mathcal{V}, \mathcal{B}(\mathcal{V}), \Theta(\mathcal{B}(\mathcal{V})))$*, il* criterio di informazione modificato di Bayesiano per Viti *(`mBIC`[7]) è definito come:*

$$mBICV(\mathcal{M}) = -2\ell(\widehat{\boldsymbol{\theta}}; \boldsymbol{u}) + K \ln d - 2 \sum_{m=1}^{n-1} \left[\widehat{q}_m \ln \psi_0^m + (n-m-\widehat{q}_m) \ln(1-\psi_0^m) \right] \tag{5.23}$$

dove K è il numero di parametri del modello e $\widehat{q}_m$ è il numero di copule non-indipendenti in T_m per il modello con parametro $\widehat{\boldsymbol{\theta}}$.

Attraverso vari casi di studio, Nagler, Bumann e Czado (2019) hanno inolte mostrato come questo nuovo criterio sia effettivamente in grado di costruire modelli sparsi in maniera più efficente di metodi alternativi.

5.7.3 Il test di comparazione di Vuong

Siccome si è interessati a comparare modelli non annidati, si introduce ora un metodo alternativo proposto da Vuong (1989). Al contrario dei criteri di informazione, il test di Vuong è un test statistico ovvero un test che assegna un certo livello di significatività trai due modelli.

[7] Modified Bayesian Information Criterion.

Prima di introdurre questo test, si segnala la necessità di introdurre un altro criterio di comparazione: il *criterio di Kullback-Leibler* (`KLIC`) introdotto in Kullback e Leibler (1951).

Il criterio misura la distanza tra un corretta ma sconosciuta densità h_0 e una densità parametrica approssimata $f(\bullet|\boldsymbol{\theta})$ con parametro $\boldsymbol{\theta}$. Nota che il parametro $\boldsymbol{\theta}$ deve coincidere con il parametro della densità h_0 nel caso in cui h_0 sia una densità parametrica. Nel seguito si seguirà l'esposizione di Czado (2019) e Vuong (1989).

Definizione 5.10. *Il* criterio di comparazione di Kullback-Leibler *(`KLIC`) tra la densità $h_0(\bullet)$ di una variabile aleatoria $\boldsymbol{X}$ e una sua approssimazione parametrica $f(\bullet, \boldsymbol{\theta})$ è definita come:*

$$KLIC(h_0, f, \boldsymbol{\theta}) = \int \ln\left[\frac{h_0(\boldsymbol{x})}{f(\boldsymbol{x}|\boldsymbol{\theta})}\right] h_0(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \tag{5.24}$$

$$= \mathbb{E}_0\big[\ln h_0(\boldsymbol{X})\big] - \mathbb{E}_0\big[\ln f(\boldsymbol{X}|\boldsymbol{\theta})\big] \tag{5.25}$$

dove $\mathbb{E}_0$ denota il valore atteso rispetto alla densità h_0.

È evidente che, per il `KLIC`, la scelta ottimale di $\boldsymbol{\theta}$ è quella che massimizza $\mathbb{E}_0\big[\ln f(\boldsymbol{X}|\boldsymbol{\theta})\big]$. Date d osservazioni indipendenti e identicamente distribuite $\{\boldsymbol{x}_k\}_{1\leq k\leq d}$ di densità h_0, il termine $\mathbb{E}_0\big[\ln f(\boldsymbol{X}|\boldsymbol{\theta})\big]$ può essere stimato usando la media campionaria:

$$\frac{1}{d}\sum_{k=1}^{d} \ln f(\boldsymbol{x}_k|\boldsymbol{\theta}). \tag{5.26}$$

e quindi si può ricavare $\widehat{\boldsymbol{\theta}}_d$ calcolando la massima verosimiglianza sull'equazione (5.26). Pertanto il minimo di $KLIC(h_0, f, \boldsymbol{\theta})$ può essere stimato da $KLIC(h_0, f, \widehat{\boldsymbol{\theta}}_d)$.

Osservazione 5.1. *Il criterio di Kullback-Leibler è un caso speciale della divergenza di Kullback-Leibler tra due densità h_0 e h_1, definita come:*

$$KLIC(h_0, h_1) = \mathbb{E}_0\left(\ln\left[\frac{h_0(\boldsymbol{X})}{h_1(\boldsymbol{X})}\right]\right) \tag{5.27}$$

per $h_1(\boldsymbol{x}) = f(\boldsymbol{x}|\boldsymbol{\theta})$.

L'obiettivo di Vuong (1989) era quello di analizzare se due classi di modelli $\mathcal{M}_1$, $\mathcal{M}_2$ definite da due densità parametriche $\{f_1(\bullet|\boldsymbol{\theta}_1)|\boldsymbol{\theta}_1 \in \Theta_1\}$ e $\{f_2(\bullet|\boldsymbol{\theta}_2)|\boldsymbol{\theta}_2 \in \Theta_2\}$ forniscono fit equivalenti per un dato *dataset* oppure se uno dei modelli era da preferire all'altro.

In particolare, Vuong (1989) utilizza il `KLIC` definito nella definizione 5.10 e costruisce un test asintotico del rapporto di verosimiglianza con l'ipotesi nulla di equivalenza tra i due modelli:

$$H_0 : KLIC(h_0, f_1, \bar{\boldsymbol{\theta}}_1) = KLIC(h_0, f_2, \bar{\boldsymbol{\theta}}_2). \tag{5.28}$$

dove $\bar{\boldsymbol{\theta}}_1$ e $\bar{\boldsymbol{\theta}}_2$ i valori che massimizzano $\mathbb{E}_0(\ln f_1(\boldsymbol{X}|\boldsymbol{\theta}_1))$ e $\mathbb{E}_0(\ln f_2(\boldsymbol{X}|\boldsymbol{\theta}_2))$. Usando equazione (5.24), l'ipotesi nulla può essere riscritta come:

$$H_0 : \mathbb{E}_0\big[\ln f_1(\boldsymbol{X}|\bar{\boldsymbol{\theta}}_1)\big] = \mathbb{E}_0\big[\ln f_2(\boldsymbol{X}|\bar{\boldsymbol{\theta}}_2)\big]. \tag{5.29}$$

Ovviamente, nel caso in cui l'ipotesi alternativa:

$$H_{m1} : \mathbb{E}_0[\ln f_1(\boldsymbol{X}|\boldsymbol{\theta}_1)] > \mathbb{E}_0[\ln f_2(\boldsymbol{X}|\boldsymbol{\theta}_2)] \tag{5.30}$$

venga soddisfatta, il modello $\mathcal{M}_1$ è migliore del modello $\mathcal{M}_2$ e il contrario vale per l'ipotesi alternativa:

$$H_{m2} : \mathbb{E}_0[\ln f_1(\boldsymbol{X}|\boldsymbol{\theta}_1)] < \mathbb{E}_0[\ln f_2(\boldsymbol{X}|\boldsymbol{\theta}_2)] \tag{5.31}$$

Si cerca ora di capire come determinare se un modello sia migliore rispetto ad un altro in modo significativo. Un modo per testare l'ipotesi nulla è utilizzare il rapporto di verosimiglianza.

Definizione 5.11. *Si considerino d osservazioni indipendenti e identicamente distribuite $\boldsymbol{x}_k$ con $1 \leq k \leq d$ e densità h_0. Siano $\widehat{\boldsymbol{\theta}}_1$ e $\widehat{\boldsymbol{\theta}}_2$ le stime di massima log-verosimiglianza associate ai modelli $\mathcal{M}_1$ e $\mathcal{M}_2$ di densità $f_1(\bullet|\boldsymbol{\theta}_1)$ e $f_2(\bullet|\boldsymbol{\theta}_2)$, allora il k-*esimo contibuto al rapporto di *log*-verosimiglianza *è definito come:*

$$m_k(\boldsymbol{x}_k) = \ln\left[\frac{f_1(\boldsymbol{x}_k|\widehat{\boldsymbol{\theta}}_1)}{f_2(\boldsymbol{x}_k|\widehat{\boldsymbol{\theta}}_2)}\right] \tag{5.32}$$

per $k = 1, \ldots, d$.

La statistica del rapporto di verosimiglianza osservato *per i due modelli approssimati $\mathcal{M}_1$ e $\mathcal{M}_2$ è definito come:*

$$LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{x}) = \sum_{k=1}^{d} m_k(\boldsymbol{x}_k) \tag{5.33}$$

Si noterà inoltre con $m_k(\boldsymbol{X}_k)$ e $LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{X})$ le variabili aleatorie associate, dove $\boldsymbol{X}_k \sim h_0$ i.i.d.

Nella distribuzione h_0 e sotto appropriate condizioni di regolarità sui due modelli, il rapporto di verosimiglianza normalizzato $\frac{1}{d}LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{x})$ è una variabile aleatoria che soddisfa la seguente legge dei grandi numeri.

Teorema 5.1 (Legge dei grandi numeri per statistiche LR)**.** *Sotto opportune ipotesi di regolarità che assicurino l'esistenza e l'unicità dei vari valori (si vedano le ipotesi A1-A3 di Vuong (1989)), si ha che*

$$\frac{1}{d}LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{x}) \xrightarrow[d\to\infty]{q.o} \mathbb{E}_0\left(\ln \frac{f_1(\boldsymbol{X}|\boldsymbol{\theta}_1)}{f_2(\boldsymbol{X}|\boldsymbol{\theta}_2)}\right). \tag{5.34}$$

Appurato il comportamento asintotico della statistica, Vuong (1989) né considera la varianza campionaria di $LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{x})$:

$$\widehat{\omega}(\boldsymbol{x})^2 = \frac{1}{d}\sum_{k=1}^{d}\left(\boldsymbol{x}_k - \overline{\boldsymbol{x}}_d\right)^2 \tag{5.35}$$

dove:

$$\overline{\boldsymbol{x}}_d = \frac{1}{d}\sum_{k=1}^{d}\boldsymbol{x}_k \tag{5.36}$$

e la usa per ottenere la distribuzione asintotica di $LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{X})$.

Teorema 5.2. *Sotto opportune ipotesi di regolarità (si veda il teorema 5.1 di Vuong (1989)) e quando i modelli* $\mathcal{M}_1$ *e* $\mathcal{M}_2$ *sono strettamente non annidati, si ha che:*

$$\nu(\boldsymbol{X}) = \frac{LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{X})}{\sqrt{d\widehat{\omega}(\boldsymbol{X})^2}} \xrightarrow[\mathcal{D}]{d\to\infty} \mathcal{N}(0,1). \tag{5.37}$$

Questo teorema porta ad un test asintotico chiamato *test di Vuong* per la selezione del modello tra modelli strettamente non-annidati. Siccome il test di Vuong è un test statistico, assegna un livello di significatività alle proprie decisioni.

Definizione 5.12. *Un livello asintotico* α *per il test di Vuong con ipotesi nulla:*

$$H_0 : \mathbb{E}_0\big[\ln f_1(\boldsymbol{X}|\boldsymbol{\theta}_1)\big] = \mathbb{E}_0\big[\ln f_2(\boldsymbol{X}|\boldsymbol{\theta}_2)\big] \tag{5.38}$$

contro l'ipotesi alternativa H_1*, è dato dal rigettare* H_0 *se e solo se:*

$$|\nu(\boldsymbol{X})| > \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big). \tag{5.39}$$

dove Φ *è la funzione di distribuzione di una normale standard.*

In particolare, se $\nu(\boldsymbol{X}) > \Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)$ allora il modello $\mathcal{M}_1$ sarà da preferire al modello $\mathcal{M}_2$. Infatti, in questo caso, il test indicherebbe che il $KLIC$ del modello $\mathcal{M}_1$ sarà significativamente minore di quello del modello $\mathcal{M}_2$. Nel caso in cui $\nu(\boldsymbol{X}) < -\Phi^{-1}\Big(1 - \frac{\alpha}{2}\Big)$, sarà invece da preferire il secondo modello.

Fino a questo momento sono stati introdotti i concetti generali del test di Vuong. Si supponga ora che $\mathcal{M}_1$ e $\mathcal{M}_2$ siano due modelli PCC. Per poter utilizzare il teorema 5.2, i due modelli devono essere strettamente non annidati.

Nel caso in cui questo non avvenga, ovvero non valga la proposizione 5.1, non si può usare il teorema 5.2 e le cose si complicano. Vuong (1989), nel teorema 6.3, mostra alcuni risultati utilizzabili in questo caso, ma l'analisi risulta senz'altro meno immediata.

Nel caso di modelli annidati, Vuong (1989) osserva che i test standard sul rapporto di verosimiglianza sono utilizzabili. Il caso particolare dei modelli troncati è stato studiato da Brechmann, Czado e Aas (2012). Siccome la distribuzione asintotica del rapporto di verosimiglianza standard non è trattabile, Brechmann, Czado e Aas (2012) hanno utilizzato il test di Vuong suggerito per il caso non annidato. Gli autori hanno motivato il loro approccio con un ampio studio simulativo e mostrato che lo strumento ha dato performance soddisfacenti per campioni finiti.

Il test definito in definizione 5.12 non tiene conto della possibilità che i due modelli possano avere un numero differente di parametri ovvero ignora la complessità del modello. Questa versione del test è quindi definita *non-aggiustata* e Vuong (1989) definisce il test *aggiustata* come:

$$\widetilde{LR}_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{X}) = LR_d(\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2)(\boldsymbol{X}) - K_d(f_1, f_2) \tag{5.40}$$

dove $K_d(f_1, f_2)$ è un *fattore di correzione* che dipende dalle caratteristiche di $\mathcal{M}_1$ e $\mathcal{M}_2$ e si assume che soddisfi:

$$\frac{K_d(f_1, f_2)}{d} = o_p(1) \tag{5.41}$$

dove $Z_d = o_p(1)$ è l'insieme delle variabili aleatore $(Z_d)_{d\in\mathbb{N}}$ corrispondenti alla convergenza a zero in probabilità, ovvero tali che $\lim_{d\to\infty} \mathbb{P}\{|Z_d| > \varepsilon\} = 0$ per ogni $\varepsilon > 0$. Quando questo avviene, il risultato asintotico del teorema 5.2 è verificato anche per $\widetilde{LR}_d$ ed è quindi possibile ridefinire il test anche per $\widetilde{LR}_d$. In particolare, Vuong (1989) osserva che in termini della statistica LR non aggiustata non è possibile rigettare l'ipotesi nulla con un livello α se e solo se

$$-\Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + \frac{K_d(f_1,f_2)}{\sqrt{d\widehat{\omega}^2}} \leq \nu(\boldsymbol{x}) \leq \Phi^{-1}\Big(1-\frac{\alpha}{2}\Big) + \frac{K_d(f_1,f_2)}{\sqrt{d\widehat{\omega}^2}} \tag{5.42}$$

dove ν è il rapporto scalato di verosimiglianza definito nel teorema 5.2.

Vuong (1989) suggerisce due possibili definizioni per K_d, ispirati rispettivamente ad `AIC` e `BIC`:

- *correzione di Akaike*: $K_d^A(f_1,f_2) = k_1 - k_2$;
- *correzione di Schwarz*: $K_d^S(f_1,f_2) = \frac{k_1}{2}\ln d - \frac{k_2}{2}\ln d$.

dove k_1 e k_2 sono il numero di parametri del modelli $\mathcal{M}_1$ e $\mathcal{M}_2$. La correzione di Schwarz porta alla scelta di modelli più parsimoniosi in termini di parametri. È utile osservare che la scelta di quelle funzioni è arbitraria e altre scelte potrebbero essere più appropriate in specifiche applicazioni.

È importante inoltre osservare che lo scopo del test di Vuong e dei criteri di `AIC` e `BIC` sia quello di comparare modelli. Questi criteri non possono però fornire indicazioni sulla effettiva bontà dei due modelli.

5.7.4 Il test di comparazione di Clarke

Un test alternativo a quello di Vuong è rappresentato da quello di Clarke, introdotto in Clarke (2007). Il test di Clarke è basato su quello di Vuong, e dunque sulla stessa teoria.

Se l'ipotesi H_0 del test di Vuong è corretta allora:

$$\mathbb{P}_0\bigg(\ln\frac{f_1(\boldsymbol{X}|\boldsymbol{\theta}_1)}{f_2(\boldsymbol{X}|\boldsymbol{\theta}_2)}\bigg) = \frac{1}{2} \tag{5.43}$$

dove $\mathbb{P}_0$ è la probabilità calcolata seconda la distribuzione effettiva di X.

Clarke (2007) quindi pone $\delta_i = \ln f_1(\boldsymbol{x}_i|\boldsymbol{\theta}_1) - \ln f_2(\boldsymbol{x}_i|\boldsymbol{\theta}_2)$ e considera la statistica che segue:

$$B = \sum_{i=1}^{d} \mathbb{1}_{(0,+\infty)}(\delta_i) \tag{5.44}$$

dove $\mathbb{1}_{(0,+\infty)}$ è la funzione indicatrice dell'insieme $(0,+\infty) \subset \mathbb{R}$. Clarke (2007) afferma che la statistica B si distribuisce asintoticamente come una bernoulliana di parametro $\theta = \frac{1}{2}$.

Se il modello $\mathcal{M}_1$ è meglio di $\mathcal{M}_2$ in maniera significativa allora B sarà significativamente superiore a $\frac{d}{2}$, ovvero il valore atteso dall'ipotesi H_0. Per il test della coda superiore, si rigetterà l'ipotesi di equivalenza quando $B \geq c_\alpha$ dove c_α è il più piccolo intero tale che:

$$\sum_{c=c^\alpha}^{d} \binom{d}{c} \leq \alpha 2^d \tag{5.45}$$

Nel test della coda inferiore, si rigetterà l'ipotesi se $B \leq c_\alpha$ dove c_α è il più piccolo c_α tale che

$$\sum_{c=0}^{c_\alpha} \binom{d}{c} \geq \alpha 2^d \tag{5.46}$$

Clarke (2007) definisce inoltre dei fattori di correzione per la sua statistica, sempre collegati ai criteri di informazione di Akaike e Bayesiano. Il fattore di correzione non può essere applicato a B, come nel caso di Vuong, ma va a correggere le log-verosimiglianze.

Il test B non presta attenzione a quanto le due log-verosimiglianze siano dissimili, ma solo a quale delle due sia migliore. Il resto di Clarke (2007) si focalizza quindi sul mostrare come il suo test sia attendibile come quello di Vuong.

Si osserva che l'ipotesi nulla di Vuong non implichi quella di Clarke e viceversa. Vi saranno quindi casi in cui i due test daranno risultati diversi.

5.8 Conclusioni

I capitoli da 8 a 10 applicano le metodologie proposte in questo capitolo a vari *dataset* attuariali ma la domanda di ricerca – ovvero: "*la metodologia* PCC *è appropriata per essere usata in ambito attuariale*?" – non può essere risposta in modo completo utilizzando solo qualche esempio, indipendentemente dalla qualità degli esempi scelti.

Questo paragrafo vuole dunque fare il punto della situazione ed evidenziare gli aspetti positivi e negativi della metodologia, ovvero vuole dare una prima risposta teorica alla domanda di ricerca (a cui vanno aggiunte le considerazioni del capitolo 6).

Le copule sono uno strumento molto versatile per lo studio della dipendenza e la modellazione inferenziale tramite copule è alla base della loro utilità. Le classiche famiglie di copule parametriche, presentate nel capitolo 3, così numerose e versatili nel caso bi-dimensionale, perdono molto della loro capacità espressiva al crescere della dimensione. La metodologia PCC, come altre metodologie concorrenti, cerca dunque di rendere i modelli basati sulle copule utilizzabili a dimensioni superiori a 2.

Alcuni dei limiti della *PCC* non sono propri di questa metodologia ma sono condivisi dalle metodologie concorrenti; non sarebbe pertanto corretto biasimarla per questi difetti.

Le difficoltà presentate nelle sezioni 5.2 e 5.4 sono condivise da tutti i modelli basati sulle copule e si è visto come la struttura della *PCC* possieda alcuni vantaggi rispetto alle copule archimedee gerarchiche. La modellazione tramite copule gaussiane multivariate, comune nelle applicazioni, richiede anch'essa la modellazione delle funzioni marginali. La stima dei parametri è generalmente più semplice nelle copule gaussiane multivariate ma non in maniera determinante.

Al variare di famiglie e *R-vine*, la metodologia *PCC* produce modelli parametrici differenti. Le difficoltà e le incertezze incontrate nelle sezioni da 5.5 a 5.7 sono dunque proprie della teoria statistica inferenziale e non delle *PCC* in senso stretto. Nella modellazione tramite copule gaussiane multivariate, il modello è stato scelto a priori e la sua comparazione con modelli concorrenti è un problema esterno. La

metodologia PCC, al contrario, è un *framework* che genera molti modelli concorrenti e cerca di selezionare il migliore in modo automatico. La versatilità nei software disponibili permettono di utilizzare un approccio misto, ma l'idea generale è quella di scegliere il modello in modo automatico.

La sezione 5.7 mostra invece come il concetto di "migliore" nella statistica inferenziale non sia univoco e certo e come l'incertezza insita nella statistica inferenziale sia condivisa da tutte quelle metodologie che mirano a scegliere un modello tra più opzioni concorrenti. Ma questo è il prezzo della statistica inferenziale.

Al di là delle problematiche presentate in questo capitolo, l'enorme numero di modelli generabili tramite questa metodologia rende la modellazione tramite *PCC* molto versatile e, per lo meno nell'analisi di fenomeni continui, garantisce la creazione di modelli sufficientemente vicini alla realtà. La semplice modellazione tramite copule gaussiane multivariate, tra l'altro rappresentabile attraverso modelli PCC, non può competere sotto questo aspetto.

In definitiva, questo capitolo mostra un quadro positivo. La *PCC* permette di generare una copula molto versatile ed espressiva che riesce a catturare le caratteristiche del *dataset* in maniera abbastanza fedele. Molti dei suoi limiti computazionali e relativi alla ricerca del modello migliore sono condivise dalle metodologie concorrenti e quindi sono problematiche con cui si deve necessariamente convivere. La *PCC* non è certo semplice da interpretare, ma l'interpretazione di un modello multivariato non-lineare con più di due dimensioni è sempre difficile. La scelta dunque è tra limitarsi a due dimensioni e perdere parte della struttura di dipendenza oppure salire di dimensioni e complessità.

Capitolo 6

La *Pair-Copula Construction* con Variabili Non-Continue

> [...] surely discrete marginal dfs $F_1, \cdots, F_d$ *must* cause problems. Believe me, they do! The *full* version of Sklar's Theorem makes this clear. See Genest e Nešlehová (2007) for an excellent primer on this issue and be prepared that everything that can go wrong, *will* go wrong.
>
> Embrechts (2009)

6.1 Introduzione

Correva l'anno 2005, lo stesso in cui veniva pubblicato A. McNeil, Frey e Embrechts (2005), che osserva:

> *The copula concept is slightly less natural for multivariate discrete distributions. This is because there is more than one copula that can be used to join the margins to form the joint df.*

Unitamente, nello stesso anno, veniva pubblicata la tesi dal titolo *Dependence of non-continuous random variables* ad opera dalla giovane dottoranda Johanna Nešlehová, che proprio quest'anno ha vinto il Krieger-Nelson Prize.

> *The Canadian Mathematical Society is pleased to announce that Dr. Johanna G. Nešlehová (McGill University) has been named the recipient of the* 2023 *Krieger-Nelson Prize. Dr. Nešlehová is recognized for her exceptional contributions to Statistics, including multivariate analysis, stochastic dependence modeling, and extreme-value theory. Dr. Nešlehová earned her PhD in Mathematics from Carl-von-Ossietzky-Universität Oldenburg in* 2004*. Since then, she has built an outstanding*

academic career with exceptional talent and a high rate of research productivity, with 43 *peer-reviewed articles, as well as numerous other publications, including book chapters, conference proceeding articles, popular science articles, editorials, and the like, not to mention a popular undergraduate textbook, written in German no less! Dr. Nešlehová is a world leader on copula models and their many ramifications in multivariate statistics, notably in relation to risk analysis and extreme-value theory, an area to which she has made numerous outstanding contributions. She is well-known for promoting statistical risk analysis in insurance and finance through her writing and through short courses. She has also made key contributions to the theory of empirical processes and has wide-ranging interests in the application of stochastic dependence and extreme values to climate and finance. In 2019, Dr. Nešlehová was the distinguished recipient of the CRM-SSC Prize in Statistics. Her work has been published in top-ranked statistical journals, including The Annals of Statistics, ASTIN Bulletin, Biometrika, and the Journal of the American Statistical Association. She is highly engaged in international collaboration, conference speaking and organization, editorial work, and service to the profession. She is currently Editor-in-Chief for The Canadian Journal of Statistics, and she has served as Associate Editor for journals such as Test, the Journal of Multivariate Analysis and Statistics and Risk Modeling. In 2011, she received the distinction of being named an Elected Member of the International Statistics Institute, and in 2020 she was named a Fellow of the Institute of Mathematical Statistics. In addition to these achievements, Dr. Nešlehová is also a generous and dedicated mentor for young researchers. In* 2019*, she was recognized for the excellence of her graduate training with the Carrie M. Derick Award for Graduate Supervision and Teaching from McGill University. In sum, Dr. Nešlehová is an outstanding mathematical statistician and an exemplary role model. Her research accomplishments are all the more impressive given her active engagement in and service to the research community, her dedication to excellent mentorship, and her many other leadership qualities. Johanna G. Nešlehová is an indispensable member of the mathematical community, and the CMS is proud to award her the* 2023 *Krieger-Nelson Prize.*

Qualche anno più tardi, sempre quella giovane dottoranda insieme a Genest (Genest e Nešlehová (2007)) ha analizzato il comportamento delle copule con dati di conteggio (*A Primer on copulas for count data*).

Si segnala che "solo" nel 2012 si inizia a trattare con un approccio assiomatico l'intero perimetro discreto coniugandolo alla teoria della *PCC*, in particolare nell'articolo *Pair Copula Constructions for Multivariate Discrete Data* (Panagiotelis, Czado e Joe (2012)) viene presentata *a new class of models for multivariate discrete data based on pair copula constructions (PCCs).* Dove in primo luogo, vengono derivate le condizioni in base alle quali qualsiasi distribuzione discreta multivariata può essere decomposta come una *PCC*. In secondo luogo, viene presentato l'onere computazionale della valutazione della verosimiglianza per una *PCC* discreta

n-dimensionale.

Si osserva che lavorare con variabili discrete equivale a confrontarsi anche con dati di natura qualitativa nei quali viene meno il concetto di *ordine* mentre per le variabili discrete di tipo ordinale i cosidetti *count data* ereditano molti degli indicatori che definiscono nel caso continuo la dipendenza tra variabili aleatorie che sono stati presentati nel capitolo 2. Ad esempio, quando si parla di variabili di tipo categoriale non esistendo un ordine e non ha senso parlare di concordanza.

Quindi quello che si andrà ad analizzare e che sin dall'introduzione di questo elaborato è stato presentato come "il problema dell'inversione", è un problema che non solo esiste e se ne prende atto ma si comprenderà essere un problema delicato, stratificato e che si propaga dalle copule fino alla *pair-copula construction*. In questo capitolo si presenteranno gli aspetti più rilevanti sia in termini di teoria delle copule che di *PCC*. In particolare i principali *topic* su cui si porrà l'attenzione saranno:

- I temi di carattere teorico-probabilistico evidenziati da Genest e Nešlehová (2007);
- I temi di carattere modellistico-computazionale prodotti dall'impossibilità di utilizzare gli stessi algoritmi nel caso continuo. Per esempio il metodo classico per determinare la verosimiglianza nei modelli di copula n-dimensionali richiede 2^n valutazioni per ogni punto del reticolo dell'intervallo unitario. Panagiotelis, Czado e Joe (2012) mostrano come sia possibile farlo usando $2n(n-1)$ valutazioni delle funzioni di copula quando tutte le variabili sono discrete. Successivamente, Stöber, Hong et al. (2015) presentano il caso di modelli con variabili miste, nell'articolo infatti scrivono: *the contribution is to introduce a novel model for multivariate data where some response variables are discrete and some are continuous.*

Si sottolinea, che lo studio delle varibili discrete costituiscono una parte centrale per rispondere alla domanda di ricerca di questo elaborato in quanto nelle scienze attuariali l'applicabilità sia dei metodologie inferenziali che fanno uso delle copule che dell'approccio di tipo *pair* sono state poco esplorate. In particolare questo capitolo tratta argomenti spesso ignorati dalla manualistica sulle copule, il problema dell'inversione e le sue conseguenze e gli approcci che possono "lenire" tale problematica sono frutto del presente lavoro ricerca. Come vedremo trovare una copula nel perimetro discreto che soddisfi il teorema di Sklar non equivale a trovare una struttura di dipendenza per la coppia di variabili aleatorie in esame, anzi.

6.2 Problema dell'inversione, mancanza di unicità e conseguenze

Il primo osservare ad la presenza di incorenze nel perimetro delle copule aventi marginali non continue è stato Marshall (1996), unitamente a questo si osserva che la definizione di una funzione "quasi-inversa" e del teorema di Sklar nella sua forma "alternativa", entrambi presentati nel corso del capitolo 1, ci riporta proprio di fronte a questo senso di "non-continuità" delle variabili marginali che entrano in gioco nella costruzione della funzione multivariata.

Come è noto, la copula è una funzione multivariata avete marginali uniformi e attraverso il Teorema di Sklar si riesce a unire una qualsiasi distribuzione congiunta con delle distribuzioni marginali univariate e se le marginali godono della proprietà di continuità allora la funzione di copula non solo esiste ma è anche unica.

Diversamente se le marginali non sono continue si è posti dinnazi al problema dell'inversione che consiste, di fatto, nel confrontarsi non più con una mappa biunivoca che fa corrispondere ai punti del dominio in modo univoco ai punti del codominio ma una "pluralità di contro-immagini": di fatto nel passaggio all'inversione i punti delle marginali corrispondono a un *plateau* di valori. Questo fenomeno lo si può vedere in modo chiaro nella figura 6.1:

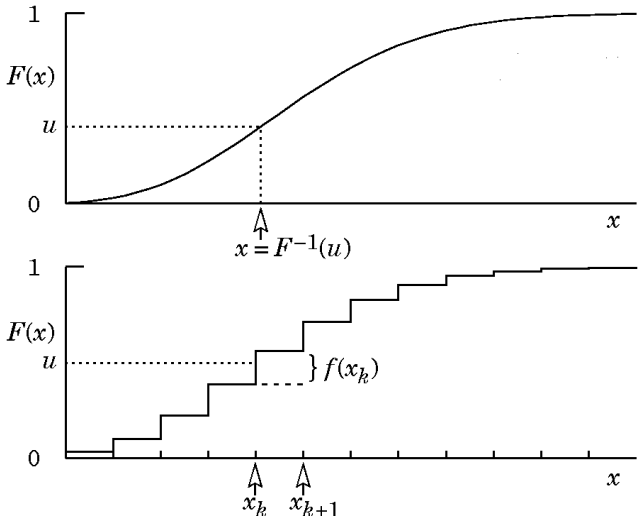


Figura 6.1. Problemi di inversione.

Infatti, nel passaggio all'inversione di una distribuzione discreta, diversamente a quanto accade a distribuzioni marginali continue, non vi è biezione tra dominio e codominio, ovvero la contro-immagine della funzione multivariata è non unica, in altre parole $F^{(-1)}(u)$ ha un intervallo di punti contenuti nella finestra di estremi $[x_k, x_{k+1}]$.

In generale, diversamente da quello che accade nelle soluzioni delle equazioni differenziali, in cui la non continuità di fatto non assicurava nemmeno l'esistenza di una soluzione, nel contesto delle copule rilasciando l'ipotesi di continuità del teorema di Abe Sklar (1959) si perde "solo" il senso di unicità ovvero la soluzione non solo esiste sempre ma il problema è che ne esistono troppe.

Esempio 6.1 (Copula di Bernulli)**.** *Siano* $X \sim \mathrm{B}(X)$, $Y \sim \mathrm{B}(Y)$ *per due probabilità* $\pi_X, \pi_Y \in (0,1)$ *e siano* X *e* Y *indipendenti allora per costruire la distribuzione bivariata di Bernoulli* F_{XY} *è sufficiente definire la copula nel modo seguente:*

$$C(1-\pi_X, 1-\pi_Y) = (1-\pi_X)(1-\pi_Y) \tag{6.1}$$

Facilmente si osserva che C *è definita su* $\operatorname{Ran} F_X \times \operatorname{Ran} F_Y = \{0, 1-\pi_X, 1\} \times \{0, 1-\pi_Y, 1\}$ *ma il comportamento di* C *lungo i lati di* $\mathbb{I}^2$ *è regolato da vincoli*

banali avendo marginali uniformi, quindi solo in corrispondenza $(1-\pi_X)(1-\pi_Y)$ *può fornire delle informazioni sulla struttura di dipendenza. Si osserva quindi che la copula prodotto soddisfa naturalmente il Teorema di Sklar, ma lo stesso vale per un'ampia gamma di altre funzioni di copula.*

Venendo meno l'unicità legittimamente si mette in discussione qualsiasi conclusione: infatti l'elemento che si suppone descriva la struttura di dipendenza tra le marginali, cioè la copula C, può caratterizzare in modo "intercambiabile" l'indipendenza o la dipendenza e la "forza" e "natura" di questo legame. In altre parole, qualsiasi misura di dipendenza basata sulla copula in contesto di questo tipo non è interpretabile, dato che la copula potrebbe caratterizzare strutture di dipendenza assolutamente diverse. L'importante è il passaggio sui punti non triviali della scacchiera qualunque funzione soddisfi tale passaggio e abbia le proprietà di una copula è legittima in tal senso.

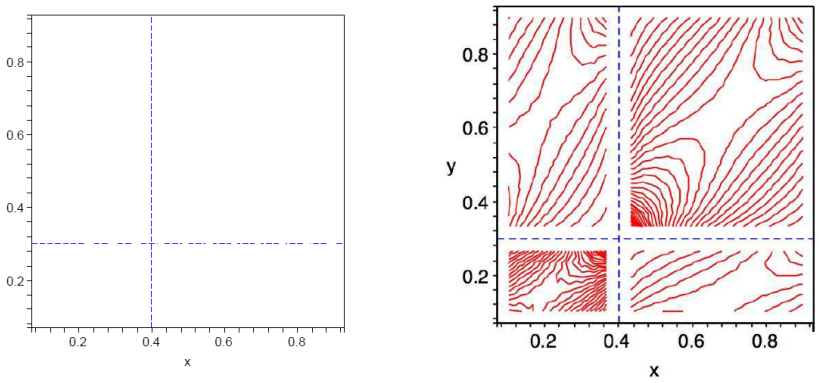


Figura 6.2. Curve di livello che soddisfano $C(1-\pi_X, 1-\pi_Y) = (1-\pi_X)(1-\pi_Y)$.

Come emerge nella figura 6.2 al netto del passaggio sui 9 punti che sono i vertici dei rettangoli sul quadrato unitario, qualunque *sub*-copula passante per i punti non triviali soddisfa Sklar e quindi a pieno titolo dovrebbe essere adeguata a rappresentare l'indipendenza tra le marginali di partenza.

Ad esempio, se si pone $\pi_X = \pi_Y = \frac{1}{2}$ si può osservare le densità di 6 funzioni di copula differenti che soddisfano l'indipendenza tra X e Y. Tuttavia, evidentemente la loro rappresentazione non coglie la "vera natura" della variabili marginali e di conseguenza qualsiasi valutazione qualitativa o quantitativa della dipendenza sottostante basata su di esse è assolutamente forviante.

In particolare Genest e Nešlehová (2007) dimostrano che non solo la funzione non è unica ma forniscono un risultato circa la dimensione di questo insieme.

