



**SAPIENZA**  
UNIVERSITÀ DI ROMA

Tesi di Dottorato

**Costo Sinistri Aggregato e Strutture di Dipendenza: Extreme Value Theory, modelli simulativi e Fast Fourier Transform**

Facoltà di Ingegneria dell'Informazione, Informatica e Statistica  
Dipartimento di Scienze Statistiche  
Dottorato di Ricerca in Scuola di Scienze Statistiche  
Curriculum in Scienze Attuariali  
XXX Ciclo

**Candidato**  
**Francesco Acri**  
**1626211**

Tutor  
Prof. Rocco Roberto Cerchiara

Coordinatore della Scuola  
Prof. Pier Luigi Conti

Referente curriculum in Scienze Attuariali  
Prof. Fabio Grasso

A/A 2016/2017

# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Teoria dei valori estremi e modelli composti</b>	<b>4</b>
1.1 Approccio classico . . . . .	4
1.2 Il metodo Peaks Over Threshold . . . . .	9
1.3 Il metodo Bayesiano . . . . .	14
1.4 Stima delle distribuzioni Extreme Value . . . . .	16
1.5 Distribuzione Kumaraswamy e generalizzazioni . . . . .	21
1.6 Modelli composti . . . . .	25
1.7 Modelli composti con soglia aleatoria . . . . .	28
<b>2 Funzioni Copula e Fast Fourier Transform</b>	<b>32</b>
2.1 Distribuzioni multivariate . . . . .	33
2.2 Funzione Copula . . . . .	34
2.2.1 Alcuni esempi . . . . .	37
2.2.2 Indici di dipendenza . . . . .	41
2.3 Extreme Value Theory multivariata e funzioni Copula . . . . .	43
2.4 Simulazioni . . . . .	48
2.5 Fast Fourier Transform . . . . .	64
2.6 Errori tipici . . . . .	67
<b>3 Misure di rischio</b>	<b>70</b>
3.1 Misure di rischio coerenti . . . . .	70
3.2 Alcune misure di rischio . . . . .	72
3.2.1 Value-at-Risk . . . . .	72
3.2.2 Expected Shortfall . . . . .	74
3.2.3 Misure di rischio distorte . . . . .	78
3.2.4 Misure di rischio spettrali . . . . .	81
3.3 Commenti . . . . .	83
<b>4 Applicazioni a dati reali</b>	<b>85</b>
4.1 Dataset monogaranzia . . . . .	85
4.1.1 Modelli composti . . . . .	87
4.1.2 Funzioni Copula . . . . .	93
4.1.3 Fast Fourier Transform . . . . .	100
4.1.4 Risultati . . . . .	102
4.2 Dataset multigaranzia . . . . .	103

---

4.2.1	Applicazione della Standard Formula . . . . .	105
4.2.2	Modelli composti . . . . .	110
4.2.3	Funzioni Copula . . . . .	113
4.2.4	Fast Fourier Transform . . . . .	118
4.2.5	Risultati . . . . .	120
	<b>Conclusioni</b>	<b>121</b>
	<b>Bibliografia</b>	<b>125</b>

# Introduzione

La corretta determinazione della distribuzione di probabilità del costo sinistri aggregato ha un ruolo fondamentale nell'analisi della rischiosità e della solvibilità di una compagnia di assicurazione e la letteratura ha sviluppato, e continua a farlo, diverse modellistiche complesse che consentono di analizzare i rischi assicurativi in modo integrato, individuando eventuali dipendenze tra gli elementi caratterizzanti il profilo di rischio dell'azienda. Alcune delle tecniche concernenti la modellizzazione del costo sinistri aggregato sono proprio l'oggetto principale del presente lavoro di tesi, in cui particolare importanza è data al concetto di funzione Copula ed alla Fast Fourier Transform.

È noto ormai da tempo, infatti, che i rischi del portafoglio di una compagnia di assicurazione non possono essere considerati del tutto indipendenti ed è sorta la necessità di tener conto dell'associazione tra i rischi come uno degli elementi fondamentali per il calcolo delle aspettative sul futuro andamento dei sinistri e per una valutazione più prudente dei requisiti di capitale in ottica solvibilità.

Intuitivamente, se pensiamo ad una compagnia di assicurazione che esercita la propria attività in almeno due rami assicurativi (o segmenti), è possibile ritenere che una dipendenza positiva e forte tra questi implichi che, se in un anno un segmento presenterà un risultato tecnico piuttosto negativo, anche l'altro risulterà essere nella medesima situazione. Ovviamente, può

anche verificarsi il caso opposto in cui i rami chiuderanno la gestione tecnica in positivo. È ovvio che in tali casi il rischio totale dell'intera compagnia aumenta, rispetto alla situazione di indipendenza. Al contrario, nel caso di una associazione forte ma inversa, avremo che i segmenti tenderanno a bilanciarsi l'un l'altro, determinando una sorta di compensazione dei costi. Di conseguenza, i risultati globali presenterebbero maggiore stabilità ed il rischio totale della compagnia diverrebbe minore, sempre rispetto al caso di indipendenza.

Il presente lavoro è così strutturato:

- il Capitolo 1 riporta i principali risultati della Extreme Value Theory, primo tra tutti il teorema di Fisher & Tippett, nonché le distribuzioni dei valori estremi. Del suddetto teorema verranno prese in considerazione due alternative, ovvero il metodo Peaks Over Threshold (POT) e l'approccio bayesiano; saranno mostrati alcuni modelli composti ed una metodologia di stima di crescente interesse in ambito attuariale e ben nota in quello prettamente statistico, ovvero lo stimatore di minima distanza. Mostriamo anche alcune composizioni basate sulla distribuzione Kumaraswamy;
- il Capitolo 2 descrive il concetto di funzione Copula mediante il quale sarà possibile definire e stimare la - eventuale - struttura di dipendenza tra i sinistri di distinti segmenti oppure tra sinistri attritional e large; le caratteristiche di alcune funzioni Copula sono investigate attraverso procedure di simulazione. Inoltre, viene descritto l'algoritmo Fast Fourier Transform, di seguito FFT;
- il Capitolo 3 è dedicato alla descrizione delle più comuni misure di rischio che vengono impiegate in ambito assicurativo e delle loro proprietà matematiche;

- nel Capitolo 4, sulla base di dati empirici e mediante l'impiego di modelli composti, funzioni Copula ed FFT, si determinerà la distribuzione del costo sinistri aggregato e la conseguente stima del Value-at-Risk (di seguito VaR), base di calcolo del Solvency Capital Requirement. In particolare, presenteremo due distinte applicazioni in cui analizzeremo, rispettivamente, la dipendenza tra sinistri attritional e large di uno stesso segmento e quella tra i sinistri di segmenti differenti.

Alla data di scrittura, alcuni dei risultati del lavoro di tesi sono raccolti in un working paper (Acri, F. & Cerchiara, R. R., 2016) del Dipartimento di Economia, Statistica e Finanza dell'Università della Calabria. Inoltre, sono stati oggetto di presentazione durante la IVASS Conference on Insurance Research, tenutasi a Roma nel luglio 2017, e ai 40° e 41° AMASES Annual Meetings, tenutisi a Catania nel 2016 e a Cagliari nel 2017, rispettivamente.

# Capitolo 1

## Teoria dei valori estremi e modelli composti

### 1.1 Approccio classico

Nel presente capitolo presentiamo la logica che è stata seguita nella prima fase di sviluppo della teoria dei valori estremi. Essa fa riferimento, principalmente, al più importante risultato ottenuto nell'ambito della EVT, il teorema di Fisher & Tippet. Prima di entrare nel merito di tale teorema è necessario introdurre la notazione di base.

Sia  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un campione casuale di variabili aleatorie (v.a.) indipendenti e identicamente distribuite (i.i.d.) con funzione di ripartizione  $F$  (fdr) e funzione di densità  $f$ . Allora le  $n$  v.a.  $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$  sono definite statistiche d'ordine corrispondenti al campione casuale  $X_1, X_2, \dots, X_n$ . Esse non sono altro che le v.a.  $X_1, X_2, \dots, X_n$  ordinate in senso crescente e sono da intendere come funzioni del campione casuale  $\mathbf{X}$  stesso. A differenza delle componenti di quest'ultimo, tuttavia, le statistiche d'ordine non sono indipendenti, in quanto se  $X_{j:n} \geq x$  allora  $X_{j+1:n} \geq x$  (si veda Mood et al., 1993).

Possiamo dire che il valore massimo di un campione di ampiezza  $n$  è una statistica d'ordine, in particolare l' $n$ -esima, cioè  $M_n = \max(X_1, \dots, X_n) = X_{n:n}$ . Per quanto riguarda il minimo vale, invece, la seguente relazione:  $m_n = \min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n) = X_{1:n}$ .

Poniamo l'attenzione su queste due statistiche d'ordine ricavandone, per prima cosa, le rispettive fdr. Relativamente ad  $M_n$ , grazie all'ipotesi di osservazioni i.i.d., avremo:

$$F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^n P(X_i \leq x) = F^n(x) \quad (1.1)$$

Per quanto riguarda, invece, il minimo  $m_n$ :

$$\begin{aligned} F_{m_n}(x) &= P(m_n \leq x) = 1 - P(m_n > x) = 1 - P(X_1 > x, \dots, X_n > x) \\ &= 1 - \prod_{i=1}^n P(X_i > x) = 1 - \bar{F}^n(x) \end{aligned} \quad (1.2)$$

dove la funzione  $\bar{F}(x) = 1 - F(x)$  è nota come *funzione di sopravvivenza*.

Calcolando la derivata prima della (1.1) e della (1.2) otteniamo anche le funzioni di densità di  $M_n$  e di  $m_n$  date, rispettivamente, da:

$$\begin{aligned} f_{M_n}(x) &= n[F(x)]^{n-1}f(x) \\ f_{m_n}(x) &= n[1 - F(x)]^{n-1}f(x) \end{aligned}$$

Ora, prima di enunciare il Teorema di Fisher & Tippett, introduciamo il cosiddetto estremo superiore destro (right-end point) della fdr  $F$ , definito come quel valore a cui tendono i valori estremi, ovvero

$$x_F = \sup\{x \in \mathfrak{R} : F(x) < 1\} \quad (1.3)$$

Si potrebbe dire che  $M_n \xrightarrow{a.s.} x_F$ ; tuttavia, per definire correttamente la

distribuzione del massimo bisogna enunciare il più volte già citato teorema di Fisher & Tippett (si veda, ad esempio, Embrechts et al., 1997):

**Teorema 1.1 (Fisher-Tippett)** *Sia  $\{X_i\}_{i=1}^n$  una successione di v.a. i.i.d.. Se esistono delle costanti  $c_n > 0$  e  $d_n \in \mathfrak{R}$  e una qualche fdr non degenere  $H$ , tale che:*

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left[ \frac{M_n - d_n}{c_n} \leq x \right] &= \lim_{n \rightarrow \infty} P(M_n \leq c_n x + d_n) \\ &= \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x) \end{aligned} \quad (1.4)$$

allora  $H$  ha la seguente espressione:

$$H(x; \mu, \varphi, \xi) = \begin{cases} \exp \left\{ - \left( 1 + \xi \frac{x - \mu}{\varphi} \right)^{-\frac{1}{\xi}} \right\} & \text{se } \xi \neq 0 \\ \exp \left\{ - \exp \left( - \frac{x - \mu}{\varphi} \right) \right\} & \text{se } \xi = 0 \end{cases} \quad (1.5)$$

con  $\mu \in \mathfrak{R}$  parametro di locazione,  $\varphi > 0$  parametro di scala,  $\xi \in \mathfrak{R}$  parametro di forma, e tali che  $1 + \xi \frac{x - \mu}{\varphi} > 0$ .

Tale famiglia di distribuzioni prende il nome di *Generalized Extreme Value Distribution* (di seguito *GEV*). In particolare, ponendo nella (1.5)  $\mu = 0$  ed  $\varphi = 1$ , otteniamo la *GEV standardizzata*:

$$H(x; \xi) = \begin{cases} \exp \left\{ - \left( 1 + \xi x \right)^{-\frac{1}{\xi}} \right\} & \text{se } \xi \neq 0 \\ \exp \left\{ - \exp(-x) \right\} & \text{se } \xi = 0 \end{cases} \quad (1.6)$$

Tra i tre parametri che caratterizzano la (1.5) quello di maggiore importanza è il parametro di forma  $\xi$ . Esso è detto anche *extreme value index* e la sua stima influenzerà notevolmente la stima degli altri due parametri,  $\mu$  e  $\varphi$  (rispettivamente, parametri di locazione e di scala), nonchè la determinazione del right end-point (1.3). Inoltre, tale parametro permetterà di “pesare” in modo appropriato la coda destra della funzione di sopravvivenza (1.2).

Potrebbero tornare utili anche la funzione quantile

$$x_p = \begin{cases} \mu + \frac{\varphi}{\xi} \left[ (-\log(p))^{-\xi} - 1 \right] & \text{se } \xi \neq 0 \\ \mu - \varphi \log(-\log(p)) & \text{se } \xi = 0 \end{cases} \quad (1.7)$$

nonché la media e lo scarto quadratico medio della (1.5):

$$E(H) = \mu + \frac{\xi}{\varphi} \left( 1 - \left( 1 + \frac{1}{\xi} \right) \right) \text{ per } \xi > -1$$

$$\sigma(H) = \frac{\xi}{\varphi} \sqrt{\left( 1 + \frac{2}{\xi} \right)^2 - \left( 1 + \frac{1}{\xi} \right)^2} \text{ per } \xi > -0.5$$

In base al valore assunto da  $\xi$ , è possibile individuare tre diversi casi particolari della (1.5), ossia:

- la distribuzione Gumbel (d'ora in poi  $\Lambda(x)$ ) definita su tutto l'asse reale e per  $\xi \rightarrow 0$ . Essa è spesso nota come “La” distribuzione dei valori estremi ed è utilizzata per modellizzare  $M_n$ ;
- la distribuzione Frèchet ( $\Phi(x)$ ), definita per  $x \geq \mu$  e per  $\xi > 0$ ;
- e, infine, la distribuzione Weibull ( $\Psi(x)$ ) definita per  $x \leq \mu$  e per  $\xi < 0$ .

Nella pratica, basterà fare inferenza sul parametro di forma per capire a quale di queste distribuzioni si riduce la GEV. Nel grafico successivo riportiamo, a titolo esemplificativo, la funzione di densità della Gumbel ( $\mu = 0$  e  $\varphi = 1$ ), della Frèchet ( $\mu = 0$ ,  $\varphi = 1$  e  $\xi = 5$ ) e della Weibull ( $\mu = 0$ ,  $\varphi = 1$  e  $\xi = 2$ ):

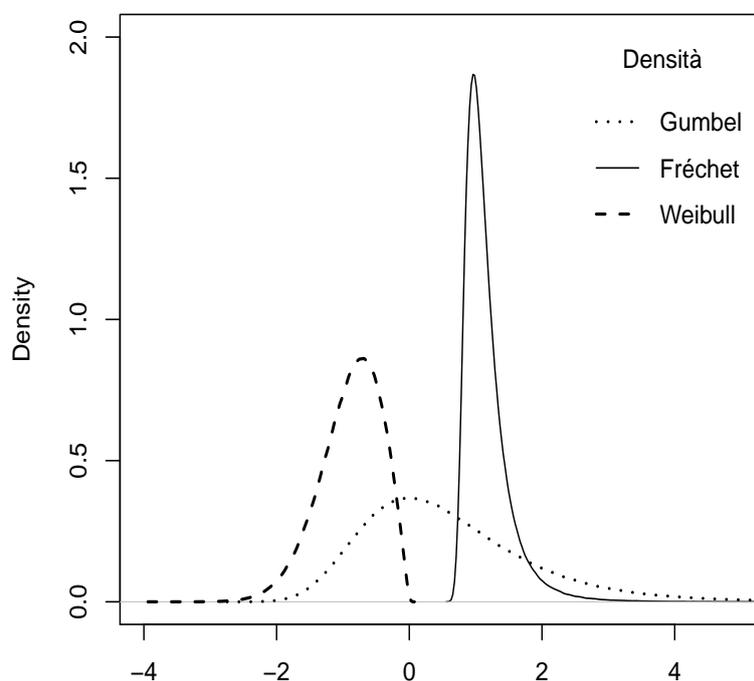


Figura 1.1: Funzioni di densità di una distribuzione Gumbel, Frèchet e Weibull negativa.

Alla luce del teorema (1.1), sono ora note le condizioni da soddisfare affinché  $M_n$  converga ad  $H$ . Tuttavia, anche se il teorema garantisce l'esistenza delle costanti  $c_n$  ed  $d_n$ , non risulta chiarissimo il modo in cui dovranno essere scelte. Una possibilità viene descritta in Embrechts et al. (1997).

Per una maggior precisione formale introduciamo la seguente definizione:

**Definizione 1.1 (Dominio di Attrazione)** *Una v.a.  $X$  (e la sua fdr  $F$ ) appartiene al dominio di attrazione della GEV se esistono delle costanti  $c_n > 0$ ,  $d_n \in \mathfrak{R}$  tali per cui valga la (1.4). In tal caso scriveremo  $X \in MDA(H)$  (o equivalentemente  $F \in MDA(H)$ ).*

In particolare, se  $X$  ha  $x_F = \infty$  e  $1 - F(x) = o(cx^{-\frac{1}{\xi}})$ <sup>1</sup>, per  $x \rightarrow x_F$ ,

<sup>1</sup>Ciò significa che  $\lim_{x \rightarrow x_F} \frac{1-F(x)}{cx^{-\frac{1}{\xi}}} = 0$

allora  $X \in MDA(H)$  con  $H$  di tipo Gumbel. La normale, la lognormale, l'esponenziale sono esempi di distribuzioni appartenenti al dominio di attrazione di una Gumbel.

## 1.2 Il metodo Peaks Over Threshold

L'approccio classico appena descritto è stato, ed è tuttora, oggetto di numerose critiche. In particolare, tali obiezioni sono legate proprio al risultato più importante della teoria: il teorema di Fisher & Tippett. Esso, infatti, richiede, per la sua applicazione *pratica*, un'ingente mole di dati derivanti da un campione i.i.d. e a tali aspetti sono strettamente legati due problemi. In primis, come è ben noto, non sempre è possibile avere osservazioni provenienti da un campione i.i.d.; secondo, anche se volessimo supporre la disponibilità di un campione siffatto in modo da rendere più semplice la trattazione del problema, dovremmo comunque fronteggiare il problema derivante dai pochi dati a disposizione. Proprio questa scarsità di osservazioni non garantisce l'affidabilità delle stime.

Ecco perché sono state proposte diverse soluzioni alternative, ma non per questo meno interessanti. Tra queste citiamo, ad esempio, il modello Poisson-GPD, la tecnica Point Process, il metodo Hill plot ed il Peaks Over Threshold (POT), metodo che descriviamo in tale sezione. Esistono inoltre metodi self-calibrating che in maniera iterativa permettono di determinare un valore (o un range) della soglia, si veda ad esempio Debbabi et al. (2016). Per altre tecniche e per ulteriori dettagli si rimanda ad Embrechts et al. (1997).

Il metodo della soglia consente di attenuare, in parte, il cruccio della scarsità di dati ai livelli estremi e si caratterizza per il fatto che tiene conto di tutte quelle osservazioni che superano una determinata soglia (da qui

il nome del metodo). Non consideriamo, dunque, soltanto i massimi (o i minimi) riferiti ad un determinato arco temporale, né tanto meno siamo interessati a tutta la distribuzione campionaria, bensì solo a una particolare “porzione” di essa, quella formata appunto dai soli dati eccedenti la soglia prescelta. Indicheremo tale valore con  $u$ , mentre  $N$  indicherà il numero di dati che eccedono  $u$  ed  $\mathbf{X}$ , come al solito, sarà il nostro campione casuale i.i.d..

In tale contesto siamo interessati a conoscere la probabilità che un valore eccedente la soglia  $u$  risulti inferiore di una certa quantità  $x$ , con  $x \geq u$ .

Sia  $F^{[u]}(x)$  la fdr di  $X$  dato che  $X > u$ , allora:

$$F^{[u]}(x) = P(X \leq x | X > u) = \frac{P(X \leq x, X > u)}{P(X > u)} = \frac{F(x) - F(u)}{1 - F(u)} \quad (1.8)$$

Sulla base della (1.8) otteniamo la fdr degli eccessi oltre la soglia  $u$ :

$$\begin{aligned} F^{(u)}(x) &= P(X - u \leq x | X > u) = P(X \leq x + u | X > u) \\ &= F^{[u]}(x + u) = \frac{F(x + u) - F(u)}{1 - F(u)} \end{aligned} \quad (1.9)$$

A questo punto, possiamo definire una funzione di primaria importanza per l'approccio POT denominata *funzione dell'eccesso medio*:

$$e_F(u) = E[X - u \leq x | X > u] = \int x dF^{(u)}(x) \quad (1.10)$$

che varia al variare di  $u$ . Essa è fondamentale in quanto permette di individuare un adeguato valore per la soglia  $u$ .

In campo assicurativo, la (1.10) viene spesso detta *funzione dell'eccesso medio di perdita* e può essere interpretata come il danno (economico) atteso oltre la soglia  $u$  (che in tal caso funge da priorità). Per priorità si intende la soglia al di sotto della quale l'impresa di assicurazione è disposta

ad assicurare un determinato rischio. Se per l'evento oggetto del contratto si stimano esborsi maggiori della suddetta priorità, l'impresa, non potendo/volendo coprire il rischio autonomamente, ricorre alla riassicurazione o, addirittura, rifiuta di stipulare il contratto. Da ciò capiamo quanto sia importante scegliere in modo appropriato tale valore limite.

In modo del tutto analogo a quanto è stato fatto per la distribuzione del massimo  $M_n$ , anche per gli eccessi possiamo definire una distribuzione *limite* che può essere approssimata dalla seguente:

$$\begin{aligned} \lim_{u \rightarrow x_F} F^{(u)}(x) &= \lim_{u \rightarrow x_F} P(X - u \leq x | X > u) \\ &= \lim_{u \rightarrow x_F} P(X \leq x + u | X > u) \\ &= G(x; \sigma_\mu, \xi) \end{aligned} \quad (1.11)$$

dove  $G$  è la distribuzione generalizzata di Pareto (di seguito GPD), definita come:

$$G(x; \mu, \sigma_\mu, \xi) = \begin{cases} \left(1 + \xi \frac{x - \mu}{\sigma_\mu}\right)^{-1/\xi} & \text{se } \xi \neq 0 \\ \exp^{-\frac{x - \mu}{\sigma_\mu}} & \text{se } \xi = 0 \end{cases} \quad (1.12)$$

dove  $x \geq 0$  se  $\xi \geq 0$  mentre  $0 \leq x \leq -\sigma_\mu/\xi$  se  $\xi < 0$ .

Pickands (1975) ha dimostrato che la (1.11) può essere considerata un'ottima approssimazione degli eccessi oltre  $u$  quando si verifica la (1.4), cioè quando vale il Teorema di Fisher & Tippett.

Infine, potremmo essere interessati a conoscere il tempo medio intercorrente tra due successivi eventi estremi. A tale scopo, definiamo inizialmente il cosiddetto *periodo di ritorno*. Consideriamo sempre il nostro campione  $\mathbf{X}$ , la soglia  $u$  e sia  $\{I_{X_i > u}\}$  la sequenza di v.a. Bernoulli i.i.d. con probabilità di successo  $p = \bar{F}(u)$ . Indichiamo con  $L(u)$  il numero di osservazioni necessarie affinché si verifichi il primo successo (cioè, il primo eccesso oltre

la soglia  $u$ ):

$$L(u) = \min\{i \geq 1 : X_i > u\}$$

È immediato constatare che  $L(u)$  si distribuisce secondo una v.a. Geometrica di parametro  $p$ :

$$P(L(u) = k) = p(1 - p)^{k-1}, k = 1, 2, \dots$$

Il periodo di ritorno degli eventi  $X_i > u$  è definito come l'aspettativa di  $L(u)$ :

$$E[L(u)] = p^{-1} = [\bar{F}(u)]^{-1} \quad (1.13)$$

Sapendo che la probabilità di avere almeno un eccesso oltre la soglia  $u$  entro  $k$  osservazioni è  $P(L(u) \leq k) = p \sum_{i=1}^k (1 - p)^{i-1} = 1 - (1 - p)^k$ , otteniamo la probabilità di avere almeno un eccesso oltre la soglia **prima** del periodo di ritorno come:

$$P(L(u) \leq E[L(u)]) = P(L(u) \leq p^{-1}) = 1 - (1 - p)^{1/p} \quad (1.14)$$

Fissata una certa soglia, sarebbe buona cosa che questa probabilità fosse prossima allo zero.

Spendiamo alcune parole sulla scelta della soglia  $u$  che caratterizza, di fatto, il metodo POT e che ne determina la capacità di restituire un buon modello. Sappiamo, infatti, che in tale contesto è necessario individuare quella distribuzione teorica che ben si adatta agli eccessi oltre il valore  $u$  e abbiamo già citato alcune tecniche grafiche utili nella determinazione di quest'ultimo; tuttavia, un grafico come quello dell'eccesso medio (1.10) o come L-moments Plot sono soggetti ad interpretazioni e, pertanto, non indicano in maniera inequivocabile l'entità della soglia. Proprio per questo motivo, diversi autori hanno proposto procedure di scelta più oggettive e che non

diano troppo spazio alle “esigenze” del ricercatore, il quale potrebbe scegliere un determinato valore solo perché di comodo. Tra queste citiamo quelle suggerite in Zhou et al. (2007) e Pigeon & Denuit (2011), rispettivamente. In particolare, i primi propongono di stimare la soglia  $u$  partendo dallo stimatore di Hill per giungere, attraverso una regressione, alla individuazione del “turning point”  $TP$ . Esso altro non è che il numero di statistiche d’ordine che popolano la coda destra; se, per esempio, abbiamo  $TP = 100$  allora è necessario che la coda destra sia composta dalle ultime 100 statistiche d’ordine. Inizialmente, per determinare il  $TP$  venne proposto il cosiddetto plot di Hill (basato anch’esso sullo stimatore di Hill) che, tuttavia, come tutte le tecniche grafiche già citate, non fa uso di alcuna regressione, lasciando spazio a soggettive interpretazioni. Riportiamo, di seguito, lo stimatore di Hill, da non confondere con lo stimatore Generalizzato di Hill già visto in precedenza:

$$H(k) = \frac{1}{k} \sum_{j=1}^k \ln(X_{n+1-j:n}) - \ln(X_{n-j:n}), k = 1, \dots, n-1 \quad (1.15)$$

dove  $k$  indica il numero di osservazioni nella coda. Il plot di Hill riporta  $k$  sull’asse delle ascisse e l’inverso dello stimatore di Hill sulle ordinate. Il  $TP$  è quel punto a partire dal quale il grafico inizia ad assomigliare a quello di una serie storica stazionaria.

Come individuiamo questo punto di stazionarietà? Anzichè limitarsi alla visione del plot di Hill e senza addentrarsi nella metodologia Box-Jenkins delle serie storiche, Zhou et al. (2007) hanno proposto la seguente regressione per  $H(k)$ :

$$H(k) = \beta_0 + \beta_1 k + \epsilon(k), k_1 \leq k \leq k_2 \quad (1.16)$$

dove  $k_1 = \lceil n/10 \rceil$ ,  $k_2 \leq \lceil n/2 \rceil$ ,  $\epsilon(k)$  è il termine di errore,  $n$  è la dimensione

campionaria. La (1.16) dovrebbe essere sufficiente per garantire la stazionarietà di  $H(k)$ . Una volta stimati i parametri della regressione, bisognerà implementare i seguenti passaggi:

- calcolare  $e(k) = H(k) - (\hat{\beta}_0 + \hat{\beta}_1 k)$ ,  $1 \leq k \leq k_2$
- calcolare  $\hat{s} = \sqrt{\frac{1}{k_2 - k_1} \sum_{k=k_1}^{k_2} e(k)^2}$
- il turning point risulta  $k^* = \max\{k : |e(k)| \geq \hat{s}\lambda, 1 \leq k \leq k_1\}$

$\lambda$  è un parametro positivo. La soglia  $u$  è quel valore tale che gli eccessi oltre essa sono esattamente  $k^*$ .

La proposta di Pigeon & Denuit (2011) verrà descritta più avanti, nel contesto dei modelli composti.

Concludiamo la sezione riportando la relazione che lega le due distribuzioni viste in precedenza, la GEV  $H$  e la GPD  $G$ :

$$G(x; \sigma_\mu, \xi) = 1 + \ln(H(x\sigma_\mu/\xi)) = 1 - (1 + x\sigma_\mu)^{-1/\xi} \quad (1.17)$$

## 1.3 Il metodo Bayesiano

Il metodo della soglia appena descritto non permette di risolvere completamente il problema legato alla scarsità dei dati ai livelli estremi; di sicuro, considerando i valori eccedenti una determinata soglia riusciamo ad avere un numero di osservazioni maggiore dei soli massimi (minimi) ma ciò potrebbe non bastare.

In termini inferenziali, se riuscissimo ad avere informazioni aggiuntive sul fenomeno analizzato potremmo ottenere stime sicuramente più affidabili. Tali informazioni, tuttavia, devono derivare da una conoscenza approfondita del fenomeno stesso per evitare di aggiungere informazioni ridondanti e “pericolose”.

L'approccio che descriviamo ora deriva dalla teoria bayesiana la quale permette di stimare i parametri di una distribuzione tenendo conto di informazioni a priori derivanti da una conoscenza pregressa, o a volte presunta tale, del fenomeno in esame. Spesso risulta necessario l'aiuto di alcuni esperti del settore (si veda in proposito Coles & Tawn, 1996a). In termini matematici e statistici tali informazioni devono essere tradotte in una qualche funzione che chiameremo *prior* e che indicheremo con  $\pi(\boldsymbol{\theta})$ , dove  $\boldsymbol{\theta}$  indica un parametro incognito (o un vettore di parametri) ed è da intendere come la determinazione della variabile casuale  $\Theta$ . Tale *prior* non può che essere, appunto, la funzione di densità della  $\Theta$  la cui determinazione  $\boldsymbol{\theta}$  racchiude tutta la nostra conoscenza (più o meno soggettiva) sul fenomeno. Essa verrà poi combinata (nel senso del teorema di Bayes) con la funzione di verosimiglianza  $L(\mathbf{x}; \boldsymbol{\theta})$  per ottenere la *densità a posteriori* del parametro incognito, ovvero

$$\pi(\boldsymbol{\theta}|\mathbf{x}) = \frac{\pi(\boldsymbol{\theta})L(\mathbf{x}; \boldsymbol{\theta})}{\int_{\Theta} \pi(\boldsymbol{\theta})L(\mathbf{x}; \boldsymbol{\theta})d\boldsymbol{\theta}} \quad (1.18)$$

Il vettore  $\boldsymbol{\theta}$  potrebbe essere stimato mediante il metodo della massima verosimiglianza, scegliendo come funzione di verosimiglianza la densità a posteriori. In particolare, potremmo stimare  $\boldsymbol{\theta}$  con la moda, la mediana o la media della distribuzione a posteriori. Parlando in termini più generali, è possibile stimare una funzione  $f(\boldsymbol{\theta})$  come media di  $f(\Theta)$  subordinata al campione osservato  $\mathbf{x}$ . Pertanto lo stimatore bayesiano a posteriori di  $f(\boldsymbol{\theta})$  rispetto alla *prior* risulta:

$$E(f(\Theta)|\mathbf{x}) = \int_{\Theta} f(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} \quad (1.19)$$

Vi sono tuttavia dei problemi: per prima cosa, non è detto che si è in grado di definire la prior sui parametri della distribuzione di interesse. Po-

tremmo avere informazioni e conoscenze a priori sul fenomeno, ma come tradurle matematicamente? Supponendo comunque di avere tale funzione, potremmo incontrare problemi nella determinazione della prior a posteriori a causa dell'integrale presente nella (1.18). Esso, a volte, deve essere valutato numericamente ed è proprio per tale motivo che l'approccio classico e il metodo POT sono stati inizialmente preferiti a quello bayesiano. Tuttavia, oggi la valutazione numerica di un integrale non è un problema e, in particolare, da quando i metodi di simulazione MCMC (Monte Carlo Markov Chain) sono stati implementati grazie alla disponibilità di calcolatori sempre più efficienti, il metodo bayesiano ha visto crescere la propria popolarità.

Diverse sono le applicazioni dell'approccio bayesiano presenti in letteratura. In particolare Coles & Tawn (1996a) e Coles & Powell (1996) lo hanno proposto, rispettivamente, per lo studio di dati relativi a precipitazioni e inondazioni e per l'analisi delle velocità estreme dei venti.

I metodi bayesiani sono molto potenti anche in quelle situazioni relative all'inferenza predittiva, cioè situazioni in cui non si è interessati alla stima di particolari parametri o alla verifica di ipotesi, bensì alla previsione di valori futuri che dipendono da  $\theta$ .

## 1.4 Stima delle distribuzioni Extreme Value

In questa sezione riportiamo le metodologie di stima dei parametri delle distribuzioni Extreme Value più impiegate, ovvero la GEV e la GPD.

