

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**ΘΕΩΡΙΑ ΤΩΝ COPULAS ΚΑΙ
ΧΡΗΣΗ ΤΟΥΣ ΓΙΑ ΠΑΡΑΓΩΓΗ
ΣΥΝΘΕΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ**

Θεοδώρα Κ. Κοκκινιώτη

Διπλωματική εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Ιούλιος 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ



ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΚΑΙ ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ

**ΘΕΩΡΙΑ ΤΩΝ COPULAS ΚΑΙ
ΧΡΗΣΗ ΤΟΥΣ ΓΙΑ ΠΑΡΑΓΩΓΗ
ΣΥΝΘΕΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ**

Θεοδώρα Κ. Κοκκινιώτη

Διπλωματική εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής
Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των
απαιτήσεων για την απόκτηση του Μεταπτυχιακού
Διπλώματος Ειδίκευσης στην Εφαρμοσμένη Στατιστική

Πειραιάς
Ιούλιος 2014

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ' αριθμόν συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική.

Τα μέλη της Επιτροπής ήταν:

- Καθηγητής: Κούτρας Μάρκος (Επιβλέπων)
- Αναπληρωτής Καθηγητής: Ηλιόπουλος Γεώργιος
- Επίκουρος Καθηγητής: Μπούτσικας Μιχαήλ

Η έγκριση της Διπλωματικής Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνώμών του συγγραφέα.

UNIVERSITY OF PIRAEUS



DEPARTMENT OF STATISTICS
AND INSURANCE SCIENCE

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

**THEORY OF COPULAS
AND THEIR USE TO GENERATE
SYNTHETIC DATA**

By
Theodora K. Kokkinioti

MSc Dissertation

submitted to the Department of Statistics and Insurance
Science of the University of Piraeus in partial fulfilment of
the requirements for the degree of Master of Science in
Applied Statistics

Piraeus, Greece
July 2014

*Στους γονείς μου
Κυριάκο και Ροζημαρή*

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς και επιβλέποντα κ. Μάρκο Κούτρα για την πολύτιμη βοήθεια του, την καθοδήγηση αλλά και την κατανόηση που έδειξε σε όλο το διάστημα εκπόνησης της παρούσας διπλωματικής εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τους καθηγητές του μεταπτυχιακού διπλώματος ειδίκευσης στην Εφαρμοσμένη Στατιστική για τις γνώσεις που μου προσέφεραν. Τέλος, θα ήθελα να ευχαριστήσω όλους εκείνους που ήταν δίπλα μου και με στήριξαν, μα πάνω από όλους την οικογένειά μου και ιδιαίτερα τους γονείς μου που μου συμπαραστέκονται και με στηρίζουν σε κάθε περίοδο της ζωής μου.

Περίληψη

Τις τελευταίες δεκαετίες, όλο και περισσότεροι επιστήμονες σε διαφορετικούς επιστημονικούς κλάδους ασχολούνται με τα Copulas καθώς είναι πολλές οι περιπτώσεις στις οποίες τα δεδομένα που μελετώνται παρουσιάζουν συσχέτιση. Στην παρούσα εργασία δίνουμε κάποιους ορισμούς οι οποίοι είναι απαραίτητοι για τη Θεωρία των Copulas και κάποιες από τις ιδιότητές τους. Εν συνεχεία, δίνουμε συντελεστές συσχέτισης ή εξάρτησης και αναπτύσσουμε δύο μεγάλες κλάσεις των Copulas, τα Elliptical και τα Archimedean Copulas. Επίσης, δίνουμε τους αλγόριθμους για την παραγωγή δεδομένων από συγκεκριμένες οικογένειες, μεθόδους εκτίμησης των Copulas και αναφέρουμε ελέγχους καλής προσαρμογής που αφορούν τα Copulas. Τέλος, προσομοιώνουμε δεδομένα από συγκεκριμένο Copula για να εφαρμόσουμε τα παραπάνω.

Abstract

Over the last few decades, an increasing number of scientists in different scientific fields are getting interested in Copulas. This is mainly due to the fact that in many areas of their research they have to deal with dependent data. In this dissertation, we give some basic definitions to present the Theory of Copulas and some of their properties. Furthermore, we present some correlation and dependence indices and we analyse two classes of Copulas, known as Elliptical and Archimedean Copulas. Moreover, we describe algorithms that generate data from specific families, methods to estimate the Copulas and Goodness-of-fit tests for Copulas. Finally, we use simulated data from a specific family of Copulas in order to apply the techniques illustrated in the dissertation.

Περιεχόμενα

	Σελίδα
Περιεχόμενα	XV
Κατάλογος Πινάκων	XIX
Κατάλογος Σχημάτων	XXI
Πρόλογος	XXV
1 COPULAS: ΟΡΙΣΜΟΙ-ΙΔΙΟΤΗΤΕΣ	1
1.1 ΕΙΣΑΓΩΓΗ	1
1.2 ΒΑΣΙΚΟΙ ΟΡΙΣΜΟΙ	1
1.3 ΒΑΣΙΚΕΣ ΙΔΙΟΤΗΤΕΣ	6
1.4 ΘΕΩΡΗΜΑ ΤΟΥ SKLAR	14
1.5 SURVIVAL COPULAS	21
1.6 ΠΟΛΥΔΙΑΣΤΑΤΑ COPULAS	24
2 ΣΥΣΧΕΤΙΣΗ-ΕΞΑΡΤΗΣΗ	29
2.1 ΕΙΣΑΓΩΓΗ	29
2.2 ΣΥΣΧΕΤΙΣΗ-ΕΞΑΡΤΗΣΗ-ΣΥΜΦΩΝΙΑ	29
2.3 ΣΥΝΤΕΛΕΣΤΕΣ ΣΥΣΧΕΤΙΣΗΣ-ΕΞΑΡΤΗΣΗΣ	31
2.3.1 Επιθυμητές ιδιότητες των συντελεστών συσχέτισης	31
2.3.2 Συντελεστής Γραμμικής Συσχέτισης του Pearson	33
2.3.3 Συντελεστές συσχέτισης τάξης	35
2.3.3.1 Tau του Kendall (τ)	36
2.3.3.2 Rho του Spearman (ρ_S)	40

2.3.4	Συντελεστής Εξάρτησης Ουρών	47
2.3.5	Θετική Τεταρτημοριακή Εξάρτηση	49
3	ΟΙΚΟΓΕΝΕΙΕΣ ΤΩΝ COPULAS	57
3.1	ΕΙΣΑΓΩΓΗ	57
3.2	Elliptical (Ελλειπτικά) Copulas	58
3.2.1	Gaussian (Normal) Copula	65
3.2.2	Student's <i>t</i> -Copula	70
3.3	Archimedean (Αρχιμήδεια) Copulas	75
3.3.1	Οικογένεια Clayton	86
3.3.2	Οικογένεια Gumbel-Hougaard	90
3.3.3	Οικογένεια Frank	93
4	ΑΛΓΟΡΙΘΜΟΙ ΠΑΡΑΓΩΓΗΣ ΔΕΔΟΜΕΝΩΝ-ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ	97
4.1	ΕΙΣΑΓΩΓΗ	97
4.2	Αλγόριθμοι για την παραγωγή δεδομένων από Copula	97
4.2.1	Elliptical Copulas	98
4.2.2	Archimedean Copulas	100
4.3	Μέθοδοι εκτίμησης των Copulas	102
4.3.1	Παραμετρική μέθοδος εκτίμησης ενός Copula	102
4.3.2	Ημιπαραμετρική μέθοδος εκτίμησης ενός Copula	105
4.3.3	Μη παραμετρική μέθοδος εκτίμησης ενός Copula	110
4.4	Έλεγχοι καλής προσαρμογής (Goodness-of-fit tests)	111
5	ΕΦΑΡΜΟΓΗ ΣΕ ΠΡΟΣΟΜΟΙΩΜΕΝΑ ΔΕΔΟΜΕΝΑ	115
5.1	ΕΙΣΑΓΩΓΗ	115
5.2	Εφαρμογή	116
A	ΑΛΓΟΡΙΘΜΟΙ	127
A.1	Υπολογισμός Rho του Spearman και Tau του Kendall με την R για 2 τυχαία δείγματα από τη $N(0, 1)$	127
A.2	Παραγωγή 1000 τυχαίων σημείων από Gaussian Copula	131
A.3	Παραγωγή 1000 τυχαίων σημείων από t-Copula	133

A.4	Παραγωγή 1000 τυχαίων σημείων από Archimedean Copulas	135
A.5	Αλγόριθμος που χρησιμοποιήθηκε για τα αποτελέσματα της Εφαρμογής . .	137
A.6	Αλγόριθμος που χρησιμοποιήθηκε για τον υπολογισμό του μέσου τετραγωνικού σφάλματος της Εφαρμογής	142

Βιβλιογραφία		145
---------------------	--	------------

Κατάλογος Πινάκων

1.1	Δύο τυχαία δείγματα μεγέθους $n = 6$ από την $N(0, 1)$	19
1.2	Τα ranks (οι τάξεις) των παρατηρήσεων.	20
3.1	Σύνοψη αποτελεσμάτων για τα Archimedean Copulas.	96
4.1	Τα W_i, \tilde{W}_i για τον υπολογισμό του συνεπή εκτιμητή της διασποράς που σχετίζεται με το Ταυ του Kendall.	109
5.1	Τα περιγραφικά στοιχεία των τυχαίων μεταβλητών X, Y	117
5.2	Αποτελέσματα για τον έλεγχο καλής προσαρμογής που βασίζεται στη στατιστική συνάρτηση των Cramér-von Mises για τις 5 οικογένειες των Copulas.	120
5.3	Υπολογισμός του μέσου τετραγωνικού σφάλματος για τους εκτιμητές ML , IFM και CML σε 1000 επαναλήψεις και σε δείγματα μεγέθους $n=100, 200, 500, 1000$	126

Κατάλογος Σχημάτων

1.1	Δισδιάστατη Τυπική Κανονική Κατανομή	2
1.2	Plot 3D: Κάτω φράγμα (3D) ενός δισδιάστατου Copula	10
1.3	Contour Plot : Κάτω φράγμα ενός δισδιάστατου Copula	10
1.4	Plot 3D: Άνω φράγμα (3D) ενός δισδιάστατου Copula	11
1.5	Contour Plot: Άνω φράγμα ενός δισδιάστατου Copula	11
1.6	Plot 3D : $C(u, v)$ μαζί με το κάτω φράγμα	11
1.7	Plot 3D : $C(u, v)$ μαζί με το άνω φράγμα	11
1.8	Contour Plot του κάτω φράγματος	12
1.9	Contour Plot του Copula	12
1.10	Contour Plot του άνω φράγματος	12
1.11	Contour Lines του Copula και των φραγμάτων	12
1.12	Plot 3D: Το Copula-γινόμενο	13
1.13	Contour Plot του Copula-γινόμενο	13
1.14	Το διάγραμμα των X, Y από τη $N(0, 1)$	19
1.15	Το διάγραμμα των $e^X, 3e^Y$	19
1.16	Το διάγραμμα των ranks (τάξεων) των X, Y	20
1.17	Το διάγραμμα των ranks (τάξεων) των $e^X, 3e^Y$	20
2.1	(τ-ρ) χωρίο	44
3.1	Περιοχές που ο H -όγκος δυο ακτινικά συμμετρικών τυχαίων μεταβλητών είναι ίδιος	59
3.2	Περιοχές που ο C -όγκος δυο ακτινικά συμμετρικών τυχαίων μεταβλητών είναι ίδιος	60
3.3	1000 τυχαία σημεία από Gaussian Copula με συντελεστή συσχέτισης $\rho = 0.7$	66

3.4	1000 τυχαία σημεία από Gaussian Copula με συντελεστή συσχέτισης $\rho = 0$.	66
3.5	1000 τυχαία σημεία από Gaussian Copula με συντελεστή συσχέτισης $\rho = -0.7$.	66
3.6	Το Contour Plot της συνάρτησης κατανομής του Gaussian Copula με $\rho = 0.7$.	67
3.7	Το Contour Plot της $\sigma.κ.$ του Gaussian Copula μαζί με τα φράγματα.	67
3.8	Η συνάρτηση πυκνότητας ενός Gaussian Copula με περιθώριες ομοιόμορφες στο $(0,1)$.	69
3.9	Contour Plot της συνάρτησης πυκνότητας ενός Gaussian Copula με $\rho = 0.7$.	69
3.10	1000 σημεία από t-Copula για $\nu=7$ και διαφορετικά ρ .	71
3.11	1000 σημεία από t-Copula για $\rho = 0.7$ και διαφορετικά ν .	71
3.12	Το Contour Plot της συνάρτησης κατανομής του t-Copula με $\rho = 0.7$ και $\nu = 7$.	73
3.13	Το Contour Plot της συνάρτησης κατανομής του t-Copula μαζί με τα φράγματα.	73
3.14	Συνάρτηση πυκνότητας ενός t-Copula για $\rho = 0.7$ και $\nu = 7$.	74
3.15	Contour Plot της συνάρτησης πυκνότητας ενός t-Copula για $\rho = 0.7$ και $\nu=7$.	74
3.16	Γραφική παράσταση ενός χαρακτηριστικού γεννήτορα μαζί με την εφαπτομένη στο σημείο $(\alpha, \varphi(\alpha))$.	81
3.17	Καμπύλες επιπέδων και μηδενική καμπύλη ενός Copula.	82
3.18	1000 τυχαία σημεία από Clayton με παράμετρο $\frac{1}{\theta} = 10^{-2}$.	87
3.19	1000 τυχαία σημεία από Clayton με παράμετρο $\frac{1}{\theta} = 10^3$.	87
3.20	1000 τυχαία σημεία από Clayton με παράμετρο $\theta = 1$.	87
3.21	Το Contour Plot της $\sigma.κ.$ του Clayton Copula για $\theta = \frac{1}{3}$.	88
3.22	Η συνάρτηση πυκνότητας του Clayton Copula για $\theta = \frac{1}{3}$.	89
3.23	Το Contour Plot της $\sigma.π.$ του Clayton Copula για $\theta = \frac{1}{3}$.	89
3.24	1000 τυχαία σημεία από Gumbel Copula με παράμετρο $\theta = 1$.	90
3.25	1000 τυχαία σημεία από Gumbel Copula με παράμετρο $\theta = 10^2$.	90
3.26	Το Contour Plot της $\sigma.κ.$ του Gumbel Copula για $\theta = 3$.	91
3.27	Η συνάρτηση πυκνότητας του Gumbel Copula για $\theta = 3$.	92
3.28	Το Contour Plot της $\sigma.π.$ του Gumbel Copula για $\theta = 3$.	92
3.29	1000 τυχαία σημεία από Frank με παράμετρο $e^{-\theta} = 10^{-15}$.	93
3.30	1000 τυχαία σημεία από Frank με παράμετρο $e^{-\theta} = 10^{30}$.	93
3.31	1000 τυχαία σημεία από Frank με παράμετρο $e^{-\theta} = 1$.	93
3.32	Το Contour Plot της $\sigma.κ.$ του Frank Copula για $e^{-\theta} = \frac{1}{3}$.	93
3.33	Η συνάρτηση πυκνότητας του Frank Copula για $e^{-\theta} = \frac{1}{3}$.	94

3.34	To Contour Plot της $\sigma.π.$ του Frank Copula για $e^{-\theta} = \frac{1}{3}$	94
5.1	Το γράφημα με τα προσομοιωμένα δεδομένα από το Clayton Copula με παράμετρο 3 και περιθώριες $N(0, 1)$	116
5.2	Το Contour Plot ενός Clayton Copula με παράμετρο 3 και περιθώριες $N(0, 1)$	117
5.3	Το $Q - Q$ plot της τυχαίας μεταβλητής X με την Κανονική Κατανομή.	118
5.4	Το $Q - Q$ plot της τυχαίας μεταβλητής Y με την Κανονική Κατανομή.	118
5.5	Το Contour plot του Gumbel Copula με περιθώριες $N(0, 1)$	121
5.6	Το Contour plot του Frank Copula με περιθώριες $N(0, 1)$	121
5.7	Το Contour plot του Normal Copula με περιθώριες $N(0, 1)$	121
5.8	Το Contour plot του t Copula με περιθώριες $N(0, 1)$	121

Πρόλογος

Τα Copulas είναι ένα σχετικά πρόσφατο αντικείμενο στην επιστήμη της στατιστικής και επομένως η ιστορία τους δεν είναι εκτενής. Αν και πλέον είναι ένας γνώριμος όρος για την επιστημονική κοινότητα, η πρώτη εμφάνιση του όρου Copula δεν είναι μακρινή. Μόλις το 1959 αναφέρθηκε για πρώτη φορά από τον Abe Sklar, ο οποίος και το χρησιμοποίησε για να περιγράψει ένα θεώρημα, που πλέον φέρει και το όνομά του, σε μια επιστολή του προς τον Frèchet. Μπορεί ο όρος να είναι πρόσφατος, ωστόσο αρκετοί επιστήμονες ασχολήθηκαν με αυτό το αντικείμενο ή πιο συγκεκριμένα με τη μελέτη πολυμεταβλητών κατανομών με γνωστές τις περιθώριες κατανομές και πλέον η στατιστική κοινότητα βρίθει από εργασίες που έχουν σχέση με αυτό το αντικείμενο.

Ο Hoeffding, για παράδειγμα, ασχολήθηκε με δισδιάστατες τυποποιημένες κατανομές, των οποίων το στήριγμα περιέχεται στο $[-\frac{1}{2}, \frac{1}{2}]^2$ και των οποίων οι περιθώριες κατανομές ομοιόμορφα στο $[-\frac{1}{2}, \frac{1}{2}]$ καθώς, επίσης, συνεισέφερε στη δημιουργία ορίων για τα Copulas. Επίσης, ο Frèchet συντέλεσε και αυτός στη δημιουργία ορίων και ανέφερε ότι, ενώ το σύνολο που περιλαμβάνει όλες τις από κοινού συναρτήσεις κατανομής δύο τυχαίων μεταβλητών είναι μη κενό, (περιλαμβάνει το γινόμενο των περιθωρίων στην περίπτωση που οι τυχαίες μεταβλητές είναι ανεξάρτητες), τα υπόλοιπα στοιχεία του συνόλου αυτού δεν είναι εκ των προτέρων γνωστά.

Επομένως ο Abe Sklar δεν είναι ο πρώτος που ασχολήθηκε με τα Copulas, αλλά είναι εκείνος που με το θεώρημά του θεμελίωσε τις βάσεις για τη δημιουργία της θεωρίας τους. Γιατί όμως Copulas; Πώς η λέξη αυτή ήταν έμπνευση για τον Sklar; Η απάντηση είναι απλή, αν ανατρέξει κανείς στη Λατινική Γλώσσα σύμφωνα με την οποία το copula σημαίνει δεσμός. Στην ουσία είναι όρος της Γραμματικής και χρησιμοποιείται σε μία πρόταση για να ενώσει το υποκείμενο με το κατηγορήμα. Αυτή ήταν και η ιδέα του Sklar, καθώς και εδώ υπάρχει ένα είδος δεσμού μεταξύ των περιθωρίων κατανομών και της από κοινού

συνάρτησης κατανομής.

Τα Copulas, επομένως, είναι μαθηματικά εργαλεία, τα οποία αντικατοπτρίζουν τη σχέση εξάρτησης μεταξύ τυχαίων μεταβλητών και χρησιμοποιούνται ευρέως από τους επιστήμονες, καθώς είναι χρήσιμα για την ανάπτυξη θεωρητικών μεθόδων, οι οποίες, όμως, έχουν πρακτική εφαρμογή π.χ. risk management, portfolio analysis, κ.τ.λ. Συγκεκριμένα, τη δεκαετία του 90 έγιναν πολύ σημαντικά βήματα για τη διάδοση των Copulas και ξεκίνησε η συγγραφή βιβλίων από τα οποία έγιναν πιο γνωστά στο ευρύ κοινό. Ο Embrechts [9] αιτιολόγησε το αυξανόμενο ενδιαφέρον της επιστημονικής κοινότητας τη δεκαετία του 90 ως το αποτέλεσμα της επιθυμίας για την ανάπτυξη της οικονομίας σε παγκόσμια κλίμακα.

Παρ' όλα αυτά, όπως όλα τα πράγματα δέχονται επικρίσεις, το ίδιο συμβαίνει και με τα Copulas. Ένας από τους λόγους για τους οποίους συμβαίνει αυτό, είναι η αλόγιστη χρήση τους από διάφορους επιστημονικούς κλάδους, κυρίως σε προβλήματα που εμφανίζουν στοχαστική εξάρτηση. Προφανώς δεν είναι αυτός ο ρόλος τους, όμως είναι ένα εργαλείο που έλυσε τα χέρια των επιστημόνων στο συγκεκριμένο τομέα και συνεχίζει να είναι ένα πολυσυζητημένο και άκρως ενδιαφέρον αντικείμενο μελέτης.

Στην παρούσα διπλωματική εργασία θα παρουσιάσουμε την έννοια των Copulas καθώς και τη χρησιμότητά τους όταν έχουμε συσχετισμένα δεδομένα. Η περίπτωση με την οποία θα ασχοληθούμε είναι κυρίως τα δισδιάστατα Copulas ώστε, πέρα από το θεωρητικό πλαίσιο το οποίο θα αναπτύξουμε, να μπορούμε να έχουμε γραφική απεικόνιση των αποτελεσμάτων που παρουσιάζονται. Συγκεκριμένα, η διάρθρωση της εργασίας έχει ως εξής:

Στο πρώτο κεφάλαιο δίνουμε κάποιους βασικούς ορισμούς ώστε να φτάσουμε τελικά στον ορισμό των Subcopulas και των Copulas. Στη συνέχεια δίνουμε κάποιες από τις βασικές ιδιότητες των Copulas και αναπτύσσουμε το Θεώρημα του Sklar, το οποίο αποτελεί τη βάση όλης αυτής της θεωρίας. Αναφέρουμε τα φράγματα των Frèchet-Hoeffding και αναπτύσσουμε και την περίπτωση των Copulas Επιβίωσης. Κλείνουμε το πρώτο κεφάλαιο με μία γενίκευση στις n διαστάσεις.

Στο δεύτερο κεφάλαιο υπενθυμίζουμε τις έννοιες της συσχέτισης και εξάρτησης δύο τυχαίων μεταβλητών και εισάγουμε την έννοια της συμφωνίας δύο τυχαίων μεταβλητών. Στη συνέχεια παρουσιάζουμε κάποιους συντελεστές συσχέτισης-εξάρτησης. Συγκεκριμένα, αναπτύσσουμε το συντελεστή γραμμικής συσχέτισης του Pearson και παρουσιάζουμε τους συντελεστές συσχέτισης τάξης Rho του Spearman και Tau του Kendall. Εν συνεχεία παρουσιάζουμε το συντελεστή εξάρτησης ουρών και κλείνουμε το κεφάλαιο με αναφορά

στη θετική τεταρτημοριακή εξάρτηση και στη μονοτονία στις ουρές.

Στο τρίτο κεφάλαιο αναπτύσσουμε δύο από τις πιο σημαντικές κλάσεις των Copulas, τα Elliptical και τα Archimedean Copulas. Από την οικογένεια των Elliptical Copulas εστιάζουμε κυρίως στις οικογένειες Gaussian Copulas και Student's t -Copulas, ενώ από την οικογένεια των Archimedean Copulas επικεντρωνόμαστε στις οικογένειες Clayton, Gumbel-Hougaard και Frank. Παράλληλα με την ανάπτυξη της θεωρίας που αφορά τις παραπάνω οικογένειες δίνουμε και διαγράμματα με προσομοιωμένα δεδομένα από τα συγκεκριμένα Copulas καθώς και Contour Plots.

Στο τέταρτο κεφάλαιο παρουσιάζουμε αλγόριθμους για την παραγωγή δεδομένων από τα Copula που δόθηκαν στο προηγούμενο κεφάλαιο. Στη συνέχεια παρουσιάζουμε μεθόδους εκτίμησης για τα Copulas. Συγκεκριμένα, δίνουμε μία παραμετρική, μια ημι-παραμετρική και μία μη παραμετρική μέθοδο εκτίμησης. Κλείνουμε το κεφάλαιο αυτό αναφέροντας δύο ελέγχους καλής προσαρμογής, δηλαδή δύο μεθόδους σύγκρισης των Copulas.

Στο τελευταίο κεφάλαιο έχουμε προσομοιώσει 150 δεδομένα από συγκεκριμένο Copula, κάνουμε έλεγχο καλής προσαρμογής ώστε να διαπιστώσουμε ποιο Copula προσαρμόζεται καλύτερα στα δεδομένα μας και στη συνέχεια εφαρμόζουμε μεθόδους εκτίμησης για την παράμετρο του Copula. Κλείνουμε το κεφάλαιο αυτό εξετάζοντας ποια μέθοδος είναι πιο αποτελεσματική, βάσει των συγκεκριμένων δεδομένων, υπολογίζοντας το μέσο τετραγωνικό σφάλμα.

Θεοδώρα Κοκκινιώτη

Πειραιάς

Ιούλιος 2014

Κεφάλαιο 1

COPULAS: ΟΡΙΣΜΟΙ-ΙΔΙΟΤΗΤΕΣ

1.1 ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό θα δώσουμε κάποιους αρχικούς ορισμούς, οι οποίοι είναι απαραίτητοι για να κατανοήσουμε τι είναι τα Copulas και για να αποτυπώσουμε το πλαίσιο στο οποίο βασίζονται. Η περίπτωση που θα μας απασχολήσει ιδιαίτερα είναι η δισδιάστατη περίπτωση. Έπειτα θα ορίσουμε τα Subcopulas και τα Copulas καθώς και θα παρουσιάσουμε κάποιες βασικές ιδιότητές τους. Εν συνεχεία θα δώσουμε τα φράγματα των Frèchet-Hoeffding για τα Copula και τις από κοινού συναρτήσεις κατανομής καθώς και το Θεώρημα του Sklar. Τέλος, θα ορίσουμε τα Copula Επιβίωσης και θα δώσουμε κάποιες βασικές ιδιότητές τους και θα κλείσουμε το κεφάλαιο αυτό παρουσιάζοντας μια γενίκευση των Copulas στις n διαστάσεις.

1.2 ΒΑΣΙΚΟΙ ΟΡΙΣΜΟΙ

Έστω δύο τυχαίες μεταβλητές X, Y με συναρτήσεις κατανομής F, G , αντίστοιχα, όπου $F(x) = P(X \leq x)$ και $G(y) = P(Y \leq y)$ και από κοινού συνάρτηση κατανομής $H(x, y) = P(X \leq x, Y \leq y)$. Επίσης, έστω $h(x, y)$ η από κοινού συνάρτηση πυκνότητας των παραπάνω τυχαίων μεταβλητών. Στη συνέχεια θα συμβολίζουμε με $\bar{\mathbf{R}}$ το επεκτεταμένο σύνολο των πραγματικών αριθμών το οποίο ορίζεται ως: $\bar{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$. Δίνουμε τους παρακάτω ορισμούς:

ΟΡΙΣΜΟΣ 1.2.1 (*H-όγκος*): Έστω S_1, S_2 υποσύνολα του $\bar{\mathbf{R}}$ και H μια δισδιάστατη πραγματική συνάρτηση με πεδίο ορισμού $S_1 \times S_2$. Θεωρούμε το ορθογώνιο παραλληλόγραμμο $B = [x_1, x_2] \times [y_1, y_2]$, του οποίου οι κορυφές ανήκουν στο $S_1 \times S_2$. Ο *H-όγκος* του B ορίζεται ως εξής:

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$

□

Με λίγα λόγια, όταν H είναι η από κοινού συνάρτηση κατανομής δύο συνεχών τυχαίων μεταβλητών, X, Y , ο *H-όγκος* είναι εύκολο να διαπιστωθεί ότι ισούται με την εξής πιθανότητα:

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} h(x, y) dx dy.$$

Πράγματι, χρησιμοποιώντας τον τύπο $P(AB^c) = P(A) - P(AB)$ μπορούμε να γράψουμε

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = P(X \leq x_2, y_1 \leq Y \leq y_2) - P(X < x_1, y_1 \leq Y \leq y_2) =$$

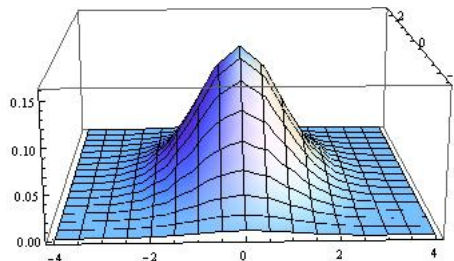
$$P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y < y_1) - P(X < x_1, y_1 \leq Y \leq y_2) =$$

$$P(X \leq x_2, Y \leq y_2) - P(X \leq x_2, Y \leq y_1) - P(X \leq x_1, Y \leq y_2) + P(X \leq x_1, Y \leq y_1) =$$

$$H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1).$$

Για παράδειγμα, έστω μια δισδιάστατη Τυποποιημένη Κανονική Κατανομή με $\rho = 0$. Για $x_1 = -4, x_2 = 4, y_1 = -2, y_2 = 2$ ο *H-όγκος* είναι ο όγκος που βρίσκεται πάνω από το ορθογώνιο με κορυφές τα $(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)$ και κάτω από την τρισδιάστατη επιφάνεια που φαίνεται στο παρακάτω σχήμα.

Σχήμα 1.1: Δισδιάστατη Τυπική Κανονική Κατανομή



Η δισδιάστατη συνάρτηση πυκνότητας της Τυποποιημένης Κανονικής Κατανομής για $\rho = 0$ είναι:

$$\phi(x, y) = \frac{1}{2\pi} \exp\left[-\frac{(x^2 + y^2)}{2}\right], \text{ όπου } (x, y) \in \mathbf{R}^2$$

και επομένως ο Φ -όγκος είναι ίσος με:

$$\begin{aligned} V_{\Phi}(B) &= \int_{-2}^2 \int_{-4}^4 \phi(x, y) dx dy = \Phi(4, 2) - \Phi(4, -2) - \Phi(-4, 2) + \Phi(-4, -2) \\ &= \text{Erf}[\sqrt{2}]\text{Erf}[2\sqrt{2}] = 0.954439 > 0 \end{aligned}$$

όπου

$$\text{Erf}[z] = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt.$$

ΟΡΙΣΜΟΣ 1.2.2 (2-αύξουσα συνάρτηση): Μια δισδιάστατη πραγματική συνάρτηση καλείται 2-αύξουσα αν $V_H(B) \geq 0$ για όλα τα ορθογώνια παραλληλόγραμμα B των οποίων οι κορυφές ανήκουν στο $S_1 \times S_2$. □

Να τονίσουμε σε αυτό το σημείο ότι η συνθήκη $H(x, y) \geq 0$ είναι ικανή αλλά όχι αναγκαία για να έχουμε μια 2-αύξουσα συνάρτηση. Ο H -όγκος μπορεί να βγει μη αρνητικός ακόμα και αν η συνάρτηση που ολοκληρώνεται δεν είναι θετική σε όλο το πεδίο ορισμού της. Όταν, όμως, έχουμε συνάρτηση πυκνότητας, και επομένως, συνάρτηση κατανομής, τότε ο H -όγκος είναι μη αρνητικός. Επίσης, η μονοτονία αυτού του τύπου δεν συνεπάγεται μονοτονία κατά συντεταγμένες ούτε αντίστροφα.

Για παράδειγμα¹, έστω H η συνάρτηση $H(x, y) = \max(x, y)$ ορισμένη στο I^2 , όπου $I = [0, 1]$. Η H είναι αύξουσα κατά συντεταγμένες: Έστω $(x_1, y_1), (x_2, y_2)$ δύο σημεία στο $[0, 1]^2$ με $x_1 < x_2, y_1 < y_2$. Τότε,

$$H(x_1, y_1) = \max(x_1, y_1) \leq \max(x_2, y_2) = H(x_2, y_2)$$

δηλαδή η H είναι αύξουσα κατά συντεταγμένες. Όμως ο H -όγκος για το $B = [0, 1]^2$ είναι ίσος με:

$$V_H(I^2) = H(1, 1) - H(1, 0) - H(0, 1) + H(0, 0) = 1 - 1 - 1 + 0 = -1 < 0.$$

¹ Το συγκεκριμένο παράδειγμα υπάρχει στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28].

Επομένως η H δεν είναι 2-άξουσα. Αντίστροφα, έστω H η συνάρτηση που δίνεται από τον τύπο $H(x, y) = (2x - 1)(2y - 1)$ ορισμένη στο I^2 . Η H είναι δύο άξουσα καθώς ο H -όγκος κάθε $B = [x_1, x_2] \times [y_1, y_2]$ με $x_1 < x_2, y_1 < y_2$ υπολογίζεται:

$$\begin{aligned} V_H(B) &= H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) = \\ &= ((2x_2 - 1) - (2x_1 - 1))((2y_2 - 1) - (2y_1 - 1)) \geq 0. \end{aligned}$$

Παρό όλα αυτά δεν είναι άξουσα κατά συντεταγμένες. Για παράδειγμα, αν θεωρήσουμε $(x_1, y_1), (x_2, y_2)$ δύο σημεία στο $[0, 1]^2$ με $x_1 < x_2, y_1 < y_2$ και $x_1, x_2, y_1 \in (0, \frac{1}{2}), y_2 \in (\frac{1}{2}, 1)$, τότε θα ισχύει

$$H(x_2, y_2) < 0 < H(x_1, y_1).$$

ΟΡΙΣΜΟΣ 1.2.3 (*Grounded συνάρτηση*): Η από κοινού συνάρτηση H με πεδίο ορισμού το $S_1 \times S_2$ ονομάζεται grounded, αν $H(x, b) = 0 = H(a, y)$, όπου a, b είναι τα ελάχιστα στοιχεία των συνόλων S_1, S_2 , αντίστοιχα. \square

Προφανώς αν είμαστε στη συνεχή περίπτωση, που είναι αυτή η οποία και θα μας απασχολήσει στην παρούσα εργασία, το παραπάνω αποτέλεσμα ισχύει καθώς θα έχουμε:

$$\begin{aligned} H(x, y_{min}) &= P(X \leq x, Y \leq y_{min}) = P(X \leq x, Y = y_{min}) = 0, \\ H(x_{min}, y) &= P(X \leq x_{min}, Y \leq y) = P(X = x_{min}, Y \leq y) = 0. \end{aligned}$$

Παρό όλα αυτά δεν ισχύει πάντα για τη διακριτή περίπτωση γι' αυτό και δόθηκε ο παραπάνω ορισμός. Να σημειώσουμε ότι αποδεικνύεται ότι η H είναι άξουσα κατά συντεταγμένες, αν είναι 2-άξουσα και grounded συνάρτηση (βλπ. Nelsen [28]).