In particolare l'insieme delle funzioni di copula che soddisfa Sklar è definito da un *range* che è funzione di τ_C di Kendall, e che può essere espressa dalla seguente espressione:

$$\tau_C = -1 + 4 \iint_{I^2} C(u,v)\, dC(u,v)$$

Dunque basta mostrare che la differenza tra le probabilità di concordanza e la probabilità della discordanza, ovvero la definizione in senso probabilistico di τ_K non

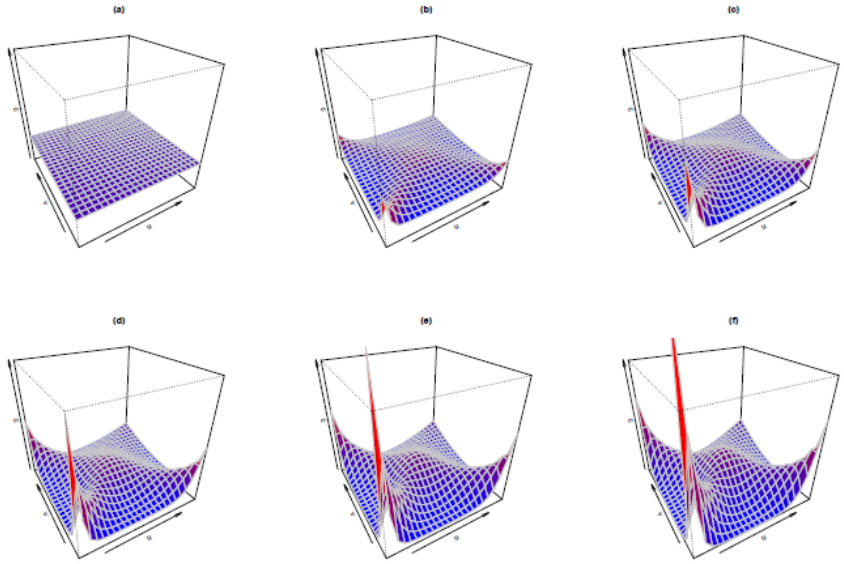


Figura 6.3. Densità che soddisfano $C(1-\pi_X, 1-\pi_Y) = (1-\pi_X)(1-\pi_Y)$.

sono pari al valore di τ_C. In modo del tutto analogo si dimostra che esiste un *range* che è funzione anche del ρ_C di Spearman:

$$\rho_C = -3 + 12 \iint_{I^2} uv \, dC(u,v)$$

Inoltre tale fenomeno si osserva è *ulteriormente* aggravato quando il numero di valori che la variabili discrete assumono è limitato ovvero il numero delle funzioni di copula è inversamente proporzionale alle realizzazioni delle variabili discrete. E questo si coglie abbastanza facilmente dalle rappresentazioni grafiche che è stato evidenziato negli esempi precedenti. Infatti intuitivamente si può intuire che maggiori sono i vincoli e minori saranno le *sub*-copule che riempono i rettangoli nel quadrato unitario[1].

6.3 Misure di concordanza per variabili casuali non continue

Nel capitolo 2 è stato osservato che il valore del τ_K di Kendall e il ρ_S di Spearman in senso probabilistico, nel perimetro delle variabili continue, coincidono, col τ_C e ρ_C nel senso della copula:

$$\tau_C = 4 \iint_{\mathbb{I}^2} C(u,v) \, \mathrm{d}C(u,v) - 1.$$

[1]Viceversa, se nel continuo accade il contrario, essendo l'insieme dei numeri reali un insieme denso corrisponderà solo una copula.

$$\rho_C = 12 \iint_{\mathbb{I}^2} (C(u,v) - uv)\, \mathrm{d}u\mathrm{d}v. \tag{6.2}$$

Tuttavia questa corrispondenza viene meno nelle discrete, Genest e Nešlehová (2007)) mostrano che non solo il τ_K di Kendall e il ρ_S di Spearman calcolati a partire dalla definizione probabilistica non corrispondono ai valori di τ_C e ρ_C ma che esiste, al variare di C un ampio *range* dei valori di τ_C e di ρ_C. Nelle figura 6.4 e figura 6.5 si osserva il *lower-bound* and *upper-bound* e la differenza tra *lower-bound* e *upper-bound* rispettivamente del τ_C e del ρ_C delle copule associate a due bernulliane:

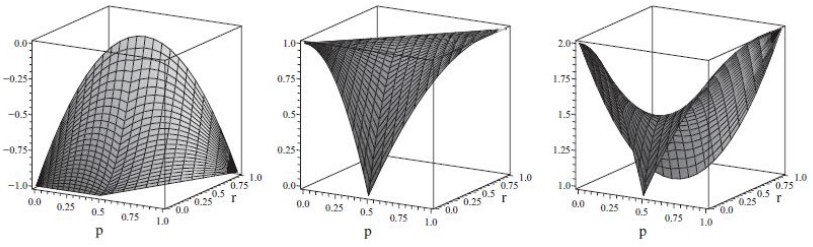


Figura 6.4. *Range* dei valori di τ_C.

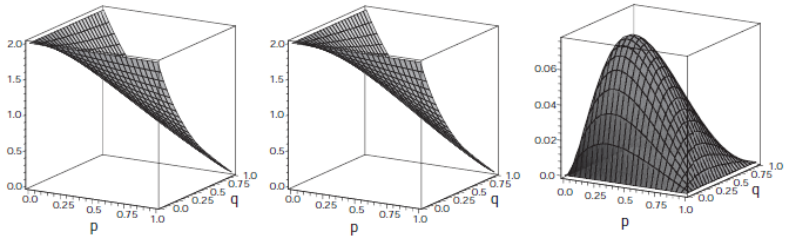


Figura 6.5. *Range* dei valori di ρ_C.

6.4 Come salvare il legame tra copula e dipendenza per variabili non continue

Malgrado le criticità rilevate nei paragrafi precedenti, la letteratura è abbastanza unanime nel ritenere che si possano utilizzare le copule anche per comprendere le informazioni riguardo la dipendenza per le variabili non continue. Infatti, in base al Teorema di Sklar, la funzione di distribuzione di probabilità di ogni vettore aleatorio può essere decomposta in modo univoco in termini di distribuzioni marginali delle sue componenti e di un'opportuna *sub*-copula: questo vale sempre, il vettore aleatorio non deve essere necessariamente continuo. Nel corso degli anni sono stati presentati diversi approcci per trattare le variabili non continue, in particolare si segnala[2]:

[2]Non verranno analizzati tutti casi, tuttavia era doveroso quantomeno citarli.

- L'estensione bilineare (Nelsen (2006), Genest, Nešlehová e Rémillard (2014));
- L'estensione di Carley (2002);
- I nuclei di Bernoulli (Geenens (2020))[3].

Si osserva che in questi approcci vengono analizzate le possibili distribuzioni bivariate in modo specifico per alcune distribuzioni discrete, come la Poisson, la Binomiale, le Bernoulli e la distribuzione Geometrica. Non esiste in generale un approccio univoco per avvicinarsi al problema, inoltre le estensioni che si vedranno riguardano, non a caso, solo il contesto bivariato. A prima vista, potrebbe sembrare possibile estendere la strategia a dimensioni superiori. Questo, purtroppo, non è possibile in quanto, il limite inferiore W^n non è una copula se $n \geq 3$: il minimo in generale non è una copula (capitolo 1).

Si presenteranno in questo elaborato l'estensione *standard* su cui hanno lavorato Schweizer e Abe Sklar (1983) Schweizer e Sklar [4], Genest e Nešlehová (2007) [5].

Si osserva che, nella realtà come nella produzione scientifica, quando un particolare argomento è lungamente dibattuto, il detto "ognuno dice la sua" ha sempre una radice di fondatezza (che in questo caso va inteso come "ognuno usa la sua (di notazione)").

Anche in questo caso, come è stato segnalato nel capitolo 1 per la funzione inversa generalizzata e la funzione quasi-inversa, si può inizialmente essere travolti dalla confusione. Infatti quella che viene detta forma bilineare e che viene indicata con $C^{\maltese}$ nel capitolo 1 è detta anche estensione *standard* di Schweizer e Sklar (Nelsen (2006)) che si ritrova indicata anche con la notazione C^* (nel libro di Nelsen) oppure con la notazione $\mathcal{C}_S$ in altre trattazioni (la s suggerisce "*standard*").

Fondamentalmente $C^{\maltese}$(utilizzata dalla Nešlehová) corrisponde all'interpolazione lineare di un'unica *sub*-copula e dell'unica copula della distribuzione congiunta *smoothed*. L'idea alla base di questo approccio è ricercare un'adeguata estensione che produca una copula che sia in grado di catturare la strutture di dipendenza della distribuzione di partenza. Ora in verità tutti parlano di "estensione" il cui termine in qualche modo rimanda ad un "allargarsi", tuttavia si osserva che il problema consiste fondamentalmente nel decomporre una copula in opportune *sub*-copule[6] (sezione 6.4). Inoltre, poichè l'interpolazione è lineare nei due parametri λ e μ (Nelsen (2006)) è detta "bilineare".

[3] Approccio molto interessante che definisce un metodo grafico chiamato *confetti di Bernulli*.

[4] Per tale ragione, talune volte si parla "di estensione di Schweizer e Sklar". Nelsen (2006) la indica con C^*.

[5] Loro parlano di "forma bilineare" e la indicano con $C^{\maltese}$.

[6] Quindi si presume che il concetto di estensione rimandi alla procedura di *smooting*.

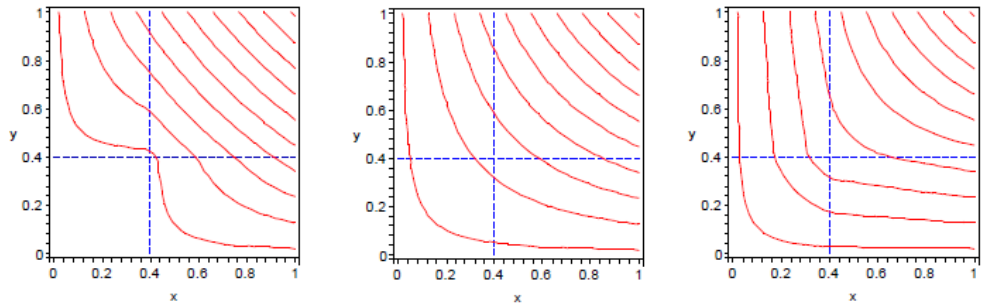


Figura 6.6. *Smoothing* dell'estensione bilineare di due bernoulliane.

Si dimostra che:

- $C'^{\maltese}$ è una copula assolutamente continua;
- $C'^{\maltese}$ conserva l'indipendenza delle marginali;
- $C'^{\maltese}$ conserva la concordanza.

$C'^{\maltese}$ è considerata la miglior metodologia per trattare le copule con variabili non continue, tuttavia si osserva che se le marginali hanno un legame di perfetta monotonia, $C'^{\maltese}$ non coincide con i limiti di Fréchet-Hoeffding.

Operativamente si definisce una tabella di contingenza e si "distribuisce uniformemente" la massa in ogni rettangolo contenuto nel quadrato unitario.

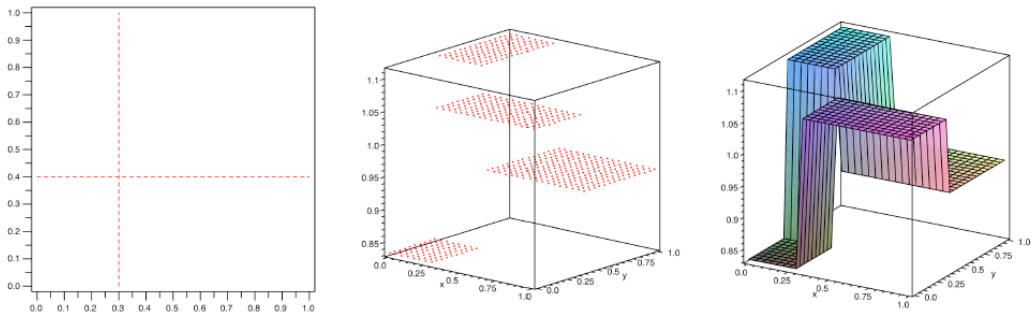


Figura 6.7. Densità dell'estensione bilineare della copula avente marginali bernuliane.

Dal punto di vista inferenziale di un campione di d osservazioni con marginali note e definite in $\mathbb{N}$, l'estensione bilineare di $C'^{\maltese}$ può essere stimata mediante l'interpolazione bilineare $B_d^{\maltese}$ della distribuzione empirica del campione:

$$\big(F(X_i); G(Y_i), 1 \geq i \geq d\big) \tag{6.3}$$

Ovvero vale il seguente teorema:

Teorema 6.1. *Sia il processo* $\mathbb{B}_d^{\maltese}$ *definito come segue:*

$$\mathbb{B}_d^{\maltese} = \sqrt{d}(B_d^{\maltese} - C_d^{\maltese}) \tag{6.4}$$

Si dimostra che $\mathbb{B}_d^{\maltese}$ *è asintoticamente normale ovvero per* $d \to \infty$*, il processo* $\mathbb{B}_d^{\maltese}$ *converge debolmente ad un processo Gaussiano centrato* $\mathbb{B}_C^{\maltese}$ *ovvero:*

$$\mathbb{B}_C^{\maltese} \coloneqq \lim_{d\to\infty} \mathbb{B}_d^{\maltese} = \lim_{d\to\infty} \sqrt{d}(B_d^{\maltese} - C_d^{\maltese}) \tag{6.5}$$

6.5 La *Pair-Copula Construction* con variabili discrete

Nei paragrafi precedenti sono stati discussi i principali problemi a cui si può incorrere quando vengono trattate copule aventi come marginali variabili discrete. Ora si vedrà cosa accade nel coniugio *pair-copula construction* e variabili discrete.

Si segnala che Panagiotelis, Czado e Joe (2012) ricostruiscono totalmente l'impianto teorico sino ad allora adottato per le variabili continue adeguandolo al perimetro delle variabili discrete. In particolare essi derivano le condizioni per cui qualsiasi distribuzione discreta multivariata può essere decomposta con approccio *pair-copula construction*. Utilizzando il teorema di Sklar, osservano che qualsiasi densità condizionale univariata può essere decomposta nel prodotto di una densità copula bivariata e di un'altra densità condizionale univariata con l'insieme condizionante ridotto di un elemento.

Lasciando che V_h sia un qualsiasi elemento scalare di V e $\boldsymbol{V}_{\setminus h}$ il suo complemento a uno e sia Y_j un elemento non contenuto in V.

L'equazione che segue si applica ricorsivamente per decomporre una distribuzione multivariata in piccoli *blocchi* nel caso discreto[7]:

$$\begin{aligned} F_{Y_j|V_h,V_{\setminus h}}(y_j|v_h,\boldsymbol{v}_{\setminus h}) = \Big[& C_{Y_j,V_h|\boldsymbol{V}_{\setminus h}}(F_{Y_j|\boldsymbol{V}_{\setminus h}}(y_j|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})) \\ & - C_{Y_j,V_h|\boldsymbol{V}_{\setminus h}}(F_{Y_j|\boldsymbol{V}_{\setminus h}}(y_j|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h-1|\boldsymbol{v}_{\setminus h}))\Big] \\ & / \, \mathbb{P}\{Z_k = z_k|\boldsymbol{Z}_{\setminus k} = \boldsymbol{z}_{\setminus k}\} \end{aligned} \tag{6.6}$$

Si enuncia il risultato più importante nel perimetro delle variabili discrete ovvero il teorema di rappresentazione (Panagiotelis, Czado e Joe (2012)):

Teorema 6.2. *(Esistenza di una rappresentazione* PCC *per una qualsiasi distribuzione discreta multivariata). Una distribuzione discreta multivariata può essere espressa attraverso la* pair-copula construction *se esiste una decomposizione in cui ogni insieme di distribuzioni condizionate bivariate della vine ha una copula condizionata* costante *a variabili rettangolari non-negative che soddisfano i vincoli delle delle seguenti forme:*

$$C_{Y_j|V_h,V_{\setminus h}}(a_{yj}|v_h,b_{vh}|\boldsymbol{V}_{\setminus h}) = p_{yj}|\boldsymbol{V}_{\setminus h} \tag{6.7}$$

[7]Per i dettagli si rimanda a Panagiotelis, Czado e Joe (2012)

$$C_{Y_j|V_h,V_{\setminus h}}(a_{yj}|1|\boldsymbol{V}_{\setminus h}) = a_{yj,v_h|\boldsymbol{V}_{\setminus h}} \tag{6.8}$$

$$C_{Y_j|V_h,V_{\setminus h}}(1|v_h, b_{vh}|\boldsymbol{V}_{\setminus h}) = b_{vh|\boldsymbol{V}_{\setminus h}} \tag{6.9}$$

Inoltre Panagiotelis, Czado e Joe (2012) mostrano che dal punto di vista computazionale non è possibile utilizzare gli stessi algoritmi del caso continuo e implementano una metodologia più parsimoniosa rispetto ad approcci tradizionali.

Il metodo implementato da Panagiotelis, Czado e Joe (2012) utilizza $2n(n-1)$ valutazioni contro le 2^n valutazioni degli algoritmi adottati per la stima della massima verosimiglianza, primo tra tutti si ricorda quello definito da Joe (1997).

6.6 Lo stimatore *kernel* con *jittering*

Da ultimo si introduce una recente classe di stimatori non parametrici per stimare la densità di distribuzioni multivariate aventi come marginali variabili discrete denominate stimatori *jittering* (*jittering estimators* (Nagler (2018))).

Lo stimatore *kernel* con approccio di *jittering* è una metodologia definita da Nagler nel 2018 e disponibile anche nella libreria `rvinecopulib` di `R` (attraverso la liberia `kde1d`). Siccome verrà usata nei prossimi capitoli, è utile darne una breve introduzione.

Cos è il *jittering*?

Il *jittering*, come il *bootstapping*, rimanda in qualche modo al significato della parola "abracadabra" che deriva dal sanscrito e significa "creare dal nulla".

Infatti, in modo del tutto analogo a quello che accade con il *bootstapping*[8] in cui si aggiunge "variabilità artificiale" ai nostri dati, nel *jittering*[9] "si aggiunge rumore" alle variabili discrete.

In particolare, il *jittering* dei dati discreti, come è stato osservato da Denuit e Lambert (2005) per gli stimatori non parametrici generici, non distrugge il comportamento dei dati.

Nagler (2018) infatti mostra, che l'aggiunta di rumore attraverso l'approccio di *jittering* non causa la perdita di informazioni, inoltre si sottolinea che diversamente dall'approccio parametrico mostrato da Nikoloulopoulos, Joe e Li (2012), lo stimatore *kernel* che fa uso del *jittering* è efficente anche asintoticamente.

[8] bootstrap s. ingl. [propr. stringa per scarpe, prob. dall'espressione idiomatica *by one's bootstraps con le proprie forze*] Termine impiegato con diversi significati nel linguaggio scientifico e tecnico, sempre con riferimento a processi o fenomeni che traggono sostentamento o alimento da sé stessi: in elettronica, qualificazione di varî circuiti, per es. di un inseguitore catodico o di un circuito di polarizzazione automatica, per caduta di tensione, di tubi e transistori; in informatica, la procedura con la quale, all'accensione, un calcolatore carica da una delle memorie il sistema operativo, mettendosi quindi da solo in grado di operare; in fisica delle particelle elementari il quale ogni particella è costituita da una sovrapposizione di tutte le altre. Nel contesto statistico quindi si richiama all'azione "del tirare fuori dai dati".

[9] jittering s. ingl. [propr. nervosismo; tremolio, come effetto che si manifesta in segnali elettrici] , usato in ital. al masch. - In elettronica, la variazione, spesso casuale e di breve durata, dell'ampiezza o della fase di un segnale elettrico. Nel contesto statistico quindi si richiama all'azione "si rendono nervosi i dati".

Definizione 6.1 (Stimatore *jittering kernel*)**.** *Si supponga che* $(\boldsymbol{Z}, \boldsymbol{X})$ *sia un vettore di vettori aleatori dove* $\boldsymbol{Z}$ *rappresenta la sua componente discreta ovvero* $\boldsymbol{Z} \in \mathbb{Z}^n$ *e* $\boldsymbol{X}$ *rappresenta la componente continua ovvero* $\boldsymbol{X} \in \mathbb{R}^n$*.*

L'obiettivo di questa trattazione è stimare la densità f *di* $(\boldsymbol{Z}, \boldsymbol{X})$ *sulla base delle "osservazioni".* $(Z_i, X_i), i = 1, \ldots, d$ *che sono vettori casuali indipendenti e identicamente distribuiti con la stessa distribuzione di* $(\boldsymbol{Z}, \boldsymbol{X})$*. In questo contesto,* f *è la densità rispetto alla misura di Lebesgue, ossia:*

$$f_{\boldsymbol{Z},\boldsymbol{X}}(\boldsymbol{z}, \boldsymbol{x}) = \frac{\partial^q}{\partial x_1 \cdots \partial x_q} \mathbb{P}(\boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{X} \leq \boldsymbol{x}) \tag{6.10}$$

Sia $\mathbb{K}$ *una funzione a valori reali definita come segue:*

$$\mathbb{K}(\boldsymbol{w}) = \prod_{j=1}^{n} K(w_j) \tag{6.11}$$

Dove $\boldsymbol{w} \in \mathbb{R}^n$ *e* $n \in \mathbb{N}$*.*

$\mathbb{K}$ *è detta* kernel[10] *e la sua versione classica è definita funzione:*

$$\hat{f}_{\boldsymbol{Z},\boldsymbol{X}}(\boldsymbol{z}, \boldsymbol{x}) = \frac{1}{d h_d^p b_d^q} \sum_{i=1}^{d} \mathbb{K}\left(\frac{\boldsymbol{Z}_i - \boldsymbol{z}}{h_d}\right) \mathbb{K}\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{b_d}\right) \tag{6.13}$$

Dove $h_d, b_d > 0$ *sono detti parametri di larghezza di banda e controllano la quantità di smoothing.*

Si definisce densità dello stimatore jittering kernel *la seguente espressione:*

$$\tilde{f}_{\boldsymbol{Z},\boldsymbol{X}}(\boldsymbol{z}, \boldsymbol{x}) = \frac{1}{d h_d^p b_d^q} \sum_{i=1}^{d} \mathbb{K}\left(\frac{\boldsymbol{Z}_i + \boldsymbol{E}_i - \boldsymbol{z}}{h_d}\right) \mathbb{K}\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{b_d}\right) \tag{6.14}$$

In pratica le variabili discrete si rendono continue aggiungendo del rumore $\boldsymbol{E}_i$ *al vettore aleatorio di partenza* $(\boldsymbol{Z}_i, \boldsymbol{X}_i)$*, dove* $\boldsymbol{E}_i, \in \mathbb{R}^n, i + 1, \ldots, d$ *sono vettori indipendenti e identicamente distribuiti di* $(\boldsymbol{Z}_i, \boldsymbol{X}_i)$*.*

In particolare si dimostra che lo stimatore così definito gode di tutte le proprietà che ci auspica possegga uno stimatore non parametrico, ovvero:

- È asintoticamente normale e asintoticamente imparziale per le variabili discrete;
- È fortemente e uniformemente consistente;
- È relativamente efficiente;
- Converge a velocità min max-ottima per un'ampia classe di densità *target.*[11]

[10] $\mathbb{K}$ è una funzione che soddisfa le proprietà, $\mathbb{K}$ è non-negativa ($\mathbb{K}(y) \geq 0$), $\mathbb{K}$ è una densità ($\int_{\mathbb{R}} \mathbb{K}(y) = 1$), $\mathbb{K}$ ha media nulla: ($\int_{\mathbb{R}} yK(y)dy = 0$), $\mathbb{K}$ ha varianza positiva ($\int_{\mathbb{R}} y^2 K(y)dy > 0$). Esistono vari tipi di *kernel* come ad esempio il *kernel* gaussiano ($\mathbb{K}(y) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{y^2}{2}\right\}$) o il *kernel* di Epanechnikov $\mathbb{K}(y) = \frac{3}{4}(1 - y^2)I(y)$, dove:

$$I(y) = \begin{cases} 1 & \text{se } |y| \leq 1 \\ 0 & \text{se } |y| > 1 \end{cases} \tag{6.12}$$

[11] Nagler è stato il primo a trovare un'approccio non parametrico di questo tipo.

Capitolo 7

Librerie R

> One should avoid solving more
> difficult intermediate problems
> when solving a target problem
>
> Vapnik (1999)

Sul sito CRAN esiste il catalogo completo ordinato o per data di pubblicazione o per nome di tutti i *packages* che consentono l'implementazione delle copule in R.

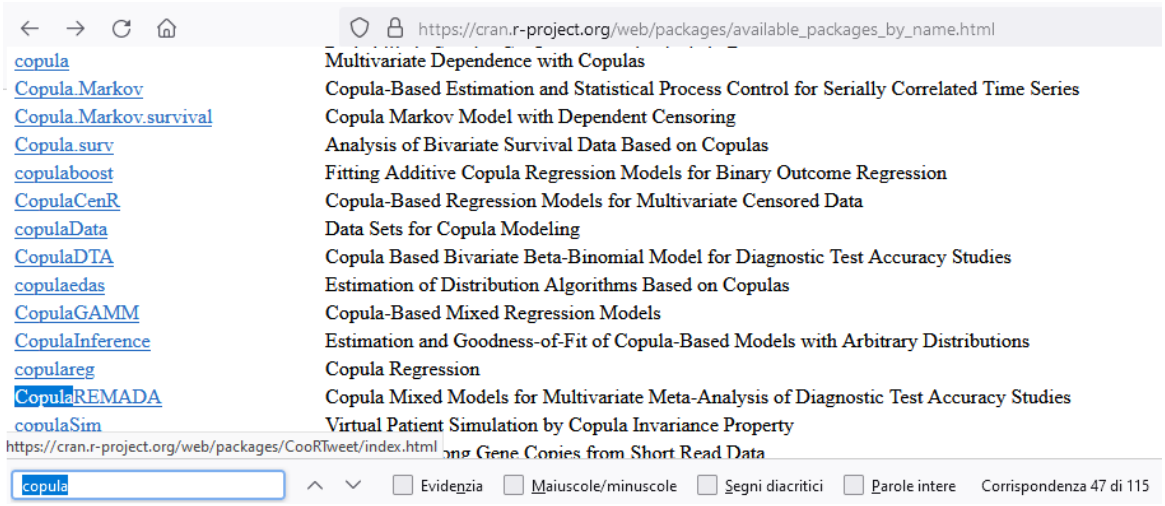


Figura 7.1. Librerie disponibili sul sito CRAN.

All'attualità, il *repository* conta circa 20.000 librerie. Per quanto riguarda il trattamento delle copule, si contano 76 pacchetti di cui 12 sono specifici *ad-hoc* per le *R-vine*. La tabella 7.1 contiene l'elenco delle 12 librerie R che sono state individuate.

Nome del Pacchetto	Descrizione del Pacchetto
`VineCopula`	Statistical Inference of Vine Copulas.
`rvinecopulib`	High Performance Algorithms for Vine Copula Modeling.
`portvine`	Vine Based (Un)Conditional Portfolio Risk Measure Estimation. Basato su `rvinecopulib`.
`pacotest`	Testing for Partial Copulas and the Simplifying Assumption in Vine Copulas. Basato su `VineCopula`.
`kdevine`	Multivariate Kernel Density Estimation with Vine Copulas. Basato su `VineCopula`.
`svines`	Stationary Vine Copula Models. Basato su `rvinecopulib`.
`vinereg`	D-Vine Quantile Regression. Basato su `rvinecopulib`.
`vccp`	Vine Copula Change Point Detection in Multivariate Time Series. Basato su `VineCopula`.
`gamCopula`	Generalized Additive Models for Bivariate Conditional Dependence Structures and Vine Copulas. Basato su `VineCopula`.
`CDVineCopulaConditional`	Sampling from Conditional C-and D-Vine Copulas. Basato su `VineCopula`.
`pencopulaCond`	Estimating Non-Simplified Vine Copulas Using Penalized Splines. Basato su `pacotest`.
`vines`	Multivariate Dependence Modeling with Vines.

Tabella 7.1. Librerie `R` che implementano le *vine*.

La disponibilità di *software* gratuiti ed efficienti per la modellazione mediante *pair-copula construction* ha avuto un ruolo chiave nel suo recente sviluppo. Dalle prime librerie `R` per la modellazione per le sole *C-vine* e *D-vine*, si è passati a soluzioni via via più complesse e potenti.

La libreria `VineCopula` in Nagler, Schepsmeier et al. (2023) è senza dubbio la più conosciuta ed utilizzata. La sua prima versione, la `1.0` è stata lanciata nel 2012 mentre l'ultima la `2.5` è stata rilasciata nel 2023. Si segnala che la Czado (2019) la usa a piene mani nel corso degli anni, facendone ampio riferimento. Inoltre, questa libreria ha il merito di aver favorito l'utilizzo e la diffusione della *pair-copula construction* nel perimetro computazionale.

Tuttavia, che gli autori stessi, affermano che `VineCopula` non è più sviluppata attivamente e consigliano l'utilizzo di `rvinecopulib`[1].

La libreria `rvinecopulib` rilasciata da Nagler e Vatter (2023a) è attualmente

[1] `https://www.math.cit.tum.de/en/math/research/groups/statistics/vine-copula-models/packages-libraries/`

alla versione `0.6.3.1.1`. Si osserva che allo stato dell'arte non implementa tutte le funzionalità di `VineCopula`.

Ad esempio, il `test di Vuong` e il `test di Clarke` non sono disponibili in `rvinecopulib` e dunque le applicazioni dei capitoli da 8 a 10 usano una implementazione che ricalca quella di `VineCopula`, fatta appositamente per questa trattazione. I vantaggi nell'utilizzo di `rvinecopulib` sono i seguenti:

- le operazioni sono implementate in `C++` e sfruttano le potenzialità di `Eigen` e `Boost`;
- la libreria include le funzionalità della libreria `kde1d` per modellare le funzioni marginali;
- si possono modellare variabili non continue;
- si possono generare *R-vine* casuali[2];
- permette di utilizzare modelli non-parametrici per le copule bivariate.

In buona sostanza, `rvinecopulib` è molto più veloce a creare i modelli, non richiede di trasformare le variabili in uniformi per essere usata e supporta più tipologie di variabili e copule.

Le funzionalità che `rvinecopulib` non supporta sono facilmente implementabili in `R`. La disponibilità del codice `R` di `VineCopula` facilita indubbiamente questo tipo di operazioni.

Funzione	Utilizzo
`mBICV`	Calcola l'$mBICV$ di un modello *PCC* calcolato con `vinecop`. Si può fornire un campione differente (il campione deve essere distribuito con marginali uniformi).
`pairs_copula_data`	Costruisce il plot `pairs` per un campione con marginali uniformi. Mostra i plot `contour` nel pannelli inferiori, le correlazione in quelli superiori e gli istogrammi sulla diagonale.
`pseudo_obs`	Usa la distribuzione empirica, ovvero la usa la funzione rango scalata per la dimensione del campione più uno, per convertire il campione in un campione con marginali uniformi.
`rvine_structure_sim`	Funzione per generare *R-vine* casuali.
`vine`	Funzione per il *fitting* dei dati che utilizza `kde1d` per le marginali. Le variabili discrete vengono riconosciute automaticamente in base al loro tipo all'interno del `data.frame`. Si tratta di fatto di un *wrapper* intorno a `kde1d` e `vinecop`. Per le variabili non continue applica il *jittering* automaticamente.
`vinecop`	Funzione di fitting per eccellenza. Richiede un campione con marginali uniformi e dei limiti destro e sinistro della marginali nel caso di variabili non continue.

Tabella 7.2. Le più importanti funzioni di `rvinecopulib`.

La tabella 7.2 elenca le principali funzioni di `rvinecopulib` che saranno usate nei capitoli applicativi. La funzione `vinecop`, quando si usano solo variabili continue

[2]Questa funzionalità sarà fondamentale nel capitolo 10

e si evitano le copule non-parametriche, produce lo stesso risultato di `VineCopula`. Infatti entrambe le librerie utilizzano l'algoritmo di Dißmann et al. (2013).

Oltre a quanto già scritto in tabella 7.2 è importante osservare che la libreria `C++ vinecopulib`[3], ovvero il cuore della libreria `rvinecopulib`, non supporta la generazione di campioni casuali a partire da copule costruite con variabili discrete. Quindi, anche se `vinecop` permette di modellare campioni non continui o misti, il modello creato con questa funzione non può essere usato per generare nuovi campioni casuali. Lo stessa cosa non è vera per la funzione `vine` per via dell'uso di `kde1d` per la modellazione delle marginali.

Va infine osservato che, quando possibile, è meglio utilizzare `pseudo_obs` e `vinecop` per la modellazione invece che `vine` perchè permette un maggiore controllo sulla modellazione delle marginali. Inoltre gran parte degli esempi in Czado (2019) e gli articoli di ricerca basati su `VineCopula` ultizzano le marginali empiriche.

La tabella 7.3 contiene le famiglie di copule supportate da `rvinecopulib` e la loro codifica. È inoltre possibile richiedere modelli che usano tutte le copule (`"all"`), solo le parametriche (`"parametric"`), solo le non parametriche (`"nonparametric"`), solo quelle ad un parametro (`"onepar"`), solo quelle a due parametri (`"twopar"`), solo quelle ellittiche (`"elliptical"`), solo quelle archimedee (`"archimedean"`), solo le `BB` (`"BB"`) e solo quelle per cui è possibile stimare i parametri usando il τ_K di Kendall (`"itau"`). Al contrario di `VineCopula`, le copule ruotate non sono viste come famiglie differenti.

Famiglia	stringa corrispondente in `rvinecopulib`
indipendenti	`"indep"`
non parametriche	`"tll"`
`B1` Gaussiane	`"gaussian"`
`B3` di Frank	`"frank"`
`B4` di Clayton	`"clayton"`
`B5` di Joe	`"joe"`
`B6` di Gumbel	`"gumbel"`
t-Student	`"t"`
`BB1`	`"bb1"`
`BB6`	`"bb6"`
`BB7`	`"bb7"`
`BB8`	`"bb8"`

Tabella 7.3. Le famiglie di copule supportate da `rvinecopulib`.

[3]Nagler e Vatter (2023b)

Capitolo 8

Prime Applicazioni

> "Whether true or not [it] is at least probable; and he who tells nothing exceeding the bounds of probability has a right to demand that they should believe him who cannot contradict him". Samuel Johnson, author of the first English dictionary, wrote this in 1735. He is referring to the Jesuit priest Jeronimo Lobo's account of the unicorns he saw during his visit to Abyssinia in the 17th century.
>
> Shepard (1930)

8.1 Introduzione

Nei capitoli precedenti è stata ripercorsa la teoria fondamentale alla base delle copule e della *pair-copula construction*, ora si inizierà a vedere qualche applicazione di questa metodologia nel perimetro delle scienze attuariali. Si analizzerà un *dataset* del settore *motor* e verrà applicato prima l'approccio classico alle copule dopodiché si confronteranno le risultanze ottenute con quelle prodotte dall'implementazione della *pair-copula construction* facendo uso all'algoritmo di ottimo locale di Dißmann.

8.2 Approccio Copula

Si considera il *dataset* `SwedishMotorInsurance` relativo al settore delle auto utilizzato da Frees (2009) e la libreria `rvinecopulib` di `R` presentata nel capitolo 7. Questi dati sono stati utilizzati dallo *Swedish Committee* per l'analisi del *Risk-Premium* per RcA. Il *dataset* è composto da 2182 osservazioni e presenta 7 variabili: 3 di tipo continuo e 4 di tipo discreto (si veda la tabella 8.1).