Iniziamo dai parametri della GEV, ossia il parametro di forma  $\xi$ , e i parametri di scala  $\mu$  e  $\varphi$ . Come al solito, consideriamo un campione causale i.i.d.  $\mathbf{X}$  e sia  $X_{k:n}$  la statistica di ordine  $k$ . Possiamo, dunque, definire le seguenti quantità (si veda Rodrigues et al., 2011),  $\forall j \geq 1$ ,  $j$  intero:

$$L_{k,n}^{(j)} := \frac{1}{k} \sum_{i=1}^k \left\{ 1 - \frac{X_{n-k:n}}{X_{n-i+1:n}} \right\}^j \quad (1.20)$$

$$M_{k,n}^{(j)} := \frac{1}{k} \sum_{i=1}^k \{ \ln X_{n-i+1:n} - \ln X_{n-k:n} \}^j \quad (1.21)$$

Grazie alle (1.20) e (1.21) è possibile definire diversi stimatori per qualunque valore di  $\xi \in \mathfrak{R}$ . Ad esempio

- Stimatore dei momenti (M):

$$\widehat{\xi}_{k,n}^M = M_{k,n}^{(1)} + \frac{1}{2} \left\{ 1 - \left( \frac{M_{k,n}^{(2)}}{[M_{k,n}^{(1)}]^2} - 1 \right)^{-1} \right\}$$

- Stimatore generalizzato di Hill (GH):

$$\widehat{\xi}_{k,n}^{GH} = M_{k,n}^{(1)} + \frac{1}{k} \sum_{i=1}^k \{ \ln M_{i,n}^{(1)} - \ln M_{k,n}^{(1)} \}$$

- Stimatore del momento misto (MM):

$$\widehat{\xi}_{k,n}^{MM} = \frac{\widehat{\phi}_{k,n} - 1}{1 + 2\min(\widehat{\phi}_{k,n} - 1, 0)} \quad \text{dove } \widehat{\phi}_{k,n} := \frac{M_{k,n}^{(1)} - L_{k,n}^{(1)}}{(L_{k,n}^{(1)})^2}$$

- Stimatore di Pickands (P):

Sia

$$U(x) = \inf\{y : F(y) \geq x\} \quad (1.22)$$

la funzione inversa generalizzata della fdr F. Considerato che

$$\lim_{t \rightarrow \infty} \frac{U(2t) - U(t)}{U(t) - U(t/2)} = 2^\xi,$$

applicando il logaritmo ad entrambi i membri, si ottiene lo stimatore:

$$\hat{\xi}^P = \frac{1}{\ln(2)} \lim_{t \rightarrow \infty} \left( \ln \left( \frac{U(2t) - U(t)}{U(t) - U(t/2)} \right) \right)$$

Sotto opportune ipotesi è possibile garantire la normalità asintotica di tutti i suddetti stimatori (si veda Embrechts et al. (1997) per le altre proprietà), nonché degli stimatori di tutti gli altri parametri di cui parliamo nel prossimo paragrafo.

Per quanto riguarda, invece, le stime di massima verosimiglianza è necessario richiamare quanto riportato in Smith (1985):

- quando  $\xi > -0,5$  lo stimatore di massima verosimiglianza è regolare e gode delle usuali proprietà, tra cui la Normalità asintotica e la consistenza. È, dunque, uno stimatore CANE (si veda Mood et al., 1993);
- quando  $-1 < \xi < -0,5$  lo stimatore esiste ma non gode delle usuali proprietà dello stimatore di massima verosimiglianza;
- quando  $\xi < -1$  lo stimatore non esiste in quanto la funzione di verosimiglianza non ha massimo.

Per completezza, riportiamo la funzione di log-verosimiglianza per un campione composto da  $n$  GEV i.i.d. (Beirlant et al., 2005):

$$l(x; \mu, \varphi, \xi) = -n \log \varphi - \left( \frac{1}{\xi} + 1 \right) \sum_{i=1}^n \log \left( 1 + \xi \frac{x_i - \mu}{\varphi} \right) - \sum_{i=1}^n \left( 1 + \xi \frac{x_i - \mu}{\varphi} \right)^{-1/\xi} \quad (1.23)$$

Quando la GEV si riduce ad una Gumbel ( $\xi = 0$ ) la log-verosimiglianza diventa:

$$l(x; \mu, \varphi, 0) = -n \log \varphi - \sum_{i=1}^n \exp \left( -\frac{x_i - \mu}{\varphi} \right) - \sum_{i=1}^n \left( -\frac{x_i - \mu}{\varphi} \right) \quad (1.24)$$

Per stimare il parametro di locazione  $\mu$  è necessario rifarsi alla funzione inversa generalizzata della fdr  $F$  riportata nella (1.22). In particolare, otteniamo il parametro di locazione nel seguente modo:

$$\hat{\mu}_{k,n} = X_{n-k:n} \stackrel{d}{=} \hat{U}(n/k).$$

mentre per il parametro di scala vale

$$\hat{\varphi}_{k,n} = X_{n-k:n} M_{k,n}^{(1)}(1 - \min(0, \hat{\xi}))$$

dove  $M_{k,n}^{(1)}$  è data in (1.21).

Senza riportare la dimostrazione, vediamo che lo stimatore del right-end point della fdr  $F$  risulta

$$\hat{x}_F = \hat{\mu} - \frac{\hat{\varphi}}{\hat{\xi}}$$

Passiamo ora alla stima della GPD: come visto nella (1.12), la distribuzione generalizzata di Pareto è corredata di tre parametri,  $\mu$ ,  $\xi$  e  $\sigma_\mu$ ; tuttavia, il parametro  $\mu$  altro non è che la soglia prescelta  $u$  e quindi non andrà stimato. Iniziamo con il metodo della massima verosimiglianza. In particolare, supponendo di avere un campione composto da  $n$  GPD i.i.d., la funzione da ottimizzare è la seguente:

$$l(x; \mu, \sigma_\mu, \xi) = -n \log \sigma_\mu - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^n \log \left(1 + \xi \frac{x_i - \mu}{\sigma_\mu}\right) \quad (1.25)$$

Per  $\xi = 0$  la (1.25) diventa

$$l(x; \mu, \sigma_\mu, 0) = -n \log \sigma_\mu - \frac{1}{\sigma_\mu} \sum_{i=1}^n x_i - \mu \quad (1.26)$$

Dalla (1.12), tuttavia, si vede che il supporto della GPD dipende dal valore assunto dal parametro incognito  $\xi$  e, pertanto, non essendo verificate le condizioni di regolarità di Cramer, lo stimatore di massima verosimiglianza potrebbe non godere di alcune proprietà. Per questi e altri motivi (si veda

Luceño (2006) per una panoramica più ampia circa le criticità legate alle stime di massima verosimiglianza) è stato proposto, tra gli altri, lo stimatore Likelihood Moment (LM) che sostituisce dignitosamente lo stimatore di massima verosimiglianza quando esso, appunto, non gode di alcune proprietà asintotiche (per esempio, efficienza e Normalità) o, addirittura, non esiste (per  $\xi < -1$ ). Per maggiori dettagli in merito si rimanda a Zhang (2007).

Oltre a massimizzare la (1.25) e allo stimatore LM, esistono altri metodi per ottenere una stima di tali parametri. Uno di questi è noto in letteratura come *metodo dei momenti pesati*. Esso si basa sui quantili  $\omega_r = \frac{\sigma_\mu}{(r+1)(r+1-\xi)}$ , dove  $r = 0, 1$ . Otteniamo le stime dei parametri  $\sigma_\mu$  e  $\xi$  nel seguente modo:

$$\sigma_\mu = \frac{2\omega_0\omega_1}{\omega_0 - 2\omega_1} \text{ ed } \xi = 2 - \frac{\omega_0}{\omega_0 - 2\omega_1}$$

Risulta evidente il carattere iterativo di tale metodo il quale necessita di un valore di partenza sia per  $\xi$  sia per  $\sigma_\mu$ . Ovviamente, è necessario anche imporre un numero massimo di iterazioni oltre ad una tolleranza. Essa potrebbe essere valutata calcolando l'errore relativo tra due successive iterazioni.

Per il valore di partenza del parametro  $\xi$  potremmo utilizzare uno degli stimatori visti nella sezione precedente.

In letteratura viene anche proposto (si veda, per esempio, Engeland et al., 2004) un metodo leggermente diverso da quello appena esposto. Esso è noto come Probability Weighted Moments (PWM) e lo stimatore che si ottiene è considerato tra i più robusti per le stime relative alla GPD. In particolare, la distorsione di questo stimatore cresce al crescere di  $\xi$  e risulta minima per  $\xi < 1$ . Il metodo è caratterizzato da una funzione di penalizzazione che associa alla stima iniziale di  $\xi$  (ottenuta mediante il metodo dei momenti pesati) un peso compreso tra zero e uno: minore è il

valore assunto da  $\xi$  maggiore è il peso ad esso associato. Nello specifico, se  $\xi \leq 0$  il suo peso sarà pari ad 1 mentre se  $\xi \geq 1$  avrà peso nullo. Per  $0 < \xi < 1$  il peso è funzione di  $\xi$ .

La funzione di penalizzazione si ritrova anche nello stimatore Penalized Maximum Likelihood (PML). In tal caso, però, la stima iniziale di  $\xi$  viene ottenuta massimizzando la funzione di verosimiglianza (si veda ancora Engeland et al., 2004). A tale stima si applica la suddetta funzione seguendo il criterio di scelta appena descritto.

Infine, in letteratura esistono diversi stimatori capaci di restituire stime attendibili anche in presenza di campioni contenenti valori anomali. Uno tra questi è il Minimum Density Power Divergence Estimator (MDPDE). Esso è indicizzato da una costante  $\alpha \geq 0$  che consente di giungere ad un compromesso tra robustezza ed efficienza: al crescere di  $\alpha$  cresce la robustezza dello MDPDE ma decresce la sua efficienza. Si rimanda a Juarez & Schucany (2004) per ulteriori dettagli.

Nonostante la disponibilità di tutte queste tecniche di stima, nel caso di studio che presenteremo sono stati preferiti i metodi “classici” dei momenti e della massima verosimiglianza in quanto si è pervenuti con quest’ultimi a risultati non peggiori di quelli ottenuti applicando le tecniche sopra descritte.

## 1.5 Distribuzione Kumaraswamy e generalizzazioni

In tale sezione descriviamo la distribuzione Kumaraswamy ed una generalizzazione della distribuzione Gumbel (Cordeiro et al., 2012). In particolare, sia

$$K(x; \alpha, \beta) = 1 - (1 - x^\alpha)^\beta, x \in (0, 1) \quad (1.27)$$

la distribuzione proposta in Kumaraswamy (1980), dove i parametri  $\alpha$  e  $\beta$  definiscono la sua forma. Se  $G$  è la distribuzione di una v.a., allora possiamo definire una nuova funzione di ripartizione dalla seguente:

$$F(x; a, b) = 1 - (1 - G(x)^a)^b \quad (1.28)$$

dove  $a > 0$  e  $b > 0$  sono parametri di forma e la loro entità dipende da curtosi e asimmetria. La distribuzione Kumaraswamy-Gumbel (Kum-Gumbel) è definita grazie alla (1.28) ed ha la seguente espressione (si veda Cordeiro et al., 2012):

$$F_{KG}(x; a, b) = 1 - (1 - \Lambda(x)^a)^b \quad (1.29)$$

dove  $\Lambda(x)$  è la distribuzione Gumbel. La corrispondente funzione quantile è ottenuta invertendo la (1.29) ed esplicitando i parametri della Gumbel ( $u$  and  $\phi$ ):

$$x_p = F^{-1}(p) = u - \phi \log \left[ -\log \left( 1 - (1 - p)^{1/b} \right)^{1/a} \right] \quad (1.30)$$

con  $p \in (0, 1)$ .

La tabella e la figura che seguono mostrano curtosi, asimmetria e densità della Kum-Gumbel al variare dei suoi quattro parametri:

$u$	$\phi$	a	b	Curtosi	Asimmetria
0	5	1	1	5.4	1.1
0	1	0.5	0.5	7.1	1.6
-5	3	2	3	3.6	0.5
1	10	5	0.7	6.4	1.4
0	15	1	0.4	7.6	1.7

Tabella 1.1: Curtosi e Asimmetria della distribuzione Kum-Gumbel.

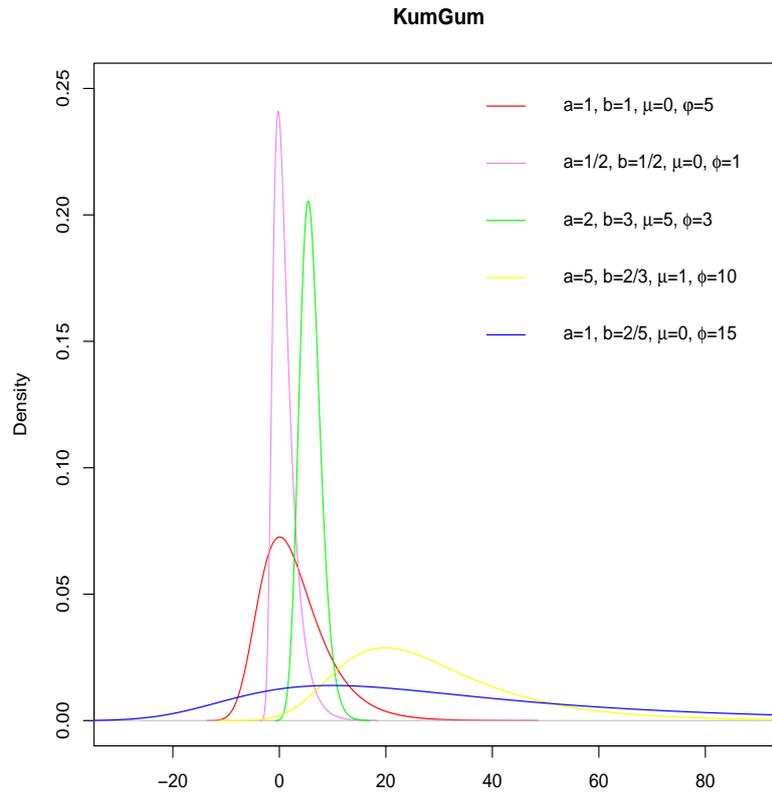


Figura 1.2: Densità della KumGum.

Un'altra generalizzazione della distribuzione Kum è la Kumaraswamy-Pareto (Kum-Pareto); in particolare, è definita sempre dalla (1.28) facendo riferimento, questa volta, alla distribuzione di Pareto  $P$  che riportiamo qui di seguito:

$$P(x; \beta, \kappa) = 1 - \left(\frac{\beta}{x}\right)^\kappa \quad (1.31)$$

per  $x \geq \beta$ , dove  $\beta > 0$  è un parametro di scala e  $\kappa \geq 0$  è quello di forma. Pertanto dalle (1.27), (1.28) e (1.31) possiamo facilmente ottenere la Kum-Pareto:

$$F_{KP}(x; \beta, \kappa, a, b) = 1 - \left\{1 - \left[1 - \left(\frac{\beta}{x}\right)^\kappa\right]^a\right\}^b \quad (1.32)$$

la cui corrispondente funzione quantile risulta

$$F^{-1}(p) = \beta \left\{ \left\{ 1 - \left[ 1 - (1-p)^{1/b} \right]^{1/a} \right\}^{1/\kappa} \right\}^{-1} \quad (1.33)$$

dove  $p \in (0, 1)$ .

Nella figura seguente è riportata la densità della Kum-Pareto al variare dei suoi parametri:

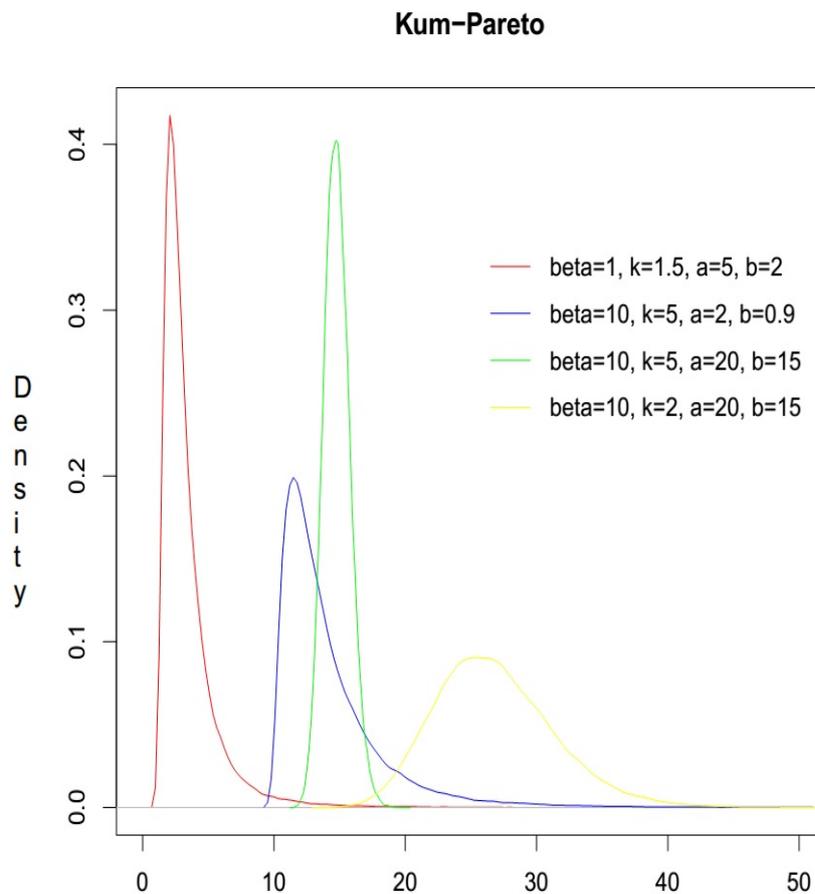


Figura 1.3: Densità della Kum-Pareto.

Per concludere la sezione riportiamo le funzioni di log-verosimiglianza delle distribuzioni generalizzate di Kumaraswamy. Per la Kum-Gumbel, sia  $\boldsymbol{\theta}_{KG} = (a, b, \mu, \varphi)^T$  il vettore dei parametri; dalla funzione di densità,

ottenuta derivando la (1.29), possiamo giungere alla seguente:

$$\begin{aligned}
 l(x; \boldsymbol{\theta}_{KG}) &= n \log(ab) - n \log(\varphi) - \sum_{i=1}^n \left( \frac{x_i - \mu}{\varphi} \right) - a \sum_{i=1}^n \exp \left( -\frac{x_i - \mu}{\varphi} \right) \\
 &\quad + (b-1) \sum_{i=1}^n \left\{ 1 - \exp \left( -a \exp \left( -\frac{x_i - \mu}{\varphi} \right) \right) \right\}
 \end{aligned} \tag{1.34}$$

Per quel che riguarda, invece, la Kum-Pareto, fissando il vettore dei parametri  $\boldsymbol{\theta}_{KP} = (a, b, \beta, \kappa)^T$ , avremo:

$$\begin{aligned}
 l(x; \boldsymbol{\theta}_{KP}) &= n(\kappa \log(\beta) + \log(\kappa) + \log(a) + \log(b)) - (\kappa + 1) \sum_{i=1}^n \log(x_i) \\
 &\quad + (a-1) \sum_{i=1}^n \log \left[ 1 - \left( \frac{\beta}{x_i} \right)^\kappa \right] + (b-1) \sum_{i=1}^n \log \left\{ 1 - \left[ 1 - \left( \frac{\beta}{x_i} \right)^\kappa \right]^a \right\}
 \end{aligned} \tag{1.35}$$

## 1.6 Modelli composti

In diverse situazioni capita di dover trattare un insieme di osservazioni che non possono essere ricondotte ad un'unica distribuzione di probabilità. È possibile avere, per esempio, osservazioni di medio-bassa entità ma con alte frequenze e osservazioni di entità maggiori, o addirittura estreme, ma che si verificano solo occasionalmente. Pertanto, la corrispondente distribuzione potrebbe essere costruita non come fatto in precedenza, facendo riferimento ad un unico modello, bensì come una combinazione di due differenti modelli: uno caratterizzato da una coda poco pesante e da impiegare per modellizzare i valori entro un certo livello limite; l'altro a coda più pesante e da trattare come modello per i valori oltre tale livello limite. Distribuzioni di questo tipo sono state proposte, inizialmente, da Cooray & Ananda (2005), da Scollnik (2007) e più recentemente da Teodorescu & Vernic (2013). Formalizzando la logica appena esposta, avremo una funzione di densità del

tipo

$$f(x) = \begin{cases} cf_1(x), & -\infty < x \leq \theta \\ cf_2(x), & \theta < x < \infty \end{cases} \quad (1.36)$$

dove  $f_1$  e  $f_2$  sono funzioni di densità e  $c$  è una costante ottenuta imponendo determinate condizioni di continuità e differenziabilità ad  $f$ . Appare ovvia l'importanza del valore limite  $\theta$ , dato che al suo variare cambia l'intero modello. Tale valore limite altro non è che la soglia  $u$  già trattata in precedenza; pertanto, tutte le considerazioni fatte per  $u$  valgono, senza alcuna eccezione, per  $\theta$ . Nel seguito, per riferirci alla soglia utilizzeremo, indistintamente,  $u$  o  $\theta$ .

In Cooray & Ananda (2005),  $f_1$  è la densità Lognormale mentre  $f_2$  è quella di Pareto (non generalizzata). Basandosi sulle stesse distribuzioni, Scollnik (2007) propose due modelli composti più generali rispetto a quelli di Cooray & Ananda (2005). Similmente, Teodorescu & Vernic (2009) proposero un modello Esponenziale-Pareto, mentre Preda & Ciumara (2006) hanno realizzato uno studio comparativo tra un composto Weibull-Pareto e uno Lognormale-Pareto. Nel seguito definiremo un modello composto generalizzato in modo da poter adottare, di volta in volta, senza troppi sforzi, le densità  $f_1$  ed  $f_2$  che preferiamo. Formalmente, la densità del modello composto è scritta come

$$f(x) = \begin{cases} rf_1^*(x), & -\infty < x \leq \theta \\ (1-r)f_2^*(x), & \theta < x < \infty \end{cases} \quad (1.37)$$

dove  $r \in [0, 1]$  mentre  $f_1^*$  ed  $f_2^*$  sono le densità troncate di  $f_1$  ed  $f_2$ , rispettivamente. Nello specifico, se  $F_i$  è la funzione di ripartizione di  $f_i$ ,  $i = 1, 2$ ,

avremo

$$\begin{cases} f_1^*(x) = \frac{f_1(x)}{F_1(\theta)}, & -\infty < x \leq \theta \\ f_2^*(x) = \frac{f_2(x)}{1-F_2\theta}, & \theta < x < \infty \end{cases} \quad (1.38)$$

È facile notare che la (1.37) è una combinazione convessa di  $f_1^*$  ed  $f_2^*$  con pesi  $r$  ed  $1-r$ . Inoltre, vogliamo che la (1.37) sia una funzione continua, derivabile e con derivata continua e per questo imponiamo due restrizioni:

$$\begin{cases} \lim_{x \rightarrow \theta} f(x) = f(\theta) \\ \lim_{x \rightarrow \theta^-} f'(x) = \lim_{x \rightarrow \theta^+} f'(x) \end{cases} \quad (1.39)$$

Dalla prima restrizione abbiamo

$$r = \frac{f_2(\theta)F_1(\theta)}{f_2(\theta)F_1(\theta) + f_1(\theta)(1-F_2(\theta))} \quad (1.40)$$

mentre dalla seconda

$$r = \frac{f_2'(\theta)F_1(\theta)}{f_2'(\theta)F_1(\theta) + f_1'(\theta)(1-F_2(\theta))} \quad (1.41)$$

La quantità  $r$  è da interpretare come il peso che associamo alla componente  $f_1$ , ovvero più è grande  $r$  maggiore sarà il numero di valori che seguiranno la distribuzione  $f_1$  e minore sarà il numero di valori che seguiranno la  $f_2$ .

A questo punto definiamo la funzione di ripartizione  $F$  della (1.37):

$$F(x) = \begin{cases} r \frac{F_1(x)}{F_1(\theta)}, & -\infty < x \leq \theta \\ r + (1-r) \frac{F_2(x)-F_2(\theta)}{1-F_2(\theta)}, & \theta < x < \infty \end{cases} \quad (1.42)$$

Supponendo che  $F_1$  ed  $F_2$  ammettano inverse, possiamo definire la funzione quantile della (1.37) attraverso il metodo dell'inversione; sia  $p \sim U(0, 1)$ ,

per simulare dalla (1.37) impieghiamo la seguente

$$F^{-1}(x) = \begin{cases} F_1^{-1}\left(\frac{p}{r}F_1(\theta)\right), & \text{se } p \leq r \\ F_2^{-1}\left(\frac{p-r+(1-p)F_2(\theta)}{1-r}\right), & \text{se } p > r \end{cases} \quad (1.43)$$

Per quanto riguarda la stima dei parametri della (1.42) procediamo prima stimando separatamente i parametri delle due distribuzioni che definiscono il modello composto (l'ipotesi implicita è che non ci sia alcuna relazione tra i sinistri "normali" e quelli gravi); tali stime sono da intendere come valori di partenza delle funzioni che permettono di massimizzare la seguente verosimiglianza (Teodorescu & Vernic, 2013)

$$L(x_1, \dots, x_n; \boldsymbol{\theta}) = r^m (1-r)^{n-m} \prod_{i=1}^m f_1^*(x_i) \prod_{j=m+1}^n f_2^*(x_j) \quad (1.44)$$

dove, come al solito,  $n$  è la dimensione campionaria,  $\boldsymbol{\theta}$  è il vettore contenente i parametri del modello composto mentre  $m$  è t.c.  $X_m \leq u < X_{m+1}$ , ovvero è il livello della statistica d'ordine immediatamente precedente (o coincidente) ad  $u$ .

## 1.7 Modelli composti con soglia aleatoria

Descriviamo in tale sezione la metodologia proposta in Pigeon & Denuit (2011), già citata in precedenza. Tale lavoro ha un raggio molto ampio in quanto non solo troviamo un approccio completamente nuovo per la trattazione della soglia  $u$  ma viene anche descritta un'alternativa alle metodologie Block Maxima e POT; in particolare, viene rivisitato il modello composto Lognormale-Pareto, proposto originariamente da Cooray & Ananda (2005) e successivamente modificato in Scollnik (2007).

Pigeon e Denuit ritengono che la soglia debba essere trattata come una

vera e propria variabile aleatoria con tanto di distribuzione di probabilità. Indichiamo, come al solito, con  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  il campione casuale; possiamo assumere che ciascuna osservazione proveniente da esso sia dotata di una propria soglia. Quindi, la generica osservazione  $x_i$  ha soglia  $u_i$ , per  $i = 1, \dots, n$ . In particolare, avremo che  $u_1, \dots, u_n$  sono le realizzazioni di una v.a.  $U$  con funzione di ripartizione  $G$ .

Ora, quali distribuzioni, tra quelle note, sono adatte per descrivere la soglia  $u$ ? Nel contesto in cui ci troviamo, siamo interessati alla coda destra della distribuzione dei costi e la soglia stessa è un costo; pertanto, è necessario che la v.a.  $U$  sia non negativa e abbia una coda pesante. Si potrebbe pensare, pertanto, ad una distribuzione Gamma o ad una Lognormale, solo per citare le più note.

Ad ogni modo, ciò che è importante cogliere è che, con tale approccio, una volta definita la distribuzione  $G$ , basterà inglobarla nella definizione del modello composto. La costruzione di un siffatto modello manifesta un aspetto completamente nuovo e innovativo; infatti, non dovendo più scegliere un preciso ed unico valore per  $u$  ma dovendo, piuttosto, definirne l'intera distribuzione, i parametri di quest'ultima vengono inglobati nella definizione dello stesso modello composto. Per esempio, partendo dalla (4.7) e dalla derivata di (1.12), definiamo la funzione di densità del modello composto Lognormale-GPD con soglia aleatoria nel modo seguente:

$$f(x) = (1 - r) \int_0^x f_2(x)g(u)du + r \int_x^\infty \frac{1}{\Phi(\xi\sigma)} f_1(x)g(u)du \quad (1.45)$$

dove  $r \in [0, 1]$ ,  $u$  è la soglia variabile la cui densità è espressa dalla  $g$ ,  $f_1$  ed  $f_2$  sono, rispettivamente, le densità della Lognormale e della GPD,  $\Phi$  è la funzione di ripartizione di una Normale standard,  $\xi$  è il parametro di forma

della GPD e  $\sigma$  è il parametro di scala della Lognormale.

Supponendo che la densità  $g$  della soglia variabile sia la (4.5) e sostituendola nella (1.45), dopo alcuni passaggi avremo

$$f(x) = -\frac{\beta^\alpha(1-r)}{\sigma_\mu\Gamma(\alpha)}\left(1+\xi\frac{x-\mu}{\sigma_\mu}\right)^{-\frac{1+\xi}{\xi}}\int_0^x u^{\alpha-1}e^{-\frac{u}{\beta}}du + \frac{r\beta^\alpha}{x\sigma\sqrt{2\pi}\Phi(\xi\sigma)\Gamma(\alpha)}\exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\mu_1}{\sigma}\right)^2\right\}\int_x^\infty u^{\alpha-1}e^{-\frac{u}{\beta}}du \quad (1.46)$$

dove, per non fare confusione, con  $\mu_1$  abbiamo indicato il parametro di locazione della Lognormale, mentre con  $\mu$  quello della GPD. Ricordiamo che, di fatto,  $\mu$  altro non è che la soglia  $u$  (o  $\theta$ ) e che è dotata di una sua densità  $g$ . Svolgendo gli integrali, l'equazione suddetta diventa

$$f(x) = \frac{\beta^\alpha(1-r)}{\sigma_\mu\Gamma(\alpha)}\left(1+\xi\frac{x-\mu}{\sigma_\mu}\right)^{-\frac{1+\xi}{\xi}}\beta^\alpha\left[\gamma\left(\alpha,\frac{x}{\beta}\right)-\gamma(\alpha,0)\right] + \frac{r\beta^\alpha}{x\sigma\sqrt{2\pi}\Phi(\xi\sigma)\Gamma(\alpha)}\exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\mu_1}{\sigma}\right)^2\right\}\beta^\alpha\left[\gamma(\alpha,\infty)-\gamma\left(\alpha,\frac{x}{\beta}\right)\right] \quad (1.47)$$

dove  $\gamma$  è la funzione Gamma incompleta inferiore. Per definizione  $\gamma(\alpha,0) = 0$ , mentre  $\gamma(\alpha,\infty) = \Gamma(\alpha)$ . Pertanto, avremo

$$f(x) = \frac{\beta^{2\alpha}(1-r)\gamma\left(\alpha,\frac{x}{\beta}\right)}{\sigma_\mu\Gamma(\alpha)}\left(1+\xi\frac{x-\mu}{\sigma_\mu}\right)^{-\frac{1+\xi}{\xi}} + \frac{r\beta^{2\alpha}\left[\Gamma(\alpha)-\gamma\left(\alpha,\frac{x}{\beta}\right)\right]}{x\sigma\sqrt{2\pi}\Phi(\xi\sigma)\Gamma(\alpha)}\exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\mu_1}{\sigma}\right)^2\right\} \quad (1.48)$$

Supponendo di avere un campione casuale i.i.d, la log-verosimiglianza della (1.48) risulta

$$\begin{aligned}
l(x; \mu_1, \sigma, \mu, \sigma_\mu, \xi, \alpha, \beta) = & n \left[ \ln \left( \frac{1-r}{r} \right) - \ln \sigma_\mu - \ln \Gamma(\alpha) + \ln \sigma + \right. \\
& \left. \frac{1}{2} \ln(2\pi) + \ln \Phi(\xi\sigma) \right] + 2 \sum_{i=1}^n \ln \gamma(\alpha, x_i/\beta) - \\
& \frac{1+\xi}{\xi} \sum_{i=1}^n \ln \left( 1 + \xi \frac{x_i - \mu}{\sigma_\mu} \right) + \sum_{i=1}^n \ln x_i + \\
& \frac{1}{2} \sum_{i=1}^n \left( \frac{\ln x_i - \mu_1}{\sigma} \right)^2
\end{aligned} \tag{1.49}$$

Massimizzare analiticamente la verosimiglianza (1.49) è pressoché impossibile mentre vi sono diverse possibilità in termini numerici; in particolare, utilizzando il pacchetto `evmix` di R<sup>2</sup> (si veda Hu & Scarrott, 2013), possiamo ottenere quei parametri che massimizzano la cosiddetta *verosimiglianza profilo*. Tale funzione può essere vista come una pseudo verosimiglianza, più facile da trattare numericamente ma che condivide con la verosimiglianza, e quindi anche con la (1.49), lo stesso punto di ottimo.

Considerato che i metodi classici per la scelta della soglia non consentono di identificare quest'ultima in maniera univoca ed obiettiva, il metodo appena esposto costituisce, di fatto, una possibile soluzione a tale problema. Inglobare la determinazione del valore spartiacque nella definizione di un modello composto, infatti, non costringe il ricercatore a dover interpretare risultanze grafiche o a dover difendere ipotesi obiettivamente poco sostenibili.

---

<sup>2</sup>Per maggiori dettagli sull'ambiente R si rimanda a R Development Core Team (2014), mentre per i package più diffusi e maggiormente sviluppati per il calcolo attuariale si veda Charpentier, A. (2014).

## Capitolo 2

# Funzioni Copula e Fast Fourier Transform

Lo studio di fenomeni multivariati e la valutazione congiunta dei rischi presuppongono una corretta stima delle interazioni esistenti tra le diverse componenti aleatorie; in alcuni ambiti applicativi, tra cui quello attuariale, la natura intrinseca dei fenomeni in esame è tale da rendere inadeguato il ricorso ad una struttura di dipendenza lineare e si rende necessario studiare una qualche altra forma di dipendenza. Un modo per farlo è quello di ricorrere alle funzioni Copula. Tali funzioni, consentendo di creare una corrispondenza diretta tra la funzione di ripartizione di una distribuzione multivariata e le funzioni di ripartizione delle marginali univariate, sono state introdotte da Sklar nel 1959, anche se sono state approfondite solo negli ultimi anni e impiegate per descrivere le dipendenze non parametriche (e non necessariamente lineari) tra coppie, terne, ..., n-uple di variabili aleatorie. Ogni funzione di ripartizione congiunta racchiude in se informazioni sia sul comportamento della singola distribuzione marginale sia sulla struttura di dipendenza che lega le diverse v.a. In questo capitolo riporteremo alcune proprietà riguardanti la distribuzione di un vettore di v.a. e vedremo come,

di fatto, la Copula altro non è che una particolare distribuzione multivariata, le cui marginali sono delle v.a. uniformi. Ci soffermeremo per lo più sulle copule bidimensionali e su quelle che vengono definite Meta Distributions (si veda McNeil et al., 2005).