Υπενθυμίζουμε ότι θεωρούμε ότι η (αθροιστική) συνάρτηση κατανομής F είναι μια συνάρτηση με πεδίο ορισμού το $\bar{\mathbf{R}}$, η οποία είναι άξουσα και έχει τις παρακάτω οριακές ιδιότητες:

$$F(\infty) = \lim_{t \rightarrow \infty} F(t) = 1 \text{ και } F(-\infty) = \lim_{t \rightarrow -\infty} F(t) = 0.$$

Επίσης, η από κοινού (αθροιστική) συνάρτηση κατανομής H είναι μια συνάρτηση με πεδίο ορισμού το $\bar{\mathbf{R}}^2$, η οποία είναι 2-άξουσα και έχει τις παρακάτω οριακές ιδιότητες:

$$\begin{aligned} H(x, -\infty) &= \lim_{y \rightarrow -\infty} H(x, y) = 0, \quad H(-\infty, y) = \lim_{x \rightarrow -\infty} H(x, y) = 0 \text{ και} \\ H(\infty, \infty) &= \lim_{x \rightarrow \infty, y \rightarrow \infty} H(x, y) = 1. \end{aligned}$$

Για το υπόλοιπο της εργασίας θα θεωρούμε ότι το πεδίο τιμών μιας συνάρτησης, έστω F , συμβολίζεται με $\mathcal{R}(F)$, ενώ το πεδίο ορισμού της με $DomF$.

ΟΡΙΣΜΟΣ 1.2.4 (*Γενικευμένη αντίστροφη*): Έστω F μια συνάρτηση κατανομής. Τότε ορίζεται ως γενικευμένη αντίστροφη, (generalised inverse), της F κάθε συνάρτηση $F^{(-1)}$ με $DomF = [0, 1]$ για την οποία ισχύει:

$$F^{(-1)}(t) = \inf\{x|F(x) \geq t\} = \sup\{x|F(x) \leq t\}.$$

□

Για τα δισδιάστατα Copulas παροτρύνουμε τον αναγνώστη να κοιτάξει το δεύτερο κεφάλαιο του βιβλίου του Nelsen, *An Introduction to Copulas*, [28] καθώς και το πρώτο κεφάλαιο του βιβλίου των Balakrishnan και Lai, *Continuous Bivariate Distributions*, [4]. Επίσης, ο αναγνώστης μπορεί να ανατρέξει και στην εργασία των Schweizer και Wolff [29] για μία συνοπτική μελέτη των δισδιάστατων Copulas, συμπεριλαμβανομένου του ορισμού που δίνεται στην παρούσα εργασία καθώς και ιδιοτήτων που αναφέρουμε στη συνέχεια.

1.3 ΒΑΣΙΚΕΣ ΙΔΙΟΤΗΤΕΣ

Στην παράγραφο αυτή δίνεται ο μαθηματικός ορισμός των Copulas και στη συνέχεια κάποιες βασικές ιδιότητές τους.

ΟΡΙΣΜΟΣ 1.3.1 (*Subcopulas και ιδιότητές τους*): Ένα δισδιάστατο Subcopula (2-Subcopula) είναι μια συνάρτηση C' η οποία έχει τις παρακάτω ιδιότητες:

1. $DomC' = S_1 \times S_2$, όπου S_1, S_2 είναι υποσύνολα του $I = [0, 1]$ τα οποία περιέχουν το 0 και το 1.
2. Η C' είναι 2-αύξουσα και grounded.
3. Για κάθε u στο S_1 και για κάθε v στο S_2 ισχύει

$$C'(u, 1) = u \text{ και } C'(1, v) = v.$$

□

Το δισδιάστατο Copula είναι μια γενίκευση του 2-Subcopula με τη διαφορά ότι το πεδίο ορισμού είναι το I^2 . Έτσι, έχουμε:

ΟΡΙΣΜΟΣ 1.3.2 (*Copulas και ιδιότητές τους*): Ένα δισδιάστατο Copula (2-Copula) είναι μια συνάρτηση $C : I^2 \rightarrow I$ με τις εξής ιδιότητες:

1. Για κάθε $u, v \in I$ ισχύουν

$$C(u, 0) = 0 = C(0, v) \text{ και } C(u, 1) = u, C(1, v) = v.$$

2. Για κάθε $u_1, u_2, v_1, v_2 \in I$ τέτοια ώστε $u_1 \leq u_2$ και $v_1 \leq v_2$ ισχύει

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

Με απλά λόγια η C είναι μια συνάρτηση που παίρνει τιμές στο $[0, 1]$. Επίσης, είναι μια grounded συνάρτηση καθώς για την ελάχιστη τιμή της μίας περιθώριας και για όλες τις τιμές της άλλης η από κοινού μηδενίζεται. Η ελάχιστη τιμή της πρώτης επιτυγχάνεται για $u = 0$ και αντίστοιχα της δεύτερης για $v = 0$. Επιπρόσθετα, η C είναι 2-αύξουσα καθώς ο C -όγκος είναι μεγαλύτερος ή ίσος από το 0. Αν η $C(u, v)$ είναι από κοινού συνάρτηση κατανομής μιας δισδιάστατης τυχαίας μεταβλητής, τότε η ιδιότητα $C(u, 1) = u, C(1, v) = v$, προκύπτει άμεσα όταν τα u, v είναι οι περιθώριες συναρτήσεις κατανομής.

Για παράδειγμα:

$$C(u, 1) = \lim_{v \rightarrow 1} P(U \leq u, V \leq v) = P(U \leq u, V \leq 1) = P(U \leq u) = u$$

και ουσιαστικά εξασφαλίζει την ύπαρξη των περιθωρίων. Βέβαια το C δεν ορίστηκε σαν από κοινού συνάρτηση κατανομής αλλά, όπως θα δούμε στην επόμενη ενότητα, αυτός είναι ο επιθυμητός στόχος.

Ας δούμε όμως ένα παράδειγμα²:

Παράδειγμα: Δίνεται η συνάρτηση $C(u, v) = \frac{uv}{u + v - uv}$ με $(u, v) \in [0, 1]^2$. Θα αποδείξουμε ότι το C είναι Copula³:

Έχουμε ότι

$$C(u, v) \geq 0 \text{ καθώς } uv \geq 0 \text{ και } u + v - uv = u(1 - v) + v \geq 0.$$

Επίσης,

$$C(u, v) \leq 1 \text{ καθώς αν } uv \leq u + v - uv$$

$$u + v - uv - uv = u(1 - v) + v(1 - u) \geq 0.$$

Επίσης,

$$C(u, 0) = 0 = C(0, v) \text{ και } C(u, 1) = u, C(1, v) = v.$$

Τέλος, για κάθε $u_1, u_2, v_1, v_2 \in [0, 1]$ με $u_1 \leq u_2$ και $v_1 \leq v_2$ ο C -όγκος υπολογίζεται από τον τύπο:

$$\begin{aligned} V_C([u_1, v_1] \times [u_2, v_2]) &= C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) = \\ &= \frac{u_2 v_2}{u_2 + v_2 - u_2 v_2} - \frac{u_2 v_1}{u_2 + v_1 - u_2 v_1} - \frac{u_1 v_2}{u_1 + v_2 - u_1 v_2} + \frac{u_1 v_1}{u_1 + v_1 - u_1 v_1} \end{aligned}$$

και είναι μη αρνητικό καθώς οι παρονομαστές είναι θετικοί άρα και το γινόμενο τους, ενώ ο αριθμητής που προκύπτει έπειτα από πράξεις είναι:

² Το συγκεκριμένο παράδειγμα βρίσκεται στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28].

³ Το συγκεκριμένο Copula είναι το Copula της διδιάστατης λογιστικής κατανομής του Gumbel.

$$v_1 v_2 (u_1 + u_2 - u_1 u_2) + u_1 u_2 (v_1 + v_2 - v_1 v_2) \geq 0.$$

Επομένως το C είναι ένα Copula, σύμφωνα με τον παραπάνω ορισμό.

Ας δούμε, όμως, και τι έχουμε αν πάρουμε τα άκρα του διαστήματος:

$$\begin{aligned} V_C(I^2) &= C(1, 1) - C(1, 0) - C(0, 1) + C(0, 0) = \\ &= 1 - 0 - 0 + \lim_{u \rightarrow 0, v \rightarrow 0} \frac{uv}{u + v - uv} = 1 + \lim_{u \rightarrow 0, v \rightarrow 0} \frac{uv}{uv \left(\frac{1}{v} + \frac{1}{u} - 1 \right)} = \\ &= 1 + \lim_{u \rightarrow 0, v \rightarrow 0} \frac{1}{\left(\frac{1}{v} + \frac{1}{u} - 1 \right)} = 1 + 0 = 1. \end{aligned}$$

Η παραπάνω πιθανότητα που υπολογίστηκε στα άκρα του διαστήματος $[0, 1]$ ήταν αναμενόμενο να είναι 1 καθώς είναι ουσιαστικά η πιθανότητα που αντιστοιχεί σε ολόκληρο το δειγματικό χώρο. \square

Μία ακόμα ιδιότητα των Copulas είναι ότι είναι ομοιόμορφα συνεχείς συναρτήσεις.⁴ Πιο συγκεκριμένα, για κάθε $(u_1, v_1), (u_2, v_2) \in \text{Dom}C$ ισχύει:

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

Φράγματα Frèchet-Hoeffding για Copulas⁵:

Όπως αναφέραμε στην εισαγωγή, οι Frèchet και Hoeffding συνέβαλαν στη δημιουργία ορίων για τα Copula. Η αλήθεια είναι πως τα αποτελέσματα αποδείχθηκαν για τα Subcopula αλλά ισχύουν και για τα Copula και για αυτά θα τα χρησιμοποιήσουμε στην παρούσα εργασία.

⁴ Η απόδειξη συμπεριλαμβάνεται στο βιβλίο *Copula Methods in Finance* των Cherubini, Luciano και Vecchiato [6], στο παράρτημα του δεύτερου κεφαλαίου.

⁵ Για τα φράγματα των Frèchet-Hoeffding παροτρύνουμε τον αναγνώστη στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28].

Πρόταση: Έστω C ένα copula. Τότε για κάθε u, v στο $DomC = I^2$ ισχύει:

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v)$$

Απόδειξη: Έστω (u, v) ένα τυχαίο σημείο στο I^2 . Ισχύει ότι

$$C(u, v) \leq C(u, 1) = u \text{ και } C(u, v) \leq C(1, v) = v.$$

Επομένως,

$$C(u, v) \leq \min(u, v).$$

Επίσης, έχουμε ότι η C είναι 2-αύξουσα και άρα

$$V_C([u, 1] \times [v, 1]) \geq 0.$$

Όμως:

$$V_C([u, 1] \times [v, 1]) = C(1, 1) - C(1, v) - C(u, 1) + C(u, v) = 1 - v - u + C(u, v).$$

Συνεπώς,

$$C(u, v) \geq u + v - 1 \text{ και } C(u, v) \geq 0,$$

άρα

$$C(u, v) \geq \max(u + v - 1, 0).$$

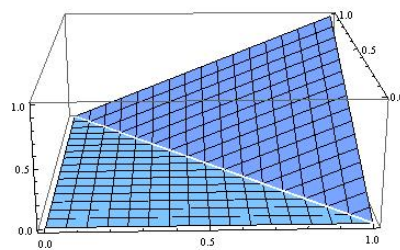
Τα φράγματα $W(u, v) = \max(u + v - 1, 0)$ και $M(u, v) = \min(u, v)$ ονομάζονται Frèchet-Hoeffding κάτω και άνω όριο, αντίστοιχα, και είναι πολύ σημαντικά καθώς, όταν επιτυγχάνονται, είμαστε σε κάποια ακραία ή ειδική περίπτωση. Αποδεικνύεται ότι τα M, W στη δισδιάστατη περίπτωση είναι Copulas.

Παρατηρούμε ότι το πεδίο τιμών του C θα είναι το $[0, 1]$ καθώς, όταν τα u, v παίρνουν τη μέγιστη τιμή τους, το άνω όριο είναι ίσο με 1, ενώ, όταν παίρνουν την ελάχιστη τιμή τους, το κάτω όριο είναι ίσο με 0. Αυτό είναι σημαντικό, αφού θέλουμε να ορίσουμε τα Copulas ως συναρτήσεις κατανομής. Επομένως, αν έχουμε δύο τυχαίες μεταβλητές X, Y με περιθώριες συναρτήσεις κατανομής F, G αντίστοιχα και H είναι η από κοινού συνάρτηση κατανομής, τα φράγματα των Frèchet-Hoeffding για τις από κοινού συναρτήσεις κατανομής είναι:

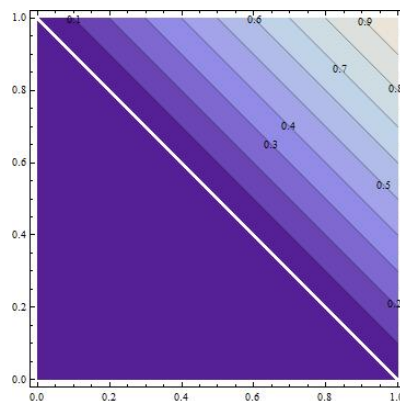
$$\max(F(x) + G(y) - 1, 0) \leq H(x, y) \leq \min(F(x), G(y))$$

Εν συνεχεία, θα δούμε τα παραπάνω φράγματα και οπτικά. Παρακάτω δίνονται τα διαγράμματα στις τρεις διαστάσεις του κάτω και του άνω ορίου (βλπ. σχήματα 1.2 και 1.4) καθώς και τα αντίστοιχα Contour Plots (βλπ. σχήματα 1.3 και 1.5). Τα Contour Plots απεικονίζουν μονοδιάστατες καμπύλες για τις οποίες θεωρούμε ότι η από κοινού, εδώ τα Copula, είναι σταθερά. Δηλαδή το δισδιάστατο σχήμα προκύπτει από το τρισδιάστατο, αν φέρουμε οριζόντιες τομές με επίπεδο παράλληλο προς το επίπεδο $x \circ y$.

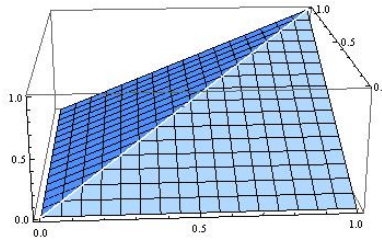
Σχήμα 1.2: Plot 3D: Κάτω φράγμα (3D) ενός δισδιάστατου Copula



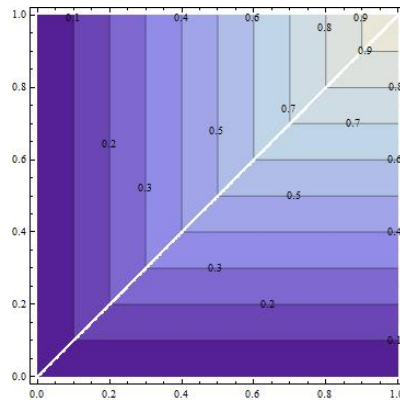
Σχήμα 1.3: Contour Plot : Κάτω φράγμα ενός δισδιάστατου Copula



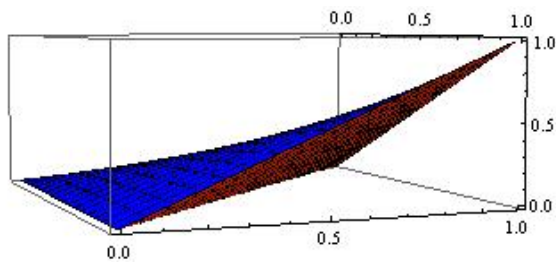
Σχήμα 1.4: Plot 3D: Άνω φράγμα (3D) ενός δισδιάστατου Copula



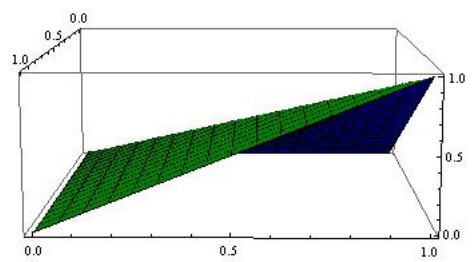
Σχήμα 1.5: Contour Plot: Άνω φράγμα ενός δισδιάστατου Copula



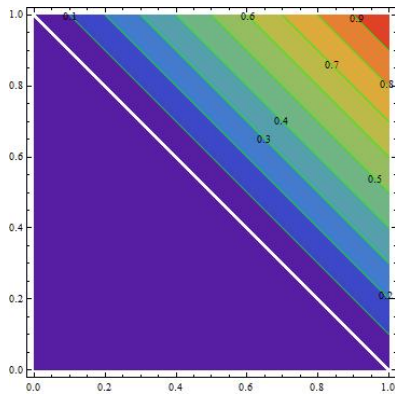
Ας δούμε όμως και κάποιο Copula μαζί με τα αντίστοιχα φράγματα. Στα παρακάτω σχήματα (Σχήματα 1.6, 1.7) έχουμε το Copula $C(u, v) = \frac{uv}{u + v - uv}$ το οποίο χρησιμοποιήσαμε προηγουμένως. Με μπλε χρώμα είναι το παραπάνω Copula, με πράσινο χρώμα το άνω φράγμα και με πορτοκαλί το κάτω φράγμα. Ακολουθούν και τα Contour Plots των φραγμάτων και του Copula, (Σχήματα 1.8, 1.9, 1.10), καθώς και ένα Contour Plot στο οποίο απεικονίζονται οι Contour lines των τριών Copulas σε ένα σχήμα (Σχήμα 1.11).



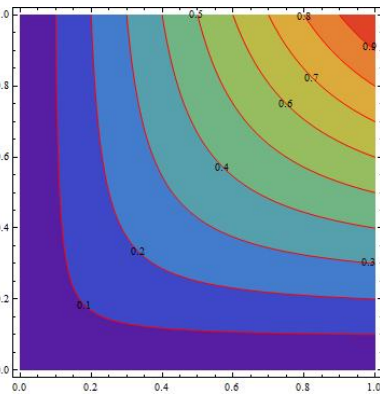
Σχήμα 1.6: Plot 3D : $C(u, v)$ μαζί με το κάτω φράγμα



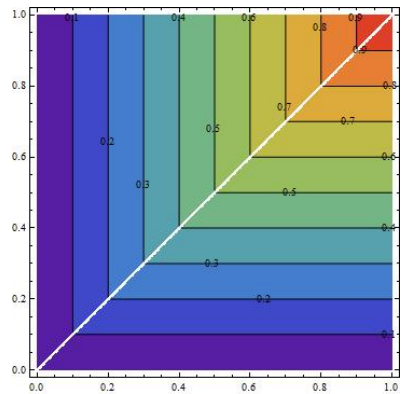
Σχήμα 1.7: Plot 3D : $C(u, v)$ μαζί με το άνω φράγμα



Σχήμα 1.8: Contour Plot του κάτω φράγματος



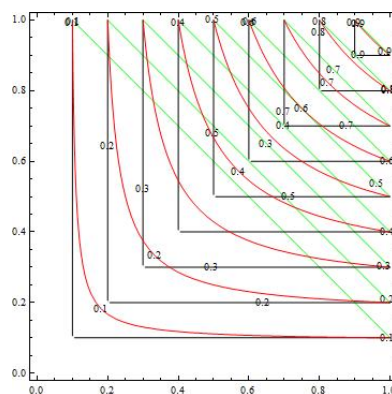
Σχήμα 1.9: Contour Plot του Copula



Σχήμα 1.10: Contour Plot του άνω φράγματος

Με πράσινο χρώμα είναι οι γραμμές του Contour Plot του κάτω φράγματος, με κόκκινο του Copula, ενώ με μαύρο είναι του άνω φράγματος. Αυτό που παρατηρούμε είναι ότι σε κάθε περιοχή (τρίγωνο) που σχηματίζουν τα φράγματα, οι Contour lines του Copula εμπεριέχονται στο τρίγωνο αυτό, όπως ήταν αναμενόμενο. Μάλιστα, όταν το Copula τείνει να πάρει την τιμή 1, πλησιάζει τη γραμμή του κάτω φράγματος, ενώ όταν τείνει να πάρει την τιμή 0, πλησιάζει τη γραμμή του άνω φράγματος.

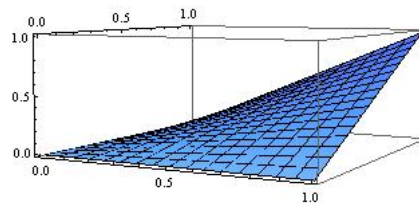
Σχήμα 1.11: Contour Lines του Copula και των φραγμάτων



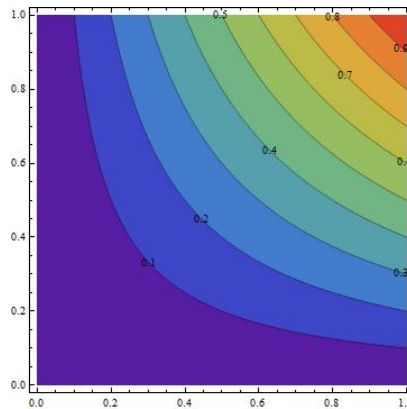
Όπως αναφέραμε και νωρίτερα, τα φράγματα των Frèchet-Hoeffding είναι πολύ σημαντικά καθώς, όταν επιτυγχάνονται βρισκόμαστε σε κάποια οριακή κατάσταση. Ένα εξίσου σημαντικό Copula είναι και το Copula-γινόμενο (Product Copula), $C(u, v) = uv$, και τούτο γιατί, μέσω αυτού εκφράζεται η ανεξαρτησία των τυχαίων μεταβλητών, αν θεωρήσουμε τις αντίστοιχες συναρτήσεις κατανομής. Συγκεκριμένα ισχύει το παρακάτω αποτέλεσμα:

Πρόταση: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με περιθώριες F, G αντίστοιχα. Τότε οι X, Y είναι ανεξάρτητες αν και μόνο αν $C(F(x), G(y)) = F(x)G(y)$. □

Σχήμα 1.12: Plot 3D: Το Copula-γινόμενο



Σχήμα 1.13: Contour Plot του Copula-γινόμενο



Το Copula-γινόμενο είναι εκείνο που συνδέεται με την ανεξαρτησία δύο τυχαίων μεταβλητών. Όπως θα αναφέρουμε και αργότερα στην παρούσα εργασία, τα Copula δίνουν τη δομή εξάρτησης των δεδομένων μας. Επομένως, όταν επιτυγχάνονται τα φράγματα θα έχουμε είτε τέλεια θετική συσχέτιση (άνω φράγμα) είτε τέλεια αρνητική συσχέτιση (κάτω φράγμα). Σύμφωνα με τους Schweizer και Wolff, το Copula-γινόμενο, (στις δύο διαστάσεις), θα βρίσκεται στη μέση (sits midway) μεταξύ των δύο φραγμάτων (βλπ. Schweizer και Wolff [29]).

1.4 ΘΕΩΡΗΜΑ ΤΟΥ SKLAR

Στην ενότητα αυτή θα παρουσιάσουμε το Θεώρημα του Sklar, το οποίο αποτελεί τη βάση της Θεωρίας των Copulas, και θα δούμε τι αλλαγές θα γίνουν στα Copulas όταν χρησιμοποιούμε γνησίως μονότονους μετασχηματισμούς των τυχαίων μεταβλητών.

Θεώρημα του Sklar:

Έστω H η από κοινού συνάρτηση κατανομής δύο τυχαίων μεταβλητών X και Y με περιθώριες F, G , αντιστοίχως. Τότε υπάρχει Copula C τέτοιο, ώστε για όλα τα $x, y \in \bar{\mathbf{R}}$ να ισχύει:

$$H(x, y) = C(F(x), G(y))$$

Αν οι F, G είναι συνεχείς τότε εξασφαλίζεται η μοναδικότητα του C , αλλιώς το C είναι μοναδικώς ορισμένο στο $\mathcal{R}(F) \times \mathcal{R}(G)$.

Αντίστροφα, αν C είναι ένα Copula και F, G είναι συναρτήσεις κατανομής, τότε η συνάρτηση H , που δίνεται από τον προηγούμενο τύπο, είναι μια από κοινού συνάρτηση κατανομής με περιθώριες F, G .⁶ □

Στο σημείο αυτό θα εξηγήσουμε την ουσία του θεωρήματος. Το σημαντικό σε αυτό το θεώρημα είναι ότι ορίζει το Copula ως από κοινού συνάρτηση κατανομής και το συνδέει με τις περιθώριες. Επομένως, αν η επιλογή του Copula είναι σωστή, και εννοείται ότι οι περιθώριες είναι γνωστές, τότε έχουμε λύσει το πρόβλημα εύρεσης από κοινού συνάρτησης κατανομής δύο τυχαίων μεταβλητών. Επίσης, το παραπάνω θεώρημα διαχωρίζει τη μελέτη της δομής εξάρτησης των δεδομένων από τη μελέτη των περιθωρίων, καθώς η επιλογή των περιθωρίων συναρτήσεων κατανομής δεν επηρεάζει την επιλογή του Copula. Για τη διαδικασία της επιλογής του Copula θα αναφερθούμε αργότερα, ας δούμε τώρα το αντίστροφο του θεωρήματος το οποίο λέει ότι μπορούμε να κατασκευάσουμε το Copula αν έχουμε τις περιθώριες και την από κοινού. Συγκεκριμένα, ισχύει:

⁶ Για την απόδειξη του Θεωρήματος καθώς και για το Πρόβλημα που ακολουθεί παροτρύνουμε τον αναγνώστη στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28].

Πόρισμα: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με περιθώριες συναρτήσεις κατανομής F, G , από κοινού συνάρτηση κατανομής H και Copula C . Επίσης, έστω $F^{(-1)}, G^{(-1)}$ οι γενικευμένες αντίστροφες των F, G , αντιστοίχως. Τότε, για κάθε $(u, v) \in I^2$ ισχύει:

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v))$$

□

Παράδειγμα⁷: Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής τη διοδιάστατη λογιστική κατανομή του Gumbel (1961) η οποία δίνεται από τον τύπο:

$$H(x, y) = (1 + e^{-x} + e^{-y})^{-1}.$$

Εύκολα αποδεικνύεται ότι οι X, Y έχουν μονοδιάστατες λογιστικές κατανομές αφού:

$$F(x) = \lim_{y \rightarrow \infty} H(x, y) = (1 + e^{-x})^{-1} \text{ και } G(y) = \lim_{x \rightarrow \infty} H(x, y) = (1 + e^{-y})^{-1}.$$

Εύκολα αποδεικνύεται επίσης ότι οι γενικευμένες αντίστροφες των F, G , (εδώ είναι οι γνωστές αντίστροφες καθώς οι F, G είναι γνησίως αύξουσες συναρτήσεις) είναι:

$$F^{-1}(u) = \ln u - \ln(1 - u), G^{-1}(v) = \ln v - \ln(1 - v) \text{ για κάθε } u, v \in I.$$

Πράγματι,

$$F(x) = u$$

αν και μόνο αν

$$1 + e^{-x} = \frac{1}{u}.$$

Άρα

$$e^{-x} = \frac{1 - u}{u} \Rightarrow x = \log u - \log(1 - u)$$

οπότε $F^{-1}(u) = \log u - \log(1 - u)$. Όμοια για την $G^{-1}(v)$.

Αφού $\mathcal{R}(F) = \mathcal{R}(G) = I \Rightarrow \text{Dom}C = I^2$ και από το πόρισμα προκύπτει:

⁷ Το συγκεκριμένο παράδειγμα υπάρχει στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28] καθώς και στην εργασία των Bouyé, Durrleman, Nikeghbali, Riboulet και Roncalli [5].

$$C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v)) = \frac{uv}{u + v - uv}.$$

Δηλαδή, κατασκευάσαμε το Copula! Το συγκεκριμένο μάλιστα ανήκει σε μια συγκεκριμένη κλάση των Copulas γνωστή ως Archimedean Copulas και στην οποία θα αναφερθούμε αργότερα. Η αλήθεια είναι πως το παραπάνω Copula δεν είναι εύχρηστο για περαιτέρω μελέτη καθώς δεν έχει κάποια παράμετρο. Παρ' όλα αυτά γενικεύεται αλλά δε θα επεκταθούμε περισσότερο στην παρούσα φάση.

Θα κλείσουμε την ενότητα αυτή με ένα θεώρημα⁸ που αφορά τη συμπεριφορά του Copula σε γνησίως μονότονους μετασχηματισμούς των X, Y . Θα θεωρούμε ότι $\mathcal{R}(X), \mathcal{R}(Y)$ είναι τα αντίστοιχα πεδία τιμών των τυχαίων μεταβλητών.

Θεώρημα: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με Copula C_{XY} και $\alpha(\cdot), \beta(\cdot)$ γνησίως μονότονες συναρτήσεις στο $\mathcal{R}(X)$ και $\mathcal{R}(Y)$, αντιστοίχως. Τότε:

1. Αν α, β είναι γνησίως αύξουσες, τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = C_{XY}(u, v)$$

2. Αν α γνησίως αύξουσα, β γνησίως φθίνουσα τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = u - C_{XY}(u, 1 - v)$$

3. Αν α γνησίως φθίνουσα, β γνησίως αύξουσα τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = v - C_{XY}(1 - u, v)$$

4. Αν α, β γνησίως φθίνουσες τότε

$$C_{\alpha(X), \beta(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v)$$

Με απλά λόγια, σε γνησίως αύξοντες μετασχηματισμούς των X, Y το Copula παραμένει αμετάβλητο αλλιώς στους άλλους γνησίως μονότονους μετασχηματισμούς το Copula προκύπτει με απλό τρόπο. Σε κάθε περίπτωση πάντως η επιλογή των α, β δεν παίζει ρόλο,

⁸ Το συγκεκριμένο θεώρημα συμπεριλαμβάνεται στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28], στην εργασία του Berthold Schweizer, *Thirty years of Copulas*, η οποία συμπεριλαμβάνεται στο βιβλίο των Dall'Aglio, Kotz και Salinetti [7], καθώς και στην εργασία των Schweizer και Wolff [29].

όπως φαίνεται από το παραπάνω θεώρημα. Θα δούμε στη συνέχεια πώς προκύπτουν τα αποτελέσματα του παραπάνω θεωρήματος μέσω ενός παραδείγματος.

Παράδειγμα:

1. Έστω $\alpha(X) = \log X, \beta(Y) = \log Y$ και F_1, G_1, F_2, G_2 οι συναρτήσεις κατανομής των $X, Y, \alpha(X), \beta(Y)$ αντίστοιχα. Τότε:

$$F_2(x) = P(\alpha(X) \leq x) = P(\log X \leq x) = P(X \leq e^x) = F_1(e^x)$$

όμοια:

$$G_2(y) = G_1(e^y).$$

Επομένως:

$$\begin{aligned} C_{\alpha(X), \beta(Y)}(F_2(x), G_2(y)) &= P(\alpha(X) \leq x, \beta(Y) \leq y) = P(\log X \leq x, \log Y \leq y) = \\ &P(X \leq e^x, Y \leq e^y) = C_{XY}(F_1(e^x), G_1(e^y)) = C_{XY}(F_2(x), G_2(y)). \end{aligned}$$

2. Έστω $\alpha(X) = \log X, \beta(Y) = \log(-Y)$ και F_1, G_1, F_2, G_2 οι συναρτήσεις κατανομής των $X, Y, \alpha(X), \beta(Y)$ αντίστοιχα. Τότε:

$$\begin{aligned} F_2(x) &= P(\alpha(X) \leq x) = P(\log X \leq x) = P(X \leq e^x) = F_1(e^x) \\ G_2(y) &= P(\beta(Y) \leq y) = P(\log(-Y) \leq y) = P(Y \geq -e^y) = \\ &1 - P(Y \leq -e^y) = 1 - G_1(-e^y) \end{aligned}$$

Επομένως:

$$\begin{aligned} C_{\alpha(X), \beta(Y)}(F_2(x), G_2(y)) &= \\ &P(\alpha(X) \leq x, \beta(Y) \leq y) = P(\log X \leq x, \log(-Y) \leq y) = \\ &P(X \leq e^x, Y \geq -e^y) = P(X \leq e^x) - P(X \leq e^x, Y \leq -e^y) = F_1(e^x) - C_{XY}(F_1(e^x), G_1(-e^y)) = \\ &F_2(x) - C_{XY}(F_2(x), 1 - G_2(y)). \end{aligned}$$

3. Έστω $\alpha(X) = \log(-X), \beta(Y) = \log(Y)$ και F_1, G_1, F_2, G_2 οι συναρτήσεις κατανομής των $X, Y, \alpha(X), \beta(Y)$ αντίστοιχα. Τότε:

$$\begin{aligned} F_2(x) &= P(\alpha(X) \leq x) = P(\log(-X) \leq x) = P(X \geq -e^x) = \\ &1 - P(X \leq -e^x) = 1 - F_1(-e^x) \\ G_2(y) &= P(\beta(Y) \leq y) = P(\log Y \leq y) = P(Y \leq e^y) = G_1(e^y). \end{aligned}$$

Επομένως:

$$\begin{aligned}
C_{\alpha(X),\beta(Y)}(F_2(x), G_2(y)) &= \\
P(\alpha(X) \leq x, \beta(Y) \leq y) &= P(\log(-X) \leq x, \log Y \leq y) = \\
P(X \geq -e^x, Y \leq e^y) &= P(Y \leq e^y) - P(X \leq -e^x, Y \leq e^y) = G_1(e^y) - C_{XY}(F_1(-e^x), G_1(e^y)) = \\
&G_2(y) - C_{XY}(1 - F_2(x), G_2(y)).
\end{aligned}$$

4. Έστω $\alpha(X) = \log(-X), \beta(Y) = \log(-Y)$ και F_1, G_1, F_2, G_2 οι συναρτήσεις κατανομής των $X, Y, \alpha(X), \beta(Y)$ αντίστοιχα. Τότε:

$$\begin{aligned}
F_2(x) &= P(\alpha(X) \leq x) = P(\log(-X) \leq x) = P(X \geq -e^x) = \\
&1 - P(X \leq -e^x) = 1 - F_1(-e^x) \\
G_2(y) &= P(\beta(Y) \leq y) = P(\log(-Y) \leq y) = P(Y \geq -e^y) = \\
&1 - P(Y \leq -e^y) = 1 - G_1(-e^y)
\end{aligned}$$

Επομένως:

$$\begin{aligned}
C_{\alpha(X),\beta(Y)}(F_2(x), G_2(y)) &= \\
P(\alpha(X) \leq x, \beta(Y) \leq y) &= P(\log(-X) \leq x, \log(-Y) \leq y) = \\
P(X \geq -e^x, Y \geq -e^y) &= P(X \geq -e^x) - P(X \geq -e^x, Y \leq -e^y) = \\
1 - P(X \leq -e^x) - P(X \geq -e^x, Y \leq -e^y) &= \\
1 - P(X \leq -e^x) - P(Y \leq -e^y) + P(X \leq -e^x, Y \leq -e^y) &= \\
1 - (1 - F_2(x)) - (1 - G_2(y)) + C_{XY}(1 - F_2(x), 1 - G_2(y)) &= F_2(x) + G_2(y) - 1 + C_{XY}(1 - F_2(x), 1 - G_2(y)).
\end{aligned}$$

Θα δώσουμε σε αυτό το σημείο ένα παράδειγμα για το ότι τα Copula παραμένουν αναλλοίωτα κάτω από γνησίως αύξοντες μετασχηματισμούς χρησιμοποιώντας προσομοιωμένα δεδομένα:

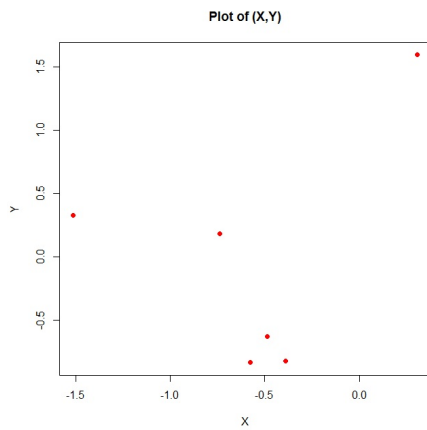
Παράδειγμα⁹: Παράγουμε δύο τυχαία δείγματα μεγέθους $n = 6$ από την $N(0, 1)$. Επίσης επιλέγουμε γνησίως αύξοντες μετασχηματισμούς, έστω $\alpha(X) = e^X$ και $\beta(Y) = 3e^Y$.