Variabile	Tipo Variabili	Descrizione
V. Assicurato	Continua	Valore Assicurato del Veicolo
Sinistri	Continua	Numero dei Sinistri
Pagamenti	Continua	Importo dei Sinistri
Chilometri	Discreta	Chilometri percorsi raggruppati in 5 categorie
Bonus	Discreta	Livello esperienza guidatore raggruppata in 7 categorie
Tipo	Discreta	Tipo di Veicolo raggruppati in 9 categorie
Zona	Discreta	Zona Geografica raggruppata in 7 categorie

Tabella 8.1. Le variabili del *dataset* `SwedishMotorInsurance`.

La matrice di correlazione tra le variabili del *dataset* è contenuta nella tabella 8.2.

	V. Assicurato	Sinistri	Pagamenti	Chilometri	Bonus	Tipo	Zona
V. Assicurato	1.00	0.91	0.93	−0.11	0.16	0.18	−0.05
Sinistri	0.91	1.00	0.99	−0.12	0.10	0.25	−0.11
Pagamenti	0.93	0.99	1.00	−0.12	0.11	0.24	−0.10
Chilometri	−0.11	−0.12	−0.12	1.00	0.00	−0.00	−0.01
Bonus	0.16	0.10	0.11	0.00	1.00	0.00	0.01
Tipo	0.18	0.25	0.24	−0.00	0.00	1.00	−0.00
Zona	−0.05	−0.11	−0.10	−0.01	0.01	−0.00	1.00

Tabella 8.2. Le correlazioni delle variabili del *dataset* `SwedishMotorInsurance`.

Mediante `corrplot` con riordino automatico si può avviare una prima esplorazione grafica della matrice di correlazione che ci aiuta nella ricerca delle variabili più "interessanti" su cui indagare[1].

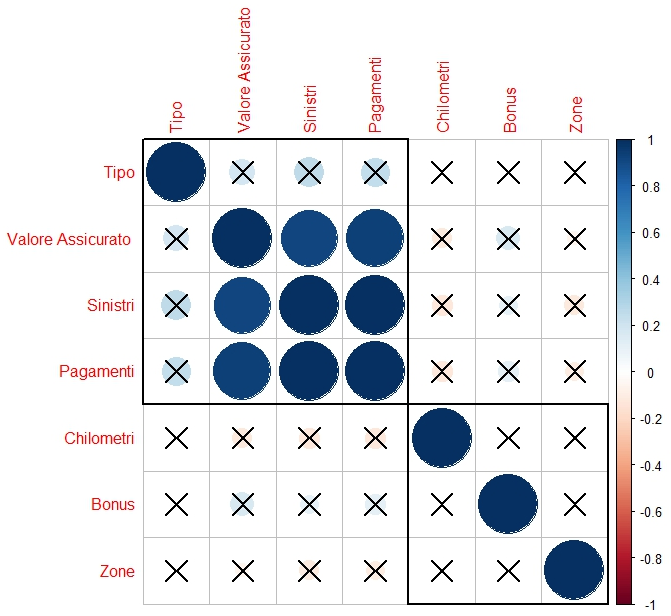


Figura 8.1. Correlazioni con riordino.

[1]Sempre forti del fatto che correlazione nulla non implica indipendenza.

Si può osservare una forte correlazione tra le seguenti variabili:

- Sinistri;
- Pagamenti;
- Valore Assicurato.

Si incomincia l'esame delle variabili analizzando l'eventuale dipendenza tra le variabili continue del *dataset*, si consideri quindi la terna `(Sinistri, Pagamenti, Valore Assicurato)`. Le correlazioni di quest'ultime sono elevate e differiscono in modo non rilevante tra le diverse misure di dipendenza quali ρ di Pearson, τ_K di Kendall e ρ_S Spearman.

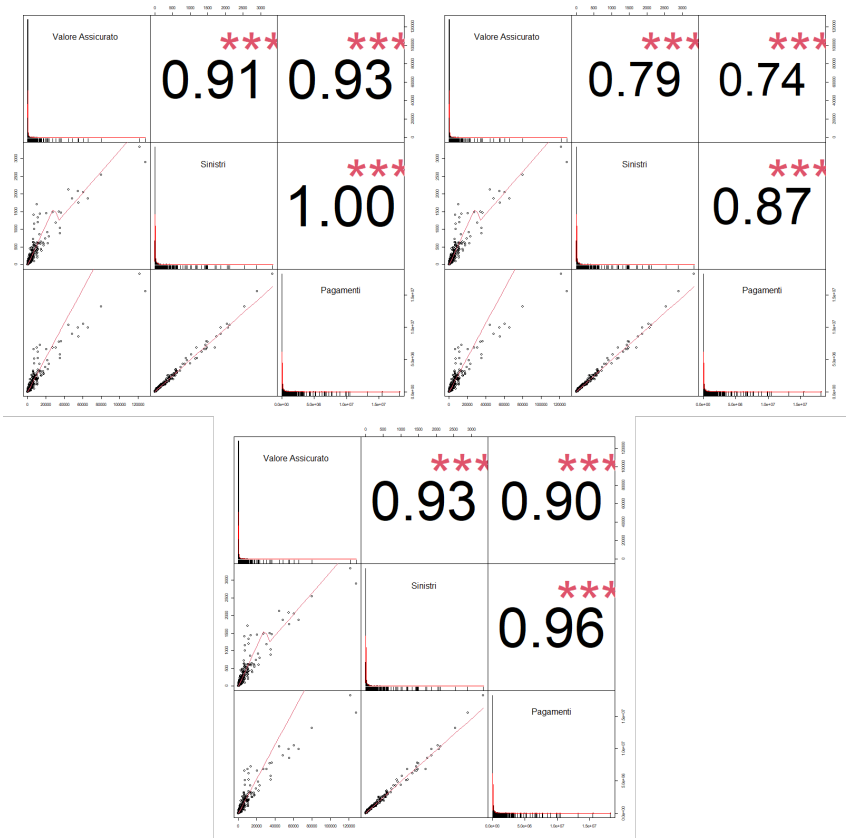


Figura 8.2. Correlazioni Pearson, Kendall e Spearman.

Inoltre dall'ossevazione dei grafici delle frequenze in figura 8.3 non emerge nulla di rilevante.

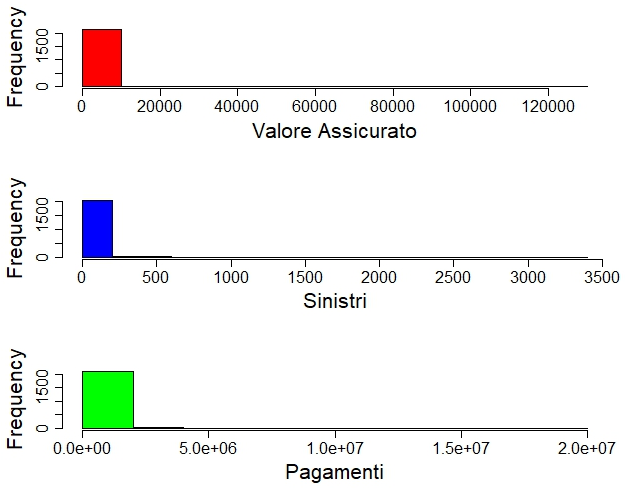


Figura 8.3. Grafici delle frequenze.

Per lavorare con le copule è necessario eseguire delle trasformazioni dei dati. In letteratura ne sono state definite diverse, le più usate sono le seguenti:

Definizione 8.1 (Trasformazioni dei dati)**.** *Si definiscono le seguenti scale di variabili (Czado (2019)):*

- Scala Originale*:* (X_1, X_2) *con densità* $f(x_1, x_2)$*;*
- Dati Copula-Data*:* (U_1, U_2) *dove* $U_i := F_i(X_i)$ *e densità copula* $c(u_1, u_2)$*;*
- Dati Normalizzati*:* (Z_1, Z_2)*, dove* $Z_i := \Phi^{-1}(U_i) = \Phi^{-1}(F_i(X_i))$ *e densità:*

$$g(z_1, z_2) = c((z_1), (z_2))\phi(z_1)\phi(z_2) \tag{8.1}$$

Nella figura 8.4 che segue (partendo da sinistra) si ossevano i dati empirici, ovvero i dati in scala originale, poi in *copula-data-scale* detti *u-scale* ed infine quelli normalizzati detti anche *z-scale*. Dove con la notazione $\Phi(\cdot)$ e ϕ si indichi rispettivamente la consueta funzione di ripartizione e funzione di densità di una variabile Normale Standard.

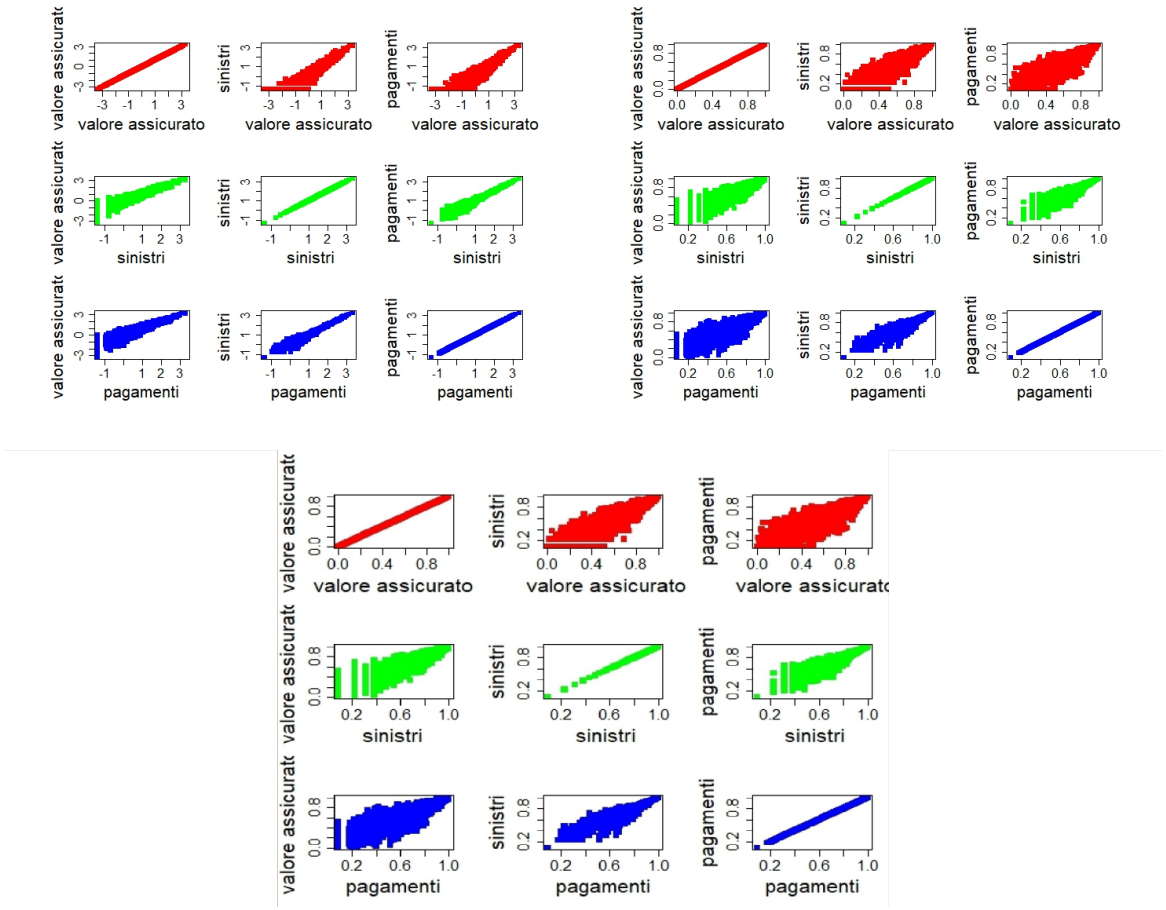


Figura 8.4. Trasformazione dei dati.

Nel tradizionale approccio inferenziale, quando si applica la metodologia delle copule si parte sempre dall'analisi approfondita del *dataset*.

In particolare l'esame dei dati richiede la disamina delle curve di livello empiriche ottenute dai dati e il confronto di quest'ultime con le curve di livello teoriche delle diverse famiglie di copule.

Si osserva che le curve di livello empiriche dei dati (visibili nella figura 8.5) presentano delle dipendenze sicuramente non appartenenti alla famiglia della copule ellittiche. Le curve di livello presentano una marcato schiacciamento lungo la bisettrice del primo quadrante.

Si ritengono da un'attenta comparazione delle forme note che per una più appropriata rappresentazione della struttura di dipendenza della terna `(Sinistri, Pagamenti, Valore Assicurato)` le copule archimedee quali le copule di Clayton, di Gumbel e di Joe appartenenti alla famiglia delle copule parametriche (capitolo 3) che presentano la particolare e caratteristica forma a goccia che si ritrova anche

nelle curve di livello empiriche. In particolare nella figura 8.5 si ritrovano anche i valori di τ_K di Kendall.

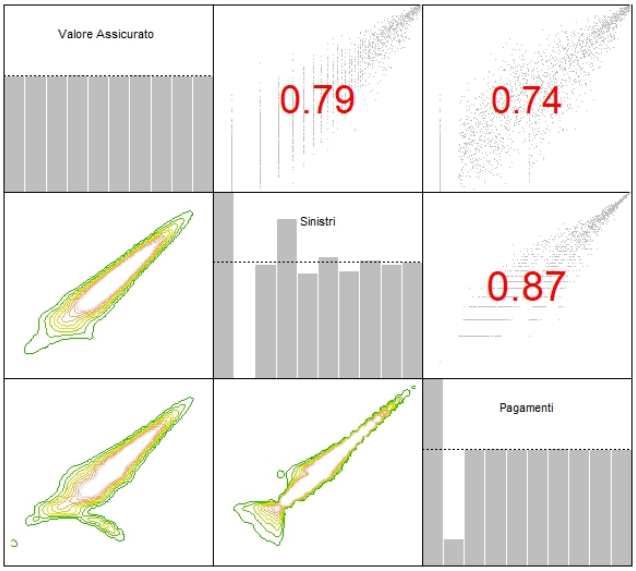


Figura 8.5. Curve di livello empiriche.

Una volta identificate le copule più appropriate alla struttura di dipendenza della terna di variabili `(Sinistri, Pagamenti, Valore Assicurato)` devono essere calcolati i valori dei parametri delle diverse copule bivariate.

Nella tabella 8.3 si riportano i valori dei parametri delle diverse copule determinati attraverso la funzione `BiCopTau2Par`, questa funzione calcola il parametro della copula partendo dal rispettivo valore del τ_K di Kendall:

Clayton	Gumbel	Joe
7.7	3.9	13.8

Tabella 8.3. Parametri delle copule `(Sinistri, Pagamenti, Valore Assicurato)`.

In questo modo è possibile ottenere le copule per descrivere la struttura di dipendenza tra le variabili. In particolare si osservano le curve di livello delle copule ottenute e confrontarle con quelle empiriche osservate precedentemente.

In particolare, si osserva (figura 8.6) che la copula di Joe utilizzata per rappresentare la dipendenza per le variabili `(Sinistri, Valore Assicurato)` sembra non adattarsi al meglio ai dati empirici.

Infine, si osserva la presenza di code pesanti (inferiori) per tutte e tre le coppie di copule bivariate (figura 8.7).

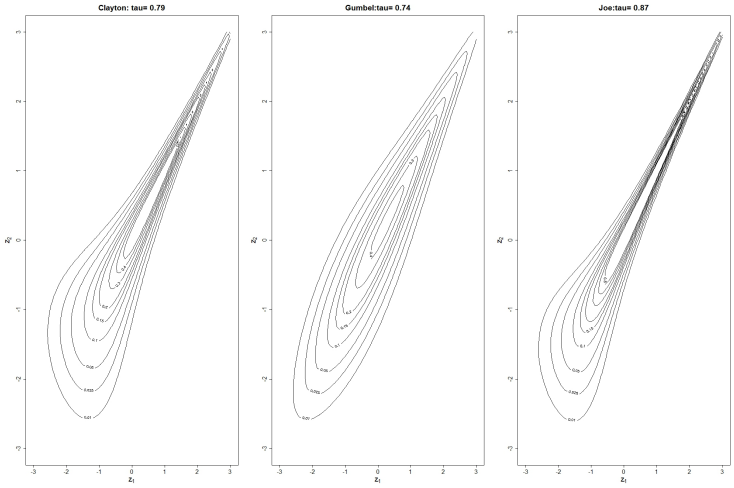


Figura 8.6. Curve di livello nell'approccio inferenziale.

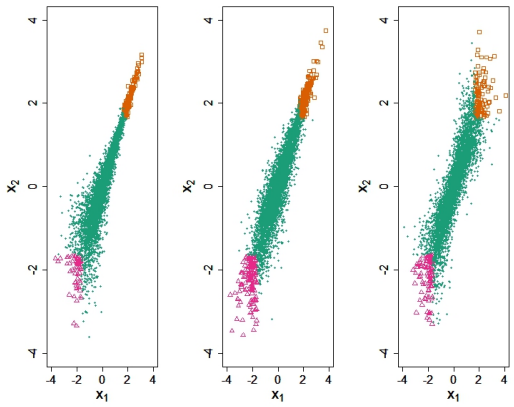


Figura 8.7. Illustrazione delle *tail*.

Si segnala che si possono trovare i modelli migliori rispetto a quello presentato attraverso o funzioni di *best fitting* che individuano la bi-copula in modo automatico oppure attraverso i tradizionali indicatori di adattamento. La finalità di questa trattazione era presentare l'approccio "base" al metodo inferenziale per la teoria delle copule.

8.3 Approccio *Pair-Copula Construction*

Si inizia ora ad applicare la metodologia *pair-copula construction* mediante l'approccio che fa uso dell'algoritmo di ottimo locale definito da Dißmann che è stato visto nel capitolo 5 alla terna `(Valore Assicurato, Sinistri, Pagamenti)`.

Nella tabella 8.4 si indicano le decodifiche delle marginali provenienti dall'*output* dell'algoritmo di Dißmann implementato con la libreria `rvinecopulib` di `R`.

Marginali	Nome
1	Valore Assicurato
2	Sinistri
3	Pagamenti

Tabella 8.4. Indici relativi alle marginali.

I tempi di calcolo sono molto ridotti, infatti la struttura determinata dall'algoritmo è stata definita in soli 1.9 secondi, mentre con l'approccio classico il tempo di implementazione è stato molto più lungo.

Si ricorda che nel tradizionale approccio copula, visto nel precedente paragrafo, l'implementazione del modello è stata effettuata mediante numerosi *step*. Inoltre si segnala la presenza di valutazioni arbitrarie, seppur "ragionevoli" comunque arbitrarie, come ad esempio la *scelta* tra le diverse famiglie di copule.

Come segnalato nel capitolo 5, la metrica di default per la scelta della *vine* nell'algoritmo di Dißmann è il τ_K di Kendall. Oltre ai risultati che si riportano in questa trattazione sono stati implementati dei modelli che utilizzano altri criteri di ottimo locale. Per questo specifico *dataset*, tutti i criteri disponibili nella libreria `rvinecopulib` presentavano lo stesso sentiero al primo albero. Si è quindi deciso di non riportare questi risultati. Si indicherà d'ora in poi con la notazione $\mathcal{M}_{\mathcal{D}}$[2] il modello implementato con l'approccio *pair-copula construction* mediante Dißmann.

L'*output* ottenuto è presentato in tabella 8.5. Trattandosi di un modello di dimensione 3, il modello generato è sia una *C-vine* che una *D-vine*. La variabile al centro del modello è `Sinistri`. In altre parole, le bi-copule del primo albero sono $(1,2)$ (`Valore Assicurato, Sinistri`) e $(2,3)$ (`Sinistri, Pagamenti`) mentre la bi-copula al secondo albero è $(1,3|2)$ ovvero `(Valore Assicurato, Pagamenti)` condizionato alla variabile sinistri. I due alberi della *vine* sono rappresentati nella figura 8.8.

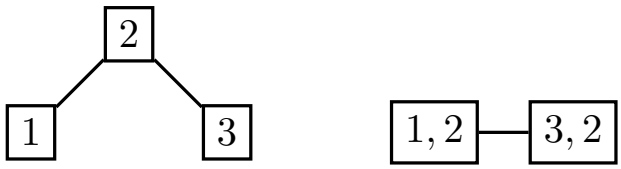
Nella figura 8.9 si può comparare le curve di livello empiriche contro quelle del modello, mentre la tabella 8.6 contiene i principali indicatori statistici di adattamento relativi al modello $\mathcal{M}_{\mathcal{D}}$ generato con l'algoritmo di Dißmann.

Albero	Arco	Vincolo	Copula	Rot.	Par.	df	τ	`loglik`
1	1	$(1,2)$	`bb7`	0	4.57, 1×10^{-6}	2	0.65	3040.34
1	2	$(2,3)$	`bb7`	0	5.99, 24.99	2	0.86	4184.04
2	1	$(1,3\|2)$	`gumbel`	0	1.027	1	0.03	54.96

Tabella 8.5. Modello di $\mathcal{M}_{\mathcal{D}}$.

Considerato il numero ridotto di variabili, $n = 3$, e che il numero di *D-vine* è pari a $n!/2 = 3!/2 = 3$. Forzando l'algoritmo sulla struttura, è possibile confrontare il modello $\mathcal{M}_{\mathcal{D}}$ con i modelli generati sugli altri due sentieri: il sentiero $(1,3,2)$ (il

[2] La $\mathcal{D}$ rimanda a *Dißmann*.



(a) Albero 1 (b) Albero 2

Figura 8.8. Alberi generati con l'algoritmo di Dißmann.

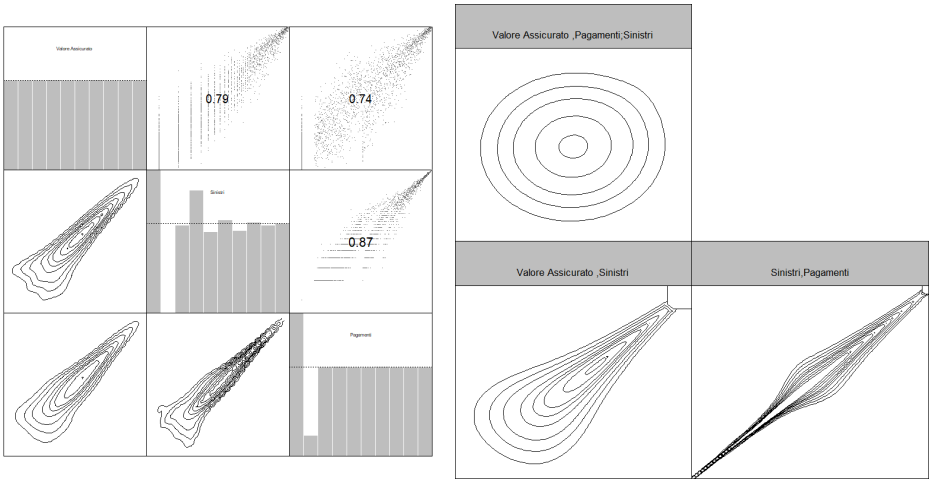


Figura 8.9. Curve di livello empiriche e quelle di $\mathcal{M}_{\mathcal{D}}$.

`log lik`	Numero Parametri	AIC	BIC	mBIC
7279.35	5	−14548.7	−14520.26	−14507.73

Tabella 8.6. Principali indicatori statistici del modello $\mathcal{M}_{\mathcal{D}}$.

cui modello è espresso in tabella 8.7) e il sentiero $(1, 3, 2)$ (il cui modello è espresso in tabella 8.8).

Albero	Arco	Vincolo	Copula	Rot.	Par.	df	τ	`log lik`
1	1	$(1, 3)$	`bb7`	0	$6, 1 \times 10^{-6}$	2	0.72	4974
1	2	$(3, 2)$	`bb7`	0	5, 25	2	0.86	4184
2	1	$(2, 3\|1)$	`frank`	0	2.8	1	0.29	204

Tabella 8.7. Modello $\mathcal{M}_2$ relativo al sentiero $(1, 3, 2)$.

Nella figura 8.10 si possono confrontare le curve di livello dei modelli $\mathcal{M}_2$ e $\mathcal{M}_3$. La tabella 8.9 contiene i valori dei principali indicatori per i tre modelli. I risultati dei `test di Voung` sono riportati in tabella 8.10 e quelli dei `test di Clarke` nella

Albero	Arco	Vincolo	Copula	Rot.	Par.	df	τ	`log lik`
1	1	$(2,1)$	`bb7`	0	4.6, 1×10^{-6}	2	0.65	3040
1	2	$(1,3)$	`bb7`	0	6, 1×10^{-6}	2	0.72	4974
2	1	$(2,3\|1)$	`bb1`	0	2.1, 1.1	2	0.54	1045

Tabella 8.8. Modello $\mathcal{M}_3$ relativo al sentiero $(2,1,3)$.

tabella 8.11.

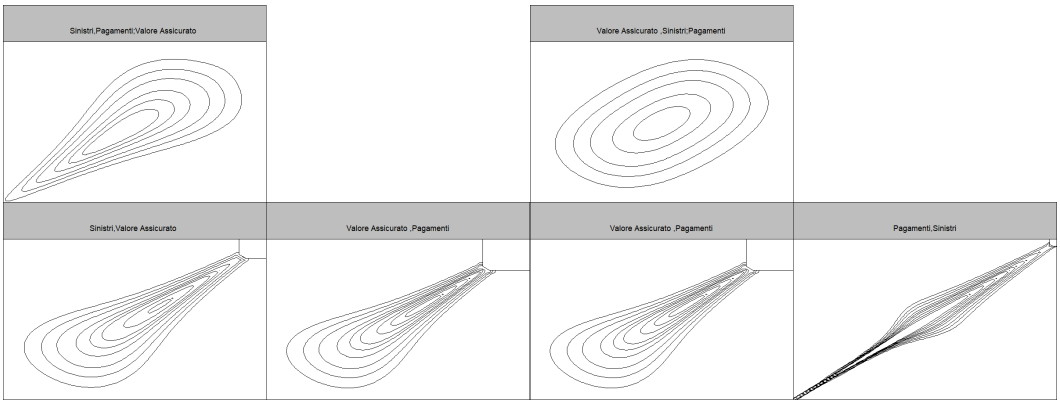


Figura 8.10. Curve di livello dei Modelli $\mathcal{M}_2$ (a sinistra) e $\mathcal{M}_3$ (a destra).

Modello	`log lik`	Numero Parametri	AIC	BIC	mBIC
$\mathcal{M}_{\mathcal{D}}$	7279.35	5	-14548.7	-14520.26	-14507.73
$\mathcal{M}_2$	9362.42	5	-18714.85	-14520.26	-18686.41
$\mathcal{M}_3$	9059.78	6	-18107.57	-18107.57	-18073.44

Tabella 8.9. Principali indicatori statistici per i 3 modelli.

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
$(\mathcal{M}_{\mathcal{D}},\mathcal{M}_2)$	-4.83	-4.83	-4.83	1.32×10^{-6}	1.32×10^{-6}	1.32×10^{-6}
$(\mathcal{M}_{\mathcal{D}},\mathcal{M}_3)$	-3.64	-3.64	-3.63	0.0002	0.0002	0.0002
$(\mathcal{M}_2,\mathcal{M}_3)$	1.009	1.01	1.021	0.31	0.31	0.306

Tabella 8.10. Risultati del test di Vuong trai 3 modelli.

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
$(\mathcal{M}_{\mathcal{D}},\mathcal{M}_2)$	755	755	755	1.58×10^{47}	1.58×10^{47}	1.58×10^{47}
$(\mathcal{M}_{\mathcal{D}},\mathcal{M}_3)$	1307	1307	1308	$\simeq 0$	$\simeq 0$	$\simeq 0$
$(\mathcal{M}_2,\mathcal{M}_3)$	1609	1610	1610	$\simeq 0$	$\simeq 0$	$\simeq 0$

Tabella 8.11. Risultati del test di Clarke trai 3 modelli.

In particolare, si osserva che il modello selezionato dall'algoritmo di Dißmann con il criterio definito di *default* (il τ_K di Kendall) seleziona la *R-vine* peggiore.

Infatti, dal confronto del modello $\mathcal{M}_{\mathcal{D}}$ sia con $\mathcal{M}_2$ che con $\mathcal{M}_3$, i valori dei principali indicatori, posizionano $\mathcal{M}_{\mathcal{D}}$ in entrambi i casi come "fanalino di coda". Inoltre si osserva, dall'analisi delle risultanze dei `test di Vuong` e `test di Clarke` tale differenza è rilevante.

In particolare, si osserva che secondo `Voung`, il modello $\mathcal{M}_{\mathcal{D}}$, ovvero quello ottenuto attraverso l'algoritmo di Dißmann, è significativamente peggiore rispetto sia a $\mathcal{M}_2$ sia a $\mathcal{M}_3$ come mostrano i valori negativi del test. Diversamente, il test di `Voung` non ritiene che la differenza tra $\mathcal{M}_2$ e $\mathcal{M}_3$ sia significativa (i *p-value* sono superiori a 0.05).

Il `test di Clarke` mostra un'altra versione: $\mathcal{M}_2$ è in maniera significativa migliore sia rispetto a $\mathcal{M}_{\mathcal{D}}$ che $\mathcal{M}_3$. Infatti si osseva che i valori sono nettamente inferiori a $\frac{2182}{2} = 1091$ per il test $(\mathcal{M}_{\mathcal{D}}, \mathcal{M}_2)$ e superiori a quel valore nel test $(\mathcal{M}_2, \mathcal{M}_3)$, mentre ritiene che $\mathcal{M}_3$ sia significativamente peggiore a $\mathcal{M}_{\mathcal{D}}$.

Come già discusso nel capitolo relativo alla modellazione inferenziale (capitolo 5), i due test non hanno la stessa ipotesi nulla. In particolare, la diversità dei risultato è legata al fatto che il `test di Clarke` ignora la differenza tra le log-verosimiglianze. Pertanto, $\mathcal{M}_{\mathcal{D}}$ avrà log-verosimiglianza maggiore rispetto a $\mathcal{M}_2$ per 1307 osservazioni, ma il `test di Vuong` considera "quanto" le log-verosimiglianze differiscono dando quindi più peso alle restanti $2182 - 1307 = 875$ osservazioni.

8.4 Livelli di complessità

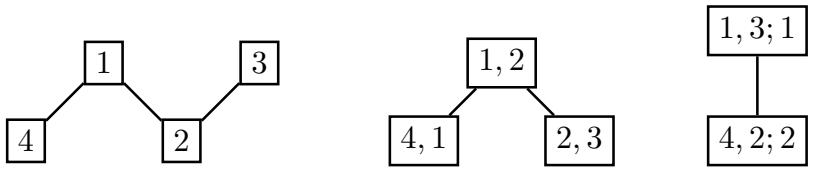
In questa sezione, si include nel modello anche la variabile `Chilometri` che rappresenta la distanza percorsa da un veicolo, che in questo caso assume valori interi $\{1, 2, 3, 4, 5\}$, si è dunque di fronte ad una variabile discreta di tipo *count data*, ovvero di tipo ordinale. Perciò non è stato necessario ricorrere all'utilizzo delle variabili *dummy* che si vedrà nel dettaglio nel capitolo 9.

Similmente al caso con 3 variabili, si riportano in tabella 8.12 gli indici relativi alle marginali nei modelli.

Marginali	Nome
1	Valore Assicurato
2	Sinistri
3	Pagamenti
4	Chilometri

Tabella 8.12. Indici relativi alle marginali.

L'algoritmo di Dißmann lanciato in modo libero determina il modello $\mathcal{M}_{\mathcal{D}}$ (questa volta di dimensione 4), definito da una *D-vine* il cui sentiero può essere riassunto dalla figura 8.11. L'*output* del modello è visibile nella tabella 8.13. I valori dei principali indicatori di adattamento sono elencati in tabella 8.14 mentre nella figura 8.12 si osservano le curve di livello empiriche e del modello $\mathcal{M}_{\mathcal{D}}$.



(a) Albero 1 (b) Albero 2 (c) Albero 3

Figura 8.11. Alberi del modello $\mathcal{M}_{\mathcal{D}}$.

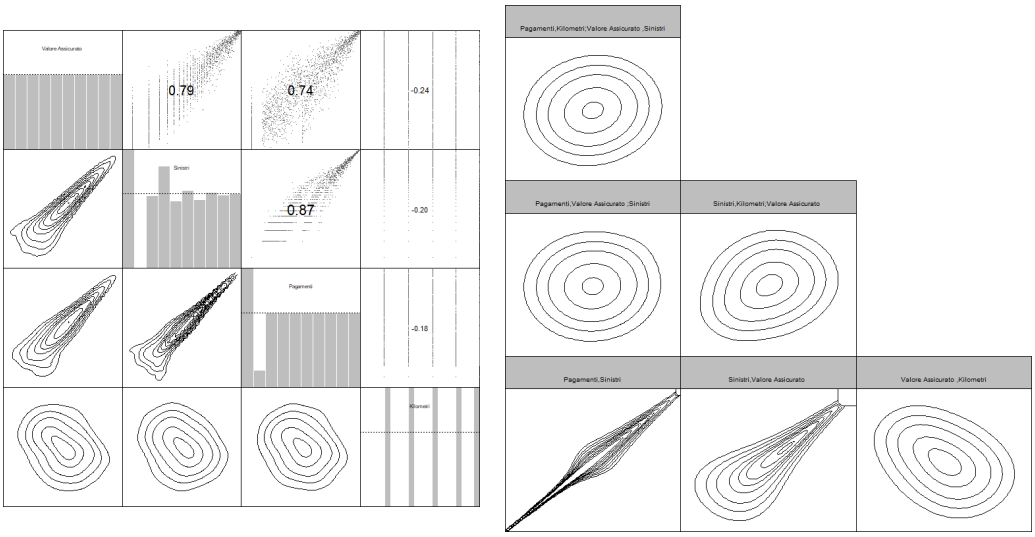


Figura 8.12. Curve di livello empiriche (a sinistra) e del modello $\mathcal{M}_{\mathcal{D}}$ (a destra).

Albero	Arco	Vincolo	Tipo	Copula	Rot.	Par.	df	τ	`log lik`
1	1	$(4,1)$	`d,c`	`bb8`	270	3.0, 0.61	2	-0.230	130.0
1	2	$(1,2)$	`c,c`	`bb7`	0	4.6, 1×10^{-6}	2	0.653	3040.3
1	3	$(2,3)$	`c,c`	`bb7`	0	6, 25	2	0.863	4184.0
2	1	$(4,2\|2)$	`d,c`	`bb8`	180	1.67, 0.79	2	0.134	29.9
2	2	$(1,3\|1)$	`c,c`	`gumbel`	0	1	1	0.026	55.0
3	1	$(4,3\|2,1)$	`d,c`	`frank`	0	0.5	1	0.055	6.1

Tabella 8.13. Modello $\mathcal{M}_{\mathcal{D}}$ con sentiero $(4,1,2,3)$.

`log lik`	Numero Parametri	AIC	BIC	mBIC
7445.33	10	-14870.66	-14813.78	-14790.71

Tabella 8.14. Principali indicatori statistici del modello $\mathcal{M}_{\mathcal{D}}$.

Osservando il modello selezionato dall'algoritmo emerge che la scelta effettuata non è del tutto sorprendente, infatti il sentiero prescelto $(4,1,2,3)$ sembra essere un emanazione del modello $\mathcal{M}_{\mathcal{D}}$ a 3 variabili ovvero $(1,2,3)$, l'ottimo locale per tre

variabili rimane l'ottimo locale anche aggiungendo la varibile `Chilometri`.

Il modello risultante presenta livelli di bontà di adattamento equivalente al modello selezionato nel caso $n = 3$, la domanda che ci si pone è se esiste, come nel caso $n = 3$, un modello migliore per rappresentare la dipendenza tra queste quattro variabili aleatorie.