## 2.1 Distribuzioni multivariate

Sia  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vettore di  $n$  v.a. e sia  $\mathbf{x} = (x_1, \dots, x_n)$  un vettore contenente un'osservazione per ciascuna  $X_i, i = 1, \dots, n$ . La funzione di ripartizione congiunta di tali v.a. è definita come:

$$F(\mathbf{x}) = Pr[X_1 \leq x_1, \dots, X_n \leq x_n] \quad (2.1)$$

mentre la funzione di sopravvivenza corrispondente ad  $F$  risulta:

$$\bar{F}(\mathbf{x}) = Pr[X_1 > x_1, \dots, X_n > x_n] = 1 - F(x_1, \dots, x_n) \quad (2.2)$$

Vediamo alcune proprietà. Sia  $F : \mathfrak{R}^n \rightarrow [0, 1]$  una funzione continua a destra; essa è una funzione di ripartizione congiunta se e solo se soddisfa le seguenti:

- $F(\mathbf{x}) = F(x_1, \dots, x_n)$  è monotona crescente in ogni componente del vettore  $\mathbf{x}$ ;
- $\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_i, \dots, x_n) = 0, i = 1, \dots, n$ ;
- $\lim_{x_i \rightarrow \infty} F(x_1, \dots, x_i, \dots, x_n) = 1, i = 1, \dots, n$
- se  $n = 2$ , deve essere  $Pr[a < X_1 \leq b, c < X_2 \leq d] = F(b, d) + F(a, c) - F(a, d) - F(b, c) \geq 0$  con  $a \leq c, b \leq d$ .

Supponiamo che  $F(\mathbf{x})$  sia proprio una distribuzione di probabilità congiunta le cui marginali  $F_1, \dots, F_n$  sono v.a. univariate; è possibile deter-

minare i cosiddetti limiti di Fréchet-Hoeffding i quali, appunto, limitano superiormente ( $F_U(\mathbf{x})$ ) e inferiormente ( $F_L(\mathbf{x})$ ) la  $F$ . Essi sono definiti nel seguente modo:

$$\begin{cases} F_U(\mathbf{x}) = \min(F_1(x), \dots, F_n(x)) = M \\ F_L(\mathbf{x}) = \max\left(\sum_{i=1}^n F_i(x) - n + 1, 0\right) = W \end{cases} \quad (2.3)$$

Avremo, dunque,  $W \leq F(\mathbf{x}) \leq M$ .

## 2.2 Funzione Copula

Come detto in precedenza, la Copula è una distribuzione multivariata con marginali uniformi nell'intervallo  $[0, 1]$ . Formalmente, sia  $\mathbf{u} = (u_1, \dots, u_n)$  un vettore la cui  $i$ -esima componente  $u_i \in [0, 1]$  e sia  $C : [0, 1]^n \rightarrow [0, 1]$ . Allora  $C(\mathbf{u})$  è una Copula se e solo se sono soddisfatte le seguenti condizioni (si veda McNeil et al., 2005):

- $C(\mathbf{u}) = C(u_1, \dots, u_n)$  è crescente rispetto ad ogni  $u_i$ ;
- $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$  per ogni  $i = 1, \dots, n, u_i \in [0, 1]$ ;
- $Pr[a_1 \leq u_1 \leq b_1, \dots, a_n \leq u_n \leq b_n] \geq 0$

L'impiego delle copule nello studio delle distribuzioni multivariate è giustificato dal seguente teorema dovuto a Sklar (1959):

**Teorema 2.1 (Sklar)** *Sia  $F$  una funzione di ripartizione congiunta con marginali  $F_1, \dots, F_n$ . Allora esiste una Copula  $C : [0, 1]^n \rightarrow [0, 1]$  tale che*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)), \forall x_1, \dots, x_n \in \mathfrak{R} \quad (2.4)$$

*Tale Copula, inoltre, è unica se le v.a. sono continue.*

Il teorema appena enunciato, la cui dimostrazione si può trovare in McNeil et al. (2005), ci informa di due aspetti: innanzitutto, capiamo che da una qualsiasi distribuzione multivariata è possibile pervenire ad una Copula; inoltre, a partire dalle distribuzioni marginali  $F_1, \dots, F_n$  ed utilizzando una Copula, si può determinare la distribuzione congiunta  $F$ . Quest'ultima proprietà ci consentirà di individuare la funzione di ripartizione del costo sinistri aggregato, avendo a disposizione la distribuzione del costo sinistri per ogni singola garanzia.

Indicando con  $F_i^{\leftarrow}(u_i)$  la funzione quantile della  $i$ -esima v.a. univariata, è possibile riscrivere la (2.4) come:

$$C(u_1, \dots, u_n) = F(F_1^{\leftarrow}(u_1), \dots, F_n^{\leftarrow}(u_n)) \quad (2.5)$$

da cui capiamo che, effettivamente, la Copula  $C$  è la distribuzione congiunta di v.a.  $U(0, 1)$ . Pertanto, per essa valgono tutte le proprietà delle distribuzioni multivariate viste in precedenza. In particolare, per qualunque Copula  $C$ , esistono i limiti di Fréchet-Hoeffding:

$$\max\left(\sum_{i=1}^n u_i - n + 1, 0\right) \leq C(\mathbf{u}) \leq \min(u_1, \dots, u_n) \quad (2.6)$$

e vale anche la seguente:

**Proposizione 2.2 (Invarianza della Copula)** *Sia  $\mathbf{X}$  un vettore di v.a. univariate e continue e sia  $C$  la Copula associata ad esso. Siano  $T_1, \dots, T_n$  funzioni strettamente crescenti. Allora, il vettore  $(T_1, \dots, T_n)$  ammette  $C$  come Copula.*

Grazie al Teorema 2.1 ed alla Proposizione 2.2, possiamo utilizzare la Copula per descrivere la struttura di dipendenza che sussiste tra le v.a. univariate  $X_1, \dots, X_n$ . Per la precisione, la Copula  $C$  che viene impiegata non

dipende solo ed esclusivamente dalle distribuzioni marginali, dalle quali è possibile avere delle indicazioni sulla forma funzionale della Copula, ma anche da un vettore di parametri sconosciuti  $\theta = (\theta_1, \dots, \theta_p)$  la cui stima fornisce una misura della dipendenza tra le v.a. e determina in maniera univoca la Copula  $C$ . Pertanto, possiamo riscrivere la tesi del teorema di Sklar come:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n); \theta) \quad (2.7)$$

È necessario dire che la funzione Copula può essere determinata anche quando le distribuzioni marginali appartengono a diverse famiglie di distribuzioni.

Il teorema di Sklar può essere visto anche in altri termini. Supponiamo di avere funzione di sopravvivenza  $S$  per il vettore  $\mathbf{X}$ , distribuzioni marginali  $F_1, \dots, F_n$  e funzioni di sopravvivenza marginali  $S_1, \dots, S_n$ . Allora avremo

$$S(x_1, \dots, x_n) = \hat{C}(S_{x_1}, \dots, S_{x_n}) \quad (2.8)$$

dove  $\hat{C} \neq C$  è detta survival Copula di  $C$ .

La seguente definizione potrebbe tornare utile:

**Definizione 2.1 (Simmetria radiale)** *Un vettore casuale  $\mathbf{X}$  è detto radialmente simmetrico rispetto ad  $\mathbf{a}$  se  $\mathbf{X} - \mathbf{a} = \mathbf{a} - \mathbf{X}$ .*

In particolare, diremo che una Copula  $C$  è simmetrica radialmente se coincide con la sua survival Copula, ovvero se  $C = \hat{C}$ .

Dato che una Copula è di per sé una distribuzione multivariata, è possibile individuarne la funzione di sopravvivenza ( $\bar{C}$ ), che non deve essere confusa con la survival Copula. In particolare, se  $\mathbf{U}$  ha come distribuzione

la Copula  $C$ , allora la funzione di sopravvivenza di  $C$  è definita come:

$$\begin{aligned}\bar{C}(\mathbf{u}) &= P(U_1 > u_1, \dots, U_d > u_d) = P(1 - U_1 \leq 1 - u_1, \dots, 1 - U_d \leq 1 - u_d) \\ &= \hat{C}(1 - u_1, \dots, 1 - u_d)\end{aligned}\tag{2.9}$$

Nel caso bivariato, è possibile definire una semplice relazione tra la funzione di sopravvivenza  $\bar{C}$  e la Copula stessa; infatti, avremo:

$$\bar{C}(u_1, u_2) = 1 - u_1 - u_2 + C(u_1, u_2)\tag{2.10}$$

Spesso è utile far riferimento alla distribuzione condizionata di una Copula; soffermandoci solo sul caso bivariato, per  $\mathbf{U} = (U_1, U_2)$  avremo

$$\begin{aligned}C_{U_2|U_1}(u_2|u_1) &= P(U_2 \leq u_2 | U_1 = u_1) = \lim_{\delta \rightarrow 0} \frac{C(u_1 + \delta, u_2) - C(u_1, u_2)}{\delta} \\ &= \frac{\partial}{\partial u_1} C(u_1, u_2)\end{aligned}\tag{2.11}$$

Un'altra importante definizione è quella di scambiabilità:

**Definizione 2.2 (Scambiabilità)** *Un vettore casuale  $\mathbf{X}$  si dice scambiabile se  $(X_1, \dots, X_d) \stackrel{d}{=} (X_{\Pi(1)}, \dots, X_{\Pi(d)})$*

per una qualsiasi permutazione  $(\Pi_1, \dots, \Pi_d)$  di  $(1, \dots, d)$ . Una Copula è detta scambiabile se è la distribuzione di un vettore scambiabile.

### 2.2.1 Alcuni esempi

Come in McNeil et al. (2005), classifichiamo le copule in tre macro categorie: *fondamentali, implicite ed esplicite*. Le prime racchiudono i tre casi estremi, ovvero indipendenza, perfetta dipendenza positiva e perfetta

dipendenza negativa. La Copula *indipendenza* risulta

$$C(u_1, \dots, u_n) = \prod_{i=1}^n u_i \quad (2.12)$$

Le v.a. univariate sono, dunque, indipendenti se e solo se la funzione di ripartizione congiunta si fattorizza nel prodotto delle funzioni di ripartizione marginali, ossia se la Copula che descrive il legame tra le variabili è la suddetta (2.12).

Nel caso di perfetta dipendenza positiva, invece, abbiamo la Copula *comonotonicità* che coincide con il limite superiore di Fréchet:

$$C(u_1, \dots, u_n) = \min(u_1, \dots, u_n) \quad (2.13)$$

Invece, in caso di dipendenza perfetta ma negativa, se  $n = 2$  avremo la Copula *contromonotonicità*:

$$C(u_1, u_2) = \max(u_1 + u_2 - 1, 0) \quad (2.14)$$

la quale, a differenza della precedente, non esiste per  $n > 2$ . Se poniamo  $n = 2$ , la Copula appena descritta coincide con il limite inferiore di Fréchet (2.3). Si dimostra (come fatto in McNeil et al. (2005), esempio 5.21) che tale limite inferiore, a differenza di quello superiore, non è una Copula per  $n > 2$  in quanto non soddisfa la terza proprietà di una Copula.

Le copule implicite, invece, possono essere estratte da una qualsiasi distribuzione congiunta con marginali continue. Per esempio, la Copula Gaussiana coincide con la funzione di ripartizione congiunta di una Normale standard d-dimensionale  $\mathbf{N}(\mathbf{0}, \Sigma)$ . In particolare, avremo

$$C_{\Sigma}^{Ga}(\mathbf{u}) = \phi_{\Sigma}(\phi^{-1}(u_1), \dots, \phi^{-1}(u_d)) \quad (2.15)$$

Se  $\Sigma$  è la matrice identità, la Copula gaussiana coincide con la Copula indipendenza. Se, invece, la matrice delle correlazioni è una matrice unitaria, allora avremo la Copula comonotonicità, mentre se  $d = 2$  e il coefficiente di correlazione è  $-1$ , allora avremo la Copula contromonotonicità. Pertanto, nella bidimensionalità la Copula gaussiana esprime una forma di dipendenza intermedia tra quella perfetta positiva e quella perfetta negativa.

Un'altra Copula implicita è data dalla funzione di ripartizione di una t-Student standardizzata d-dimensionale.

A differenza delle copule implicite, quelle esplicite hanno spesso una forma funzionale chiusa e per questo più semplici da trattare analiticamente. Tra le più famose copule esplicite troviamo la Copula di Gumbel, definita come

$$C_{\theta}^{Gu}(u_1, u_2) = \exp\{-[\ln(u_1)^{-\theta} + \ln(u_2)^{-\theta}]^{1/\theta}\}, 1 \leq \theta < \infty \quad (2.16)$$

Essa coincide con la (2.12) se  $\theta = 1$ , con la (2.13) se  $\theta \rightarrow \infty$ . Per questo motivo, è adatta solo per esprimere una forma positiva di dipendenza, ovvero descrive un caso intermedio tra la completa indipendenza e la perfetta dipendenza positiva.

Un altro esempio di Copula esplicita è quella di Clayton, definita come

$$C_{\theta}^{Cl}(u_1, u_2) = \max\{u_1^{-\theta} + u_2^{-\theta} - 1; 0\}^{-1/\theta}, \theta \geq -1, \theta \neq 0 \quad (2.17)$$

Sia la Copula di Gumbel sia quella di Clayton appartengono alla famiglia delle copule archimedee. Tale famiglia comprende altre copule tra cui la Copula di Frank, quella di Joe e quella di Ali-Mikhail-Haq, riportate di seguito:

$$C_{\theta}^{Fr}(u_1, u_2) = -\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right), \theta \in \mathfrak{R} \setminus \{0\} \quad (2.18)$$

$$C_{\theta}^{Joe}(u_1, u_2) = 1 - \left( (1-u_1)^{\theta} + (1-u_2)^{\theta} - (1-u_1)^{\theta}(1-u_2)^{\theta} \right)^{1/\theta}, \theta \geq 1 \quad (2.19)$$

$$C_{\theta}^{AMH}(u_1, u_2) = \frac{u_1 u_2}{1 - \theta(1-u_1)(1-u_2)}, -1 \leq \theta < 1 \quad (2.20)$$

La famiglia delle copule archimedee assume particolare importanza nella pratica in quanto semplici da costruire. Sono ottenibili, infatti, attraverso un'unica funzione unidimensionale caratterizzata da alcune proprietà. Nel caso bivariato, la Copula archimedeica può scindere la funzione di ripartizione congiunta  $F$  delle due marginali attraverso l'utilizzo di una funzione  $\lambda(t) > 0$  con  $t \in (0, 1)$ :  $\lambda(F(x_1, x_2)) = \lambda(F_1(x_1))\lambda(F_2(x_2))$ . Ponendo  $\phi(t) = -\ln(\lambda(t))$ , è possibile scrivere  $\phi(F(x_1, x_2)) = \phi(F_1(x_1)) + \phi(F_2(x_2))$  mentre, in termini di copule, avremo  $\phi(C(u, v)) = \phi(u) + \phi(v)$ .

Sklar e Schweizer dimostrano che se  $C(u, v) = \phi^{-1}(\phi(u) + \phi(v))$ , allora essa è una Copula archimedeica. Pertanto, la Copula archimedeica risulta essere generata dalla funzione  $\phi$ , detta, appunto, generatore, e dalla sua pseudo-inversa  $\phi^{-1}$ ; si può dimostrare che la convessità della funzione generatrice è condizione sufficiente affinché  $C$  sia una Copula.

Di seguito riportiamo alcune proprietà della Copula archimedeica  $C$ :

- è simmetrica:  $C(u, v) = C(v, u) \quad \forall u, v \in (0, 1)$ ;
- è associativa:  $C(C(u, v), w) = C(u, C(v, w)) \quad \forall u, v \in (0, 1)$ ;
- se  $\phi$  è una funzione generatrice per  $C$ , allora anche  $c\phi$  lo è,  $\forall c > 0$ ;
- $C$  è differenziabile almeno due volte ed esiste una funzione integrabile  $f$  tale che:  $f(v) \frac{dC(u, v)}{dv} = f(u) \frac{dC(u, v)}{du}$ .

Infine, un'altra Copula molto diffusa, ma non archimedea, è quella di Plackett:

$$C(u_1, u_2) = \begin{cases} \frac{L - \sqrt{L^2 - 4\theta(\theta - 1)u_1u_2}}{2(\theta - 1)}, & \theta > 0 \\ u_1u_2, & \theta = 1 \end{cases} \quad (2.21)$$

dove  $L = 1 + (\theta - 1)(u_1 + u_2)$ . Così come la Copula di Frank, anche quella di Plackett è in grado di descrivere diverse tipologie di dipendenza tra due marginali, basta far variare opportunamente il parametro. Per soddisfare il vincolo della positività del parametro spesso le procedure di stima vengono impostate sul  $\ln(\theta)$  per poi applicare l'esponenziale ottenendo, dunque, il valore cercato; spesso si fa ricorso al delta method oppure a quanto riportato in Rockinger & Jondeau (2006).

### 2.2.2 Indici di dipendenza

Esistono alcune misure di dipendenza strettamente legate al concetto di Copula e alternative al coefficiente di correlazione  $\rho$ . In particolare, queste sono le misure di “rank correlation” e i coefficienti di dipendenza nelle code che, a differenza di  $\rho$ , dipendono esclusivamente dalla Copula della distribuzione multivariata e non dalle marginali. Due sono le principali misure di correlazione di tipo rank, ovvero il tau di Kendall ( $\tau_K$ ) e il rho di Spearman ( $\rho_S$ ). Il primo può essere visto come una misura di concordanza tra due v.a. Due punti di  $R^2$ , diciamo  $(x, y)$  e  $(\bar{x}, \bar{y})$ , sono concordanti se  $(x - \bar{x})(y - \bar{y}) > 0$ , discordanti altrimenti. Consideriamo un vettore di due v.a.  $(X, Y)$  ed una sua copia indipendente  $(\bar{X}, \bar{Y})$ ; il  $\tau_K$  altro non è che la probabilità di concordanza meno la probabilità di discordanza tra i due

vettori; in particolare, avremo:

$$\begin{aligned}\tau_K(X, Y) &= P((X - \bar{X})(Y - \bar{Y}) > 0) - P((X - \bar{X})(Y - \bar{Y}) < 0) \\ &= E[\text{sign}((X - \bar{X})(Y - \bar{Y}))]\end{aligned}\quad (2.22)$$

Per quanto riguarda, invece, il  $\rho_S$  avremo:

$$\rho_S(X, Y) = \rho(F_1(X), F_2(Y)) = \rho(C(F_1, F_2)) \quad (2.23)$$

dove  $F_1(X)$  è la funzione di ripartizione marginale della v.a.  $X$ , mentre  $F_2(Y)$  è quella di  $Y$ . In altre parole, il rho di Spearman è il coefficiente di correlazione calcolato per la Copula  $C$ . Le misure appena descritte sono entrambe simmetriche, assumono valori nell'intervallo  $[-1, 1]$  ed indicano indipendenza se assumono valore nullo (nonostante una rank correlation nulla non implichi indipendenza). Sono pari ad 1 se le v.a.  $X$  ed  $Y$  sono co-monotone, -1 se sono, invece, contro-monotone. Possono essere riscritte anche nel seguente modo, supponendo di avere marginali continue e, dunque, Copula unica  $C$ :

$$\begin{cases} \tau_K(X, Y) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1 \\ \rho_S(X, Y) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2 \end{cases} \quad (2.24)$$

Passiamo ora ai coefficienti di dipendenza di coda che, come le misure appena viste, misurano la dipendenza tra due v.a. continue e dipendono unicamente dalla Copula che lega tali v.a. Questi coefficienti sono in grado di misurare l'intensità della dipendenza che sussiste, appunto, tra le code delle distribuzioni marginali. Nello specifico, per quel che concerne le sole code destre delle distribuzioni marginali, avremo il cosiddetto *up tail dependence index*, definito come la probabilità che la v.a.  $Y$  ecceda il suo p-esimo percentile, dato che la v.a.  $X$  eccede il suo p-esimo percentile (e viceversa),

per  $p \rightarrow 1$ . Formalmente,

$$\lambda_u(X, Y) = \lim_{p \rightarrow 1^-} P(Y > F_2^{\leftarrow}(p) | X > F_1^{\leftarrow}(p)) \quad (2.25)$$

Facendo riferimento, invece, alle code sinistre, avremo:

$$\lambda_l(X, Y) = \lim_{p \rightarrow 0^+} P(Y \leq F_2^{\leftarrow}(p) | X \leq F_1^{\leftarrow}(p)) \quad (2.26)$$

Supponendo di avere marginali continue, potremo riscrivere i coefficienti di tail dependence come:

$$\begin{cases} \lambda_u(X, Y) = \lim_{p \rightarrow 1^-} \frac{\hat{C}(1-p, 1-p)}{1-p} = \lim_{p \rightarrow 0^+} \frac{\hat{C}(p, p)}{p} \\ \lambda_l(X, Y) = \lim_{p \rightarrow 0^+} \frac{C(p, p)}{p} \end{cases} \quad (2.27)$$

I due coefficienti coincidono se  $C = \hat{C}$ , ovvero se si tratta di una Copula simmetrica radialmente (si veda McNeil et al., 2005).

## 2.3 Extreme Value Theory multivariata e funzioni Copula

In questa sezione riportiamo alcuni importanti concetti della teoria dei valori estremi in più dimensioni, estendendo la EVT unidimensionale attraverso l'impiego della funzione Copula. Quest'ultima può anche descrivere il comportamento congiunto di quelle v.a. i cui massimi (o minimi) periodali o le cui code (da qui il legame ai valori estremi) mostrano una certa dipendenza (si veda Galambos, 1987).

Siano  $\mathbf{X}_1, \dots, \mathbf{X}_n$  vettori casuali i.i.d. definiti in  $\mathfrak{R}^d$  con distribuzione congiunta  $F$  e distribuzioni marginali  $F_1, \dots, F_d$ . Questi vettori i.i.d. formano, dunque, una matrice  $\mathbb{X}$  con  $n$  righe e  $d$  colonne, ovvero per ciascuna v.a.,

riportate per colonna, abbiamo  $n$  osservazioni (righe). Definiamo il massimo della  $j$ -esima v.a. come  $M_{n,j} = \max(X_{1,j}, \dots, X_{n,j})$ ,  $\forall j = 1, 2, \dots, d$ ; nella EVT classica multivariata l'oggetto di interesse è l'insieme dei valori massimi, ovvero  $\mathbf{M}_n = (M_{n,1}, \dots, M_{n,d})'$ , la cui distribuzione limite esiste se vale

$$\lim_{n \rightarrow \infty} P\left(\frac{\mathbf{M}_n - \mathbf{d}_n}{\mathbf{c}_n} \leq \mathbf{x}\right) = \lim_{n \rightarrow \infty} F^n(\mathbf{x}\mathbf{c}_n + \mathbf{d}_n) = H(\mathbf{x}) \quad (2.28)$$

dove  $\mathbf{d}_n = (d_{n,1}, \dots, d_{n,d})'$  e  $\mathbf{c}_n = (c_{n,1}, \dots, c_{n,d})' > \mathbf{0}$  sono vettori contenenti delle costanti di normalizzazione (come nel caso univariato). In particolare, se tale limite esiste allora  $H$  è la distribuzione multivariata dei valori estremi (MEV distribution) e si può scrivere  $F \in MDA(H)$ , ovvero  $F$  appartiene al massimo dominio di attrazione di  $H$ . Ciascuna marginale di  $H$  avrà distribuzione continua di tipo GEV, ovvero Gumbel, Fréchet o Weibull e, dal teorema di Sklar (2.1), esiste unica la funzione Copula  $C$  che definisce il legame tra tali marginali. In particolare, vale il seguente:

**Teorema 2.3 (Copula EV)** *Se esiste una certa  $H$  con marginali di tipo GEV, allora  $\exists!$  la Copula  $C$  tale che  $C(\mathbf{u}^t) = C^t(\mathbf{u})$ ,  $\forall t > 0$ .*

Una Copula che soddisfa tale teorema è detta Copula dei valori estremi (Copula EV) e può essere impiegata come distribuzione MEV. Esempi di copule EV sono la Copula indipendenza (2.12) e quella comonotonicità (2.13).

Un altro teorema fondamentale in tale contesto è il seguente:

**Teorema 2.4**  *$C$  è una Copula EV se e solo se ha la forma seguente*

$$C(\mathbf{u}) = \exp\left\{B\left(\frac{\ln u_1}{\sum_{k=1}^d \ln u_k}, \dots, \frac{\ln u_d}{\sum_{k=1}^d \ln u_k}\right) \sum_{i=1}^d \ln u_i\right\}$$

dove  $B(\mathbf{w}) = \int_{S_d} \max(x_1 w_1, \dots, x_d w_d) dH(\mathbf{x})$  è la *funzione di dipendenza di coda*, anche detta *funzione di dipendenza di Pickands*, ed  $H$  è una misura finita del semplice<sup>1</sup>  $d$ -dimensionale  $S_d = \{\mathbf{x} : x_i \geq 0, i = 1, 2, \dots, d, \sum_{i=1}^d x_i = 1\}$ . La funzione  $B$  soddisfa le seguenti proprietà:

- è convessa;
- $\max(w_1, \dots, w_d) \leq B(\mathbf{w}) \leq 1$ ;
- si ha indipendenza se e solo se  $B(\mathbf{w}) = 1$ ;
- al decrescere di  $B(\mathbf{w})$  cresce la dipendenza.

Nel caso bidimensionale la funzione  $B(w)$  è ridefinita come  $B((w, 1 - w))$  con  $w \in [0, 1]$ . Solo ed esclusivamente nel caso bidimensionale, una Copula è una Copula EV se e solo se ha la seguente forma:

$$C(u_1, u_2) = \exp\left\{\ln(u_1 u_2) B\left(\frac{\ln u_1}{\ln(u_1 u_2)}\right)\right\} \quad (2.29)$$

dove, ora,  $B(w) = \int_0^1 \max((1 - x)w, x(1 - w)) dH(x)$  ed  $H$  è una misura di  $[0, 1]$ . Per ulteriori dettagli si rimanda a McNeil et al. (2005).

Riportiamo ora alcuni risultati relativi all'approccio POT multivariato facendo uso, come prima, del concetto di Copula. L'idea è quella di poter definire dei modelli multivariati in grado di cogliere l'andamento congiunto degli eccessi oltre determinate soglie attraverso la distribuzione generalizzata di Pareto (GPD) e, appunto, un'opportuna funzione Copula. Supponiamo che i vettori  $\mathbf{X}_1, \dots, \mathbf{X}_n$  abbiano una distribuzione congiunta  $F(\mathbf{x}) = C(F_1(x_1), \dots, F_d(x_d))$  sconosciuta in quanto non è nota la Copula  $C$ . Siamo interessati a descrivere il comportamento della coda destra

<sup>1</sup>Un  $d$ -politopo è l'analogo di un poligono nel piano ( $d=2$ ) e di un poliedro nello spazio ( $d=3$ ) generalizzato ad uno spazio euclideo reale  $\mathfrak{R}^d$ . Un semplice  $d$ -dimensionale è il politopo  $d$ -dimensionale col minor numero di vertici. Il semplice di dimensione zero è un singolo punto, il semplice bidimensionale è un triangolo, quello tridimensionale un tetraedro, etc.

di  $F$ , oltre certe soglie  $\mathbf{u} = (u_1, \dots, u_d)'$ . Dalla EVT univariata sappiamo che, per una soglia opportunamente grande, la distribuzione della  $j$ -esima v.a. marginale coincide con quella di una GPD (1.12). Per brevità la indichiamo con  $G_j(x_d)$ . Per tale motivo è possibile considerare la seguente approssimazione

$$F(\mathbf{x}) \approx C(G_1(x_1), \dots, G_d(x_d)) \quad (2.30)$$

Ancora non conosciamo la Copula  $C$  ma, tuttavia, possiamo approssimarla con la Copula limite  $C_0$  (si rimanda a McNeil et al. (2005) per i dettagli). In particolare,  $\forall \mathbf{v} \in [0, 1]^d$  e  $t$  sufficientemente grande, possiamo scrivere  $C(\mathbf{v}^{1/t}) \approx C_0^{1/t}(\mathbf{v})$ ; ponendo  $\mathbf{w} = \mathbf{v}^{1/t}$ , per la proprietà scalare della Copula, avremo:

$$C(\mathbf{w}) \approx C_0^{1/t}(\mathbf{w}^t) = C_0(\mathbf{w}) \quad (2.31)$$

La distribuzione multivariata  $F$  potrà, dunque, essere approssimata come:

$$F(\mathbf{x}) \approx C_0(G_1(x_1), \dots, G_d(x_d)), \mathbf{x} \geq \mathbf{u}. \quad (2.32)$$

Avendo accennato al concetto di funzione Copula dei valori estremi, ne riportiamo di seguito alcuni esempi. In particolare, definiremo 4 differenti copule EV, ovvero la Galambos, la Husler-Reiss, la Tawn e la t-EV. Iniziamo proprio da quest'ultima, la quale può essere intesa come la Copula che descrive la dipendenza tra i valori massimi (o anche tra valori oltre una soglia  $u$ ) di v.a. aventi distribuzione  $d$ -variata  $t_d(\mu, \Sigma; \nu)$  di tipo t-Student. Soffermandoci sul caso bidimensionale, riportiamo prima la Copula t-Student, da cui poi potremo ottenere la t-EV; in particolare, sappiamo che una Copula è invariante rispetto ad una standardizzazione delle marginali e, pertanto, la Copula per una distribuzione bivariata  $t_2(\mu, \Sigma; \nu)$ , con  $\mu$  vettore delle medie e  $\Sigma$  matrice di dispersione, è uguale alla Copula per la distribuzione bivariata  $t_2(\mathbf{0}, \mathbf{P}; \nu)$ , dove  $\mathbf{P}$  è la matrice di correlazione indotta da  $\Sigma$ .

Avremo, quindi

$$C_{\nu, \mathbf{P}}^t(u_1, u_2) = \frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\nu/2)\sqrt{(\pi\nu)^2|\mathbf{P}|}} \int_{-\infty}^{t_{\nu}^{-1}(u_1)} \int_{-\infty}^{t_{\nu}^{-1}(u_2)} \left(1 + \frac{\mathbf{x}'\mathbf{P}^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+2}{2}} d\mathbf{x} \quad (2.33)$$

dove  $t_{\nu}^{-1}$  è la funzione quantile di una t-Student univariata con  $\nu$  gradi di libertà. Il seguente teorema (si veda Demarta & McNeil, 2005) definisce la Copula t-EV:

**Teorema 2.5** *Sia  $\rho$  il coefficiente di correlazione delle due marginali. La Copula bivariata  $C_{\nu, \mathbf{P}}^t(u_1, u_2)$  è attratta al limite EV dato da*

$$C_{\nu, \rho}^{tEV}(u_1, u_2) = \exp\left(\ln(u_1 u_2) B_{\nu, \rho}\left(\frac{\ln(u_1)}{\ln(u_1 u_2)}\right)\right) \quad (2.34)$$

$$B_{\nu, \rho}(w) = w t_{\nu+1}\left(\frac{\left(\frac{w}{1-w}\right)^{1/\nu} - \rho}{\sqrt{1-\rho^2}} \sqrt{\nu+1}\right) + (1-w) t_{\nu+1}\left(\frac{\left(\frac{1-w}{w}\right)^{1/\nu} - \rho}{\sqrt{1-\rho^2}} \sqrt{\nu+1}\right)$$

e  $t_{\nu+1}$  è la funzione di ripartizione di una t-Student univariata standardizzata.  $B_{\nu, \rho}(w)$  è la funzione di dipendenza per la Copula t-EV. Se confrontiamo la (2.34) con la (2.29) ci accorgiamo che sono identiche a meno della definizione della funzione  $B$ ; questo conferma quanto affermato dalla (2.29), ovvero che una Copula EV è definita unicamente dalla funzione di dipendenza  $B$ .

La Copula t-EV non è particolarmente conveniente per fini pratici, a causa del fatto che la sua densità, necessaria per ottenere le stime di massima verosimiglianza, è decisamente scomoda da trattare analiticamente; inoltre, studi empirici hanno dimostrato che la Copula t-EV, per ogni scelta di  $\nu$  e  $\rho$ , coincide praticamente o con la Copula di Gumbel (2.16), che è anch'essa una Copula EV, o con quella di Galambos, la quale è definita dalla (2.29) e dalla funzione di dipendenza

$$B_{\theta}(w) = 1 - \left(w^{-1/\theta} + (1-w)^{-1/\theta}\right)^{\theta} \quad (2.35)$$

Per completezza, riportiamo anche la funzione di dipendenza della Copula Gumbel (2.16)

$$B_\theta(w) = \left( w^{1/\theta} + (1-w)^{1/\theta} \right)^\theta \quad (2.36)$$

La Copula di Husler-Reiss (1989), invece, è definita come

$$C_\theta^{HR}(u_1, u_2) = \exp \left\{ \Phi \left[ \frac{\theta}{2} + \frac{1}{\theta} \ln \left( \frac{\ln u_2}{\ln u_1} \right) \right] \ln u_1 + \Phi \left[ \frac{\theta}{2} + \frac{1}{\theta} \ln \left( \frac{\ln u_1}{\ln u_2} \right) \right] \ln u_2 \right\} \quad (2.37)$$

dove  $\Phi$  è la funzione di ripartizione di una Normale standard.