⁹ Το συγκεκριμένο παράδειγμα, (προφανώς με άλλα δεδομένα καθώς προκύπτουν από προσομοίωση), συμπεριλαμβάνεται στην εργασία των Genest και Favre [15].

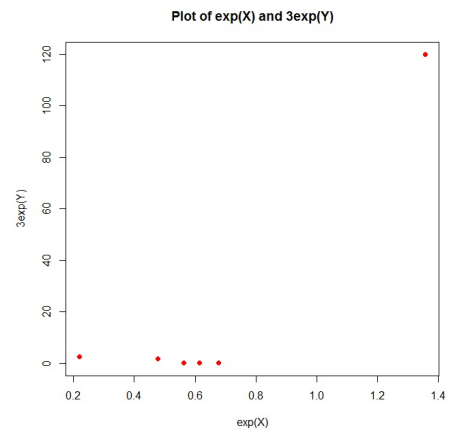
Προφανώς το διάγραμμα των (X, Y) και το διάγραμμα των $(\alpha(X), \beta(Y))$ είναι τελείως διαφορετικά.

Πίνακας 1.1: Δύο τυχαία δείγματα μεγέθους $n = 6$ από την $N(0, 1)$.

i	1	2	3	4	5	6
X_i	-0.4874	-0.7383	-0.5758	0.3054	-1.5118	0.3898
Y_i	-0.6265	0.1836	-0.8356	1.5953	0.3295	-0.8205



Σχήμα 1.14: Το διάγραμμα των X, Y από τη $N(0, 1)$.



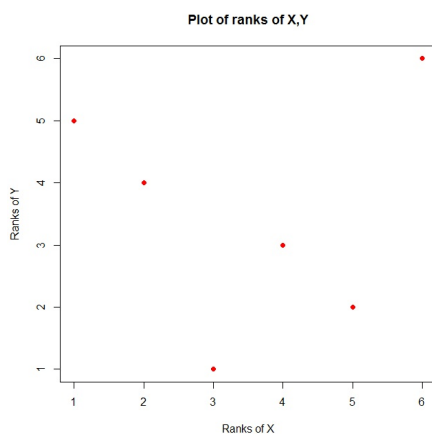
Σχήμα 1.15: Το διάγραμμα των $e^X, 3e^Y$.

Πίνακας 1.2: Τα ranks (οι τάξεις) των παρατηρήσεων.

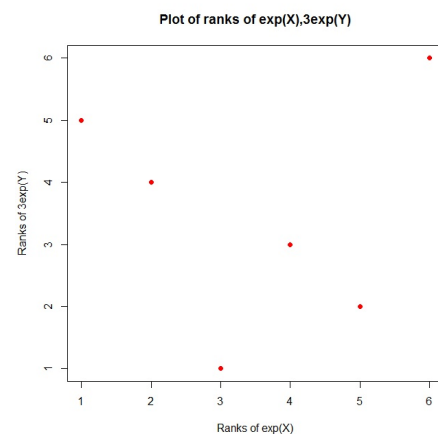
i	1	2	3	4	5	6
R_i	1	2	3	4	5	6
S_i	5	4	1	3	2	6

Διατάσσουμε τα αρχικά μας δεδομένα επιλέγοντας αύξουσα διάταξη της πρώτης τυχαίας μεταβλητής και βρίσκουμε τα αντίστοιχα ranks (τάξεις). Φτιάχνουμε ουσιαστικά δύο νέες τυχαίες μεταβλητές με τα ranks των αρχικών μας δεδομένων. Κάνουμε το ίδιο με τα μετασχηματισμένα δεδομένα. Όλα τα παραπάνω δίνονται σε κώδικα στα Παραρτήματα.

Το Copula είναι, όπως έχουμε αναφέρει, η συνάρτηση που, αν τις δώσεις τις περιθώριες, σου δίνει την από κοινού συνάρτηση κατανομής. Ουσιαστικά, η εξάρτηση των δύο τυχαίων μεταβλητών εμφανίζεται μέσω του Copula. Επομένως, μια γραφική αναπαράσταση της εξάρτησης των δύο τυχαίων μεταβλητών θα έπρεπε να έχει την ίδια ιδιότητα, δηλαδή να παραμένει αναλλοίωτη κάτω από γνησίως αύξοντα μετασχηματισμό. Αυτός είναι ο λόγος που βάζουμε τα ranks, τα οποία είναι ένα μέτρο που μας δίνει την εξάρτηση των δεδομένων μας. Τα διαγράμματα διασποράς των ranks δίνονται παρακάτω και όπως αναμέναμε είναι ίδια.



Σχήμα 1.16: Το διάγραμμα των ranks (τάξεων) των X, Y .



Σχήμα 1.17: Το διάγραμμα των ranks (τάξεων) των $e^X, 3e^Y$.

1.5 SURVIVAL COPULAS

Στην ενότητα αυτή θα ασχοληθούμε με τα Copula επιβίωσης, στα οποία θεωρούμε ότι οι τυχαίες μεταβλητές αναπαριστούν τους χρόνους ζωής μονάδων, ανθρώπων κ.α. σε κάποιο πληθυσμό. Ας υπενθυμίσουμε, αρχικά, κάποια στοιχεία που αφορούν την περίπτωση αυτή: Όταν μια τυχαία μεταβλητή εκφράζει το χρόνο ζωής μονάδων, τότε ορίζεται στο $[0, \infty]$. Η συνάρτηση επιβίωσης, $\bar{F}(x)$, (αναφέρεται και ως συνάρτηση αξιολοπιστίας), εκφράζει την πιθανότητα να επιβιώσει η μονάδα πέραν κάποιου χρονικού σημείου x . Επομένως,

$$\bar{F}(x) = P(X > x) = 1 - F(x),$$

όπου η F είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής X . Στη διοδιάστατη περίπτωση, όμως, τα πράγματα διαφέρουν λίγο.

Για τις ανάγκες των Copulas θα θεωρήσουμε αρχικά ότι οι χρόνοι ζωής μπορούν να πάρουν τιμές στο $\bar{\mathbf{R}}$. Έτσι για ένα ζεύγος τυχαίων μεταβλητών, (X, Y) , με από κοινού συνάρτηση κατανομής H , η από κοινού συνάρτηση επιβίωσης ορίζεται ως

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

και εκφράζει την πιθανότητα και οι δυο μονάδες να έχουν επιβιώσει πέραν κάποιων χρονικών σημείων, έστω x και y αντίστοιχα. Οι περιθώριες της \bar{H} είναι οι

$$\bar{F}(x) = \bar{H}(x, -\infty) \text{ και } \bar{G}(y) = \bar{H}(-\infty, y).$$

Και θέτουμε λοιπόν ένα ερώτημα: Αφού το Θεώρημα του Sklar συνδέει την από κοινού συνάρτηση κατανομής με τις περιθωρίες της μέσω του Copula, μήπως συμβαίνει κάτι αντίστοιχο και στην περίπτωση των συναρτήσεων επιβίωσης; Ας το υπολογίσουμε:

$$\begin{aligned} \bar{H}(x, y) &= 1 - F(x) - G(y) + H(x, y) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(F(x), G(y)) \\ &= \bar{F}(x) + \bar{G}(y) - 1 + C(1 - \bar{F}(x), 1 - \bar{G}(y)). \end{aligned}$$

Επομένως, αν ορίσουμε τη συνάρτηση $\hat{C} : I^2 \rightarrow I$ με τύπο:

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

θα έχουμε:

$$\bar{H}(x, y) = \hat{C}(\bar{F}(x), \bar{G}(y)).$$

Αποδεικνύεται ότι το \hat{C} είναι Copula, (βλπ. Nelsen [28]), και συγκεκριμένα είναι το Copula επιβίωσης. Παρατηρούμε από την παραπάνω σχέση ότι συνδέει την από κοινού συνάρτηση επιβίωσης με τις περιθώριες με τρόπο ανάλογο με το Θεώρημα του Sklar. Θα πρέπει όμως να τονίσουμε το εξής: το Copula επιβίωσης \hat{C} δεν είναι η από κοινού συνάρτηση επιβίωσης \bar{C} δύο τυχαίων μεταβλητών που κατανέμονται ομοιόμορφα στο $(0,1)$ και με από κοινού συνάρτηση κατανομής το Copula C. Κι αυτό γιατί:

$$\bar{C}(u, v) = P(U > u, V > v) = 1 - u - v + C(u, v) = \hat{C}(1 - u, 1 - v).$$

Τελειώνοντας την ενότητα αυτή, θα αναφέρουμε ότι υπάρχουν περιπτώσεις στις οποίες το Copula επιβίωσης έχει πιο απλό τύπο και άρα πιο εύχρηστο σε σχέση με την από κοινού συνάρτηση κατανομής και προτιμάται.

Παράδειγμα¹⁰:

Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής τη διοδιάστατη εκθετική κατανομή του Gumbel η οποία ορίζεται από τον τύπο:

$$H_{\theta}(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)}, & x \geq 0, y \geq 0 \\ 0, & \text{αλλιώς} \end{cases}$$

όπου θ είναι μια παράμετρος εξάρτησης στο $[0,1]$. Εύκολα αποδεικνύεται ότι οι X, Y έχουν μονοδιάστατες εκθετικές κατανομές:

$$F(x) = \lim_{y \rightarrow \infty} H_{\theta}(x, y) = 1 - e^{-x} \text{ και } G(y) = \lim_{x \rightarrow \infty} H_{\theta}(x, y) = 1 - e^{-y}.$$

Θα υπολογίσουμε τώρα τις γενικευμένες αντίστροφες των F, G , οι οποίες δεν είναι άλλες από τις γνωστές αντίστροφες γιατί είναι μονότονες συναρτήσεις. Έστω $F(x) = u, G(y) = v$. Τότε θα έχουμε:

$$1 - e^{-x} = u$$

αν και μόνο αν

¹⁰ Το συγκεκριμένο παράδειγμα συμπεριλαμβάνεται στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28].

$$e^{-x} = 1 - u \Rightarrow x = -\log(1 - u).$$

Όμοια:

$$y = -\log(1 - v).$$

Επομένως οι γενικευμένες αντίστροφες των F, G είναι οι

$$F^{-1}(u) = -\log(1 - u) \text{ και } G^{-1}(v) = -\log(1 - v)$$

και το Copula:

$$C_{\theta}(u, v) = H(F^{-1}(u), G^{-1}(v)) = u + v - 1 + (1 - u)(1 - v)e^{-\theta \log(1-u) \log(1-v)}.$$

Το Copula Επιβίωσης υπολογίζεται από τον τύπο:

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

και άρα έχουμε:

$$\hat{C}_{\theta}(u, v) = uve^{-\theta \log u \log v}.$$

Παρατηρούμε ότι το Copula Επιβίωσης έχει πιο απλό τύπο και επομένως θα είναι πιο εύχρηστο. Αν θέλαμε την από κοινού συνάρτηση επιβίωσης τότε από τη σχέση:

$$\bar{C}(u, v) = \hat{C}(1 - u, 1 - v)$$

θα είχαμε:

$$\bar{C}_{\theta} = (1 - u)(1 - v)e^{-\theta \log(1-u) \log(1-v)}.$$

1.6 ΠΟΛΥΔΙΑΣΤΑΤΑ COPULAS

Στην ενότητα αυτή θα παρουσιάσουμε την επέκταση των αποτελεσμάτων που δόθηκαν στις προηγούμενες παραγράφους, για την περίπτωση που έχουμε διάσταση $n \geq 3$.¹¹

ΟΡΙΣΜΟΣ 1.6.1 (H-όγκος): Έστω S_1, S_2, \dots, S_n μη κενά υποσύνολα του $\bar{\mathbf{R}}$ και H μια n -διάστατη πραγματική συνάρτηση με πεδίο ορισμού το $S_1 \times S_2 \times \dots \times S_n$ (δηλαδή το πεδίο ορισμού είναι υποσύνολο του επεκτεταμένου πραγματικού n -χώρου $\bar{\mathbf{R}}^n$). Έστω ακόμη $B = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times \dots \times [\alpha_n, \beta_n]$ με $\alpha \leq \beta$ ένα n -διάστατο παραλληλεπίπεδο (n -box) με τις κορυφές του στο $DomH$. Τότε, ο H -όγκος του B δίνεται από τον τύπο:

$$V_H(B) = \sum \text{sgn}(\mathbf{c})H(\mathbf{c})$$

όπου το άθροισμα υπολογίζεται για όλες τις κορυφές \mathbf{c} του B και το $\text{sgn}(\mathbf{c})$ υπολογίζεται ως εξής:

$$\text{sgn}(\mathbf{c}) = \begin{cases} 1, & \text{αν } c_k = \alpha_k \text{ για } k \text{ άρτιο} \\ -1, & \text{αν } c_k = \alpha_k \text{ για } k \text{ περιττό} \end{cases}$$

Εναλλακτικά, ο H -όγκος ενός n -διάστατου παραλληλεπιπέδου $B = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times \dots \times [\alpha_n, \beta_n]$ είναι η n -οστής τάξης διαφορά της H πάνω στο B :

$$V_H(B) = \Delta_{\alpha}^{\beta} H(\mathbf{t}) = \Delta_{\alpha_n}^{\beta_n} \Delta_{\alpha_{n-1}}^{\beta_{n-1}} \dots \Delta_{\alpha_1}^{\beta_1} H(\mathbf{t})$$

□

Ο τελεστής διαφορών Δ υπολογίζεται ως εξής:

$$\Delta_a^b F(x) = F(b) - F(a)$$

¹¹ Για τη γενίκευση στις n διαστάσεις παροτρύνουμε τον αναγνώστη στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28], στο τέταρτο κεφάλαιο του βιβλίου των Cherubini, Luciano και Vecchiato [6], στην εργασία των Durante και Sempì, *Copula Theory: An Introduction*, η οποία συγκαταλέγεται στο βιβλίο *Copula Theory and its Applications* των Jaworski, Durante, Härdle και Rychlik [23], στην εργασία των Embrechts, Lindskog και McNeil [10] με τίτλο *Modelling Dependence with Copulas and Applications to Risk Management*, καθώς και στην εργασία του Berthold Schweizer, *Thirty years of Copulas*, η οποία συμπεριλαμβάνεται στο βιβλίο των Dall'Aglio, Kotz και Salinetti [7].

Στις δύο διαστάσεις:

$$\Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} H(x, y) = \Delta_{a_2}^{b_2} (H(b_1, y) - H(a_1, y)) = H(b_1, b_2) - H(a_1, b_2) - H(b_1, a_2) + H(a_1, a_2)$$

Στις τρεις διαστάσεις:

$$\begin{aligned} \Delta_{a_3}^{b_3} \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} H(x, y) &= \Delta_{a_3}^{b_3} \Delta_{a_2}^{b_2} (H(b_1, y, z) - H(a_1, y, z)) = \\ \Delta_{a_3}^{b_3} (H(b_1, b_2, z) - H(a_1, b_2, z) - H(b_1, a_2, z) + H(a_1, a_2, z)) &= H(b_1, b_2, b_3) - H(a_1, b_2, b_3) - \\ H(b_1, a_2, b_3) + H(a_1, a_2, b_3) - H(b_1, b_2, a_3) + H(a_1, b_2, a_3) + H(b_1, a_2, a_3) - H(a_1, a_2, a_3) \end{aligned}$$

κ.ο.κ

Με λίγα λόγια, ο H -όγκος ισούται με την εξής πιθανότητα:

$$P(\alpha_1 \leq X_1 \leq \beta_1, \alpha_2 \leq X_2 \leq \beta_2, \dots, \alpha_n \leq X_n \leq \beta_n) = \int_{\alpha_n}^{\beta_n} \dots \int_{\alpha_2}^{\beta_2} \int_{\alpha_1}^{\beta_1} h(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

την οποία και προφανώς δεν μπορούμε να τη δούμε οπτικά.

ΟΡΙΣΜΟΣ 1.6.2 (n -αύξουσα συνάρτηση): Μια n -διάστατη πραγματική συνάρτηση H ονομάζεται n -αύξουσα αν $V_H(B) \geq 0$ για όλα τα n -διάστατα παραλληλεπίπεδα B των οποίων οι κορυφές ανήκουν στο $DomH$.

ΟΡΙΣΜΟΣ 1.6.3 (n -grounded συνάρτηση): Μια n -διάστατη πραγματική συνάρτηση H ονομάζεται grounded όταν $H(\mathbf{t}) = 0$ για όλα τα $\mathbf{t} \in DomH$ τέτοια, ώστε ένα τουλάχιστον από τα t_k ισούται με την ελάχιστη τιμή του.

Να σημειώσουμε σε αυτό το σημείο ότι για να βρούμε τις μονοδιάστατες περιθώριες βάζουμε σε όλες τις άλλες μεταβλητές τη μέγιστη τιμή τους, ενώ για τις k -διάστατες περιθώριες βάζουμε τη μέγιστη τιμή στις υπόλοιπες $n - k$. Προφανώς, ισχύουν τα παραπάνω αν τα σύνολα S_1, S_2, \dots, S_n είναι μη κενά και έχουν ελάχιστο και μέγιστο στοιχείο.

ΟΡΙΣΜΟΣ 1.6.4 (*n*-διάστατο Copula): Ένα *n*-διάστατο Copula, (*n*-Copula), είναι μια συνάρτηση $C : I^n \rightarrow I$ με τις εξής ιδιότητες:

1. Για κάθε $\mathbf{u} \in I^n$,

$$C(\mathbf{u}) = 0,$$

αν τουλάχιστον μία συνιστώσα του \mathbf{u} είναι μηδενική (δηλαδή η C είναι grounded συνάρτηση) και

2. Αν όλες οι συνιστώσες του \mathbf{u} είναι μονάδες εκτός της u_k , τότε

$$C(\mathbf{u}) = u_k,$$

δηλαδή ορίζονται οι μονοδιάστατες περιθώριες.

3. Για κάθε $a, b \in I^n$ με $a \leq b$ ισχύει

$$V_C([\mathbf{a}, \mathbf{b}]) \geq 0,$$

δηλαδή είναι *n*-αύξουσα. □

Αποδεικνύεται ότι για οποιαδήποτε *n*-Copula, $n \geq 3$, κάθε *k*-διάστατη περιθώρια της C είναι μια *k*-Copula, $2 \leq k \leq n$.

Μία ακόμα ιδιότητα του *n*-Copula είναι ότι είναι ομοιόμορφα συνεχής συνάρτηση καθώς για κάθε $\mathbf{u}, \mathbf{v} \in \text{Dom}C$ ισχύει:

$$|C'(\mathbf{v}) - C'(\mathbf{u})| \leq \sum_{k=1}^n |v_k - u_k|.$$

ΟΡΙΣΜΟΣ 1.6.5(*n*-διάστατη συνάρτηση κατανομής): Μία *n*-διάστατη συνάρτηση κατανομής είναι μια συνάρτηση H με πεδίο ορισμού το $\bar{\mathbf{R}}^n$ έτσι, ώστε:

1. Η H είναι *n*-αύξουσα

2. $H(\mathbf{t}) = 0$ για όλα τα $\mathbf{t} \in \bar{\mathbf{R}}^n$ ώστε $t_k = -\infty$ για τουλάχιστον ένα *k* και

$$H(\infty, \infty, \dots, \infty) = 1$$

ΘΕΩΡΗΜΑ (*Sklar*): Έστω H είναι μια *n*-διάστατη συνάρτηση κατανομής με περιθώριες F_1, F_2, \dots, F_n . Τότε, υπάρχει ένα *n*-Copula τέτοιο, ώστε για κάθε $\mathbf{x} \in \bar{\mathbf{R}}^n$, ισχύει:

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)).$$

Αν οι F_1, F_2, \dots, F_n είναι όλες συνεχείς, τότε το C είναι μοναδικό, αλλιώς το C είναι μοναδικώς ορισμένο στο $\mathcal{R}(F_1) \times \mathcal{R}(F_2) \times \dots \times \mathcal{R}(F_n)$.

Αντίστροφα, αν C είναι ένα n -Copula και F_1, F_2, \dots, F_n είναι συναρτήσεις κατανομής, τότε η H όπως ορίζεται παραπάνω είναι μια n -διάστατη συνάρτηση κατανομής με περιθώριες τις F_1, F_2, \dots, F_n . (Το Θεώρημα είναι αντίστοιχο και όταν έχουμε n τυχαίες μεταβλητές). \square

Επομένως αν έχουμε n περιθώριες και επιλέξουμε το σωστό Copula τότε δημιουργούμε την n -διάστατη από κοινού. Η διαισθητική απεικόνιση σε αυτήν την περίπτωση είναι αδύνατη, γι' αυτό στην παρούσα εργασία θα αναφερθούμε μόνο στη δισδιάστατη περίπτωση.

Πόρισμα: Έστω $H, F_1, F_2, \dots, F_n, C$ όπως ορίστηκαν παραπάνω και $F_1^{(-1)}, F_2^{(-1)}, \dots, F_n^{(-1)}$ οι γενικευμένες αντίστροφες συναρτήσεις των F_1, F_2, \dots, F_n , αντίστοιχα. Τότε, για κάθε $\mathbf{u} \in \mathbf{I}^n$,

$$C(u_1, u_2, \dots, u_n) = H(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2), \dots, F_n^{(-1)}(u_n)).$$

\square

Φράγματα Frèchet-Hoeffding για n -Copulas

Όπως είδαμε στις δύο διαστάσεις υπάρχουν φράγματα, τα οποία αποδεικνύεται ότι είναι και τα ίδια Copula. Η επέκτασή τους στις n -διαστάσεις είναι:

$$\text{Το κάτω όριο: } W^n(\mathbf{u}) = \max(u_1 + u_2 + \dots + u_n - n + 1, 0)$$

$$\text{Το άνω όριο: } M^n(\mathbf{u}) = \min(u_1, u_2, \dots, u_n)$$

Επίσης το Copula-γινόμενο δίνεται από τον τύπο:

$$\Pi^n(\mathbf{u}) = u_1 u_2 \cdots u_n$$

Να τονίσουμε σε αυτό το σημείο ότι στις n -διαστάσεις το κάτω όριο δεν είναι Copula για διάσταση $n \geq 3$.¹² Παρ' όλα αυτά χρησιμοποιείται η αριστερή ανισότητα, καθώς αποδεικνύεται ότι για διάσταση $n \geq 3$, υπάρχει Copula C τέτοιο, ώστε:

¹² Παράδειγμα μπορεί να βρει ο αναγνώστης στην εργασία των Embrechts, Lindskog και McNeil [10].

$$C(\mathbf{u}) = W^n(\mathbf{u}), \text{ για κάθε } \mathbf{u} \in \mathbf{I}^n. ^{13}$$

Επίσης, για το Copula Γινόμενο και το άνω φράγμα ισχύει το εξής:

Θεώρημα: Για $n \geq 2$ και για X_1, X_2, \dots, X_n συνεχείς τυχαίες μεταβλητές, ισχύει:

1. οι X_1, X_2, \dots, X_n είναι ανεξάρτητες αν και μόνο αν το Copula είναι αυτό του γινομένου και
2. κάθε μία ξεχωριστά από τις τυχαίες μεταβλητές X_1, X_2, \dots, X_n είναι μια γνησίως αύξουσα συνάρτηση οποιασδήποτε άλλης αν και μόνο αν το Copula είναι ίσο με το άνω φράγμα.

Όταν είμαστε στις δύο διαστάσεις, αν επιτυγχάνεται το άνω φράγμα, θα λέμε ότι οι τυχαίες μεταβλητές έχουν τέλεια θετική συσχέτιση ενώ όταν επιτυγχάνεται το κάτω φράγμα θα λέμε ότι έχουν τέλεια αρνητική συσχέτιση. Στις n -διαστάσεις, όπου $n \geq 3$, όπως αναφέραμε, δεν ορίζεται η τέλεια αρνητική συσχέτιση.

¹³ Η απόδειξη υπάρχει στο δεύτερο κεφάλαιο του βιβλίου του Nelsen [28].

Κεφάλαιο 2

ΣΥΣΧΕΤΙΣΗ-ΕΞΑΡΤΗΣΗ

2.1 ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό θα παρουσιάσουμε τις επιθυμητές ιδιότητες που θέλουμε να έχουν οι συντελεστές συσχέτισης και θα ορίσουμε την έννοια της “συμφωνίας” και της “ασυμφωνίας”. Η περίπτωση που θα εξετάσουμε είναι εκείνη των δύο διαστάσεων. Οι δείκτες συσχέτισης ή εξάρτησης που θα παρουσιάσουμε είναι: ο συντελεστής γραμμικής συσχέτισης του Pearson, οι συντελεστές συσχέτισης τάξης Rho του Spearman και Tau του Kendall και ο συντελεστής εξάρτησης ουρών. Τέλος θα κλείσουμε το κεφάλαιο αυτό παρουσιάζοντας τη μονοτονία στις ουρές και τη θετική τεταρτημοριακή εξάρτηση.

2.2 ΣΥΣΧΕΤΙΣΗ-ΕΞΑΡΤΗΣΗ-ΣΥΜΦΩΝΙΑ

Αρχικά θα υπενθυμίσουμε τις έννοιες της συσχέτισης και της εξάρτησης δύο τυχαίων μεταβλητών X, Y και πώς διαφοροποιούνται μεταξύ τους. Συγκεκριμένα, η συσχέτιση συνδέεται με τη συνδιακύμανση $Cov(X, Y)$. Οι X, Y ονομάζονται θετικά συσχετισμένες αν $Cov(X, Y) > 0$ και αρνητικά συσχετισμένες αν $Cov(X, Y) < 0$. Στην πρώτη περίπτωση η X και η Y θα συμπεριφέρονται όμοια γύρω από τη μέση τιμή, δηλαδή μεγάλες (μικρές) τιμές της X ως προς τον μέσο της συνεπάγονται μεγάλες (μικρές) τιμές της Y ως προς τον μέσο της, ενώ για τη δεύτερη περίπτωση έχουμε ότι μεγάλες (μικρές) τιμές της X ως προς τον μέσο της συνεπάγονται μικρές (μεγάλες) τιμές της Y ως προς τον μέσο της. Στην περίπτωση που $Cov(X, Y) = 0$ λέμε ότι οι X, Y είναι ασυσχέτιστες.

Όσον αφορά την εξάρτηση δύο τυχαίων μεταβλητών, ή καλύτερα την ανεξαρτησία τους, γνωρίζουμε ότι δύο τυχαίες μεταβλητές είναι ανεξάρτητες, αν η από κοινού συνάρτηση κατανομής τους γράφεται ως το γινόμενο των περιθωρίων. Αν δύο τυχαίες μεταβλητές είναι ανεξάρτητες τότε είναι και ασυσχέτιστες, ενώ το αντίστροφο εν γένει δεν ισχύει.

Παράδειγμα:

Έστω δύο τυχαίες μεταβλητές X, Y , όπου $X \sim N(0, 1)$ και $Y = X^2$. Τότε οι X, Y δεν είναι ανεξάρτητες είναι όμως ασυσχέτιστες αφού:

$$\text{Cov}(X, Y) = \text{Cov}(X, X^2) = E(X^3) - E(X)E(X^2) = E(X^3) - 0 \cdot E(X^2) = E(X^3) = 0.$$

Το τελευταίο ισχύει καθώς η ροπή τρίτης τάξης της X υπολογίζεται από το ολοκλήρωμα:

$$E(X^3) = \int_{-\infty}^{\infty} x^3 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

το οποίο είναι συγκλίνον ολοκλήρωμα περιττής συνάρτησης σε συμμετρικό διάστημα και άρα είναι ίσο με μηδέν. Ή πολύ πιο απλά, στην συγκεκριμένη περίπτωση έχουμε ουσιαστικά περιττή τάξης κεντρική ροπή του X , όπου η κατανομή του X είναι συμμετρική και άρα ίση με 0. ¹ Επομένως, οι X, Y είναι ασυσχέτιστες αλλά όχι ανεξάρτητες. \square

Για δύο τυχαίες μεταβλητές θα μπορούσαμε ενδεικτικά να πούμε ότι είναι σε συμφωνία, (*concordant*), όταν μεγάλες (μικρές) τιμές της μίας τείνουν να συσχετιστούν με μεγάλες (μικρές) τιμές της άλλης, ενώ λέμε ότι είναι σε διαφωνία, (*discordant*), όταν μεγάλες (μικρές) τιμές της μίας τείνουν να συσχετιστούν με μικρές (μεγάλες) τιμές της άλλης.

Ουσιαστικά, αυτό που μας ενδιαφέρει είναι, αν και οι δύο τυχαίες μεταβλητές παίρνουν ταυτόχρονα μεγάλες ή ταυτόχρονα μικρές τιμές. Η συμφωνία σαν έννοια θα γίνει πιο κατανοητή όταν παρουσιάσουμε τους συντελεστές συσχέτισης τάξης. Εδώ θα ορίσουμε αρχικά τη διάταξη συμφωνίας δύο συναρτήσεων κατανομής ή δύο Copulas.

¹ Υπάρχουν περιπτώσεις που οι ροπές k τάξης δεν ορίζονται, όταν, δηλαδή, το $\int_{-\infty}^{\infty} |x^k| f(x) dx$ δεν είναι πεπερασμένο. Εφόσον υπάρχει, ένας άλλος τρόπος υπολογισμού της ροπής k τάξης μίας τυχαίας μεταβλητής είναι η k -οστή παράγωγος της ροπογεννήτριας στο σημείο μηδέν.

ΟΡΙΣΜΟΣ(Διάταξη συμφωνίας): Έστω F, G δύο n -διάστατες συναρτήσεις κατανομής και \bar{F}, \bar{G} οι αντίστοιχες συναρτήσεις επιβίωσης. Τότε η G είναι περισσότερο concordant από την F , (συμβολικά θα γράφουμε $F < G$), αν ισχύουν τα εξής:

$$F(\mathbf{y}) \leq G(\mathbf{y}) \text{ και } \bar{F}(\mathbf{y}) \leq \bar{G}(\mathbf{y}) \text{ για κάθε } \mathbf{y} \in \mathbf{R}^n.$$

Η πρώτη ανισότητα αφορά στο κάτω ορθομόριο, ενώ η δεύτερη στο άνω. Αν ισχύουν και οι δύο, τότε μικρές και μεγάλες τιμές τείνουν να είναι περισσότερο συσχετισμένες μέσω της G παρά μέσω της F . Αυτό που παρατηρούμε, επίσης, είναι ότι συγκρίνουμε κατανομές με ίδιες περιθώριες.

Αν είμαστε στη διδιάστατη περίπτωση και έχουμε δύο διαφορετικά Copula C_1, C_2 , τότε οι ανισότητες

$$C_1(F(x), G(y)) \leq C_2(F(x), G(y)) \text{ και } \bar{C}_1(F(x), G(y)) \leq \bar{C}_2(F(x), G(y))$$

σημαίνουν πως οι τυχαίες μεταβλητές συσχετίζονται πιο ισχυρά μέσω του Copula C_2 .

2.3 ΣΥΝΤΕΛΕΣΤΕΣ ΣΥΣΧΕΤΙΣΗΣ-ΕΞΑΡΤΗΣΗΣ

Στην ενότητα αυτή θα παρουσιάσουμε κάποιες επιθυμητές ιδιότητες που θέλουμε να έχουν οι συντελεστές συσχέτισης και θα παρουσιάσουμε κάποιους από αυτούς. Επίσης θα δώσουμε και κάποια πλεονεκτήματα αλλά και μειονεκτήματά τους.

2.3.1 Επιθυμητές ιδιότητες των συντελεστών συσχέτισης

Όπως αναφέραμε προηγουμένως, δύο τυχαίες μεταβλητές είναι εξαρτημένες ή συσχετισμένες, αν δεν είναι ανεξάρτητες. Έστω X, Y δύο συνεχείς μεταβλητές με συναρτήσεις κατανομής F, G αντίστοιχα και από κοινού συνάρτηση κατανομής H . Τότε αυτές είναι εξαρτημένες αν

$$H(x, y) \neq F(x)G(y) \text{ για κάποιο } x, y \in \mathbf{R}$$

Για κάθε συντελεστή συσχέτισης, $\delta(X, Y)$, είναι επιθυμητές οι επόμενες ιδιότητες (βλπ. Embrechts, McNeil και Straumann [11] και Trivedi και Zimmer [30]):

1. Συμμετρία: $\delta(X, Y) = \delta(Y, X)$.

2. Κανονικότητα: $-1 \leq \delta(X, Y) \leq 1$.

3. Comonotonicity: $\delta(X, Y) = 1 \Leftrightarrow (X, Y)$ comonotonic, δηλαδή τέλεια θετική συσχέτιση των X, Y , όταν ο συντελεστής παίρνει τη μέγιστη τιμή του.

4. Countermonotonicity: $\delta(X, Y) = -1 \Leftrightarrow (X, Y)$ countermonotonic, δηλαδή τέλεια αρνητική συσχέτιση των X, Y , όταν ο συντελεστής παίρνει την ελάχιστη τιμή του.

5. Για ένα γνησίως μονότονο μετασχηματισμό $T : \mathbf{R} \rightarrow \mathbf{R}$ του X ισχύει:

$$\delta(T(X), Y) = \begin{cases} \delta(X, Y), & \text{αν η } T \text{ είναι αύξουσα} \\ -\delta(X, Y), & \text{αν η } T \text{ είναι φθίνουσα} \end{cases}$$

Ουσιαστικά, ένα από τα πράγματα που επιθυμούμε για ένα Copula είναι να καλύπτει όλο το χώρο ανάμεσα από τα φράγματα και, καθώς η παράμετρος εξάρτησης θ πλησιάζει το άνω ή το κάτω όριο του εύρους του, το Copula να πλησιάζει το άνω ή κάτω φράγμα των Fréchet-Hoeffding. Παρ' όλα αυτά, τα Copulas, από τη μορφή τους και μόνο, μπορεί να βάζουν κάποιους περιορισμούς. Αυτά που ισχύουν, όμως, είναι τα εξής :

1. Οι X, Y είναι ανεξάρτητες αν και μόνο αν

$$C(u, v) = uv, \text{ όπου } u = F(x), v = G(y) \text{ για κάθε } x, y \in \mathbf{R} \text{ και για κάθε } u, v \in I.$$

2. Η Y είναι μια αύξουσα συνάρτηση της X αν και μόνο αν

$$C(u, v) = M(u, v), \text{ όπου } u = F(x), v = G(y) \text{ για κάθε } x, y \in \mathbf{R}, \text{ για κάθε } u, v \in I$$

το οποίο αντιστοιχεί στην τέλεια θετική συσχέτιση, (comonotonicity).

3. Η Y είναι φθίνουσα συνάρτηση της X αν και μόνο αν

$$C(u, v) = W(u, v), \text{ όπου } u = F(x), v = G(y) \text{ για κάθε } x, y \in \mathbf{R} \text{ και για κάθε } u, v \in I$$

το οποίο αντιστοιχεί στην τέλεια αρνητική συσχέτιση, (countermonotonicity).