Nel caso $n = 3$ sono stati implementati 3 modelli che erano equivalenti a tutti i possibili sentieri selezionabili sia nel perimetro delle *C-vine* che nel caso delle *D-vine* e in generale per le *R-vine*, tuttavia si ricorda che il numero di *R-vine* come già detto nei capitoli precedenti è pari a $n! \cdot 2^{\binom{n-2}{2}}$ mentre il numero di *C-vine* e *D-vine* è pari a $\frac{n!}{2}$.

Osservando la tabella 8.15[3] emerge che un *dataset*, come quello utilizzato, avente 4 variabili, ha circa 24 possibili *R-vine*, mentre il *dataset* completo ne possiede 2.5 milioni (ignorando la necessità di creare variabili *dummy*).

n	# *C-vine*/*D-vine*	# *R-vine*
2	1	1
3	3	3
4	12	24
5	60	480
6	360	23040
7	2520	2580480

Tabella 8.15. Numero di viti.

Invece di considerare tutte le 24 *R-vine*, si userà il risultato dei precedenti test per identificare modelli potenzialmente migliori in maniera più efficiente. Quindi, tenendo conto della filtrazione del caso $n = 3$ si possono implementare ulteriori due modelli $\mathcal{M}_2$ e $\mathcal{M}_3$ che "forzando la *vine*" secondo il modello ottimo della dimensione $n = 3$, definiti dalle seguenti *D-vine*:

- $\mathcal{M}_2$ definito dal sentiero $(4, 1, 3, 2)$;
- $\mathcal{M}_3$ definito dal sentiero $(1, 3, 2, 4)$.

I due modelli implementati presentano le curve di livello rappresentate nella figura 8.13 mentre le tabelle con l'*output* dell'algoritmo sono state omesse.

Come si può vedere dal confronto dei principali indicatori presenti nella tabella 8.16 si evince che i due sentieri scelti facendo uso della filtrazione del modello di dimensione 3 ha fornito sia nel caso del $\mathcal{M}_2$ che del $\mathcal{M}_3$ valori migliori per l'intera batteria di test statistici: migliore $log\,lik$, migliore `AIC`, `BIC` e migliore `mBIC`.

Confrontando i tre modelli con `Voung Test` e `Clarke Test`, i cui risultati sono riportati nelle tabelle 8.17 e 8.18, si osserva che $\mathcal{M}_\mathcal{D}$ è significativamente peggiore degli altri due modelli. Mentre dal confronto di $\mathcal{M}_2$ con $\mathcal{M}_3$, $\mathcal{M}_2$ è migliore al netto della correzione di Akaike (0.61). Da questo si desume che la variabile `Chilometri` abbia una migliore livello di concordanza con la variabile valore assicurato, questo fenomeno era già colto dalla selezione implementata dall'algoritmo di *greedy* di

[3]La tabella 8.15 non è altro che un estratto della tabella 4.1 della sezione 4.2.4.

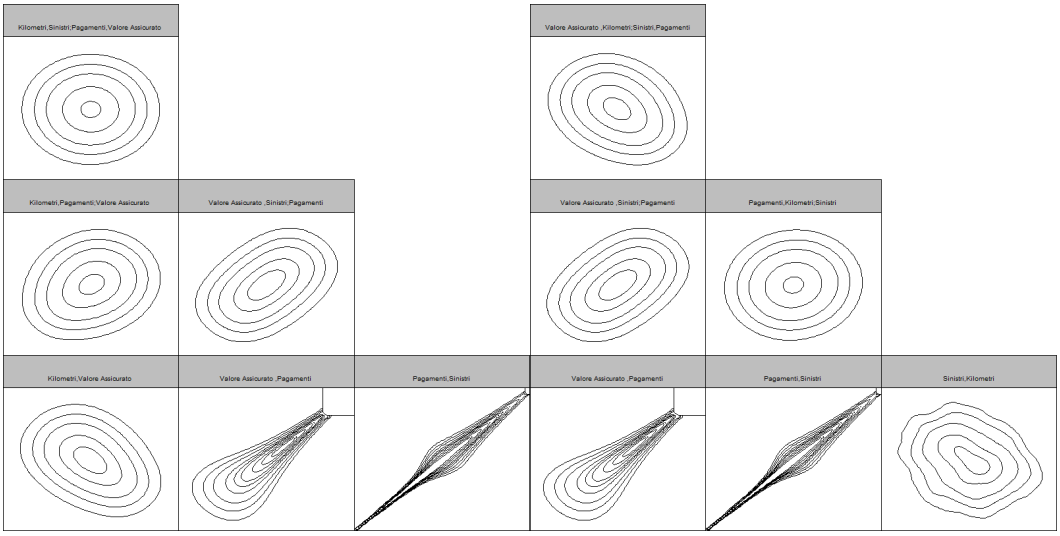


Figura 8.13. Curve di livello dei Modelli $\mathcal{M}_2$ (a sinistra) e $\mathcal{M}_3$ (a destra).

Confronto dei Principali Indicatori					
Modello	`log lik`	Numero Parametri	AIC	BIC	mBIC
$\mathcal{M}_\mathcal{D}$	7445.33	10	−14870.66	−14813.78	−14790.71
$\mathcal{M}_2$	9527.68	9	−19037.35	−18986.16	−18965.07
$\mathcal{M}_3$	9548.41	26	−19044.97	−18897.5	−18874.43

Tabella 8.16. Principali indicatori statistici per i 3 modelli.

Dißmann $(4, 1, 2, 3)$ che generava come primo arco del primo albero proprio l'arco $(4, 1)$.

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
$(\mathcal{M}_\mathcal{D},\mathcal{M}_2)$	−4.83	−4.84	−4.84	1.33×10^{-6}	1.31×10^{-6}	1.27×10^{-6}
$(\mathcal{M}_\mathcal{D},\mathcal{M}_3)$	−4.88	−4.85	−4.74	1.02×10^{-6}	1.23×10^{-6}	2.09×10^{-6}
$(\mathcal{M}_2,\mathcal{M}_3)$	−2.73	−0.5	5.84	0.006	0.62	5.08×10^{-9}

Tabella 8.17. Risultati del test di Vuong trai 3 modelli.

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
$(\mathcal{M}_\mathcal{D},\mathcal{M}_2)$	780	779	776	7.28×10^{-41}	4.03×1^{-41}	6.8×10^{-42}
$(\mathcal{M}_\mathcal{D},\mathcal{M}_3)$	785	800	846	1.34×10^{-39}	6.34×10^{-36}	7.55×10^{-26}
$(\mathcal{M}_2,\mathcal{M}_3)$	968	1024	1168	1.51×10^{-7}	0.004	0.001

Tabella 8.18. Risultati del test di Clarke trai 3 modelli.

8.5 Conclusioni

Da quanto presentato in questo capitolo, emerge in modo più che evidente l'enorme potenzialità della metodologia *pair-copula construction* facente uso dell'algoritmo *greedy* di Dißmann e già attraverso questo primo *case study* si riesce già a cogliere seppur in modo parziale agli aspetti relativi della domanda di ricerca.

Infatti, già da una prima analisi, considerando un numero di variabili limitate, rispetto all'approccio tradizionale si segnalano degli evidenti vantaggi rispetto agli aspetti relativi alla riduzione dei tempi di implementazione.

Il cosidetto "approccio inferenziale classico" mostrato nei primi paragrafi del capitolo mostra un assorbimento dei tempi molto più sfidante (e il modello in approccio tradizionale è stato guardato solo considerando tre variabili).

Di concerto si sottolinea la flessibilità dello strumento, infatti i modelli di tipo misto che sono stati presentati, e che rappresentano un contributo originale a questa metodologia, anche definiti durante questa trattazione modelli di tipo *mixed-tailored* migliorano la stima di Dißmann in modo concreto, e questo vale sia nel perimetro di ridotte dimensioni che in grandi dimensioni.

L'esercizio di calibrazione dei dati è e rimarrà sempre un approccio migliorabile nel continuo attraverso l'uso di metodologie e approcci congiunti.

Infatti dall'analisi di questo *dataset* è emerso che l'algoritmo effettuando "solo" una selezione "localmente ottimale" ad ogni passo, ignora l'impatto di tale selezione sia negli alberi precedenti e sia nei successivi.

Si ricorda che Gruber e Czado (2015) hanno dimostrato che in scenari di simulazione a 6 dimensioni l'algoritmo *greedy* di Dißmann ottiene almeno il 75% della vera log-verosimiglianza che significa una buona prestazione se si considerano dimensioni molto elevate[4].

Il problema della dimensionalità acquista un'importanza fondamentale nelle problematiche di tipo attuariale, questo vale sia per i prodotti danni che vita.

In queste analisi si evidenzia che tale asserto è confermato in quanto le verosimiglianze ottenute sia dal modello a 3 e 4 variabili, sebbene nel caso $n = 4$ non siano stati implementate tutte le possibili *R-vine*, presentano livelli di verosimiglianza analoghi. I risultati dei rapporti sono rappresentati nelle tabelle 8.19 e 8.20.

Modello $n = 3$	`log lik`	Numero Parametri
$\mathcal{M}_{\mathcal{D}}$	7279.35	–
$\mathcal{M}_2$	9362.42	$\frac{7279.35}{9362.42} = 0.78\%$
$\mathcal{M}_3$	9059.78	$\frac{7279.35}{9059.78} = 0.80\%$

Tabella 8.19. Confronto del rapporto di verosimiglianza per i modelli a 3 variabili.

Modello $n = 4$	`log lik`	Numero Parametri
$\mathcal{M}_{\mathcal{D}}$	7445.33	–
$\mathcal{M}_2$	9527.68	$\frac{7445.33}{9527.68} = 0.78\%$
$\mathcal{M}_3$	9548.41	$\frac{7445.33}{9548.41} = 0.78\%$

Tabella 8.20. Confronto del rapporto di verosimiglianza per i modelli a 4 variabili.

[4]Si vedrà il caso dimensione 11 nel capitolo 10

Capitolo 9

Caso studio: distribuzioni con variabili discrete

> I asked Giorgio Dall'Aglio, Ingram Olkin, and Abe Sklar if they thought that there might indeed be interest in the statistical community for such a book. Encouraged by their responses and knowing that I would soon be eligible for a sabbatical, I began to give serious thought to writing an introduction to copulas.
>
> Nelsen (2006)

9.1 Introduzione

Nel capitolo 6 si è visto quali sono le problematiche legate alle variabili discrete. In particolare è stata fatta una distinzione tra:

- le variabili di tipo *count-data* ovvero la variabile discrete di tipo ordinale;
- le variabili *categoriali "pure"* ovvero le variabili in cui "non ha senso" parlare di "ordine".

Inoltre, nel capitolo 8 è stata trattata una variabile discreta di tipo *count-data*, la variabile `Chilometri` che aveva valori nell'insieme $\{1, 2, 3, 4, 5\}$.

In questo capitolo si vedrà cosa accade quando si lavora con variabili di tipo *categoriali "pure"*. Prima vengono presentate alcune osservazioni sulle variabili *dummy*.

Si ricorda che questa trattazione ha un'importanza rilevante nel campo delle scienze attuariali, in quanto le variabili categoriali ricorrono spesso nella modellizzazione dei rischi legati al *business* assicurativo. Si sottolinea che i *dataset* che si stanno utilizzando per le applicazioni sono tutti all'interno del perimetro *non-life*, tuttavia si segnala che anche nel perimetro *life* ricorrono variabili *categoriali "pure"*.

9.2 Le variabili *Dummy*

Come è noto dalla teoria delle probabilità, le variabili aleatorie assumono generalmente valori in sottoinsiemi di $\mathbb{R}$, nella pratica delle analisi statistiche e attuariali non tutte le variabili estratte da un campione hanno un valore numerico intrinseco. Si distinguerà perciò tra variabili *quantitative*, ovvero variabili associate ad un valore numerico misurabile e variabili *qualitative*, ovvero quelle variabili che descrivono qualità del campione che non sono misurabili numericamente.

All'interno delle variabili quantitative è inoltre usuale distinguere tra *variabili quantitative continue* e *variabili quantitative discrete*, a seconda se i valori delle variabili siano ben distinti o distribuiti su un qualche intervallo reale. Molte delle variabili quantitative discrete (anche dette *count-data*) sono prodotte per conteggio di determinati oggetti o eventi. Per esempio, i numero dei sinistri è una variabile quantitativa discreta mentre i valori dei pagamenti legati a quei sinistri è una variabile quantitativa continua.

Le variabili qualitative possono invece essere distinte tra *qualitative ordinali* e *qualitative nominali*. Le prime sono variabili qualitative su cui esiste un ordine naturale (come per esempio le classi di rischio, il *rating* finanziario o il livello di istruzione) mentre le seconde si limitano a descrivere caratteristiche come, per esempio, il genere o la residenza.

Per prima cosa è utile osservare che è sempre possibile assegnare ad una variabile qualitativa ordinale dei valori numerici. Perciò, le variabili qualitative ordinali si distinguono dalle quantitative discrete per l'importanza data agli esatti valori numerici assegnati. Non avrà quindi senso calcolare la correlazione tra una variabile qualitativa ordinale e una qualche altra variabile ma è senz'altro possibile analizzare la concordanza usando il τ_K di Kendall o il ρ_s di Spearman (nella loro forma probabilistica).

Inoltre si ricorda che le copule non "danno importanza" ai valori assunti dalle variabili, infatti si ha spesso detto in questa trattazione che lavorano su scala quantilica (capitolo 1).

Poichè le variabili nominali non possiedono un ordinamento naturale, non ha senso parlare di concordanza o correlazione tra loro e altre variabili. Questi concetti vengono alle volte sostituiti con quelli di bontà di classificazione, ovvero si interpreta la variabile nominale come una classificazione dei dati e si misura quanto questa classificazione sia espressa nelle altre variabili.

Va però osservato che quando la variabile ha due sole realizzazioni[1], ovvero quando la variabile è dicotomica, l'assegnazione di un ordine non presenta un problema.

Importante evidenziare che quando il numero di variabili *dummy* è uguale al numero di valori che la variabile può assumere, si ricade nella perfetta *multicollinearità* delle variabili il che provoca un calcolo errato delle misure di condordanza e dei valori di p-value come afferma Suits (1957):

> *If dummy variables for all categories were included, their sum would equal 1 for all observations, which is identical to and hence perfectly*

[1] Come ad esempio il genere.

> *correlated with the vector-of-ones variable whose coefficient is the constant term; if the vector-of-ones variable were also present, this would result in perfect multicollinearity, so that the matrix inversion in the estimation algorithm would be impossible. This is referred to as the dummy variable trap.*

Dove con *the dummy variable trap* s'intende proprio la multicollinearità artificiale indotta dalla trasformazione di una variabile con più realizzazioni a più variabili dicotomiche, quindi proprio "una trappola" innescata da queste variabili fittizie. Questa criticità si risolve utilizzando $n-1$ variabili anzichè n.

Di fatto un metodo assolutamente efficace per gestire il fenomeno della collinearità artificiale è quello di trasformare una variabile qualitativa categoriale avente n realizzazioni con $n-1$ variabili dicotomiche: queste variabili sono dette variabile *dummy* o di *comodo*.

Esempio 9.1 (Variabili *Dummy*)**.** *Si considera un caso molto semplice a titolo esemplificativo, si supponga di avere un* dataset *formato dalle seguenti* 3 *variabili:*

1. *Reddito;*
2. *Genere ("Maschio", "Femmina");*
3. *Stato Civile ("Sposato", "Single", "Divorziato").*

E si supponga di voler utilizzare lo stato civile e l'età per studiare gli effetti sul reddito. La tabella 9.1 contiene un possibile una possibile realizzazione di questo dataset*.*

Reddito	Genere	Stato Civile
45000€	M	Single
48000€	F	Single
54000€	M	Single
57000€	M	Single
65000€	F	Sposata
69000€	F	Single
78000€	F	Sposata
83000€	M	Divorziato
98000€	M	Divorziato
104000€	F	Sposata
107000€	M	Sposato

Tabella 9.1. Dataset di esempio.

Quando si utilizzano variabili categoriali si osserva che non ha comunque senso "semplicemente" trasformare le variabili categoriali in numeri, quindi nella presente trattazione le realizzazioni del genere "Sposato", "Single", "Divorziato" in `{1, 2, 3}` *perché non significa nulla dire che il "Single" vale il doppio dello "Sposato" o che il "Divorziato" vale il triplo dello "Sposato". Inoltre, si osserva che utilizzando l'approccio "ad ogni variabile con n realizzazioni si fa corrispondere n*

variabili dicotomiche" determina un caso di perfetta correlazione per le variabili "Single" e "Sposato", come mostrato in tabella 9.2.

Reddito	Maschio	Donna	Single	Sposato/a	Divorziato/a
45000€	1	0	1	0	0
48000€	0	1	1	0	0
54000€	1	0	1	0	0
57000€	1	0	1	0	0
65000€	0	1	0	1	0
69000€	0	1	1	0	1
78000€	0	1	0	1	0
83000€	0	0	1	0	1
98000€	0	0	0	0	1
104000€	0	1	0	1	0
107000€	1	0	0	1	0

Tabella 9.2. Dataset di esempio con variabili *dummy* per ogni valore.

Infatti, le variabili "Single" e "Sposato" hanno un coefficiente di correlazione pari a -1*. Pertanto, quando si andrà a eseguire i calcoli dei coefficienti di correlazione i valori non saranno corretti. Per evitare la trappola delle variabili* dummy *è sufficiente ricordare una sola regola "a* n *realizzazioni corrispondono* $n-1$ *variabili* dummy*". Dunque se una variabile categorica può assumere* n *valori diversi si devono creare "solo"* $n-1$ *variabili* dummy *da utilizzare. Nel caso in esame, essendo* 3 *le realizzazioni della variabile "Stato Civile" sarà quindi semplicemente formato da* 2 *variabili* dummy*.*

Le variabili dummy *saranno semplicemente:*

$$X_i = \begin{cases} 1 \text{ se sposato} \\ 0 \text{ altrimenti} \end{cases}$$

Per ogni $i = 1, 2$*.*

La tabella 9.3 presenta il dataset *convertito con variabili* dummy *evitando il problema della multicollinearità.*

In questo modo il problema di multicollinearità è superato, questo approccio con variabili *dummy* viene ampiamente utilizzato nella libreria `rvinecopulib` di `R` nell'uso delle variabili discrete: in questo caso, la libreria sostituirà la variabile con un numero adeguato di variabili *dummy*. Si segnala che dai test effettuati per l'implementazione che seguirà è emerso una maggiore variabilità dei risultati nell'analisi di variabili discrete rispetto al caso continuo ovvero lanciando lo stesso modello i risultati presentano una varibilità, seppur contenuta comunque da portare all'attenzione perchè il valore di tale "finestra mobile" delle risultanze influisce in maniera diretta sui principali test di confronto.

Questo fenomeno comunque non ci sorprende in quanto la libreria `rvinecopulib` di `R` ricorre ad algoritmi *randomizzati*, quali la metodologia della stima *kernel* con approccio di *jittering* la cui logica è stata presenta nel capitolo 6.

Reddito	Donna	Sposato/a	Divorziato/a
45000€	0	0	0
48000€	1	0	0
54000€	0	0	0
57000€	0	0	0
65000€	1	1	0
69000€	1	0	0
78000€	1	1	0
83000€	0	0	1
98000€	0	0	1
104000€	1	1	0
107000€	0	1	0

Tabella 9.3. Dataset di esempio con variabili *dummy* senza multicollinearità.

9.3 Applicazione: variabili discrete non *count-data*

Si consideri un *dataset* diverso da quello usato nel capitolo precedente che conteneva variabili discrete di tipo ordinali. In questo capitolo si tratteranno variabili di tipo discreto non *count-data*. Si userà il *dataset* `dataCar` contenuto nella libreria `insuranceData`, implementato dalla *Macquarie University*[2] e utilizzato da Jong e Heller (2008).

Il *dataset* contiene informazioni riguardanti un portafoglio di 67.856 contratti stipulati nella formula mono-annuale tra il 2004 e il 2005 e presenta 10 variabili: 3 di tipo continuo e le rimanenti sono di tipo discreto. Una descrizione completa del *dataset* `dataCar` è contenuto nella tabella 9.4.

Variabile	Tipo	Descrizione	Valori
`veh_value`	c	Valore del Veicolo	$[0, 35]$
`exposure`	c	Esposizione	$[0, 1]$
`clm`	d	Esistenza di Sinistro	$\{0, 1\}$
`numclaims`	d	Numero di Sinistri	$\{1, 2, 3, 4\}$
`claimcst0`	c	Ammontare	$[0, 56000]$
`veh_body`	d	Tipo di Veicolo	{`BUS`, `CONVT`, `COUPE`, `HBACK`, `HDTOP`, `MCARA`, `MIBUS`, `PANVN RDSTR`, `SEDAN`, `STNWG`, `TRUCK`, `UTE`}
`veh_age`	d	Età del Veicolo	$\{1, 2, 3, 4\}$
`gender`	d	Sesso	{`M`, `F`}
`area`	d	Area Geografica	{`A`,`B`,`C`,`D`,`E`,`F`}
`agecat`	d	Gruppo di Età	$\{1, 2, 3, 4, 5, 6\}$

Tabella 9.4. Descrizione *dataset* `dataCar`.

Dove la variabile `numclaims` è definita nel seguente modo:

$$\texttt{numclaims} = \begin{cases} 1 & \text{Non si è verificato alcun sinitro.} \\ 0 & \text{Si è verificato almeno un sinistro.} \end{cases} \tag{9.1}$$

[2]Department of Applied Finance and Actuarial Studies.

9.4 Modello a 3 variabili

Per esplorare il funzionamento dell'algoritmo con variabili discrete e prenderne familiarità, si considerino le variabili della tabella 9.5:

Marginali	Nome Variabile
1	`numclaims`
2	`exposure`
3	`xM`

Tabella 9.5. Tabella delle variabili marginali.

Questo caso studio verrà esplorato implementando quattro modelli distinti che differiscono tra loro o per le modalità con cui viene trattata la variabile `gender` o rispetto ai vincoli sulle famiglie di bi-copule. In particolare si indicherà con:

- $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ il modello che tratta `gender` come variabile categoriale;
- $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ il modello che tratta `gender` come variabile categoriale e che si avvale in modo esclusivo di bi-copule parametriche;
- $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ il modello che tratta `gender` come variabile continua;
- $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$ il modello che tratta `gender` come variabile continua e che si avvale in modo esclusivo di bi-copule parametriche.

In questo modo si avrà la possibilità di esplorare lo strumento della *pair-copula construction* in campo categoriale e analizzare come il modello varia al variare delle modalità attraverso cui la variabile `gender` viene trattata.

Nel caso dei modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ si tratteranno le realizzazioni della variabile `gender` come variabile categoriale pura, ovvero si consideri l'insieme `{M,F}` come un insieme contenente due elementi `{M}` e `{F}` sul quale non vi è definito alcun concetto di ordine.

Diversamente con i modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$ si tratteranno le realizzazioni della variabile `gender` come variabile continua, ovvero si considererà l'insieme `{M,F}` come l'insieme `{0,1}`.

I modelli $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$ rappresentano gli omologhi di $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ vincolati all'approccio parametrico. Le motivazioni di questa scelta acquisteranno un senso di ragionevolezza quando si osserverà il numero di parametri dei modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$. In questi approcci quindi, verranno escluse le copule non-parametriche calcolate con la *transformation kernel*[3].

[3] Anche questo è un aspetto è stato discusso nel capitolo 6.

9.4.1 Caso 1: gender come variabile categoriale

In questo primo modello, si osserva che i tempi computazionali sono elevati se si considera il numero di variabili: 102 secondi. Come annunciato, il trattamento delle variabili categoriali non solo risulta più delicato dal punto di vista teorico ma presenta un assorbimento di calcolo computazionale superiore[4]. Come si nota, il modello è comunque costituito da 3 variabili, in quanto le "famose" $n-1$ realizzazioni del `gender` sono pari a $2-1=1$.

Di seguito i consueti *output* del modello: la tabella 9.6 contiene la descrizione del modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ e la figura 9.1 il grafico del primo albero del modello.

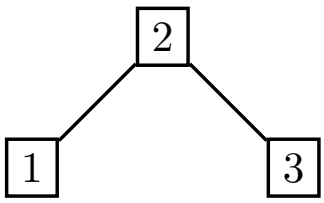


Figura 9.1. Sentiero del primo Albero T_1 del modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$.

Albero	Arco	Vincolo	Tipo	Copula	Rot.	Par.	df	τ	`log lik`
1	1	$(1,2)$	`c,c`	`gaussian`	0	0.17	1	0.110	511
1	2	$(2,3)$	`c,d`	`tll`	0	$[30\times 30]$	1	0.005	23
2	1	$(1,3\|2)$	`c,d`	`tll`	0	$[30\times 30]$	30	0.110	7694

Tabella 9.6. Modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$.

9.4.2 Caso 2: gender come variabile categoriale con approccio parametrico

In questo primo approccio, si vede che i tempi computazionali sono ancora più elevati: 130 secondi. Evidentemente nel caso `1`, l'algoritmo esclude determinate famiglie attraverso test statistici, riuscendo quindi a ridurre i tempi di modellazione delle copule.

Il modello trovato è riportato nella tabella 9.7 e la figura 9.2 presenta le curve di livello dei due modelli. Il modello $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$, similmente al modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$, possiede figura 9.1 come primo albero T_1.

Albero	Arco	Vincolo	Tipo	Copula	Rot.	Par.	df	τ	`log lik`
1	1	$(1,2)$	`c,c`	`gaussian`	0	0.17	1	0.11	511
1	2	$(2,3)$	`c,d`	`joe`	0	1	1	0.01	8.8
2	1	$(1,3\|2)$	`c,d`	`indep`	0		0	0.00	0.00

Tabella 9.7. Modello $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$.

[4]Va comunque osservato che il *dataset* usato nel capitolo 8 possiede molte meno osservazioni e si sono modellate le marginali con le marginali empiriche.

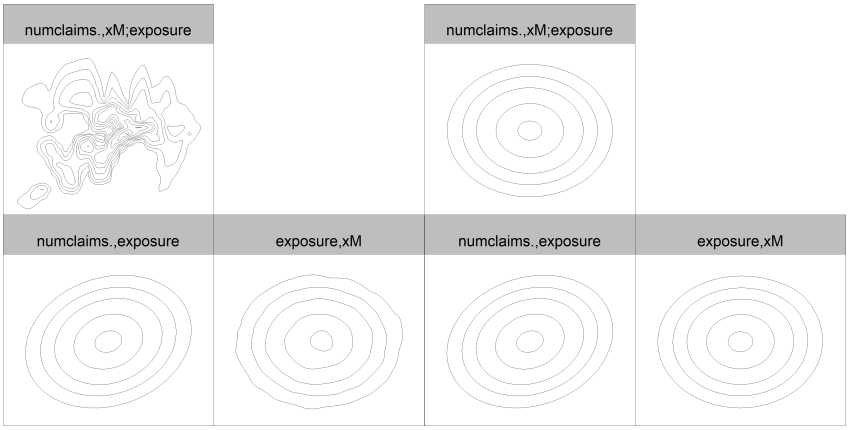


Figura 9.2. Curve di livello (A sinistra il modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ a destra il modello $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$).

I valori dei principali indicatori di adattamento dei modelli $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ sono riportati in tabella 9.8. La tabella 9.9 fa invece riferimento ai risultati dei test statistici.

Modello	`log lik`	`mBIC`	`nPars`
$\mathcal{M}_{\mathcal{D}}(\mathbb{D})$	32436.24	-16091.01	32
$\mathcal{M}_{\mathcal{D}}(\mathcal{D})$	24727.7	-1008.511	2

Tabella 9.8. Indicatori statistici per i modelli $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$.

`Test`	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
`Clarke`	30354	30354	30248	5×10^{-166}	5×10^{-166}	6×10^{-176}
`Voung`	66.86	66.6	65.42	$\simeq 0$	$\simeq 0$	$\simeq 0$

Tabella 9.9. Risultati dei test di Clarke e Voung per $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ `v.s` $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$.

Guardando i risultati di `log lik` e `mBIC`, sembrerebbe che il modello parametrico $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ abbia un *fitting* decisamente peggiore del modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$. Questa analisi è confermata dal `test di Vuong`, tuttavia il `test di Clarke` non è concorde ($67856/2 = 33928 > 30354$). Sorprendentemente, anche se la differenza nel numero dei parametri è notevole, i fattori di correzione di Akaike e Schwarz non arrivano mai a modificare i risultati dei due test.

Si segnala che in considerazione dell'approccio di *jittering* utilizzato per il calcolo delle marginali, sono stati generati diversi modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ e in molti casi $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ risulta significativamente migliore in entrambi i test.

9.4.3 Caso 3: `gender` come variabile continua

Come anticipato con i modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$ si esamina la variabile `gender` come continua. Operativamente è stata effettuata una biezione, trasformando la realizzazione `{M}` nella realizzazione `{1.0}` e la realizzazione `{F}` nella realizzazione

{0.0}. In questo modo si conferisce un "ordine artificiale" alla variabile gender (che non gli appartiene).

I tempi di elaborazione per questo modello sono stati nettamente inferiori: 42 secondi[5].

Inoltre diversamente dai casi precedenti, è stato possibile calcolare la correlazione tra le variabili e come si osserva nella tabella 9.10 risulta essere molto debole.

	Numero Sinistri	Esposizione	gender
Numero Sinistri	1	0.134	−0.002
Esposizione	0.134	1	0.014
gender	−0.002	0.014	1

Tabella 9.10. Matrice di Correlazione tra le tre variabili con gender continua.

Il modello ottenuto usando l'algoritmo di Dißmann è riportato in tabella 9.11.

Albero	Arco	Vincolo	Tipo	Copula	Rot.	Par.	df	τ	log lik
1	1	(1, 2)	c,c	tll	0	$[30 \times 30]$	178	0.110	0.038
1	2	(2, 3)	c,c	clayton	180	0.019	1	0.005	0.009
2	1	(1, 3|2)	c,c	tll	0	$[30 \times 30]$	1099	0.110	0.304

Tabella 9.11. Modello $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$.

9.4.4 Caso 4: gender come variabile continua con approccio parametrico

I tempi di elaborazione sono analoghi al caso non parametrico: 42 secondi. Il risultato dell'elaborazione è visibile nella tabella 9.12.

Albero	Arco	Vincolo	Tipo	Copula	Rot.	Par.	df	τ	log lik
1	1	(1, 2)	c,c	gaussian	0	0.17	1	0.11	511
1	2	(2, 3)	c,c	clayton	180	0.019	1	0.01	8
2	1	(1, 3|2)	c,c	bb7	180	$(1.6, 1 \times 10^{-6})$	2	0.26	2189

Tabella 9.12. Modello $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$.

Le curve di livello dei due modelli con gender come variabile continua sono rappresentati in figura 9.3; mentre i principali indicatori e i risultati dei test statistici possono essere consultati nelle tabelle 9.13 e 9.14.

I valori dei principali indicatori di adattamento e dei test di Voung e Clarke sono riportati nelle tabelle 9.13 e 9.14. In questo caso, il modello non parametrico $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ è unanimemente considerato migliore del modello continuo con approccio parametrico $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$. Infatti, questo aspetto emerge sia dagli indicatori di bontà di adattamento, quali log lik e mBIC, che dai test di Voung e di Clarke.

[5]Si ottiene un abbattimento dei tempi computazionali maggiori del 50%.

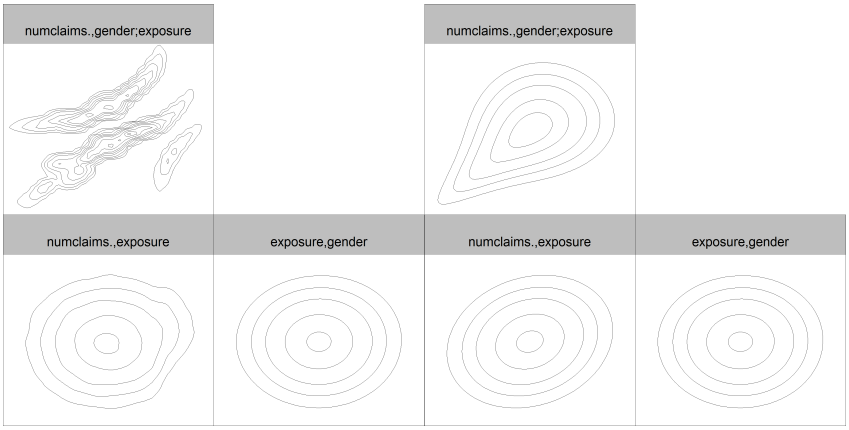


Figura 9.3. Curve di livello (A sinistra il modello $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ a destra il modello $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$).

Modello	`log lik`	`mBIC`	`nPars`
$\mathcal{M}_{\mathcal{D}}(\mathbb{C})$	72562.08	-77320.56	32
$\mathcal{M}_{\mathcal{D}}(\mathcal{C})$	29496.68	-5360.589	2

Tabella 9.13. Indicatori statistici per i modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$.

`Test`	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
`Clarke`	65849	65835	65169	$\simeq 0$	$\simeq 0$	$\simeq 0$
`Voung`	302	293	252	$\simeq 0$	$\simeq 0$	$\simeq 0$

Tabella 9.14. Risultati dei test di Clarke e Voung per $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ `v.s` $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$.

9.5 Modello a 4 variabili

Per studiare in modo più approfondito il funzionamento della *pair-copula construction* nel perimetro delle variabili categoriali si è ritenuto importante trattare anche un modello che avesse variabili categoriali con un numero di realizzazioni maggiori di 2.

A tal fine è stata aggiunta al modello precedente anche la variabile `area`, avente 6 realizzazioni e definita dal seguente insieme:

```
{A,B,C,D,E,F}
```

Fondamentalmente il modello in esame ha 4 variabili, 2 delle quali verranno trasformate in variabili *dummy* (`gender` e `area`).

Un modello così definito equivale a risolvere un problema avente dimensione 8:

- 2 dimensioni per le variabili continue (`numclaims` e `exposure`);
- 1 afferente al `gender` (che avendo 2 realizzazioni definisce nel perimetro delle variabili *dummy* un problema di dimensione $2-1$);
- 5 livelli dimensionali per la variabile `area` (che avendo 6 realizzazioni definisce nel perimetro delle variabili *dummy* un problema di dimensione $6-1$).

Marginali	Nome Variabile
1	`numclaims`
2	`exposure`
3	`xM`
4	`area.xB`
5	`area.xC`
6	`area.xD`
7	`area.xE`
8	`area.xF`

Tabella 9.15. Tabella delle variabili marginali.

La tabella 9.15 riporta gli indici usati per le marginali nel modello.

Ci si soffermerà solo sul caso con entrambe le variabili categoriali, e si analizzerà il modello parametrico contro quello senza restrizioni sui modelli di copula. Il tempo di elaborazione del modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ senza restrizioni è stato di circa di 13 minuti, mentre sono stati sufficienti 11 minuti per $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$.

Si segnala che, poichè il numero di variabili dei modelli è elevato, si presenterà gli *output* dei due modelli solo del primo albero T_1. In entrambi i casi, la struttura al primo livello risulta essere quella rappresentata in figura 9.4.

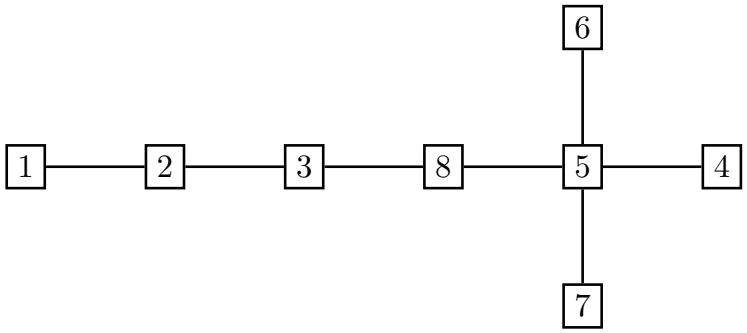


Figura 9.4. T_1 dei modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ a 4 Variabili.