La Copula di Tawn, infine, è una Copula EV asimmetrica, estensione della Copula di Gumbel. Essa è definita dalla seguente funzione di dipendenza:

$$B_{\theta, \psi_1, \psi_2}(w) = 1 - \psi_1 + w(\psi_1 - \psi_2) + \left( (\psi_1 w)^{1/\theta} + (\psi_2(1-w))^{1/\theta} \right)^\theta \quad (2.38)$$

dove  $\psi_1, \psi_2 \in [0, 1]$

## 2.4 Simulazioni

In questa sezione investighiamo le caratteristiche delle copule appena definite attraverso procedure di simulazione; per il momento, dunque, non faremo riferimento ad un particolare dataset in quanto l'obiettivo è quello di comprendere le potenzialità delle diverse copule e di cogliere le differenze strutturali che sussistono tra di esse. Iniziamo dalla Copula più semplice, ovvero quella di indipendenza che potrà essere utilizzata come termine di confronto per le altre; in particolare, la riportiamo nel seguente grafico, mentre nel successivo abbiamo le sue curve di livello:

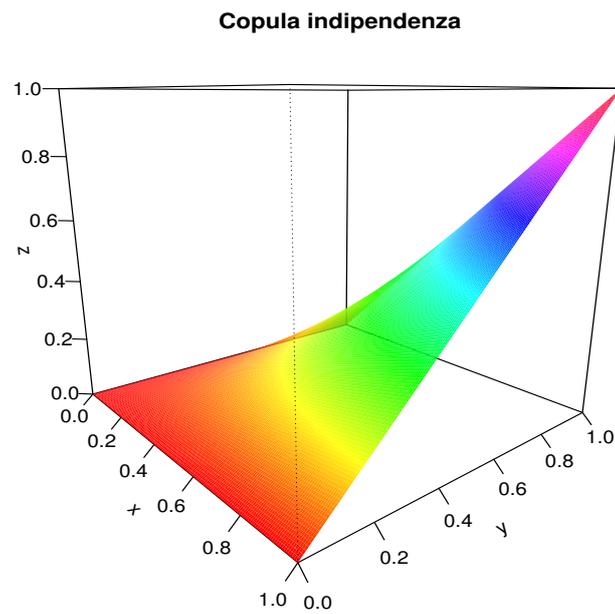


Figura 2.1: Copula di indipendenza.

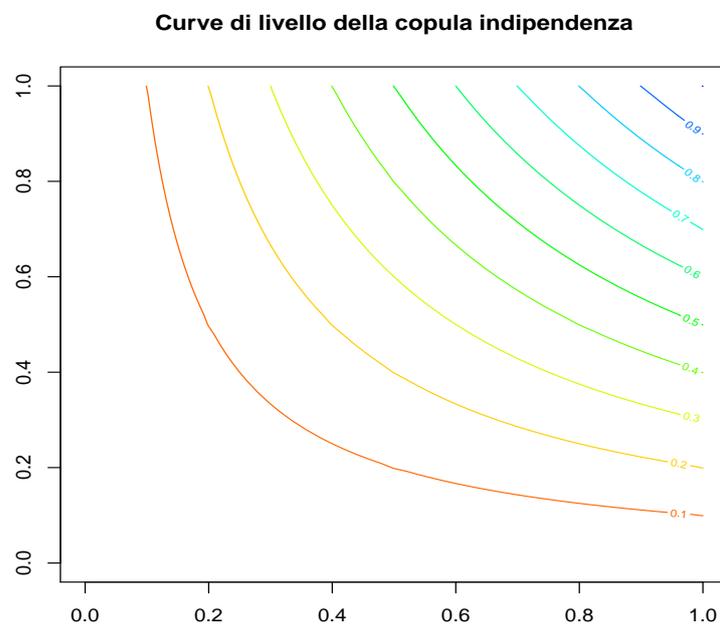


Figura 2.2: Curve di livello della Copula di indipendenza.

Anche le copule definite dai limiti di Fréchet ((2.13) e (2.14)) possono

essere utilizzate come riferimento per le altre copule:

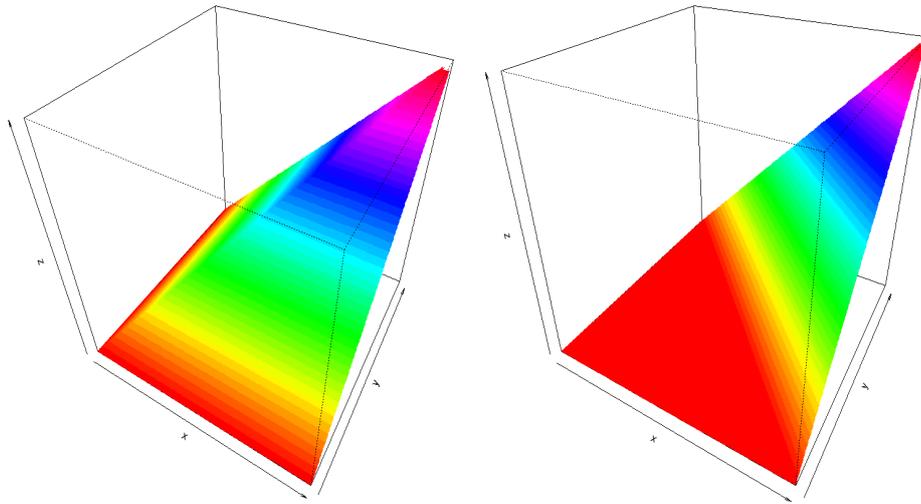


Figura 2.3: Copula massima (a sinistra) e Copula minima (a destra).

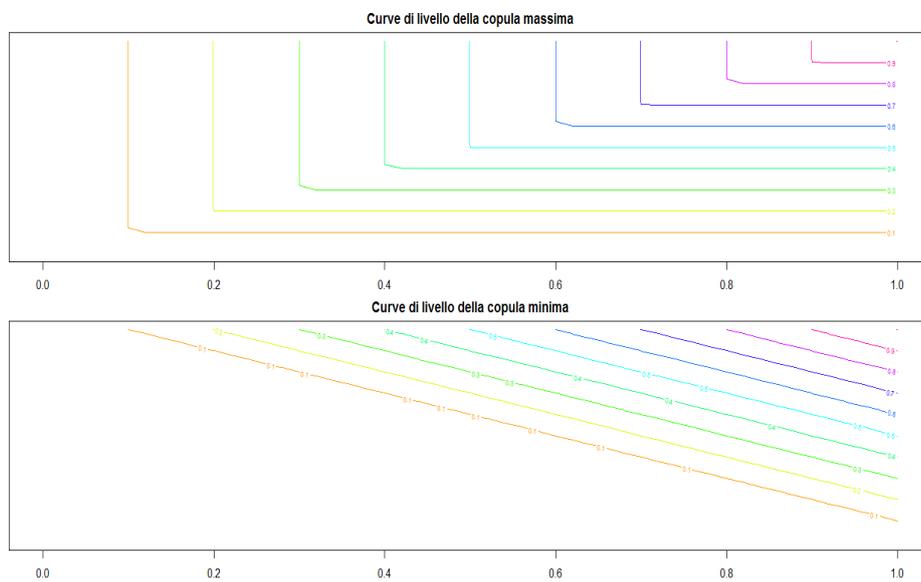


Figura 2.4: Curve di livello delle copule massima e minima.

Osserviamo come la Copula massima sia caratterizzata da curve di livello convesse che formano un angolo retto, mentre quelle della Copula minima sono dei segmenti. Capiamo, dunque, che tanto più prossimo a  $90^\circ$  ( $180^\circ$ ) è l'angolo formato dalle curve di livello di una Copula tanto maggiore sarà la dipendenza positiva (negativa) tra le marginali; l'indipendenza si può inquadrare, dunque, come caso intermedio.

Passiamo alla Copula gaussiana (2.15) di cui osserviamo la densità al variare del parametro  $\theta$  e, dunque, al variare delle misure di associazione  $\tau_K$  e  $\rho_S$ ; in particolare, sull'asse delle quote sono riportate le frequenze assolute prodotte dalla simulazione. Per valori negativi del  $\tau_K$  osserviamo una dipendenza negativa tra le due marginali; infatti, la densità risulta maggiore per piccoli valori di una marginale ( $y$ ) e grandi per l'altra ( $x$ ) oppure per valori grandi di  $y$  e piccoli di  $x$ . Tale relazione va scemando man mano che il valore del  $\tau_K$  cresce, annullandosi, ovviamente, per  $\tau_k = 0$  (caso indipendenza). Dal momento in cui il  $\tau_K$  assume valori positivi, riemerge una dipendenza ma nel verso opposto al precedente, ovvero osserviamo una dipendenza positiva la quale, tra l'altro, è molto accentuata tra le code delle distribuzioni marginali.

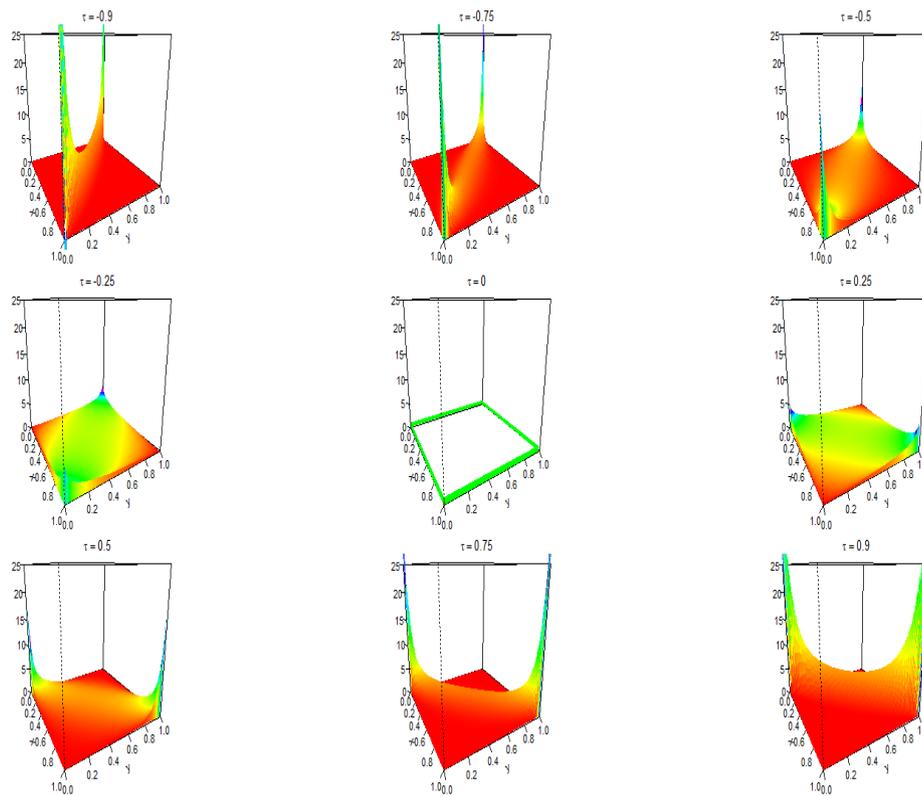


Figura 2.5: Funzione di densità della Copula gaussiana al variare di  $\tau_K$ .

Per quel che riguarda, invece, le curve di livello della Copula gaussiana, queste tendono a coincidere con quelle della Copula minima per  $\tau_K$  prossimo a -1, coincidono con quelle della Copula indipendenza per  $\tau_K = 0$ , mentre tendono a formare un angolo di  $90^\circ$  per  $\tau_K \rightarrow 1$ . In ogni caso, risultano essere curve convesse.

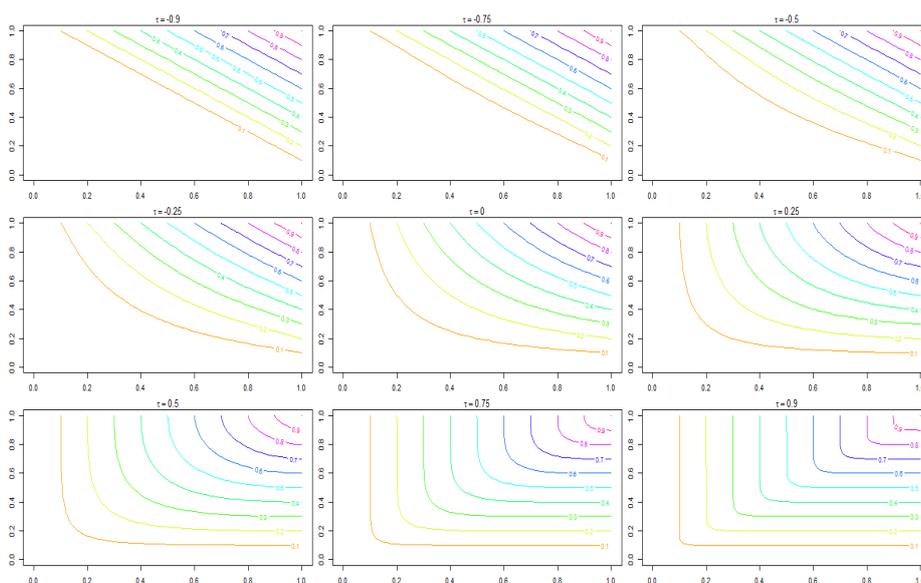


Figura 2.6: Curve di livello della Copula gaussiana.

Di seguito, invece, alcuni grafici relativi alla Copula t-Student (2.33); fissiamo 1 grado di libertà. La sua densità, come quella della gaussiana, mette in risalto una dipendenza negativa quando  $\tau_K$  è negativo, positiva quando esso è positivo. Tuttavia, a differenza della gaussiana, la t-Student non tende alla Copula indipendenza per  $\tau_K \rightarrow 0$ : in tal caso, infatti, la densità si mantiene maggiore dell'unità negli angoli del grafico, ovvero per valori grandi sia per x sia per y, per valori piccoli sia per x sia per y (dipendenza positiva), per valori grandi di x ma piccoli per y e per piccoli valori di x e grandi per y (dipendenza negativa). Tali relazioni sono confermate anche dalle curve di livello della densità stessa. Da quanto detto, dunque, si può dedurre che tale Copula è in grado di cogliere una eventuale doppia forma di dipendenza tra le marginali.

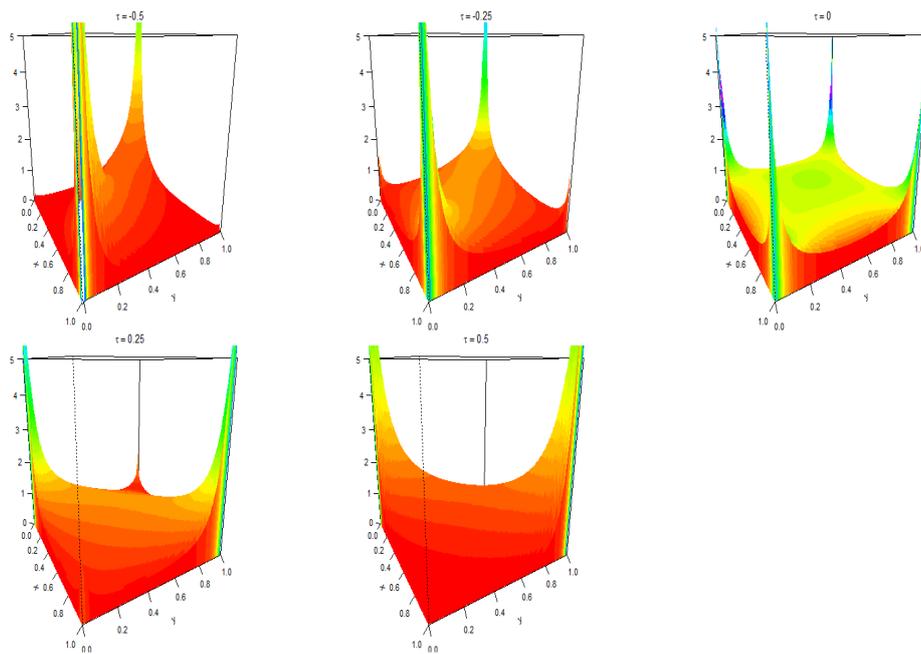


Figura 2.7: Funzione di densità della Copula t-Student al variare di  $\tau_K$

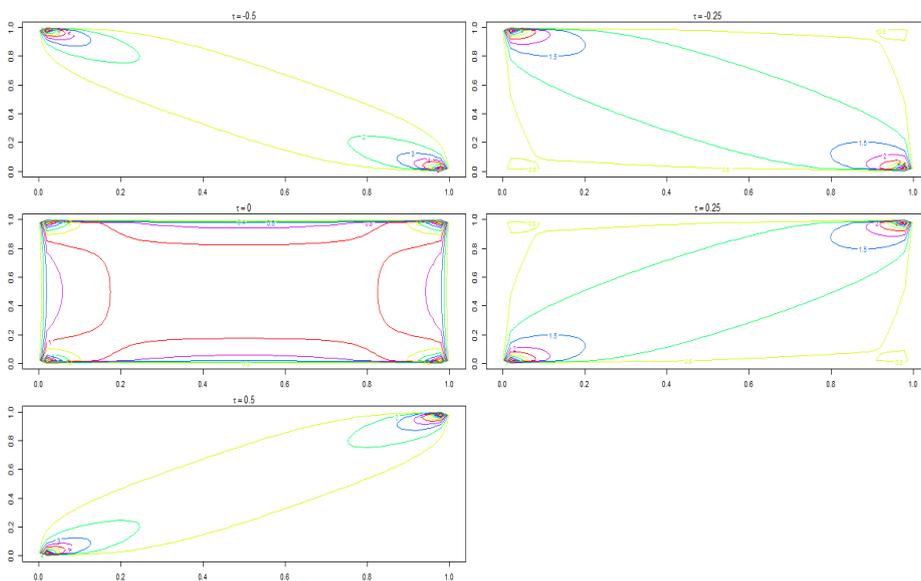


Figura 2.8: Curve di livello della densità della Copula t-Student.

La Copula t-Student presenta sempre un doppio trend, sia positivo che negativo, e assume, quindi, forma di croce.

Per la Copula di Gumbel (2.16) sono stati ottenuti, al variare, come al solito, di  $\tau_K$ , i seguenti grafici relativi alla sua densità e alle sue curve di livello. Per quanto riguarda la densità, è evidente che c'è coincidenza con quella della Copula indipendenza per  $\tau_K = 0$ : in tal caso, infatti, la frequenza è costantemente pari ad 1, ovvero, eseguendo  $n$  simulazioni, per ciascuna coppia  $(u_1, u_2)$  avremo una densità congiunta pari a  $\frac{1}{n}$ ; una densità uniforme è esattamente ciò che si ottiene nel caso di indipendenza. All'aumentare di  $\tau_K$ , la dipendenza positiva ha un'intensità via via crescente; caratteristica molto importante di tale Copula è che non è in grado strutturalmente di descrivere un'eventuale forma negativa di dipendenza. Inoltre, è classificata anche come una Copula EV dato che risalta la dipendenza di coda delle marginali, in particolare quella tra le code destre.

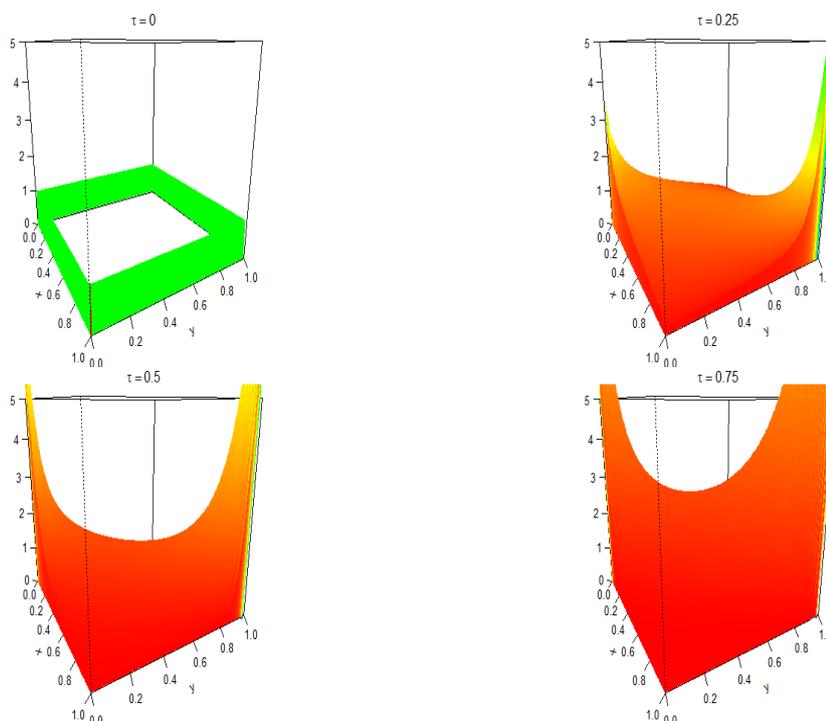


Figura 2.9: Funzione di densità della Copula Gumbel al variare di  $\tau_K$

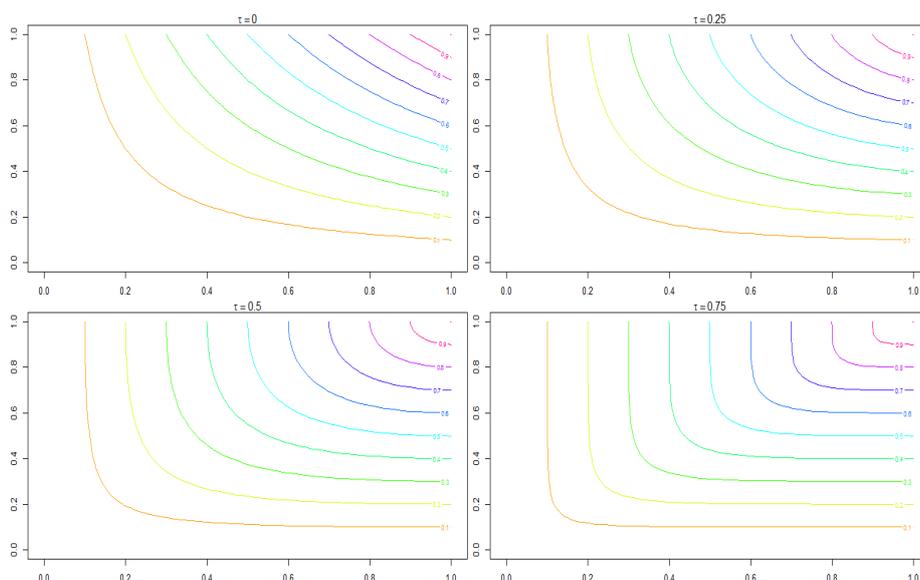


Figura 2.10: Curve di livello della Copula Gumbel.

Questa peculiarità è particolarmente interessante in ambito attuariale; infatti, ipotizzando che le due marginali rappresentino il costo dei sinistri attritional e gravi, rispettivamente (oppure i costi dei sinistri di due differenti linee di business), l'adozione della Gumbel comporterebbe una dipendenza molto forte per i sinistri di importi più elevati, anche in presenza di una correlazione di rango piuttosto contenuta; infatti, dal suddetto grafico già per  $\tau_K = 0.25$  si osserva una dipendenza tra le code destre non irrilevante. Tra l'altro, non potremmo nemmeno trascurare una certa dipendenza tra i sinistri di ammontare basso.

La Copula di Frank (2.18), invece, come si può osservare dalle seguenti figure, sembra essere maggiormente indicata a descrivere quei casi in cui c'è una forte dipendenza (positiva o negativa) tra le marginali. Guardando soprattutto alle curve di livello della sua densità, ci accorgiamo che tale Copula spazia tra la Copula minima (primo riquadro in alto a sinistra) e

quella massima (ultimo riquadro in basso a destra), passando per il caso di indipendenza (riquadro centrale).

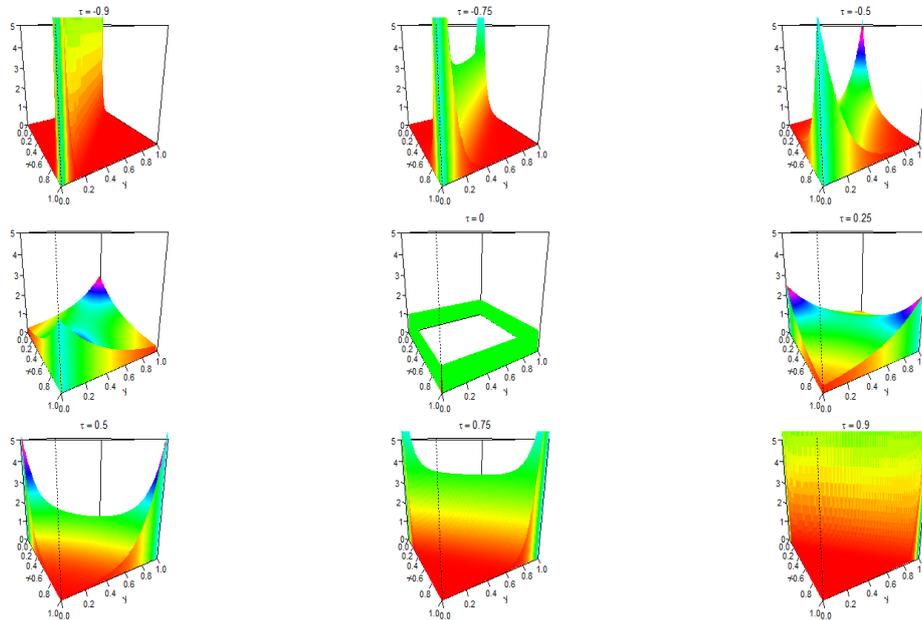


Figura 2.11: Funzione di densità della Copula Frank al variare di  $\tau_K$

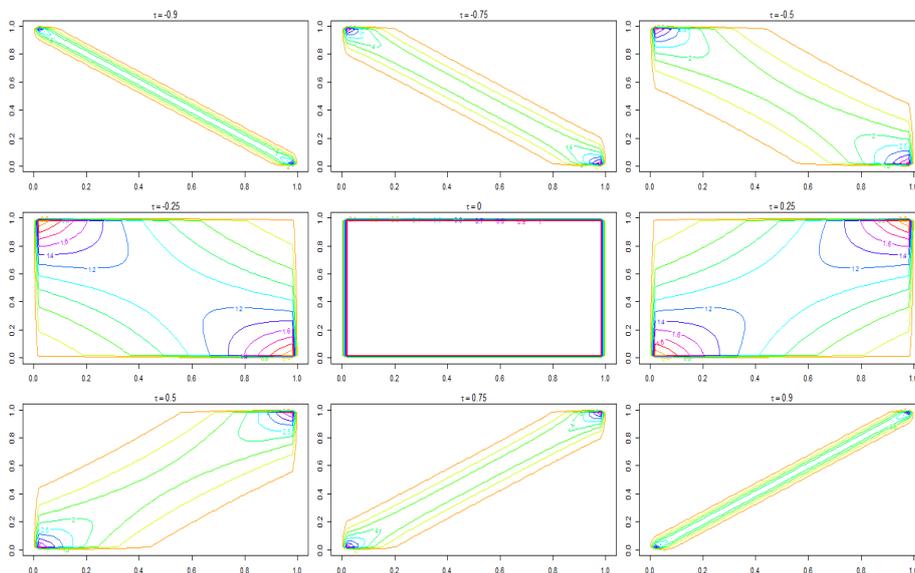


Figura 2.12: Curve di livello della densità Frank al variare di  $\tau_K$ .

Al contrario della Copula di Gumbel, la AMH (2.20) pare privilegiare una dipendenza positiva tra le code sinistre delle marginali, ovvero la densità è maggiore nei casi in cui entrambe le marginali assumono valori piccoli (lo si vede chiaramente in entrambi i grafici seguenti). Tuttavia, per valori negativi di  $\tau_K$ , il quale, in tal caso, risulta sempre  $\tau_K \geq -0.18$ , la AMH prevede una leggera dipendenza negativa.

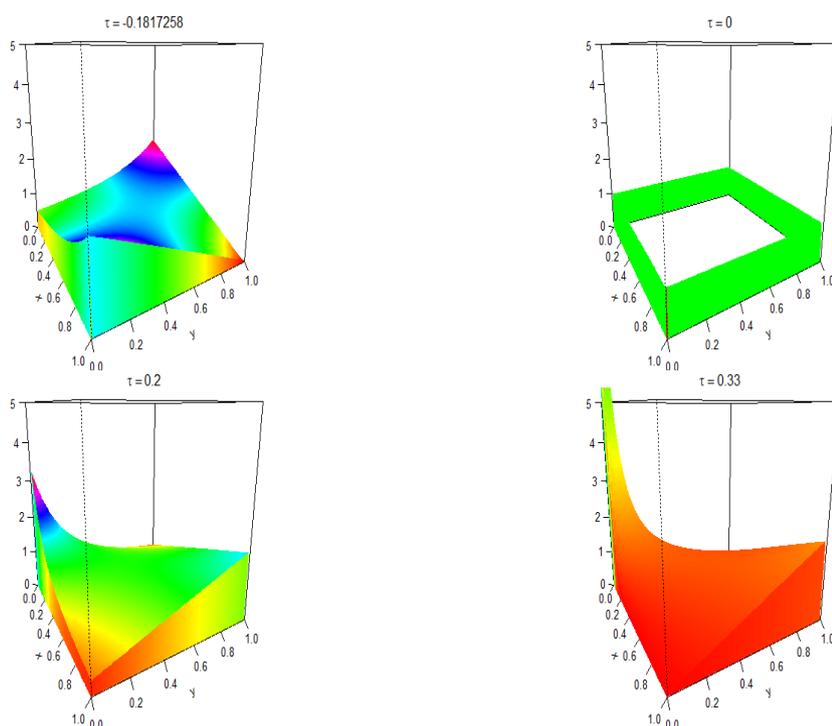


Figura 2.13: Funzione di densità della Copula AMH al variare di  $\tau_K$

Dalle curve di livello della sua densità, vediamo che al crescere di  $\tau_K$  la dipendenza diviene via via più concentrata nell'angolo in basso a sinistra, ovvero per piccoli valori per entrambe le v.a. marginali. Un comportamento simile si può ritrovare con la Copula di Clayton (2.17).

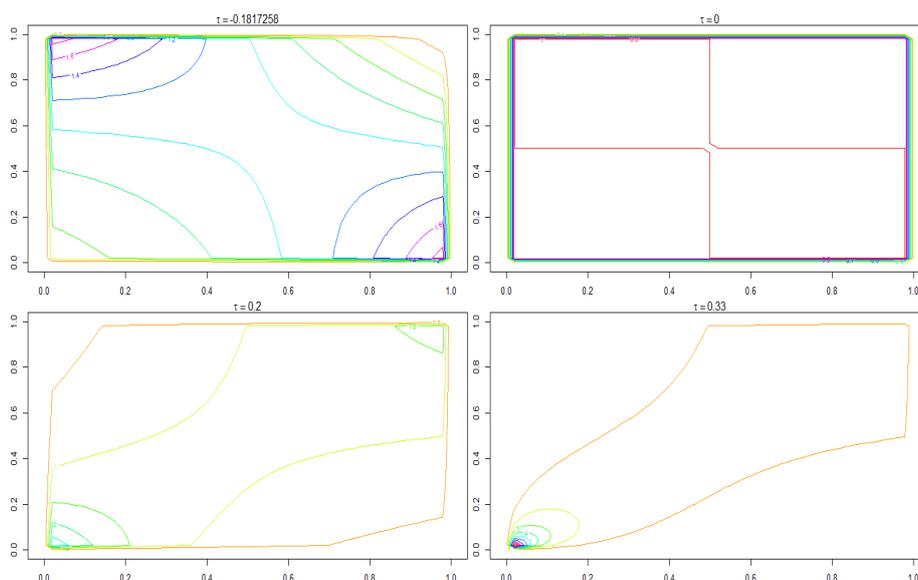


Figura 2.14: Curve di livello della densità AMH al variare di  $\tau_K$ .

Nel seguito della sezione osserveremo alcune copule EV di cui una (la Gumbel) è stata già investigata. Iniziamo dalla Galambos, definita dalla (2.29) e dalla funzione di dipendenza (2.35). Dalla forma della sua densità, è evidente il motivo per cui viene classificata come una Copula EV: la dipendenza che descrive è positiva e di maggiore intensità per le code destre delle distribuzioni marginali. Un comportamento del tutto simile lo abbiamo già osservato, appunto, per la Copula di Gumbel. Così come per quest'ultima, anche per la Galambos il  $\tau_K$  può assumere valori solo in  $[0, 1]$  e dalle sue curve di livello notiamo che si spazia tra la Copula indipendenza e quella massima.