Ας επανέλθουμε, όμως, στους συντελεστές συσχέτισης ή εξάρτησης και ας δούμε κάποιους από τους πιο σημαντικούς.

2.3.2 Συντελεστής Γραμμικής Συσχέτισης του Pearson

Ο συντελεστής αυτός εισήχθη για να εξαληφθεί η επίδραση των μονάδων μέτρησης των τυχαίων μεταβλητών, κατά τον υπολογισμό της συνδιακύμανσης. Ουσιαστικά είναι η συνδιακύμανση των τυποποιημένων τυχαίων μεταβλητών. Αν X, Y είναι δύο τυχαίες μεταβλητές, τότε ο συντελεστής γραμμικής συσχέτισης του Pearson δίνεται από τον τύπο:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

όπου σ_X, σ_Y είναι οι τυπικές αποκλίσεις των X, Y , αντίστοιχα.

Όπως αναφέρει και το όνομά του, ο συντελεστής αυτός αφορά γραμμικές και μόνο συσχετίσεις των τυχαίων μεταβλητών. Από τις επιθυμητές ιδιότητες που αναφέρουμε παραπάνω πληροί τις πρώτες δύο μόνο, δηλαδή ισχύουν τα εξής:

1. $\rho_{XY} = \rho_{YX}$,
2. $-1 \leq \rho_{XY} \leq 1$.

Για τις επόμενες δύο ισχύουν:

3. $\rho_{XY} = 1 \Rightarrow$ τέλεια θετική συσχέτιση αλλά δεν ισχύει η αντίστροφη συνεπαγωγή,
4. $\rho_{XY} = -1 \Rightarrow$ τέλεια αρνητική συσχέτιση αλλά δεν ισχύει η αντίστροφη συνεπαγωγή.

Όσον αφορά την πέμπτη ιδιότητα, έχει την εξής διαφοροποίηση:

5. Ο συντελεστής ρ_{XY} παραμένει αναλλοίωτος μόνο σε γραμμικούς μετασχηματισμούς των X, Y .

Ο συντελεστής γραμμικής εξάρτησης του Pearson είναι πολύ χρήσιμος όταν έχουμε τη δισδιάστατη Κανονική Κατανομή, γιατί τότε το ρ_{XY} έχει όλη την πληροφορία για την από κοινού εξάρτηση των X, Y και επίσης $\rho_{XY} = 0 \Leftrightarrow X, Y$ είναι ανεξάρτητα. Στη γενική περίπτωση, όμως, η μηδενική συνδιακύμανση $Cov(X, Y) = 0$ και, επομένως, ο μηδενισμός του συντελεστή $\rho_{XY} = 0$ δεν συνεπάγεται ανεξαρτησία. Αυτό που ισχύει σε αυτήν την περίπτωση είναι ότι οι X, Y είναι ασυσχέτιστες. Για να ίσχυε η ανεξαρτησία, θα έπρεπε (βλπ. Trivedi και Zimmer[30])

$$\text{Cov}(\phi_1(X), \phi_2(Y)) = 0, \forall \phi_1, \phi_2.$$

Επομένως ο συντελεστής ρ_{XY} δεν μας δίνει πληροφορία για την εξάρτηση των τυχαίων μεταβλητών εκτός από ειδικές περιπτώσεις δισδιάστατων κατανομών, όπως π.χ. η κανονική. Μια ακόμη αδυναμία του ρ_{XY} είναι ότι δεν ορίζεται για κάποιες κατανομές με βαριές ουρές (heavy-tailed distributions) για τις οποίες δεν υπάρχουν ροπές δευτέρας τάξης. Επίσης, σε μη γραμμικούς μετασχηματισμούς δεν παραμένει αναλλοίωτος, αλλά διαφοροποιείται και αυτό τον καθιστά όχι και τόσο εύχρηστο. Τέλος, οι εφικτές τιμές του ρ_{XY} στο $[-1, 1]$ εξαρτώνται από τις περιθώριες F, G κι επομένως απαιτείται η γνώση τους.

Συνοψίζοντας, τα πλεονεκτήματα του συντελεστή συσχέτισης είναι τα ακόλουθα:

1. Για ελλειπτικές κατανομές², η δομή εξάρτησης καθορίζεται πλήρως από τον πίνακα συσχέτισης.
2. Για την πολυμεταβλητή κανονική κατανομή (περίπτωση των ελλειπτικών κατανομών), όχι μόνο η δομή εξάρτησης καθορίζεται πλήρως από τον πίνακα συσχέτισης, αλλά και η μηδενική συσχέτιση συνεπάγεται ανεξαρτησία και αντίστροφα.

Τα μειονεκτήματα του ρ_{XY} , (βλπ. *Correlation and Dependence in Risk Management: properties and pitfalls* [11], καθώς και *Correlation: Pitfalls and Alternatives* [12], των Embrechts, McNeil και Straumann), είναι τα εξής:

1. Εκτός από ειδικές περιπτώσεις, όπως για παράδειγμα οι ελλειπτικές κατανομές, ο συντελεστής γραμμικής συσχέτισης δεν έχει εν γένει όλες τις πληροφορίες για τη δομή εξάρτησης των τυχαίων μεταβλητών.
2. Οι τιμές του εξαρτώνται από τις περιθώριες κατανομές.
3. Υπάρχουν περιπτώσεις στις οποίες δεν είναι εφικτές όλες οι τιμές από το -1 έως το 1 και έτσι δεν μπορούμε να πάρουμε τέλεια συσχέτιση στα άκρα του διαστήματος $[-1, 1]$.³
4. Η μηδενική συσχέτιση δεν συνεπάγεται πάντα ανεξαρτησία. Αυτό ισχύει για όλους τους συντελεστές συσχέτισης που πληρούν την επθυμητή ιδιότητα 5.

² Παραπέμπουμε τον αναγνώστη στη σελίδα 61 της παρούσας εργασίας για τον ορισμό των ελλειπτικών κατανομών.

³ Ένα ενδιαφέρον παράδειγμα υπάρχει στην εργασία του Embrechts [9], με τίτλο *Copulas: A personal view*.

5. Σε μη γραμμικούς μετασχηματισμούς ο ρ_{XY} δεν παραμένει αναλλοίωτος
6. Απαιτεί πεπερασμένες διακυμάνσεις, το οποίο δε συμβαίνει πάντα. Για παράδειγμα, σε κατανομές με βαριές ουρές όπως η δισδιάστατη κατανομή t-student με βαθμούς ελευθερίας $n \leq 2$.

Έτσι έχουν προταθεί και άλλοι συντελεστές τους οποίους και δίνουμε στη συνέχεια.

2.3.3 Συντελεστές συσχέτισης τάξης

Στην κατηγορία της συσχέτισης τάξης (rank correlation) ανήκουν οι συντελεστές Rho του Spearman και Tau του Kendall, (βλπ. πέμπτο κεφάλαιο του βιβλίου του Nelsen [28] και τέταρτο κεφάλαιο του βιβλίου των Balakrishnan και Lai [4]), οι οποίοι έχουν το πλεονέκτημα ότι δεν εξαρτώνται από τις περιθώριες κατανομές, όπως ο ρ_{XY} . Και οι δύο συντελεστές βασίζονται στην έννοια της συμφωνίας ή ασυμφωνίας των τυχαίων μεταβλητών. Πριν τους αναπτύξουμε, όμως, θα δώσουμε έναν ορισμό που αφορά τα μέτρα συμφωνίας, όπως αποδεικνύεται ότι είναι και οι δύο που θα παρουσιάσουμε.(βλπ. Nelsen [28]).

ΟΡΙΣΜΟΣ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές και κ ένα μέτρο συσχέτισής τους. Θα λέμε ότι το κ είναι μέτρο συμφωνίας των X, Y και θα το συμβολίζουμε με $\kappa_{X,Y} = \kappa_C$, όπου C είναι το Copula των X, Y , αν ικανοποιεί τα εξής:

1. Το κ ορίζεται για κάθε ζεύγος X, Y συνεχών τυχαίων μεταβλητών.
2. $-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,X} = 1$ και $\kappa_{X,-X} = -1$.
3. $\kappa_{X,Y} = \kappa_{Y,X}$.
4. Αν οι X, Y είναι ανεξάρτητες τότε $\kappa_{X,Y} = \kappa_{\Pi} = 0$.
5. $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$.
6. Αν C_1, C_2 είναι Copulas για τα οποία ισχύει ότι $C_1 < C_2$ τότε $\kappa_{C_1} \leq \kappa_{C_2}$.
7. Αν $\{(X_n, Y_n)\}$ είναι μια ακολουθία συνεχών τυχαίων μεταβλητών με Copulas C_n και η C_n συγκλίνει στη C , τότε $\lim_{n \rightarrow \infty} \kappa_{C_n} = \kappa_C$.

Παρακάτω παρουσιάζουμε τα δύο μέτρα συμφωνίας Tau του Kendall και Rho του Spearman.

2.3.3.1 Ταυ του Kendall (τ)

ΟΡΙΣΜΟΣ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές και $(x_1, y_1), (x_2, y_2)$ δύο ζεύγη παρατηρούμενων τιμών τους. Τότε έχουμε συμφωνία (*concordance*) των $(x_1, y_1), (x_2, y_2)$, αν:

$$x_1 < x_2 \text{ και } y_1 < y_2 \text{ ή } x_1 > x_2 \text{ και } y_1 > y_2,$$

ενώ έχουμε διαφωνία (*discordance*) των $(x_1, y_1), (x_2, y_2)$ αν:

$$x_1 < x_2 \text{ και } y_1 > y_2 \text{ ή } x_1 > x_2 \text{ και } y_1 < y_2.$$

Επομένως, τα $(x_1, y_1), (x_2, y_2)$ συμφωνούν αν:

$$(x_1 - x_2)(y_1 - y_2) > 0$$

και διαφωνούν αν:

$$(x_1 - x_2)(y_1 - y_2) < 0.$$

□

Ο συντελεστής του Kendall⁴ εισήχθη πρώτη φορά από τον Fechner περί το 1900 και ανακαλύφθηκε πάλι από τον Kendall το 1938. Για να καταλάβουμε, όμως, τι υπολογίζει ο συντελεστής αυτός, ας θεωρήσουμε ένα τυχαίο δείγμα n παρατηρήσεων από μια διδιάστατη κατανομή, έστω $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Τότε εξετάζουμε ανά δύο τις παρατηρήσεις για να δούμε αν υπάρχει συμφωνία ή όχι. Άρα έχουμε $\binom{n}{2}$ διαφορετικά ζεύγη τα οποία εξετάζουμε αν είναι σύμφωνα ή ασύμφωνα. Έστω c, d το πλήθος των ζευγών που συμφωνούν και που διαφωνούν αντίστοιχα. Τότε ο συντελεστής τ του Kendall για το δείγμα δίνεται από το λόγο:

$$t = \frac{c - d}{c + d} = \frac{c - d}{\binom{n}{2}}$$

⁴ Παροτρύνουμε τον αναγνώστη να κοιτάξει και την εργασία του Nelsen, *Copulas and Association*, η οποία συμπεριλαμβάνεται στο βιβλίο των Dall'Aglio, Kotz και Salinetti [7].

Με απλά λόγια, ο συντελεστής είναι ίσος με το πλήθος των ζευγών που συμφωνούν μείον το πλήθος των ζευγών που διαφωνούν προς το πλήθος των διαφορετικών ζευγών.

Ας ορίσουμε, τώρα, το συντελεστή αυτόν όταν έχουμε κάποιο πληθυσμό:

ΟΡΙΣΜΟΣ (Kendall's Tau): Έστω (X_1, Y_1) και (X_2, Y_2) δύο ανεξάρτητα και ισόνομα τυχαία διανύσματα που το καθένα έχει από κοινού συνάρτηση κατανομής H και τα X_1, X_2 έχουν συνάρτηση κατανομής F , ενώ τα Y_1, Y_2 έχουν συνάρτηση κατανομής G . Τότε το τ του Kendall είναι η πιθανότητα της συμφωνίας μείον την πιθανότητα της ασυμφωνίας:

$$\tau = \tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

□

Θα δώσουμε και άλλη μία σχέση που υπολογίζει το τ όταν έχουμε το Copula:

Θεώρημα: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με Copula C . Τότε το τ του Kendall των X, Y δίνεται από τον τύπο:

$$\tau_C = 4 \int \int_{\mathcal{I}^2} C(u, v) dC(u, v) - 1$$

Παρατηρούμε ότι το παραπάνω διπλό ολοκλήρωμα δεν είναι άλλο από την αναμενόμενη τιμή του $C(U, V)$, όπου U, V είναι τυχαίες μεταβλητές ομοιόμορφα κατανομημένες στο $(0, 1)$ και έχουν από κοινού συνάρτηση κατανομής τη C . Για να γίνει, όμως, αυτό κατανοητό θα πρέπει να ορίσουμε το Copula ως πυκνότητα και όχι ως συνάρτηση κατανομής. Έτσι έχουμε:

ΟΡΙΣΜΟΣ: Έστω X_1, X_2, \dots, X_n συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F_1, F_2, \dots, F_n , αντίστοιχα, και Copula C . Αν οι F_1, F_2, \dots, F_n, C είναι n φορές παραγωγίσιμες και f_1, f_2, \dots, f_n είναι οι αντίστοιχες πυκνότητες, τότε η από κοινού συνάρτηση πυκνότητας των X_1, X_2, \dots, X_n είναι:

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) \times f_2(x_2) \times \dots \times f_n(x_n) c[F_1(x_1), \dots, F_n(x_n)]$$

όπου το c ονομάζεται Copula πυκνότητας και είναι ίσο με

$$c(F_1(x_1), \dots, F_n(x_n)) = \frac{\partial^n C(F_1(x_1), \dots, F_n(x_n))}{\partial F_1 \cdots \partial F_n}.$$

Επομένως

$$\int \int_{I^2} C(u, v) dC(u, v) = E(C(U, V)) \text{ και } \tau_C = 4E(C(U, V)) - 1$$

καθώς

$$E(C(U, V)) = \int \int_{I^2} C(u, v) \frac{\partial^2 C(u, v)}{\partial u \partial v} dudv = \int \int_{I^2} C(u, v) dC(u, v).$$

Θα πρέπει να τονιστεί ότι τα παραπάνω ισχύουν μόνο όταν το C είναι απόλυτα συνεχής συνάρτηση.

Παράδειγμα⁵: Έστω το Copula

$$C(u, v) = uv + \theta uv(1 - u)(1 - v)$$

της οικογένειας των Farlie-Gumbel-Morgenstern, όπου $\theta \in [-1, 1]$. Επειδή το C είναι απόλυτα συνεχής συνάρτηση μπορούμε να υπολογίσουμε το τ του Kendall από τη σχέση:

$$\tau_{C_\theta} = 4 \int \int_{I^2} C_\theta(u, v) dC_\theta(u, v) - 1$$

όπου

$$dC_\theta(u, v) = \frac{\partial^2 C_\theta(u, v)}{\partial u \partial v} dudv = (1 + \theta(1 - 2u)(1 - 2v))dudv.$$

Άρα

$$\tau_{C_\theta} = 4 \int \int_{I^2} [uv(1 - \theta(1 - u)(1 - v))][(1 + \theta(1 - 2u)(1 - 2v))]dudv - 1$$

$$\tau_{C_\theta} = 4\left(\frac{1}{4} + \frac{\theta}{18}\right) - 1 = \frac{2\theta}{9}.$$

□

Το παραπάνω Copula δεν είναι πολύ εύχρηστο καθώς για $\theta \in [-1, 1]$ παίρνει τιμές στο $[-\frac{2}{9}, \frac{2}{9}]$ το οποίο είναι ένα αρκετά μικρό εύρος τιμών για συντελεστή συσχέτισης.

Όλα τα παραπάνω απαιτούν το Copula να είναι απόλυτα συνεχές. Παρ' όλα αυτά υπάρχει μία ακόμα σχέση που χρησιμοποιείται όταν υπάρχουν και ιδιάζουσες συνιστώσες και είναι η εξής:

⁵ Το συγκεκριμένο παράδειγμα συμπεριλαμβάνεται στο πέμπτο κεφάλαιο του βιβλίου του Nelsen [28].

$$\tau_C = 1 - 4 \int \int_{I^2} \frac{\partial C(u, v)}{\partial u} \frac{\partial C(u, v)}{\partial v} dudv.$$

Οι δύο παραπάνω σχέσεις για το τ του Kendall είναι ισοδύναμες στην περίπτωση που η C είναι απόλυτα συνεχής και αυτό προκύπτει από το παρακάτω λήμμα:

ΛΗΜΜΑ: Έστω C_1, C_2 δύο Copulas. Τότε:

$$\int \int_{I^2} C_1(u, v) dC_2(u, v) + \int \int_{I^2} \frac{\partial C_1(u, v)}{\partial u} \frac{\partial C_2(u, v)}{\partial v} dudv = \frac{1}{2}.$$

Απόδειξη: Για την ισοδυναμία θα θεωρήσω τις C_1, C_2 απόλυτα συνεχείς. Έτσι έχουμε:

$$\int \int_{I^2} C_1(u, v) dC_2(u, v) = \int_0^1 \left[\int_0^1 C_1(u, v) \frac{\partial^2 C_2(u, v)}{\partial u \partial v} du \right] dv.$$

Το εσωτερικό ολοκλήρωμα εφαρμόζοντας ολοκλήρωση κατά παράγοντες είναι ίσο με:

$$\begin{aligned} & C_1(u, v) \frac{\partial C_2(u, v)}{\partial v} \Big|_{u=0}^{u=1} - \int_0^1 \frac{\partial C_1(u, v)}{\partial u} \frac{\partial C_2(u, v)}{\partial v} du \\ &= v1 - 0 - \int_0^1 \frac{\partial C_1(u, v)}{\partial u} \frac{\partial C_2(u, v)}{\partial v} du = v - \int_0^1 \frac{\partial C_1(u, v)}{\partial u} \frac{\partial C_2(u, v)}{\partial v} du. \end{aligned}$$

Αντικαθιστώντας την τελευταία σχέση στο εξωτερικό ολοκλήρωμα παίρνουμε:

$$\int_0^1 [v - \int_0^1 \frac{\partial C_1(u, v)}{\partial u} \frac{\partial C_2(u, v)}{\partial v} du] dv = \frac{1}{2} - \int_0^1 \int_0^1 \frac{\partial C_1(u, v)}{\partial u} \frac{\partial C_2(u, v)}{\partial v} dudv.$$

Αφού είδαμε τον τύπο του τ του Kendall, ας δούμε τώρα και ποιες από τις επιθυμητές ιδιότητες έχει. Αυτές είναι:

1. Το $\tau_{X,Y}$ είναι συμμετρικό, δηλαδή $\tau_{X,Y} = \tau_{Y,X}$,
2. $-1 \leq \tau_{X,Y} \leq 1$,
3. Για $\tau_{X,Y} = 1$ έχουμε τέλεια θετική συσχέτιση και αντίστροφα,
4. Για $\tau_{X,Y} = -1$ έχουμε τέλεια αρνητική συσχέτιση και αντίστροφα,
5. το $\tau_{X,Y}$ είναι αναλλοίωτο σε γνησίως μονότονους μετασχηματισμούς των X, Y .

Έχει, επομένως, όλες τις επιθυμητές ιδιότητες και επιπλέον δεν εξαρτάται από τις περιθώριες, παρά μόνο από το Copula. Τα πλεονεκτήματα και τα μειονεκτήματά του θα τα δώσουμε στο τέλος, αφού θα παρουσιάσουμε και έναν δεύτερο συντελεστή συσχέτισης τάξης, το Rho του Spearman (ρ_S).

2.3.3.2 Rho του Spearman (ρ_S)

Ο συντελεστής συσχέτισης ρ_S του Spearman⁶ προτάθηκε το 1904 από τον ψυχολόγο Charles Spearman και, όπως ο τ του Kendall, έτσι και αυτός βασίζεται στην έννοια της συμφωνίας. Συγκεκριμένα, ο συντελεστής συσχέτισης του Spearman, είναι συντελεστής γραμμικής συσχέτισης (πιθανοτικών)-μετασχηματισμών των τυχαίων μεταβλητών, δηλαδή μπορεί να γραφεί στη μορφή:

$$\rho_S(X, Y) = \rho(F(X), G(Y))$$

Αυτό θα γίνει κατανοητό, αφού ορίσουμε το ρ_S και το γράψουμε σε μορφή παρόμοια του συντελεστή γραμμικής συσχέτισης του Pearson. Ας δώσουμε αρχικά τον ορισμό του όταν έχουμε κάποιο πληθυσμό.

ΟΡΙΣΜΟΣ (Spearman's Rho): Έστω $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3)$ τρία ανεξάρτητα τυχαία διανύσματα που το καθένα έχει από κοινού συνάρτηση κατανομής H και τα X_1, X_2, X_3 έχουν συνάρτηση κατανομής F , ενώ τα Y_1, Y_2, Y_3 έχουν συνάρτηση κατανομής G . Τότε το ρ_S του Spearman ορίζεται να είναι ανάλογο της πιθανότητας της συμφωνίας μείον την πιθανότητα της διαφωνίας των διανυσμάτων (X_1, Y_1) και (X_2, Y_3) :

$$\rho_S = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0])$$

Ο συντελεστής ρ_S χρησιμοποιείται για να παίρνει ο συντελεστής τιμές μεταξύ -1 και 1. Και έρχεται εδώ το ερώτημα: Μόνο για αυτά τα δύο διανύσματα ισχύει ο παραπάνω τύπος; Προφανώς και όχι. Στην ουσία το ρ_S , ισχύει για τα ζεύγη διανυσμάτων, τα οποία έχουν κοινές περιθώριες και το ένα διάνυσμα έχει από κοινού συνάρτηση κατανομής την H , ενώ το άλλο έχει το $F(x)G(y)$. Με απλά λόγια, το δεύτερο διάνυσμα θα έχει ανεξάρτητες συνιστώσες.

Όταν έχουμε δεδομένα ο συντελεστής συσχέτισης του Spearman μοιάζει πολύ με του

⁶ Παροτρύνουμε τον αναγνώστη να κοιτάξει και την εργασία του Nelsen, *Copulas and Association*, η οποία συμπεριλαμβάνεται στο βιβλίο των Dall'Aglio, Kotz και Salinetti [7].

Pearson με τη διαφορά ότι χρησιμοποιούμε τα ranks και όχι τις παρατηρήσεις. Ας υποθέσουμε ότι έχουμε δύο συνεχείς τυχαίες μεταβλητές X, Y και ότι έχουμε βάλει τα αντίστοιχα ranks R_i, S_i στην i παρατηρούμενη τιμή. Τότε, σε αυτήν την περίπτωση, ο συντελεστής συσχέτισης του Spearman όταν έχουμε κάποιο δείγμα είναι:

$$\rho_n = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}}$$

όπου

$$\bar{R} = \frac{\sum_{i=1}^n R_i}{n}$$

και

$$\bar{S} = \frac{\sum_{i=1}^n S_i}{n}.$$

Ένας τύπος που χρησιμοποιείται στην πράξη είναι:

$$\rho_n = 1 - 6 \frac{\sum_{i=1}^n (R_i - S_i)^2}{(n^3 - n)}.$$

Ας επιστρέψουμε στην περίπτωση που έχουμε κάποιο πληθυσμό και ας δούμε στη συνέχεια πώς συνδέεται ο συντελεστής ρ_S με το Copula.

ΘΕΩΡΗΜΑ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με Copula C . Τότε το ρ_S του Spearman δίνεται από τις σχέσεις:

$$\rho_S = 12 \int \int_{I^2} uv dC(u, v) - 3 \quad (a)$$

$$= 12 \int \int_{I^2} C(u, v) dudv - 3 \quad (b)$$

Ας δούμε, τώρα, γιατί το ρ_S είναι συντελεστής γραμμικής συσχέτισης τάξης. Από τον τύπο (a) στο παραπάνω θεώρημα έχουμε:

$$\begin{aligned} \rho_S &= 12 \int \int_{I^2} uv dC(u, v) - 3 = 12 \int \int_{I^2} uv \frac{\partial^2 C(u, v)}{\partial u \partial v} dudv - 3 \\ &= 12E(UV) - 3 = \frac{E(UV) - \frac{1}{4}}{\frac{1}{12}} = \frac{E(UV) - E(U)E(V)}{\sqrt{\text{Var}(U)} \sqrt{\text{Var}(V)}}. \end{aligned}$$

Στον παραπάνω τύπο έγινε η χρήση του γεγονότος ότι οι $U = F(X)$, $V = G(Y)$ είναι ομοιόμορφες στο $(0, 1)$ και, επομένως, έχουν μέσες τιμές ίσες με $\frac{1}{2}$ και διακυμάνσεις ίσες με $\frac{1}{12}$.

Από τον τύπο (b) του θεωρήματος μπορούμε εύκολα να συμπεράνουμε ότι το ολοκλήρωμα που χρησιμοποιεί δεν είναι άλλο από τον όγκο που σχηματίζεται μεταξύ του Copula και του επιπέδου I^2 . Αν ξαναγράψουμε τον τύπο ως εξής:

$$\begin{aligned} \rho_S &= 12 \int \int_{I^2} C(u, v) dudv - 3 = 12 \left(\int \int_{I^2} C(u, v) dudv - \frac{1}{4} \right) = \\ &12 \left(\int \int_{I^2} C(u, v) dudv - \int \int_{I^2} uv dudv \right) = 12 \int \int_{I^2} (C(u, v) - uv) dudv \end{aligned}$$

έχουμε ότι ο συγκεκριμένος συντελεστής είναι και ανάλογος του όγκου μεταξύ της διαφοράς του Copula των τυχαίων μεταβλητών και του Copula-γινόμενο.⁷ Δηλαδή, είναι ένα μέτρο της μέσης απόστασης της κατανομής των X, Y και της ανεξαρτησίας.⁸

Παράδειγμα⁹:

Έστω το Copula

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v)$$

της οικογένειας των Farlie-Gumbel-Morgenstern, όπου $\theta \in [-1, 1]$. Τότε ο συντελεστής συσχέτισης του Spearman υπολογίζεται ως εξής:

$$\begin{aligned} \rho_S &= 12 \int \int_{I^2} C_\theta(u, v) dudv - 3 = 12 \left(\int_0^1 \int_0^1 uv + \theta uv(1 - u)(1 - v) dudv \right) - 3 = \\ &12 \left(\int_0^1 \left[\left(\frac{1}{2} + \frac{\theta}{2} - \frac{\theta}{3} \right) v - \left(\frac{\theta}{2} - \frac{\theta}{3} \right) v^2 \right] dv \right) - 3 = 12 \left(\frac{1}{4} + \frac{\theta}{36} \right) - 3 = \frac{\theta}{3}. \end{aligned}$$

Παρατηρούμε ότι και εδώ το εύρος τιμών είναι μικρό για το συγκεκριμένο συντελεστή συσχέτισης και διαφέρει ελάχιστα από το αντίστοιχο του Kendall.

⁷ Ο αναγνώστης μπορεί να απευθυνθεί και στην εργασία των Frees και Valdez [14] καθώς και στο δεύτερο κεφάλαιο των Trivedi και Zimmer [30].

⁸ Οι Schweizer και Wolff στην εργασία τους με τίτλο *On Nonparametric Measures of Dependence for random variables* [29] ορίζουν νέα μέτρα εξάρτησης χρησιμοποιώντας αποστάσεις του Copula από το Copula-γινόμενο. Οι ιδιότητες που πρέπει να πληρούν οι συγκεκριμένοι συντελεστές συμπεριλαμβάνονται στην εργασία τους.

⁹ Το συγκεκριμένο παράδειγμα συμπεριλαμβάνεται στο πέμπτο κεφάλαιο του βιβλίου του Nelsen [28].

Οι ιδιότητες που έχει το ρ_S του Spearman είναι όμοιες με του τ του Kendall. Πιο συγκεκριμένα ισχύουν τα εξής:

1. Το ρ_S είναι συμμετρικό,
2. $-1 \leq \rho_S \leq 1$,
3. Για $\rho_S = 1$ έχουμε τέλεια θετική συσχέτιση και αντίστροφα,
4. Για $\rho_S = -1$ έχουμε τέλεια αρνητική συσχέτιση και αντίστροφα,
5. το ρ_S είναι αναλλοίωτο σε γνησίως μονότονους μετασχηματισμούς των X, Y .

Έχει, επομένως, όλες τις επιθυμητές ιδιότητες και επιπλέον δεν εξαρτάται από τις περιθώριες, παρά μόνο από το Copula. Συνοψίζοντας, επομένως, οι συντελεστές συσχέτισης τάξης, τ του Kendall και ρ_S του Spearman, έχουν τα εξής πλεονεκτήματα σε σύγκριση με το συντελεστή συσχέτισης του Pearson:

1. Θεωρούνται μη παραμετρικοί συντελεστές καθώς δεν εξαρτώνται από τις περιθώριες συναρτήσεις κατανομής, παρά μόνο από το Copula C .
2. Παραμένουν αναλλοίωτοι σε γνησίως μονότονους μετασχηματισμούς των τυχαίων μεταβλητών και όχι μόνο σε γραμμικούς μετασχηματισμούς.
3. Δεν απαιτούν πεπερασμένη διακύμανση για τον υπολογισμό τους.
4. Επιτυγχάνουν τέλεια θετική συσχέτιση όταν παίρνουν την τιμή 1 και τέλεια αρνητική συσχέτιση όταν παίρνουν την τιμή -1 και αντίστροφα.

Παρά όλα αυτά έχουν και μειονεκτήματα, όπως:

1. Είναι συντελεστές συσχέτισης και όχι εξάρτησης, επομένως δεν έχουν όλες τις πληροφορίες για τη δομή εξάρτησης.
2. Ο μηδενισμός τους δεν συνεπάγεται ανεξαρτησία.
3. Έχουν πολύπλοκους τύπους, οι οποίοι σε κάποιες περιπτώσεις είναι δύσκολα υπολογίσιμοι και δε βασίζονται στις ροπές και τις ιδιότητές τους.

Οι συντελεστές συσχέτισης τάξης, αν και δεν είναι ακόμα γνωστό κατά πόσον κάποιος από τους δυο είναι καλύτερος, συνδέονται μεταξύ τους με κάποιες ανισοτικές σχέσεις. Και οι δύο συντελεστές μετρούν την πιθανότητα συμφωνίας μεταξύ δύο τυχαίων μεταβλητών,

αλλά μπορεί να έχουν πολύ διαφορετικές τιμές.¹⁰

ΘΕΩΡΗΜΑ¹¹: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές και τ_C, ρ_S οι συντελεστές συσχέτισης των Kendall και Spearman, αντίστοιχα. Τότε ισχύουν τα εξής:

$$-1 \leq 3\tau_C - 2\rho_S \leq 1$$

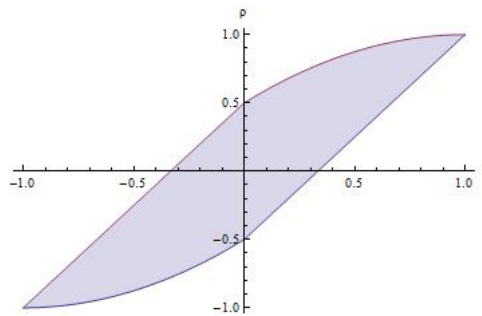
$$\frac{3\tau_C - 1}{2} \leq \rho_S \leq \frac{1 + 2\tau_C - \tau_C^2}{2}, \quad \text{για } \tau_C \geq 0$$

και

$$\frac{\tau_C^2 + 2\tau_C - 1}{2} \leq \rho_S \leq \frac{1 + 3\tau_C}{2}, \quad \text{για } \tau_C \leq 0$$

Οι δύο τελευταίες ανισότητες, οι οποίες συμπεριλαμβάνουν την πρώτη, προτάθηκαν το 1951 από τους Durbin και Stuart και δημιουργούν την $(\tau - \rho)$ -περιοχή, η οποία δίνεται στο παρακάτω γράφημα:

Σχήμα 2.1: $(\tau - \rho)$ χωρίο



Με απλά λόγια, για κάθε ζεύγος συνεχών τυχαίων μεταβλητών (X, Y) , οι τιμές των συντελεστών των Kendall και Spearman πρέπει να βρίσκονται μέσα στο παραπάνω σκιασμένο

¹⁰ Γενικεύσεις των συντελεστών συσχέτισης τάξης ρ_S του Spearman και τ του Kendall στις n διαστάσεις μπορεί να βρει ο αναγνώστης στην εργασία των Schmid, Schmidt, Blumentritt, Gaißer και Ruppert με τίτλο *Copula-Based Measures of Multivariate Association*, η οποία συγκαταλέγεται στο βιβλίο των Jaworski, Durante, Härdle και Rychlik [23].

¹¹ Παροτρύνουμε τον αναγνώστη να απευθυνθεί στην εργασία του Nelsen με τίτλο *Copulas and Association*, η οποία συμπεριλαμβάνεται στο βιβλίο των Dall'Aglio, Kotz και Salinetti [7].

γράφημα. Να σημειώσουμε σε αυτό το σημείο ότι, ενώ οι γραμμικές σχέσεις στις ανισώσεις δεν βελτιώνονται περαιτέρω, οι δευτεροβάθμιες σχέσεις, (δηλαδή τα παραβολικά κομμάτια του σχήματος), αποδεικνύεται πως επιδέχονται βελτιώσεις. Παρ' όλα αυτά, δε θα επιμείνουμε άλλο στα πλαίσια της παρούσας εργασίας.

Σε αυτό το σημείο θα παρουσιάσουμε με ποιον τρόπο μπορούμε να διενεργήσουμε τον έλεγχο υπόθεσης $H_0 : \rho = 0$ έναντι της H_1 : τα δεδομένα είναι συσχετισμένα (ή έχουν θετική ή αρνητική συσχέτιση για μονόπλευρο έλεγχο). Για να γίνει αυτό θα πρέπει να ορίσουμε τη στατιστική συνάρτηση με την οποία θα κάνουμε τον έλεγχο.