9.5.1 Caso 1: entrambe le variabili come categoriali

Il modello definito dall'algoritmo di Dißmann senza vincoli ovvero l'algoritmo di ottimo locale ha generato l'*output* della tabella tabella 9.16. Le curve di livello corrispondenti sono riportate nella figura 9.5

Albero	Arco	Vincolo	Tipo	Famiglia	Rot.	Par.	df	τ	`log lik`
1	1	$(6,5)$	`d,d`	`bb8`	90	$(7.99, 0.99)$	2	-0.77	3797
1	2	$(1,2)$	`c,c`	`tll`	0	$[30 \times 30]$	195.7	0.04	796
1	3	$(4,5)$	`d,d`	`bb8`	90	$(8.0, 0.99)$	2	-0.77	6121
1	4	$(7,5)$	`d,d`	`bb8`	90	$(8.0, 0.99)$	2	-0.77	2872
1	5	$(5,8)$	`d,d`	`bb8`	270	$(8.0, 0.99)$	2	-0.77	1966
1	6	$(2,3)$	`c,d`	`tll`	0	$[30 \times 30]$	1.2	0.007	27
1	7	$(3,8)$	`d,d`	`joe`	180	1.2	1	0.08	19

Tabella 9.16. *Output* del modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ a 4 Variabili.

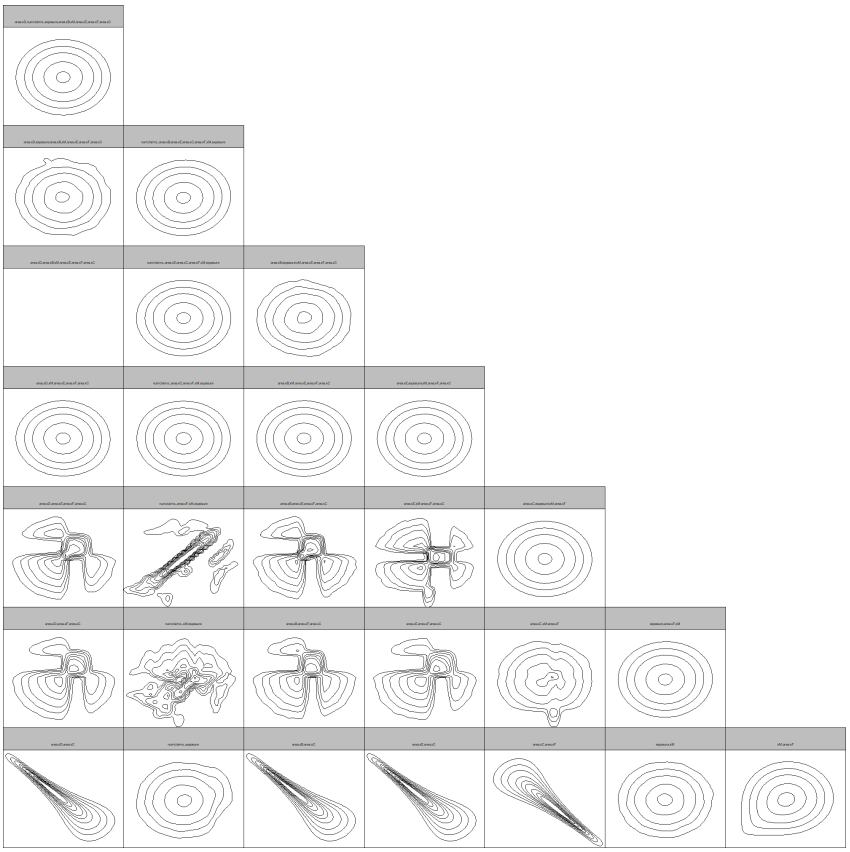


Figura 9.5. Curve di livello del modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ a 4 Variabili.

9.5.2 Caso 2: entrambe le variabili come categoriali e copule parametriche

Il modello parametrico è in tabella 9.17 mentre le curve sono in figura 9.6. Confrontando la tabella 9.17 con la tabella 9.16, afferente al precedente modello, si osserva che al netto dei due archi evidenziati in giallo le rimaneti bi-copule del primo albero sono le stesse nei due modelli.

I valori dei principali indicatori di adattamento sono in tabella 9.18 e i risultati dei test in tabella 9.19. Similmente a quanto avvenuto nel caso in dimensione 3, anche in dimensione 4 il modello non parametrico $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ è decisamente migliore

Albero	Arco	Vincolo	Tipo	Famiglia	Rot.	Par.	df	τ	`log lik`
1	1	$(4,5)$	`d,d`	`bb8`	90	$(8.0, 0.99)$	2	-0.77	6121
1	2	$(7,5)$	`d,d`	`bb8`	90	$(8.0, 0.99)$	2	-0.77	2872
1	3	$(1,2)$	`c,c`	`gaussian`	0	1.17	1	0.1094	544
1	4	$(6,5)$	`d,d`	`bb8`	90	$(7.99, 0.99)$	2	-0.77	3797
1	5	$(5,8)$	`d,d`	`bb8`	270	$(8.0, 0.99)$	2	-0.77	1966
1	6	$(2,3)$	`c,d`	`joe`	0	1	1	0.0102	8
1	7	$(3,8)$	`d,d`	`joe`	180	1.2	1	0.08	19

Tabella 9.17. *Output* del modello $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$ a 4 Variabili.

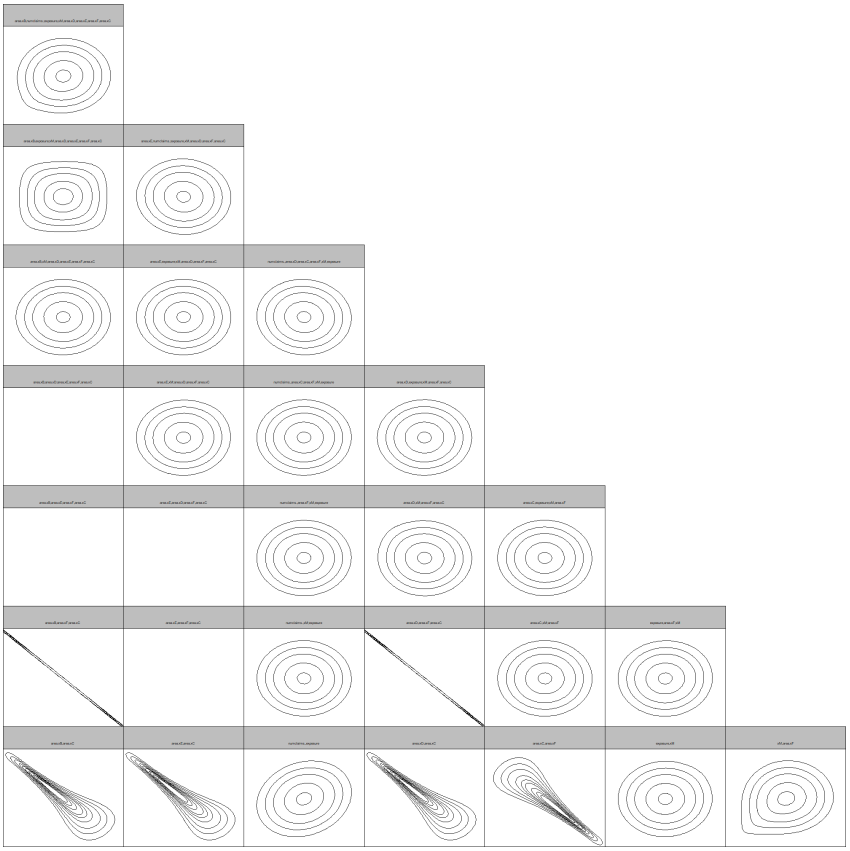


Figura 9.6. Curve di livello del modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ a 4 Variabili.

del modello continuo parametrico $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$. Si osserva infatti, che il modello $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ presenta indicatori statistici `log lik` e `mBIC` migliori ed vince in maniera indiscussa sia rispetto al test di `Clarke` che al test di `Voung`.

Modello	`log lik`	`mBIC`	`nPars`
$\mathcal{M}_{\mathcal{D}}(\mathbb{D})$	53449.61	-384336	486
$\mathcal{M}_{\mathcal{D}}(\mathcal{D})$	-115842.7	-50841.89	29

Tabella 9.18. Indicatori statistici per i modelli $\mathcal{M}_{\mathcal{D}}(\mathbb{C})$ e $\mathcal{M}_{\mathcal{D}}(\mathcal{C})$.

Test	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
Clarke	66874	66863	66837	$\simeq 0$	$\simeq 0$	$\simeq 0$
Voung	682.8691	681.0248	672.6099	$\simeq 0$	$\simeq 0$	$\simeq 0$

Tabella 9.19. Risultati dei test di Clarke e Voung per $\mathcal{M}_{\mathcal{D}}(\mathbb{D})$ `v.s` $\mathcal{M}_{\mathcal{D}}(\mathcal{D})$.

9.6 Conclusioni

Da un'analisi approfondita delle variabili discrete di tipo categoriale con metodologia *pair-copula construction* emerge che l'adozione di modelli non parametrici e dunque di bi-copule `ttl` è senz'altro consigliata. Infatti, tali modelli presentano livelli migliori sia per i principali indicatori di adattamento che nei test di confronto e selezione.

Unitamente si segnala che la presenza dell'approccio di *jittering* per le marginali definisce un perimetro di variabilità dei risultati, diversamente a ciò che accade quando si trattano le variabili continue. In tal senso si consiglia sempre l'implementazione di una pluralità di modelli per aumentare il livello di fiducia rispetto alle risultanze fornite dall'algoritmo di Dißmann nel caso discreto.

Inoltre, si osserva che la necessità di utilizzare modelli non parametrici basati su funzioni *kernel* rende la *pair-copula construction* per il trattamento delle variabili categoriali uno strumento, seppur flessibile, di difficile comprensione. Tuttavia si ripete, come già segnalato, che nel perimetro delle variabili non continue ricercare nella funzione di copula "il vero legame" tra le marginali e la distribuzione congiunta, che è colto invece dalla funzione di copula per le variabili continue, è illegittimo. Infatti la rappresentazione delle bi-copule non coglie la "vera natura" della variabili marginali e di conseguenza qualsiasi valutazione qualitativa o quantitativa della dipendenza sottostante basata su di esse è assolutamente forviante, come già indicato nel capitolo 6.

Capitolo 10

Il Ruolo del *Greedy Algorithm* nella costruzione di un modello interno

> No more sophisticated then is needed in order to reach this objective.
>
> Direttiva Solvency II

10.1 Il Solvency 2 Règime

Il riconoscimento dell'importanza del controllo del rischio nel governo d'impresa nell'ultimo decennio altro non è che il riconoscimento del ruolo primario dell'incertezza nei mercati finanziari. Il Regolatore Europeo per curare il disagio da "incertezza" ha riscritto il linguaggio della solvibilità, tradizionalmente dominato da un atteggiamento contabile e lo ha tradotto con le parole della probabilità.

Di fatto, il *framework* di vigilanza europeo introdotto da Solvency II ha definito un processo di calcolo per l'*SCR* che si basa sull'impianto di valutazioni *market-consistent* adeguate alla *natura*, alla *rilevanza* e alla *complessità* dei rischi dell'impresa, non a caso il Regolamento 2015/35/CE afferma *no more sophisticated then is needed in order to reach this objective* (Commissione Europea (2014)).

> *In base al "principio di proporzionalità" le imprese possono decidere se utilizzare la cosiddetta* formula standard*, eventualmente con i parametri specifici; o il modello interno, completo o parziale. Le soluzioni di calcolo ammissibili si differenziano per come viene trattata l'incertezza. Nella* formula standard *è l'Autorità che definisce il processo di calcolo (fornendo l'intero schema degli algoritmi), "tratta l'incertezza" (formulando le sue "opinioni medie" sul futuro dei* risk driver *e delle loro interrelazioni), e la surroga nel valore di parametri - i "*key parameters*" da dare in* input *agli algoritmi di calcolo dell'SCR. Diversamente col modello interno è l'impresa ad avere tutta la responsabilità sulle tecniche di va-*

lutazione: algoritmiche, distribuzioni di probabilità e correlazioni *(De Felice e Moriconi (2011)).*

Il presente capitolo ha lo scopo di mostrare un caso di studio in cui si determina l'intera struttura di dipendenza tra i rischi tecnici per una compagnia di assicurazione che opera nel mercato *non-life* facendo ricorso all'approccio *pair-copula construction* mediante l'algoritmo di Dißmann. I dati sono provenienti da una compagnia sotto egida di EIOPA, quindi una compagnia che risponde a pieno titolo al Solvency 2 règime e dunque a pieno titolo ha il diritto di sfidare la *standard formula* non solo per la "mera" valutazione del requisito di capitale ma anche ai fini un approccio di validazione a cui le compagnie sono chiamate a rispondere nel continuo. In tal senso si ricorda l'articolo 124 della Direttiva Europea, "Standard di convalida":

- *Le imprese di assicurazione e di riassicurazione prevedono un ciclo regolare di convalida del loro modello interno, che include il monitoraggio del suo buon funzionamento, il riesame della continua adeguatezza delle sue specifiche e il raffronto delle sue risultanze con i dati tratti dall'esperienza.*
- *La procedura di convalida del modello interno include un processo statistico efficace che consenta alle imprese di assicurazione e di riassicurazione di dimostrare alle loro autorità di vigilanza che i requisiti patrimoniali che risultano da tale modello sono appropriati.*
- *I metodi statistici utilizzati verificano l'appropriatezza della distribuzione di probabilità prevista non soltanto rispetto all'esperienza passata, ma anche rispetto a tutti i nuovi dati rilevanti e alle nuove informazioni relativi ad essa.*
- *La procedura di convalida del modello interno include un'analisi della sua stabilità ed in particolare la verifica della sensibilità delle sue risultanze a variazioni delle principali ipotesi sottostanti. Essa include altresì la valutazione dell'accuratezza, della completezza e dell'adeguatezza dei dati utilizzati nel modello interno.*

Dunque è fondamentale per l'intero *management* d'impresa e in particolare le funzioni attuariali della compagnia, che esse lavorino *on-going* per aumentare *il grado di fiducia* riguardo i risultati, sia nella logica della *standard formula* che di un modello interno avendo sempre riguardo che i requisiti patrimoniali risultanti siano appropriati.

Nella seguente trattazione si farà uso dell'algoritmo di Dißmann per implementere l'intera struttura di dipendenza della compagnia definendo la distribuzione di probabilità congiunta di dimensione 11, il risultato di tale processo porterà a una struttura multivariata costituita da 55 bi-copule.

Successivamente, attraverso l'approccio simulativo "si sfiderà" l'algoritmo di ottimo locale, il "*greedy algorithm*" e verrà analizzata la bontà del modello.

Si osserva che raramente le compagnie gestiscono un solo ramo di attività, dunque sarebbe sempre necessario considerare la dipendenza tra i diversi rischi anche "solo" per finalità di quantificazione e controllo del *reserve-risk*.

Pertanto, al di là del rilascio "altisonante" di potenziale modello interno che "sfida" e "convalida" la struttura di dipendenza tra i rischi tecnici sarebbe sempre indicato considerare la possibile dipendenza tra distinti rami di attività anche per la modellizzazione delle riserve: in altre parole, i modelli di riserva stocastica andrebbero sempre rivisti all'interno di quadro multivariato.

A tal proposito si citano alcuni studi nella letteratura di stampo attuariale che analizzano l'estensione del *chain-ladder* nel caso multivariato, tra i principali si ricorda Schmidt (2006), Merz e Wüthrich (2008), Shi e Frees (2011) e Shi e Yang (2018). In genere in questi elaborati si trova la distribuzione normale bivariata per modellare al più due linee di *business* in maniera simultanea.

La novità di questi articoli è l'aver coniugato l'uso delle copule al "problema" delle di riserve, in quanto esse consentono di considerare la distribuzione marginale e la struttura di dipendenza in modo separato.

Si segnala che in Ricotta (2019) viene implementata una struttura di dipendenza per 4 *linee di business*, quindi una struttura multivariata di dimensione 4. L'approccio non facendo uso di algoritmi automatici prevede l'implementazione di tutte le configurazioni in dimensione 4 ovvero 24. In particolare vengono calcolate tutte le *R-vine* in dimensione 4 che sono rispettivamente 12 *D-vine* e 12 *C-vine*. Si sottolinea che un'approccio di questo tipo, ovvero che non ricorre ad algoritmi automatici nel perimetro dell'*high-dimensional applications* non sarebbe stato percorribile.

In questa trattazione la *pair-copula construction* verrà utilizzata per definire la dipendenza esistente tra tutti i rami di attività del comparto danni.

Questo è un tema che si sottolinea fin dall'introduzione, la *pair-copula construction* in alte dimensioni acquista un senso solo in un *framework* operativo che considera non solo la teoria dei grafi ma prende a piene mani dalle soluzioni *software*[1] per l'ottimizzazione e nella disponibilità di elaboratori che presentano una potenza di calcolo sufficiente[2].

In dimensione 11 si hanno circa 1.3×10^{18} *R-vine*[3]; ognuna di esse può essere la struttura alla base di un modello inferenziale. Siccome `rvinecopulib` permette di associare ognuno dei $\binom{11}{2} = 55$ archi della *R-vine* con 12 famiglie differenti di bi-copule, il numero di effettivi modelli generabili dalla libreria si avvicina a 2.9×10^{77}.

Con questi numeri, la modellazione di un modello *PCC* è un problema di difficile risoluzione anche con l'alta tecnologia. Inoltre va considerato che trovare il miglior modello nell'universo dei modelli implementabili è un problema che si dipana su due livelli: il primo relativo ai limiti computazionali, il secondo si riferisce ai criteri di selezione di ottimo globale. Entrambe le problematiche sono state analizzate in dettaglio nel capitolo 5, ma è utile riprenderli in questa sede.

Il problema dell'ottimo globale, ovvero l'identificazione del modello migliore, è un problema di tipo puramente teorico: in campo inferenziale non si conoscono infatti metodologie di confronto di modelli che siano esenti di difetti e che permettano di identificare in maniera deterministica ed oggettiva il modello migliore tra due modelli concorrenti. Un valore più alto in un indicatore non sempre è una garanzia di

[1] Si ricorrerà a Dißmann ma in generale non importa quale sia l'algoritmo *greedy* tra i tanti.

[2] Essa si rivelerà una vera e propria frontiera operativa.

[3] Il numero si riduce di molto se si usano *C-vine* o *D-vine*, sono circa 20 milioni per le *vine* di tipo *drawable* e 20 per le *vine* di tipo *canonical*.

migliore qualità e i test statistici possono sia essere inconcludenti che discordanti tra di loro. Questa indeterminazione si traduce nell'impossibilità di costruire un ordine definitivo tra i modelli. Quindi, a meno di non conoscere la distribuzione che ha generato i dati, non è possibile identificare con certezza e precisione il modello *PCC* che si avvicina di più alla distribuzione obiettivo. Come evidenziato nel capitolo 5, in letteratura si è cercato di ovviare a questa limitazione attraverso la comparazione di funzioni obiettivo differenti in simulazioni controllate ed applicazioni reali. I *software* considerati in questa tesi permettono di selezionare criteri di scelta differenti.

Il problema computazionale è invece collegato al fatto che gli algoritmi conosciuti per la ricerca dell'ottimo tra le *R-vine* sono o approssimati o inefficienti. In particolare, come è stato osservato nella sezione 5.6, anche nel contesto più semplice, ovvero facendo uso di *vine* di tipo *drawable*, ossia le *D-vine* la scelta dell'ottimo è un NP-*hard problem*. In teoria della complessità, i problemi NP-*hard* (ovvero *non-determistic polynomial-time hard problem*, "problema non deterministico difficile in un tempo polinomiale") sono una classe di problemi che può essere definita informalmente come la classe dei problemi almeno difficili come i più difficili problemi delle classi di complessità P e NP. Il problema delle classi P e NP sono un problema tuttora aperto nella teoria della complessità computazionale.[4]

Queste limitazioni non vanno però confuse come limitazioni della metodologia PCC: il problema della scelta del modello è infatti un problema di statistica inferenziale. La ragione per cui affrontiamo questa problematica per le *PCC* è legata al fatto che in altre metodologie la selezione del modello avviene manualmente e a posteriori.

La matrice di correlazione attualmente indicata dalla Direttiva per aggregare il rischio di tariffazione e riservazione delle differenti linee di attività è riportato in tabella 10.1.

	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2	0.50	1										
3	0.50	0.25	1									
4	0.25	0.25	0.25	1								
5	0.50	0.25	0.25	0.25	1							
6	0.25	0.25	0.25	0.25	0.50	1						
7	0.50	0.50	0.25	0.25	0.50	0.50	1					
8	0.25	0.50	0.50	0.50	0.25	0.25	0.25	1				
9	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	1			
10	0.25	0.25	0.25	0.25	0.50	0.50	0.50	0.25	0.25	1		
11	0.25	0.25	0.50	0.50	0.25	0.25	0.25	0.25	0.50	0.25	1	
12	0.25	0.25	0.25	0.50	0.25	0.25	0.25	0.50	0.25	0.25	0.25	1

Tabella 10.1. Matrice di Correlazione tra le *LoB* indicata nella Direttiva.

[4]Si segnala che NP-*hard problem* rientra tra i problemi del millennio alla pari dell'ipotesi di Riemann.

Il numero delle correlazioni nella matrice è 78, come suggerisce la formula di Gauss per i numeri triangolari $\frac{n(n+1)}{2} = \frac{12(13)}{2} = 78$.

10.2 Il Modello

Verranno utilizzati come *dataset* i triangoli di *run-off* degli importi per sinistri pagati aventi tutti la stessa profondità storica di 15 anni nella finestra temporale $2008 - 2022$. Le linee di *business* considerate sono 11 e sono tutte quelle che nel comparto danni possono essere esercitate in Italia, poichè la *LoB* 3 ovvero *Workers' compensation insurance* di fatto non esiste nei confini nazionali, in quanto è l'INAIL[5] ad avere riguardo dell'assicurazione obbligatoria contro gli infortuni sul lavoro e le malattie professionali. Dunque la struttura di dipendenza che verrà implementata coinvolgerà le seguenti *LoB* elencate in tabella 10.2.

Numero *LoB*	Attività
LoB 1	Medical Expense
LoB 2	Income Protection
LoB 4	Motor Vehicle
LoB 5	Other Motor
LoB 6	Marine
LoB 7	Fire
LoB 8	General Liability
LoB 9	Credit
LoB 10	Legal Expenses
LoB 11	Assistance
LoB 12	Miscellaneous

Tabella 10.2. Elenco delle *LoB* trattate dalle compagnie italiane.

La stima della riserva sinistri è stata ottenuta in modo analogo per tutte le attività tramite la classica metodologia del *chain-ladder* applicando un approccio di tipo *bootstrapping*. Si ricorda che in generale il *bootstrapping* consente di superare le problematiche connesse alla limitata numerosità dei dati ed attraverso ad essa si riesce a costruire quella che nel gergo attuariale è detta variabilità "artificiale". In questo caso si osserva che il *il bootstrapping* gioca un doppio ruolo, costruire la variabilità fittizia legata al calcolo di una valutazione della riserva con approccio stocastico e congiuntamente "coglie" l'eventuale rapporto di dipendenza tra le diverse linee di attività.

Si osserva che la letteratura attuariale è stata generosa anche a tal riguardo, infatti nel corso degli anni si è ricorsi a questa tecnica per diverse questioni come ad esempio, la stima dell'errore di previsione sulla riserva sinistri, la costruzione per esso di un limite superiore e dei relativi intervalli di confidenza. Si segnala che in genere l'applicazione della tecnica *bootstrap* non è sempre immediata, la *issue* principale riguarda la definizione dei residui da considerare nella procedura (Pinheiro, Andrade e Silva e Centeno (2003)). Nel seguente caso si utilizzerà una

[5]Istituto Nazionale Assicurazione Infortuni sul Lavoro

forma "speciale" rispetto al classico *bootstraping* detta *bootstrap dipendente* che ci consentirà appunto di estrarre l'eventuale dipendenza tra le diverse linee di attività.

Essa di fatto, è implicitamente contenuta nei triangoli *run-off* del pagato delle diverse *LoB* della compagnia e verrà estratta ricorrendo ad un'approccio *bootstrap dipendente* in modo analogo a come ha fatto Ricotta (2019).

Si osserva che nell'approccio *bootstrap dipendente* il ricampionamento dei residui avviene effettuato congiuntamente per ogni *LoB* del modello in esame[6]. Ad ogni passo del ricampionamento viene eseguito il *bootstrapping* di tutte le celle aventi la stessa posizione nei rispettivi triangoli *run-off* delle 11 linee di *business* del modello, detto appunto *bootstrap cell-by-cell*. Il ricampionamento ha considerato 10.000 iterazioni. Si trova in Taylor e McGuire (2007) il procedimento a questa tecnica di *bootstrapping*.

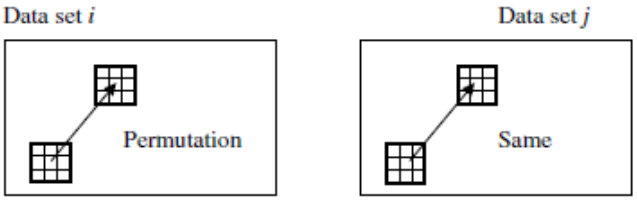


Figura 10.1. Ricampionamento *cell-by-cell* del *bootstraping dipendente*.

Alla fine della procedura di *bootstraping dipendente* si ottiene una matrice di 10.000×11 ovvero 10.000 righe e 11 colonne dove la cella (i, j) rappresenta di fatto l'i-esima stima *bootstrapping* con approccio dipendente per la j-esima linea di attività.

10.3 La scelta del "*Greedy Algorithm*"

Da una prima analisi esplorativa dell'*output* (figura 10.2) ottenuto attraverso l'approccio di *bootstrapping dipendente* emerge che al netto della correlazione positiva tra la *LoB* 4 (*Motor Vehicle*) con la *LoB* 8 (*General Liability*) i livelli delle correlazioni tra le altre linee di attività sono dell'ordine di 10^{-2}. Dunque al netto delle citate *LoB* si può affermare che i *pattern* del pagato non presenta dipendenza di tipo lineare. Come sempre è stato affermato: correlazione nulla non implica indipendenza ed è proprio su questo che si vuole indagare.

Si osserva inoltre che la correlazione positiva tra la *LoB* 4 con la *LoB* 8 è pari a 0.36. Questo è un fenomeno che ci sorprende per la natura delle *LoB* in esame, in quanto il *business* di tali attività non si "intersecano". In particolare si osserva che nella *LoB* 4 confluisce tutto quello che è il *motor* ovvero l'R.c.A (quindi Ramo 10) e quello che in senso *local* è chiamato "Altri danni ai beni dei veicoli " (Ramo 13).

Si osserva che la polizza R.c.A in senso Solvency II va intesa come tutto ciò che riguarda l'autovettura (Ramo 9 e Ramo 10) quindi il rischio rappresentato dalla

[6]Operativamente questo è stato possibile imponendo lo stesso seme nell'implementazione della simulazione

responsabilità civile del autoveicolo congiutamente a ogni tipo di danno subito dal veicolo (quindi la grandine che colpisce l'autovettura, piuttosto che il furto dell'autovettura, piuttosto che l'incendio dell'autovettura) mentre R.C. generale (Ramo 13) riguarda la responsabilità civile nei confronti dei terzi ovvero ogni responsabilità diversa da quelle menzionate nel Rami 10 (Responsabilità civile autoveicoli terrestri), 11 (R.C. aeromobili) e 12 (R.C. natanti, compresa la responsabilità del vettore).

Potrebbe esserci un grado di dipendenza nella gestione delle due linee di attività, essendo le due *LoB* di matrice R.C., quindi le due linee di attività potrebbero essere entrambe di tipo *long-tail* dal punto di vista della gestione dei sinistri. Infatti nel *pattern* del pagamento tra il sinistro R.c.A e relativo processo di liquidazione, difficilmente ci può essere un legame anche presunto con il rimborso relativo ad un intervento chirugico eseguito in modo inadeguato.

Quello che si può supporre è che essendo i due rami caratterizzati da un approccio *long tail* in termini di gestione, effettivamente sono attività aventi le maggiori "questioni aperte", la velocità con cui si smonta il triangolo del R.c.A e la velocità con cui si smonta il triangolo dell'R.c.G potrebbe essere "simile" in questo senso. Diversamente accade per gli altri Rami, ad esempio il Ramo Malattia in cui al terzo anno generalmente il più del 60% viene liquidato sia in termini di numeri che di importi non può evidenziare un fenomeno di questo tipo nè con *LoB* 4 e nè con la *LoB* 8.

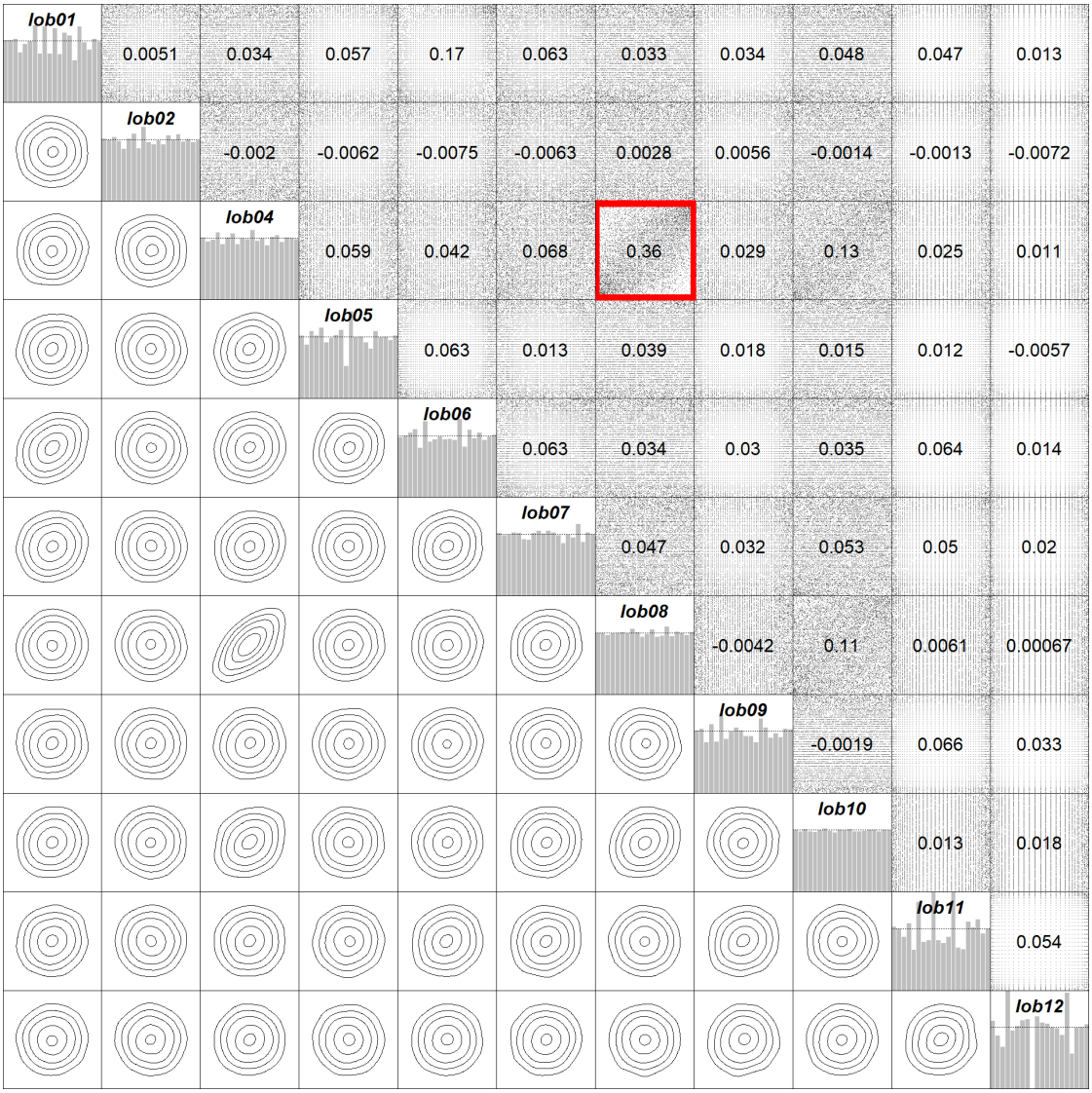


Figura 10.2. Curve empiriche e correlazioni tra le *lines of business*.

Dunque, a seguito dell'implementazione del modello che utilizza l'algoritmo *greedy* di Dißmann si ottiene l'*output* dei risultati che può essere osservato nella tabella 10.4.

Unitamente si osserva il *pattern* definito da Dißmann al primo albero (figura 10.3). Si segnala che per non appesantire la trattazione si evita di inserire in questo elaborato i rimanenti 9 alberi. Si sottolinea che si può vedere nell'*output* finale *step-by-step* l'intera costruzione *building-blocks* della distribuzione multivariata risultante nella ricerca dell'ottimo locale, ovvero ottenuta albero per albero.

È stato ottenuto un modello a 61 parametri; i livelli dei principali indicatori sono riportati in tabella 10.3. Il modello $\mathcal{M}_{\mathcal{D}}$ implementato attraverso l'algoritmo

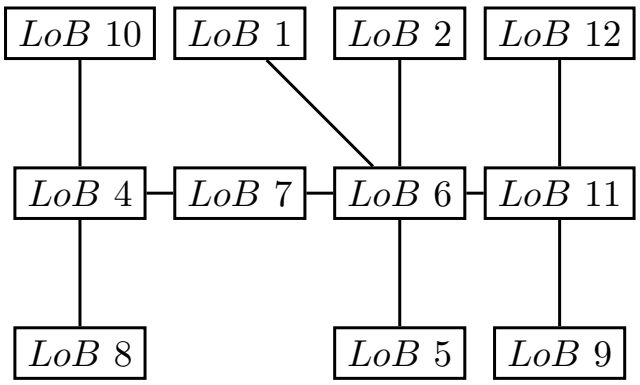


Figura 10.3. Albero T_1 del modello creato con Dißmann.

di Dißmann è formato da 55 copule bivariate di cui:

- 13 presentano struttura indipendente;
- 11 sono delle copule T-Student;
- 31 sono copule parametriche.

`log lik`	Numero Parametri	AIC	BIC	mBIC
2853.18	61	−5584.36	−5144.53	−5012.507

Tabella 10.3. I principali indicatori statistici relativi al modello $\mathcal{M}_{\mathcal{D}}$.

Vedendo l'*output* ottenuto, riportato in tabella 10.4, acquista un senso lo zelo e la solerzia con cui sono state descritte le diverse famiglie nel capitolo 3.

Da una prima analisi esplorativa delle curve di livello emerge che esse presentano una forma abbastanza regolare, si può vedere che anche attraverso l'approccio *pair-copula construction* le due linee di attività che presentavano correlazione positiva non trascurabile, ovvero la *Lob* 4 con la *Lob* 8, mostrano ancora una forma di dipendenza funzionale, tuttavia questa appare non di tipo lineare poichè anche in una prima analisi le curve di livello assumono una forma più schiacciata come si osseva nella figura 10.4. Inoltre da un'analisi dell'*output* risultante emerge che la copula che lega queste due linee di attività è la copula T, quindi emerge fin da subito che il modello implementato, più raffinato e frutto di impianto metodologico imponente, sebbene sacrifichi la semplicità, risulta di fatto più efficace nella rappresentazione dei legami di dipendenza per i rischi congiunti. Si ricorda infatti che al netto di una correlazione tra la Lob 4 e la Lob 8, tutte le altre correlazioni figuravano nell'ordine di 10^{-2}.