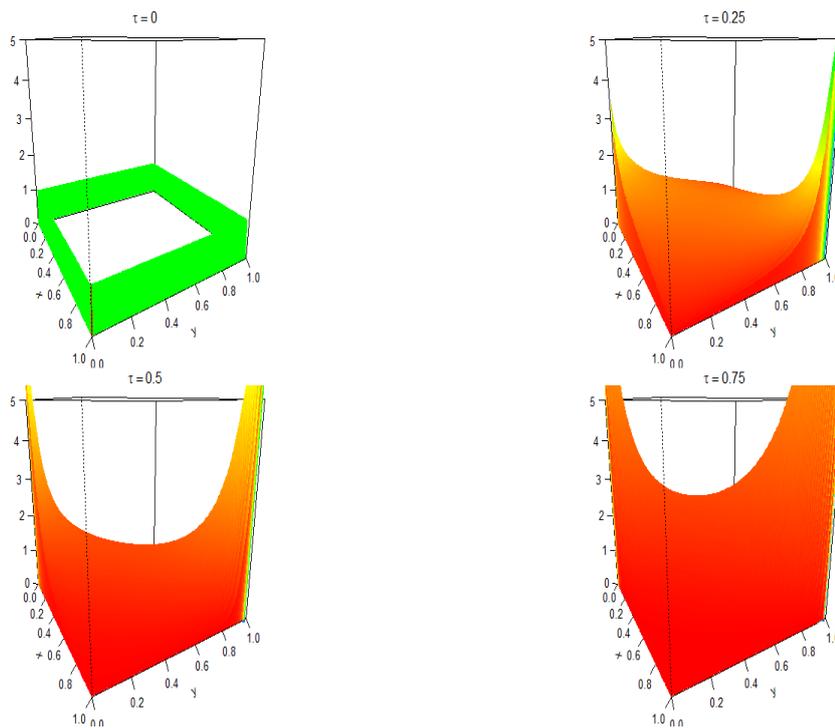


Figura 2.15: Funzione di densità della Copula di Galambos al variare di  $\tau_K$

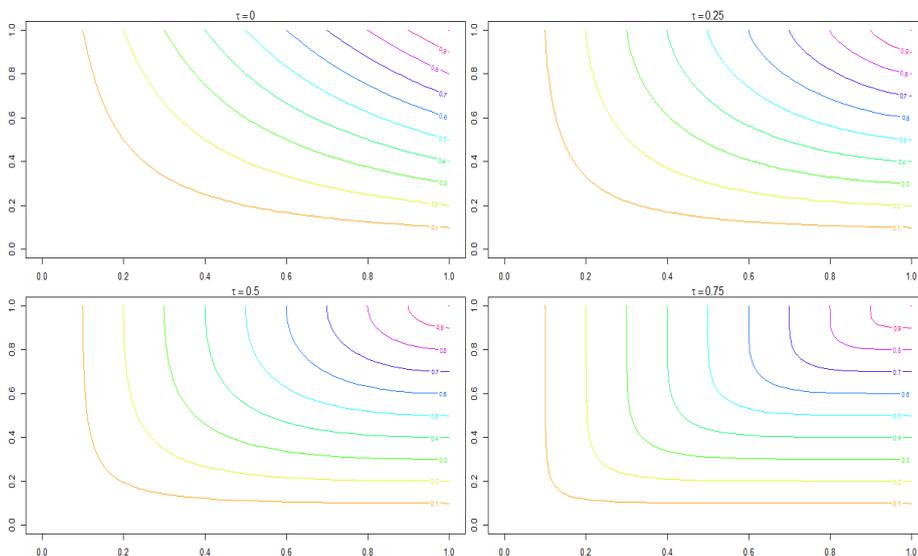


Figura 2.16: Curve di livello della Copula Galambos.

Qui di seguito, infine, osserviamo come varia la funzione di dipendenza

caratterizzante la Copula di Galambos al variare del suo parametro  $\theta$ :

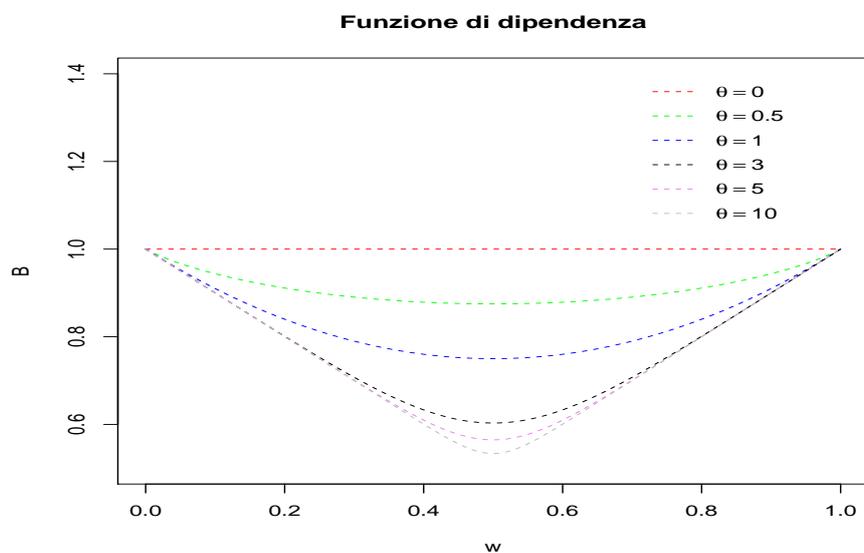


Figura 2.17: Funzione di dipendenza della Copula Galambos al variare del parametro.

Come già anticipato, la funzione di dipendenza è sempre convessa, decrescente al crescere dell'intensità della dipendenza ed è ha come valore massimo l'unità. Capiamo, dunque, che tanto maggiore è il parametro della Galambos, tanto minore è la sua funzione di dipendenza e, dunque, tanto maggiore è la dipendenza che tale Copula descrive. Questa caratteristica accomuna un pò tutte le copule EV.

Proseguiamo con la Copula di Tawn, definita dalla (2.38), della cui densità riportiamo le curve di livello:

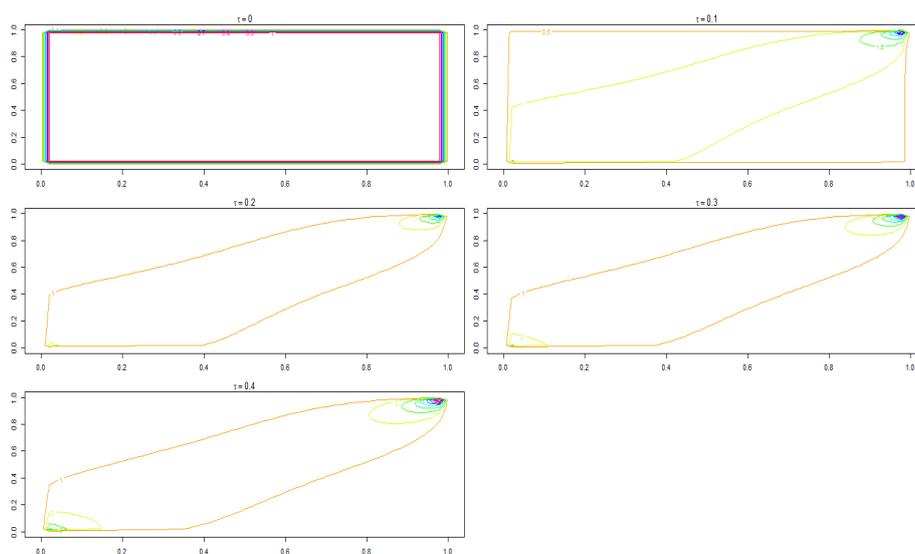


Figura 2.18: Curve di livello della densità della Copula di Tawn.

Per tale Copula il  $\tau_K$  può variare nell'intervallo chiuso  $[0, 0.42]$ . Nel grafico soprastante, si parte dal caso di totale indipendenza ( $\tau_K = 0$ ) per arrivare ad una forte dipendenza positiva tra le sole code destre delle marginali ( $\tau_K = 0.4$ ), ovvero al crescere di  $\tau_K$  cresce la probabilità che entrambe le marginali assumano valori estremi.

Concludiamo la sezione osservando le ultime due copule EV descritte in precedenza, ovvero la Husler-Reiss e la t-EV; per quest'ultima (2.34), fissando 1 grado di libertà, abbiamo ottenuto le seguenti densità:

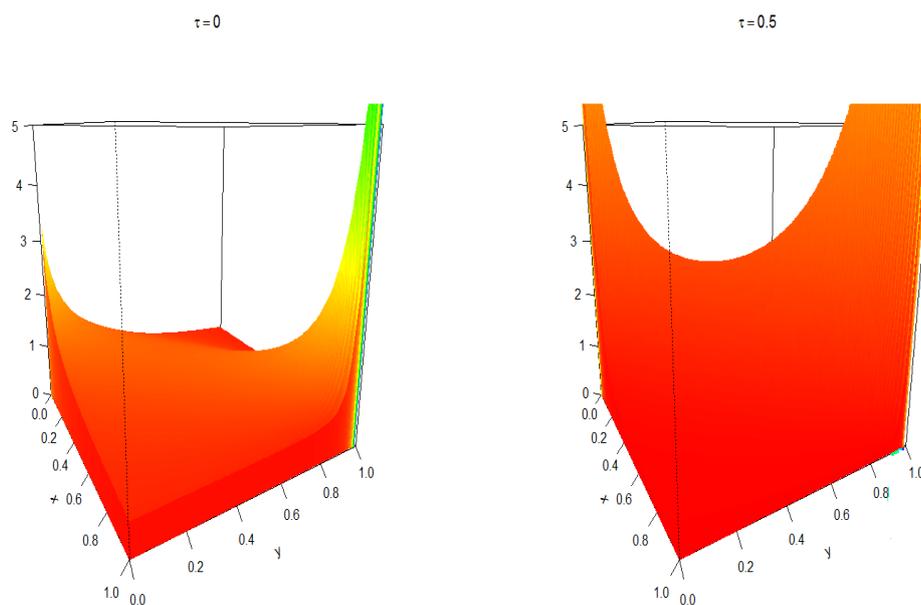


Figura 2.19: Funzione di densità della Copula t-EV con 1 grado di libertà.

Per tale Copula abbiamo anche calcolato il coefficiente di dipendenza di coda destra (2.25), in corrispondenza del valore di  $\tau_K$ :

$\tau_K$	$\lambda_u$
0	0,293
0,10	0,374
0,25	0,556
0,50	0,749
0,75	0,888
0,90	0,958
0,95	0,979

che conferma quanto già suggerito dalle densità, ovvero che al crescere del parametro  $\theta$  (o di  $\tau_K$ ) diviene più intensa la dipendenza tra i valori estremi e positivi delle marginali.

Infine, per quanto concerne la Copula di Husler-Reiss (2.37), riportiamo la funzione di dipendenza al variare del suo unico parametro  $\theta$ :

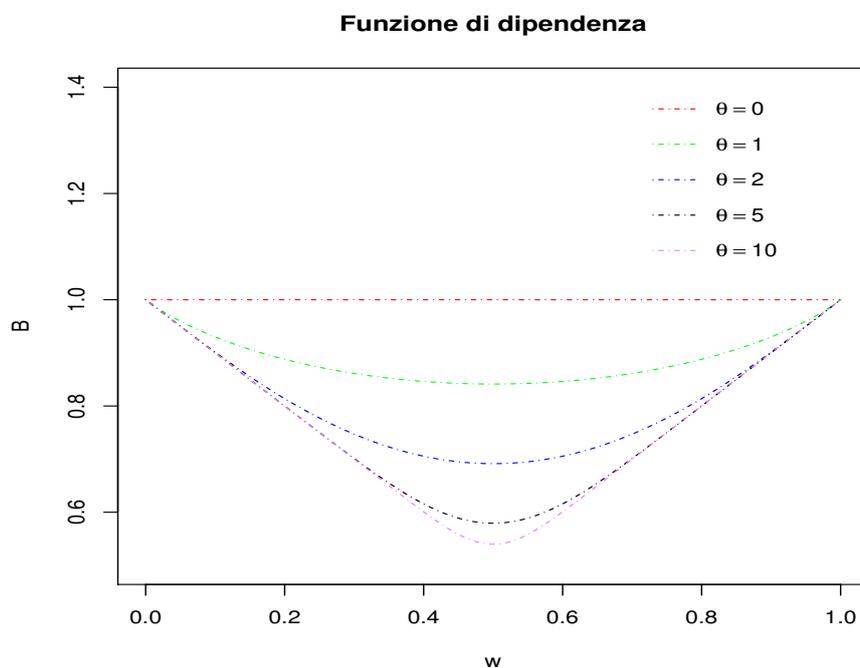


Figura 2.20: Funzione di dipendenza della Copula Husler-Reiss al variare del parametro.

## 2.5 Fast Fourier Transform

La Fast Fourier Transform, come vedremo, è una valida alternativa per approssimare la distribuzione del costo sinistri aggregato. In letteratura, l'efficacia dell'algoritmo è stata studiata e verificata per modelli univariati e bivariati ma nel prosieguo vedremo anche come ben si presta a valutazioni di problemi a dimensioni maggiori.

Secondo la teoria del rischio, il costo sinistri aggregato  $X$  è dato dalla somma di un numero aleatorio  $k$  di variabili aleatorie  $Z$ , ciascuna rappresentante il

costo del singolo sinistro (si veda Vannucci, 2010). Formalmente,

$$X = \sum_{i=1}^k Z_i \quad (2.39)$$

dove  $k$  rappresenta il numero dei sinistri che si verificano ed è, pertanto, pensato in termini aleatori.

Avendo definito  $X$  come somma di un numero aleatorio di v.a., l'applicazione dell'approccio collettivo della TdR richiederebbe il ricorso alla convoluzione di variabili aleatorie che può essere realizzata mediante simulazione o ricorsione o mediante la Fast Fourier Transform<sup>2</sup>. In quest'ottica la FFT può essere intesa come un'alternativa alle funzioni Copula viste in precedenza. Un aspetto importante nell'utilizzo della procedura FFT, è che con essa è possibile superare una delle due ipotesi caratterizzanti l'approccio collettivo della TdR, ovvero quella della reciproca indipendenza tra i sinistri. Tale ipotesi, invece, è necessaria quando si vogliono aggregare i singoli sinistri mediante convoluzione<sup>3</sup>.

Introduciamo formalmente la FFT univariata (per maggiori dettagli rimandiamo a Wang (1998) e ad Homer & Clark, 2003). Per costruire la loss distribution attraverso la FFT, è necessario, per prima cosa, discretizzare la distribuzione del costo del singolo sinistro  $Z$  e ottenere il vettore  $\mathbf{z} = (z_0, z_1, z_2, \dots, z_{n-1})$ , il cui elemento  $z_i$  è la probabilità che il singolo sinistro produca un costo pari a  $ic$ , dove  $c$  è una costante prefissata tale che, data la lunghezza  $n$  del vettore  $\mathbf{z}$ , il costo  $cn$  ha una probabilità trascurabile. Teniamo conto, inoltre, della distribuzione della frequenza dei sinistri  $k$  mediante la funzione generatrice di probabilità (probability generating

<sup>2</sup>La FFT è un'implementazione della Discrete Fourier Transform, DFT.

<sup>3</sup>L'altra ipotesi è quella di identica distribuzione per le severity dei singoli sinistri.

function, PGF), definita come

$$PGF_k(t) = \sum_{j=0}^{\infty} t^j Pr(k = j) = E[t^k] \quad (2.40)$$

In particolare, indicando la FFT e la sua inversa rispettivamente con  $FFT(\mathbf{z})$  e  $IFFT(\mathbf{z})$ , otteniamo la distribuzione di probabilità discretizzata per il costo sinistri aggregato  $X$  come

$$(x_0, x_1, \dots, x_{n-1}) = IFFT(PGF(FFT(\mathbf{z}))) \quad (2.41)$$

Sia la  $FFT(\mathbf{z})$  che la  $IFFT(\mathbf{z})$  sono dei vettori  $n$ -dimensionali i cui elementi generici sono, rispettivamente <sup>4</sup>,  $\hat{z}_k = \sum_{j=0}^{n-1} z_j \exp(\frac{2\pi i}{n} jk)$  e  $z_k = \frac{1}{n} \sum_{j=0}^{n-1} \hat{z}_j \exp(-\frac{2\pi i}{n} jk)$ ,  $i = \sqrt{-1}$ . Notiamo, dunque, che la  $IFFT$  è identica alla  $FFT$  eccetto che per un segno meno nell'esponente e per la divisione per  $n$ .

È possibile dimostrare (si veda Wang, 1998) che una trasformata di Fourier di lunghezza  $n = 2^r$  si può scrivere come somma di due distinte trasformate di Fourier, ciascuna di lunghezza  $n/2 = 2^{r-1}$ ; a loro volta, queste possono essere viste come somma di altre due trasformate di lunghezza  $2^{r-2}$  e così via... Data questa proprietà, l'algoritmo FFT lavora alla velocità massima solo quando il vettore  $\mathbf{z}$  ha una lunghezza pari ad un multiplo di 2, ovvero  $n = 2^r$ . Se la discretizzazione della distribuzione del costo del singolo sinistro produce un vettore con una lunghezza  $n \neq 2^r$ , allora basterà aggiungere un opportuno numero di zeri alla destra del vettore  $\mathbf{z}$  ed ottenere  $\mathbf{z} = (z_0, z_1, z_2, \dots, z_{n-1}, 0, 0, \dots, 0)$ .

Da un punto di vista teorico, la FFT (o meglio, la DFT) può esser vista come una versione discretizzata della trasformata di Fourier o funzione

<sup>4</sup>In R i segni sono invertiti.

caratteristica:

$$\phi(z) = \int_{-\infty}^{+\infty} f(x) \exp(izx) dx \quad (2.42)$$

La funzione caratteristica, di fatto, ad una densità di probabilità continua associa una funzione complessa continua, mentre la DFT ad un vettore di  $n$  valori associa un vettore di  $n$  valori nel campo complesso; la FFT, come detto, è l'algoritmo che ci consente di realizzare questa associazione one-to-one nel discreto.

Per quel che riguarda il caso multidimensionale, è necessario avere come input non più un vettore bensì una matrice  $\mathbf{M}_z$  contenente le probabilità congiunte delle v.a. che si vogliono aggregare. Nota tale matrice, sarà necessario applicare le versioni multivariate della FFT e della sua inversa nella stessa sequenza di prima; infatti, la matrice  $\mathbf{M}_x$ , contenente la distribuzione del costo sinistri aggregato, sarà data da

$$\mathbf{M}_x = IFFT(PGF(FFT(\mathbf{M}_z))) \quad (2.43)$$

Così come nel caso unidimensionale, anche qui la *PGF* viene applicata elemento per elemento. Più avanti vedremo un'applicazione della FFT multidimensionale i cui risultati potranno essere confrontati con quelli ottenuti attraverso l'impiego di funzioni Copula a più dimensioni; un confronto, non solo applicativo, tra tali approcci multidimensionali è riportato in Wang (1998), il cui fine ultimo è l'aggregazione dei costi di differenti rami assicurativi.

## 2.6 Errori tipici

Concludiamo il capitolo passando in rassegna alcuni errori che si devono fronteggiare se si sceglie di impiegare la metodologia delle funzioni Copu-

la o la FFT. Relativamente alle prime si incontrano difficoltà legate alla valutazione delle distribuzioni marginali, al limitato spazio parametrico e all'ottimizzazione vincolata. Errori numerici nella valutazione della funzione di ripartizione congiunta sorgono prevalentemente in prossimità dei bordi del dominio della Copula di interesse, mentre per la densità possono manifestarsi anche quando il parametro della Copula tende a certi valori limite in quanto si manifestano complicità nella procedura di massimizzazione della verosimiglianza.

Per quanto riguarda, invece, la Fast Fourier Transform, essendo una procedura puramente numerica, è stato osservato che può generare errori soprattutto nelle applicazioni relative all'analisi delle onde armoniche (processing di immagini e segnali fisici). Nel campo attuariale, seppur il primo tentativo di impiegare la FFT risale al 1983 (si veda Heckman & Meyers, 1983), solo nel 1999 Grubel & Hermesmeier condussero un'analisi approfondita di tali errori proponendo anche una miglioria della procedura basata su un cambio di misura esponenziale. Relativamente alla stessa questione, Schaller & Temnov (2008) notarono che tali tipici errori possono esser cruciali per il risultato finale, specialmente quando oggetto di interesse è una distribuzione a coda pesante.

Ci si riferisce, in sostanza, agli errori dovuti alla discretizzazione delle informazioni di input e ai cosiddetti aliasing errors. I primi sono dovuti alla scelta del passo  $h$  con cui si sceglie, appunto, di discretizzare le distribuzioni empiriche ed è possibile ridurne l'impatto valutando accuratamente i limiti inferiori e superiori di tali distribuzioni e, successivamente, riducendo di volta in volta il passo  $h$  fintanto che il beneficio marginale della stima non è piccolo abbastanza. Gli aliasing errors, invece, sono specifici della Discrete Fourier Transform dovuti al necessario troncamento delle distribuzioni che porta ad addensare attorno il valore di troncamento tutti quei valori che,

naturalmente, starebbero oltre tale soglia. Per ridurre l'impatto di tale errore si potrebbe incrementare sensibilmente il punto di troncamento oppure applicare il cosiddetto exponential tilting che genera un decadimento esponenziale della coda della distribuzione (per maggiori dettagli si rimanda a Robe-Voinea & Vernic, 2016).

# Capitolo 3

## Misure di rischio

Uno degli aspetti determinanti per la stima del requisito patrimoniale è rappresentato dall'individuazione della più idonea misura di rischio. Il concetto di Value-at-Risk (VaR) ha assunto ormai il ruolo di misura standard da utilizzare per la valutazione dell'esposizione al rischio. In realtà la necessità di determinare quali proprietà dovrebbe possedere un'idonea misura, ha prodotto una vasta letteratura riguardante lo studio delle misure di rischio portando alla definizione del concetto di misure di rischio coerenti (si veda McNeil et al., 2005); in particolare, l'introduzione del concetto di subadditività e l'utilizzo di una misura non subadditiva, quale il VaR, hanno portato all'introduzione di misure alternative fra cui l'Expected Shortfall (ES).

### 3.1 Misure di rischio coerenti

La necessità di disporre di misure di rischio coerenti è naturale quando si vogliono confrontare situazioni diverse con un criterio di scelta omogeneo. In particolare, introducendo lo spazio di probabilità  $(\Omega, F, P)$ , l'orizzonte temporale  $\Delta$ , l'insieme finito  $L^0(\Omega, F, P)$  di tutti i rischi e il sottoinsieme dei rischi finanziari  $M \subset L^0$ , generalmente ipotizzato un cono convesso, si definisce misura di rischio la funzione  $\rho(x) : M \rightarrow \Re$  che soddisfa le quattro

proprietà che verranno definite nel seguito. La funzione  $\rho(L)$  può essere interpretata come l'ammontare di capitale che dovrebbe essere aggiunto ad una posizione con perdita  $L$  in modo che la posizione divenga accettabile ad un regolatore esterno od interno.

Allo scopo di definire una misura di rischio  $\rho(L)$  coerente, occorre soddisfare le seguenti proprietà:

- invarianza per traslazione: per qualsiasi  $L \in M$  e per qualsiasi  $l \in \mathfrak{R}$  si ha  $\rho(L + l) = \rho(L) + l$ ;
- subadditività: per qualsiasi  $L_1, L_2 \in M$  si ha  $\rho(L_1) + \rho(L_2) \geq \rho(L_1 + L_2)$ ;
- positiva omogeneità: per qualsiasi  $L \in M$  e ogni  $\lambda > 0$  si ha  $\rho(\lambda L) = \lambda \rho(L)$ ;
- monotonicità: per qualsiasi  $L_1, L_2 \in M$  tali che  $L_1 \geq L_2$  quasi sicuramente si ha  $\rho(L_1) \geq \rho(L_2)$ .

L'assioma di invarianza prevede che sottraendo o aggiungendo una quantità deterministica  $l$  ad una posizione, si altera il requisito patrimoniale esattamente di quella quantità. L'assioma è fondamentale per l'interpretazione della misura di rischio come requisito patrimoniale: infatti, considerando una posizione  $L$  con  $\rho(L) > 0$  ed aggiungendo la quantità di capitale  $\rho(L)$  si ottiene la posizione accettabile  $L_M$  con  $\rho(L_M) = \rho(L - \rho(L)) = \rho(L) - \rho(L) = 0$ .

L'assioma di subadditività è sicuramente il più discusso, probabilmente anche perché esclude il VaR dalle misure di rischio coerenti, ma risulta, in ogni caso, un assioma ragionevole: tale principio riflette, infatti, l'idea che il rischio può essere ridotto mediante diversificazione. L'utilizzo di misure di rischio non subadditive per la determinazione del requisito patrimoniale in-

centiverebbe la scissione in più sotto imprese per ridurre il proprio requisito patrimoniale.

L'assioma di positiva omogeneità è una diretta conseguenza della subadditività: considerando infatti  $n$  rischi identici si ottiene  $\rho(nL) = \rho(L + L + \dots + L)$  e per la subadditività  $n\rho(L) \geq \rho(nL)$ .

Infine, l'assioma di monotonicità appare ovvio da un punto di vista economico: posizioni che possono provocare perdite più elevate devono prevedere più elevati requisiti di capitale.

L'assioma di positiva omogeneità è stato criticato suggerendo che per grandi valori di  $\lambda$  dovrebbe valere  $\rho(\lambda L) > \lambda\rho(L)$  per penalizzare la concentrazione dei rischi e superare i problemi di liquidità. Tale disuguaglianza come mostrato in precedenza viola però anche l'assioma di subadditività. Questi aspetti hanno portato all'introduzione di alternative proprietà per le misure di rischio, definite misure di rischio convesse; le ipotesi di subadditività e positiva omogeneità vengono sostituite dal seguente assioma di convessità debole:  $\lambda\rho(L_1) + (1 - \lambda)\rho(L_2) \geq \rho(\lambda L_1 + (1 - \lambda)L_2)$ .

Una misura di rischio  $\rho(L)$  è definita convessa se è debolmente convessa, cioè se soddisfa gli assiomi di invarianza, monotonicità e convessità, e se  $\rho(0) = 0$ .

## 3.2 Alcune misure di rischio

### 3.2.1 Value-at-Risk

Come anticipato in precedenza, la più comune misura di rischio utilizzata negli ultimi anni è rappresentata dal VaR. Considerando un portafoglio rischioso in un orizzonte temporale prefissato e assumendo che la perdita aggregata  $X$  abbia funzione di ripartizione  $F(x)$  al termine del periodo considerato, viene definito Value at Risk (VaR) l'ammontare tale per cui la

perdita di portafoglio sarà inferiore di tale ammontare con una probabilità  $\alpha$ :  $VaR_\alpha(X) = \inf\{x|F(x) \geq \alpha\} = \inf\{x|P(X \geq x) < 1 - \alpha\}$  (si veda McNeil et al., 2005).

Il VaR può essere quindi definito come la massima perdita ad un livello di confidenza prefissato, ovvero il capitale necessario per far fronte alla massima perdita con probabilità  $\alpha$ . Analogamente, accantonando un capitale pari al VaR, la probabilità di rovina non potrà essere superiore di  $1 - \alpha$ . Nel caso di distribuzioni ellittiche il VaR è semplicemente un multiplo della standard deviation (ad esempio per una distribuzione normale ad un livello di confidenza del 99% il VaR è pari a 2,33 volte la standard deviation); si dimostra, infatti, che qualora  $X$  sia distribuito secondo una Normale o una t-Student con media  $\mu$  e scarto quadratico medio  $\sigma$  vale la seguente relazione:  $VaR_\alpha = \mu\sigma F^{-1}(\alpha)$ .

Nel caso di distribuzioni non ellittiche il VaR presenta invece numerosi limiti; i principali sono:

- il VaR potrebbe violare la dominanza di second'ordine e non sempre descrivere l'avversione al rischio in senso tradizionale;
- il VaR non considera l'entità degli eventi superiore al VaR stesso trascurando la coda estrema della distribuzione;
- il VaR viola la subadditività e non risulta dunque una misura di rischio coerente (si veda McNeil et al., 2005).

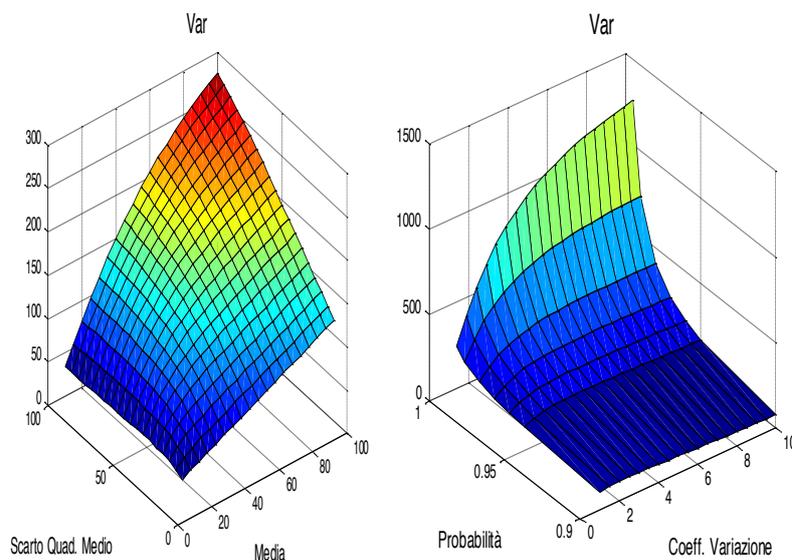


Figura 3.1: VaR al 95° percentile per una distribuzione Lognormale al variare di media e scarto quadratico medio, ed al variare del livello di confidenza e del coefficiente di variazione (media = 50).

### 3.2.2 Expected Shortfall

Come detto, la principale misura utilizzata generalmente nella pratica per la valutazione del requisito di capitale è rappresentato dal VaR, ma tale misura, violando la subadditività, può portare ad un'errata valutazione del requisito in presenza di aggregazione dei rischi. Per superare la principale limitazione del VaR, molti autori hanno proposto differenti misure di rischio che hanno assunto diverse denominazioni in letteratura, ma che coincidono sotto determinate condizioni.

Una prima misura alternativa è rappresentata dalla Tail Conditional Expectation (TCE) o Tail-VaR, che trascurando il fattore di sconto, è definita nel modo seguente:  $TCE_\alpha = E[X|X \geq VaR_\alpha(X)]$ . Tale misura non soddisfa però, nel caso di distribuzioni discrete, la subadditività. Artzner et al.

(1999) propongono così una seconda misura denominata Worst Conditional Expectation (WCE) definita come:  $WCE_\alpha = -\inf\{E[X|A : P(A) > \alpha]\}$ . Questa ulteriore misura risulta coerente al rischio ma risente della non trascurabile complessità derivante dalla dipendenza dallo spazio di probabilità  $P$  oltre che dalla variabile  $X$ .

Infine, comparando diverse definizioni proposte dalla letteratura e ottenendo una definizione basata sulla  $\alpha$ -tail mean (TM), è stato proposto il concetto di Expected Shortfall che soddisfa il requisito di subadditività anche in presenza di distribuzioni discrete o discontinue:

$$ES_\alpha(X) = -TMC_\alpha(X) = (1 - \alpha)^{-1}\{E[X|X \geq VaR_\alpha(X)] + VaR_\alpha(X)[1 - \alpha - P(X \geq VaR_\alpha(X))]\} \quad (3.1)$$

Facendo riferimento a distribuzioni continue, la formula dell'ES può essere sensibilmente semplificata nel modo seguente:

$$ES_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_u(X) du \quad (3.2)$$

In tal caso l'ES è pari alla media dei VaR per livelli di confidenza superiori ad  $\alpha$ , può essere cioè definito come la media di tutte le perdite che si hanno con probabilità al massimo pari ad  $\alpha$ . L'ES dipende dunque unicamente dalla distribuzione di  $X$  e per costruzione varrà sempre:  $ES_\alpha(X) \geq VaR_\alpha(X)$ .

La figura mostra il requisito di capitale calcolato utilizzando l'ES con una distribuzione Lognormale al variare dei parametri della distribuzione e del livello di probabilità. Come era lecito attendersi per livelli di probabilità più elevata un incremento di volatilità mostra una notevole variazione in termini di ES. Vengono inoltre ottenute abbastanza facilmente in letteratura le equazioni semplificate dell'ES in presenza di distribuzioni ellittiche. Infatti,

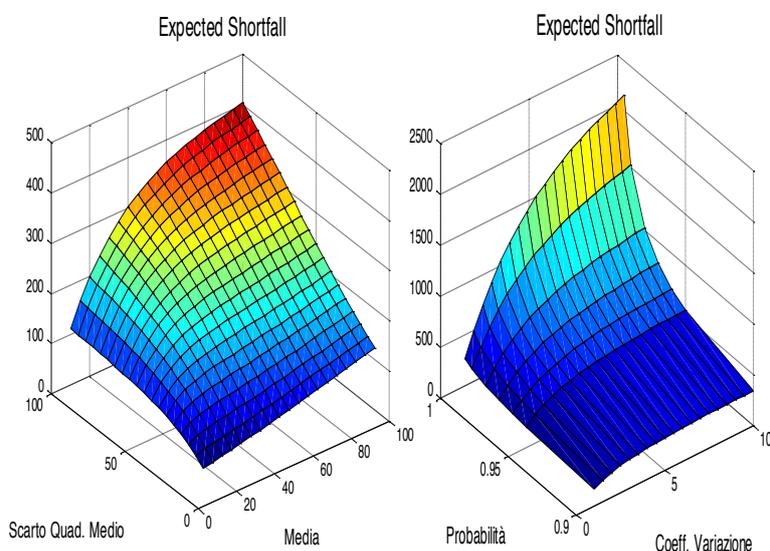


Figura 3.2: Expected Shortfall al 95° percentile per una distribuzione Lognormale al variare di media e scarto quadratico medio, ed al variare del livello di confidenza e del coefficiente di variazione (media = 50).

per una distribuzione Normale si ottiene la seguente espressione funzione dei parametri della normale e della densità della normale standard:

$$ES_{\alpha} = \mu + \sigma \frac{\phi(\phi^{-1}(\alpha))}{1 - \alpha} \quad (3.3)$$

Nel caso di distribuzioni continue senza discontinuità si dimostra inoltre che la CTE, denominata spesso in campo attuariale Tail-VaR ed identificato dall'espressione

$$\begin{aligned} Tail - VaR_{\alpha}(X) &= VaR_{\alpha} \\ &+ \frac{Pr(X > VaR_{\alpha}(X))}{1 - \alpha} E[X - VaR_{\alpha}(X) | X > VaR_{\alpha}(X)] \end{aligned} \quad (3.4)$$

coincide con l'Expected Shortfall definito in precedenza. La figura sottostante mostra un esempio di come sia possibile ottenere VaR e Tail-VaR dalla distribuzione di probabilità del costo (perdita).