α. Για το τ του Kendall, και εφόσον έχουμε ένα σχετικά μικρό δείγμα, $n \leq 10$, δεν μπορούμε να χρησιμοποιήσουμε το Κεντρικό Οριακό Θεώρημα. Για αυτόν τον λόγο υπάρχουν πίνακες που δίνουν τις κρίσιμες τιμές, είτε για μονόπλευρο είτε για αμφίπλευρο έλεγχο. Στη δεύτερη περίπτωση θα απορρίψουμε τη μηδενική υπόθεση, αν η κατά απόλυτο τιμή, η οποία υπολογίζεται από τα δεδομένα μας, είναι μεγαλύτερη από την κρίσιμη τιμή του πίνακα στο επίπεδο στατιστικής σημαντικότητας που έχουμε επιλέξει. Όταν έχουμε ένα σχετικά μεγάλο δείγμα, ($n \geq 30$), τότε μπορούμε να θεωρήσουμε ότι το $\tau_n \sim N(0, \frac{2(2n+5)}{9n(n-1)})$. Επομένως, για τον αμφίπλευρο έλεγχο, θα απορρίψουμε τη μηδενική υπόθεση σε επίπεδο στατιστικής σημαντικότητας α όταν:

$$\frac{|\tau_n|}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} > z_{\alpha/2}.$$

β. Για το ρ_S του Spearman όταν έχουμε ένα σχετικά μικρό δείγμα, $n \leq 10$, μπορούμε να χρησιμοποιήσουμε αντίστοιχους πίνακες με κρίσιμες τιμές και ανάλογα με το α και τον έλεγχο (αμφίπλευρο ή μονόπλευρο) να απορρίψουμε ή να μην απορρίψουμε την H_0 . Όταν έχουμε ένα σχετικά μεγάλο δείγμα, ($n \geq 30$), τότε μπορούμε να θεωρήσουμε ότι η στατιστική συνάρτηση $T = \rho_n \sqrt{\frac{n-2}{1-\rho_n^2}} \sim t_{n-2}$. Αποδεικνύεται, όμως, ότι για ένα αρκετά μεγάλο δείγμα $\rho_n \sim N(0, \frac{1}{n-1})$ και επομένως για δίπλευρο έλεγχο απορρίπτω τη μηδενική υπόθεση όταν:

$$\sqrt{n-1} |\rho_n| > z_{\alpha/2}.$$

Ας δούμε, όμως, ένα παράδειγμα, (συνέχεια από παράδειγμα του Κεφαλαίου 1, το οποίο υπάρχει στην εργασία των Genest και Favre [15]):

Παράδειγμα: Έχουμε παραγάγει δύο τυχαία δείγματα μεγέθους $n = 6$ από την $N(0, 1)$ (Πίνακας 1.1). Έχουμε διατάξει τα δεδομένα έχοντας επιλέξει αύξουσα διάταξη ως προς τη πρώτη τυχαία μεταβλητή και έχουμε βάλει τα αντίστοιχα ranks (Πίνακας 1.2).

Η εκτίμηση που παίρνουμε για το ρ_n είναι 0.02857. Επομένως, σε επίπεδο στατιστικής σημαντικότητας 0.05 υπολογίζω $\sqrt{5} \times 0.02857 = 0.06389 < 1.96$ και άρα δεν απορρίπτω την αρχική μου υπόθεση, ότι δηλαδή τα δεδομένα είναι ασυσχέτιστα. Το $p - value$ υπολογίζεται ως $2P(Z > \sqrt{n-1} | \rho_n |) = 0.949 > 0.05$. Βέβαια όλα τα παραπάνω θα ίσχυαν αν είχαμε μεγάλο δείγμα. Επειδή το δείγμα είναι μικρό ανατρέχουμε σε πίνακες με τις κρίσιμες τιμές του συντελεστή και για $n = 6$ και $\alpha=0.05$ η κρίσιμη τιμή είναι $0.886 > | \rho_n |$ και επομένως πάλι καταλήγουμε ότι δεν απορρίπτουμε την H_0 . Επομένως, μπορούμε να αποφανθούμε ότι τα δεδομένα μας είναι ασυσχέτιστα. Είναι προφανές ότι είναι ασυσχέτιστα, και μάλιστα είναι και ανεξάρτητα, καθώς έτσι τα έχουμε παραγάγει. Απλά εφαρμόζουμε τους τύπους που δώσαμε παραπάνω.

Η εκτίμηση που παίρνουμε για το τ_n είναι -0.06667. Επομένως, για $\alpha=0.05$ διαπιστώνουμε ότι $\sqrt{\frac{9n(n-1)}{2(2n+5)}} | \tau_n | = 0.18786 < 1.96$ και δεν απορρίπτω την αρχική υπόθεση, ότι δηλαδή τα δεδομένα είναι ασυσχέτιστα. Το $p - value$ είναι ίσο με $2P(Z > 0.18786) = 0.85 > 0.05$. Για μικρό δείγμα για $n = 6$ και $\alpha=0.05$ η κρίσιμη τιμή είναι $0.867 > | \tau_n |$. Και σε αυτήν την περίπτωση μπορούμε να αποφανθούμε ότι τα δεδομένα μας είναι ασυσχέτιστα σε επίπεδο στατιστικής σημαντικότητας $\alpha=0.05$. Ο αλγόριθμος που χρησιμοποιήθηκε για τα παραπάνω δίνεται στα Παραρτήματα.

2.3.4 Συντελεστής Εξάρτησης Ουρών

Όπως είδαμε παραπάνω, οι συντελεστές συσχέτισης τάξης στηρίζονται στη συμφωνία δύο συνεχών τυχαίων μεταβλητών. Στις περιπτώσεις, όμως που μας ενδιαφέρει η συμφωνία μεταξύ ακραίων τιμών των τυχαίων μεταβλητών και, επομένως, η εξάρτηση στις ουρές των κατανομών, υπάρχουν άλλοι συντελεστές οι οποίοι είναι πλέον κατάλληλοι. Αυτού του είδους η εξάρτηση σχετίζεται με τη δεσμευμένη πιθανότητα ότι η μία τυχαία μεταβλητή είναι μεγαλύτερη κάποιας τιμής, δεδομένου ότι η άλλη τυχαία μεταβλητή έχει ξεπεράσει κι αυτή μια συγκεκριμένη τιμή. Αν έχουμε συνεχείς τυχαίες μεταβλητές, η οποία είναι και η περίπτωση που θα εξετάσουμε, τότε το μέτρο εξάρτησης για τις ουρές που θα ορίσουμε είναι συνάρτηση του Copula και, επομένως, είναι αναλλοίωτο ως προς γνησίως αύξοντες μετασχηματισμούς.

Ο συντελεστής εξάρτησης ουρών, όπως αναφέρει και το όνομά του, είναι ένα μέτρο εξάρτησης για τις ουρές των κατανομών. Με απλά λόγια, μετρά την εξάρτηση των μεταβλητών που παίρνουν τιμές στο πάνω-δεξιά και στο κάτω-αριστερά τεταρτημόριο του I^2 . Ας δώσουμε, όμως, το μαθηματικό ορισμό.

ΟΡΙΣΜΟΣ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με συναρτήσεις κατανομής F, G αντίστοιχα. Ο δείκτης εξάρτησης της άνω ουράς, (upper tail dependence parameter) λ_U είναι το όριο, (αν υπάρχει), καθώς το $t \rightarrow 1$, της δεσμευμένης πιθανότητας ότι το Y είναι μεγαλύτερο του $100t$ -οστού ποσοστημορίου της G , δεδομένου ότι το X είναι μεγαλύτερο του $100t$ -οστού ποσοστημορίου της F , δηλαδή

$$\lambda_U = \lim_{t \rightarrow 1^-} P(Y > G^{(-1)}(t) | X > F^{(-1)}(t)).$$

Όμοια, ο δείκτης εξάρτησης της κάτω ουράς, (lower tail dependence parameter), λ_L είναι το όριο, (αν υπάρχει), καθώς το $t \rightarrow 0$, της δεσμευμένης πιθανότητας ότι το Y είναι μικρότερο ή ίσο του $100t$ -οστού ποσοστημορίου της G , δεδομένου ότι το X είναι μικρότερο ή ίσο του $100t$ -οστού ποσοστημορίου της F , δηλαδή

$$\lambda_L = \lim_{t \rightarrow 0^+} P(Y \leq G^{(-1)}(t) | X \leq F^{(-1)}(t)).$$

Οι δείκτες αυτοί είναι μη παραμετρικοί καθώς δεν εξαρτώνται από τις περιθώριες κατανομές παρά μόνο από το Copula των X, Y .¹² Συγκεκριμένα, μπορούμε να γράψουμε τις παραπάνω σχέσεις χρησιμοποιώντας το Copula και την από κοινού συνάρτηση επιβίωσης:

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}$$

και

$$\lambda_U = \lim_{t \rightarrow 1^-} \frac{\bar{C}(t, t)}{1 - t}.$$

Αυτό συμβαίνει γιατί:

$$C(t, t) = P(U \leq t, V \leq t) = P(V \leq t | U \leq t)P(U \leq t)$$

και επειδή η U είναι ομοιόμορφη το $(0,1)$ έχουμε ότι $P(U \leq t) = t$.

Όμοια

$$\bar{C}(t, t) = P(U > t, V > t) = P(V > t | U > t)P(U > t).$$

Επίσης, ο μηδενισμός τους συνεπάγεται ανεξαρτησία, στην άνω ή στην κάτω ουρά. Αυτό σημαίνει πως ανεξάρτητα από το πόσο συσχετισμένες είναι οι δύο τυχαίες μεταβλητές, το να πάρει η μία ακραία τιμή δεν εξαρτάται από το αν θα πάρει και η άλλη ακραία τιμή. Ας δούμε όμως ένα παράδειγμα:

Παράδειγμα:

Έστω $a, b \in \mathbf{I}$ με $a + b \leq 1$ και το Copula

$$C_{a,b}(u, v) = aM(u, v) + (1 - a - b)\Pi(u, v) + bW(u, v),¹³$$

όπου M, Π, W είναι το άνω φράγμα, το Copula-γινόμενο και το κάτω φράγμα, αντίστοιχα. Θα υπολογίσουμε τους δείκτες εξάρτησης ουρών.

Έχουμε:

¹² Γενικεύσεις των συντελεστών εξάρτησης ουρών στις n διαστάσεις μπορεί να τις βρει ο αναγνώστης στην εργασία των Schmid, Schmidt, Blumentritt, Gaißer και Ruppert με τίτλο *Copula-Based Measures of Multivariate Association*, η οποία συγκαταλέγεται στο βιβλίο των Jaworski, Durante, Härdle και Rychlik [23].

¹³ Γενικά, αποδεικνύεται ότι οποιοσδήποτε κυρτός συνδυασμός Copula είναι επίσης Copula.

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{at + (1 - a - b)t^2}{t} = \lim_{t \rightarrow 0^+} a + (1 - a - b)t = a$$

και

$$\begin{aligned} \lambda_U &= \lim_{t \rightarrow 1^-} \frac{1 - 2t + C(t, t)}{1 - t} = 2 - \lim_{t \rightarrow 1^-} \frac{1 - C(t, t)}{1 - t} = \\ &= 2 - \lim_{t \rightarrow 1^-} \frac{(1 - t^2) - at(1 - t) + b(1 - t)^2}{1 - t} = 2 - \lim_{t \rightarrow 1^-} ((1 + t) - at + b(1 - t)) = a. \end{aligned}$$

Αυτό που παρατηρούμε στο παραπάνω παράδειγμα είναι ότι η συγκεκριμένη οικογένεια επιτυγχάνει και τα δύο φράγματα, καθώς έτσι έχει κατασκευαστεί. Αυτό δεν είναι πολύ σύνηθες καθώς τα Copula έχουν κάποιους περιορισμούς εκ κατασκευής ή λόγω του παραμετρικού χώρου και συνήθως επιτυγχάνουν το ένα από τα δύο φράγματα. Τα Copula τα οποία επιτυγχάνουν και τα δύο φράγματα και προφανώς και το γινόμενο ονομάζονται εκτενή (comprehensive). Η παραπάνω εκτενή οικογένεια προτάθηκε από τον Frèchet το 1958 και όπως διαπιστώνουμε από τους δείκτες για την εξάρτηση των ουρών έχει κάποια συμμετρία.

2.3.5 Θετική Τεταρτημοριακή Εξάρτηση

Αφού ορίσαμε τους συντελεστές εξάρτησης ουρών, θα παρουσιάσουμε και κάποιες άλλες ιδιότητες που σχετίζονται με την εξάρτηση ή τη συσχέτιση δύο μεταβλητών και θα αναφέρουμε ποιες από αυτές θεωρούνται πιο ισχυρές έννοιες. Συγκεκριμένα θα παρουσιάσουμε τη θετική τεταρτημοριακή εξάρτηση (positively quadrant dependence), (βλπ. Nelsen [28] και Joe [24]), και δύο πιο ισχυρές έννοιες από αυτήν, τη συνάφεια (association), (βλπ. Balakrishnan και Lai [4] και Joe [24]), και τη μονοτονία στις ουρές (βλπ. Nelsen [28], Balakrishnan, Lai [4] και Joe [24]).

ΟΡΙΣΜΟΣ: Έστω X, Y δύο τυχαίες μεταβλητές. Αυτές θα ονομάζονται θετικά τεταρτημοριακά εξαρτημένες, (PQD), αν για κάθε $(x, y) \in \mathbf{R}^2$ ισχύει

$$P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y)$$

ή ισοδύναμα, στην περίπτωση της συνάρτησης επιβίωσης

$$P(X > x, Y > y) \geq P(X > x)P(Y > y).$$

Αυτό σημαίνει πως, όταν οι τυχαίες μεταβλητές είναι ταυτόχρονα μικρότερες από κάποια τιμή, (ή μεγαλύτερες στη δεύτερη περίπτωση), τότε η από κοινού πιθανότητα να συμβεί αυτό είναι μεγαλύτερη ή το πολύ ίση με την από κοινού πιθανότητα στην περίπτωση της ανεξαρτησίας.

Επομένως, αν X, Y είναι δύο συνεχείς τυχαίες μεταβλητές με περιθώριες συναρτήσεις κατανομής F, G αντίστοιχα, από κοινού συνάρτηση κατανομής H και Copula C , τότε είναι PQD αν και μόνο αν ισχύει:

$$H(x, y) \geq F(x)G(y) \text{ για κάθε } (x, y) \in \mathbf{R}^2.$$

Αντίστοιχα για το Copula η ιδιότητα PQD οδηγεί στην ανισότητα:

$$C(u, v) \geq uv \text{ για κάθε } (u, v) \in \mathbf{I}^2.$$

Στη δεύτερη περίπτωση ουσιαστικά έχουμε μια ανισοτική σχέση μεταξύ του Copula των X, Y και του Copula-γινόμενο. Δηλαδή έχουμε μια διάταξη μεταξύ δύο Copula, όπως είχαμε και στην περίπτωση των φραγμάτων των Frèchet-Hoeffding στις δύο διαστάσεις, τη διάταξη συμφωνίας.

Να τονίσουμε σε αυτό το σημείο ότι η συγκεκριμένη διάταξη δεν είναι πλήρης με την έννοια ότι δεν μπορούμε να συγκρίνουμε οποιοδήποτε ζεύγος Copulas. Για την περίπτωση των PQD έχουμε μια διάταξη συμφωνίας μεταξύ του Copula των τυχαίων μεταβλητών και του Copula-γινόμενου. Για παράδειγμα αν X, Y εκφράζουν χρόνους ζωής δύο μονάδων, τότε η πιθανότητα οι χρόνοι ζωής των δύο μονάδων να μην ξεπεράσουν δύο συγκεκριμένες χρονικές στιγμές που έχουμε ορίσει, είναι μεγαλύτερη όταν οι χρόνοι ζωής των μονάδων εξαρτώνται μεταξύ τους από ότι αν είναι ανεξάρτητοι.

Πώς συνδέεται, όμως, αυτή η έννοια της εξάρτησης με την πολύ γνωστή έννοια της συσχέτισης; Για να απαντήσουμε σε αυτό το ερώτημα, θα χρειαστεί να γράψουμε τις παραπάνω σχέσεις υπό μορφή συσχέτισης η απόδειξη του οποίου χρησιμοποιεί ένα Λήμμα του Hoeffding (1940)(βλπ. Balakrishnan και Lai [4], καθώς και Lehmann [26]).

Ο Lehmann (1966) [26], απέδειξε ότι οι παραπάνω σχέσεις γράφονται στη μορφή:

$$\text{cov}(g_1(X), g_2(Y)) \geq 0$$

για κάθε ζεύγος αυξουσών συναρτήσεων g_1, g_2 στο \mathbf{R} . Άρα, αν δύο μεταβλητές είναι PQD τότε είναι συσχετισμένες, (θεωρούμε τις g_1, g_2 ως ταυτοτικές συναρτήσεις), ενώ το αντίθετο εν γένει δεν ισχύει. Βέβαια, για να είναι συσχετισμένες οι τυχαίες μεταβλητές οι παραπάνω ανισότητες, θα πρέπει να είναι γνήσιες για ένα τουλάχιστον ζεύγος (x, y) , καθώς οι ανεξάρτητες τυχαίες μεταβλητές, όπως έχουμε ήδη αναφέρει, είναι και ασυσχέτιστες.

Στη συνέχεια θα ορίσουμε τη συνάφεια (association)¹⁴ μεταξύ δύο τυχαίων μεταβλητών.

ΟΡΙΣΜΟΣ: Δύο τυχαίες μεταβλητές X, Y ονομάζονται θετικά συναφείς, (positively associated, PA), όταν

$$E(g_1(X, Y)g_2(X, Y)) \geq E(g_1(X, Y))E(g_2(X, Y))$$

για οποιοσδήποτε πραγματικές, αύξουσες κατά συντεταγμένες συναρτήσεις g_1, g_2 για τις οποίες ορίζονται οι παραπάνω αναμενόμενες τιμές.

Η παραπάνω σχέση σε μορφή συσχετίσεων γράφεται:

$$cov(g_1(X, Y), g_2(X, Y)) \geq 0.$$

Αποδεικνύεται ότι η θετική συνάφεια συνεπάγεται θετική τεταρτημοριακή εξάρτηση και απλή συσχέτιση, ενώ το αντίθετο εν γένει δεν ισχύει.

Ισχύουν, επομένως, οι επόμενες συνεπαγωγές (βλπ. τρίτο κεφάλαιο των Balakrishnan και Lai [4], καθώς και δεύτερο κεφάλαιο του βιβλίου του Joe [24]):

$$\text{Τέλεια θετική συσχέτιση} \Rightarrow PA \Rightarrow PQD \Rightarrow \rho(X, Y) \geq 0, \rho_S(X, Y) \geq 0, \tau_{X, Y} \geq 0.$$

Για την αρνητική τεταρτημοριακή συσχέτιση (NQD) ισχύουν όλα τα παραπάνω με αλλαγμένες τις φορές των ανισοτήτων και, επομένως, δεν θα επεκταθούμε περαιτέρω. Ας δούμε, όμως, τι γίνεται στις ουρές, δηλαδή όταν έχουμε ακραίες περιπτώσεις.

¹⁴ Τον ορισμό καθώς και ιδιότητες της έννοιας της συνάφειας (association) μπορεί να τις βρεί ο αναγνώστης στην εργασία των Esary, Proschan και Walkup [13].

ΟΡΙΣΜΟΣ: Έστω X, Y δύο τυχαίες μεταβλητές. Τότε:

1. Η Y είναι φθίνουσα στην αριστερή ουρά στο X (left tail decreasing $LTD(Y|X)$) αν η δεσμευμένη πιθανότητα $P(Y \leq y|X \leq x)$ είναι γνησίως φθίνουσα συνάρτηση του x για κάθε y .
2. Η X είναι φθίνουσα στην αριστερή ουρά στο Y ($LTD(X|Y)$) αν η δεσμευμένη πιθανότητα $P(X \leq x|Y \leq y)$ είναι γνησίως φθίνουσα συνάρτηση του y για κάθε x .
3. Η Y είναι αύξουσα στη δεξιά ουρά στο X (right tail increasing $RTI(Y|X)$) αν η δεσμευμένη πιθανότητα $P(Y > y|X > x)$ είναι γνησίως αύξουσα συνάρτηση του x για κάθε y .
4. Η X είναι αύξουσα στη δεξιά ουρά στο Y ($RTI(X|Y)$) αν η δεσμευμένη πιθανότητα $P(X > x|Y > y)$ είναι γνησίως αύξουσα συνάρτηση του y για κάθε x .

Για να καταλάβουμε τον παραπάνω ορισμό, θα δώσουμε ένα παράδειγμα:

Έστω X, Y δύο τυχαίες μεταβλητές για τις οποίες η Y είναι LTD στο X , και έστω $x_1 < x_2$.

Τότε:

$$P(Y \leq y|X \leq x_1) \geq P(Y \leq y|X \leq x_2).$$

Αν τα X, Y εκφράζουν χρόνους ζωής μονάδων, τότε διαισθητικά η παραπάνω ανισοτική σχέση εκφράζει το ότι η πιθανότητα η δεύτερη μονάδα να έχει μικρό χρόνο ζωής φθίνει όσο ο χρόνος ζωής της πρώτης αυξάνει. Αυτό ισχύει στην αριστερή ουρά, δηλαδή όσο αυξάνει το x η πιθανότητα το Y να πέσει στην αριστερή ουρά μικραίνει.¹⁵

Οι παραπάνω ανισοτικές σχέσεις συνεπάγονται θετική τεταρτημοριακή εξάρτηση. Για παράδειγμα,

$$P(Y \leq y|X \leq x) \geq P(Y \leq y|X \leq \infty) = P(Y \leq y) \Rightarrow$$

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y|X \leq x) \geq P(X \leq x)P(Y \leq y).$$

Όμοια αποδεικνύονται και για τις υπόλοιπες. Επίσης αποδεικνύεται ότι η RTI και η LTD συνεπάγονται συνάφεια. Τα αντίστροφα των παραπάνω εν γένει δεν ισχύουν.

¹⁵ Για τη γεωμετρική ερμηνεία των LTD, RTI παροτρύνουμε τον αναγνώστη να κοιτάξει την εργασία του Nelsen, *Copulas and Association*, η οποία συμπεριλαμβάνεται στο βιβλίο των Dall'Aglio, Kotz και Salinetti [7].

Θα κλείσουμε τη μονοτονία στις ουρές με δύο θεωρήματα τα οποία αναφέρονται και στα Copulas (βλπ. Nelsen [28]). Για να εισαγάγουμε τις συναρτήσεις κατανομής και, επομένως, τα Copulas, μπορούμε να γράψουμε,

$$P(Y \leq y | X \leq x) = \frac{P(Y \leq y, X \leq x)}{P(X \leq x)} = \frac{H(x, y)}{F(x)}$$

όπου X, Y είναι τυχαίες μεταβλητές με συναρτήσεις κατανομής F, G και από κοινού συνάρτηση κατανομής H και, επίσης, ισχύει ότι $F(x) > 0$.

ΘΕΩΡΗΜΑ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με Copula C . Τότε:

1. $LTD(Y|X)$ αν και μόνο αν για κάθε $v \in \mathbf{I}$ η συνάρτηση $C(u, v)/u$ είναι γνησίως φθίνουσα ως προς u .
2. $LTD(X|Y)$ αν και μόνο αν για κάθε $u \in \mathbf{I}$ η συνάρτηση $C(u, v)/v$ είναι γνησίως φθίνουσα ως προς v .
3. $RTI(Y|X)$ αν και μόνο αν για κάθε $v \in \mathbf{I}$ η συνάρτηση $1 - u - v + C(u, v)/(1 - u)$ είναι γνησίως αύξουσα ως προς u .
4. $RTI(X|Y)$ αν και μόνο αν για κάθε $u \in \mathbf{I}$ η συνάρτηση $1 - u - v + C(u, v)/(1 - v)$ είναι γνησίως αύξουσα ως προς v .

Επειδή, όμως, οι παραπάνω σχέσεις είναι αρκετά δύσκολα υπολογίσιμες, στην πράξη προτιμάται το παρακάτω θεώρημα που αφορά τις μερικές παραγώγους.

ΘΕΩΡΗΜΑ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με Copula C . Τότε:

1. $LTD(Y|X)$ αν και μόνο αν για κάθε $v \in \mathbf{I}$ ισχύει $\frac{\partial C(u, v)}{\partial u} \leq \frac{C(u, v)}{u}$ για κάθε u .
2. $LTD(X|Y)$ αν και μόνο αν για κάθε $u \in \mathbf{I}$ ισχύει $\frac{\partial C(u, v)}{\partial v} \leq \frac{C(u, v)}{v}$ για κάθε v .
3. $RTI(Y|X)$ αν και μόνο αν για κάθε $v \in \mathbf{I}$ ισχύει $\frac{\partial C(u, v)}{\partial u} \leq \frac{v - C(u, v)}{1 - u}$ για κάθε u .
4. $RTI(X|Y)$ αν και μόνο αν για κάθε $u \in \mathbf{I}$ ισχύει $\frac{\partial C(u, v)}{\partial v} \leq \frac{u - C(u, v)}{1 - v}$ για κάθε v .

Σε αυτό το σημείο θα δώσουμε μία ακόμα σχέση η οποία συνδέει τη μονοτονία στις ουρές με τους δείκτες συσχέτισης τάξης.

ΠΡΟΤΑΣΗ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές. Αν η Y είναι *LTD* και *RTI* ως προς X , τότε $\rho_{X,Y} \geq \tau_{X,Y} \geq 0$. Όμοια για την X .

Η παραπάνω σχέση δεν ισχύει για θετικά τεταρτημοριακά συσχετισμένες τυχαίες μεταβλητές, καθώς είναι ανεπαρκής ως έννοια για να αποδειχθούν τα παραπάνω (βλπ. Nelsen [28]). Επομένως, αν θέλουμε να διατάξουμε και τη μονοτονία στις ουρές σε σχέση με τις προηγούμενες έννοιες, έχουμε (βλπ. τρίτο κεφάλαιο των Balakrishnan και Lai [4], καθώς και δεύτερο κεφάλαιο του βιβλίου του Joe [24]):

$$\text{Τέλεια θετική συσχέτιση} \Rightarrow LTD, RTI \Rightarrow PA \Rightarrow PQD \Rightarrow \rho(X, Y) \geq 0, \rho_S(X, Y) \geq 0, \tau_{X,Y} \geq 0.$$

Δίνουμε παρακάτω μια μικρή γενίκευση στις n διαστάσεις για τη θετική τεταρτημοριακή συσχέτιση χωρίς όμως να επεκταθούμε περισσότερο.

ΟΡΙΣΜΟΣ: Έστω $\mathbf{X} = (X_1, \dots, X_n)$ ένα τυχαίο διάνυσμα. Θα ονομάζουμε το \mathbf{X} :

α. θετικά εξαρτημένο στο κάτω ορθομόριο (positively lower orthant dependent *PLOD*), αν για κάθε $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}^n$, ισχύει:

$$P(\mathbf{X} \leq \mathbf{x}) \geq \prod_{i=1}^n P(X_i \leq x_i)$$

β. θετικά εξαρτημένο στο άνω ορθομόριο (positively upper orthant dependent *PUOD*), αν για κάθε $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}^n$, ισχύει:

$$P(\mathbf{X} > \mathbf{x}) \geq \prod_{i=1}^n P(X_i > x_i)$$

γ. θετικά ορθομοριακά εξαρτημένο (positively orthant dependent *POD*), αν για κάθε $\mathbf{x} \in \mathbf{R}^n$ είναι *PLOD* και *PUOD*.

Για αρνητική εξάρτηση στο άνω ή στο κάτω ορθομόριο απλά αλλάζουμε τις φορές των ανισοτήτων. Να σημειώσουμε σε αυτό το σημείο πως οι δύο πρώτες σχέσεις είναι ισοδύναμες μόνο στην δισδιάστατη περίπτωση, η οποία είναι και η θετική τεταρτημοριακή εξάρτηση. Για $n \geq 3$ δεν είναι ισοδύναμες.

ΠΑΡΑΔΕΙΓΜΑ¹⁶: Έστω \mathbf{X} μία τρισδιάστατη τυχαία μεταβλητή η οποία παίρνει τις εξής τέσσερις τιμές, $(1,1,1)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, με πιθανότητα $\frac{1}{4}$ για την κάθε μία. Η X είναι *PUOD*, όμως δεν είναι *PLOD*. Για την *PUOD* ενδεικτικά υπολογίζουμε:

$$P(\mathbf{X} > \mathbf{0}) = \frac{1}{4} \geq \frac{2}{4} \frac{2}{4} \frac{2}{4} = \frac{1}{8}.$$

Η ανισότητα ισχύει και για όλους τους υπόλοιπους συνδυασμούς. Η \mathbf{X} δεν είναι *PLOD* καθώς:

$$P(\mathbf{X} \leq \mathbf{0}) = 0 \text{ και } P(X_1 \leq 0)P(X_2 \leq 0)P(X_3 \leq 0) = \frac{1}{8}.$$

Αποδεικνύεται πως η γενίκευση της έννοιας της συνάφειας συνεπάγεται τις ιδιότητες *PLOD* και *PUOD* (βλπ. Joe [24]).

¹⁶ Το συγκεκριμένο παράδειγμα συμπεριλαμβάνεται στο πέμπτο κεφάλαιο του βιβλίου του Nelsen [28].

Κεφάλαιο 3

ΟΙΚΟΓΕΝΕΙΕΣ ΤΩΝ COPULAS

3.1 ΕΙΣΑΓΩΓΗ

Όπως αναφέραμε στα προηγούμενα κεφάλαια, τα Copulas είναι πολύ σημαντικά για τη δημιουργία από κοινού συναρτήσεων κατανομής και, γενικά, στη δημιουργία μοντέλων με πολύπλοκες ιδιότητες, όπως οι βαριές ουρές. Στην παράγραφο αυτή θα παρουσιάσουμε διάφορες οικογένειες Copulas, η επιλογή των οποίων έγινε βάσει της χρησιμότητάς τους και των ιδιοτήτων τους.

Όσον αφορά την επιλογή του Copula, υπάρχουν ερευνητές, οι οποίοι επιλέγουν να προσαρμόζουν οποιοδήποτε μοντέλο από τις γνωστές και εύχρηστες οικογένειες και να εξετάζουν, αν η προσαρμογή είναι ικανοποιητική. Η συγκεκριμένη διαδικασία, όμως, δεν είναι απόλυτα σωστή, καθώς υπάρχουν περιπτώσεις στις οποίες κανείς πρέπει να λαμβάνει υπ' όψιν τις ιδιότητες των Copulas, όπως η συμπεριφορά στις ουρές, η συμμετρία κ.τ.λ.

Για το λόγο αυτό, όταν επιλέγουμε μια οικογένεια Copulas, θα πρέπει να εξετάζουμε αν έχει κάποιες “καλές” ιδιότητες. Για παράδειγμα, θα πρέπει τα μέλη της εκάστοτε οικογένειας να έχουν κάποια φυσική ερμηνεία, να περιγράφουν διαφορετικού τύπου εξαρτήσεις και να συμπεριλαμβάνουν την περίπτωση της ανεξαρτησίας. Επίσης, η εξάρτηση στις ουρές είναι μία ακόμα επιθυμητή ιδιότητα.

Στο κεφάλαιο αυτό, επομένως, θα παρουσιάσουμε διαφορετικές οικογένειες Copulas, όπως τα Elliptical Copulas και τα Archimedean Copulas και κάποια από τα πιο σύννηθη μέλη τους. Εν συνεχεία θα χρησιμοποιήσουμε στις οικογένειες αυτές τους δείκτες εξάρτησης – συσχέτισης που παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Θα κλείσουμε το κεφάλαιο

παρουσιάζοντας τις πιο σημαντικές ιδιότητες κάθε οικογένειας.

3.2 Elliptical (Ελλειπτικά) Copulas

Τα Elliptical Copulas¹, όπως άλλωστε αναφέρει και το όνομά τους, σχετίζονται με τις ελλειπτικές κατανομές, όπως είναι η Κανονική Κατανομή και η Κατανομή t -Student. Οι ελλειπτικές κατανομές στις δύο διαστάσεις έχουν το χαρακτηριστικό ότι, αν τμήσουμε τις επιφάνειες στο χώρο με οριζόντια επίπεδα και δώσουμε τη γραφική αναπαράσταση των καμπυλών που δημιουργούνται, τότε αυτές θα είναι ελλείψεις. Ας ορίσουμε, όμως, τις ελλειπτικές κατανομές, αφού πρώτα ορίσουμε και τις σφαιρικές κατανομές.

ΟΡΙΣΜΟΣ (Σφαιρική κατανομή):

Μια d -διάστατη τυχαία μεταβλητή $\mathbf{Y} = (Y_1, Y_2, \dots, Y_d)^t$, με τιμές στον Ευκλείδειο χώρο \mathbf{R}^d , $d \geq 2$, λέγεται ότι ακολουθεί σφαιρική κατανομή, αν η \mathbf{Y} και η $P\mathbf{Y}$ έχουν την ίδια κατανομή για κάθε ορθογώνιο $p \times p$ πίνακα P . □

Επειδή πολύ συχνά υπολογίζουμε τη χαρακτηριστική συνάρτηση μιας τυχαίας μεταβλητής ώστε να καταλάβουμε ποια κατανομή ακολουθεί, θα δώσουμε τη χαρακτηριστική συνάρτηση την οποία θα πρέπει να έχει η παραπάνω d -διάστατη τυχαία μεταβλητή $\mathbf{Y} = (Y_1, Y_2, \dots, Y_d)^t$. Αυτή θα είναι της μορφής:

$$\phi_{\mathbf{Y}}(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{Y}}) = g(\|\mathbf{t}\|^2)$$

όπου $\mathbf{t} \in \mathbf{R}^d$ και $\|\mathbf{t}\| = (t_1^2 + \dots + t_d^2)^{\frac{1}{2}}$ είναι η ευκλείδεια νόρμα και η g είναι μια μη αρνητική πραγματική συνάρτηση η οποία ονομάζεται χαρακτηριστικός γεννήτορας. Τότε θα συμβολίζουμε με $\mathbf{Y} \sim \Sigma\mathbf{K}_d(\boldsymbol{\mu}, g)$, όταν η \mathbf{Y} ακολουθεί τη σφαιρική κατανομή με παράμετρο θέσης $\boldsymbol{\mu}$ και γεννήτρια συνάρτηση g .² Ο ορισμός των σφαιρικών κατανομών ουσιαστικά καταδεικνύει πως οι σφαιρικές κατανομές παραμένουν αναλλοίωτες σε περιστροφές και

¹ Για τη θεωρία των Elliptical Copulas παροτρύνουμε τον αναγνώστη στο τέταρτο κεφάλαιο του βιβλίου SIMULATING COPULAS: Stochastic Models, Sampling Algorithms and Applications [27].

² Για να δείξει κάποιος ότι μια d -διάστατη τυχαία μεταβλητή $\mathbf{Y} = (Y_1, Y_2, \dots, Y_d)^t$ ακολουθεί σφαιρική κατανομή αρκεί να δείξει ότι για κάθε $\boldsymbol{\alpha} \in \mathbf{R}^d$ ισχύει: $\boldsymbol{\alpha}'\mathbf{Y} =_{d\|} \boldsymbol{\alpha} \| Y_1$

ανακλάσεις και πως είναι ακτινικά συμμετρικές.

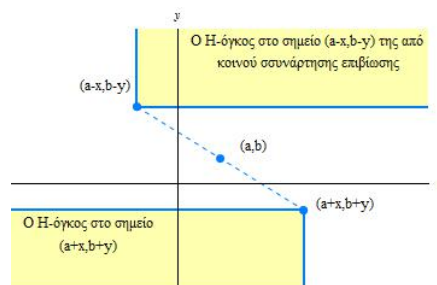
ΟΡΙΣΜΟΣ (Ακτινική Συμμετρία): Μία d -διάστατη τ.μ $\mathbf{X} = (X_1, X_2, \dots, X_d)$ ονομάζεται ακτινικά συμμετρική, αν η $X_1 - x_1$ έχει την ίδια κατανομή με την $x_1 - X_1$, η $X_2 - x_2$ έχει την ίδια κατανομή με την $x_2 - X_2, \dots$, η $X_d - x_d$ έχει την ίδια κατανομή με την $x_d - X_d$, όπου $x_1, x_2, \dots, x_d \in \mathbf{R}$. Δηλαδή η $(X_1 - x_1, \dots, X_d - x_d)$ έχει την ίδια κατανομή με την $(x_1 - X_1, \dots, x_d - X_d)$ ή αντίστοιχα η (X_1, \dots, X_d) έχει την ίδια κατανομή με την $(2x_1 - X_1, \dots, 2x_d - X_d)$.