Albero	Arco	Vincolo	Famiglia	Rot.	Par.	df	τ
1	1	$(5,6)$	bb1	0	$(0.079, 1.021)$	2	0.058
1	2	$(12,11)$	t	0	$(0.078, 21.089)$	2	0.049
1	3	$(2,6)$	indep	0		0	0
1	4	$(9,11)$	t	0	$(0.094, 23.425)$	2	0.060
1	5	$(11,6)$	t	0	$(0.092, 32.057)$	2	0.059
1	6	$(8,4)$	t	0	$(0.530, 10.100)$	2	0.356
1	7	$(10,4)$	t	0	$(0.205, 12.335)$	2	0.132
1	8	$(1,6)$	t	0	$(0.247, 25.466)$	2	0.159
1	9	$(4,7)$	bb1	180	$(0.048, 1.0440)$	2	0.064
1	10	$(6,7)$	t	0	$(0.093, 23.122)$	2	0.059
2	1	$(5,1\|6)$	bb8	0	$(1.176, 0.834)$	2	0.043
2	2	$(12,9\|11)$	bb8	180	$(1.121, 0.828)$	2	0.029
2	3	$(2,1\|6)$	gumbel	180	1.008	1	0.008
2	4	$(9,6\|11)$	gaussian	0	0.038	1	0.024
2	5	$(11,7\|6)$	t	0	$(0.068, 28.165)$	2	0.043
2	6	$(8,10\|4)$	frank	0	0.441	1	0.049
2	7	$(10,7\|4)$	gumbel	180	1.040	1	0.038
2	8	$(1,7\|6)$	gaussian	0	0.075	1	0.048
2	9	$(4,6\|7)$	t	0	$(0.053, 49.999)$	2	0.034
3	1	$(5,2\|1,6)$	clayton	90	0.015	1	0
3	2	$(12,6\|9,11)$	gumbel	180	1.009	1	0.009
3	3	$(2,7\|1,6)$	indep	0		0	0
3	4	$(9,7\|6,11)$	t	0	$(0.0369, 49.999)$	2	0.023
3	5	$(11,1\|7,6)$	frank	0	0.266	1	0.029
3	6	$(8,7\|10,4)$	clayton	180	0.022	1	0.011
3	7	$(10,6\|7,4)$	bb1	180	$(0.031, 1.005)$	2	0.020
3	8	$(1,4\|7,6)$	bb8	180	$(1.048, 0.960)$	2	0.020
4	1	$(5,7\|2,1,6)$	indep	0		0	0
4	2	$(12,7\|6,9,11)$	frank	0	0.131	1	0.014
4	3	$(2,4\|7,1,6)$	indep	0		0	0
4	4	$(9,1\|7,6,11)$	frank	0	0.203	1	0.022
4	5	$(11,4\|1,7,6)$	frank	0	0.152	1	0.016
4	6	$(8,6\|7,10,4)$	clayton	180	0.019	1	0.009
4	7	$(10,1\|6,7,4)$	bb8	180	$(1.192, 0.711)$	2	0.032
5	1	$(5,4\|7,2,1,6)$	t	0	$(0.084, 49.654)$	2	0.053
5	2	$(12,1\|7,6,9,11)$	indep	0		0	0
5	3	$(2,10\|4,7,1,6)$	indep	0		0	0
5	4	$(9,4\|1,7,6,11)$	frank	0	0.208	1	0.023
5	5	$(11,10\|4,1,7,6)$	indep	0		0	0
5	6	$(8,1\|6,7,10,4)$	gumbel	180	1.009	1	0.009
6	1	$(5,10\|4,7,2,1,6)$	indep	0		0	0
6	2	$(12,4\|1,7,6,9,11)$	joe	180	1.012	1	0.007
6	3	$(2,8\|10,4,7,1,6)$	indep	0		0	0
6	4	$(9,10\|4,1,7,6,11)$	frank	0	-0.113	1	-0.01
6	5	$(11,8\|10,4,1,7,6)$	clayton	270	0.0206	1	-0.01
7	1	$(5,8\|10,4,7,2,1,6)$	frank	0	0.098	1	0.010
7	2	$(12,10\|4,1,7,6,9,11)$	frank	0	0.134	1	0.014
7	3	$(2,11\|8,10,4,7,1,6)$	indep	0		0	0
7	4	$(9,8\|10,4,1,7,6,11)$	clayton	90	0.044	1	-0.02
8	1	$(5,11\|8,10,4,7,2,1,6)$	indep	0		0	0
8	2	$(12,8\|10,4,1,7,6,9,11)$	joe	270	1.009	1	0
8	3	$(2,9\|11,8,10,4,7,1,6)$	indep	0		0	0
9	1	$(5,9\|11,8,10,4,7,2,1,6)$	bb7	0	$(1.006, 0.018)$	2	0.012
9	2	$(12,2\|8,10,4,1,7,6,9,11)$	indep	0		0	0
10	1	$(5,12\|9,11,8,10,4,7,2,1,6)$	gumbel	90	1.010	1	-0.01

Tabella 10.4. $\mathcal{M}_{\mathcal{D}}$ generato dall'algoritmo di Dißmann.

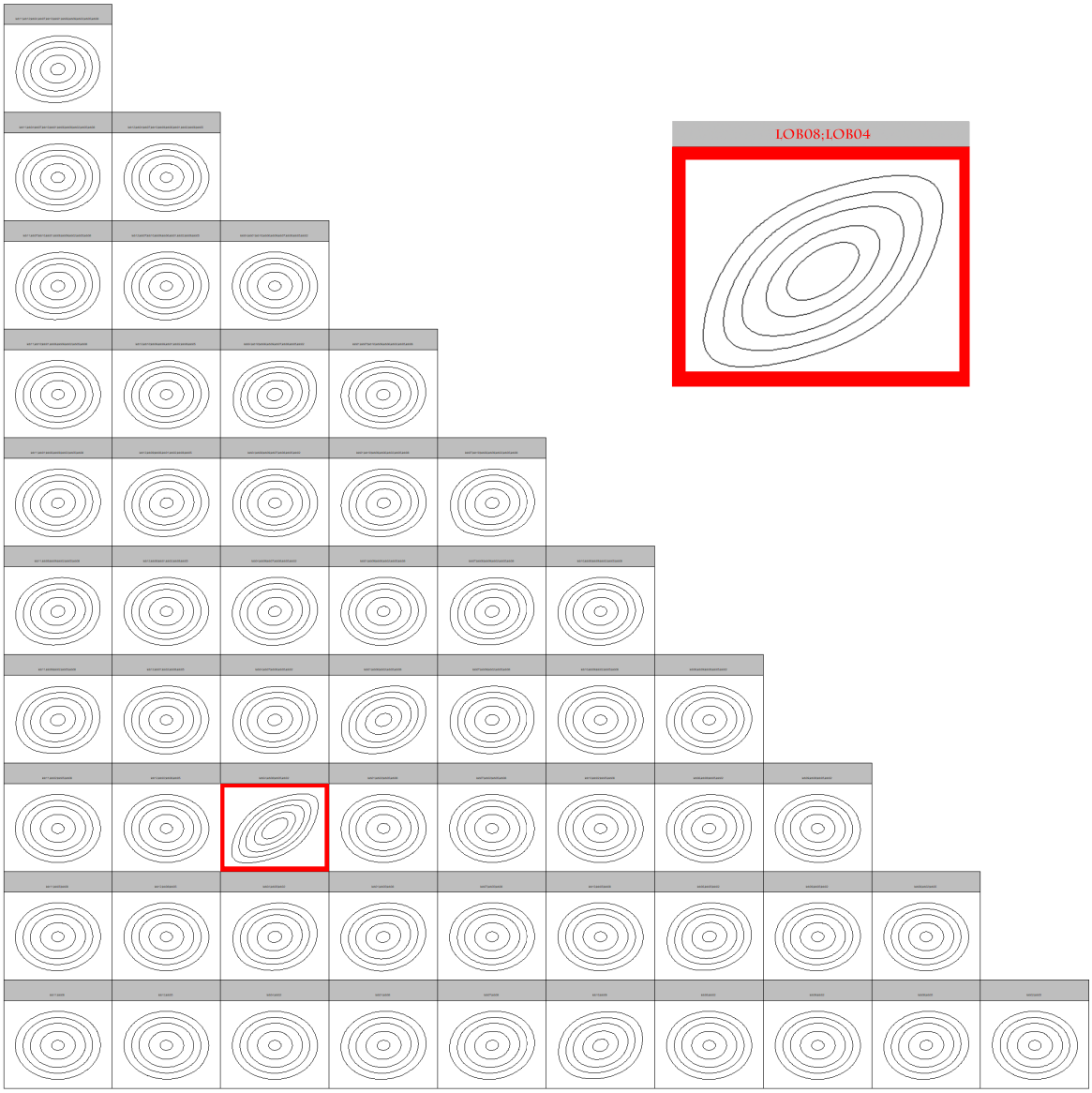


Figura 10.4. Curve di livello del Modello $\mathcal{M}_{\mathcal{D}}$.

Si osserva che per quanto riguarda la possibilità di fare un'analisi completa dell'applicabilità dell'ipotesi semplificate in dimensioni elevate è operativamente e computazionalmente molto costoso, in quanto andrebbe fatto per ogni bi-copula del *pattern*, in questo caso il modello è formato da 55 bi-copule.

Si segnala che nei lavori della Czado e di Nagler, che sono i maggiori esperti nell'alveo dell'utilizzo della *pair-copula construction* in contesto multivariato, affermano entrambi che indipendentemente dall'ipotesi semplificante, la *PCC* in dimensioni elevate offre comunque una migliore rappresentazione rispetto all'approccio multi-normale, altresì l'impiego a posteriori di una buona batteria di indicatori per la verifica dell'adattamento si rileva uno strumento efficace per verifica degli aspetti di adattamento del modello. In particolare si ricorda la definizione dei un indicatore

ad-hoc come `mBIC`. Si osserva che il modello $\mathcal{M}_{\mathcal{D}}$ non è il risultato di un algoritmo di ottimo globale ma bensì locale dunque non si esclude la possibilità che vi siano delle soluzioni migliori rispetto a quella offerta dall'algoritmo di Dißmann. In questo caso si può analizzare il problema seguendo una metodologia "mista" ovvero quella utilizzata nel capitolo 8, quindi un approccio di tipo *mixed-tailored* cioè ottenuto troncando il *pattern* su dimensioni più piccole di 11 e poi lavorare aggiungendo le ulteriori variabili *step-by-step* facendo uso dei principali criteri per valutare la bontà di adattamento del modello ottenuto. Osserviamo che l'approccio misto in dimensioni basse come si è visto nel capitolo 8 può offrire buoni margini di miglioramento, tuttavia in grandi dimensioni i tempi di una siffatta analisi non sono percorribili. Si vedranno nei prossimi paragrafi che verrà utilizzato un'approccio simulativo poichè in dimensione 11, nemmeno l'alta tecnologia può generare in tempi ragionevoli tutte le possibili traiettorie e scegliere la migliore tra esse.

10.4 10^{18} Possibili Scelte

Prima di parlare dei problemi di *model selection* che emergono lavorando con 11 variabili, si rievocano le parole di Nagler, Bumann e Czado (2019) in cui ringraziavano il Leibniz Rechenzentrum per aver utilizzato il super-computer posseduto dal centro di ricerca:

> *Authors gratefully acknowledge the compute and data resources provided by the Leibniz Supercomputing Centre (www.lrz.de).*

Ed è fondamentale confrontarsi con questo ringraziamento quando si intraprende la strada della *pair-copula construction* in grandi dimensioni. Si può capire il perchè di questa affermazione, se si riporta all'attenzione la riga per la dimensione 11 della "solita" tabella 4.1:

Dimensione	Numero di *R-vine*	Numero di *C-vine* o *D-vine*
11	1371530804487782400	19958400

Tabella 10.5. Riga per la dimensione 11 della tabella 4.1.

Un modello *PCC* in dimensione 11 dispone di circa 1.3×10^{18} modelli selezionabili mentre esistono rispettivamente circa 20 milioni di *C-vine* e circa 20 milioni di *D-vine*. Nel caso del modello ($\mathcal{M}_{\mathcal{D}}$) utilizzando il *greedy algorithm* visto nel paragrafo precedente, l'elaboratore utilizzato possedeva 14 `core` fisici e 20 `core` logici ha richiesto circa 60 secondi, un minuto, un tempo ragionevole minore persino del calcolo del *bootstrap* per le singole *Lob*. Si calcoli ora, quanto tempo occorre allo stesso elaboratore per calcolare 39916800 di traiettorie, ovvero la somma di *C-vine* e *D-vine*: 39916800 minuti equivalgono a 665280 ore, 665280 ore equivalgono a 27720 giorni, 27720 giorni equivalgono a 75.9 anni. Ci vorrebbero quindi circa 75 anni per lanciare tutte le traiettorie in termini di *D-vine* e *C-vine*. Per quando riguarda le *R-vine* il problema non verrà considerato per ovvie ragioni di tipo computazionali[7].

[7]Si afferma sin dall'introduzione che parlare di *pair-copula construction* nel contesto in *high-dimensional* non ha senso se non si coniuga con l'alta tecnologia.

Dopo aver fatto questo conto della serva si capisce che simulare tutti i modelli e scegliere il “migliore” (migliore poi rispetto a quele criterio?) non è una strada percorribile (almeno con le risorse computazionali attuali).

Unitamente a questo si comprende perchè la Czado è riuscita a tracciare delle soglie di validità in termini di `log lik` per il *greedy algorithm* solo “fino” in dimensione $n = 7$.

E infine, si capisce perchè per i modelli *vine-copula* è stato costruito (direi “non a caso”) un nuovo indicatore *ad-hoc* di adattamento che fosse *tailored* per i problemi di *model-selection* in grandi dimensioni nella *pair copula construction* ovvero l'`mBIC`.

10.5 “Dißmann Contro Mille”

Le ultime soluzioni *software* che sono state recentemente rilasciate ci consentono non solo di implementare Dißmann ma anche di simulare le *R-vine*. Ad ogni traiettoria è possibile quindi associare un modello e per ognuno di essi è possibile confrontarlo con quello ottenuto nel paragrafo precedente calcolato attraverso l'algoritmo *greedy* che è stato indicato con la notazione $\mathcal{M}_{\mathcal{D}}$. Le simulazioni sono state ottenute attraverso la funzione `rvine_structure_sim` della libreria `rvinecopulib`.

Si segnala che per effettuare una simulazione di 1000 *R-vine* l'elaboratore di cui si disponeva ha impiegato circa un giorno.

In tal senso l'obiettivo di questo paragrafo è comprendere la qualità della soluzione offerta dal *greedy algorithm* di Dißmann.

Si osserva che in questo *case-study* effettivamente si riescono a cogliere tutte le problematiche di *model-selection* esposte nel corso di questo elaborato e in particolare il problema dell'enumerazione lungamente segnalato sin dall'inizio della trattazione.

In particolare, congiuntamente alla simulazione è stato intrapresa un'analisi di tipo *multi-criteria decision* ovvero selezionare uno o più criteri, tra la pluralità degli indicatori a disposizione per verificare la bontà di adattamento e decidere se vi fosse almeno una tra 1000 traiettorie implementate che avesse degli aspetti tali che la potessero farla ritenere migliore rispetto al *pattern* ottenuto mediante Dißmann. Come è stato osservato nell'alveo di tutta la tesi, nel perimetro della *pair-copula construction* esiste una batteria molto ricca per valutare la bontà di adattamento di un modello, in particolare un modello potrebbe essere migliore rispetto ad un altro rispetto ai seguenti criteri:

- `log lik`
- numero di parametri;
- `AIC`;
- `BIC`;
- `mBIC`;
- `Voung-Test`;
- `Clarke-Test`.

Che in buona sostanza sono le stesse riflessioni fatte dai diversi i padri del *greedy-algorithm*, primo tra tutti Dißmann e poi la Czado:

> *Clearly, this algorithm only makes a locally optimal selection in each step, since the impact on previous and subsequent trees is ignored. The strategy is however reasonable with regard to statistical modeling with regular vine copulas. Important dependencies as measured by the weights are modeled first. The maximum spanning tree in lines 2 and 10 can be found, e.g., using the classical algorithms by Prim or Kruskal (see for example Cormen et al. (2009)) Later Gruber and Czado (2015) have shown in six-dimensional simulation scenarios, that the greedy Dißmann algorithm gets about at least 75% of the true log-likelihood, which indicates a good performance. Possible choices for the weight ω are, for example:*
>
> - *the absolute empirical Kendall's τ as proposed by Dißmann et al. (2013), Czado et al. (2012);*
> - *the AIC of each pair copula;*
> - *the (negative) estimated degrees of freedom of Student's t pair copulas as proposed by Mendes et al. (2010);*
> - *the p-value of a copula goodness of fit test and variants as proposed by Czado et al. (2013).*

Considerato che il `Voung-Test` e il `Clark-Test` sono ritenuti in letteratura i test fondamentali per scegliere tra due modelli in termini di bontà di adattamento per le *R-vine* sono stati utilizzati proprio questi come metrica di selezione.

Si segnala che la funzioni per questi test ancora non sono disponibili in `rvinecopulib` e quindi sono state implementate le funzioni in `R` sia nel caso del `Voung-Test` che il `Clarke-Test`. Si osserva che i risultati di Voung e Clarke non è detto si allineino sempre, in quanto uno *preferisce* un modello anzichè un altro in termini assoluti mentre l'altro lavora in termini di conteggi positivi.

In particolare, ai fini di questa trattazione, verrà considerato un modello migliore rispetto a $\mathcal{M}_{\mathcal{D}}$ e verrà indicato con $\mathcal{M}_{\mathcal{V}}$[8] se esso è migliore sia in termini di `Test di Voung` che per il `Test di Clarke`, e nel caso ci siano più modelli nell'approccio simulativo che presentano tali caratteristiche verrà considerato quello che presenta il livello di `log lik` più elevato.

Secondo quanto sopra descritto è stata implementata la simulazione generando 1000 *pattern* distinti e il *best-fit* ottenuto secondo il criterio di selezione appena presentato è riportato nella tabella 10.6.

Il modello $\mathcal{M}_{\mathcal{V}}$ implementato stavolta con approccio simulativo è formato da 66 copule bivariate di cui:

- 11 presentano struttura indipendente;
- 9 sono delle copule T-Student;
- 46 sono copule parametriche.

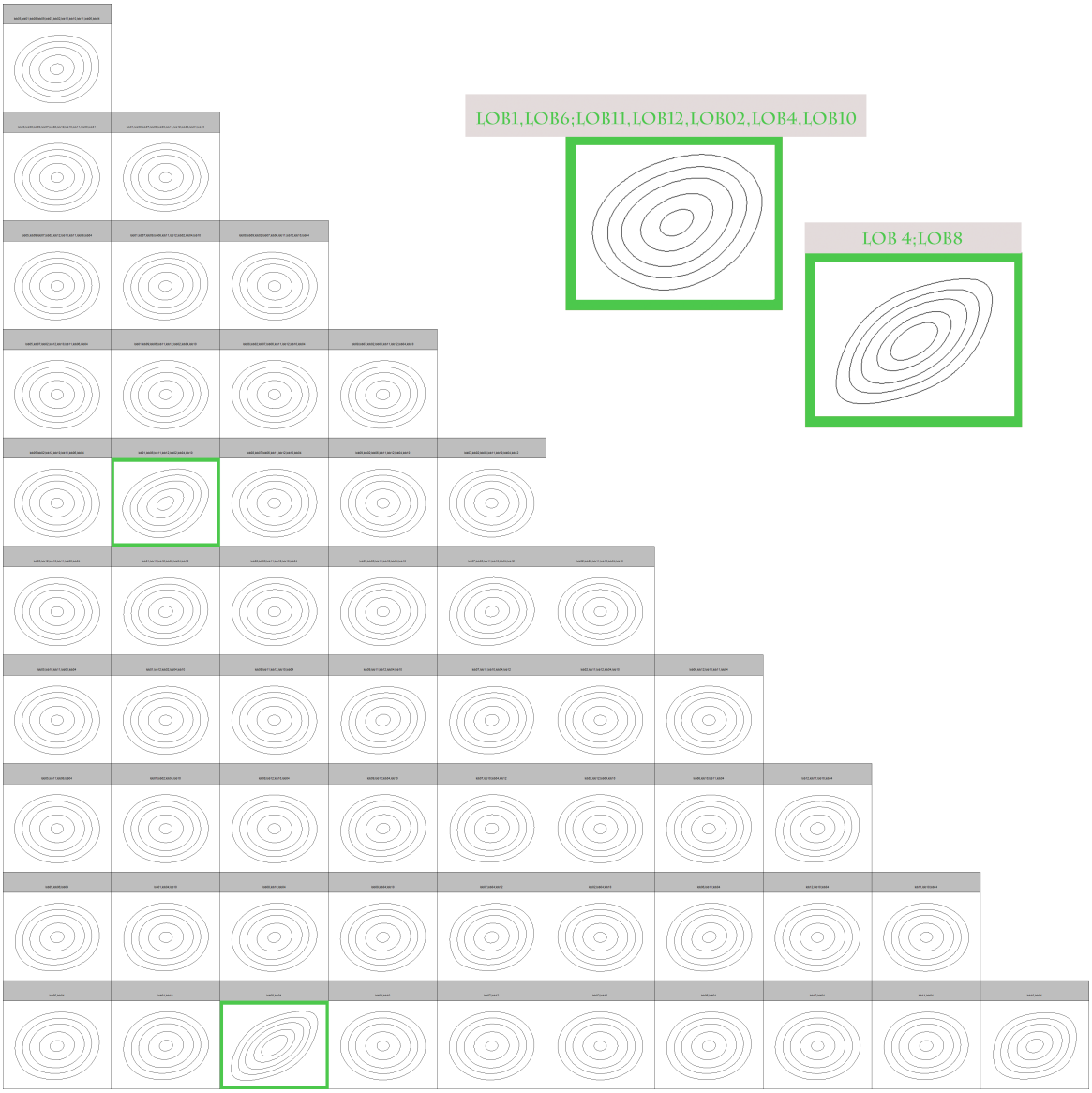


Figura 10.5. Curve Di Livello del Modello $\mathcal{M}_{\mathcal{V}}$.

Quello che si osserva è che il modello risultante dalla simulazione che seleziona il modello migliore sia in termini di `Voung-Test` che di `Clark-Test` avente la `log lik` più elevata, fornisce di fatto un *output* simile a $\mathcal{M}_{\mathcal{D}}$ ovvero quello ottenuto di *default* mediante Dißmann. Infatti la composizione del “paniere delle copule” è sempre circa 30% copule con struttura ellittica (ripartite in modo abbastanza omogeneo tra copule indipendenti ed T-Student) e il resto del paniere da copule parametriche. Dalle curve di livello di $\mathcal{M}_{\mathcal{V}}$ si nota a colpo d’occhio (figura 10.5) che anche in questo caso le *LoB* 4 e *LoB* 8 hanno una dipendenza funzionale non banale e la si osserva rappresentata da curve di livello più schiacciate. Si osserva che anche in questo caso

[8] La $\mathcal{V}$ richiama il concetto di *vincente*.

risulta viene adottata una copula T-Student.

Unitamente a questo, emerge un'ulteriore bi-copula che rappresenta una dipendenza positiva più marcata per la variabile:

$$(LoB\ 1, LoB\ 6; LoB\ 11, LoB\ 12, LoB\ 2, LoB\ 4, LoB\ 10) \tag{10.1}$$

Osservando congiuntamente gli *output* dei modelli $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{V}}$ emerge, al di là di un grado di raffinatezza superiore per quanto riguarda $\mathcal{M}_{\mathcal{V}}$, entrambi presentano numerose copule appartenenti alle famiglie `B` e `BB`. In particolare, come è stato osservato nel capitolo capitolo 3, la copula di Joe, Gumbel che appartengono alla famiglia delle copule a un parametro sono caratterizzate da dipendenza di coda superiore mentre la copula di Frank, anch'assa appartenente alla famiglia delle `B` è caratterizzata da dipendenza negativa. Unitamente per quanto riguarda le copule che utilizzano 2 parametri possono essere utilizzate per catturare più di una tipo di dipendenza, ad esempio un parametro può essere adottato per cogliere la dipendenza dalla coda superiore mentre il secondo per rappresentare la concordanza, oppure un parametro per la dipendenza dalla coda superiore e uno per la dipendenza dalla coda inferiore. Nel caso specifico *output* ottenuto si osserva che le copule in questione presentano una doppia dipendenza di coda, superiore e inferiore (`BB1` e `BB7`) oppure sono sensibili alla concordanza (`BB8`). Evidentemente questo strumento metodologico riesce a cogliere la natura più complessa di dipendenza tra rischi multivariati, una complessità che non riusciva a essere rappresentata in modo adeguato con le tradizionali strutture "semplici".

Dalla simulazione inoltre emerge che il modello $\mathcal{M}_{\mathcal{V}}$ supera sia per Clarke che per Voung $\mathcal{M}_{\mathcal{D}}$, inoltre ha una migliore verosimiglianza, ha un numero di parametri maggiore 66 contro 61.

Inoltre, per capire lo spettro dei risultati della simulazione sono stati raccolti nel corso dalla simulazione i valori di principali indicatori per tutti i *pattern* implementati in modo da non disperdere le principali informazioni riguardanti gli altri modelli implementati.

In particolare nelle figure 10.6 e 10.7 si può osservare che i valori della terna `(loglik,mBIC,nPars)` per ogni *pattern* implementato. Dall'analisi dei metodi grafici presentati emerge che la volatilità di questi indicatori è contenuta in una finestra dei valori particolarmente ristretta.

Si rappresenta attraverso una la linea tratteggiata in rosso i valori del modello ottenuto mediante Dißmann, ovvero $\mathcal{M}_{\mathcal{D}}$, mentre in verde si presentano le risultaze prodotte attraverso il processo *multi-criterio* che ha selezionato il modello $\mathcal{M}_{\mathcal{V}}$.

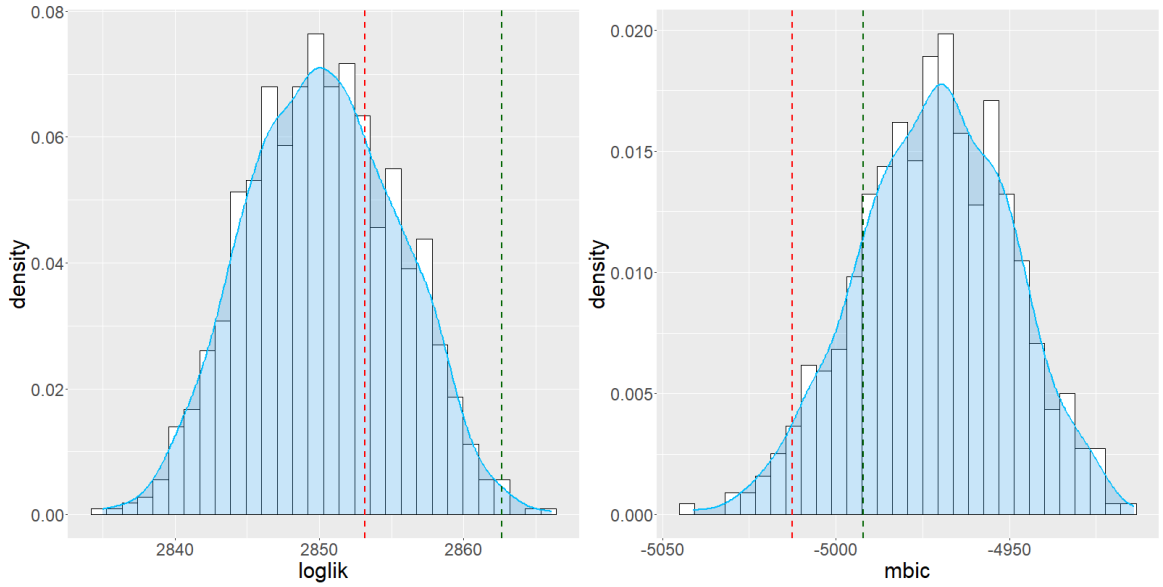


Figura 10.6. Distribuzione `log lik` e `mBIC` nei modelli simulati.

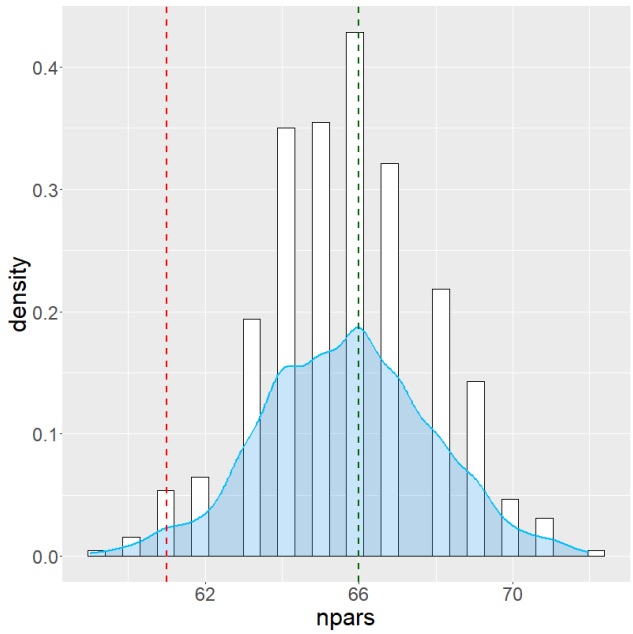


Figura 10.7. Distribuzione del numero di paramatri nei modelli simulati.

Albero	Arco	Vincolo	Famiglia	Rot.	Par.	df	τ
1	1	$(5, 4)$	bb1	180	$(0.055, 1.029)$	2	0.05
1	2	$(1, 10)$	bb8	180	$(1.2, 0.8)$	2	0.04
1	3	$(8, 4)$	t	0	$(0.53, 10.1)$	2	0.35
1	4	$(9, 10)$	indep	0		0	0
1	5	$(7, 12)$	frank	0	0.176	1	0.01
1	6	$(2, 10)$	indep	0		0	0
1	7	$(6, 4)$	t	0	$(0.06, 49.99)$	2	0.03
1	8	$(12, 4)$	clayton	0	0.02	1	0.01
1	9	$(11, 4)$	t	0	$(0.035, 39.67)$	2	0.02
1	10	$(10, 4)$	t	0	$(0.205, 12.335)$	2	0.13
2	1	$(5, 6\|4)$	bb1	0	$(0.07, 1.02)$	2	0.05
2	2	$(1, 4\|10)$	bb8	180	$(1.05, 0.95)$	2	0.02
2	3	$(8, 10\|4)$	frank	0	0.44	1	0.04
2	4	$(9, 4\|10)$	frank	0	0.26	1	0.02
2	5	$(7, 4\|12)$	bb1	180	$(0.05, 1.04)$	2	0.06
2	6	$(2, 4\|10)$	joe	90	1.01	1	-0.005
2	7	$(6, 11\|4)$	bb8	180	$(1.33, 0.73)$	2	0.06
2	8	$(12, 10\|4)$	frank	0	0.14	1	0.01
2	9	$(11, 10\|4)$	clayton	0	0.017	1	0.008
3	1	$(5, 11\|6, 4)$	indep	0		0	0
3	2	$(1, 2\|4, 10)$	joe	180	1.01	1	0.005
3	3	$(8, 12\|10, 4)$	joe	90	1.01	1	-0.005
3	4	$(9, 12\|4, 10)$	bb8	180	$(1.16, 0.78)$	2	0.03
3	5	$(7, 10\|4, 12)$	gumbel	180	1.04	1	0.038
3	6	$(2, 12\|4, 10)$	indep	0		0	0
3	7	$(6, 10\|11, 4)$	bb1	180	$(0.03, 1.01)$	2	0.02
3	8	$(12, 11\|10, 4)$	t	0	$(0.07, 21.27)$	2	0.04
4	1	$(5, 10\|11, 6, 4)$	indep	0		0	0
4	2	$(1, 12\|2, 4, 10)$	clayton	180	0.02	1	0.01
4	3	$(8, 11\|12, 10, 4)$	clayton	90	0.02	1	-0.01
4	4	$(9, 11\|12, 4, 10)$	t	0	$(0.09, 22.72)$	2	0.05
4	5	$(7, 11\|10, 4, 12)$	t	0	$(0.07, 28)$	2	0.04
4	6	$(2, 11\|12, 4, 10)$	indep	0		0	0
4	7	$(6, 12\|10, 11, 4)$	gumbel	180	1.01	1	0.01
5	1	$(5, 12\|10, 11, 6, 4)$	gumbel	90	1.01	1	-0.01
5	2	$(1, 11\|12, 2, 4, 10)$	t	0	$(0.06, 33.24)$	2	0.04
5	3	$(8, 6\|11, 12, 10, 4)$	frank	0	0.12	1	0.01
5	4	$(9, 6\|11, 12, 4, 10)$	gaussian	0	0.03	1	0.02
5	5	$(7, 6\|11, 10, 4, 12)$	bb8	180	$(1.27, 0.76)$	2	0.05
5	6	$(2, 6\|11, 12, 4, 10)$	indep	0		0	0
6	1	$(5, 2\|12, 10, 11, 6, 4)$	clayton	90	0.01	1	-0.007
6	2	$(1, 6\|11, 12, 2, 4, 10)$	bb8	0	$(2.39, 0.59)$	2	0.15
6	3	$(8, 7\|6, 11, 12, 10, 4)$	clayton	180	0.02	1	0.01
6	4	$(9, 2\|6, 11, 12, 4, 10)$	indep	0		0	0
6	5	$(7, 2\|6, 11, 10, 4, 12)$	indep	0		0	0
7	1	$(5, 7\|2, 12, 10, 11, 6, 4)$	indep	0		0	0
7	2	$(1, 9\|6, 11, 12, 2, 4, 10)$	frank	0	0.21	1	0.02
7	3	$(8, 2\|7, 6, 11, 12, 10, 4)$	indep	0		0	0
7	4	$(9, 7\|2, 6, 11, 12, 4, 10)$	t	0	$(0.03, 49.99)$	2	0.02
8	1	$(5, 9\|7, 2, 12, 10, 11, 6, 4)$	bb7	0	$(1.007, 0.02)$	2	0.01
8	2	$(1, 7\|9, 6, 11, 12, 2, 4, 10)$	bb1	180	$(0.05, 1.01)$	2	0.03
8	3	$(8, 9\|2, 7, 6, 11, 12, 10, 4)$	clayton	270	0.04	1	-0.02
9	1	$(5, 8\|9, 7, 2, 12, 10, 11, 6, 4)$	frank	0	0.1	1	0.01
9	2	$(1, 8\|7, 9, 6, 11, 12, 2, 4, 10)$	gumbel	180	1.01	1	0.01
10	1	$(5, 1\|8, 9, 7, 2, 12, 10, 11, 6, 4)$	bb8	0	$(1.13, 0.88)$	2	0.03

Tabella 10.6. Modello $\mathcal{M}_{\mathcal{V}}$ determinato con approccio simulativo.

Ora si iniziano a fare un pò di valutazioni in termini di *model selection*. La comparazione dei valori dei principali indicatori per $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{V}}$ sono in tabella 10.7.

Confronto dei Principali Indicatori					
Modello	`log lik`	Numero Parametri	AIC	BIC	mBIC
$\mathcal{M}_{\mathcal{D}}$	2853.18	61	-5584.36	-5144.53	-5012.507
$\mathcal{M}_{\mathcal{V}}$	2862.659	66	-5593.32	-5117.44	-4992.181

Tabella 10.7. Confronto principali indicatori statistici tra $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{V}}$.