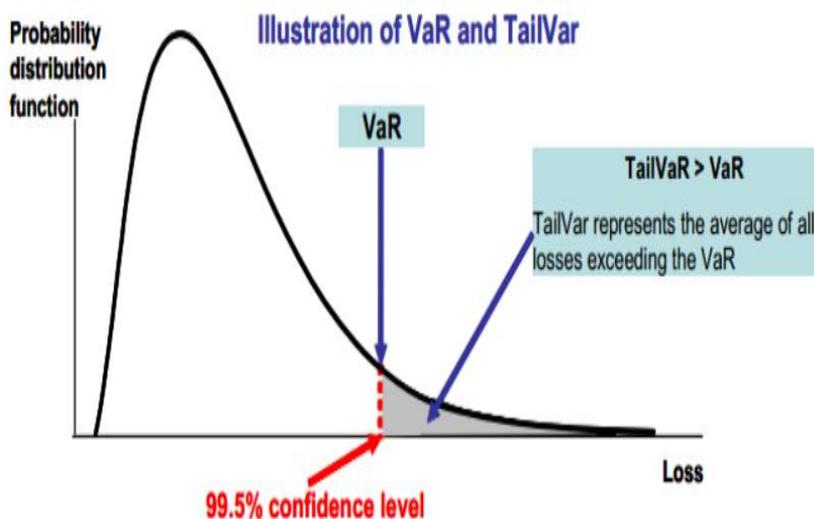


Figura 3.3: VaR e Tail-VaR a confronto.

L'area evidenziata in grigio rappresenta l'intera coda della distribuzione di cui il Tail-VaR ne è il valor medio; il VaR, invece, è solo un particolare quantile della distribuzione. Appare evidente che il Tail-VaR è una misura di rischio che meglio rispecchia il profilo di rischio di una compagnia, proprio perché tiene conto di tutte le situazioni estreme che possono manifestarsi.

### Relazione asintotica tra VaR ed Expected Shortfall

In questo paragrafo riportiamo un'importante relazione asintotica che lega i concetti di Value-at-Risk ed Expected Shortfall. Nello specifico, se  $N$  indica la frequenza sinistri ed  $X_F$  è l'insieme delle variabili aleatorie che hanno distribuzione  $F$ , abbiamo

$$\lim_{N \rightarrow \infty} \frac{\sup_{Z \subset X_F} VaR_\alpha(X)}{N} = ES_\alpha(Z) \quad (3.5)$$

dove  $\alpha \in (0, 1)$ ,  $Z$  è il costo del singolo sinistro ed  $X$  è il costo aggregato che, ovviamente, dipende dalla frequenza sinistri  $N$ . Tale risultato può anche essere riscritto come

$$\lim_{N \rightarrow \infty} \frac{\sup_{Z \subset X_F} \text{VaR}_\alpha(X)}{\sup_{Z \subset X_F} \text{ES}_\alpha(X)} = 1 \quad (3.6)$$

noto come *asymptotic equivalence* tra VaR ed ES (si veda Liu & Wang, 2017).

Tale equivalenza, di fatto, permette di ottenere una stima del VaR del costo aggregato  $X$  semplicemente calcolando l'ES del costo  $Z$  del singolo sinistro. Tuttavia, è necessario che sia soddisfatta la stringente condizione sulla frequenza sinistri: se la distribuzione di  $N$  ha parametro  $k$ , allora deve essere  $N(k)/k \rightarrow 1$  per  $k \rightarrow \infty$ .

### 3.2.3 Misure di rischio distorte

Le differenti tipologie di misure di rischio introdotte nel paragrafo precedente si basano sull'analisi del comportamento della variabile  $X$  quando assume valori maggiori del VaR. Lo stesso Tail-VaR, infatti, sebbene soddisfi l'assioma di coerenza, riflette unicamente i sinistri di entità superiore al VaR causando una possibile mancanza di incentivo in caso di mitigazione dei sinistri inferiori al VaR. Wang (1998) sostiene dunque che un'adeguata misura di rischio debba andare oltre l'assioma di coerenza e sulla base di questa considerazione introduce una famiglia di misure di rischio coerenti basate su funzioni di distorsione.

Si dimostra, infatti, che data una qualsiasi funzione  $g : [0, 1][0, 1]$  crescente e tale che  $g(0) = 0$  e  $g(1) = 1$ , allora la trasformata  $F^*(X) = g(F(X))$  definisce una distribuzione di probabilità distorta e  $g$  viene definita funzione di distorsione.

Viene così facilmente definita una famiglia di misure di rischio distorte nel modo seguente:  $p(X) = E^*(X) = -\int_{-\infty}^0 g(F(x))dx + \int_0^{\infty} [1 - g(F(x))]dx$ ; tale misura risulta inoltre coerente nel caso in cui  $g$  è una funzione continua.

Oltre alla revisione delle misure classiche, come VaR e Tail-VaR, riscritte in funzione delle funzioni di distorsione, sono state proposte alcune misure innovative basate su diverse funzioni  $g$ . Fra queste occorre segnalare le funzioni di distorsione basate sulla famiglia Beta proposte da Wirch & Hardy (2001) e la Wang Transform (WT).

Quest'ultima, basata sulla definizione della seguente funzione di distorsione  $g : g(u) = \phi[\phi^{-1}(u) - \lambda]$  con  $\phi$  ripartizione della Normale standard, permette di ottenere una nuova misura di rischio coerente. Data una variabile  $X$  con funzione di ripartizione  $F$ , è possibile infatti definire la nuova misura di rischio basata sulla Wang Transform nel modo seguente:

- per un livello di confidenza prefissato  $\alpha$  calcolare  $\lambda = \phi^{-1}(\alpha)$ ;
- applicare la trasformata di Wang:  $F^*(X) = \phi[\phi^{-1}(F(X)) - \lambda]$ ;
- calcolare il valore atteso  $p(X) = E^*(X)$  utilizzando la distribuzione distorta.

Nel caso la  $X$  abbia una distribuzione Normale, la misura WT assume valore identico al VaR, viceversa nel caso la distribuzione della  $X$  non sia Normale, WT può corrispondere ad un percentile più elevato o più basso di  $\alpha$ . Si dimostra inoltre che nel caso di una distribuzione Lognormale, la misura WT viene calcolata semplicemente nel modo seguente:  $WT(\alpha) = \exp(\mu + \lambda\sigma + \sigma^2/2)$  con  $\lambda = \phi^{-1}(\alpha)$ . In questo caso, come si può osservare dalla figura seguente, l'utilizzo della misura WT prevede un requisito di capitale superiore al VaR e molto simile a quello previsto dall'ES.

I grafici seguenti mostrano inoltre la sensibilità della misura WT al variare dei primi due cumulanti della Lognormale ed al variare del livello di

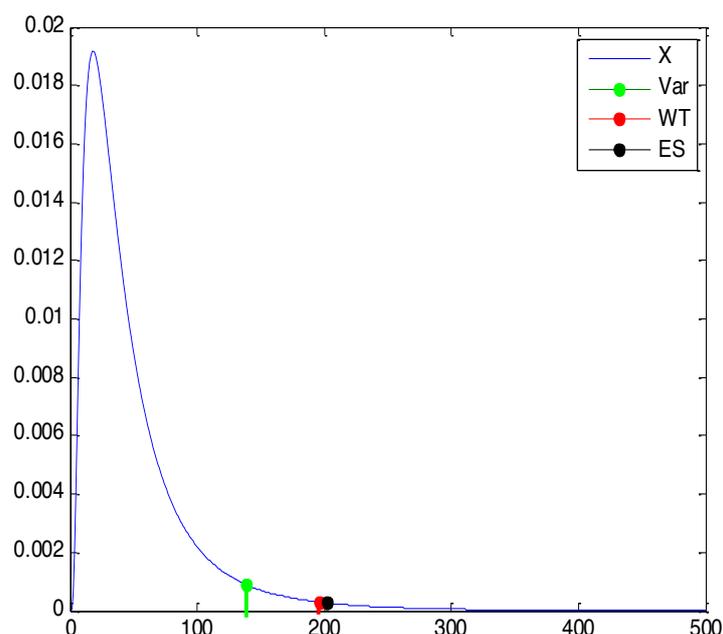


Figura 3.4: Confronto tra tre misure di rischio (VaR, WT e ES) ad un livello di  $\alpha$  pari a 0,95 per una variabile  $X$  distribuita secondo una Lognormale con coefficiente di variabilità pari 1 e media pari a 50.

confidenza  $\alpha$  prefissato. Si può osservare come la misura WT risulti particolarmente sensibile alla variazione dello scarto quadratico medio della distribuzione in esame. Infatti, se per i livelli più bassi di variabilità la WT mostra un requisito di capitale molto simile a quello ottenuto col VaR (con media e scarto quadratico pari a 20 si ha 27,81 per il VaR contro i 39,33 per la WT e 40,5 per l'ES), per i livelli di variabilità più elevati si ottiene un requisito di capitale addirittura doppio rispetto a quello ottenuto utilizzando il VaR (con media pari a 50 e scarto quadratico pari a 100 si ha 180,19 per il VaR, 332,87 per l'ES e 402,92 per la WT). Inoltre, non solo la WT cresce in modo significativo al crescere del coefficiente di variabilità, ma raggiunge livelli elevatissimi quando le variabilità elevate coincidono con livelli di confidenza particolarmente estremi (0,99 e successivi).

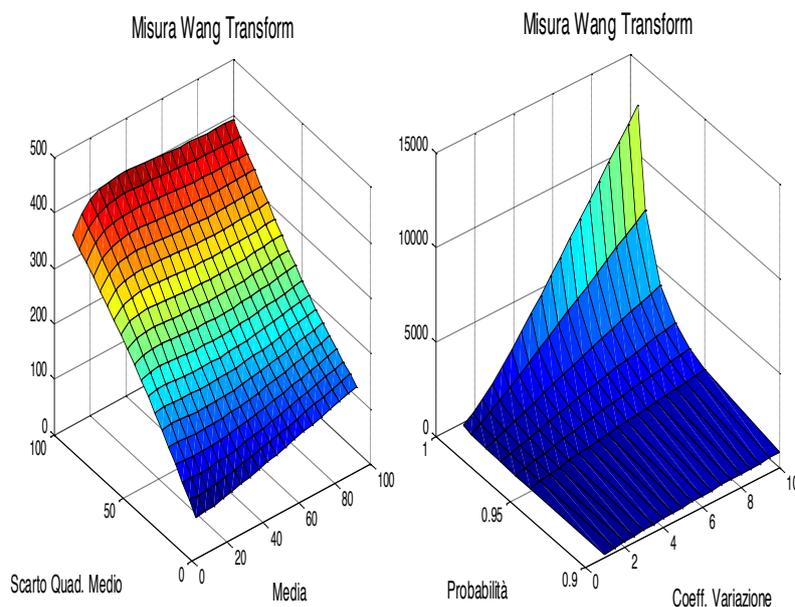


Figura 3.5: Misura Wang Transform al 95° percentile per una distribuzione Lognormale al variare di media e scarto quadratico medio, ed al variare del livello di confidenza e del coefficiente di variazione (media =50).

### 3.2.4 Misure di rischio spettrali

Uno sviluppo nel campo delle misure di rischio è rappresentato dall'introduzione delle misure spettrali. La caratteristica distintiva di questo tipo di misure è rappresentata dal collegamento esplicito con la funzione di avversione al rischio. L'idea base è fondata sull'estensione del concetto che sta alla base dell'Expected Shortfall, infatti come quest'ultimo può essere interpretato come la media dei  $(1-\alpha)\%$  peggiori sinistri o perdite di portafoglio, analogamente si può supporre l'esistenza di altre misure di rischio coerenti calcolate attraverso la media ponderata dei peggiori sinistri. Definendo così una classe di misure di rischio  $M_\phi$ , pari alla media ponderata dei quantili della distribuzione e fissando il livello di confidenza  $\alpha$ , si ottiene la seguente misura di rischio spettrale:

$$M_\phi(X) = \int_0^1 \phi(\alpha) q_\alpha d\alpha \quad (3.7)$$

dove  $q_\alpha$  rappresenta il quantile della distribuzione ad un livello  $\alpha$  e  $\phi$  prende il nome di risk spectrum.

Si può osservare che l'ES ed il VaR sono anch'essi delle particolari tipologie di misure spettrali: l'ES infatti non è altro che la media dei quantili con livello di confidenza superiore o pari ad  $\alpha$  tutti con pesi identici, mentre il VaR può essere interpretato come un caso degenere in quanto attribuisce peso pari a 1 al quantile di livello  $\alpha$  e peso pari a 0 a tutti gli altri quantili. La funzione  $\phi$  viene generalmente interpretata come funzione di avversione al rischio in quanto ogni investitore razionale potrebbe ipotizzare un diverso profilo della funzione. In realtà si dimostra che per mantenere la coerenza della misura al rischio il peso deve essere scelto opportunamente in modo da soddisfare i seguenti vincoli:

- non negatività:  $\phi(p)$  può assumere valori maggiori o uguali a zero;
- normalizzazione: la somma delle probabilità pesate  $\phi(p)$  deve essere pari a 1;
- pesi crescenti: se una probabilità  $\alpha_2$  è maggiore della probabilità  $\alpha_1$  allora  $\phi(\alpha_2)$  deve assumere un valore maggiore o uguale a  $\phi(\alpha_1)$ .

Le prime due condizioni appaiono ovvie, ma la terza assume particolare importanza. Tale condizione riflette infatti il concetto di avversione al rischio imponendo che sinistri di entità più rilevante devono essere associati a probabilità pesate non inferiori ai sinistri di entità inferiore. Il tasso di crescita sarà strettamente correlato al grado di avversione: se un decisore ha una debole avversione al rischio, allora i pesi cresceranno lentamente al crescere dei sinistri, viceversa maggiore sarà l'avversione al rischio e maggiore sarà

la crescita dei pesi al crescere dell'ammontare dei sinistri. Di conseguenza, secondo questo ragionamento, l'utilizzo dell'ES porta ad ipotizzare una situazione di neutralità al rischio nei confronti dei sinistri superiori al VaR. Esso, infine, viola la proprietà di incremento dei pesi in quanto attribuisce peso 0 a quantili con probabilità superiore ad  $\alpha$ .

Per ottenere una misura di rischio spettrale occorre, dunque, formulare un'ipotesi in merito alla forma della funzione di avversione al rischio. Tale scelta risulta chiaramente soggettiva in quanto dipende dall'avversione al rischio dell'investitore ma una semplice formulazione può essere rappresentata dalla seguente funzione esponenziale di avversione al rischio:

$$\phi(\alpha) = \frac{\exp[-(1-p)/\lambda]}{\lambda(1 - \exp^{-1/\lambda})} \quad (3.8)$$

dove il parametro positivo  $\lambda$  riflette il grado di avversione al rischio: minore è il valore di  $\lambda$ , maggiore è il grado di avversione.

### 3.3 Commenti

La letteratura relativa alle misure di rischio è molto florida e con l'introduzione del concetto di coerenza e la dimostrazione della violazione della subadditività da parte del VaR, ci si è indirizzati alla ricerca di misure di rischio che da un lato soddisfacessero i requisiti di coerenza descritti in precedenza, dall'altro lato fossero in grado di meglio considerare l'intero comportamento delle distribuzioni delle variabili esaminate.

L'introduzione di misure condizionate al VaR (dal Tail-VaR all'Expected Shortfall), molte delle quali apparse con nomi diversi ma coincidenti in caso di distribuzioni continue senza discontinuità, vanno lette proprio nel senso della ricerca di coerenza e di analisi anche della coda della distribuzione. Mentre i principali organismi di sorveglianza hanno optato su una delle mi-

sure tradizionali, ovvero il VaR, per individuare l'idoneo requisito di capitale, la letteratura attuariale e finanziaria ha invece allargato la schiera delle misure di rischio attraverso i concetti di distorsione e di misure spettrali. Mentre le prime si muovono sulla scia del Tail-VaR e dell'Expected Shortfall e hanno il principale obiettivo di considerare anche i sinistri inferiori al VaR, le seconde, creando un importante legame tra misura di rischio e funzione di avversione, introducono un importante e delicato concetto di soggettività nelle valutazioni. Utilizzando una misura spettrale infatti, due individui con identico portafoglio ed identica funzione pesi, potrebbero ottenere un diverso requisito di capitale funzione della propria avversione al rischio. Questa innovazione, sicuramente in linea con il concetto soggettivo di rischio, potrebbe invece rivelarsi particolarmente difficoltosa e problematica in sede di valutazione e di comparazione dei rischi.

# Capitolo 4

## Applicazioni a dati reali

Nel presente Capitolo mostriamo i risultati delle tecniche precedentemente esposte in relazione a due distinti dataset; uno è il dataset contenente i costi sostenuti per la garanzia incendio nel mercato danese, raccolti da Copenaghen Re, l'altro invece è un dataset multigaranzia riguardante coperture assicurative legate agli immobili.

### 4.1 Dataset monogaranzia

In letteratura è possibile trovare diversi lavori in cui i dati sui costi per incendio avvenuti in Danimarca (raccolti dalla compagnia di riassicurazione Copenaghen Re) vengono analizzati e modellizzati attraverso approcci parametrici; per esempio, Cooray & Ananda (2005), Scollnik (2007), Teodorescu & Vernic (2009 e 2013), Pigeon & Denuit (2011) e Nadarajah & Bakar (2014) mostrano che alcuni modelli composti, con soglia fissa o variabile, ben si adattano alle fire losses danesi. Esistono anche altri approcci di modellizzazione basati sull'impiego della distribuzione Burr (Burnecki et al., 2004), Hybrid Pareto (Carreau & Bengio, 2009), beta kernel quantile estimation (Charpentier & Oulidi, 2010), processi Poisson composti bivariati

(Esmacili & Kluppelberg, 2010). Esistono anche esempi di modellizzazione non parametrica, come quello mostrato in Guillotte et al. (2011).

Nella tabella che segue riportiamo alcune statistiche descrittive dei costi e delle frequenze del dataset danese, rimandando ad Embrechts et al. (1997) per una review più dettagliata:

Statistica	Costo (€)	Frequenza
Min	27.730	154
Media	613.100	299
Mediana	327.000	310
Q3	532.800	381
Max	55.240.000	447
Curtosi	505,592	-
Varianza	-	8.482
Asimmetria	17,635	-0,12

Per quanto riguarda i costi, osserviamo un valore massimo pari a circa 55 milioni di euro ed un valor medio di circa 613.000€; la distribuzione empirica, inoltre, risulta decisamente leptocurtica e con asimmetria positiva. Per le frequenze, invece, osserviamo una leggera asimmetria negativa ed un valor medio minore della varianza.

Portiamo all'attenzione del lettore che, per evitare effetti stagionali dovuti all'utilizzo dell'intera serie storica che parte dal 1980, i costi sono stati inflazionati al 2002, anno più recente per cui si hanno osservazioni.

In tale sezione riportiamo quanto ottenuto usando i modelli composti, le funzioni Copula e la FFT esposti in precedenza. Separiamo i costi al di sotto di una certa soglia (sinistri attritional) da quelli che la superano (sinistri estremi) e li supponiamo, inizialmente, indipendenti tra loro investigando l'adattabilità di alcuni modelli composti a soglia fissa o variabile. Successivamente, abbandoneremo l'ipotesi di indipendenza e cercheremo di

stimare un modello che tenga conto della eventuale dipendenza tra di essi, misurando l'impatto che questa può avere sulla stima del Value-at-Risk del costo aggregato.

### 4.1.1 Modelli composti

In questo paragrafo mostriamo i risultati a cui siamo giunti impiegando alcuni modelli composti sui dati danesi; come anticipato, per le composizioni a soglia fissa facciamo riferimento alla metodologia descritta in Teodorescu & Vernic (2009 and 2013), avendo ottenuto un valore per  $u$  pari a 1.022.125€, attraverso l'impiego del metodo POT.

Iniziamo con un Lognormale-KumPareto; dalla tabella seguente possiamo confrontare alcuni quantili empirici con i corrispondenti teorici:

Livello	50%	75%	90%	95%	99%	99,5%
Quantile empirico	327.016	532.757	1.022.213	1.675.219	5.484.150	8.216.877
Quantile teorico	333.477	462.852	642.196	840.161	2.616.338	4.453.476

Tabella 4.1: Confronto tra quantili empirici e quantili della Lognormale-KumPareto.

Solo il 50-esimo percentile della distribuzione teorica è molto prossimo a quello osservato, poi le differenze crescono nettamente. La figura seguente mostra un confronto tra le sole code destre delle distribuzioni (empirica e teorica):

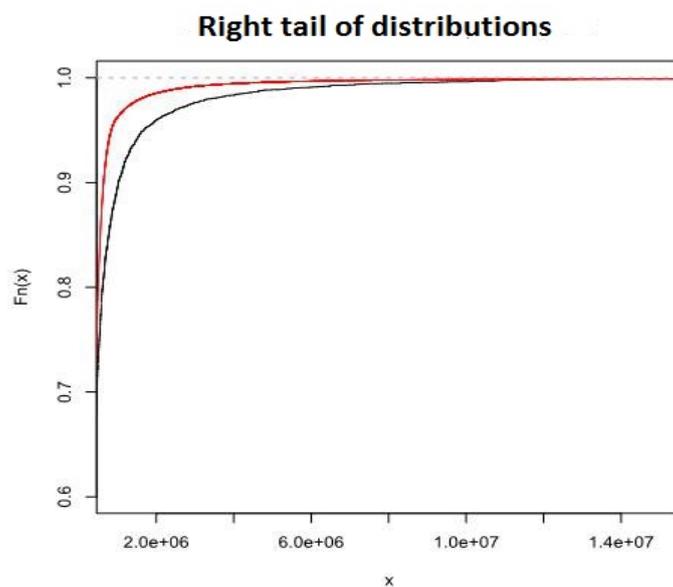


Figura 4.1: Code destre delle distribuzioni Lognormale-KumPareto (linea rossa) ed empirica (linea nera).

La linea rossa giace sempre sopra quella nera, pertanto la Lognormale-KumPareto sottostima la coda destra della distribuzione osservata.

Proseguiamo mostrando ora l'adattamento di un Lognormale-Burr, come suggerito in Nadarajah and Bakar (2014). La tabella seguente confronta alcuni percentili teorici ed empirici:

Livello	50%	75%	90%	95%	99%	99,5%
Quantile empirico	327.016	532.757	1.022.213	1.675.219	5.484.150	8.216.877
Quantile teorico	199.681	332.341	634.531	1.029.262	3.189.937	5.181.894

Tabella 4.2: Confronto tra quantili empirici e quelli della Lognormale-Burr.

Il modello sembra essere maggiormente adattabile alla coda destra empirica rispetto al precedente Lognormale-KumPareto, come possiamo notare anche dalla seguente:

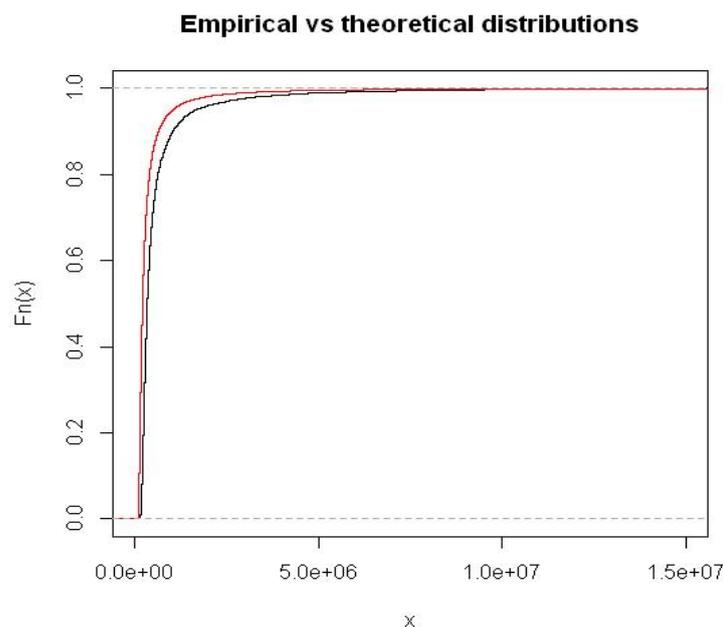


Figura 4.2: Funzioni di ripartizione Lognormale-Burr ed empirica (linee rossa e nera).

Come la Lognormale-KumPareto, anche la Lognormale-Burr sta sempre sopra quella empirica ma non sempre alle stesse distanze.

Proseguiamo con un modello composto da una Lognormale e da una distribuzione Generalizzata di Pareto (GPD), di cui ne riportiamo i parametri stimati e il QQ-plot:

	estremo inferiore	best estimate	estremo superiore
$\mu_1$	12,82	12,84	12,86
$\sigma$	0,59	0,61	0,62
$\sigma_\mu$	1.113.916	1.115.267	1.116.617
$\xi$	0,33	0,45	0,56

Tabella 4.3: Stime dei parametri della Lognormale-GPD

$\mu_1$  e  $\sigma$  sono i parametri della Lognormale, mentre  $\sigma_\mu$  e  $\xi$  sono quelli della GPD.

Osserviamo che tale modello composto ha un buono adattamento ai dati; infatti, eccetto alcuni, i percentili teorici sono molto simili ai corrispondenti

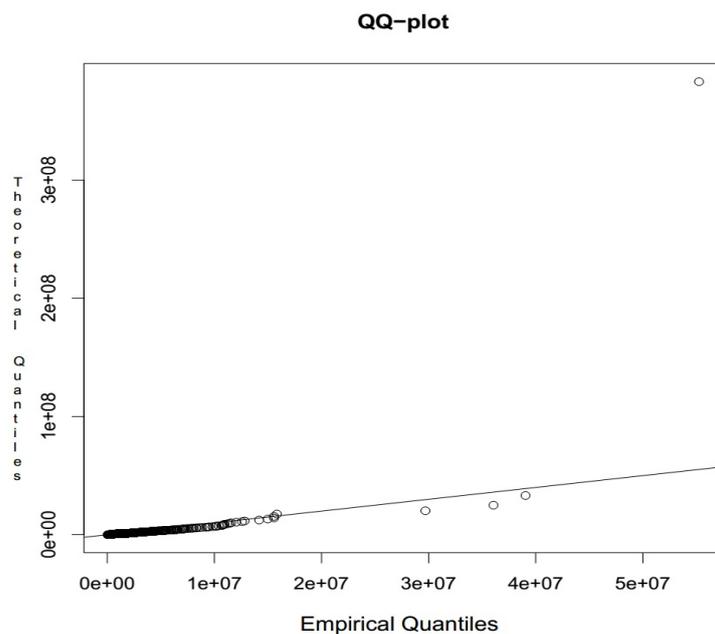


Figura 4.3: QQ-plot distribuzione empirica e distribuzione Lognormale-GPD.

percentili empirici. Nella figura seguente compariamo le funzioni di densità troncate superiormente e le code destre:

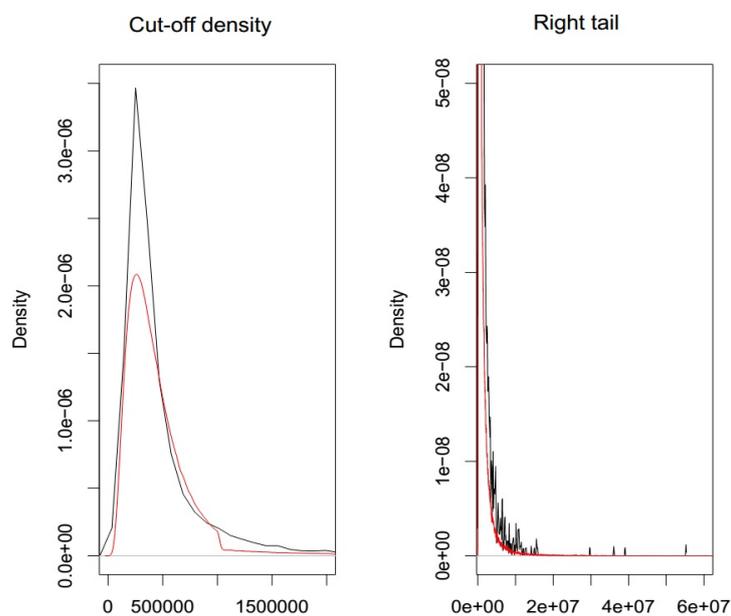


Figura 4.4: A sinistra, un confronto tra le densità troncate. A destra, invece, le corrispondenti code destre empirica e teorica (linea rossa).

Il modello presenta, inoltre, una non trascurabile coda destra (indice di curtosi pari a 115.656) che può essere analizzata anche osservando la ripartizione empirica e quella teorica stimata (si veda figura seguente).

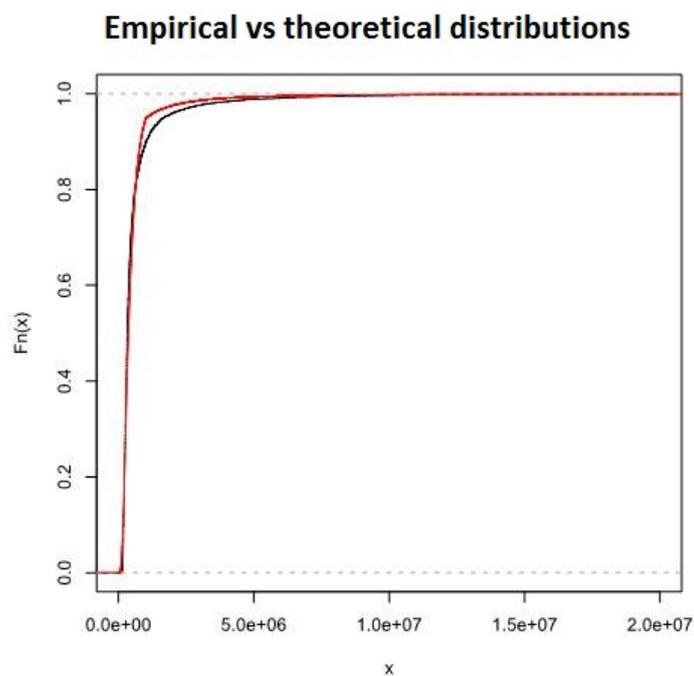


Figura 4.5: Funzione di ripartizione del Lognormale-GPD (linea rossa) e quella empirica (linea nera).

Il corrispondente test di Kolmogorov-Smirnov ha restituito un p-value pari a 0,8590423, ottenuto usando 50.000 campioni bootstrap.

Mostriamo ora il modello composto Lognormale-GPD con soglia aleatoria distribuita secondo una Gamma (1.45) di cui ne riportiamo le stime puntuali e gli intervalli di confidenza al 99%:

	estremo inferiore	best estimate	estremo superiore
$\mu_1$	12,78	12,79	12,81
$\sigma$	0,52	0,54	0,55
$u$ (soglia)	629.416	630.768	632.121
$\sigma_\mu$	1.113.915	1.115.266	1.116.616
$\xi$	0,22	0,29	0,37

Tabella 4.4: Stime puntuali e intervalli di confidenza al 99% della distribuzione Lognormale-GPD-Gamma.

La soglia  $u$  non è una costante ed il suo valore dipende ora dai parametri della distribuzione Gamma. Di seguito riportiamo un confronto tra i percentili empirici e teorici:

Livelli	50%	75%	90%	95%	99%	99,5%
Quantile empirico	327.016	532.757	1.022.213	1.675.219	5.484.150	8.216.877
Quantile teorico	360.574	517.996	1.103.309	2.077.792	5.266.116	7.149.253

Tabella 4.5: Confronto tra quantili empirici e quelli della Lognormale-GPD-Gamma.

Dalla seguente figura possiamo osserva come il modello composto Lognormale-GPD-Gamma possa essere considerato adatto ai dati oggetto di analisi:

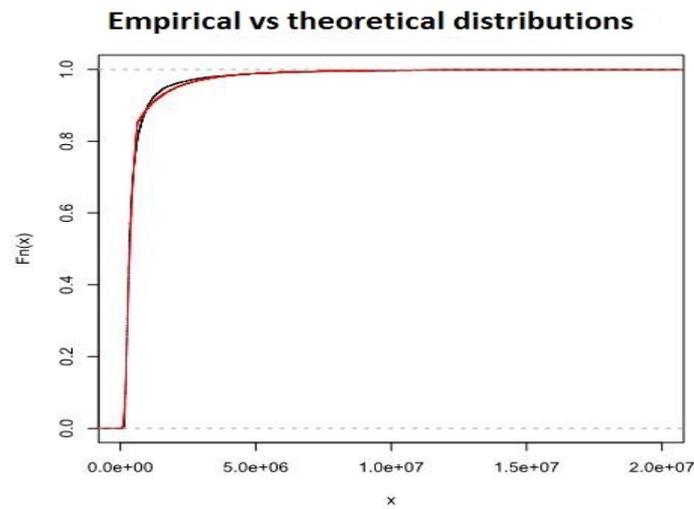


Figura 4.6: Funzioni di ripartizione Lognormale-GPD-Gamma (linea rossa) ed empirica (nera).

Il test di adattamento di Kolmogorov-Smirnov ha un p-value pari a 0,1971361; pertanto, non possiamo rigettare l'ipotesi nulla secondo cui il modello è adattabile ai dati danesi.