Στα πλαίσια της παρούσας εργασίας θα ασχοληθούμε στη συνέχεια μόνο με τη διοδιάστατη περίπτωση και θα εκφράσουμε την ακτινική συμμετρία δύο συνεχών τυχαίων μεταβλητών με χρήση της από κοινού συνάρτησης κατανομής τους και της από κοινού συνάρτησης επιβίωσής τους. Συγκεκριμένα, ισχύει το παρακάτω θεώρημα:

ΘΕΩΡΗΜΑ: Έστω X, Y δύο συνεχείς τυχαίες μεταβλητές με περιθώριες συναρτήσεις κατανομής F, G , αντίστοιχα, και από κοινού συνάρτηση κατανομής H . Έστω, επίσης, (a, b) ένα σημείο στο \mathbf{R}^2 . Τότε, η διοδιάστατη τυχαία μεταβλητή (X, Y) είναι ακτινικά συμμετρική ως προς το (a, b) αν και μόνο αν ισχύει:

$$H(a + x, b + y) = \bar{H}(a - x, b - y), \text{ για κάθε } (x, y) \in \mathbf{R}^2.$$

Σχήμα 3.1: Περιοχές που ο H -όγκος δυο ακτινικά συμμετρικών τυχαίων μεταβλητών είναι ίδιος



Ο H -όγκος στις δύο σκιασμένες περιοχές είναι ίδιος και αυτό ισχύει για όλες τις συμμετρικά ακτινικές τυχαίες μεταβλητές (βλπ. Nelsen [28]). Επίσης, όπως φαίνεται και στο γράφημα, τα σημεία $(a - x, b - y)$, $(a + x, b + y)$ είναι πάνω σε μία ακτίνα εκατέρωθεν του (a, b) . Έτσι προέκυψε και ο όρος της ακτινικής συμμετρίας.

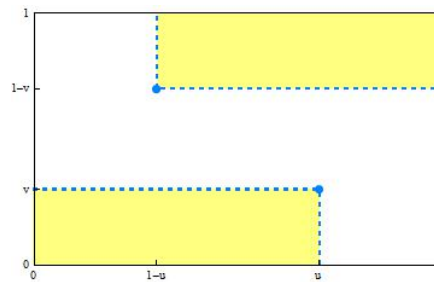
Μία διδιάστατη τυχαία μεταβλητή η οποία είναι ακτινικά συμμετρική είναι και περιθωριακά συμμετρική³, ενώ το αντίστροφο εν γένει δεν ισχύει. Ένα απλό παράδειγμα είναι να πάρουμε δύο τυχαίες μεταβλητές που κατανέμονται ομοιόμορφα στο $(0, 1)$ και επομένως είναι συμμετρικές ως προς το $\frac{1}{2}$ αλλά η από κοινού τους να μην ικανοποιεί το παραπάνω θεώρημα.

Θα χρειαστούμε και την έννοια της ακτινικής συμμετρίας για τα Copula.

ΟΡΙΣΜΟΣ (Ακτινική Συμμετρία των Copulas): Ένα Copula θα θεωρείται ακτινικά συμμετρικό αν είναι ίσο με το Copula Επιβίωσης, δηλαδή αν για κάθε $(u, v) \in [0, 1]^2$ ισχύει:

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

Σχήμα 3.2: Περιοχές που ο C -όγκος δυο ακτινικά συμμετρικών τυχαίων μεταβλητών είναι ίδιος



Όπως και στο προηγούμενο γράφημα, ο C -όγκος στις δύο σκιασμένες περιοχές είναι ίδιος και αυτό ισχύει για όλες τις συμμετρικά ακτινικές τυχαίες μεταβλητές (βλπ. Nelsen [28]).

Όσον αφορά τις ελλειπτικές κατανομές, είναι και αυτές ακτινικά συμμετρικές και ορίζονται ως γραμμικοί μετασχηματισμοί των σφαιρικών κατανομών. Συγκεκριμένα:

³ Μια διδιάστατη τυχαία μεταβλητή (X, Y) είναι περιθωριακά συμμετρική ή συμμετρική ως προς σημείο (a, b) αν η X είναι συμμετρική ως προς το a και η Y συμμετρική ως προς το b .

ΟΡΙΣΜΟΣ (Ελλειπτική κατανομή):

Μια d -διάστατη τυχαία μεταβλητή $\mathbf{X} = (X_1, X_2, \dots, X_d)$, με τιμές στον Ευκλείδειο χώρο \mathbf{R}^d , $d \geq 2$, λέγεται ότι ακολουθεί ελλειπτική κατανομή, όταν γράφεται στη μορφή:

$$\mathbf{X} = \boldsymbol{\mu} + \Sigma^{\frac{1}{2}} \mathbf{Y}$$

όπου η \mathbf{Y} ακολουθεί σφαιρική κατανομή, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)' \in \mathbf{R}^d$ και $\Sigma = (\sigma_{ij})$ ένας $d \times d$ θετικά ορισμένος πίνακας με πραγματικά στοιχεία και η ρίζα του, δηλαδή ο πίνακας $\Sigma^{\frac{1}{2}}$, υπολογίζεται από τη σχέση $E\Lambda^{\frac{1}{2}}E'$, όπου E είναι ο πίνακας ο οποίος έχει στήλες τα ιδιοδιανύσματα του Σ και $\Lambda^{\frac{1}{2}}$ είναι ο διαγώνιος πίνακας με στοιχεία τις ρίζες των ιδιοτιμών του πίνακα Σ .

Στη συνέχεια θα χρησιμοποιούμε το συμβολισμό $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g)$ για να δηλώσουμε ότι η d -διάστατη τυχαία μεταβλητή \mathbf{X} ακολουθεί ελλειπτική κατανομή με παράμετρο θέσης $\boldsymbol{\mu}$, παράμετρο κλίμακος Σ και $g(\cdot)$ είναι ο χαρακτηριστικός γεννήτορας της σφαιρικής κατανομής. Το διάνυσμα $\boldsymbol{\mu}$ είναι μοναδικό αλλά τα g, Σ δεν είναι μοναδικώς καθορισμένα από την κατανομή. Είναι εύκολο να διαπιστώσει κανείς πως οι σφαιρικές κατανομές είναι μία ειδική περίπτωση των ελλειπτικών, αν θέσουμε $\boldsymbol{\mu} = \mathbf{0}$ και $\Sigma = I_d$.

Ας δούμε, όμως, και κάποιες ενδιαφέρουσες ιδιότητες των ελλειπτικών κατανομών.⁴ Συγκεκριμένα, αν $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g)$ ισχύουν τα εξής:

1) Η χαρακτηριστική συνάρτηση του \mathbf{X} έχει τη μορφή:

$$\phi_{\mathbf{X}}(\mathbf{t}) = \exp(it'\boldsymbol{\mu})g(\mathbf{t}'\Sigma\mathbf{t}), \mathbf{t} \in \mathbf{R}^d.$$

2) Αν η \mathbf{X} έχει συνάρτηση πυκνότητας, τότε αυτή θα είναι της μορφής:

$$f_{\mathbf{X}}(\mathbf{x}) = k_d |\Sigma|^{-\frac{1}{2}} h((\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))$$

⁴ Οι ιδιότητες των ελλειπτικών κατανομών έχουν χρησιμοποιηθεί από το πρώτο κεφάλαιο της εργασίας του Κούτρας Β. Μάρκος [1] και το τέταρτο κεφάλαιο του βιβλίου των Mai και Scherer [27].

όπου k είναι μια κατάλληλη σταθερά⁵ και h είναι μια μη αρνητική πραγματική συνάρτηση για την οποία ισχύει ότι

$$\int_0^\infty t^{d-1} h(t^2) dt < \infty.$$

3) Αν υπάρχουν οι ροπές πρώτης και δεύτερης τάξης της \mathbf{X} ισχύουν τα εξής:

$$E(\mathbf{X}) = \boldsymbol{\mu} \text{ και } \text{Cov}(\mathbf{X}, \mathbf{X}) = \frac{1}{2\pi} k' \Sigma$$

όπου $k' = \frac{k_d}{k_{d+2}}$ είναι μια κατάλληλη σταθερά που εξαρτάται από τη συνάρτηση h .

4) Οι ελλειπτικές κατανομές είναι αναλλοίωτες κάτω από γραμμικούς μετασχηματισμούς με την εξής έννοια:

Αν $A \in \mathbf{R}^{d \times m}$, $\mathbf{v} \in \mathbf{R}^m$, τότε:

$$\mathbf{v} + A' \mathbf{X} \sim E_m(\mathbf{v} + A' \boldsymbol{\mu}, A' \Sigma A, g).$$

5) Για οποιοδήποτε υποδιάνυσμα $\mathbf{X}^* = (X_{i_1}, \dots, X_{i_p})$ του $\mathbf{X} = (X_1, X_2, \dots, X_d)$, όπου $1 \leq i_1 \leq \dots \leq i_p \leq d$, το \mathbf{X}^* ακολουθεί την ελλειπτική κατανομή $E_p(\boldsymbol{\mu}_p, \Sigma_p, g)$, όπου $\boldsymbol{\mu}_p = (\mu_{i_1}, \dots, \mu_{i_p})$, δηλαδή το αντίστοιχο υποδιάνυσμα του $\boldsymbol{\mu}$, και $\Sigma_p \in \mathbf{R}^{p \times p}$ είναι το μπλοκ του πίνακα Σ που περιλαμβάνει τα αντίστοιχα στοιχεία $\{\Sigma_{mj}\}_{m,j=i_1, \dots, i_p}$.

Οι ελλειπτικές κατανομές έχουν την ιδιότητα να παραμένει η συνάρτηση πυκνότητάς τους σταθερή πάνω σε ελλείψεις της μορφής:

$$(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c$$

όπου c είναι μια θετική σταθερά. Επίσης, οι ελλειπτικές κατανομές έχουν και την ιδιότητα της ακτινικής συμμετρίας όπως και οι σφαιρικές κατανομές.

Πριν δούμε τι είναι τα Elliptical Copulas θα ορίσουμε κάποιες οικογένειες ελλειπτικών κατανομών τις οποίες θα χρειαστούμε στη συνέχεια.

⁵ Αποδεικνύεται ότι η σταθερά k_d δίνεται από τον τύπο $k_d = \frac{\Gamma(\frac{d}{2})}{2\pi^{\frac{d}{2}} \int_0^\infty t^{d-1} h(t^2) dt}$.

ΟΡΙΣΜΟΣ (Πολυδιάστατη Κανονική Κατανομή): Μία d -διάστατη τυχαία μεταβλητή \mathbf{X} ακολουθεί την πολυδιάστατη Κανονική Κατανομή, $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, αν

$$\mathbf{X} = \boldsymbol{\mu} + \Sigma^{\frac{1}{2}}\mathbf{Z}$$

όπου $\boldsymbol{\mu}$ είναι η παράμετρος θέσης, Σ είναι ο πίνακας διακυμάνσεων συνδιακυμάνσεων και $\mathbf{Z} \sim N_d(\mathbf{0}, I)$ του οποίου όλες οι μονοδιάστατες περιθώριες ακολουθούν την Τυποποιημένη Κανονική Κατανομή. \square

Κάποιες από τις ιδιότητες της πολυμεταβλητής Κανονικής Κατανομής είναι:

1) Η χαρακτηριστική συνάρτηση του \mathbf{X} δίνεται από τον τύπο:

$$\phi_{\mathbf{X}}(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{X}}) = e^{i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}$$

όπου $\mathbf{t} \in \mathbf{R}^d$.

2) Για θετικά ορισμένο πίνακα Σ , η πυκνότητα του \mathbf{X} δίνεται από τον τύπο:

$$f(\mathbf{x}) = \frac{e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}}$$

όπου $\mathbf{x} \in \mathbf{R}^d$.

3) Ένα τυχαίο διάνυσμα με πολυμεταβλητή Κανονική Κατανομή που έχει ασυσχέτιστα στοιχεία έχει και ανεξάρτητα στοιχεία.

4) Η πολυμεταβλητή Κανονική Κατανομή με $\boldsymbol{\mu} = \mathbf{0}$ και $\Sigma = I_d$ ανήκει στην οικογένεια των σφαιρικών κατανομών.

Πράγματι, αν $\mathbf{X} \sim N_d(\mathbf{0}, I_d)$, τότε για κάθε ορθογώνιο $p \times p$ πίνακα P ισχύει

$$P\mathbf{X} \sim N(P\mathbf{0}, P'I_dP) = N(\mathbf{0}, I_d).$$

ΟΡΙΣΜΟΣ(Πολυδιάστατη κατανομή t -Student): Μια d -διάστατη τυχαία μεταβλητή \mathbf{X} ακολουθεί την πολυδιάστατη κατανομή t -Student, $\mathbf{X} \sim t_d(\boldsymbol{\mu}, \Sigma, \nu)$, αν

$$\mathbf{X} = \boldsymbol{\mu} + \Sigma^{\frac{1}{2}}\sqrt{W}\mathbf{Z}$$

όπου Σ είναι ένας θετικά ορισμένος πίνακας, $\mathbf{Z} \sim N_d(\mathbf{0}, I)$ και $W \sim \text{InvG}(\frac{\nu}{2}, \frac{\nu}{2})$.⁶

Κάποιες από τις ιδιότητες της πολυμεταβλητής κατανομής t -Student είναι:

1) Η χαρακτηριστική συνάρτηση του \mathbf{X} είναι:

$$\phi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu}} g(\mathbf{t}'\Sigma\mathbf{t})$$

όπου

$$g(x) = \frac{2\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} \int_0^\infty \cos(s\sqrt{vx})(1+s^2)^{-\frac{\nu+1}{2}} ds$$

2) Για θετικά ορισμένο πίνακα Σ , η πυκνότητα του \mathbf{X} δίνεται από τον τύπο:

$$f(\mathbf{x}) = \frac{\Gamma(\frac{\nu+d}{2})(1 + \frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\nu})^{-\frac{\nu+d}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}}$$

όπου $\mathbf{x} \in \mathbf{R}^d$.

3) Ένα τυχαίο διάνυσμα με πολυμεταβλητή t -Student κατανομή που έχει ασυσχέτιστα στοιχεία δεν συνεπάγεται ότι έχει ανεξάρτητα στοιχεία.

4) Ο πίνακας συνδιακυμάνσεων του \mathbf{X} είναι ο $\frac{\nu}{\nu-2}\Sigma$, για $\nu > 2$. Ο πίνακας συνδιακυμάνσεων είναι ο Σ όταν $\nu \rightarrow \infty$.

5) Όταν $\nu \rightarrow \infty$ η κατανομή $t_d(\boldsymbol{\mu}, \Sigma, \nu)$ συγκλίνει στην κατανομή $N_d(\boldsymbol{\mu}, \Sigma)$. Δηλαδή όσο αυξάνει το ν μικραίνει η πιθανότητα να γίνονται ταυτόχρονα ακραία γεγονότα.

Να σημειώσουμε στο σημείο αυτό ότι η μορφή $E_d(\boldsymbol{\mu}, \Sigma, g)$ δεν είναι μοναδική με την έννοια ότι, αν παίρναμε μέσω κατάλληλου μετασχηματισμού ένα νέο πίνακα συνδιακυμάνσεων, έστω Σ' , και μία νέα συνάρτηση, έστω g' , τότε θα μπορούσαμε να πάρουμε μία κατανομή, έστω $E_d(\boldsymbol{\mu}, \Sigma', g')$, η οποία είναι ίδια με την αρχική.

Τα Elliptical Copulas ορίζονται να είναι τα Copula της αντίστοιχης ελλειπτικής κατανομής. Οι δύο οικογένειες που θα παρουσιάσουμε στην παρούσα εργασία σχετίζονται με τις

⁶ Αν $X \sim G(\alpha, \beta)$ τότε η $Y = \frac{1}{X} \sim \text{InvG}(\alpha, \beta)$. Η συνάρτηση πυκνότητας της Y είναι $f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp(-\frac{\beta}{y}) I_{(0, \infty)}(y)$.

δύο ελλειπτικές κατανομές που δόθηκαν παραπάνω. Με άλλα λόγια, ένα Elliptical Copula ορίζεται ως η από κοινού συνάρτηση κατανομής της αντίστοιχης ελλειπτικής κατανομής όταν τυποποιήσουμε τις περιθώριες. Για τα συγκεκριμένα Copulas θα θεωρούμε ως παράμετρο κλίμακος της ελλειπτικής κατανομής των πίνακα συσχετίσεων και ως παράμετρο θέσης το μηδενικό διάνυσμα, καθώς όπως αποδείξαμε στο Κεφάλαιο 1, τα Copulas παραμένουν αναλοιώτα ως προς γνησίως αύξοντες μετασχηματισμούς των τυχαίων μεταβλητών.

Τα Elliptical Copulas που θα μας απασχολήσουν στην παρούσα εργασία είναι το Gaussian Copula και το t -Copula. Να υπενθυμίσουμε σε αυτό το σημείο ότι, στη δισδιάστατη περίπτωση, αν ένα Copula, C , είναι απόλυτα συνεχές, τότε ορίζεται η πυκνότητά του, c . Μπορούμε, επομένως, να γράψουμε το Copula στην παρακάτω μορφή:

$$C(u, v) = P(U \leq u, V \leq v) = \int_{-\infty}^u \int_{-\infty}^v c(s, t) dt ds = \int_0^u \int_0^v c(s, t) dt ds.$$

3.2.1 Gaussian (Normal) Copula

ΟΡΙΣΜΟΣ (Gaussian Copula): Έστω η d -διάστατη τ.μ $\mathbf{X} \sim N_d(\mathbf{0}, R)$, όπου R είναι ο πίνακας συσχετίσεων. Τότε το Gaussian Copula δίνεται από τον τύπο:

$$C_R(u_1, \dots, u_d) = F_R(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$$

όπου $(u_1, \dots, u_d) \in [0, 1]^d$, η F_R είναι η από κοινού συνάρτηση κατανομής της \mathbf{X} και Φ^{-1} είναι η αντίστροφη της συνάρτησης κατανομής της τυπικής κανονικής κατανομής. \square

Στη δισδιάστατη περίπτωση το Copula στη μορφή με το διπλό ολοκλήρωμα και την πυκνότητα είναι της μορφής:

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} dx dy$$

όπου ρ είναι η παράμετρος του Copula.

Η πυκνότητα⁷ είναι η ποσότητα μέσα στο διπλό ολοκλήρωμα και προκύπτει από την πυ-

⁷ Για τη γενίκευση στις n διαστάσεις παροτρύνουμε το αναγνώστη στην εργασία των Bouyè, Durrleman, Nikeghbali, Riboulet και Roncalli, καθώς και στο βιβλίο των Cherubini, Luciano, Vecchiato [6].

κνότητα της διοδιάστατης Κανονικής Κατανομής δύο τυχαίων μεταβλητών, $X \sim N(\mu_X, \sigma_X)$ και $Y \sim N(\mu_Y, \sigma_Y)$, η οποία είναι:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}Q(x, y)\right], \quad x, y \in \mathbf{R},$$

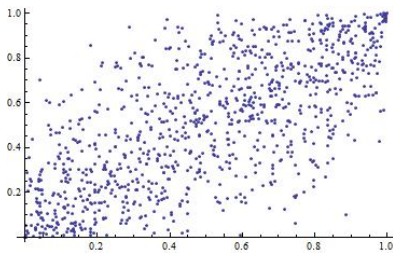
όπου

$$Q(x, y) = \left(\frac{x - \mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x - \mu_X}{\sigma_X}\right)\left(\frac{y - \mu_Y}{\sigma_Y}\right) + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2.$$

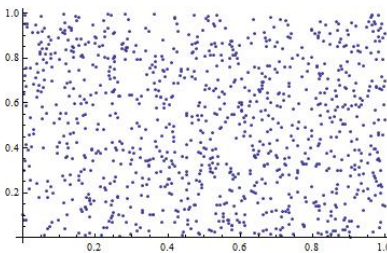
Στην πυκνότητα του Copula έχουμε τυποποιήσει τις τυχαίες μεταβλητές και επομένως $\mu_X = \mu_Y = 0$ και $\sigma_X = \sigma_Y = 1$. Με ρ συμβολίζουμε το συντελεστή συσχέτισης των X, Y και επομένως ο πίνακας συσχέτισης σε αυτήν την περίπτωση είναι:

$$R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

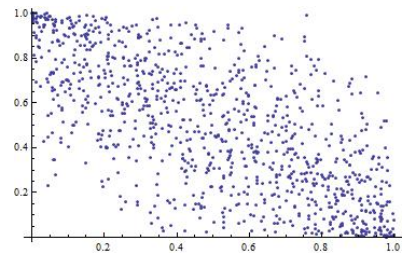
Στη συνέχεια θα ήταν δόκιμο να δούμε και οπτικά κάποια αποτελέσματα. Αρχικά δίνονται τρία διαγράμματα στα οποία έχουμε παραγάγει 1000 σημεία από Gaussian Copula με διαφορετικούς συντελεστές γραμμικής συσχέτισης.



Σχήμα 3.3: 1000 τυχαία σημεία από Gaussian Copula με συντελεστή συσχέτισης $\rho = 0.7$.



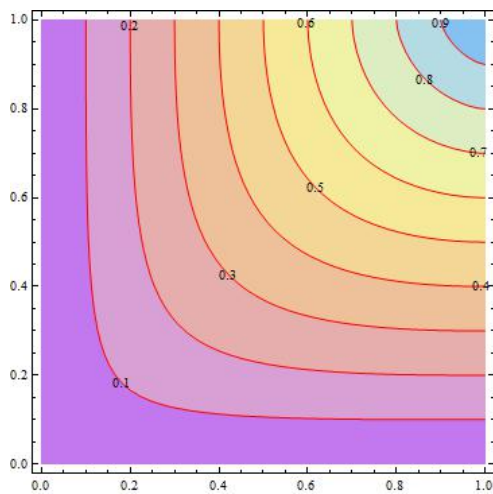
Σχήμα 3.4: 1000 τυχαία σημεία από Gaussian Copula με συντελεστή συσχέτισης $\rho = 0$.



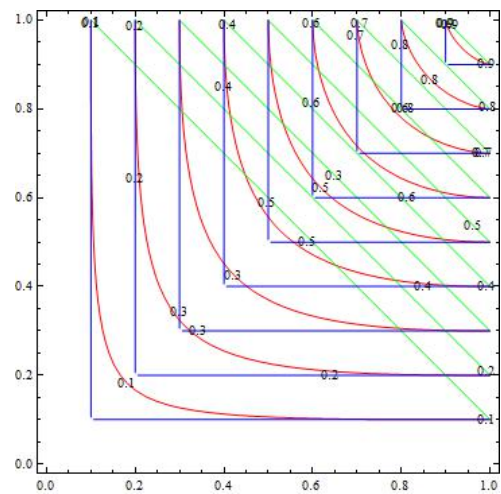
Σχήμα 3.5: 1000 τυχαία σημεία από Gaussian Copula με συντελεστή συσχέτισης $\rho = -0.7$.

Αυτό που παρατηρούμε είναι πως όταν ο συντελεστής μηδενίζεται τα σημεία είναι διάσπαρτα στο γράφημα (Σχήμα 3.4) και επομένως είναι ασυσχέτιστα. Επειδή τα Gaussian Copula κατασκευάζονται βάσει της κανονικής κατανομής είναι και ανεξάρτητα. Στο πρώτο και στο τρίτο γράφημα έχουμε μια σχετικά ισχυρή θετική και αρνητική γραμμική συσχέτιση

αντιστοίχως.⁸



Σχήμα 3.6: Το Contour Plot της συνάρτησης κατανομής του Gaussian Copula με $\rho = 0.7$.



Σχήμα 3.7: Το Contour Plot της σ.κ. του Gaussian Copula μαζί με τα φράγματα.

Στα Σχήματα 3.6 και 3.7 δίνονται: το Contour Plot της συνάρτησης κατανομής ενός Gaussian Copula το οποίο έχει περιθώριες ομοιόμορφες στο $(0,1)$ και συντελεστή γραμμικής συσχέτισης $\rho = 0.7$ (Σχήμα 3.6) καθώς επίσης και το παραπάνω Contour Plot μαζί με τα φράγματα (Σχήμα 3.7). Όπως ήταν αναμενόμενο τα Contour Lines είναι εντός των αντιστοίχων γραμμών του άνω και του κάτω φράγματος, αφού αυτά ισχύουν για κάθε Copula ανεξαρτήτως του σε ποια οικογένεια ανήκουν ή ποιες ιδιότητες έχουν. Σημειώνουμε ότι στο (Σχήμα 3.7) με μπλε γραμμές έχουμε τα Contour Lines του άνω φράγματος ενώ με πράσινο του κάτω φράγματος.

Γενικά αυτό που ενδιαφέρει πολλούς επιστήμονες για τα Copulas είναι η συμπεριφορά τους στις ουρές. Αυτή την πληροφορία θα μας τη δώσουν οι συντελεστές εξάρτησης ουρών, τους οποίους παρουσιάζουμε στη συνέχεια. Συγκεκριμένα, για ένα διδιάστατο Gaussian Copula ισχύει η παρακάτω πρόταση για τους συντελεστές εξάρτησης ουρών:

⁸ Για τα γραφήματα που χρησιμοποιούμε στο συγκεκριμένο κεφάλαιο έχει χρησιμοποιηθεί το mathematica9 το οποίο έχει έτοιμες εντολές για τα Copulas και για αυτόν τον λόγο δε συμπεριλαμβάνονται στα Παραρτήματα.

ΠΡΟΤΑΣΗ: Για ένα διδιάστατο Gaussian Copula οι δυο συντελεστές εξάρτησης ουρών είναι ίσοι με μηδέν αφού

$$\lambda_L = \lambda_U = 2 \lim_{t \rightarrow -\infty} \Phi\left(t \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}\right) = 0.^9$$

□

Αυτό σημαίνει, ότι ανεξάρτητα του πόσο συσχετισμένες είναι οι δύο τυχαίες μεταβλητές, το να πάρει η μία κάποια ακραία τιμή είναι ανεξάρτητο από το αν θα πάρει και η άλλη ακραία τιμή.

Να αναφέρουμε σε αυτό το σημείο ότι η πυκνότητα ενός Elliptical Copula υπολογίζεται ως ο λόγος της από κοινού πυκνότητας της αντίστοιχης πολυδιάστατης ελλειπτικής κατανομής προς το γινόμενο των περιθώριων πυκνοτήτων:

$$c(\mathbf{u}) = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \cdots f_d(F_d^{-1}(u_d))}.$$

Για το Gaussian Copula η πολυδιάστατη ελλειπτική κατανομή στην οποία βασίζεται δεν είναι άλλη από την πολυδιάστατη Κανονική Κατανομή με παράμετρο θέσης το μηδενικό διάνυσμα και παράμετρο κλίμακος τον πίνακα συσχετίσεων. Επομένως η πυκνότητά του υπολογίζεται ως εξής:

$$c(\mathbf{u}) = \frac{\frac{1}{(2\pi)^{\frac{d}{2}} |R|^{\frac{1}{2}}} e^{-\frac{1}{2} \mathbf{x}' R^{-1} \mathbf{x}}}{\prod_{i=1}^d \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2} x_i^2}} = \frac{e^{-\frac{1}{2} \mathbf{x}' R^{-1} \mathbf{x}}}{|R|^{\frac{1}{2}} e^{-\frac{1}{2} \sum_{i=1}^d x_i^2}}$$

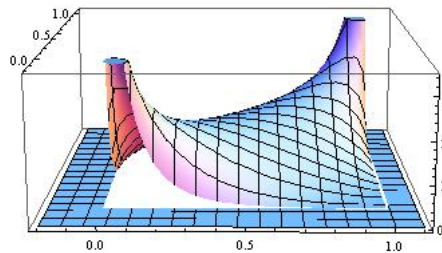
όπου $\mathbf{x} = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$. Παρατηρούμε πως η μόνη παράμετρος του Gaussian Copula είναι ο πίνακας συσχετίσεων R .¹⁰

Αφού δώσαμε τον τύπο που υπολογίζει την πυκνότητα των Elliptical Copulas θα ήταν χρήσιμο να έχουμε και εδώ κάποιο οπτικό αποτέλεσμα. Επομένως, πριν κλείσουμε την υποενότητα με τα Gaussian Copulas θα δώσουμε δύο ακόμα γραφήματα. Το γράφημα που δίνεται παρακάτω είναι η συνάρτηση πυκνότητας ενός Gaussian Copula με περιθώριες ομοιόμορφα κατανεμημένες στο $(0,1)$ και με συντελεστή γραμμικής συσχέτισης $\rho = 0.7$.

⁹ Η απόδειξη βρίσκεται στην εργασία των Embrechts, Lindskog και McNeil [10].

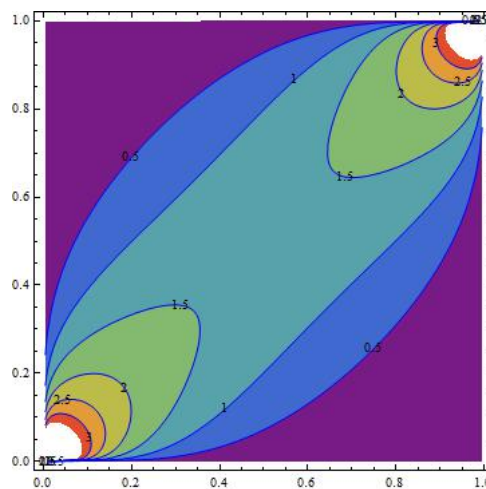
¹⁰ Για τη διδιάστατη περίπτωση δόθηκε προηγουμένως ο τύπος της πυκνότητας του Copula.

Σχήμα 3.8: Η συνάρτηση πυκνότητας ενός Gaussian Copula με περιθώριες ομοιόμορφες στο (0,1).



Στο παραπάνω γράφημα ο άξονας των x αλλά και των y είναι μεταξύ -0.2 και 1.1 για καλύτερα οπτικά αποτελέσματα και, όπως ήταν αναμενόμενο, η πυκνότητα εκεί μηδενίζεται. Δίνουμε και το αντίστοιχο Contour Plot (Σχήμα 3.8) για το οποίο υπενθυμίζουμε ότι όπου υπάρχει ο ίδιος χρωματισμός συνεπάγεται ότι το σχήμα έχει ίδιο ύψος και όταν τα Contour Lines συγκλίνουν προς κάποια περιοχή τότε έχουμε πιθανό μέγιστο ή ελάχιστο (εξαρτάται από την τιμή που έχει η γραμμή). Στο συγκεκριμένο γράφημα το ίδιο χρώμα συνεπάγεται περιοχή με ίδια πυκνότητα και στην άνω δεξιά και κάτω αριστερή γωνία έχουμε μέγιστο ύψος του σχήματος, όπως φαίνεται και στο τρισδιάστατο γράφημα.

Σχήμα 3.9: Contour Plot της συνάρτησης πυκνότητας ενός Gaussian Copula με $\rho = 0.7$.



3.2.2 Student's t -Copula

ΟΡΙΣΜΟΣ (t -Copula): Έστω η d -διάστατη τ.μ $\mathbf{X} \sim t_d(\mathbf{0}, R, \nu)$, όπου R είναι ο πίνακας συσχετίσεων. Τότε το t -Copula δίνεται από τον τύπο:

$$C_{R,\nu}(u_1, \dots, u_d) = t_{\nu,R}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d))$$

όπου $(u_1, \dots, u_d) \in [0, 1]^d$, η $t_{\nu,R}$ είναι η από κοινού συνάρτηση κατανομής της \mathbf{X} και t_{ν}^{-1} είναι η αντίστροφη της συνάρτησης κατανομής της μονοδιάστατης κατανομής t -Student με ν βαθμούς ελευθερίας, (η οποία ως γνωστόν έχει μέση τιμή 0 και διασπορά ίση με $\frac{\nu}{\nu-2}$).

Στη διδιάστατη περίπτωση το Copula στη μορφή με το διπλό ολοκλήρωμα και την πυκνότητα είναι της μορφής:

$$C_{\varrho,\nu}(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{1}{2\pi(1-\varrho^2)^{\frac{1}{2}}} \left\{ 1 + \frac{x^2 - 2\varrho xy + y^2}{\nu(1-\varrho^2)} \right\}^{-\frac{(\nu+2)}{2}} dx dy$$

όπου ϱ, ν , είναι οι παράμετροι του Copula.

Η πυκνότητα είναι η ποσότητα μέσα στο διπλό ολοκλήρωμα και προκύπτει από την πυκνότητα της διδιάστατης κατανομής t -Student, η οποία είναι:

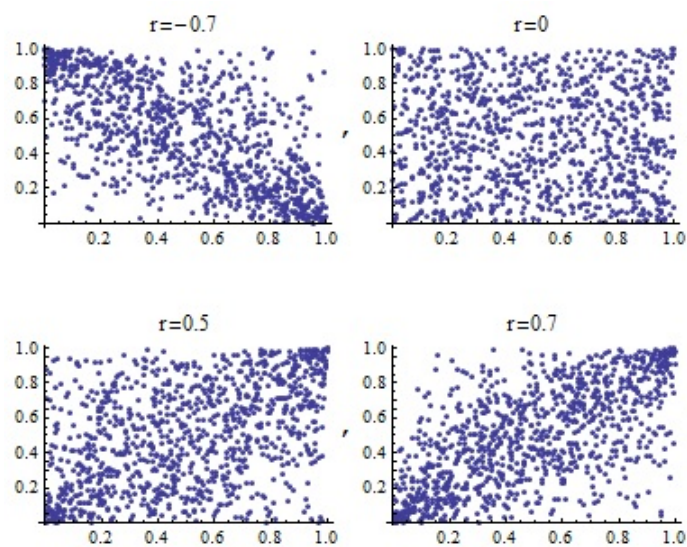
$$f(\mathbf{x}) = \frac{1}{2\pi(1-\varrho^2)^{\frac{1}{2}}} \left[1 + \frac{1}{\nu}(\mathbf{x} - \boldsymbol{\mu})^t R^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right]^{-\frac{(\nu+2)}{2}}$$

όπου

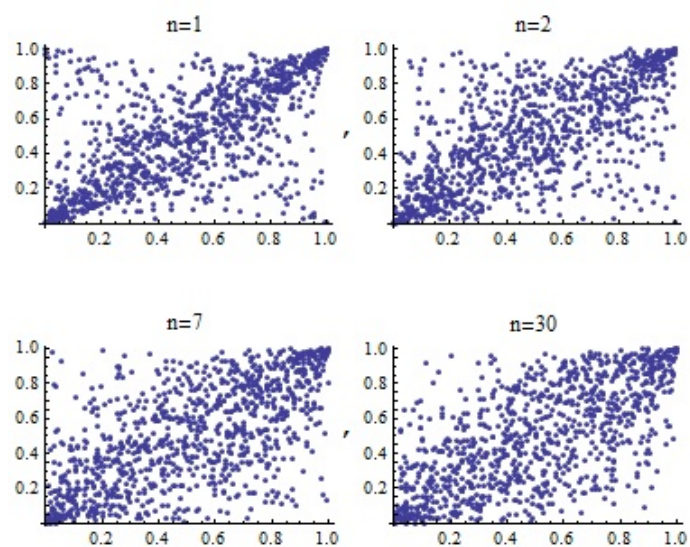
$$R = \begin{bmatrix} 1 & \varrho \\ \varrho & 1 \end{bmatrix}$$

είναι ο πίνακας συσχετίσεων και το διάνυσμα $\boldsymbol{\mu}$ είναι το μηδενικό διάνυσμα για την πυκνότητα του Copula καθώς έχουμε τυποποιήσει τις τυχαίες μεταβλητές.