Si osserva è che il modello $\mathcal{M}_{\mathcal{D}}$ in termini di (`AIC`, `BIC`, `mBIC`) è migliore rispetto a $\mathcal{M}_{\mathcal{V}}$ ottenuto con approccio *multi-criterio* basato sul `Voung-Test` e il `Clark-Test` (i valori dei test sono riportati nelle tabelle 10.8 e 10.9).

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
$(\mathcal{M}_{\mathcal{D}}, \mathcal{M}_{\mathcal{V}})$	0.76	0.36	-1.09	0.44	0.71	0.27

Tabella 10.8. Confronto test di Voung tra $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{V}}$.

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
$(\mathcal{M}_{\mathcal{D}}, \mathcal{M}_{\mathcal{V}})$	5305	5280	5191	1×10^{-9}	2×10^{-8}	0.0001

Tabella 10.9. Confronto test di Voung tra $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{V}}$.

Modello	`log lik`	Rapporto
$\mathcal{M}_{\mathcal{D}}$	2853.18	$-$
$\mathcal{M}_{\mathcal{V}}$	2862.659	$\frac{2839.13}{2862.659} = 0.99\%$

Tabella 10.10. Rapporto di verosimiglianza tra $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{V}}$.

10.6 Approccio *Pair Copula* contro approccio *Standard Formula*

Per fare un confronto tra *formula standard* e il modello interno indotto attraverso l'elaborato presentato è necessario ripercorrere le ipotesi alla base *formula standard*. La normativa definisce il requisito patrimoniale di solvibilità (*Solvency Capital Requirement, SCR*) che deve detenere un'impresa di assicurazione per potere esercitare l'attività assicurativa, come l'ammontare delle perdite potenziali a cui è esposto l'assicuratore, stimate ad un livello di confidenza del 99.5% e valutate su un orizzonte temporale annuo. In altre parole, l'SCR complessivo deve corrispondere a una specifica misura di rischio, il Value-at-Risk (VaR), soggetto a un livello di confidenza del 99.5%. L'SCR può essere stimato facendo uso dell'approccio metodologico, denominato *formula standard*, espressamente previsto dal legislatore. Tale metodo prevede di quantificare l'SCR facendo uso di un modello aggregato attraverso correlazioni

lineari, in formule:

$$BasicSCR = \sqrt{\sum_{i,j} Corr_{i,j} \cdot SCR_i \cdot SCR_j} + SCR_{intangibles} \tag{10.2}$$

Dove i pedici i e j individuano i macro-moduli relativi ai rischi.

Tuttavia, ricordiamo che la *formula standard* esiste sotto l'assunzione che i rischi si distribuiscano come una particolare classe di distribuzioni: la classe delle distribuzioni ellittiche. Duque la normativa, si basa sulla semplificazione che prevede di aggregare i rischi come se fossero degli scarti quadratici medi. Come noto, tale approssimazione non è necessariamente soddisfatta nella pratica. In particolare, le matrici di correlazione utilizzate per l'aggregazione di dei rischi e dei sotto moduli di rischi sono stimate per minimizzare l'errore di aggregazione attraverso la seguente formulazione:

$$\min_{\rho} \Big| VaR(X+Y)^2 - VaR(X)^2 - VaR(Y)^2 - 2\rho VaR(X)VaR(Y) \Big| \tag{10.3}$$

Dove X e Y sono variabili casuali che rappresentano due rischi diversi. Quindi, l'unica ipotesi che sorregge l'impianto teorico della *formula standard* è che i rischi siano multi-normali. Infatti, la radice quadrata di una forma quadratica produce una corretta aggregazione dei quantili solo nel caso di distribuzioni ellittiche centrate come una distribuzione normale multivariata.

Dunque, la *formula standard* è di fatto corretta implicitamente sotto tali assunzioni, e partendo proprio da questo presupposto è stato ipotizzato di fittare i dati secondo una distribuzione multi-normale a cui fa riferimento la *formula standard* con la finalità di comprendere se la stima ottenuta durante la trattazione fosse più robusta rispetto al classico approccio *formula standard*.

Per rendere ancora più "robusti" i risultati in modo da enfatizzare maggiormente la trattazione presentata, è stato implementato un modello che fa uso come nella *formula standard* di una struttura di tipo multi-normale che indicheremo con la notazione $\mathcal{M}_{\mathcal{S}}$.

Operativamente è stato utilizzato l'approccio esplorato nel capitolo 8, in cui il modello ricerca l'ottimo all'interno di una famiglia selezionata, nel nostro caso la funzione di ottimo ha determinato il modello facendo uso solo di copule di tipo gaussiano, questo attraverso la funzione `vinecop(data, keep_data = TRUE, family_set = c("gaussian") )`. Presente che la ricerca dell'ottimo locale è solo ristretta a un *basket* di opzioni minori e che l'indipendenza come l'utilizzo di copule gaussiane è già contemplato nell'universo di famiglie selezionabili dall'algoritmo di Dißmann. Tuttavia un confronto diretto può motivare ancor di più le risultaze ottenute e aumentarne la robustezza. Da un'attenta analisi dei principali indicatori a supporto per la stima della bontà di adattamento del modello risultante emerge che il modello $\mathcal{M}_{\mathcal{S}}$, ovvero quello ottenuto attraverso le ipotesi sottostanti la *formula standard* è peggiore rispetto sia $\mathcal{M}_{\mathcal{D}}$ che a quello implementato attraverso processo simulativo $\mathcal{M}_{\mathcal{V}}$ se teniamo conto dell'usuale batteria di indicatori presentati e utilizzati nel corso di tutta la trattazione. In particolare, si evidenzia un peggiore livello della `log lik` e un peggiore `mBIC` rispetto a due modelli determinati nei paragrafi precedenti:

Confronto dei Principali Indicatori		
Modello	`log lik`	`mBIC`
$\mathcal{M}_{\mathcal{D}}$	2853.18	−5012.51
$\mathcal{M}_{\mathcal{V}}$	2862.659	−4992.18
$\mathcal{M}_{\mathcal{S}}$	2662.486	−4676.93

Tabella 10.11. Confronto principali indicatori statistici tra $\mathcal{M}_{\mathcal{D}}$, $\mathcal{M}_{\mathcal{V}}$ e $\mathcal{M}_{\mathcal{S}}$.

Di concerto emerge che anche i principali test, Voung-Test e Clarke-Test, presentano il modello selezionato da Dißmann $\mathcal{M}_{\mathcal{D}}$ e quello implementato attraverso processo simulativo $\mathcal{M}_{\mathcal{V}}$ come migliori rispetto a $\mathcal{M}_{\mathcal{S}}$:

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
($\mathcal{M}_{\mathcal{D}}$,$\mathcal{M}_{\mathcal{S}}$)	8.83	8.56	7.55	9×10^{-19}	1×10^{-17}	1×10^{-14}

Tabella 10.12. Confronto test di Vuong tra $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{S}}$.

Modelli	Statistic	S.Akaike	S.Schwarz	p.value	p.value.Akaike	p.value.Schwarz
($\mathcal{M}_{\mathcal{D}}$,$\mathcal{M}_{\mathcal{S}}$)	5554	5534	5475	$\simeq 0$	$\simeq 0$	$\simeq 0$

Tabella 10.13. Confronto test di Clarke tra $\mathcal{M}_{\mathcal{D}}$ e $\mathcal{M}_{\mathcal{S}}$.

10.7 Conclusioni

In questo capitolo è stato possibile fornire un confronto al tradizionale approccio Solvency II rispetto ad una metodologia non pre-confezionata che consente di costruire un modello *tailored* per una compagnia che esercita tutti i rami di attività *non-life*.

Questo è stato possibile attraverso la *pair-copula construction* che ha definito una struttura di dipendenza *ad-hoc* estrapolando la dipendenza intrinseca dai triangoli *run-off* delle diverse linee di attività facendo uso del *bootstraping* condizionato. Inoltre, una volta determinato il modello mediante Dißmann è stata lanciata una simulazione che ha avuto lo scopo di confrontare il *pattern* ottenuto mediante l'algoritmo di *greedy*, che si ricorda essere un algoritmo di ottimo locale, rispetto a altri mille *pattern* ottenuti attraverso l'approccio simulativo.

Di fatto, attraverso questo processo è stata definita una metodologia che allo stesso tempo "sfida" Dißmann e lo "migliora". Unitamente a questo si è osservato che i valori presentati da $\mathcal{M}_{\mathcal{D}}$ rispetto a $\mathcal{M}_{\mathcal{V}}$ differiscono in modo non rilevante; infatti, la finestra dei valori ottenuti per i principali indicatori di adattamento `(log lik,mBIC,nPars)` presentano uno spettro di risultati non molto distanti rispetto al modello determinato mediante Dißmann.

Nel contesto delle scienze attuariali quanto presentato è un risultato rilevante nel governo d'impresa di una compagnia, infatti le bi-copule ottenute spesso non sono normali e dunque non rappresentano una dipendenza di tipo lineare che il Regolatore richiama all'osservanza nella *standard formula*.

Ulteriomente è stato osservato che le copule definite nei modelli spesso rappresentano legami sulla coda e che spesso tale dipendenza di coda è addirittura doppia. Diversamente si rammenta che il termine "linearità" deriva dalla parola "linea" che richiama ad un *trend* di coerenza nel continuo. Nel mondo la volatilità a grappolo questo non dovrebbe stupirci.

Dunque quanto osservato in questo capitolo è cruciale in un contesto prudenziale: il *buffer* di rischio risultante sia in termini di riserve che per il requisito di capitale dovrebbe essere aumentato sia nell'alveo di valutazioni *market consistent* delle *Best Estimate Liability (BEL)* che per *Solvency Capital Requirement (SCR)*.

L'attività valutativa dell'esperto deve essere sempre finalizzata ad una *fairness opinion* e di fatto se si considera Solvency II "l'immersione" ad una logica del rischio *embedded* si osserva che la struttura di dipendenza ottenuta si discosta in modo evidente da una matrice di correlazione popolata da valori polarizzati pari o a 0.25 o 0.50.

Si segnala che tale approccio con le dovute modificazioni può essere adottato anche nel perimetro di compagnie vita.

Capitolo 11

Conclusioni

> gli allòr ne sfronda, ed alle genti
> svela di che lagrime grondi e di che
> sangue
>
> Foscolo

In questo lavoro di ricerca è stata analizzata la metodologia della *pair-copula construction* mediante l'algoritmo di Dißmann per valutarne l'applicabilità nel contesto delle scienze attuariali. In tal senso, è stato necessario ripercorrere e coniugare l'impianto teorico alla base della *PCC* e l'*alta tecnologia* che ne consente l'implementazione. Si osserva che questa ricerca acquista un senso per la rilevanza che le strutture multivariate occupano nell'analisi dei fenomeni di natura attuariale.

Questa ricerca traccia le potenzialità e i limiti di questo approccio, che non sono legati specificamente alla natura dei rischi assicurativi: alcuni sono ereditati dalla teoria delle copule (come il problema dell'inversione), altre dalla teoria dei grafi (come il problema dell'enumerazione), altre ancora dalle soluzioni *software* disponibili e dai limiti computazionali.

La ricerca ha mostrato che la *PCC* decreta il supermento del problema, lungamento "aperto" e dibattuto, relativo alla rigidità delle n-copule. Ovvero la *pair-copula construction* riesce a cogliere anche in alta dimensione l'effettiva struttura di dipendenza tra variabili aleatorie, al contrario delle tradizionali copule parametriche. Anche quando l'ipotesi semplificante non è soddisfatta, la letteratura mostra che la grande flessibilità del metodo, consente di approssimare la distribuzione congiunta in modo soddisfacente. Unitamente la *PCC* possiede una formulazione della funzione di log-verosimiglianza semplice da utilizzare e analizzare, a differenza dei metodi gerarchici.

L'introduzione degli algoritmi *greedy*, in particolare l'algoritmo di Dißmann, hanno rimosso il principale ostacolo all'adozione della *pair-copula construction* nell'analisi inferenziale di fenomeni in grandi dimensioni. Infatti il tradizionale approccio inferenziale basato sull'enumerazione di tutte le possibilità non è percorribile per la maggiorparte dei problemi reali e soprattutto per quelli di natura attuariale. Seppur i limiti relativi ai tempi computazionali ancora perdurino, i nuovi algoritmi e la disponibilità di *software* ne riducono la rilevanza di tale problematicità.

I diversi casi studio che sono stati presentati mostrano in modo chiaro gli enormi

vantaggi all'uso della *pair-copula construction* nel perimetro delle scienze attuariali. Infatti considerate le attuali richieste normative questa metodologia consente di fornire un'intera struttura di dipendenza per quanto riguarda tutte le *lines of business* dell'attività danni. In particolare si è avuto modo di confrontare il tradizionale approccio Solvency II rispetto ad una metodologia non pre-confezionata che permette la costruzione di un modello *ad-hoc* per una compagnia che esercita tutti i rami di attività *non-life*. In tal senso si osserva che una struttura "alternativa" così fatta può essere considerata una tecnica che mira a consolidare i risultati forniti dall'approccio Solvency II, dunque uno strumento di validazione dei risultati oppure un'alternativa alla formula pre-costrituita, quindi una risposta nella logica del modello interno.

La *PCC* è particolarmente consigliata nel contesto multivariato con variabili continue, ad alte dimensioni e avente una lunga profondità storica. Questo è il tipico *ambiente* attuariale, consueto sia per le problematiche dei rischi *life* che *non-life*. In particolare, nello stesso modo in cui la *PCC* fornisce una risposta valida alla struttura di dipendenza per l'attività dei rischi *non-life*, potrebbe essere adottata in modo del tutto analogo per fornire le stesse risposte del perimetro vita.

Inoltre non si esclude l'adozione dei questa metodologia per quanto riguarda il *value for money* di un prodotto assicurativo. Sotto le giuste condizioni potrebbe essere implementato un approccio a tale problema. Unitamente si segnala che diversi articoli hanno presentato la *pair-copula construction* come metodologia per costruire alcuni indici di portafoglio (come la volatilità) e per alcune misure a rischio (come il VaR). E in tal senso se si guarda ad ulteriori passi per questa ricerca nel contesto attuariale, non si può non considerare l'adozione della *pair-copula construction* congiuntamente alla tecnica di regressione e più in generale in un contesto dinamico.

Tuttavia se in grandi dimensioni l'utilizzo dell'approccio di ottimo "locale" sancisce non solo gli aspetti di enorme vantaggio ma proprio nella natura dell'*high dimensional* la *pair-copula construction* mostra i suoi limiti *the curse of dimensionality* (la maledizione della dimensionalità). In particolare il problema dell'enumerazione può essere "risolto" solo attraverso l'alta tecnologia e risolve il problema non in maniera diretta ma in modo indiretto ovvero non si possono generare tutti le *R-vine* possibili ma è necessario considerare l'approccio simulativo come valida alternativa. E in tal senso si segnala che l'adottare la *PCC* mediante l'algoritmo di Dißmann equivale sempre "accettare" soluzioni "ottime localmente". Gruber e Czado (2015) hanno dimostrato che in scenari di simulazione a 6 dimensioni il *greedy algorithm* di Dißmann ottiene almeno il 75% della vera log-verosimiglianza ed è stato mostrato nel corso di questa ricerca che in dimensione 11 l'approccio simulativo fornisce una stima migliore in termini di `log lik` non significativa.

Segnaliamo che in "piccole" dimensioni i modelli di tipo misto che è stato presentato rappresentano un contributo originale all'applicazione della *PCC* e risulta evidente che un approccio di tipo *mixed-tailored* migliora la stima di Dißmann in modo concreto.

Tuttavia la logica della *pair-copula construction* per essere implementata necessita di una profonda conoscenza del perimetro teorico che dei limiti computazionali. Infatti, le analisi sviluppate per stimare l'utilizzo della *PCC* mostrano che le variabili non continue, largamente utilizzate per lo studio dei rischi assicurativi, non

si presentano come frontiere invalicabili ma comunque necessitano di una maggiore attenzione, rendendo fondamentale la presenza dell'*expert-judgment*.

In particolare, l'analisi approfondita delle variabili discrete di tipo categoriale con metodologia *pair-copula construction* emerge che l'adozione di modelli non parametrici e dunque di bi-copule `ttl` è senz'altro consigliata. Infatti, tali modelli risultano migliori rispetto a modelli che fanno uso esclusivo di copule parametriche, sia per i principali indicatori di adattamento, che per i test per il confronto e la selezione.

Unitamente si segnala che la presenza dell'approccio di *jittering* per le marginali definisce un perimetro di variabilità dei risultati, diversamente a ciò che accade quando si trattano le variabili continue. In tal senso si consiglia sempre l'implementazione di una pluralità di modelli per aumentare il livello di fiducia rispetto alle risultanze fornite dall'algoritmo di Dißmann nel caso discreto.

Inoltre si osserva che la necessità di utilizzare modelli non parametrici basati su funzioni *kernel* rende la *pair-copula construction* per il trattamento delle variabili categoriali uno strumento, seppur flessibile, di difficile comprensione. Tuttavia si ripete, come già segnalato, che nel perimetro delle variabili non continue ricercare nella funzione di copula "il vero legame" tra le marginali e la distribuzione congiunta, che come è noto viene colto per le variabili continue, è illegittimo. Infatti la rappresentazione delle bi-copule nel caso discreto non coglie la "vera natura" della variabili marginali e di conseguenza qualsiasi valutazione qualitativa o quantitativa della dipendenza sottostante basata su di esse è assolutamente forviante.

Appendice A

Appendice

> The use of copulas is however challenging in higher dimensions where standard multivariate copulas suffer from rather inflexible structures.
>
> Brechmann e Schepsmeier

Il presente capitolo ha lo scopo di fare una rassegna sulle principali variabili aleatorie ritenute utili ai fini di questa trattazione.

A.1 Variabili Aleatorie Unidimensionali

In generale si useranno le lettere maiuscole per le variabili casuali unidimensionali X, lettere minuscole per i valori osservati x e per i vettori di variabili aleatorie si userà il grasseto $\boldsymbol{X}$. Come di consueto accade, $\boldsymbol{X}$ rappresenterà il tradizionale vettore riga:

$$\boldsymbol{X} = \begin{pmatrix} X_1, & X_2, & \cdots, & X_n \end{pmatrix}$$

Mentre il rispettivo trasposto verrà indicato con un vettore colonna $\boldsymbol{X}^T$:

$$\boldsymbol{X}^T = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

Sia nel caso di variabili continue che discrete si utilizzerà la lettera f per indicare o la densità o la funzioni di massa di probabilità mentre con la lettera F si denoterà la corrispondente funzione di ripartizione.

A.1.1 Distribuzione Normale

Definizione A.1 (Distribuzione Normale)**.** *Una variabile X si dice v.c.* Normale *con parametri μ e σ^2 e si indica con $X \sim \mathcal{N}(\mu, \sigma^2)$ se è definita su tutto l'asse reale e ha funzione di densità:*

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

dove i parametri μ e σ^2 sono detti media *e* varianza *della distribuzione. Nel caso $\mu = 1$ e $\sigma^2 = 0$ allora la distribuzione è detta* Normale standard *e si denota con $X \sim \mathcal{N}(1, 0)$.*

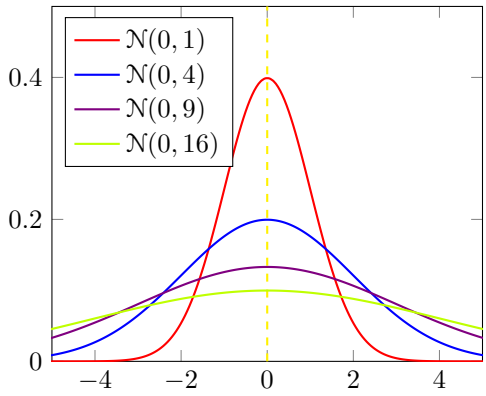


Figura A.1. Distribuzioni Normali con $\mu = 0$ e diverse σ^2.

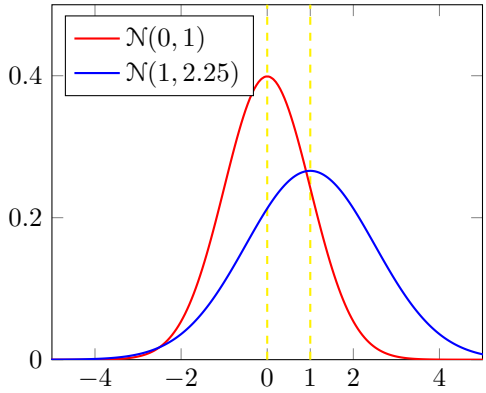


Figura A.2. Una distribuzione Normale confrontata con una non standard.

Tale distribuzione ha numerose proprietà e da essa derivano innumerevoli teoremi che esulano da questa trattazione, tuttavia si richiama solo l'enunciato di un teorema ad esso collegato che ha numerose implicazioni sia in campo teorico che applicativo[1]. Esso dimostra che la combinazione lineare di v.c Normali e indipendenti

[1] Anche la v.c. Gamma gode di questa proprietà.

è ancora una v.c Normale e ha per valore medio la combinazione lineare dei valori medi e per varianza la combinazione lineare dei quadrati dei coefficenti[2].

Teorema A.1 (Proprietà Riproduttiva della v.c Normale)**.** *Siano $X_1, \dots, X_n$ v.c distribuite normalmente tali che $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ per $1 \leq i \leq n$, allora vale la seguente relazione:*

$$\sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\Big(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\Big) .$$

A.1.2 Distribuzione T-Student

Definizione A.2 (Distribuzione T-Student)**.** *Una variabile X si dice v.c.* T_g-Student *di parametri μ e σ^2 con g* gradi di libertà *e si indica con $X \sim \mathrm{T}_g(\mu, \sigma^2)$ se è definita su tutto l'asse reale e ha funzione di densità:*

$$f_g(x; \mu, \sigma^2) = \frac{\Gamma\left(\frac{g+1}{2}\right)}{\sqrt{g\pi}\,\Gamma\left(\frac{g}{2}\right)} \left(1 + \frac{x^2}{g}\right)^{-\frac{g+1}{2}}$$

dove $\Gamma(x)$ è la funzione Gamma di Eulero. Media e varianza sono date da:

$$\mathbb{E}(X) = \mu \qquad\qquad \mathbb{V}\mathrm{ar}(X) = \frac{\mu}{\mu - 1}\sigma^2 .$$

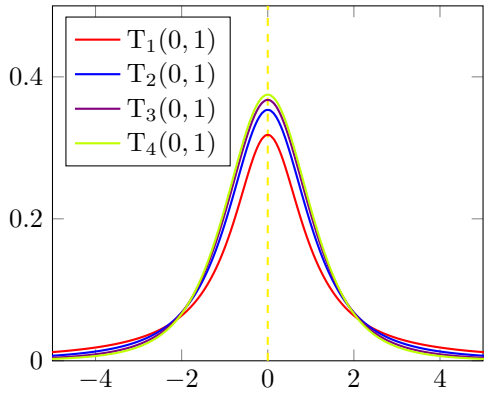


Figura A.3. Distribuzioni $\mathrm{T}_g(0,1)$ con diversi gradi di libertà g.

La distribuzione T-Student ha code più pesanti della distribuzione Normale e si osserva che i dati finanziari sono spesso rappresentati da distribuzioni che presentano code più pesanti della distribuzione Normale, ovvero distribuzioni che assegnano una probabilità maggiore della distribuzione Normale al verificarsi di eventi estremi. Questi aspetti verranno approfonditi nei capitoli sucessivi.

[2]Si ricorda che la somma di g v.c Normali al quadrato è una v.c. χ-quadrato ed è indicata con χ^2_g e inoltre che la v.c $\mathcal{F}$-di Fischer viene "creata" a partire da una v.c Normale, nel caso della $\mathcal{F}$ di Fischer non si parla di combinazioni lineari di v.c. Normali ma rapporto tra due Normali.

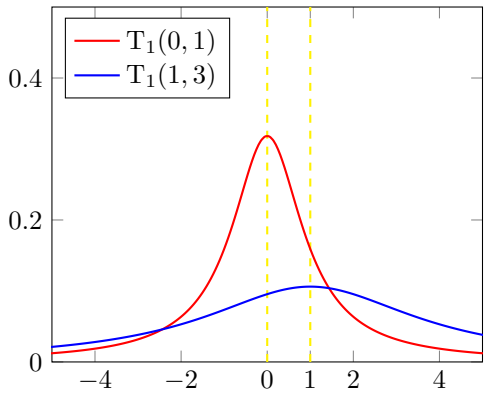


Figura A.4. Una distribuzione $\mathrm{T}_1(0,1)$ affiancata da una T-Student generalizzata.

A.1.3 Il Ruolo Dei Parametri

Si ricorda che esistono principalmente tre tipologie di parametri:

- I parametri di posizione;
- I parametri di scala;
- I parametri di forma.

Nel caso della Normale il parametro di posizione è la media μ e variare esso equivale a fare una traslazione lungo l'asse delle ascisse. I parametri di scala come ad esempio lo scarto quadratico medio o la varianza sono dei parametri che ci consentono di quantificare la dispersione della distribuzione e dunque forniscono una misura della variabilità della dispersione. In questo caso non si parla di dilatazione in quanto la densità di una distribuzione è sempre pari a 1, quindi quello che accade è avere una diminuzione della massa di probabilità nella pancia della distribuzione e un aumento sulle code.

In generale si può dire che a è un parametro di scala per la variabile aleatoria X se:

$$X \to aX$$

Cioè se a la variabile aleatoria si mantiene omogenea per una costante moltiplicativa positiva. Nel caso della Normale accade:

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

Sapete che aY si distribuisce come una Normale:

$$aY \sim \mathcal{N}(a\mu, a^2\sigma^2)$$

La Normale quindi ha soltanto due parametri, un parametro di posizione e un parametro di scala non possiede parametri di forma. Si considerino due distribuzioni Normali:

$$Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$$

$$Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

Qualunque distribuzione Normale la quantità compresa tra media e deviazione standard è sempre una quantità fissa, inoltre è noto che tale quantità risulta pari a $0,65\%$, questo accade qualunque sia la media e qualunque sia il valore della deviazione standard e questo implica che non si interviene sulla probabilità del verificarsi di un dato evento[3]:

$$Prob\{\mu_1 - \sigma_1 \leq Y_1 \leq \mu_1 + \sigma_1\} = Prob\{\mu_2 - \sigma_2 \leq Y_2 \leq \mu_2 + \sigma_2\} = 0,68\%$$

Il parametro di forma agisce ad esempio nella distribuzione T-Student, infatti accade che date due distribuzioni T-Student tali che $\nu_1 \leq \nu_2$:

$$Y_1 \sim T_{\nu_1}$$

$$Y_1 \sim T_{\nu_1}$$

Quello che si ha è che la stessa probabilità che si è visto nel caso precedente, ovvero che gli scarti siano compresi tra la media e più o meno la deviazione standard:

$$Prob\{\mu_1 - \sigma_1 \leq Y_1 \leq \mu_1 + \sigma_1\} \neq Prob\{\mu_2 - \sigma_2 \leq Y_2 \leq \mu_2 + \sigma_2\}$$

Le due probabilità sono diverse e in particolare si ha una relazione di questo tipo:

$$Prob\{\mu_1 - \sigma_1 \leq Y_1 \leq \mu_1 + \sigma_1\} \leq Prob\{\mu_2 - \sigma_2 \leq Y_2 \leq \mu_2 + \sigma_2\}$$

Quindi una probabilità è minore dell'altra, il che discende dal fatto che più sono bassi i gradi di libertà più le code sono pesanti e di conseguenza meno massa di probabilità si ritrova intorno alla media. Diversamente maggiori sono i gradi di libertà più si abbassarano la massa sulle code[4].
Quindi la T-Student è caratterizzata da un terzo parametro, parametro di forma, che è il grado di libertà ν, e il parametro di forma agisce quindi sul quella probabilità che non più è la stessa, appunto perché dà una differente forma funzionale.

Lo strumento solitamente utilizzato per investigare quanto una distribuzione abbia code pesanti è la curtosi, cioè misura la massa di probabilità è concentrata

[3] Ampliando il discorso è noto che:

- l'intervallo [media $\pm 1,96$ volte la deviazione standard] comprende il 95% circa dei dati;
- l'intervallo [media $\pm 2,57$ volte la deviazione standard] comprende il 99% dei dati.

Ovvero:

$$Prob\{\mu - 1,96 \cdot \sigma_1 \leq Y_1 \leq \mu_1 + 1,96 \cdot \sigma_1\} = 0,95\%$$
$$Prob\{\mu - 2,57 \cdot \sigma_1 \leq Y_1 \leq \mu_1 + 2,57 \cdot \sigma_1\} = 0,99\%$$

[4] Non a caso se i gradi di libertà tendono all'infinito mano a mano che i gradi di aumentano la distribuzione T tende alla Normale, quindi le code diventano sempre più leggere man mano che i gradi di libertà aumenteranno.

nel centro della distribuzione. La curtosi per una variabile aleatoria per cui si definito un momento di ordine quarto si valuta nel seguente modo:

$$KUR(Y) = \mathbb{E}\left[\left[\frac{Y - \mathbb{E}[Y]}{\sigma}\right]^4\right] = \frac{\mathbb{E}[[Y - \mathbb{E}[Y]]^4]}{\sigma^4}$$

In corrispondenza di un campione osservato si definisce:

$$K\hat{U}R(\underline{Y}) = \frac{1}{n}\frac{\sum_1^n (y_i - \bar{y})^4}{S^4}$$

E dove S è la stima non distorta dello scarto quadratico medio definito come:

$$S = \sqrt{\frac{1}{n-1}\sum_1^n (y_i - \bar{y})^2}$$

Il valore della Curtosi di una distribuzione Normale è pari a 3, quindi:

$$Y \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow Kur(Y) = 3$$

Dunque distribuzioni che hanno:

- Curtosi maggiore di 3 hanno code più pesanti rispetto alla Normale e si definiscono distribuzioni leptocurtiche.
- Curtosi minore di 3 hanno code più leggere rispetto alla Normale e si definiscono distribuzioni platicurtiche.

Si ricorda che il minimo valore della Curtosi è 1 e questo si verifica quando Y prende due valori ciascun valore con probabilità $\frac{1}{2}$, quindi Y si distribuisce come una Bernulli di parametro $\frac{1}{2}$:

$$Y \sim Bern(\frac{1}{2})$$

Se Y si distribuisce come una T-Student allora:

$$\mathrm{T} \sim T_\nu \Rightarrow Kur(Y) = 3 + \frac{6}{\nu - 4}$$

Dalla relazione sopra si evince che tale momento è definito per $\nu \geq 4$ ed inoltre è una quantità maggiore di 3 quindi la T-Student ha code più pesanti di una distribuzione Normale. Inoltre si osserva che nel momento in cui i gradi di libertà dalla T-Student aumentano essa tende ad una Normale coerentemente con il valore della curtosi che tende a 3.

$$\lim_{\nu \to \infty} T_\nu = \mathcal{N}$$

Esaminando il caso univariato il comportamento all'infinito delle code per una Normale sono:

$$Y \sim \mathcal{N} \Rightarrow f(y) \propto exp\left(-\frac{1}{2}\frac{y}{\sigma}\right)$$

Mentre Y è una T di Student con ν gradi di libertà, guardando la formula della densità della T di Student, potete verificare che la formula è proporzionale ad:

$$Y \sim T_\nu \Rightarrow f(y) \propto \frac{1}{\left(1+\frac{y^2}{\nu}\right)^{\frac{\nu+1}{2}}} \approx \frac{1}{\left(\frac{y^2}{\nu}\right)^{\frac{\nu+1}{2}}} \propto |y|^{-(\nu+1)}$$

Si può scrivere a meno di costanti moltiplicative le densità sulle code di una Normale e di una T di Student, studiando il loro comportamento all'infinito in valore assoluto si può vedere cosa accade:

$$\lim_{|y|\to\infty} \mathcal{N} = \lim_{|y|\to\infty} \mathrm{T} = 0$$

Per T si ricorda innanzitutto che un numero di gradi di libertà maggiori di 0, oovero $\nu > 0$, garantisce che l'integrale della densità si 1:

$$\lim_{|y|\to\infty} T = 0$$

La cosa interessante da ricordare è il limite del rapporto delle due quantità in esame:

$$\lim_{|y|\to\infty} \frac{T}{N} = \infty$$

Quello che succede che l'esponenziale va a 0 più velocemente della foma polinomiale. Quello che si ha è che le code *polinomiali* della T-Student vanno a 0 più lentamente delle code *esponenziali* della Normale. In generale posso definire distribuzioni:

- con una coda polinomiale destra quando la densità:

  $$f(y) \propto Ay^{-(a+1)}$$

  Dove sia A che a sono quantità maggiori di 0, questo quando y a $+\infty$.

- con una coda esponenziale destra, quando la densità:

  $$f(y) \propto Aexp\left(-\frac{y}{\theta}\right)$$

  Dove sia A che θ sono quantità maggiori di 0, questo quando y a $+\infty$.

- con una coda polinomiale sinistra, quando la densità:

  $$f(y) \propto Ay^{-(a+1)}$$

  Dove sia A che a sono quantità maggiori di 0, questo quando y a $-\infty$.

- con una coda esponenziale sinistra, quando la densità:

$$f(y) \propto Aexp\left(-\frac{y}{\theta}\right)$$

  Dove sia A che θ sono quantità maggiori di 0, questo quando y a $-\infty$.

E in termini di momenti valogono questi risultati:

- Se la distribuzione $f(y)$ ha code esponenziali esistono tutti momenti
- Se la distribuzione $f(y)$ ha code polinomiali esistono momenti fino al momento k-esimo, dove $k < min\{a_L, a_R\}$ dove a_L è l'indice della coda sinistra e a_R è l'indice della coda destra

L'analisi delle code è centrale nello studio dei fenomeni assicurativi e assume un'importanza centrale anche nella trattazione delle copule e nella *pair-copula construction*.

A.2 Modelli Parametrici

Il modello parametrico per rappresentare la v.a X viene indicato con l'espressione:

$$X \sim f(-;\theta) \quad \text{con} \quad \theta \in \Theta$$

dove θ è il vettore dei parametri nello spazio campionario Θ.

Generalmente il vettore dei parametri viene stimato con il metodo della massima verosimiglianza, quindi risolvendo il seguente sistema di equazioni:

$$\hat{\theta} \coloneqq \arg\max_{\theta\in\Theta} \prod_{i=1}^{n} f(x_i, \theta)$$

La funzione da massimizzare è chiamata *funzione di verosimiglianza*, mentre la funzione di ripartizione stimata si indica con $\hat{F}(-;\Theta)$ e rappresenta uno stimatore per $F(-;\Theta)$.

A.3 Distribuzioni Empiriche

Nel caso si voglia percorrere una strada diversa dall'adozione dei modelli statistici parametrici si può utilizzare la funzione di distribuzione empirica.

Definizione A.3 (Funzione di distribuzione empirica)**.** *Sia $x_1, \ldots, x_n$ è un campione i.i.d. da una funzione di distribuzione F, allora la funzione di distribuzione empirica è definita come segue:*

$$\hat{F}(x) \coloneqq \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{x_1 \leq x} \quad \textit{per tutte le } x$$

Come è noto dividendo per $n+1$ anziché per n per evitare problemi al contorno per lo stimatore $\hat{F}(x)$.

Definizione A.4 (Distribuzioni empiriche)**.** *Sia R_i il rango dell'osservazione x_i, cioè $R_i = k$ se l'osservazione x_i è la k-esima osservazione più grande tra le osservazioni $x_1, \ldots, x_n$. In questo caso, ne consegue che $\hat{F}(x_i) = R_i$.*

Per caratterizzare la dipendenza tra diverse variabili casuali è necessario standardizzare le variabili e a tale scopo si utilizza la trasformata integrale di probabilità.