Concludiamo il paragrafo con un confronto, in termini di indici AIC e BIC, dei modelli visti finora, ovvero Lognormale-KumPareto, Lognormale-Burr, Lognormale-GPD con soglia fissa ed Lognormale-GPD con soglia di tipo Gamma:

Indice	KumPareto	Burr	GPD	GPD-Gamma
AIC	193.374	191.459	191.172	190.834
BIC	193.409	191.494	191.207	190.882

Tabella 4.6: Indici AIC and BIC per un confronto tra differenti modelli.

Dalle precedenti analisi e dalle indicazioni AIC e BIC, possiamo considerare il modello Lognormale-GPD-Gamma la migliore alternativa analizzata finora per i dati danesi.

### 4.1.2 Funzioni Copula

Nel precedente paragrafo abbiamo ristretto la nostra analisi al caso di indipendenza tra i sinistri attritional e large del dataset danese. Ora proviamo ad analizzare un'eventuale struttura di dipendenza tra essi; per prima cosa, definiremo un modello composto la cui definizione prevede l'impiego di una funzione Copula.

Abbiamo già detto che alcuni indici di dipendenza, come il  $\tau_K$  di Kendall o il  $\rho_S$  di Spearman, sono in grado di fornire informazioni circa le caratteristiche di cui deve essere dotata la Copula, nonché informazioni riguardanti i parametri della Copula stessa. Supponendo di avere una matrice  $(n \times d)$  la cui generica colonna è formata dalle  $n$  osservazioni di una v.a.  $X_i = (x_{1,i}, \dots, x_{n,i})'$  per  $i = 1, 2, \dots, d$ , presa una coppia di v.a.  $(X_i, X_j)$ ,

riportiamo lo stimatore di  $\tau_K$  (si veda McNeil et al., 2005):

$$\hat{\tau}_K(X_i, X_j) = \binom{n}{2}^{-1} \sum_{1 \leq t < s \leq n} \text{sign}((x_{t,i} - x_{s,i})(x_{t,j} - x_{s,j})) \quad (4.1)$$

Calcolando tale quantità per ogni coppia  $(i, j)$  di v.a., otteniamo una matrice  $R^K$  semidefinita positiva.

Per alcune copule esiste una precisa relazione tra  $\tau_K$  e il parametro della Copula stessa; in particolare, per le copule di Gumbel, Clayton e Frank valgono le seguenti:

Copula	$\tau_K$
$C_\theta^{Gu}$	$1 - 1/\theta$
$C_\theta^{Cl}$	$\theta/(\theta + 2)$
$C_\theta^{Fr}$	$1 - 4\theta^{-1}(1 - D_1(\theta))$

dove  $D_1(\theta) = \frac{1}{\theta} \int_0^\theta t/(\exp(t) - 1) dt$ . Da queste relazioni è banale individuare il valore del parametro; tuttavia, non è detto che una di queste copule sia adatta per descrivere la dipendenza tra le distribuzioni marginali e pertanto dovremo, prima di tutto, capire quale tipo di Copula impiegare.

Di seguito mostriamo, invece, lo stimatore di  $\rho_S$ :

$$\hat{\rho}_S(X_i, X_j) = \frac{12}{n(n^2 - 1)} \sum_{t=1}^n \left( \text{rank}(x_{t,i}) - 0.5(n+1) \right) \left( \text{rank}(x_{t,j}) - 0.5(n+1) \right) \quad (4.2)$$

dove con  $\text{rank}(x_{t,i})$  si intende la posizione, all'interno del campione ordinato, della t-esima osservazione  $x_{t,i}$ . Anche stavolta, tale stimatore va calcolato per ogni coppia  $(i, j)$  di v.a.; il risultato sarà una matrice  $R^S$  semidefinita positiva.

Procediamo descrivendo la fase di stima di una Copula parametrica attraverso l'impiego della massima verosimiglianza; nello specifico, è possibile scomporre tale fase in tre parti:

- stima delle distribuzioni marginali;
- generazione di “pseudo-osservazioni” della Copula;
- stima dei parametri della Copula attraverso la massima verosimiglianza e lo pseudo-campione generato al passaggio precedente.

Per quanto riguarda la stima delle distribuzioni marginali, è possibile procedere come se queste fossero del tutto indipendenti, impiegando il metodo di stima che si preferisce (massima verosimiglianza, momenti, ecc.).

Le pseudo-osservazioni della Copula coincidono, invece, con le funzioni di ripartizione marginali. Formalmente, siano  $\hat{F}_1, \dots, \hat{F}_d$  le funzioni di ripartizione stimate per le  $d$  v.a. marginali; il campione delle pseudo-osservazioni della Copula è formato da  $d$  vettori colonna  $\hat{U}_1, \dots, \hat{U}_1, \dots, \hat{U}_d$ , dove  $\hat{U}_i = (\hat{F}_i(x_{i,1}), \dots, \hat{F}_i(x_{i,n}))$ , per  $i = 1, \dots, d$ , mentre  $(x_{i,1}, x_{i,2}, \dots, x_{i,n})$  è il campione che vogliamo modellizzare con la distribuzione della  $i$ -esima v.a.

Il terzo ed ultimo passaggio consiste nella ottimizzazione della funzione di log-verosimiglianza della Copula di interesse; indicando quest'ultima con  $C_\theta$ , dove  $\theta$  è il vettore dei parametri, la funzione da massimizzare numericamente rispetto a  $\theta$  è

$$\ln L(\theta; \hat{U}_1, \dots, \hat{U}_n) = \sum_{i=1}^n \ln c_\theta(\hat{U}_1, \dots, \hat{U}_n) \quad (4.3)$$

dove  $c_\theta(\hat{U}_1, \dots, \hat{U}_n)$  è la funzione di densità della Copula  $C_\theta$  (per i dettagli si rimanda al paragrafo 5.1.5 di McNeil et al., 2005). Il primo e il terzo passaggio costituiscono quel procedimento che prende il nome di inference functions for margins (IFM, si veda Joe e Xu, 1996).

Può essere interessante esaminare alcuni metodi di costruzione che permettano di giungere alla definizione di numerose funzioni Copula. Un primo metodo, denominato metodo di inversione, permette di determinare

la Copula attraverso l'inversione della funzione di ripartizione congiunta  $C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2))$ . Se le marginali sono distribuite secondo una esponenziale, tale metodo restituisce la famiglia di copule di Marshall-Olkin. Una seconda famiglia di metodi, definiti geometrici, utilizza le proprietà matematiche che caratterizzano le copule, anziché considerare il legame con le marginali. I metodi geometrici sfruttano le seguenti proprietà:

- somma convessa. La somma, dato che ogni combinazione lineare convessa di copule è una Copula, permette di ottenere la Copula  $C$  come somma convessa della famiglia  $C_\theta$  rispetto alla mixing distribution  $\Lambda := C(u_1, u_2) = \int_{\mathfrak{R}} C_\theta(u_1, u_2) d\Lambda_\theta$ ;
- copule con sezioni verticali o orizzontali. Si possono definire famiglie di copule imponendo specifiche sezioni lineari  $C(u_1, u_2) = a(u_2)u_1 + b(u_2)$ , quadratiche  $C(u_1, u_2) = a(u_2)u_1^2 + b(u_2)u_1 + c(u_2)$ , ecc.

Un'ultima possibilità di costruzione è rappresentata dall'utilizzo di metodi algebrici che usano le relazioni esistenti tra la ripartizione e le marginali. Alla categoria di copule ottenibili attraverso tale metodo appartiene, per esempio, la Copula di Plackett (2.21).

Spostiamo ora l'attenzione sui costi danesi cercando di coglierne l'eventuale dipendenza; come distribuzioni marginali faremo riferimento alla Lognormale per i sinistri attritional e alla GPD per quelli estremi. Di seguito riportiamo la matrice di correlazione empirica  $\mathbf{R}$ :

$$\mathbf{R} = \begin{pmatrix} 1 & 0,01259155 \\ 0,01259155 & 1 \end{pmatrix}$$

e le misure di associazione  $\tau_K$  di Kendall ed il  $\rho_S$  di Spearman:

$$\mathbf{K} = \begin{pmatrix} 1 & 0,002526672 \\ 0,002526672 & 1 \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} 1 & 0,003730766 \\ 0,003730766 & 1 \end{pmatrix}$$

che suggeriscono una debole ma positiva dipendenza tra le due tipologie di sinistri.

Per tale motivo, individuare un'appropriata Copula non sarà semplice. Precisiamo inoltre che, per i dati danesi, una struttura di dipendenza empirica è indotta dalla distinzione che facciamo tra i sinistri attritional e sinistri estremi; infatti, non c'è un unico evento che causa simultaneamente sia una piccola perdita sia una estrema, ma il verificarsi dell'evento assicurato produce o una perdita al di sotto della soglia o una al di sopra. Pertanto, i risultati mostrati nel prosieguo devono essere considerati come un mero esercizio che sottolinea gli effetti importanti che la dipendenza può produrre sulla distribuzione del costo sinistri aggregato.

Di seguito riportiamo i parametri della Copula Gumbel stimati attraverso differenti metodi sulla base dei dati a disposizione:

Metodo	$\hat{\theta}$	Errore standard
Maximum pseudo-likelihood	1,11	0,008
Canonical maximum pseudo-likelihood	1,11	0,008
Simulated maximum likelihood	1,11	-
Minimum distance	1,09	-
Moments based on Kendall's tau	1,13	-
Bootstrap	1,11	0,008

Tabella 4.7: Differenti metodi per la stima dei parametri della Copula Gumbel.

Ricordiamo che il parametro  $\theta$  della Gumbel assume valori in  $[1, \infty)$  e per  $\theta \rightarrow 1$  tende alla Copula Indipendenza. Osserviamo che le stime sono significativamente diverse da 1 e pertanto la Copula Gumbel non coincide con quella di indipendenza. Possiamo affermarlo in quanto abbiamo verificato che, usando il campionamento bootstrap (si veda Efron & Tibshi-

rani, 1993), il parametro  $\theta$  ha una distribuzione Normale. Infatti, il test di Shapiro-Wilk ha restituito un p-value pari a 0,08551 e quindi, fissato un livello di significatività del 5%, non è possibile rigettare l'ipotesi nulla. Inoltre, l'intervallo di confidenza al 99% ottenuto con il metodo della massima pseudo-verosimiglianza è (1,090662; 1,131003) che non include, appunto, il valore 1; lo stesso intervallo ottenuto con la procedura bootstrap risulta (1,090662; 1,131003). Nella seguente figura riportiamo la distribuzione del parametro della Gumbel ottenuta con procedura bootstrap:

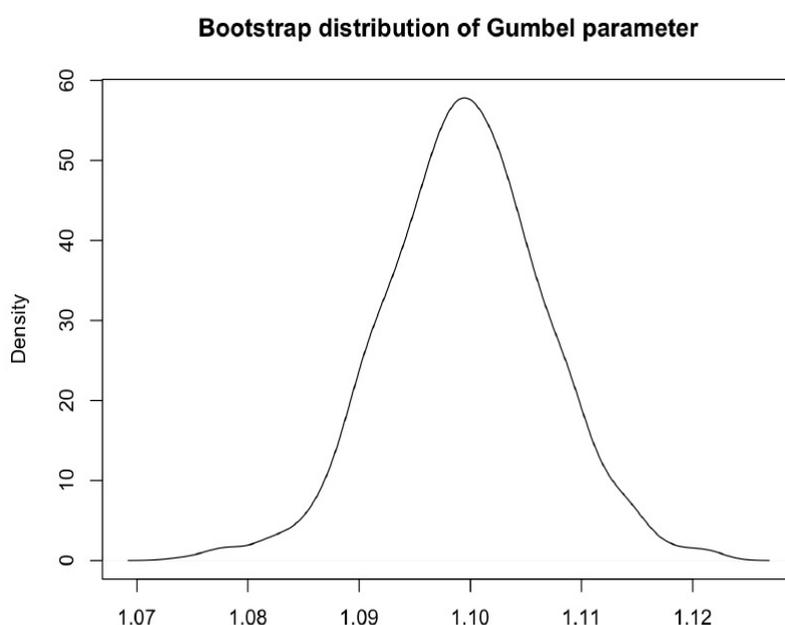


Figura 4.7: Distribuzione Normale del parametro Gumbel ottenuta dal campionamento bootstrap.

Riportiamo inoltre due utili grafici ottenuti dalla simulazione della Gumbel stimata.

La funzione di densità assume valori maggiori in corrispondenza di valori grandi per *entrambe* le distribuzioni marginali; in altre parole, impiegando tale Copula Gumbel, la probabilità che sinistri attritional producano una perdita molto prossima al valore soglia e che i sinistri large producano perdite estreme è maggiore della probabilità di un qualsiasi altro evento

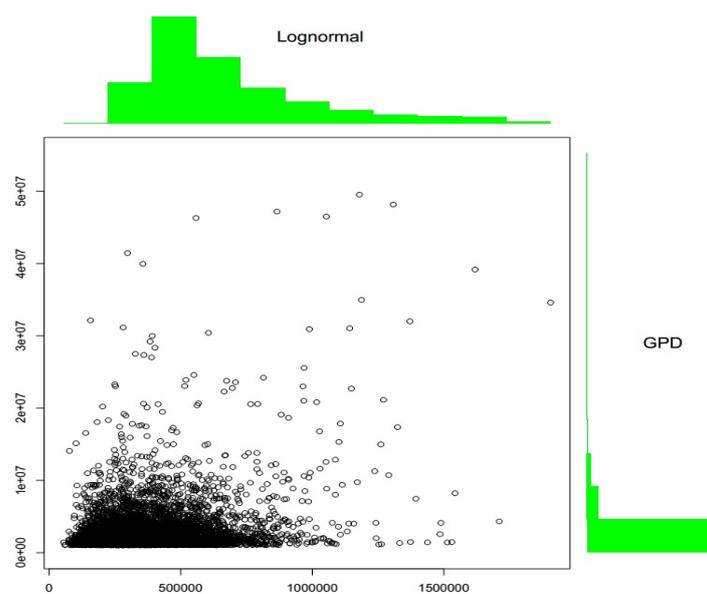


Figura 4.8: Istogrammi delle marginali Lognormale (sopra) e GPD (destra) e valori simulati della Gumbel stimata.

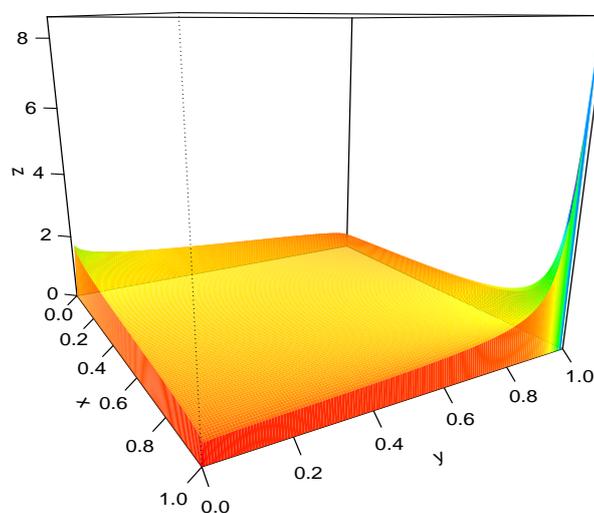


Figura 4.9: Funzione di densità della Gumbel stimata. Sinistri attrittional sull'asse X, sinistri large sull'asse Y.

congiunto.

Limitatamente all'analisi dei dati danesi, ci riferiremo alla sola Copula Gum-

bel nonostante siano state stimate e indagate altre funzioni Copula per le quali non rileviamo differenze significative.

### 4.1.3 Fast Fourier Transform

Considerando il fatto che non è semplice definire una appropriata Copula, ora cercheremo di modellizzare direttamente il costo sinistri aggregato con la Fast Fourier Transform (FFT) che può essere intesa come un'alternativa alle funzioni Copula viste in precedenza. Tuttavia, mentre con le copule siamo giunti all'individuazione di alcune possibili distribuzioni per il costo del singolo sinistro, ora con la FFT arriveremo, per la prima volta, direttamente alla distribuzione del costo sinistri aggregato. Un aspetto importante nell'utilizzo della procedura FFT, è che con essa è possibile superare una delle due ipotesi caratterizzanti l'approccio collettivo della TdR, ovvero quella della reciproca indipendenza tra i sinistri.

Inoltre, non richiede ipotesi circa la eventuale struttura di dipendenza tra i dati ma, essendo fortemente basata sulle evidenze empiriche, ne tiene comunque conto. In sostanza, la FFT può anche essere interpretata come sorta di Copula empirica ed essere considerata, pertanto, il **benchmark** per tutte le funzioni Copula teoriche.

Per costruire la distribuzione del costo aggregato attraverso la FFT è necessario, prima di tutto, rendere discreta la distribuzione  $Z$  della severity (si veda Klugman et al., 2010), ottenendo il vettore  $z = (z_0, \dots, z_{n-1})$  il cui generico elemento  $z_i$  è la probabilità che il singolo sinistro produca una perdita pari a  $ic$ , dove  $c$  è una costante tale che, data  $n$  la lunghezza del vettore  $z$ , la perdita  $cn$  ha una probabilità trascurabile. Considereremo anche la distribuzione della frequenza sinistri  $k$  tramite la funzione generatrice delle probabilità (FGP), definita dalla (2.40).

Per il caso bidimensionale è necessario avere come input la matrice  $M_Z$ ;

essa contiene le probabilità congiunte dei sinistri normali e large ed è tale che è possibile ottenere le corrispondenti distribuzioni marginali sommando lungo le righe e le colonne, rispettivamente. Per esempio, sia

$$\mathbf{M}_z = \begin{pmatrix} 0,5 & 0 & 0 \\ 0,2 & 0,25 & 0 \\ 0 & 0,05 & 0 \end{pmatrix}$$

tale matrice. Il vettore  $(0,5, 0,45, 0,05)$ , ottenuto sommando lungo le tre righe, contiene la distribuzione dei sinistri attritional, mentre il vettore  $(0,7, 0,3, 0)$ , ottenuto sommando lungo le tre colonne, contiene quella dei sinistri large. Il singolo elemento della matrice, invece, è la probabilità congiunta. La distribuzione del costo aggregato sarà la matrice  $M_X$  data dalla (2.41).

Per maggiori dettagli matematici, rimandiamo a Robe-Voinea & Vernic (2016) in cui la FFT è estesa anche al contesto multivariato e vengono illustrati diversi esempi numerici.

Discretizziamo la distribuzione impiegando la funzione *discretize* disponibile nel package R *actuar* (si veda Dutang et al., 2008). Tale discretizzazione ci permette di costruire la matrice  $M_Z$  a cui applicare la versione bidimensionale della FFT. In tal modo, abbiamo una nuova matrice  $FFT(M_Z)$  che agisce come input per la funzione generatrice delle probabilità della variabile aleatoria  $k$ .

Necessitiamo a questo punto di definire una distribuzione per la  $k$ . Riportiamo di seguito alcune statistiche descrittive delle frequenze empiriche:

Min	Max	Q1	Mean
154	447	238	299
Median	Q3	Variance	Skewness
310	381	8.482	-0,12

Tabella 4.8: Statistiche descrittive della distribuzione della frequenza sinistri.

Notiamo che il 50% delle frequenze sono incluse tra 238 e 381 sinistri e c'è una leggera asimmetria negativa. Inoltre, la varianza è maggiore del valor medio, pertanto è possibile supporre una distribuzione Binomiale Negativa per la frequenza sinistri; la corrispondente funzione generatrice delle probabilità è definita come

$$PGF(t) = \left( \frac{1-p}{1-pt} \right)^m \quad (4.4)$$

Ne abbiamo stimato i parametri ( $m = 5$  e  $p = 0,82$ ) e ottenuta la matrice  $PGF_k(FFT(\mathbf{M}_z))$ . Come ultimo step abbiamo applicato la IFFT il cui output è la matrice  $M_X$  che individua la distribuzione discretizzata del costo sinistri aggregato, mantenendo la distinzione tra sinistri attritional e large e, soprattutto, preservando la struttura di dipendenza.

#### 4.1.4 Risultati

Mostriamo di seguito i risultati ottenuti per il dataset monogaranzia (si veda Acri & Cerchiara, 2016). Ricordiamo che, tra le metodologie esposte, solo la FFT restituisce direttamente il costo aggregato; per tutte le altre (modelli composti e funzioni Copula), tenendo conto delle statistiche della frequenza sinistri osservata nei dati danesi, possiamo assumere una frequenza sinistri costante e pari a  $k = 300$ . Una singola simulazione del costo aggregato può essere ottenuta sommando le perdite di  $k$  singoli sinistri

e, ripetendo la procedura un numero elevato di volte (1.000.000 nel nostro caso), otteniamo la distribuzione del costo aggregato.

Nella tabella seguente, riportiamo i  $VaR_{0,995}$  prodotti dall'impiego dei modelli composti Lognormale-Burr, Lognormale-GPD-Gamma, Copula Gumbel ed FFT:

Modello	Frequenza Sinistri	$VaR_{99,5}$
Lognormale-Burr	300	205.727.356
Lognormale-GPD-Gamma	300	209.057.172
Copula Gumbel	300	649.006.035
FFT	Binomiale Negativa	703.601.564

Tabella 4.9: Stime del VaR al livello 99% con differenti modelli.

## 4.2 Dataset multigaranzia

In questa sezione poniamo l'attenzione su un altro dataset che raccoglie i costi sostenuti da una compagnia di assicurazione italiana dal 2006 al 2012 per la copertura di 4 distinte garanzie sugli immobili: incendio ed elementi naturali al fabbricato, danni ai beni dovuti a fenomeni elettrici, danni da acqua più spese di ricerca ed RC proprietà; di seguito ci riferiremo ad esse come garanzia 1, 2, 3 e 4, rispettivamente.

Nelle tabelle seguenti riportiamo i costi delle suddette garanzie e le frequenze di accadimento suddivisi per anno, oltre che alcune statistiche descrittive:

Anno / Costo	Garanzia 1	Garanzia 2	Garanzia 3	Garanzia 4
2006	59.293	54.927	333.215	10.278
2007	25.022	25.667	287.927	9.362
2008	37.582	32.012	244.877	16.369
2009	77.293	113.903	175.725	11.385
2010	35.718	2.647	120.465	44.032
2011	33.644	7.051	74.613	2.235
2012	7.031	4.714	81.988	4.800

Anno / Frequenza	Garanzia 1	Garanzia 2	Garanzia 3	Garanzia 4
2006	33	49	329	9
2007	20	31	283	15
2008	27	32	259	14
2009	26	25	183	13
2010	18	8	122	14
2011	6	9	74	4
2012	8	7	71	2

Statistica / Costo	Garanzia 1	Garanzia 2	Garanzia 3	Garanzia 4
Minimo	7.031	2.647	74.613	2.235
Media	39.369	34.417	188.401	14.066
Mediana	35.718	25.667	175.725	10.278
Q3	48.438	43.470	266.402	13.877
Massimo	77.293	113.903	333.215	44.032
Curtosi	0,345	2,568	1,751	4,788
Asimmetria	0,497	1,607	0,261	2,074

Statistica / Frequenza	Garanzia 1	Garanzia 2	Garanzia 3	Garanzia 4
Minimo	6	7	71	2
Q1	13	9	98	7
Media	20	23	189	10
Mediana	20	25	183	13
Q3	27	32	271	14
Massimo	33	49	329	15
Varianza	100	250	10.835	28
Asimmetria	-0,285	0,507	0,105	-0,828

Osserviamo delle distribuzioni leptocurtiche e con asimmetria positiva per i costi empirici, mentre le frequenze mostrano valori medi minori delle corrispondenti varianze.

### 4.2.1 Applicazione della Standard Formula

In questa sezione mostriamo un'applicazione della Standard Formula, come recepita dal Regolamento Delegato (UE) 2015/35 del 10 ottobre 2014 (di seguito Regolamento), sui dati di nostro interesse. Prima di mostrare quanto ottenuto, facciamo un breve richiamo di ciò che è previsto dalla normativa vigente Solvency 2 relativamente al calcolo del requisito patrimoniale di solvibilità; questo deve essere calcolato attraverso una scomposizione in moduli dei vari rischi a cui un'impresa assicurativa può essere esposta. In particolare, il SCR è ottenuto con un approccio bottom-up, ovvero è l'aggregazione di differenti risultati parziali, relativi ai diversi moduli di rischio considerati, come si può vedere dalla seguente figura:

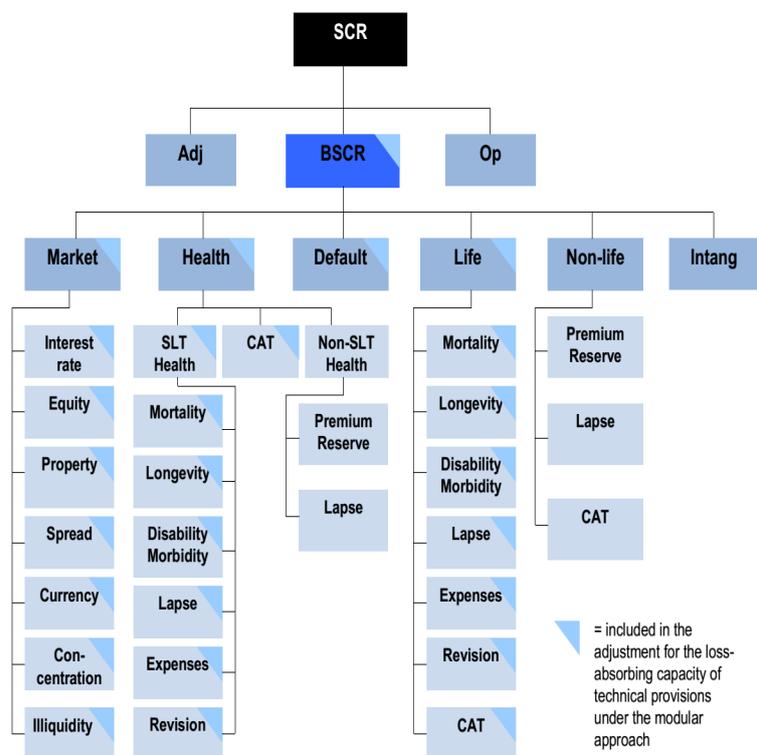


Figura 4.10: Schema modulare per il calcolo del SCR con formula standard.

Tali moduli sono costruiti sulla base della seguente classificazione dei rischi assicurativi:

- rischio sottoscrizione vita;
- rischio sottoscrizione danni;
- rischio sottoscrizione salute;
- rischio di mercato;
- rischio di default;
- rischi intangibili.

Come mostrato nella figura, i vari moduli di rischio sono scomposti in vari sub-moduli, ognuno facente riferimento ad un rischio specifico; per esempio, per il solo modulo di rischio sottoscrizione danni, i sub-moduli sono relativi a

- premium/reserve risk (rischio che i premi incassati non siano sufficienti a coprire le spese e i sinistri, ovvero è il rischio che il combined ratio sia  $> 100\%$ , ovvero che le riserve accumulate non siano sufficienti);
- rischio abbandono;
- rischio catastrofe (rischio di perdita, o incremento delle passività assicurative, causate dal manifestarsi di eventi estremi e non attesi).

Per il generico modulo, la quantificazione del rischio può seguire due differenti approcci:

- approccio factor based, ovvero un approccio basato sull'applicazione di un fattore ad una misura del rischio (Value-at-Risk, tra tutte); tale fattore è ottenuto tenendo conto della coda della distribuzione della v.a. considerata;
- approccio scenario based, ovvero il rischio è misurato confrontando il Net Asset Value (NAV, differenza tra attivi e passivi) sotto lo scenario best estimate (BE) e il NAV sotto uno scenario shock. Per scenario si intende l'insieme delle ipotesi circa lo sviluppo temporale della mortalità di una popolazione, del mercato in cui si opera e così via.

Il secondo approccio è tipico delle assicurazioni sulla vita, mentre il primo è quello a cui faremo riferimento in quanto generalmente utilizzato per portafogli danni e che fa uso dei seguenti coefficienti di correlazione

per l'aggregazione dei differenti rischi (per maggiori dettagli si rimanda al Regolamento):

i \ j	Market	Default	Life	Health	Non-life
Market	1				
Default	0.25	1			
Life	0.25	0.25	1		
Health	0.25	0.25	0.25	1	
Non-life	0.25	0.5	0	0	1

Figura 4.11: Coefficienti di correlazione tra i vari moduli di rischio.

Data la scomposizione dei moduli di rischio in sub-moduli, vengono impiegati altri coefficienti di correlazione per l'aggregazione di quei sub-rischi appartenenti alla stessa famiglia. Nella figura che segue riportiamo i coefficienti adottati per i rischi Non-Life:

$CorrNL$	$NL_{pr}$	$NL_{lapse}$	$NL_{CAT}$
$NL_{pr}$	1		
$NL_{lapse}$	0	1	
$NL_{CAT}$	0.25	0	1

Figura 4.12: Coefficienti di correlazione tra sub-moduli di rischio (modulo sottoscrizione danni).

Nel presente lavoro adottiamo un approccio simil Standard Formula, ovvero basato sulla dipendenza lineare tra i sinistri, per la stima del  $VaR_{99,5}$  del costo sinistri aggregato. Nella tabella seguente mostriamo una sensitivity analysis condotta sul VaR al variare della matrice di correlazione tra i costi delle garanzie di interesse:

Correlazione	$VaR_{99,5}$
$r$ Pearson	77.402
$\tau$ Kendall	87.932
$\rho$ Spearman	88.424
massima	105.779
nulla	32.655

La matrice di correlazione lineare di Pearson, così come i  $\tau$  di Kendall e  $\rho$  di Spearman, sono empirici e riportati nelle tabelle che seguiranno.

La sensitivity analysis ha lo scopo di mostrare l'impatto sull'output delle variazioni della dipendenza lineare: notiamo che non c'è una differenza sostanziale tra i casi in cui si adotti la dipendenza di tipo Kendall o quella di Spearman, mentre avremmo una esposizione poco minore in caso di adozione della dipendenza di Pearson. Supponendo, invece, i costi completamente incorrelati avremmo un VaR decisamente più basso, al contrario di quello che otterremo se i costi fossero perfettamente dipendenti linearmente.

Una sensitivity analysis relativa alla matrice di correlazione, simile a quella appena esposta, si può trovare in Bermúdez et al. (2013); in tale lavoro, infatti, gli autori mostrano il VaR ottenuto a partire dalle correlazioni empiriche, da correlazioni nulle e dalle correlazioni proposte dal Regolamento. Tali risultati della Standard Formula sono poi confrontati con un modello interno basato sulle Copule Gaussiana e t-Student.

Mostrati i risultati della Standard Formula, nel prosieguo riporteremo le applicazioni dei modelli "interni" esposti in precedenza, dapprima sotto

l'ipotesi di indipendenza tra i costi delle garanzie e successivamente tenendo conto della (eventuale) dipendenza tra essi.

Con tali modelli stimeremo ancora il Value-at-Risk del costo sinistri aggregato; avendo volutamente tralasciato, nell'applicazione della Standard Formula, le poste relative a premi e spese, i risultati che otterremo con i modelli interni saranno confrontabili con quanto appena ottenuto.

### 4.2.2 Modelli composti

Nella presente sezione mostreremo separatamente i modelli, composti e non, stimati per le diverse garanzie, con il fine di calcolare il Value-at-Risk del costo aggregato sotto l'ipotesi di indipendenza.

Iniziamo dalla Garanzia 1 per la quale è stato scelto un modello composto di tipo Gamma-GPD con soglia fissa; quest'ultima, come detto, ci permette di distinguere tra sinistri attritional e sinistri large, individuata attraverso lo stimatore di Hill, e risulta pari a 150€.

La distribuzione Gamma è caratterizzata da due parametri  $\alpha$  e  $\beta$  e la sua densità risulta

$$f_1(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-x\beta} \quad (4.5)$$

dove  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$  è la funzione Gamma. La ripartizione, invece, è data da

$$F_1(x; \alpha, \beta) = \frac{\gamma(\alpha, x\beta)}{\Gamma(\alpha)} \quad (4.6)$$

dove  $\gamma(\alpha, x\beta) = \int_0^{x\beta} t^{\alpha-1} e^{-t} dt$  è la la funzione Gamma incompleta inferiore.

I parametri stimati, insieme agli intervalli di confidenza al 99,5%, per il modello composto (1.37) con  $f_1 = \text{Gamma}$  ed  $f_2 = \text{GPD}$  sono riportati di seguito:

	estremo inferiore	stima puntuale	estremo superiore
$\alpha$	0,123	0,129	0,135
$1/\beta$	40,386	46,838	53,291
$\sigma_{mu}$	673,441	1.250,392	1.287,343
$\xi$	0,113	0,474	0,834

Dall'approccio aggregato della Teoria del Rischio, sappiamo che oltre ad individuare un modello per il costo del singolo sinistro, per giungere alla distribuzione del costo aggregato è necessario stimare anche una distribuzione teorica per la frequenza sinistri. Per la Garanzia 1 abbiamo stimato, in particolare, una Binomiale Negativa con parametro 4,514 e probabilità di successo 0,1863226.

Con tali informazioni siamo in grado di simulare, a questo punto, la distribuzione del costo sinistri aggregato  $X$  ed, in particolare, generare il suo quantile al livello 99,5%. Per la Garanzia 1 questo risulta circa pari a 4.300€.

Procediamo con la Garanzia 2 per la quale, a differenza di quanto fatto per la prima garanzia, è stata scelta una distribuzione teorica di tipo Lognormale semplice, senza fare ricorso ai modelli composti.