Σχήμα 3.10: 1000 σημεία από t-Corula για $\nu=7$ και διαφορετικά ρ .



Σχήμα 3.11: 1000 σημεία από t-Corula για $\rho = 0.7$ και διαφορετικά ν .



Στα (Σχήματα 3.10 και 3.11) έχουμε παραγάγει 1000 τυχαία σημεία από το t -Copula με διαφορετικούς συντελεστές γραμμικής συσχέτισης και σταθερούς βαθμούς ελευθερίας $\nu = 7$, αλλά και με σταθερό συντελεστή συσχέτισης $\rho = 0.7$ και διαφορετικούς βαθμούς ελευθερίας.

Αυτό που παρατηρούμε στο (Σχήμα 3.10) είναι πως όταν ο συντελεστής συσχέτισης μηδενίζεται τα δεδομένα είναι ασυσχέτιστα. Η διαφορά με τα Gaussian, όσον αφορά τις αλλαγές του ρ , είναι πως ενώ και εδώ έχουμε πιο ισχυρή συσχέτιση στο κέντρο της κατανομής, έχουμε κάποια παραπάνω συσχέτιση στις ουρές. Στο (Σχήμα 3.11) βλέπουμε ότι όσο μεγαλώνει το ν δεν έχουμε μεγάλες διαφοροποιήσεις στις ουρές. Παρ' όλα αυτά θα πρέπει να λαμβάνουμε υπ' όψιν και τις δύο παραμέτρους πριν βγάλουμε κάποιο συμπέρασμα. Την πληροφορία για το τι γίνεται στις ουρές θα μας τη δώσουν οι συντελεστές εξάρτησης ουρών για τους οποίους ισχύει η παρακάτω πρόταση:

ΠΡΟΤΑΣΗ: Για ένα διδιάστατο t -Copula και οι δύο συντελεστές εξάρτησης ουρών είναι ίσοι με:

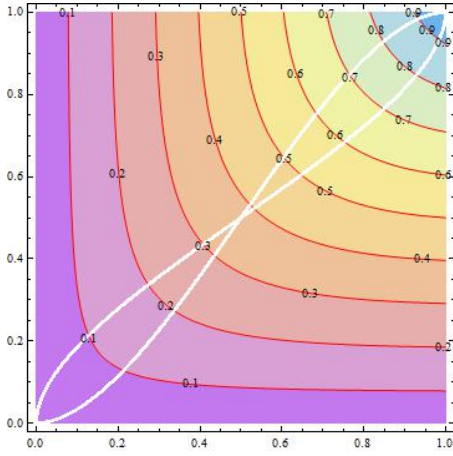
$$\lambda_L = \lambda_U = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$$

όπου η $t_{\nu+1}$ είναι η συνάρτηση κατανομής μίας μονοδιάστατης κατανομής t -Student με $\nu + 1$ βαθμούς ελευθερίας και ρ είναι ο συντελεστής γραμμικής συσχέτισης του Pearson. \square

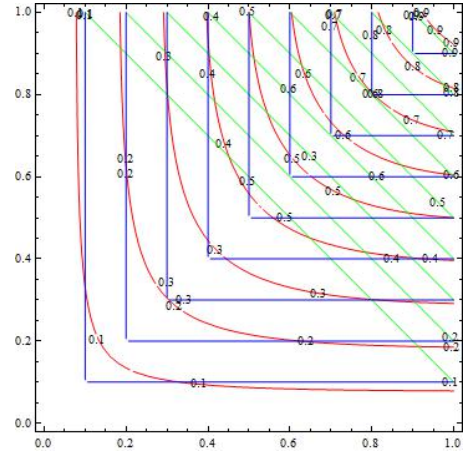
Όσο μεγαλώνει το ρ και μικραίνει το ν τόσο μεγαλώνει και η εξάρτηση στις ουρές. Με απλά λόγια, το t -Copula επιτρέπει να αυξάνεται η πιθανότητα για την ύπαρξη από κοινού ακραίων γεγονότων, εν αντιθέσει με το Gaussian Copula.

Παρακάτω δίνουμε το Contour Plot της συνάρτησης κατανομής ενός t -Copula με $\nu = 7$ και $\rho = 0.7$ (Σχήμα 3.12) καθώς και το Contour plot μαζί με τα φράγματα (Σχήμα 3.13). Οι περιθώριες του παρακάτω Copula είναι ομοιόμορφες στο $(0,1)$ και όπως φαίνεται στο Contour Plot (Σχήμα 3.12) η συνάρτηση κατανομής δεν ορίζεται παντού. Παρατηρούμε ότι τα Contour Lines στο (Σχήμα 3.13) ξεφεύγουν των φραγμάτων κάτι το οποίο οφείλεται μάλλον στον τρόπο υπολογισμού με το mathematica, αφού χρειάζεται να υπολογίσει τα ζευγάρια (u, v) που ικανοποιούν τη σχέση:

$$C_{\rho, \nu}(u, v) = \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \left\{ 1 + \frac{x^2 - 2\rho xy + y^2}{\nu(1-\rho^2)} \right\}^{-\frac{(\nu+2)}{2}} dx dy = c.^{11}$$



Σχήμα 3.12: Το Contour Plot της συνάρτησης κατανομής του t -Copula με $\rho = 0.7$ και $\nu = 7$.



Σχήμα 3.13: Το Contour Plot της συνάρτησης κατανομής του t -Copula μαζί με τα φράγματα.

Συνεχίζοντας τη μελέτη για το t -Copula, είναι προφανές ότι η πολυδιάστατη ελλειπτική κατανομή στην οποία βασίζεται είναι η πολυδιάστατη κατανομή t -Student με παράμετρο θέσης το μηδενικό διάνυσμα και παράμετρο κλίμακος τον πίνακα συσχετίσεων. Επομένως, η πυκνότητα του t -Copula αποδεικνύεται ότι δίνεται από τον τύπο:

$$c(\mathbf{u}) = \frac{\Gamma(\frac{\nu+d}{2})\Gamma(\frac{\nu}{2})^{d-1} (1 + \frac{\mathbf{x}'R^{-1}\mathbf{x}}{\nu})^{-\frac{(\nu+d)}{2}}}{|R|^{\frac{1}{2}} \Gamma(\frac{\nu+1}{2})^d \prod_{i=1}^d (1 + \frac{x_i^2}{\nu})^{-\frac{\nu+1}{2}}} \quad 12$$

όπου $\mathbf{x} = (t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d))$.

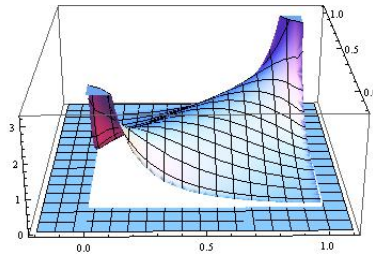
Παρακάτω δίνουμε το γράφημα της συνάρτησης πυκνότητας του t -Copula με περιθώριες ομοιόμορφες στο $(0,1)$, συντελεστή γραμμικής συσχέτισης $\rho = 0.7$ και $\nu = 7$ βαθμούς ελευθερίας καθώς και το αντίστοιχο Contour plot. Να σημειώσουμε ότι για τη συνάρτηση

¹¹ Σημειώνουμε ότι το t -Copula δόθηκε στο mathematica με έτοιμη συνάρτηση *CopulaDistribution*[{"MultivariateT", {{1, 7/10}, {7/10, 1}}, 7], {UniformDistribution[{0, 1}], UniformDistribution[{0, 1}]}

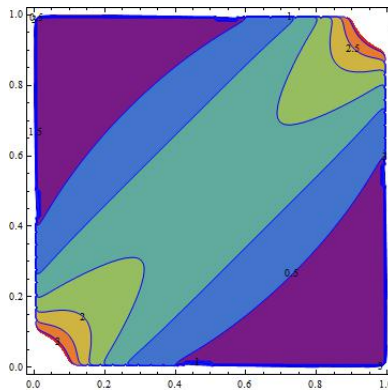
¹² Για τη γενίκευση στις n διαστάσεις παροτρύνουμε το αναγνώστη στην εργασία των Bouyè, Durrleman, Nikeghbali, Riboulet και Roncalli, καθώς και στο βιβλίο των Cherubini, Luciano, Vecchiato [6].

πυκνότητας έχουμε επιλέξει $x, y \in (-0.2, 1.1)$ για καλύτερα οπτικά αποτελέσματα και, όπως αναφέραμε νωρίτερα, το ίδιο χρώμα στο Contour Plot σημαίνει πως έχουμε ίδια πυκνότητα. Έτσι από αυτά τα γραφήματα είναι εμφανές και στα δυο Elliptical Copula ότι υπάρχει συμμετρία.

Σχήμα 3.14: Συνάρτηση πυκνότητας ενός t -Copula για $\rho = 0.7$ και $\nu = 7$.



Σχήμα 3.15: Contour Plot της συνάρτησης πυκνότητας ενός t -Copula για $\rho = 0.7$ και $\nu=7$.



Υπενθυμίζουμε ότι ακτινικά συμμετρικά Copulas σημαίνει ότι $C_R = \hat{C}_R$. Αυτό ισχύει για όλα τα Copulas που ανήκουν στην κλάση των Elliptical. Η παρακάτω πρόταση (βλπ. Nelsen [28]), όμως, που αφορά τους συντελεστές συσχέτισης ρ_S του Spearman και τ του Kendall δεν ισχύει για όλα τα Elliptical Copulas. Ισχύει, όμως, για τις δύο οικογένειες που μελετάμε στην παρούσα εργασία.

ΠΡΟΤΑΣΗ 1. (Σχέση μεταξύ του ρ του Pearson και του ρ_S του Spearman): Η σχέση που συνδέει το ρ του Pearson με το ρ_S του Spearman για το Gaussian Copula και το t -Copula είναι:

$$\rho = 2 \sin\left(\frac{\pi}{6}\rho_S\right).$$

2.(Σχέση μεταξύ του ρ του Pearson και του τ του Kendall): Η σχέση που συνδέει το ρ του Pearson με το τ του Kendall για το Gaussian Copula και το t -Copula είναι:

$$\rho = \sin\left(\frac{\pi}{2}\rho_\tau\right).$$

Επομένως, αν από τα δεδομένα μας είναι πιο εύκολο να εκτιμήσουμε τον έναν συντελεστή τότε μπορούμε να έχουμε και την εκτίμηση του άλλου. Απλά να τονίσουμε ότι ο πίνακας που θα προκύψει από τους παραπάνω μετασχηματισμούς θα έχει ως στοιχεία συσχετίσεις, ωστόσο είναι πιθανό να μην είναι θετικά ορισμένος. Σε αυτήν την περίπτωση, θα πρέπει να χρησιμοποιηθεί κατάλληλος αλγόριθμος ώστε να τροποποιηθεί κατάλληλα ο πίνακας και να πληροί και αυτήν την ιδιότητα.

3.3 Archimedean (Αρχιμήδεια) Copulas

Τα Archimedean Copulas¹³ εν αντιθέσει με τα Elliptical Copulas δεν απαιτούν την ακτινική συμμετρία, αν και υπάρχει μέλος της κλάσης αυτής που έχει αυτή την ιδιότητα. Επίσης, έχουν συνήθως κλειστό τύπο και δεν στηρίζονται σε κάποια από τις γνωστές πολυμεταβλητές κατανομές, όπως τα Elliptical. Στην παρούσα εργασία θα ασχοληθούμε με μονοπαραμετρικές οικογένειες αυτής της κλάσης κι επομένως όταν χρησιμοποιείται η παράμετρος θ αυτή θα είναι μονοδιάστατη και θα ανήκει σε κάποιο παραμετρικό χώρο Θ .

Τα Archimedean Copulas χρησιμοποιούνται ευρέως και αυτό οφείλεται κυρίως σε τρεις λόγους:

- α) Υπάρχουν πολλές οικογένειες που ανήκουν στην κλάση τους,
- β) οι οικογένειες αυτές έχουν κάποιες καλές ιδιότητες, όπως, για παράδειγμα, την εξάρτηση στις ουρές και
- γ) κατασκευάζονται εύκολα.

Η κατασκευή τους βασίζεται σε μία συνάρτηση την οποία θα την συμβολίζουμε με φ και θα την ονομάζουμε χαρακτηριστικό γεννήτορα του Copula. Στην ουσία η χρησιμότητά

¹³ Για τα Archimedean Copulas βλπ. τέταρτο κεφάλαιο του βιβλίου του Nelsen [28] καθώς και την εργασία των Embrechts,Lindskog,McNeil [10].

της φαίνεται από το γεγονός ότι, αντί να εξετάζουμε κάθε φορά το Archimedean Copula το οποίο είναι συνάρτηση πολλών μεταβλητών, θα εξετάζουμε τη συνάρτηση φ που είναι συνάρτηση μίας μεταβλητής. Στη συνέχεια δίνουμε τη σχέση της φ και των Archimedean Copulas αφού πρώτα ορίσουμε την αντίστροφη της φ .

ΟΡΙΣΜΟΣ(Ψευδοαντίστροφη): Έστω φ μια συνεχής, γνησίως φθίνουσα συνάρτηση, $\varphi : I \rightarrow [0, \infty]$, ώστε $\varphi(1) = 0$. Η ψευδοαντίστροφη της φ , $\varphi^{[-1]} : [0, \infty] \rightarrow I$, ορίζεται ως εξής:

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & \text{αν } 0 \leq t \leq \varphi(0) \\ 0, & \text{αν } \varphi(0) \leq t \leq \infty \end{cases}$$

Η $\varphi^{[-1]}$ είναι συνεχής, φθίνουσα στο $[0, \infty]$ και γνησίως φθίνουσα στο $[0, \varphi(0)]$. □

Για την ψευδοαντίστροφη της φ ισχύουν τα εξής:

1) $\varphi^{[-1]}(\varphi(u)) = u$, όπου $u \in I$.

2)

$$\varphi(\varphi^{[-1]}(t)) = \begin{cases} t, & \text{αν } 0 \leq t \leq \varphi(0) \\ \varphi(0), & \text{αν } \varphi(0) \leq t \leq \infty \end{cases} = \min(t, \varphi(0))$$

3) Αν $\varphi(0) = \infty$ τότε η ψευδοαντίστροφη της φ είναι η κλασική αντίστροφη.

Αποδεικνύεται το παρακάτω Θεώρημα το οποίο και θα μας βοηθήσει να ορίσουμε τα Archimedean Copulas. Μία απόδειξη του θεωρήματος η οποία δόθηκε από τους Alsina, Frank και Schweizer μπορεί να βρεθεί στο βιβλίο του Nelsen. [28]

ΘΕΩΡΗΜΑ: Έστω φ μία συνεχής, γνησίως φθίνουσα συνάρτηση, $\varphi : I \rightarrow [0, \infty]$, ώστε $\varphi(1) = 0$ και $\varphi^{[-1]}$ η ψευδο-αντίστροφή της. Τότε, η συνάρτηση $C : I^2 \rightarrow I$ με τύπο:

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$$

είναι Copula αν και μόνο αν η φ είναι κυρτή. □

Δηλαδή χρησιμοποιώντας τη συνάρτηση φ , η οποία πληροί τις παραπάνω ιδιότητες, κατασκευάσαμε ένα Copula. Τα Copulas τα οποία δημιουργούνται από τέτοιου είδους συναρτήσεις ανήκουν στην κλάση των Archimedean Copulas.

ΟΡΙΣΜΟΣ (*Archimedean Copulas*): Τα Copulas της μορφής:

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v))$$

ονομάζονται Archimedean Copulas και η συνάρτηση φ είναι ο χαρακτηριστικός γεννήτορας τους. Αν $\varphi(0) = \infty$, τότε έχουμε Archimedean Copula με αυστηρό χαρακτηριστικό γεννήτορα φ . □

Παράδειγμα¹⁴: Τα Archimedean Copulas τα οποία παράγονται από τους χαρακτηριστικούς γεννήτορες $\varphi(t) = 1 - t$ και $\varphi(t) = -\log t$ για κάθε $t \in [0, 1]$ είναι το κάτω φράγμα και το Copula ανεξαρτησίας αντίστοιχα.

Πράγματι, έχουμε:

1) Για $\varphi(t) = 1 - t$, η ψευδοαντίστροφη είναι:

$$\varphi^{[-1]}(t) = \begin{cases} 1 - t, & \text{αν } 0 \leq t \leq 1 \\ 0, & \text{αν } t > 1 \end{cases} = \max(1 - t, 0)$$

και επομένως το Copula που παράγεται είναι:

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \max(1 - (1 - u) - (1 - v), 0) = \max(u + v - 1, 0) = W(u, v).$$

2) Για $\varphi(t) = -\log t$, η ψευδοαντίστροφη είναι:

$$\varphi^{[-1]}(t) = \varphi^{-1}(t) = \exp(-t).$$

Αυτό ισχύει γιατί $\varphi(0) = \infty$ και άρα ο χαρακτηριστικός γεννήτορας είναι αυστηρός και το Copula που παράγεται από αυτόν είναι:

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \exp(-(-\log u - \log v)) = uv = \Pi(u, v).$$

Μία εύλογη απορία θα ήταν, ποιος είναι ο λόγος που ονομάστηκαν Archimedean τα συγκεκριμένα Copulas. Σύμφωνα με την Αρχιμήδεια Ιδιότητα για τους πραγματικούς αριθμούς έχουμε ότι για $a, b \in \mathbf{R}$ όπου $a, b > 0$ υπάρχει $n \in \mathbf{Z}$ ώστε $na > b$. Κάτι αντίστοιχο

¹⁴ Το συγκεκριμένο παράδειγμα συμπεριλαμβάνεται στο τέταρτο κεφάλαιο του βιβλίου του Nelsen [28].

ισχύει και για τα συγκεκριμένα Copulas. Για να το αποδείξουμε θα χρειαστούμε τον παρακάτω ορισμό:

ΟΡΙΣΜΟΣ: Έστω $u, v \in (0, 1)$ και έστω n ένας θετικός ακέραιος αριθμός. Θα ονομάζουμε C -δύναμη του u και θα τη συμβολίζουμε με u_C^n τη συνάρτηση του u η οποία ορίζεται από τον αναδρομικό τύπο

$$u_C^{n+1} = C(u, u_C^n) \text{ με αρχική συνθήκη } u_C^1 = u.$$

ΘΕΩΡΗΜΑ: Έστω C ένα Archimedean Copula με χαρακτηριστικό γεννήτορα φ . Για $u, v \in (0, 1)$ υπάρχει θετικός ακέραιος n , ώστε $u_C^n < v$.

Απόδειξη: Έστω $u, v \in (0, 1)$. Τότε η n -οστή δύναμη του C είναι ίση με $\varphi^{[-1]}(n\varphi(u))$. Πράγματι, για $n = 2$, έχουμε ότι

$$u_C^2 = C(u, u_C^1) = C(u, u) = \varphi^{[-1]}(2\varphi(u))$$

Έστω ότι ισχύει για $n = k$. Θα αποδείξουμε ότι ισχύει για $n = k + 1$.

$$u_C^{k+1} = C(u, u_C^k) = C(u, \varphi^{[-1]}(k\varphi(u))) = \varphi^{[-1]}(\varphi(u) + k\varphi(u)).$$

Αφού τα $\varphi(u), \varphi(v)$ είναι θετικοί πραγματικοί αριθμοί σύμφωνα με την Αρχιμήδεια ιδιότητα υπάρχει ακέραιος n , ώστε $n\varphi(u) > \varphi(v)$. Όμως, έχουμε $v > 0, \varphi(v) < \varphi(0)$. Επομένως ισχύει:

$$u_C^n = \varphi^{[-1]}(n\varphi(u)) < \varphi^{[-1]}(\varphi(v)) = v$$

(Δεν χρειάζεται στη συγκεκριμένη απόδειξη η υπόθεση ότι η φ είναι κυρτή.) □

ΘΕΩΡΗΜΑ: Έστω C ένα Archimedean Copula με χαρακτηριστικό γεννήτορα φ . Τότε ισχύουν τα παρακάτω:

- 1) $C(u, v) = C(v, u)$ για κάθε $u, v \in [0, 1]$.
- 2) $C(C(u, v), w) = C(u, C(v, w))$ για κάθε $u, v, w \in [0, 1]$.
- 3) Αν c είναι μια θετική σταθερά τότε το $c\varphi$ είναι και αυτό χαρακτηριστικός γεννήτορας του C .

Απόδειξη:

Για την (1) έχουμε ότι:

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)) = \varphi^{[-1]}(\varphi(v) + \varphi(u)) = C(v, u).$$

Για την (2) έχουμε:

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

Τέλος για την (3) έχουμε:

$$c\varphi(C(u, v)) = c(\varphi(u) + \varphi(v)) = c\varphi(u) + c\varphi(v).$$

□

Τα Archimedean Copulas δεν είναι όλα απόλυτα συνεχή. Υπάρχουν κάποια τα οποία έχουν ιδιάζοντα στοιχεία (singular components) με την έννοια ότι η πυκνότητά τους ή μηδενίζεται ή δεν ορίζεται. Πριν δούμε, όμως, πότε συμβαίνει αυτό θα πρέπει να ορίσουμε την πυκνότητα ενός Archimedean Copula. Θα επιλέξουμε να γράψουμε το Copula σαν κατανομή για να δώσουμε και τη φυσική του ερμηνεία (βλπ. Genest και MacKay [17]).

Για το υπόλοιπο της εργασίας θα συμβολίζουμε με Φ το σύνολο που περιλαμβάνει όλες τις συναρτήσεις της μορφής $\varphi : [0, 1] \rightarrow [0, \infty]$, για τις οποίες ισχύουν τα εξής:

$$\varphi(1) = 0, \quad \varphi'(t) < 0, \quad \varphi''(t) > 0, \quad \text{για κάθε } t \in (0, 1).^{15}$$

ΠΡΟΤΑΣΗ: Η πυκνότητα της διδιάστατης συνάρτησης κατανομής δύο συνεχών τυχαίων μεταβλητών X, Y με από κοινού συνάρτηση κατανομής H , η οποία γεννάται από την παραπάνω συνάρτηση $\varphi \in \Phi$ είναι της μορφής:

$$h(x, y) = -\frac{\varphi''(H(x, y))\varphi'(x)\varphi'(y)}{[\varphi'(H(x, y))]^3}$$

και είναι γνήσια θετική για όλα τα (x, y) για τα οποία ισχύει $\varphi(x) + \varphi(y) < \varphi(0)$.

¹⁵ Οι παραπάνω ιδιότητες εξασφαλίζουν ότι υπάρχει η αντίστροφη της φ , καθώς είναι συνεχής και γνήσιως φθίνουσα, καθώς και ότι υπάρχουν η πρώτη κι η δεύτερη παράγωγος της αντίστροφης, γιατί η αντίστροφη θα είναι μια φθίνουσα, κοίλη συνάρτηση.

Απόδειξη: Έστω (X, Y) ένα τυχαίο διάνυσμα με συνάρτηση κατανομής:

$$H(x, y) = \begin{cases} \varphi^{[-1]}(\varphi(x) + \varphi(y)), & \text{αν } \varphi(x) + \varphi(y) \leq \varphi(0) \\ 0, & \text{οπουδήποτε αλλού} \end{cases}$$

Αν $\varphi(0) = \infty$, τότε $\varphi^{[-1]} = \varphi^{-1}$ και η $H(x, y)$ είναι γνήσια θετική, εκτός από τα σημεία που έχουν είτε $x = 0$, είτε $y = 0$. Παραγωγίζοντας την $\varphi(x) + \varphi(y) = \varphi(H(x, y))$ ως προς x παίρνουμε

$$\varphi'(x) = \varphi'(H(x, y)) \frac{\partial H(x, y)}{\partial x}.$$

Ξαναπαραγωγίζοντας ως προς y προκύπτει

$$0 = \varphi''(H(x, y)) \frac{\partial H(x, y)}{\partial x} \frac{\partial H(x, y)}{\partial y} + \varphi'(H(x, y)) \frac{\partial^2 H(x, y)}{\partial x \partial y}$$

Λύνουμε ως προς $\frac{\partial^2 H(x, y)}{\partial x \partial y}$ και αντικαθιστούμε τα $\frac{\partial H(x, y)}{\partial x}$, $\frac{\partial H(x, y)}{\partial y}$ με τα $\frac{\varphi'(x)}{\varphi'(H(x, y))}$ και $\frac{\varphi'(y)}{\varphi'(H(x, y))}$ αντίστοιχα και προκύπτει το ζητούμενο. \square

Αφού ορίσαμε την πυκνότητα, θα δούμε πότε έχουμε ιδιαίζον στοιχείο και αν έχουμε, πόση μάζα πιθανότητας είναι συγκεντρωμένη εκεί.

ΘΕΩΡΗΜΑ: Η κατανομή $H(x, y)$ η οποία παράγεται από τη συνάρτηση φ έχει ιδιαίζον στοιχείο αν και μόνο αν $\frac{\varphi(0)}{\varphi'(0)} \neq 0$. Αν έχει ιδιαίζον στοιχείο τότε αυτό θα βρίσκεται στην καμπύλη $\varphi(x) + \varphi(y) = \varphi(0)$ με πιθανότητα $-\frac{\varphi(0)}{\varphi'(0)}$.

Απόδειξη: Θα ολοκληρώσουμε την πυκνότητα $h(x, y)$, όπου εκείνη ορίζεται, δηλαδή για τα x, y για τα οποία ισχύει $\varphi(x) + \varphi(y) < 0$. Κάνουμε την αλλαγή των μεταβλητών x, y σε u, v όπου:

$$\begin{cases} u = H(x, y) = \varphi^{-1}(\varphi(x) + \varphi(y)) \\ v = x \end{cases} \Rightarrow \begin{cases} x = v \\ y = \varphi^{-1}(\varphi(u) - \varphi(v)) \end{cases}$$

Η Ιακωβιανή ορίζουσα είναι ίση με:

$$J = \begin{vmatrix} 0 & 1 \\ \varphi'(u) & -\varphi'(v) \\ \varphi'(y) & \varphi'(y) \end{vmatrix} = -\frac{\varphi'(H)}{\varphi'(y)}.$$

Επομένως, έχουμε:

$$h(x, y)|J| = -\frac{\varphi''(H)\varphi'(x)}{[\varphi'(H)]^2} = -\frac{\varphi''(u)\varphi'(v)}{[\varphi'(u)]^2}.$$

Η μάζα πιθανότητας που εξηγείται από την πυκνότητα είναι:

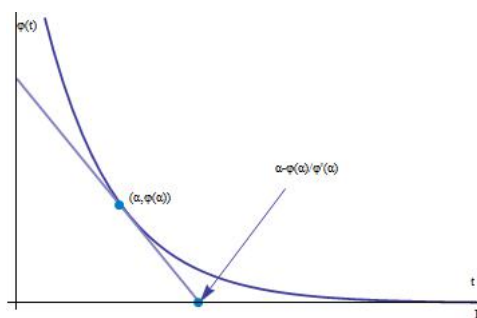
$$p = \int \int_{\varphi(x)+\varphi(y)<\varphi(0)} h(x, y)dxdy = \int \int_{0<u<v<1} -\frac{\varphi''(u)\varphi'(v)}{[\varphi'(u)]^2} dudv = -\int_0^1 \frac{\varphi''(u)}{[\varphi'(u)]^2} \varphi(v)|_u^1 du = \int_0^1 \frac{\varphi''(u)}{[\varphi'(u)]^2} \varphi(u) du = \int_0^1 \left(-\frac{1}{\varphi'(u)}\right)' \varphi(u) du = \left. \frac{-\varphi(u)}{\varphi'(u)} \right|_0^1 + \int_0^1 \frac{\varphi'(u)}{\varphi'(u)} du = 1 - \frac{\varphi(u)}{\varphi'(u)} \Big|_0^1$$

Στη γραφική παράσταση της συνάρτησης φ , η ποσότητα $a - \frac{\varphi(a)}{\varphi'(a)}$ είναι το σημείο τομής της εφαπτομένης της γραφικής παράστασης στο σημείο a με τον οριζόντιο άξονα. Καθώς $a \rightarrow 1$, η ποσότητα $-\frac{\varphi(a)}{\varphi'(a)} \rightarrow 0$. Επομένως η παραπάνω πιθανότητα είναι ίση με:

$$p = 1 + \frac{\varphi(0)}{\varphi'(0)}.$$

Αν είχαμε μια απόλυτα συνεχή συνάρτηση, τότε το παραπάνω διπλό ολοκλήρωμα θα έπρεπε να ήταν ίσο με τη μονάδα και η πυκνότητα θα εξηγούσε όλη τη μάζα πιθανότητας. Για να έχουμε, λοιπόν, ιδιάζον στοιχείο πρέπει $p \neq 1$, δηλαδή $\frac{\varphi(0)}{\varphi'(0)} \neq 0$. Αυτό θα συμβαίνει με πιθανότητα $-\frac{\varphi(0)}{\varphi'(0)}$ και μόνο στην καμπύλη $\varphi(x)+\varphi(y) = \varphi(0)$, καθώς στη συγκεκριμένη καμπύλη δεν ορίζονται οι μερικές παράγωγοι. □

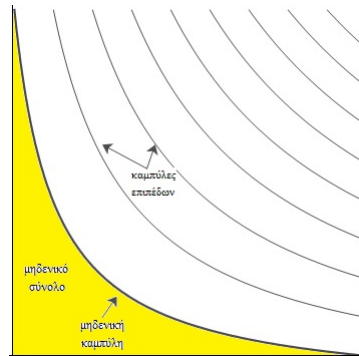
Σχήμα 3.16: Γραφική παράσταση ενός χαρακτηριστικού γεννήτορα μαζί με την εφαπτομένη στο σημείο $(\alpha, \varphi(\alpha))$.



Τα σύνολα επιπέδων (level sets) ενός Copula είναι τα $\{(u, v) \in \mathbf{I}^2 | C(u, v) = t\}$. Για ένα Archimedean Copula, όπου $t > 0$, αυτό το σύνολο αποτελείται από τα σημεία που είναι πάνω στις καμπύλες επιπέδων (level curves) της μορφής $\varphi(u) + \varphi(v) = \varphi(t)$. Εκεί όπου μηδενίζεται το Copula, δηλαδή για $t = 0$, θεωρούμε ότι έχουμε το μηδενικό σύνολο του C ,

(zero set), δηλαδή η κίτρινη περιοχή στο (Σχήμα 3.17). Η οριακή καμπύλη $\varphi(u) + \varphi(v) = \varphi(0)$ ονομάζεται μηδενική καμπύλη (zero curve) του C και, όπως αναφέραμε και πριν, αν υπάρχει κάποιο ιδιάζον στοιχείο τότε σε αυτή την καμπύλη θα βρίσκεται η μάζα πιθανότητας που δεν έχει εξηγηθεί από την πυκνότητα. Για τα Archimedean Copulas αποδεικνύεται ότι οι καμπύλες επιπέδων είναι κυρτές.

Σχήμα 3.17: Καμπύλες επιπέδων και μηδενική καμπύλη ενός Copula.



ΘΕΩΡΗΜΑ: Έστω U, V δύο τυχαίες μεταβλητές που κατανομούνται ομοιόμορφα στο $(0,1)$ με από κοινού συνάρτηση κατανομής ένα Archimedean Copula C το οποίο παράγεται από τη φ . Έστω

$$S = \frac{\varphi(U)}{\varphi(U) + \varphi(V)} \text{ και } T = C(U, V).$$

Τότε ισχύουν τα εξής:

- 1) Η S κατανέμεται ομοιόμορφα στο $(0, 1)$.
- 2) Η T έχει συνάρτηση κατανομής την $K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}$,¹⁶ για $t \in (0, 1)$.
- 3) Οι S, T είναι ανεξάρτητες τυχαίες μεταβλητές.

Το παραπάνω θεώρημα¹⁷ θα μας είναι αρκετά χρήσιμο αργότερα για τη μελέτη των Archimedean Copulas καθώς μας δίνει τη μορφή που έχει η συνάρτηση κατανομής των

¹⁶ Θεωρούμε ότι η παράγωγος της φ είναι συνεχής αλλιώς παίρνουμε το όριο από τα δεξιά.

¹⁷ Οι αποδείξεις των 1,2,3 του θεωρήματος βρίσκονται στα παραρτήματα της εργασίας των Genest και Rivest [20].

Copulas της κλάσης αυτής, (εδώ η τυχαία μεταβλητή T). Επίσης, βάσει αυτού του θεωρήματος μπορούμε να παράγουμε δύο ανεξάρτητες τυχαίες μεταβλητές βάσει δύο άλλων που είναι ομοιόμορφες στο $(0,1)$ και έχουν από κοινού συνάρτηση κατανομής το Archimedean Copula.

Πριν αναπτύξουμε κάποιες από τις πλέον γνωστές και χρήσιμες οικογένειες που ανήκουν στην κλάση των Archimedean Copulas, ας δούμε τι γίνεται με τη διάταξη συμφωνίας και πότε είμαστε σε κάποια οριακή κατάσταση, με την έννοια είτε ότι επιτυγχάνεται κάποιο από τα όρια των Frèchet-Hoeffding, είτε ότι επιτυγχάνεται η ανεξαρτησία. Σημειώνουμε και πάλι ότι ασχολούμαστε με μονοπαραμετρικές οικογένειες των Archimedean Copulas. (βλπ. τέταρτο κεφάλαιο στο βιβλίο του Nelsen [28]).

Για τη διάταξη συμφωνίας δύο Archimedean Copulas θα χρειαστούμε τα παρακάτω:

ΟΡΙΣΜΟΣ: Μία συνάρτηση f ορισμένη στο $[0, \infty]$ θα ονομάζεται υποπροσθετική (sub-additive) αν για κάθε $x, y \in [0, \infty]$ ισχύει:

$$f(x + y) \leq f(x) + f(y).$$

□

Στη συνέχεια θα χρησιμοποιήσουμε το ότι αν μια συνάρτηση f ορισμένη στο $[0, \infty]$ είναι κοίλη και $f(0) = 0$ τότε είναι υποπροσθετική.

Ας δούμε τι γίνεται με τη διάταξη συμφωνίας μεταξύ δύο Archimedean Copulas. Και σε αυτήν την περίπτωση θα ασχοληθούμε με τους χαρακτηριστικούς γεννήτορες των Copulas αν και μπορεί να χρησιμοποιηθεί και ο ορισμός που δόθηκε στο δεύτερο κεφάλαιο.