Definizione A.5 (Trasformazione integrale di probabilità)**.** *Se $X \sim F$ è una variabile casuale continua e x è un valore osservato di X, allora la trasformazione $u \coloneqq F(x)$ si chiama* trasformazione integrale di probabilità *(PIT) in corrispondenza di x.*

Definizione A.6 (Distribuzione della trasformata integrale di probabilità)**.** *Se $X \sim F$ allora $U \coloneqq F(X)$ è uniformemente distribuita, poiché, per ogni $u \in [0,1]$, risulta*

$$\begin{aligned}\mathbb{P}(U \leq u) &= \mathbb{P}(F(X) \leq u) \\ &= \mathbb{P}(X \leq F - 1(u)) \\ &= F(F^{-1}(u)) = u\,.\end{aligned}$$

Se F è stimata da $F(-;\hat{\theta})$ o dalla distribuzione empirica $\hat{F}$ allora questo vale solo in modo approssimativo.

A.4 Modelli Multivariati

Le distribuzioni multivariate descrivono il comportamento di un fenomeno definito da più variabili casuali, si può in tal senso distinguere distribuzioni congiunte, marginali e distribuzione condizionate derivanti dalla distribuzione multivariata generata dal fenomeno che si vuole rappresentare.

A.5 Distribuzioni Marginali, Congiunte e Condizionate

Si indicherà le distribuzioni marginali con il pedice j mentre le distribuzioni condizionali di X_j dato X_k sono indicate con i pedici $j|k$, la tabella che segue riporta le diverse casistiche.

	Funzione di Densità	Funzione di Ripartizione
Marginali	$f_j(x_j)$, con $1 \leq j \leq d$	$F_j(X_j)$, con $1 \leq j \leq d$
Congiunta	$f(x_1, \ldots, x_d)$, con $1 \leq j \leq d$	$F(X_1, \ldots, X_d)$, con $1 \leq j \leq d$
Condizionata	$f_{j\|k}(x_1, \ldots, x_d)$, con $j \neq k$	$F_{j\|k}(X_j\|X_k)$, con $j \neq k$

A.5.1 Distribuzione Normale Multivariata

Definizione A.7 (Distribuzione Normale Multivariata)**.** *Un vettore di variabili aleatorie $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ di dimensione d si dice v.c* Normale multivariata *con vettore della media $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T \in \mathbb{R}^d$ e matrice di covarianza definita positiva $\Sigma = [\sigma_{i,j}]_{1 \leq i,j \leq d} \in \mathbb{R}^{d \times d}$ se ha densità:*

$$f(\boldsymbol{X}; \boldsymbol{\mu}, \Sigma) = \frac{\sqrt{|\Sigma|}}{(2\pi)^{\frac{1}{d}}} \exp((\boldsymbol{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu}))$$

Ponendo:

$$g(t) = \exp^{-\frac{t}{2}}, \qquad k_d = (2\pi)^{-\frac{d}{2}}$$

si dimostra che la distribuzione Normale multivariata rispetta l'equazione (3.3) e dunque appartiene di diritto alla classe di distribuzioni ellittiche.

Si illustra la densità normale bivariata e le sue linee di contorno (linee solide) con media zero e varianza unitaria e correlazione ρ nei due pannelli più a sinistra.

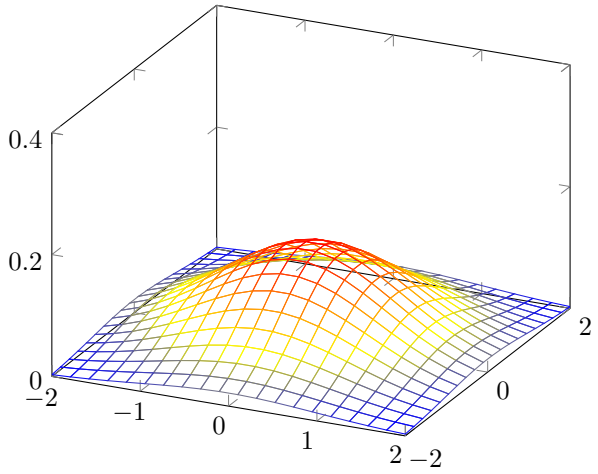


Figura A.5. Funzione di densità di $\mathcal{N}(\mathbf{0}, I_2)$.

A.5.2 Distribuzione Condizionale della Normale

Sia $\boldsymbol{X} \in \mathbb{R}^{d_1+d_2}$ un vettore aleatorio di dimensione $d_1 + d_2$ allora l'algebra lineare ci consente di scriverlo come due vettori $\boldsymbol{X}_1$ e $\boldsymbol{X}_2$ tali che:

$$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)^T$$

dove

$$\boldsymbol{X}_1 = (X_1, \cdots, X_{d_1})^T$$

e

$$\boldsymbol{X}_2 = (X_{d_1+1}, \cdots, X_{d_2})^T$$

Assumendo ora che $\boldsymbol{X}$ sia distribuito come una Normale multivariata quindi avente come la matrice di covarianza la matrice Σ definita positiva della forma:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{d_1+d_2} \\ \vdots & \ddots & \vdots \\ \sigma_{d_1+d_2} & \cdots & \sigma_{d_1+d_2} \end{pmatrix}$$

Allora si può decomporre anche la matrice Σ in una forma a blocchi:

$$\sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix} \tag{A.1}$$

Dove $\Sigma_{11} \in \mathbb{R}^{d_1 \times d_1}$, $\Sigma_{12} \in \mathbb{R}^{d_2 \times d_2}$, e $\Sigma_{22} \in \mathbb{R}^{d_1 \times d_2}$.

Allora si dimostra un risultato notevole ovvero che la distribuzione condizionale di $\boldsymbol{X}_2|\boldsymbol{X}_1$ è ancora una variabile Normale multivariata:

$$\boldsymbol{X}_2|\boldsymbol{X}_1 = x_1 \sim \mathcal{N}_d(\boldsymbol{\mu}_{2|1}, \Sigma_{2|1}) \tag{A.2}$$

Dove il vettore della media condizionale è determinato da:

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \Sigma_{12}^T \Sigma_{11}^T (x_1 - \mu_1)$$

e la matrice di covarianza condizionata da

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}\,.$$

A.5.3 Distribuzione T-Student Multivariata

Definizione A.8 (Distribuzione T-Student Multivariata)**.** *Un vettore di variabili aleatorie* $\boldsymbol{X} = (X_1, \cdots, X_d)^T$ *di dimensione* d *si dice v.c* T- *Student multivariato con* ν *gradi di libertà, con media* $\mu \in \mathbb{R}^d$ *e* Σ *matrice dei parametri di scala, dove* Σ *è una matrice* $d \times d$ *simmetrica e definita positiva con elementi* ρ_{ij}*, indicata con* $\mathrm{T}_d(\nu, \boldsymbol{\mu}, \Sigma)$ *se ha densità:*

$$f(\boldsymbol{X}; \nu, \boldsymbol{\mu}, \Sigma) = \frac{\Gamma(\frac{\nu+d}{2})\sqrt{|\Sigma|}}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{1}{d}}} \exp\Big(1 + \frac{1}{\nu}(\boldsymbol{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{X} - \boldsymbol{\mu})\Big)^{-\frac{\nu+d}{2}}$$

La distribuzione T-Student è detta centrale se $\boldsymbol{\mu} = \mathbf{0}$.

Infine ponendo:

$$g(t) = \Big(1 + \frac{t}{\nu}\Big)^{-\frac{\nu+d}{2}} \qquad\qquad k_d = \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})}$$

si dimostra che la distribuzione anche la T-Student multivariata rispetta l'equazione (3.3) e dunque anche lei appartiene di diritto alla classe di distribuzioni ellittiche.

In generale sia per il caso bivariato, così come nel caso multidimensionale, si dimostra che la distribuzione standard T-Student ha code più pesanti rispetto alla distribuzione gaussiana bivariata con media zero, varianza unitaria e correlazione ρ.

A.5.4 Distribuzione Condizionale della T-Student

Definizione A.9 (Distribuzione Condizionale della T-Student)**.** *Anche la distribuzione* T-*Student ci consente di poter fare il discorso analogo al caso precedente ovvero se il vettore aleatorio* $\boldsymbol{X}$ *si distribuisce come una* T-*Student multivariata* $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)^T \sim \mathrm{T}_d(\nu, \mathbf{0}, \Sigma)$*, e* $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)^T$ *e* Σ *è definita come partizione come nella equazione* (A.1) *allora si può riscrivere anche in questo caso la distribuzione condizionale di* $\boldsymbol{X}_2$ *dato* $\boldsymbol{X}_1$ *come nel caso della Normale:*

$$\boldsymbol{X}_2|\boldsymbol{X}_1 = x_1 \sim \mathrm{T}_d(\nu_{2|1}, \boldsymbol{\mu}_{2|1}, \Sigma_{2|1})$$

dove i paramatri sono univocamente definiti dalle seguenti espressioni:

$$\begin{aligned}
\nu_{2|1} &= \nu + d_1 \\
\boldsymbol{\mu}_{2|1} &= \boldsymbol{\mu}_2 + \Sigma_{12}^T \Sigma_{11}^{-1}(x_1 - \mu_1) \\
\Sigma_{2|1} &= \frac{\nu + (x_1 - \boldsymbol{\mu}_1)^T \Sigma_{11}^{-1}(x_1 - \boldsymbol{\mu}_1)}{\nu + d_1} - (\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12})\,.
\end{aligned}$$

A.6 Distribuzione Empirica Multivariata

Ora se si estende la nozione di distribuzione empirica univariata al caso multivariato che si ritiene utile introdurre in quanto nel corso della trattazione vogliamo evitare le assunzioni parametriche.

Si supponga che $x_i = (x_{1i}, \cdots, x_{di})$ sia un campione di prove ripetute indipendenti e identicamente distribuite con funzione di ripartizione F di dimensione d allora la funzione di ripartizione empirica multivariata è definita nel seguente modo:

$$\hat{F}(x_1, \cdots, x_d) \coloneqq \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}_{\{x_{1i} \leq x_1 \leq x_{di} \leq x_d\}}$$

A.7 Famiglie delle Trasformate di Laplace

Seguendo Joe (1997), si elencano qui di seguito alcune famiglie ad un parametro di LTs $\phi_\theta(s)$ che sono state usate nel capitolo 3:

LTA: per $\theta \geq 1$,

$$\begin{aligned}
\phi_\theta(s) &= \exp\{-s^{\frac{1}{\theta}}\}\,, \\
\phi_\theta^{-1}(t) &= \big(-\log t\big)^\theta\,;
\end{aligned}$$

LTB: per $\theta \geq 0$,

$$\begin{aligned}
\phi_\theta(s) &= (1+s)^{\frac{1}{\theta}}\,, \\
\phi_\theta^{-1}(t) &= t^{-\theta} - 1\,;
\end{aligned}$$

LTC: per $\theta \geq 1$,

$$\begin{aligned}
\phi_\theta(s) &= 1 - \big(1 - \exp\{-s\}\big)^{\frac{1}{\theta}}\,, \\
\phi_\theta^{-1}(t) &= -\log\{1 - (1-t)^\theta\}1\,;
\end{aligned}$$

LTD: per $\theta > 0$,

$$\begin{aligned}
\phi_\theta(s) &= -\theta^{-1} \log\{1 - (1 - \exp\{-\theta\}) \exp\{-s\}\}\,, \\
\phi_\theta^{-1}(t) &= -\log\left\{\frac{1 - \exp\{-\theta t\}}{1 - \exp\{-\theta\}}\right\}.
\end{aligned}$$

La combinazione di LTs della forma $\phi(-\log\psi(s))$, dove ϕ è una famiglia ad un parametro di LTs e ψ è un'altra, porta alla creazione di famiglie a due parametri di LTs. Un esempio di questa costruzione è fornito dalla seguente:

LTE: Per $\delta \geq 1$ e $\theta > 0$,

$$\phi_{\delta,\theta}(s) = \left(1 + s^{\frac{1}{\delta}}\right)^{-\frac{1}{\theta}},$$
$$\phi^{-1}_{\delta,\theta}(t) = (t^{-\theta} - 1)^{\delta}.$$

Joe (1997) definisce altre 4 famiglie di questo tipo ma non sono state usate in questa tesi.

Altre famiglie a due parametri utilizzate sono le seguenti:

LTJ: per $0 < \delta \leq 1$ e $\theta \geq 1$,

$$\phi_{\delta,\theta}(s) = \delta^{-1}\left[1 - \left\{1 - \left[1 - (1-\delta)^{\theta}\right]\exp\{-s\}\right\}^{\frac{1}{\theta}}\right],$$
$$\phi^{-1}_{\delta,\theta}(t) = -\log\frac{1-(1-\delta t)^{\theta}}{1-(1-\delta)^{\theta}};$$

LTL: per $\alpha \geq 0$ e $\theta \geq 1$,

$$\phi_{\alpha,\theta}(s) = \exp\{-(\alpha^{\theta} + s)^{\frac{1}{\theta}} + \alpha\},$$
$$\phi^{-1}_{\alpha,\theta}(t) = (\alpha - \log t)^{\theta} - \alpha^{\theta};$$

LTM: per $\delta \geq 1$ e $\theta > 0$,

$$\phi_{\delta,\theta}(s) = \left[\frac{1-\theta}{\exp(s) - \theta}\right]^{\alpha},$$
$$\phi^{-1}_{\alpha,\theta}(t) = \log\{(1-\theta)t^{-\frac{1}{\alpha}} + \theta\}.$$

Bibliografia

Aas, Kjersti et al. (2009). «Pair-copula constructions of multiple dependence». In: *Insurance: Mathematics and Economics* 44.2, pp. 182–198.

Acar, Elif F., Christian Genest e Johanna G. Nešlehová (2012). «Beyond simplified pair-copula constructions». In: *Special Issue on Copula Modeling and Dependence* 110, pp. 74–90.

Acerbi, Carlo (2002). «Spectral measures of risk: A coherent representation of subjective risk aversion». In: *Journal of Banking & Finance* 26.7, pp. 1505–1518.

Akaike, Hirotugu (1973). «Information Theory and an Extension of the Maximum Likelihood Principle». In: *2nd International Symposium on Information Theory* 73, pp. 1033–1055.

Balakrishnan, Narayanaswamy e Chin-Diew Lai (2009). *Continuous Bivariate Distributions*. Springer New York.

Bedford, Tim e Roger M. Cooke (2002). «Vines: A new graphical model for dependent random variables». In: *Annals of Statistics* 30.4, pp. 1031–1068.

Brechmann, Eike C. (2010). «Truncated and simplified regular vines and their applications». Master's Thesis. Technische Universität München.

Brechmann, Eike C., Claudia Czado e Kjersti Aas (2012). «Truncated regular vines in high dimensions with application to financial data». In: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 40.1, pp. 68–85.

Brechmann, Eike C. e Ulf Schepsmeier (2013). «Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine». In: *Journal of Statistical Software* 52.3, pp. 1–27.

Carley, Holly (2002). «Maximum and Minimum Extensions of Finite Subcopulas». In: *Communications in Statistics - Theory and Methods* 31.12, pp. 2151–2166.

Carriére, Jacques F. (1994). «Dependent Decrement Theory». In.

«A theorem on trees» (2009). In: *The Collected Mathematical Papers*. A cura di Arthur Cayley. Vol. 13. Cambridge Library Collection – Mathematics. Cambridge: Cambridge University Press, pp. 26–28.

Claeskens, Gerda e Nils L. Hjort (2008). *Model Selection and Model Averaging*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.

Clarke, Kevin (2007). «A Simple Distribution-Free Test for Nonnested Model Selection». In: *Political Analysis* 15.

Commissione Europea (10 ott. 2014). *Regolamento Delegato (UE) 2015/35 della Commissione del 10 ottobre 2014 che integra la direttiva 2009/138/CE del Parlamento europeo e del Consiglio in materia di accesso ed esercizio delle attività*

di assicurazione e di riassicurazione (Solvency II). URL: `http://data.europa.eu/eli/reg_del/2015/35/oj`.

Cooke, Roger M. (1997). «Markov and entropy properties of tree-and vine-dependent variables». In: *Proceedings of the Section on Bayesian Statistical Science*.

Cooke, Roger M., Dorota Kurowicka e K. Wilson (2015). «Sampling, conditionalizing, counting, merging, searching regular vines». In: *High-Dimensional Dependence and Copulas* 138, pp. 4–18.

Cormen, Thomas H. et al. (2022). *Introduction to Algorithms*. 4ª ed. MIT Press.

Czado, Claudia (2019). *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R*. Lecture Notes in Statistics. Springer International Publishing.

Czado, Claudia, Stephan Jeske e Mathias Hofmann (2013). «Selection strategies for regular vine copulae». In: *Journal de la société française de statistique* 154.1. Publisher: Société française de statistique, pp. 174–191.

Czado, Claudia, Rainer Kastenmeier et al. (2012). «A mixed copula model for insurance claims and claim sizes». In: *Scandinavian Actuarial Journal* 2012.4, pp. 278–305.

Czado, Claudia, Ulf Schepsmeier e Aleksey Min (2012). «Maximum likelihood estimation of mixed C-vines with application to exchange rates». In: *Statistical Modelling* 12.3, pp. 229–255.

Dall'Aglio, Giorgio, Samuel Kotz e Gabriella Salinetti (1991). *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*. Mathematics and Its Applications. Springer Netherlands.

De Felice, Massimo e Franco Moriconi (2011). *Una nuova finanza d'impresa. Le imprese di assicurazione, Solvency II, le autorità di vigilanza*. Itinerari, Economia / il Mulino. Il Mulino.

Denuit, Michel e Philippe Lambert (2005). «Constraints on concordance measures in bivariate discrete data». In: *Journal of Multivariate Analysis* 93.1, pp. 40–57.

Diestel, Reinhard (2017). *Graph Theory*. 5ª ed. Springer Graduate Texts in Mathematics. Springer-Verlag, © Reinhard Diestel.

Dißmann, Jeffrey F. (2010). «Statistical Inference for Regular Vines and Application». Master's Thesis. Technische Universität München.

Dißmann, Jeffrey F. et al. (2013). «Selecting and estimating regular vine copulae and application to financial returns». In: *Computational Statistics & Data Analysis* 59, pp. 52–69.

Doruet Mari, Dominique e Samuel Kotz (2001). *Correlation and Dependence*. published by Imperial College Press e distributed by World Scientific Publishing Co. 236 pp.

Embrechts, Paul (2009). «Copulas: A Personal View». In: *The Journal of Risk and Insurance* 76.3, pp. 639–650.

Embrechts, Paul, Alexander J. McNeil e Daniel Straumann (2002). «Correlation and Dependence in Risk Management: Properties and Pitfalls». In: *Risk Management: Value at Risk and Beyond*. A cura di Michael A. H. Dempster. Cambridge University Press, pp. 176–223.

Fang, Hong-Bin, Kai-Tai Fang e Samuel Kotz (2002). «The Meta-elliptical Distributions with Given Marginals». In: *Journal of Multivariate Analysis* 82.1, pp. 1–16.

Féron, Robert (1956). «Sur les tableaux de corrélation dont les marges sont données. Cas de l'espace à trois dimensions». In: *Annales de l'ISUP* V.1, pp. 3–12.

Fischer, Matthias (2010). «Multivariate Copulae». In: *Dependence Modeling*. WORLD SCIENTIFIC, pp. 19–36.

Frahm, Gabriel, Markus Junker e Alexander Szimayer (2003). «Elliptical copulas: applicability and limitations». In: *Statistics & Probability Letters* 63.3, pp. 275–286.

Frank, Maurice J. (1979). «On the simultaneous associativity of F(x, y) and x+y-F(x, y)». In: *aequationes mathematicae* 19.1, pp. 194–226.

Frees, Edward W. (2009). *Regression Modeling with Actuarial and Financial Applications*. International Series on Actuarial Science. Cambridge: Cambridge University Press.

Galambos, Janos (1975). «Order Statistics of Samples from Multivariate Distributions». In: *Journal of the American Statistical Association* 70.351. Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 674–680.

Geenens, Gery (2020). «Copula modeling for discrete random vectors». In: *Dependence Modeling*. Dependence Modeling 8.1, pp. 417–440.

Genest, Christian, Kilani Ghoudi e Louis-Paul Rivest (1995). «A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions». In: *Biometrika* 82.3, pp. 543–552.

Genest, Christian e Johanna G. Nešlehová (2007). «A Primer on Copulas for Count Data». In: *ASTIN Bulletin: The Journal of the IAA* 37.2, pp. 475–515.

Genest, Christian, Johanna G. Nešlehová e Bruno Rémillard (1 ago. 2014). «On the empirical multilinear copula process for count data». In: *Bernoulli* 20.3, pp. 1344–1371. DOI: 10.3150/13-BEJ524. URL: https://doi.org/10.3150/13-BEJ524.

Genest, Christian, Jean-François Quessy e Bruno Rémillard (2006). «Goodness-of-Fit Procedures for Copula Models Based on the Integral Probability Transformation». In: *Scandinavian Journal of Statistics* 33, pp. 337–366.

Genest, Christian, Bruno Rémillard e David Beaudoin (2009). «Goodness-of-fit tests for copulas: A review and a power study». In: *Insurance: Mathematics and Economics* 44.2, pp. 199–213.

Gruber, Lutz e Claudia Czado (2015). «Sequential Bayesian Model Selection of Regular Vine Copulas». In: *Bayesian Analysis* 10.4, pp. 937–963.

Gumbel, Emil J. (1935). «Les valeurs extrêmes des distributions statistiques». In: *Annales de l'institut Henri Poincaré* 5.2, pp. 115–158.

Hobæk Haff, Ingrid (2013). «Parameter estimation for pair-copula constructions». In: *Bernoulli* 19.2, pp. 462–491.

Hobæk Haff, Ingrid, Kjersti Aas e Arnoldo Frigessi (2010). «On the simplified pair-copula construction — Simply useful or too simplistic?» In: *Journal of Multivariate Analysis* 101.5, pp. 1296–1310.

Hult, Henrik e Filip Lindskog (2002). «Multivariate extremes, aggregation and dependence in elliptical distributions». In: *Advances in Applied Probability* 34, pp. 587–608.

Hüsler, Jürg e Rolf-Dieter Reiss (1989). «Maxima of normal random vectors: Between independence and complete dependence». In: *Statistics & Probability Letters* 7.4, pp. 283–286.

Hutchinson, T. Paul e Chin-Diew Lai (1990). *Continuous Bivariate Distributions, Emphasising Applications*. Rumsby Scientific Publishing.

International Actuarial Association (2004). *A Global Framework for Insurer Solvency Assessment*. International Actuarial Association.

Joe, Harry (1993). «Parametric Families of Multivariate Distributions with Given Margins». In: *Journal of Multivariate Analysis* 46.2, pp. 262–282.

— (1996). «Families of m-Variate Distributions with Given Margins and m(m-1)/2 Bivariate Dependence Parameters». In: *Lecture Notes-Monograph Series* 28, pp. 120–141.

— (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

— (2005). «Asymptotic efficiency of the two-stage estimation method for copula-based models». In: *Journal of Multivariate Analysis* 94.2, pp. 401–419.

Jogdeo, Kumar (2014). «Dependence, Concepts of». In: *Wiley StatsRef: Statistics Reference Online*.

Jong, Piet de e Gillian Z. Heller (2008). *Generalized Linear Models for Insurance Data*. International Series on Actuarial Science. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511755408.

Kendall, Maurice G. (1938). «A New Measure of Rank Correlation». In: *Biometrika* 30.1, pp. 81–93.

Kim, Gunky, Mervyn Silvapulle e Paramsothy Silvapulle (2007). «Comparisons of Semiparametric and Parametric Methods for Estimating Copulas». In: *Computational Statistics & Data Analysis* 51, pp. 2836–2850.

Kimeldori, George e Allan Sampson (1975). «Uniform representations of bivariate distributions». In: *Communications in Statistics* 4.7. Publisher: Taylor & Francis, pp. 617–627.

Kleinberg, Jon e Éva Tardos (2006). *Algorithm Design*. Pearson/Addison-Wesley.

Kocherlakota, Subrahmaniam e Kathleen Kocherlakota (1992). *Bivariate Discrete Distributions*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis.

Krämer, Nicole et al. (2013). «Total loss estimation using copula-based regression models». In: *Insurance: Mathematics and Economics* 53.3, pp. 829–839.

Kullback, Solomon e Richard A. Leibler (1951). «On Information and Sufficiency». In: *The Annals of Mathematical Statistics* 22.1, pp. 79–86.

Kurowicka, Dorota (2009). «Some results for different strategies to choose optimal vine truncation based on wind speed data.» 3rd Vine Copula Workshop. Oslo.

— (2010). «Optimal Truncation of Vines». In: World Scientific Book Chapters. World Scientific Publishing Co. Pte. Ltd., pp. 233–247.

Kurowicka, Dorota e Harry Joe, cur. (2010). *Dependence Modeling*. World Scientific Publishing. 368 pp.

Ledford, Anthony W. e Jonathan A. Tawn (1996). «Statistics for near independence in multivariate extreme values». In: *Biometrika* 83.1, pp. 169–187.

Lehmann, Erich L. (1966). «Some Concepts of Dependence». In: *The Annals of Mathematical Statistics* 37.5, pp. 1137–1153.

— (2010). *Elements of Large-Sample Theory*. Springer Texts in Statistics. Springer New York.

Lehmann, Erich L. e George Casella (1998). *Theory of point estimation*. 2ª ed. Springer texts in statistics. New York: Springer New York.

Liebscher, Eckhard (2008). «Construction of asymmetric multivariate copulas». In: *Journal of Multivariate Analysis* 99.10, pp. 2234–2250.

Manner, Hans (2007). *Estimation and model selection of copulas with an application to exchange rates*. Maastricht University, Maastricht Research School of Economics of Technology e Organization (METEOR).

Marshall, Albert W. (1996). «Copulas, Marginals, and Joint Distributions». In: *Lecture Notes-Monograph Series* 28, pp. 213–222.

McLeish, Don L. e Christopher G. Small (1988). *The Theory and Applications of Statistical Interference Functions*. Lecture Notes in Statistics. Springer New York.

McNeil, Alexander, Rüdiger Frey e Paul Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Vol. 101.

McNeil, Alexander e Johanna G. Nešlehová (2009). «Multivariate Archimedean copulas d-monotone functions and». In: *The Annals of Statistics* 37, pp. 3059–3097.

Mendes, Beatriz V. M., Mariângela M. Semeraro e Ricardo P. C. Leal (2010). «Pair-copulas modeling in finance». In: *Financial Markets and Portfolio Management* 24.2, pp. 193–213.

Merz, Michael e Mario V. Wüthrich (2008). «Prediction Error of the Multivariate Chain Ladder Reserving Method». In: *North American Actuarial Journal* 12.2, pp. 175–197.

Morales Napoles, Oswaldo (2010). «Bayesian Belief Nets and Vines in Aviation Safety and other Applications». Doctoral Thesis. Technische Universiteit Delft. 169 pp.

— (2011). «Counting Vines». In: *Dependence Modeling. Vine Copula Handbook*. A cura di Dorota Kurowicka e Harry Joe. Singapore: World Scientific Publishing, pp. 189–218.

Morgenstern, Dietrich (1956). «Einfache Beispiele zweidimensionaler Verteilungen». In: *Mitteilungsblatt für Mathematische Statistik* 8, pp. 234–235.

Morillas, Patricia M. (2005). «A method to obtain new copulas from a given one». In: *Metrika* 61.2, pp. 169–184.

Nagler, Thomas (2018). «Asymptotic Analysis of the Jittering Kernel Density Estimator». In: *Mathematical Methods of Statistics* 27.1, pp. 32–46.

Nagler, Thomas, Christian Bumann e Claudia Czado (2019). «Model selection in sparse high-dimensional vine copula models with an application to portfolio risk». In: *Dependence Models* 172, pp. 180–192.

Nagler, Thomas e Claudia Czado (2016). «Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas». In: *Journal of Multivariate Analysis* 151, pp. 69–89.

Nagler, Thomas, Christian Schellhase e Claudia Czado (2017). «Nonparametric estimation of simplified vine copula models: comparison of methods». In: Dependence Modeling 5.1, pp. 99–120.

Nagler, Thomas, Ulf Schepsmeier et al. (10 lug. 2023). *VineCopula: Statistical Inference of Vine Copulas*. Ver. 2.5.0. URL: `https://cran.r-project.org/web/packages/VineCopula/index.html`.

Nagler, Thomas e Thibault Vatter (23 feb. 2023a). *rvinecopulib: High Performance Algorithms for Vine Copula Modeling*. Ver. 0.6.3.1.1. URL: `https://cran.r-project.org/web/packages/rvinecopulib/index.html`.

— (20 feb. 2023b). *vinecopulib: a C++ library for vine copula modeling*. Ver. 0.6.3. URL: `https://vinecopulib.github.io/vinecopulib/`.

Nelsen, Roger B. (2006). *An Introduction to Copulas*. 2ª ed. Springer Series in Statistics. New York: Springer New York. XIV, 272.

Nikoloulopoulos, Aristidis K., Harry Joe e Haijun Li (2012). «Vine copulas with asymmetric tail dependence and applications to financial return data». In: *1st issue of the Annals of Computational and Financial Econometrics* 56.11, pp. 3659–3673.

Okhrin, Ostap, Yarema Okhrin e Wolfgang Schmid (2013). «On the structure and estimation of hierarchical Archimedean copulas». In: *Journal of Econometrics* 173.2, pp. 189–204.

Panagiotelis, Anastasios, Claudia Czado e Harry Joe (2012). «Pair Copula Constructions for Multivariate Discrete Data». In: *Journal of the American Statistical Association* 107.499, pp. 1063–1072.

Pearson, Karl (1895). «Note on Regression and Inheritance in the Case of Two Parents». In: *Proceedings of the Royal Society of London* 58, pp. 240–242.

Pinheiro, Paulo J. R., João M. Andrade e Silva e Maria de Lourdes C. Centeno (2003). «Bootstrap Methodology in Claim Reserving». In: *The Journal of Risk and Insurance* 70.4, pp. 701–714.

Plackett, Robin L. (1965). «A Class of Bivariate Distributions». In: *Journal of the American Statistical Association* 60.310, pp. 516–522.

Puccetti, Giovanni e Marco Scarsini (2010). «Multivariate comonotonicity». In: *Journal of Multivariate Analysis* 101.1, pp. 291–304.

Raftery, Adrian E. (1984). «A continuous multivariate exponential distribution». In: *Communications in Statistics - Theory and Methods* 13.8, pp. 947–965.

Rényi, Alfréd (1959). «On measures of dependence». In: *Acta Mathematica Academiae Scientiarum Hungarica* 10.3, pp. 441–451.

Ricotta, Alessandro (2019). «La valutazione del Non-Life Reserve Risk tramite le Vine Copule e il Collective Risk Model». Tesi di Dottorato. Università di Roma "La Sapienza". 206 pp.

Ripley, Brian D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

Savu, Cornelia e Mark Trede (2010). «Hierarchies of Archimedean copulas». In: *Quantitative Finance* 10.3, pp. 295–304.

Scarsini, Marco (1984). «On measures of concordance.» In: *Stochastica* 8.3, pp. 201–218.

Schmeidler, David (1986). «Integral Representation Without Additivity». In: *Proceedings of The American Mathematical Society* 97, pp. 255–255.

Schmidt, Klaus D. (2006). «Optimal and additive loss reserving for dependent lines of business». In: *Casualty Actuarial Society Forum* 2006 (Fall), pp. 319–351.

Schwarz, Gideon (1978). «Estimating the Dimension of a Model». In: *The Annals of Statistics* 6.2, pp. 461–464.

Schweizer, Berthold e Abe Sklar (1983). *Probabilistic Metric Spaces*. North Holland series in probability and applied mathematics. North Holland.

Schweizer, Berthold e Edward F. Wolff (1981). «On Nonparametric Measures of Dependence for Random Variables». In: *The Annals of Statistics* 9.4, pp. 879–885.

Shepard, Odell (1930). *The Lore of the Unicorn*. George Allen & Unwin.

Shi, Peng e Edward W. Frees (2011). «Dependent Loss Reserving using Copulas». In: *ASTIN Bulletin*. ASTIN Bulletin 41.2. Publisher: Cambridge University Press, pp. 449–486.

Shi, Peng e Lu Yang (2018). «Pair Copula Constructions for Insurance Experience Rating». In: *Journal of the American Statistical Association* 113.521, pp. 122–133.

Shih, Joanna H. e Thomas A. Louis (1995). «Inferences on the Association Parameter in Copula Models for Bivariate Survival Data». In: *Biometrics* 51.4, pp. 1384–1399.

Sibuya, Masaaki (1960). «Bivariate extreme statistics, I». In: *Annals of the Institute of Statistical Mathematics* 11.3, pp. 195–210.

Sklar, A. (1996). «Random Variables, Distribution Functions, and Copulas: A Personal Look Backward and Forward». In: *Lecture Notes-Monograph Series* 28, pp. 1–14.

Sklar, Abe (1959). *Fonctions de Répartition À N Dimensions Et Leurs Marges*. Université Paris 8.

Spearman, Charles (1904). «The Proof and Measurement of Association between Two Things». In: *The American Journal of Psychology* 15.1, pp. 72–101.

— (1906). «A 'Footrule' for Measuring Correlation». In: *British Journal of Psychology, 1904-1920* 2.1, pp. 89–108.

Stöber, Jakob e Claudia Czado (2016). «Pair Copula Constructions». In: *Simulating Copulas*. Vol. Volume 6. Series in Quantitative Finance Volume 6. World Scientific Publishing, pp. 185–230.

Stöber, Jakob, Hyokyoung G. Hong et al. (2015). «Comorbidity of chronic diseases in the elderly: Patterns identified by a copula design for mixed responses». In: *Computational Statistics & Data Analysis* 88, pp. 28–39.

Stöber, Jakob, Harry Joe e Claudia Czado (2013). «Simplified pair copula constructions—Limitations and extensions». In: *Journal of Multivariate Analysis* 119, pp. 101–118.

Stöber, Jakob e Ulf Schepsmeier (2013). «Estimating standard errors in regular vine copula models». In: *Computational Statistics* 28.6, pp. 2679–2707.

Suits, Daniel B. (1957). «Use of Dummy Variables in Regression Equations». In: *Journal of the American Statistical Association* 52.280, pp. 548–551.

Taylor, Greg e Gráinne McGuire (2007). «A Synchronous Bootstrap to Account for Dependencies Between Lines of Business in the Estimation of Loss Reserve Prediction Error». In: *North American Actuarial Journal* 11.3, pp. 70–88.

Tsukahara, Hideatsu (2005). «Semiparametric Estimation in Copula Models». In: *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 33.3, pp. 357–375.

Vapnik, Vladimir (1999). *The Nature of Statistical Learning Theory*. Information Science and Statistics. Springer New York.

Vatter, Thibault e Thomas Nagler (2018). «Generalized Additive Models for Pair-Copula Constructions». In: *Journal of Computational and Graphical Statistics* 27.4. Publisher: Taylor & Francis, pp. 715–727.

Volokh, Eugene (6 gen. 2015). «Zero correlation between state homicide rate and state gun laws». In: *The Washington Post's website*. URL: `https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/10/06/zero-correlation-between-state-homicide-rate-and-state-gun-laws/`.

Vuong, Quang H. (1989). «Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses». In: *Econometrica* 57.2, pp. 307–333.

Wang, Shaun S. (1998). «Aggregation of Correlated Risk Portfolios: Models and Algorithms». In: *Proceedings of the Casualty Actuarial Society* 85, pp. 848–937.

Whelan, Niall (2004). «Sampling from Archimedean copulas». In: *Quantitative Finance* 4.3, pp. 339–352.

Xu, James J. (1996). «Statistical modelling and inference for multivariate and longitudinal discrete response data». Tesi di dott. University of British Columbia.

Yaari, Menahem E. (1987). «The Dual Theory of Choice under Risk». In: *Econometrica* 55.1, pp. 95–115.