Di seguito riportiamo le funzioni di densità e ripartizione di una Lognormale:

$$f_1(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right], x > 0 \quad (4.7)$$

$$F_1(x; \mu, \sigma) = \Phi_{\mu, \sigma}(\log x) \quad (4.8)$$

dove  $\Phi_{\mu, \sigma}$  è la funzione di ripartizione di una Normale con media  $\mu$  e deviazione standard  $\sigma$ .

Similmente a quanto fatto in precedenza, mostriamo i parametri stimati per la distribuzione Lognormale:

	estremo inferiore	stima puntuale	estremo superiore
$\mu$	0,976	0,978	0,981
$\sigma$	2,489	2,508	2,732

Per la frequenza sinistri è stata scelta anche ora una distribuzione Binomiale Negativa con parametro stimato pari a 2,443 e probabilità di successo pari a 0,096016592. Il VaR al 99,5% del costo sinistri aggregato ottenuto attraverso simulazioni dalla suddetta Lognormale risulta, per la Garanzia 2, pari a circa 19.700€.

Procediamo nell'ordine con la Garanzia 3 che risulta essere, da quanto riportato precedentemente, la garanzia maggiormente impattante sul portafoglio della compagnia di assicurazione, sia per i costi prodotti sia per la frequenza con cui si manifestano i relativi sinistri.

Anche per tale garanzia, il modello teorico individuato è il Gamma-GPD con soglia fissa e pari a 1.197,30€. Di seguito i parametri stimati:

	estremo inferiore	stima puntuale	estremo superiore
$\alpha$	0,095	0,099	0,103
$1/\beta$	3.425,845	4.195,609	4.965,374
$\sigma_{mu}$	946,127	1.195,670	1.445,213
$\xi$	0,052	0,218	0,383

Per quanto riguarda la frequenza sinistri, facciamo riferimento ad una Binomiale Negativa con parametro 3,396 e probabilità di successo 0,017677124. Il VaR al 99,5% del costo sinistri aggregato è di circa 200.400€.

Infine, per i costi della Garanzia 4 è stata individuata una distribuzione Gamma senza distinzione tra sinistri normali ed estremi i cui parametri  $\alpha$  e  $\beta$  risultano

	estremo inferiore	stima puntuale	estremo superiore
$\alpha$	0,002	0,004	0,008
$\beta$	6.489,378	6.654,918	6.856,707

Per la frequenza sinistri è stata scelta anche ora una distribuzione Binomiale Negativa con parametro stimato pari a 4,738 e probabilità di successo pari a 0,31840676. Il VaR al 99,5% del costo sinistri aggregato risulta circa 8.800€.

Per concludere la sezione, il VaR al 99,5% aggregato per tutte e quattro le garanzie, ottenuto semplicemente come somma dei VaR sopra riportati, è pari a 233.200€.

Vedremo nel prosieguo come, abbandonando l'ipotesi di indipendenza tra i sinistri delle garanzie, il VaR del costo sinistri aggregato subirà una notevole variazione.

### 4.2.3 Funzioni Copula

Come detto, rilasciamo l'ipotesi di indipendenza e iniziamo col riportare la matrice di correlazione empirica:

$$\mathbf{R} = \begin{pmatrix} 1,000000000 & -0,002757362 & 0,023719841 & 0,002220844 \\ -0,002757362 & 1,000000000 & 0,002076243 & -0,001749285 \\ 0,023719841 & 0,002076243 & 1,000000000 & 0,037334148 \\ 0,002220844 & -0,001749285 & 0,037334148 & 1,000000000 \end{pmatrix}$$

che suggerisce una scarsa ma positiva correlazione tra i sinistri delle 4 garanzie.

Di seguito le matrici contenenti il  $\tau_K$  di Kendall ed il  $\rho_S$  di Spearman, rispettivamente:

$$\mathbf{K} = \begin{pmatrix} 1,000000000 & 0,05003731 & 0,04133437 & 0,05179667 \\ 0,05003731 & 1,000000000 & 0,04467803 & 0,01183049 \\ 0,04133437 & 0,04467803 & 1,000000000 & 0,07219142 \\ 0,05179667 & 0,01183049 & 0,07219142 & 1,000000000 \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} 1,00000000 & 0,05092775 & 0,04392973 & 0,05246527 \\ 0,05092775 & 1,00000000 & 0,04757146 & 0,01199865 \\ 0,04392973 & 0,04757146 & 1,00000000 & 0,07647369 \\ 0,05246527 & 0,01199865 & 0,07647369 & 1,00000000 \end{pmatrix}$$

Entrambe le misure di associazione alternative al coefficiente di correlazione di Pearson suggeriscono che c'è dipendenza positiva tra i sinistri, la cui intensità, tuttavia, non è molto elevata. Per tale motivo, l'individuazione di un'opportuna Copula non sarà semplice; addirittura, potrebbe essere necessario considerare i campioni indipendenti, rendendo più che sufficiente l'impiego di un modello costruito con l'ipotesi di indipendenza e, soprattutto, inappropriato il ricorso ad una funzione Copula.

Per le distribuzioni marginali dei costi abbiamo stimato delle v.a. Gamma per le garanzie 1, 3 e 4 ed una Lognormale per la garanzia 2. Nelle tabelle che seguono ne riportiamo i parametri ottenuti attraverso il metodo dei momenti:

Gamma	$\alpha$	$1/\beta$
Garanzia 1	0,009	0,0001
Garanzia 3	0,111	0,0003
Garanzia 4	0,004	0,0001
Lognormale	$\mu$	$\sigma$
Garanzia 2	0,978	2,508

Non sono stati presi in considerazione i modelli composti di cui abbiamo parlato in precedenza in quanto non compatibili, ad oggi, con gli algoritmi di simulazione delle funzioni Copula.

Detto questo, passiamo alla rassegna delle funzioni Copula analizzate; iniziamo dalla Gumbel (2.16) per la quale, utilizzando il metodo della maximum pseudo-likelihood, abbiamo stimato il seguente parametro:

Stimatore	$\hat{\theta}$	Errore standard
Maximum pseudo-likelihood	3,094365	0,02659297

La stima è significativamente diversa da 1, quindi la Copula di Gumbel non coincide con la Copula Indipendenza. Possiamo affermare ciò in quanto abbiamo verificato, tramite procedura bootstrap, che il parametro  $\hat{\theta}$  è significativamente distribuito come una Normale; infatti, il test di Shapiro-Wilk ha restituito un p-value pari 0,18376 e, dunque, fissato il livello di significatività al 5%, non è consentito rifiutare l'ipotesi di Normalità. L'intervallo di confidenza al 99% è (3,025735 ; 3,162995).

In generale, qualunque sia la Copula  $C$  oggetto di studio, quando i parametri sono significativamente tali da far convergere all'indipendenza, molto probabilmente è stato commesso un errore di specificazione della Copula stessa, ovvero c'è una certa dipendenza ma non è corretto descriverla attraverso  $C$ .

Passiamo alla stima ottenuta per il parametro della Copula di Clayton (2.17):

Stimatore	$\hat{\theta}$	Errore standard
Maximum pseudo-likelihood	7,005055	0,04279126

Ricordiamo che il parametro può variare nell'intervallo  $[-1, \infty)$ , con  $\theta \neq 0$ ; tuttavia, per  $\theta \rightarrow 0$  la Clayton tende alla Copula Indipendenza; data la Normalità della distribuzione bootstrap dello stimatore di massima pseudo-verosimiglianza<sup>1</sup>, l'intervallo di confidenza al 99% è (6,894621; 7,115489). Pertanto, possiamo considerare la stima puntuale (7,005055) significativa-

<sup>1</sup>Il test di Shapiro-Wilk per tale metodo di stima ha restituito un p-value pari a 0,5806.

mente diversa dallo zero e, quindi, la Clayton significativamente diversa dalla Copula Indipendenza.

Per quanto riguarda la Copula di Frank (2.18), la stima del parametro risulta:

Stimatore	$\hat{\theta}$	Errore standard
Maximum pseudo-likelihood	13,96241	0,05018706

Per la formulazione di Frank il parametro può assumere un qualsiasi valore reale eccetto lo zero, un valore  $\theta > 0$  indica una dipendenza positiva e, inoltre, solo per  $\theta \rightarrow 0$  si ha la Copula Indipendenza (si veda Assunção, 2004). L'intervallo di confidenza (13,83289; 14,09193) per il parametro stimato con la prima metodologia è stato costruito considerando una distribuzione Normale per  $\theta$  stesso, come si può intuire dalla sua densità bootstrap (figura seguente) e dal test di Shapiro-Wilk, il cui p-value è risultato pari a 0,4473.

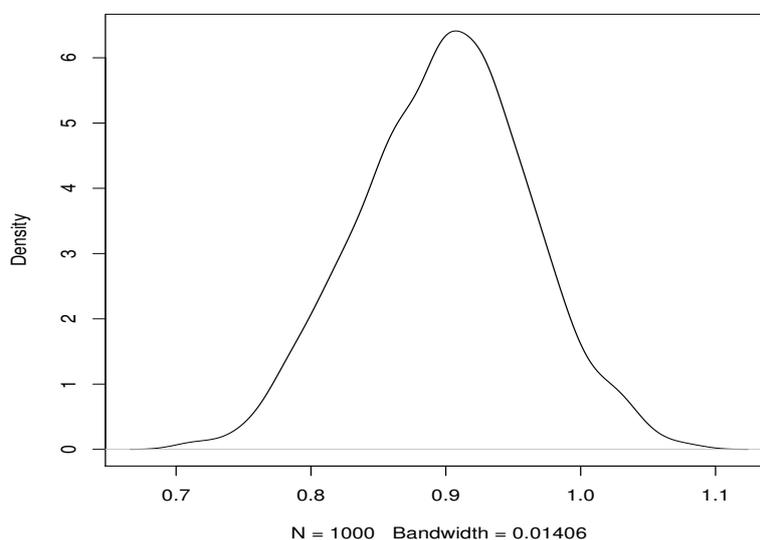


Figura 4.13: Funzione di densità, ottenuta mediante la procedura bootstrap, per il parametro della Frank.

Pertanto, la Copula di Frank appena stimata si può considerare significativamente diversa dalla Copula Indipendenza e sembra cogliere la dipenden-

za positiva tra i sinistri (ricordiamo, comunque, che tale dipendenza non è eccessivamente intensa).

Per quanto riguarda, infine, la Copula di Joe (2.19), il cui parametro è sempre almeno pari ad 1 e se tende ad 1 ci si riconduce alla Copula Indipendenza, la stima ottenuta con il solito metodo della maximum pseudo-likelihood è pari a 3,2814 con errore standard pari a 0,03291382. Sempre con la tecnica bootstrap, è stato possibile generare la distribuzione di tale parametro che non risulta significativamente diversa da quella di una Normale; il relativo intervallo di confidenza al 99% è (3,196457; 3,366343).

Se diverse funzioni copule vengono considerate, come nel nostro caso, allora sorge naturale l'esigenza di capire quale tra queste si adatti meglio ai dati. A tal proposito, è possibile far riferimento a dei test statistici basati, per lo più, su procedure di tipo bootstrap; per quel che ci riguarda, prendiamo in considerazione il test statistico basato sul cosiddetto Kendall's process. Tale procedura può essere impiegata per una qualsiasi Copula archimedeica e viene spiegata nel dettaglio in Genest & Rivest (1993) ed in Trivedi & Zimmer (2005); ci limitiamo a dire soltanto che, sulla base di  $B$  campioni bootstrap estratti dal (dai) campione(i) di partenza, vengono misurate due differenti statistiche test, quella di Cramer-von Mises e quella di Kolmogorv-Smirnov, e i corrispondenti p-value. Ricordiamo che tali statistiche test sono anche delle funzioni "distanza" e che, quindi, bisognerebbe preferire il modello con distanza minima dalla Copula empirica. Tutte le copule esaminate mettono in risalto, chi più chi meno, una dipendenza positiva tra le code delle marginali e tale aspetto risulta di indubbio interesse in ambito attuariale. Per scegliere quale tra questi modelli è il più vicino alla Copula empirica, ovvero quale meglio descrive la dipendenza osservata, abbiamo impiegato il Kendall's process con riferimento alla statistica test di Cramer-von Mises,  $S_n^C$ , definita in base all'equazione (9) di Genest et al.

(2009). Detta  $C_n$  la Copula empirica e  $C$  quella teorica <sup>2</sup>, l'ipotesi testata è  $H_0 : C_n = C$ . Di seguito riportiamo i p-value ottenuti:

Copula	p-value CvM ( $S_n^C$ )
Gumbel	0,4604
Clayton	0,00495
Frank	0,0473
Joe	0,5337

Fissato un livello di significatività del 5%, è possibile rifiutare l'ipotesi nulla nel caso delle copule Clayton e Frank, ovvero è possibile affermare che tali copule sono significativamente diverse dalla Copula empirica.

Un altro modo di scegliere tra diverse alternative è calcolare i coefficienti AIC e BIC che riportiamo in ordine crescente:

Copula	AIC	BIC
Joe	2.788	2.782
Gumbel	7.802	7.795
Frank	16.644	16.638
Clayton	19.589	19.583

Preferendo il modello con AIC e BIC minori, la migliore alternativa sarebbe la Copula Joe, a seguire Gumbel e Frank. La Clayton, invece, sembra essere la meno adatta a catturare la dipendenza tra i costi delle quattro distinte garanzie.

I criteri AIC e BIC, di fatto, confermano i risultati del test statistico.

#### 4.2.4 Fast Fourier Transform

In questa sezione cercheremo di cogliere la dipendenza empirica tra i costi sostenuti dalla compagnia per le quattro garanzie attraverso l'impiego

<sup>2</sup>In Genest et al. (2009) viene presa in considerazione una distanza tra la Copula empirica e quella di indipendenza; le procedure implementate in R, tuttavia, danno la possibilità di misurare la distanza tra la Copula empirica e una Copula teorica  $C$  che non coincide necessariamente con la Copula Indipendenza.

della Fast Fourier Transform che, come detto, ci consente di non assumere specifiche ipotesi circa la dipendenza tra sinistri, necessarie invece per la definizione di una Copula.

Di seguito riportiamo i risultati ottenuti attraverso la FFT multidimensionale, rinviando a Robertson (1992), Wang (1998), Homer & Clark (2003) e Cerchiara (2006) per ulteriori argomentazioni teoriche e applicative. Abbiamo deciso di discretizzare la funzione di ripartizione empirica dei costi senza far riferimento ad una particolare distribuzione teorica; tale discretizzazione ci ha permesso di costruire la matrice  $\mathbf{M}_z$  a cui abbiamo applicato l'algoritmo FFT. In questo modo è stata ottenuta una nuova matrice  $FFT(\mathbf{M}_z)$  che funge da input per la funzione generatrice di probabilità della v.a.  $k$ .

Ma qual è la distribuzione di  $k$ ? Avendo a disposizione il numero dei singoli sinistri suddivisi per anno e per garanzia, è stato possibile ricondursi alla frequenza complessiva anno per anno e calcolarne alcune statistiche descrittive (riportate all'inizio della sezione). Il 50% delle frequenze è compreso tra 127 e 340 sinistri e che c'è una leggera asimmetria positiva. Inoltre, la serie storica presenta un trend decrescente: si va da 420 sinistri nel 2006 a 88 nel 2012. Con una varianza nettamente maggiore della media, è possibile supporre una distribuzione Binomiale Negativa la cui funzione generatrice di probabilità è definita dalla (4.4).

Con i dati a disposizione, ne abbiamo stimato il parametro  $m$  risultato pari a 4,009486 con probabilità di successo di 0,016326537 ed ottenuto la matrice  $PGF_k(FFT(\mathbf{M}_z))$ .

Come ultimo passo, è stata applicata l'inversa IFFT della FFT il cui risultato è la matrice  $\mathbf{M}_x$  contenente, di fatto, la distribuzione congiunta dei sinistri delle garanzie, ovvero la distribuzione di probabilità discretizzata del costo sinistri aggregato, che preserva la struttura di dipendenza presente

nei dati. Avendo simulato la distribuzione di  $X$  per la prima volta, siamo in grado di stimarne il VaR al 99,5% che risulta pari a circa € 400.000.

### 4.2.5 Risultati

Per concludere riportiamo una tabella riepilogativa contenente i Value-at-Risk ottenuti per il dataset multigaranzia con la Standard Formula e con i modelli proposti, sia sotto l'ipotesi di indipendenza sia con alcune funzioni Copula<sup>3</sup>, oltre che con l'algoritmo della Fast Fourier Transform.

Modello	$VaR_{99,5}$
Standard Formula	87.932
Indipendenza	233.200
Copula Gaussiana	248.028
Copula Gumbel	396.291
Copula Joe	396.345
FFT	400.052

Tabella 4.10: Stime del VaR al livello 99% con differenti modelli.

Anche qui è evidente quanto sia importante tener conto dell'eventuale dipendenza e, se presente, della sua forma; si nota, infatti, che i VaR delle Copule Gumbel e Joe e della FFT sono drasticamente maggiori sia rispetto al caso di completa indipendenza (+70%) sia rispetto alla dipendenza lineare (+60%) della Copula Gaussiana. Relativamente, invece, alla formula standard, la variazione è addirittura del 350%<sup>4</sup>.

<sup>3</sup>Alla base della Copula Gaussiana vi è l'ipotesi di dipendenza lineare tra le marginali che, a loro volta, sono di tipo Normale.

<sup>4</sup>Il VaR riportato per l'approccio simil standard formula è quello ottenuto a partire dal coefficiente  $\tau$  di Kendall che risulta essere l'input principale per la stima dei parametri di una funzione Copula.

# Conclusioni

Come anticipato nell'introduzione del presente lavoro, lo scopo prefissato era quello di modellizzare il costo aggregato  $X$  sia per giungere alla stima del percentile al 99,5% sia, soprattutto, per quantificare l'impatto che la dipendenza tra sinistri può avere su di esso.

Abbiamo visto i principali risultati della Extreme Value Theory, in primis il metodo POT con cui è stata stimata la soglia discriminante tra sinistri attritional e large del dataset danese; abbiamo descritto e applicato alcuni modelli composti, con soglia fissa e non, anche stimati con il metodo della minima distanza, in netta crescita in ambito attuariale. Abbiamo inoltre accennato alle distribuzioni che si possono comporre a partire dalla Kumaraswamy, non ancora molto diffuse ma che potrebbero tornare utili in determinate circostanze.

Abbiamo parlato delle funzioni Copula, mostrandone gli aspetti teorici e, soprattutto, il forte legame con la EVT; ne abbiamo investigato le proprietà attraverso simulazioni ad hoc che ci hanno permesso di generare le funzioni di densità e le curve di livello e di calcolare, per esempio, gli annessi indici di dipendenza. Abbiamo, inoltre, mostrato una procedura numerica che permette, senza necessitare di particolari ipotesi di lavoro, di tener conto della interrelazione dei dati, ovvero la Fast Fourier Transform.

Tali metodologie sono state applicate a due distinti dataset, portandoci a risultati interessanti; infatti, come si evince dalle tabelle 4.9 e 4.10, sotto

l'ipotesi di indipendenza la distribuzione del costo aggregato può avere un VaR significativamente più basso di quelli calcolati assumendo ipotesi di dipendenza.

È necessario ricordare, tuttavia, che il VaR è stato, ed è tutt'oggi, soggetto a diverse critiche metodologiche, come l'assenza della proprietà di sub-additività (McNeil et al., 2005) e la delicata scelta dell'orizzonte temporale rispetto al quale calcolarlo (Cocozza et al., 2008). Una misura di rischio alternativa, come detto in precedenza, è il Tail-VaR, o Expected Shortfall, già impiegato, per esempio, nel test svizzero sulla solvibilità e proposto in diversi articoli accademici. Per definizione, è una misura di rischio maggiormente rappresentativa della coda di una distribuzione rispetto al corrispondente VaR e, pertanto, la sua adozione potrebbe condurre ad una valutazione più realistica del rischio, pur non applicando un modello interno.

La Standard Formula, basata sulle ipotesi di correlazione e distribuzione di tipo Lognormale semplice del costo sinistri aggregato, potrebbe condurre ad una sottostima del Value-at-Risk e, di conseguenza, del requisito di capitale richiesto dalla normativa. L'assunzione di correlazione di sicuro semplifica i calcoli, ma espone al pericolo di trascurare eventuali e più complesse forme di dipendenza; per di più, non è prevista alcuna correlazione tra i sinistri attritional ed i sinistri estremi.

Tuttavia, nessuno degli alternativi approcci proposti è esente da limitazioni. Con i modelli composti, costruiti sotto l'ipotesi di indipendenza, possiamo adattare accuratamente le distribuzioni marginali senza ovviamente identificare una eventuale dipendenza tra esse. Con le funzioni Copula possiamo modellare la dipendenza, ma come detto non è facile individuare quella che meglio si adatta ai dati.

Una evoluzione futura del presente lavoro è proprio legato alla scelta ottimale della Copula da impiegare, attraverso quello che viene definito **patch-**

**work construction method** di cui una dettagliata descrizione può essere trovata in Durante et al. (2013): brevemente, il metodo permette di costruire una Copula "composta"; per esempio, possono presentarsi casistiche per le quali è necessario definire due distinte funzioni Copula, una per la parte centrale della distribuzione di interesse ed una per la coda destra della stessa.

È, inoltre, al vaglio la possibilità di collaborare con IVASS per un'analisi della prassi del mercato italiano relativamente all'impiego delle funzioni Copula; in particolare, potrebbe essere interessante studiare situazioni in cui la più diffusa Copula in ambito professionale, quella Gaussiana, mostra un adattamento ai dati non sufficiente (anche in relazione al comportamento asintotico).

Inoltre, sia nell'applicazione dei modelli composti che delle funzioni Copula, per rendere più snella la procedura di simulazione, è stata utilizzata una frequenza sinistri costante; una possibile miglioria del presente lavoro sarebbe ottenibile impiegando una frequenza sinistri aleatoria anche per i modelli composti e le funzioni Copula, come fatto per la FFT.

Quest'ultima, del resto, ci permette di non dover simulare le marginali né di stimare una soglia né di assumere ipotesi circa la dipendenza, ma include distorsioni dovute alla discretizzazione ed al troncamento delle distribuzioni, come gli aliasing errors che qui non sono stati approfonditi e per i quali si potrebbe condurre un'analisi specifica (si veda Robe-Voinea & Vernic, 2016). Infine, i risultati mostrati sono stati ottenuti su dataset relativamente piccoli, al contrario di quelli con cui si ha che fare quotidianamente, e sono al lordo della riassicurazione; se considerassimo i costi al netto di coperture riassicurative, le code delle distribuzioni sarebbero meno pesanti e le correlazioni previste dalla Standard Formula potrebbero essere più ragionevoli<sup>5</sup>.

---

<sup>5</sup>Sottolineiamo comunque che l'analisi condotta in tale lavoro è valida anche per la definizione di un opportuno trattato riassicurativo multiramo.

In definitiva, i modelli proposti potrebbero sicuramente portare ad un miglior fitting del profilo di rischio, pur aggravando il SCR, ma resta da sottolineare che la Standard Formula rimane comunque un approccio pragmatico.

Si ricorda che le calibrazioni adottate hanno riguardato il rischio di sottoscrizione dei contratti di tipo non vita (Underwriting Risk Non-Life), al lordo di coperture riassicurative, ed in particolare il Premium Risk; i modelli presentati costituiscono, dunque, una possibile via per la costruzione di un modello interno parziale.

La scelta dei vari modelli, la stima dei loro parametri, il controllo in termini di adattamento, le ipotesi sulle strutture di dipendenza e la verifica dei requisiti di capitale sono tutti aspetti che necessitano di continue verifiche e calibrazioni. Con questo lavoro di tesi, si spera di essere riusciti a mostrare come un'attenta analisi di tutti questi elementi possa condurre all'individuazione di un profilo di rischio e di una misura patrimoniale significativamente espressivi della complessa realtà sottostante.

# Bibliografia

- [1] Acri, F. & Cerchiara, R. R. (2016). Loss Distribution and Dependence Structure: Composite and Copula Models. Working Paper n. 08 - 2016, Department of Economics, Statistics and Finance, University of Calabria. Link: [http://www.unical.it/portale/portalmidia/2016-09/WP08\\_2016.pdf](http://www.unical.it/portale/portalmidia/2016-09/WP08_2016.pdf)
- [2] Artzner, P. & Delbaen, F. & Eber, J. M. & Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, **9(3)**, 203-228.
- [3] Assunção, R. (2004). A note on testing parameters of Frank's copula models. *Belgian Actuarial Bulletin*, **4**, No. 1.
- [4] Beirlant, J. & Goegebeur, Y. & Segers, J. & Teugels, J. (2005). Statistics of Extremes: Theory and Applications, Wiley series in probability and statistics.
- [5] Bermúdez, L. & Ferri, A. & Guillén, M. (2013). A Correlation Sensitivity Analysis of non-Life Underwriting Risk in Solvency Capitale Requirement Estimation. *ASTIN Bulletin*, **43**, 21-37.
- [6] Burnecki, K. & Härdle, W. & Weron, R. (2004). Simulation of risk processes. *Wiley StatsRef: Statistics Reference Online*.
- [7] Carreau, J. & Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, **12(1)**, 53-76.

- 
- [8] Charpentier, A. (2014). Computational actuarial science with R. CRC Press.
- [9] Charpentier, A. & Oulidi, A. (2010). Beta kernel quantile estimators of heavy-tailed loss distributions. *Statistics and computing*, **20(1)**, 35-55.
- [10] Cerchiara, R. R. (2006). Metodo Simulativo, Fast Fourier Transform ed Extreme Value Theory per l'analisi del Costo Sinistri Aggregato nelle Assicurazioni Danni. Tesi di Dottorato in Scienze Attuariali, Università La Sapienza di Roma.
- [11] Coccozza, R. & Di Lorenzo, E. & Orlando, A. & Sibillo, M. (2008). The Value at Risk of the Mathematical Provision: Critical Issues. *Journal of Risk Management in Financial Institutions*, **1**, 311-319. Available at SSRN: <https://ssrn.com/abstract=1736963>
- [12] Coles, S.G. & Powell, E.A. (1996). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review*, **64**, 119-136.
- [13] Coles, S.G. & Tawn, J.A. (1996a). A Bayesian analysis of extreme rainfall data, *Applied Statistics*. **45**, 463-478.
- [14] Cooray, K. & Ananda, M. M. A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, **5**, 321-334.
- [15] Cordeiro, G. M. & Nadarajah, S. & Ortega, E.M.M. (2011). The Kumaraswamy Gumbel distribution. *Statistical Methods & Applications*, **21**, 139-168.
- [16] Debbabi, N. & Kratz, M. & Mboup, M. (2016). A Self-Calibrating Method for Heavy Tailed Data Modeling: Application in Neuroscience

- and Finance. *ESSEC WORKING PAPER 1619*. Available at SSRN: <https://ssrn.com/abstract=2898731>
- [17] Demarta, S. & McNeil, A.J. (2005). The t Copula and Related Copulas. *International Statistical Review*, **73**(1), 111-129.
- [18] Durante, F. & Sánchez, J. F. & Sempi, C. (2013). Multivariate patchwork copulas: a unified approach with applications to partial comonotonicity. *Insurance: Mathematics and Economics*, **53**(3), 897-905.
- [19] Dutang, C. & Goulet, V. & Pigeon, M. (2008). actuar: An R package for actuarial science. *Journal of Statistical software*, **25**(7), 1-37.
- [20] Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman & Hall.
- [21] Embrechts, P. & Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer.
- [22] Engeland, K. & Hisdal, H. & Frigessi, A. (2004). Practical extreme value modelling of hydrological floods and droughts: a case study. *Extremes*, **7**, 5-30.
- [23] Esmaeili, H. & Klüppelberg, C. (2010). Parameter estimation of a bivariate compound Poisson process. *Insurance: mathematics and economics*, **47**(2), 224-233.
- [24] Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, Melbourne: Krieger.
- [25] Genest, C. & Rémillard, B. & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**, 199-213.

- [26] Genest, C. & Rivest, L.P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, **88** (423), 1034-1043.
- [27] Grubel, R. & Hermesmeier, R. (1999). Computation of compound distributions I: aliasing errors and exponential tilting. *ASTIN Bulletin*, **29** (2), 197-214.
- [28] Guillothe, S. & Perron, F. & Segers, J. (2011). Nonparametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(3), 377-406.
- [29] Heckman, P. E. & Meyers, G. G. (1983). The calculation of aggregate loss distributions from claim severity and claim count distributions. *Proceedings of the Casualty Actuarial Society*, **LXX**, 22-61.
- [30] Homer, D.L. & Clark, D.R. (2003). Insurance Applications of Bivariate Distributions. *Proceedings of the Casualty Actuarial Society*, **90**, 274-307.
- [31] Hu, Y. & Scarrott, C.J. (2013). evmix: Extreme Value Mixture Modelling, Threshold Estimation and Boundary Corrected Kernel Density Estimation. Available on CRAN.
- [32] Joe, H. & Xu, J. (1996). The Estimation Method of Inference Functions for Margins for Multivariate Models. *Technical Report 166*, Department of Statistics, University of British Columbia.
- [33] Juarez, S. & Schucany, W. (2004). Robust and efficient estimation for the generalized Pareto distribution. *Extremes*, **7**, 237-251.

- [34] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, **46(1-2)**, 79-88.
- [35] Liu, H. & Wang, R. (2017). Collective Risk Models with Dependence Uncertainty. *ASTIN Bulletin: The Journal of the IAA*, **47(2)**, 361-389.
- [36] Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis*, **51**, 904-917.
- [37] McNeil, A. J. & Frey, R. & Embrechts, P. (2005). *Quantitative Risk Management. Concepts, Techniques and Tools*, Princeton University Press.
- [38] Mood, A.M. & Graybill, F.A. & Boes, D.C. (1993). *Introduzione alla Statistica*, McGraw-Hill.
- [39] Nadarajah, S. & Bakar, S. A. A. (2014). New composite models for the Danish fire insurance data. *Scandinavian Actuarial Journal*, **2**, 180-187.
- [40] Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, **3**, 119-131.
- [41] Pigeon, M. & Denuit, M. (2011). Composite Lognormal-Pareto model with random threshold. *Scandinavian Actuarial Journal*, **3**, 177-192.
- [42] Preda, V. & Ciumara, R. (2006). On composite models: Weibull-Pareto and Lognormal-Pareto. A comparative study. *Rom. J. Econ. Forecast.*, **3**, 32-46.
- [43] R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>

- [44] Regolamento Delegato (UE) 2015/35 del 10 ottobre 2014. Gazzetta ufficiale dell'Unione Europea, 17/01/2015.
- [45] Robertson, J. P. (1992). The Computation of Aggregate Loss Distributions. *PCAS*, **LXXIX**, 57-133.
- [46] Robe-Voinea, E.G. & Vernic, R. (2016). Fast Fourier Transform for multivariate aggregate claims, *Computational and Applied Mathematics, Springer International Publishing*, 1-15.
- [47] Robe-Voinea, E.G. & Vernic, R. (2016). On the recursive evaluation of a certain multivariate compound distribution, *Acta Mathematicae Applicatae Sinica, English Series, Springer International Publishing*, **32(4)**, 913-920.
- [48] Rockinger, M. & Jondeau, E. (2006). The Copula-GARCH Model of Conditional Dependencies: An International Stock-Market Application. *Journal of International Money and Finance*, **25(5)**, 827-853.
- [49] Rodrigues, L.H. & Gomes, M.I & Pestana, D. (2011). Statistics of extremes in athletics. *Statistical Journal*, **2**, 127-153.
- [50] Sklar, A.W. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229-231.
- [51] Scollnik, D. P. M. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, **1**, 20-33.
- [52] Schaller, P. & Temnov, G. (2008). Efficient and precise computation of convolutions: applying fft to heavy tailed distributions. *Computational Methods in Applied Mathematics Comput. Methods Appl. Math.*, **8(2)**, 187-200.

- [53] Smith, R.L. (1985). *Statistics of extreme values*, Bull. ISI, LI (Book 4), Paper 26.1.
- [54] Teodorescu, S. & Vernic, R. (2009). Some composite Exponential-Pareto models for actuarial prediction. *Rom. J. Econ. Forecast.*, **12**, 82-100.
- [55] Teodorescu, S. & Vernic, R. (2013). On Composite Pareto Models. *MATH. REPORTS*, **15(65)**, 1, 11-29.
- [56] Trivedi, P. K. & Zimmer, D. M. (2005). Copula Modeling: An Introduction for Practitioners, *Foundations and Trends<sup>®</sup> in Econometrics*, **1(1)**, 1-111.
- [57] Vannucci, L. (2010). *Teoria del rischio e tecniche attuariali contro i danni*, Ed. Pitagora, Bologna.
- [58] Wang, S.S. (1998). Aggregation of Correlated Risk Portfolios: Models and Algorithms. *Proceedings of the Casualty Actuarial society*, **85(163)**, 848-939.
- [59] Wirch, J. L. & Hardy, M. R. (2001). Distortion risk measures: coherence and stochastic dominance. *International Congress on Insurance: Mathematics and Economics*, 15-17.
- [60] Zhang, J. (2007). Likelihood moment estimation for the Generalized Pareto distribution. *Australian & New Zealand Journal of Statistics*, **49**, 69-77.
- [61] Zhou, C. & Wu, C. & Liu, H. & Liu, F. (2007). A New Method to Choose the Threshold in the POT Model. Available at SSRN: <http://ssrn.com/abstract=987796>

# Ringraziamenti

Ringrazio il Professor Rocco Roberto Cerchiara per avermi costantemente supportato nella preparazione della presente tesi e lungo tutto il mio percorso universitario. Un sentito grazie al referente del dottorato in Scienze Attuariali, Professor Fabio Grasso, ed alla Professoressa Paola Verico che ha accompagnato i miei esordi da dottorando, stimolandomi nell'attività di ricerca.