ΘΕΩΡΗΜΑ: Έστω C_1, C_2 δύο Archimedean Copulas με χαρακτηριστικούς γεννήτορες τις συναρτήσεις $\varphi_1, \varphi_2 \in \Phi$. Τότε $C_1 < C_2$ αν και μόνο αν η σύνθεση συναρτήσεων $\varphi_1 \circ \varphi_2^{[-1]}$ είναι υποπροσθετική συνάρτηση. □

Ο ορισμός της διάταξης συμφωνίας, αλλά και το παραπάνω θεώρημα είναι καμιά φορά δύσκολο να εφαρμοστούν. Υπάρχουν, όμως, κάποιες άλλες προτάσεις για να δείξουμε πιο εύκολα τη διάταξη των Archimedean Copulas. Συγκεκριμένα, αποδεικνύεται πως υπό τις υποθέσεις του παραπάνω θεωρήματος ισχύουν τα εξής:

- 1) Αν η σύνθεση συναρτήσεων $\varphi_1 \circ \varphi_2^{[-1]}$ είναι κοίλη συνάρτηση τότε $C_1 < C_2$.
- 2) Αν η συνάρτηση $\frac{\varphi_1}{\varphi_2}$ είναι αύξουσα στο $(0, 1)$, τότε $C_1 < C_2$.
- 3) Αν οι φ_1, φ_2 είναι συνεχώς παραγωγίσιμες στο $(0, 1)$ και η $\frac{\varphi_1'}{\varphi_2'}$ είναι αύξουσα στο $(0, 1)$, τότε $C_1 < C_2$.

Ας δούμε, όμως, τι γίνεται και με τις οριακές περιπτώσεις. Όπως έχουμε ήδη αναφέρει σε προηγούμενο κεφάλαιο, επιθυμία μας είναι το Copula να καλύπτει όλο το χώρο ανάμεσα από τα φράγματα και καθώς η παράμετρος τείνει προς κάποιο άκρο του εύρους του τότε το Copula να πλησιάζει κάποιο από τα φράγματα. Για τα Archimedean Copulas ισχύει το εξής:

ΘΕΩΡΗΜΑ: Έστω C_θ το Archimedean Copula με παράμετρο $\theta \in \Theta$ του οποίου ο γεννήτορας φ_θ ανήκει στο σύνολο όλων των χαρακτηριστικών γεννητόρων Φ και είναι διαφορίσιμος. Τότε το $C = \lim C_\theta$ είναι ένα Archimedean Copula αν και μόνο αν υπάρχει συνάρτηση $\varphi \in \Phi$ ώστε για όλα τα $s, t \in (0, 1)$ να ισχύει:

$$\lim \frac{\varphi_\theta(s)}{\varphi_\theta'(t)} = \frac{\varphi(s)}{\varphi'(t)}.$$

Το όριο στην παραπάνω περίπτωση είναι το κατάλληλο μονόπλευρο όριο, όταν το θ τείνει σε ένα από τα δύο άκρα του Θ .¹⁸

ΘΕΩΡΗΜΑ: Έστω C_θ είναι το Archimedean Copula με παράμετρο $\theta \in \Theta$ και του οποίου ο γεννήτορας φ_θ ανήκει στο σύνολο όλων των χαρακτηριστικών γεννητόρων Φ και είναι διαφορίσιμος. Τότε $\lim C_\theta(u, v) = M(u, v)$ αν και μόνο αν για $t \in (0, 1)$ ισχύει:

$$\lim \frac{\varphi_\theta(t)}{\varphi_\theta'(t)} = 0.$$

¹⁸ Ουσιαστικά το Θεώρημα αυτό ισχύει για τις περιπτώσεις που έχουμε είτε το κάτω φράγμα είτε την περίπτωση της ανεξαρτησίας. Δεν ισχύει για την περίπτωση του άνω φράγματος καθώς δεν είναι Archimedean Copula.

Το όριο στην παραπάνω περίπτωση είναι το κατάλληλο μονόπλευρο όριο, όταν το θ τείνει σε ένα από τα δύο άκρα του Θ .

Για τους συντελεστές συσχέτισης-εξάρτησης για τα Archimedean Copulas αποδεικνύονται τα εξής:

ΠΡΟΤΑΣΗ: Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής ένα Archimedean Copula C , το οποίο παράγεται από την $\varphi \in \Phi$. Τότε το τ του Kendall για τα X, Y δίνεται από τον τύπο:

$$\tau_C = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

ΠΡΟΤΑΣΗ: Έστω X, Y δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής ένα Archimedean Copula C , το οποίο παράγεται από την $\varphi \in \Phi$. Τότε οι συντελεστές εξάρτησης της άνω και της κάτω ουράς δίνονται από τους τύπους:

$$\lambda_U = 2 - \lim_{t \rightarrow 1^-} \frac{1 - \varphi^{[-1]}(2\varphi(t))}{1 - t} = 2 - \lim_{s \rightarrow 0^+} \frac{1 - \varphi^{[-1]}(2s)}{1 - \varphi^{[-1]}(s)}$$

και

$$\lambda_L = \lim_{t \rightarrow 0^+} \frac{\varphi^{[-1]}(2\varphi(t))}{t} = \lim_{s \rightarrow \infty} \frac{\varphi^{[-1]}(2s)}{\varphi^{[-1]}(s)}.$$

Θα παρουσιάσουμε στη συνέχεια τις τρεις πιο συνηθεις οικογένειες της κλάσης των Archimedean Copulas.¹⁹

¹⁹ Για τη γενίκευση των τριών οικογενειών στις n διαστάσεις παροτρύνουμε το αναγνώστη να κοιτάξει το τέταρτο κεφάλαιο του βιβλίου των Cherubini, Luciano και Vecchiato [6].

3.3.1 Οικογένεια Clayton

Η οικογένεια Clayton είναι της μορφής:

$$C_{\theta}(u, v) = [\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-\frac{1}{\theta}}$$

όπου $\theta \in [-1, \infty) \setminus \{0\}$. Ο χαρακτηριστικός γεννήτορας της οικογένειας αυτής είναι:

$$\varphi_{\theta}(t) = \frac{1}{\theta}(t^{-\theta} - 1).$$

Ο παραπάνω χαρακτηριστικός γεννήτορας γίνεται αυστηρός αν θεωρήσουμε ότι $\theta \geq 0$ και τότε η οικογένεια Clayton γράφεται στη μορφή:

$$C_{\theta}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}.$$

Οι οριακές περιπτώσεις που προκύπτουν στην συγκεκριμένη οικογένεια είναι: $C_{-1} = W$, $C_0 = \Pi$, $C_1 = \frac{\Pi}{\Sigma - \Pi}$ ²⁰ και $C_{\infty} = M$.

Ενδεικτικά θα αποδείξουμε τις περιπτώσεις για τα δύο φράγματα. Για το κάτω φράγμα, έχει αποδειχτεί ότι είναι Archimedean Copula με χαρακτηριστικό γεννήτορα $\varphi(t) = 1 - t$. Για να χρησιμοποιήσουμε το θεώρημα για τη σύγκλιση των Archimedean Copulas υπολογίζουμε το λόγο $\frac{\varphi(s)}{\varphi'(t)}$ για το κάτω φράγμα και είναι ίσος με $s - 1$. Άρα χρησιμοποιώντας το θεώρημα της σύγκλισης θα πρέπει το αποτέλεσμα να βγεί ίσο με $s - 1$ για να έχω το κάτω φράγμα. Πράγματι:

Η πρώτη παράγωγος του $\varphi_{\theta}(t) = \frac{1}{\theta}(t^{-\theta} - 1)$ είναι $\varphi'_{\theta}(t) = -\frac{1}{t^{\theta+1}}$. Επομένως, έχουμε:

$$\lim_{\theta \rightarrow -1^+} \frac{\varphi_{\theta}(s)}{\varphi'_{\theta}(t)} = \lim_{\theta \rightarrow -1^+} \frac{t^{\theta+1}}{\theta}(1 - s^{-\theta}) = -1(1 - s) = s - 1.$$

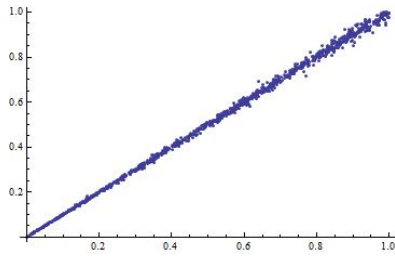
Άρα πράγματι όταν η παράμετρος τείνει να πάρει την τιμή -1 , το Clayton Copula πλησιάζει το κάτω φράγμα. Για το άνω φράγμα, χρησιμοποιούμε το αντίστοιχο θεώρημα και έχουμε:

$$\lim_{\theta \rightarrow \infty} \frac{\varphi_{\theta}(t)}{\varphi'_{\theta}(t)} = \lim_{\theta \rightarrow \infty} \frac{t^{\theta+1} - t}{\theta} = 0.$$

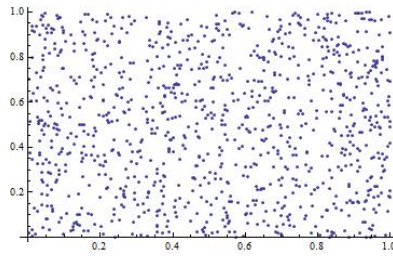
²⁰ Το Copula $C = \frac{\Pi}{\Sigma - \Pi}$ είναι το Copula της διδιάστατης λογιστικής κατανομής του Gumbel που χρησιμοποιήθηκε και στο πρώτο κεφάλαιο. Γενικά στα Archimedean Copulas εμφανίζεται συχνά και αυτός είναι ο λόγος που αναφέρεται στην παρούσα εργασία.

Επομένως, όταν η παράμετρος τείνει να πάρει μια πολύ μεγάλη τιμή, το Clayton Copula πλησιάζει το άνω φράγμα.

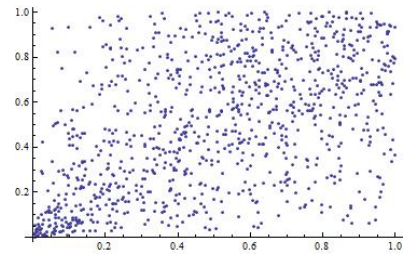
Θα δώσουμε σε αυτό το σημείο κάποια γραφήματα τα οποία είναι ουσιαστικά 1000 σημεία που έχουν παραχθεί από το Clayton Copula για διαφορετικές τιμές της παραμέτρου θ .



Σχήμα 3.18: 1000 τυχαία σημεία από Clayton με παράμετρο $\frac{1}{\theta} = 10^{-2}$.



Σχήμα 3.19: 1000 τυχαία σημεία από Clayton με παράμετρο $\frac{1}{\theta} = 10^3$.



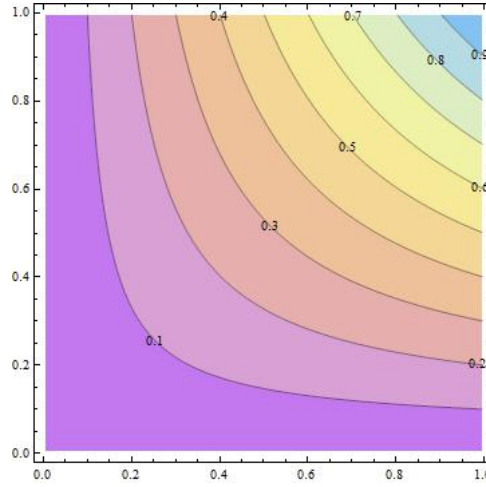
Σχήμα 3.20: 1000 τυχαία σημεία από Clayton με παράμετρο $\theta = 1$.

Να σημειώσουμε ότι συνήθως η συγκεκριμένη οικογένεια χρησιμοποιείται όταν $\theta \in [0, \infty)$.²¹ Στα παραπάνω διαγράμματα έχουμε θεωρήσει $\theta \in (0, \infty)$ και $c = \frac{1}{\theta}$ καθώς έτσι είναι ορισμένες οι εντολές στο mathematica. Τότε, προφανώς, δεν μπορούμε να επιτύχουμε το κάτω φράγμα αλλά είναι πιο εύκολοι οι υπολογισμοί. Αυτά που παρατηρούμε από τα συγκεκριμένα διαγράμματα είναι αυτά που αναφέραμε προηγουμένως στη θεωρία. Θα προσθέσουμε απλά ότι στην κάτω ουρά φαίνεται να έχουμε μεγαλύτερη εξάρτηση από ό,τι στην άνω ουρά.

Σημαντικό στην περαιτέρω μελέτη μας θα είναι και το Contour Plot της συνάρτησης κατανομής του Clayton Copula και για αυτόν τον λόγο το δίνουμε παρακάτω. Έχουμε θεωρήσει ότι η παράμετρος είναι $\theta = \frac{1}{3}$ και οι περιθώριες ομοιόμορφες στο $(0,1)$.

²¹ Για το φ και το C η τιμή 0 του θ ουσιαστικά εκφράζει την περίπτωση της ανεξαρτησίας, καθώς τα όρια όταν $\theta \rightarrow 0$ είναι $-\log t$, uv αντίστοιχα.

Σχήμα 3.21: Το Contour Plot της σ.κ. του Clayton Copula για $\theta = \frac{1}{3}$.



Το Clayton Copula είναι ουσιαστικά μια μη συμμετρική θετικά διατεταγμένη²² συνάρτηση κατανομής που επιτυγχάνει μεγαλύτερη εξάρτηση στην κάτω ουρά. Συγκεκριμένα, χρησιμοποιώντας το Πόρισμα 2 από το θεώρημα που αφορά τη διάταξη συμφωνίας των Archimedean Copula έχουμε:

$$\frac{\varphi'_{\theta_1}(t)}{\varphi'_{\theta_2}(t)} = t^{\theta_2 - \theta_1}$$

όπου για $\theta_1 \leq \theta_2$ έχουμε $(t^{\theta_2 - \theta_1})' = (\theta_2 - \theta_1)t^{\theta_2 - \theta_1 - 1} \geq 0$. Άρα η $\frac{\varphi'_{\theta_1}(t)}{\varphi'_{\theta_2}(t)} = t^{\theta_2 - \theta_1}$ είναι γνησίως αύξουσα και επομένως $C_1 < C_2$.

Για το συντελεστή συσχέτισης τ του Kendall έχουμε τα εξής για το Clayton Copula:

$$\tau = 1 + 4 \int_0^1 \frac{t^{\theta+1} - t}{\theta} dt = 1 + \frac{4}{\theta} \left[\frac{t^{\theta+2}}{\theta+2} - \frac{t^2}{2} \right]_0^1 = \frac{\theta}{\theta+2}.$$

Επειδή έχουμε περιορίσει το Copula ώστε να είναι αυστηρό, η ψευδοαντίστροφη του γεννήτορα είναι ίση με τη συνήθη αντίστροφη συνάρτηση. Άρα για $\varphi(t) = \frac{t^{-\theta} - 1}{\theta}$, έχουμε $\varphi^{-1}(t) = (\theta t + 1)^{-\frac{1}{\theta}}$. Επομένως, έχουμε:

$$\lambda_U = 2 - \lim_{s \rightarrow 0^+} \frac{1 - (\theta 2s + 1)^{-\frac{1}{\theta}}}{1 - (\theta s + 1)^{-\frac{1}{\theta}}} = 2 - 2 \lim_{s \rightarrow 0^+} \frac{(\theta 2s + 1)^{-\frac{1}{\theta} - 1}}{(\theta s + 1)^{-\frac{1}{\theta} - 1}} = 2 - 2 = 0$$

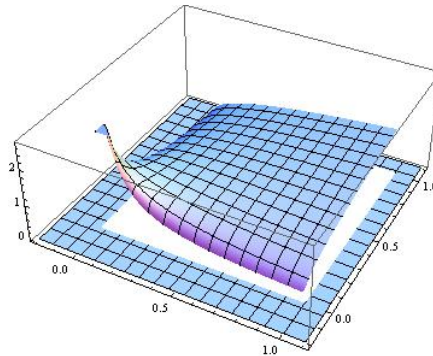
²² Θετικά διατεταγμένη συνάρτηση C_θ έχουμε όταν για $\theta_1 \leq \theta_2$ ισχύει $C_{\theta_1} < C_{\theta_2}$.

και

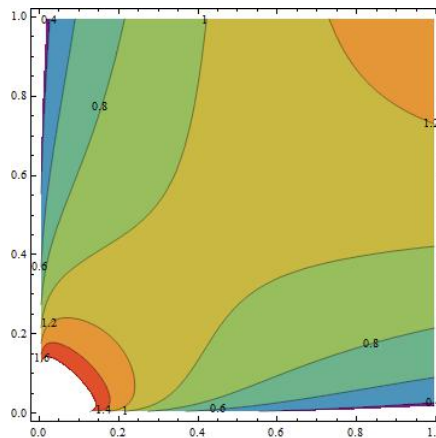
$$\lambda_L = \lim_{s \rightarrow \infty} \frac{(\theta 2s + 1)^{-\frac{1}{\theta}}}{(\theta s + 1)^{-\frac{1}{\theta}}} = 2 \lim_{s \rightarrow 0^+} \frac{(2\theta s + 1)^{-\frac{1}{\theta}-1}}{(\theta s + 1)^{-\frac{1}{\theta}-1}} = 2\left(\frac{2\theta}{\theta}\right)^{-\frac{1}{\theta}-1} = 2^{-\frac{1}{\theta}}.$$

Θα κλείσουμε την υποενότητα αυτή δίνοντας το γράφημα για τη συνάρτηση πυκνότητας του Clayton Copula με παράμετρο $\theta = \frac{1}{3}$ και το αντίστοιχο Contour Plot.

Σχήμα 3.22: Η συνάρτηση πυκνότητας του Clayton Copula για $\theta = \frac{1}{3}$.



Σχήμα 3.23: Το Contour Plot της σ.π. του Clayton Copula για $\theta = \frac{1}{3}$.



Αυτό που παρατηρούμε στο Contour Plot της συνάρτησης πυκνότητας είναι πως δεν έχουμε τον ίδιο χρωματισμό στις δύο ουρές και επομένως είναι εμφανές ότι δεν υπάρχει συμμετρία. Επίσης έχουμε θεωρήσει ότι $x, y \in (-0.2, 1.1)$ για καλύτερα οπτικά αποτελέσματα και προφανώς εκεί η πυκνότητα του Copula μηδενίζεται.

3.3.2 Οικογένεια Gumbel-Hougaard

Η οικογένεια Gumbel-Hougaard είναι της μορφής:

$$C_\theta(u, v) = \exp(-[(-\log u)^\theta + (-\log v)^\theta]^{\frac{1}{\theta}})$$

όπου $\theta \in [1, \infty)$. Ο χαρακτηριστικός γεννήτορας της οικογένειας αυτής είναι:

$$\varphi_\theta(t) = (-\log t)^\theta.$$

Ο παραπάνω γεννήτορας είναι αυστηρός και οι οριακές περιπτώσεις που ισχύουν στη συγκεκριμένη οικογένεια είναι: $C_1 = \Pi$ και $C_\infty = M$.

Πράγματι, η πρώτη παράγωγος του χαρακτηριστικού γεννήτορα είναι

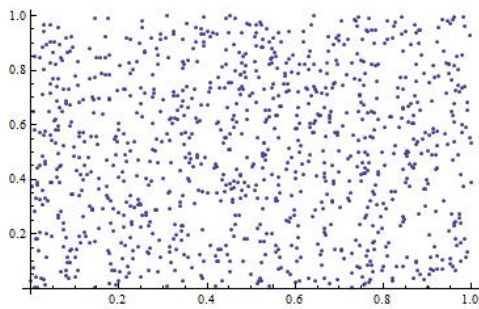
$$\varphi'_\theta(t) = -\frac{\theta}{t}(-\log t)^{\theta-1}.$$

και από το θεώρημα σύγκλισης έχουμε:

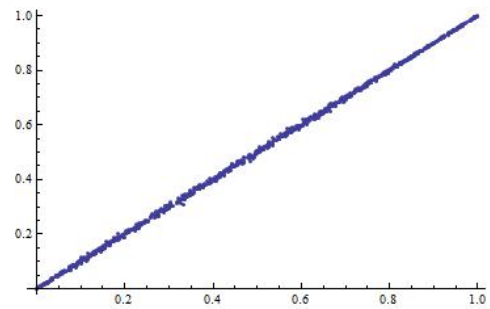
$$\lim_{\theta \rightarrow 1^+} \frac{\varphi_\theta(s)}{\varphi'_\theta(t)} = \lim_{\theta \rightarrow 1^+} -\frac{t(-\log s)^\theta}{-\theta(-\log t)^{\theta-1}} = t \log s.$$

Άρα όταν η παράμετρος τείνει να πάρει την τιμή 1, το Gumbel Copula πλησιάζει το Copula ανεξαρτησίας. Αυτό ισχύει καθώς για το Copula ανεξαρτησίας έχουμε $\varphi(s) = -\log s$ και $\varphi'(t) = -\frac{1}{t}$. Άρα ο λόγος $\frac{\varphi(s)}{\varphi'(t)}$ είναι ίσος με $t \log s$. Για το άνω φράγμα, χρησιμοποιούμε το αντίστοιχο θεώρημα και έχουμε:

$$\lim_{\theta \rightarrow \infty} \frac{\varphi_\theta(t)}{\varphi'_\theta(t)} = -t \lim_{\theta \rightarrow \infty} \frac{-\log t}{\theta} = 0.$$



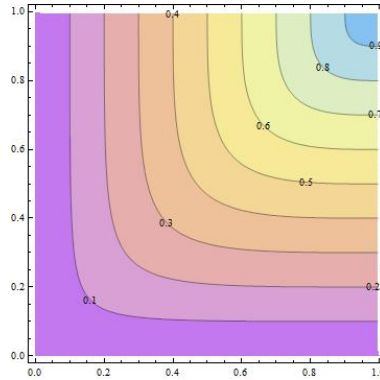
Σχήμα 3.24: 1000 τυχαία σημεία από Gumbel Copula με παράμετρο $\theta = 1$.



Σχήμα 3.25: 1000 τυχαία σημεία από Gumbel Copula με παράμετρο $\theta = 10^2$.

Τα παραπάνω σχήματα απεικονίζουν 1000 τυχαία σημεία από το Gumbel-Hougaard Copula με διαφορετικές παραμέτρους και με περιθώριες ομοιόμορφες στο (0,1). Δίνουμε και το Contour plot της συνάρτησης κατανομής του Gumbel-Hougaard Copula με παράμετρο $\theta = 3$. Αυτό είναι:

Σχήμα 3.26: Το Contour Plot της σ.κ. του Gumbel Copula για $\theta = 3$.



Το Gumbel Copula είναι μια μη συμμετρική θετικά διατεταγμένη συνάρτηση κατανομής, που επιτυγχάνει μεγαλύτερη εξάρτηση στην άνω ουρά. Από το Πόρισμα 1 του θεωρήματος για τη διάταξη συμφωνίας έχουμε:

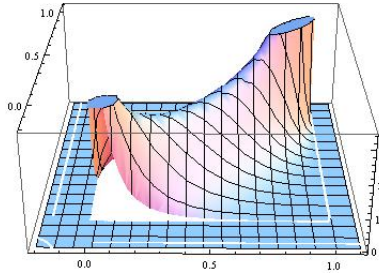
$$\varphi_{\theta_1} \circ \varphi_{\theta_2}^{-1}(t) = t^{\theta_1/\theta_2},$$

όπου για $\theta_1 \leq \theta_2$ η δεύτερη παράγωγος της παραπάνω ποσότητας είναι μικρότερη ή ίση του μηδενός και, άρα είναι κοίλη συνάρτηση. Επομένως, $C_1 < C_2$. Για το συντελεστή συσχέτισης τ του Kendall έχουμε τα εξής για το Gumbel Copula:

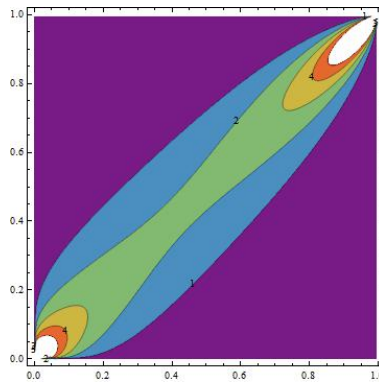
$$\begin{aligned} \tau &= 1 + 4 \int_0^1 \frac{t \log t}{\theta} dt = 1 + \frac{4}{\theta} \left[\frac{t^2}{2} \log t \right]_0^1 - \frac{4}{\theta} \int_0^1 \frac{t}{2} dt = 1 + \frac{4}{\theta} \left[0 - \lim_{t \rightarrow 0} \frac{\log t}{\frac{t^2}{2}} \right] - \frac{4}{\theta} \frac{1}{4} = \\ &= 1 - \frac{4}{\theta} \lim_{t \rightarrow 0} (-4t^2) - \frac{1}{\theta} = 1 - \frac{1}{\theta}. \end{aligned}$$

Θα δώσουμε σε αυτό το σημείο το γράφημα της συνάρτησης πυκνότητας του Gumbel-Hougaard Copula με παράμετρο $\theta = 3$ καθώς και το αντίστοιχο Contour Plot.

Σχήμα 3.27: Η συνάρτηση πυκνότητας του Gumbel Copula για $\theta = 3$.



Σχήμα 3.28: Το Contour Plot της σ.π. του Gumbel Copula για $\theta = 3$.



Και εδώ έχουμε θεωρήσει ότι $x, y \in (-0.2, 1.1)$ για καλύτερα οπτικά αποτελέσματα. Στα δύο παραπάνω γραφήματα είναι εμφανές ότι δεν υπάρχει συμμετρία και ότι η πυκνότητα στην άνω ουρά είναι μεγαλύτερη από ό,τι στην κάτω. Είναι, επομένως, χρήσιμοι οι συντελεστές εξάρτησης ουρών, οι οποίοι είναι ίσοι με:

$$\lambda_U = 2 - \lim_{s \rightarrow 0^+} \frac{1 - \exp((-2s)^{\frac{1}{\theta}})}{1 - \exp((-s)^{\frac{1}{\theta}})} = 2 + \lim_{s \rightarrow 0^+} \exp[(-2s)^{\frac{1}{\theta}} - (-s)^{\frac{1}{\theta}}](-2)^{\frac{1}{\theta}} = 2 - 2^{\frac{1}{\theta}}$$

και

$$\lambda_L = \lim_{s \rightarrow \infty} \frac{\exp((-2s)^{\frac{1}{\theta}})}{\exp((-s)^{\frac{1}{\theta}})} = \lim_{s \rightarrow \infty} \exp[(-s)^{\frac{1}{\theta}}(2^{\frac{1}{\theta}} - 1)] = 0.$$

3.3.3 Οικογένεια Frank

Η οικογένεια Frank είναι της μορφής:

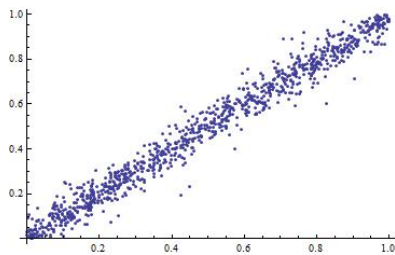
$$C_{\theta}(u, v) = -\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right)$$

όπου $\theta \in \mathbf{R} \setminus \{0\}$. Ο χαρακτηριστικός γεννήτορας της οικογένειας αυτής είναι:

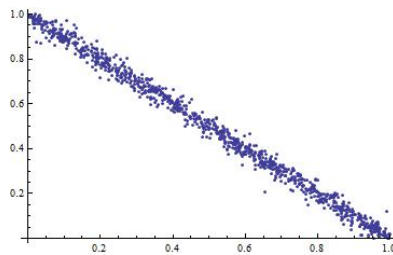
$$\varphi_{\theta}(t) = -\log \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}.$$

Ο παραπάνω γεννήτορας είναι αυστηρός και οι οριακές περιπτώσεις που ισχύουν στη συγκεκριμένη οικογένεια είναι: $C_{-\infty} = W$, $C_0 = \Pi$ και $C_{\infty} = M$.

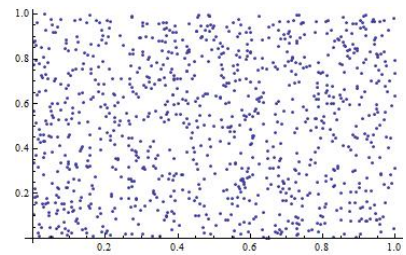
Στα παρακάτω γραφήματα έχουμε παράγει 1000 τυχαία σημεία από το Frank Copula με διαφορετικές παραμέτρους $c = e^{-\theta}$ και με περιθώριες ομοιόμορφες στο $(0,1)$. Επίσης, δίνουμε το Contour Plot της συνάρτησης κατανομής του Frank Copula με παράμετρο $e^{-\theta} = \frac{1}{3}$.



Σχήμα 3.29: 1000 τυχαία σημεία από Frank με παράμετρο $e^{-\theta} = 10^{-15}$.

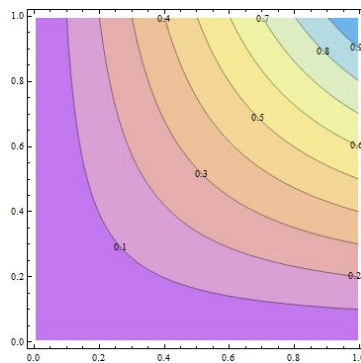


Σχήμα 3.30: 1000 τυχαία σημεία από Frank με παράμετρο $e^{-\theta} = 10^{30}$.



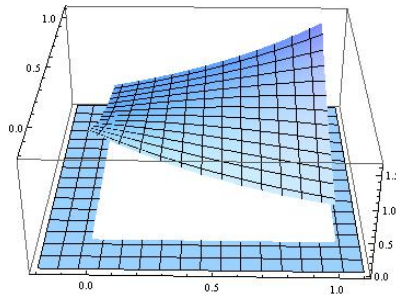
Σχήμα 3.31: 1000 τυχαία σημεία από Frank με παράμετρο $e^{-\theta} = 1$.

Σχήμα 3.32: Το Contour Plot της σ.κ. του Frank Copula για $e^{-\theta} = \frac{1}{3}$.

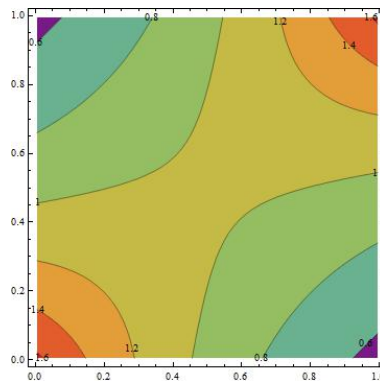


Το Frank Copula είναι ένα συμμετρικό Copula, με την έννοια της ακτινικής συμμετρίας και επομένως $C = \hat{C}$. Αν πάρουμε το αυστηρό Clayton Copula και το Gumbel Copula, θα δούμε ότι το όριο στο κάτω άκρο του διαστήματος που ορίζεται η παράμετρος είναι η περίπτωση της ανεξαρτησίας και επομένως κανένα από τα δύο δεν επιτρέπει αρνητική συσχέτιση, εν αντιθέσει με το Frank Copula το οποίο επιτρέπει την αρνητική συσχέτιση. Τέλος να αναφέρουμε ότι, αν και η συγκεκριμένη οικογένεια έχει την ιδιότητα της ακτινικής συμμετρίας, η εξάρτηση στις ουρές είναι μικρή σε σχέση με τα Elliptical Copulas που αναφέραμε, ενώ έχει πιο ισχυρή συσχέτιση στο κέντρο της κατανομής.

Σχήμα 3.33: Η συνάρτηση πυκνότητας του Frank Copula για $e^{-\theta} = \frac{1}{3}$.



Σχήμα 3.34: Το Contour Plot της σ.π. του Frank Copula για $e^{-\theta} = \frac{1}{3}$.



Παραπάνω δίνουμε το γράφημα της συνάρτησης πυκνότητας του Frank Copula με παράμετρο $e^{-\theta} = \frac{1}{3}$ και περιθώριες ομοιόμορφες στο $(0,1)$ καθώς και το αντίστοιχο Contour Plot. Και εδώ έχουμε θεωρήσει $x, y \in (-0.2, 1.1)$ για καλύτερα οπτικά αποτελέσματα. Είναι εμφανής η συμμετρία από το Contour Plot καθώς και ότι, παρ' όλο που είναι ακτινικά συμμετρικό, διαφέρει αρκετά από τα αντίστοιχα plots των Elliptical Copulas.

Για το συντελεστή συσχέτισης τ του Kendall αποδεικνύονται τα εξής για το Frank Copula:

$$\tau = 1 - \frac{4}{\theta}(1 - D_1(\theta))$$

όπου

$$D_1(a) = \frac{1}{a} \int_0^a \frac{t}{e^t - 1} dt$$

είναι η συνάρτηση Debye, η οποία δίνεται γενικά από την ολοκληρωτική έκφραση

$$D_n(x) = \frac{n}{x^n} \int_0^x \frac{t^n}{e^t - 1} dt.$$

Για τη συνάρτηση Debye ισχύει:

$$D_n(-x) = D_n(x) + \frac{nx}{n+1}.$$

Για τους συντελεστές εξάρτησης ουρών έχουμε:

$$\lambda_U = \lambda_L = \lim_{s \rightarrow \infty} \frac{\log[1 - e^{-2s}(1 - e^{-\theta})]}{\log[1 - e^{-s}(1 - e^{-\theta})]} = \lim_{s \rightarrow \infty} \frac{2e^{-s}(1 - e^{-s}(1 - e^{-\theta}))}{(1 - e^{-2s}(1 - e^{-\theta}))} = 0.$$

Η μηδενική τιμή των συντελεστών αυτών σημαίνει ότι, ανεξάρτητα από το πόσο είναι συσχετισμένες δύο τυχαίες μεταβλητές, το να πάρει η μία ακραία τιμή είναι ανεξάρτητο από το αν θα πάρει και η άλλη.

Πίνακας 3.1: Σύνοψη αποτελεσμάτων για τα Archimedean Copulas.

Οικογένεια	Copula	Γεννήτορας	Παράμετρος	Οριακές περιπτώσεις	Kendall's Tau	λ_U	λ_L
Clayton	$[\max(u^{-\theta} + v^{-\theta} - 1, 0)]^{-\frac{1}{\theta}}$	$\frac{1}{\theta}(r^{-\theta} - 1)$	$[-1, \infty) \setminus \{0\}$	$C_{-1} = W, C_0 = \Pi, C_{\infty} = M$	$\frac{\theta}{\theta + 2}$	0	$2^{-\frac{1}{\theta}}$
Gumbel	$\exp(-[(-\log u)^\theta + (-\log v)^\theta]^{\frac{1}{\theta}})$	$(-\log t)^\theta$	$[1, \infty)$	$C_1 = \Pi, C_{\infty} = M$	$1 - \frac{1}{\theta}$	$2 - 2^{\frac{1}{\theta}}$	0
Frank	$-\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right)$	$-\log \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$	$(-\infty, \infty) \setminus \{0\}$	$C_{-\infty} = W, C_0 = \Pi, C_{\infty} = M$	$1 - \frac{4}{\theta}(1 - D_1(\theta))$	0	0

Κεφάλαιο 4

ΑΛΓΟΡΙΘΜΟΙ ΠΑΡΑΓΩΓΗΣ ΔΕΔΟΜΕΝΩΝ-ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ

4.1 ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο αρχικά θα δώσουμε κάποιους αλγόριθμους για την παραγωγή δεδομένων από τα Copula που δόθηκαν στο προηγούμενο κεφάλαιο. Οι κώδικες που θα χρησιμοποιηθούν θα ενταχθούν στα Παραρτήματα στο τέλος της εργασίας. Επίσης θα παρουσιάσουμε τρόπους για την εκτίμηση των Copulas. Συγκεκριμένα, θα δώσουμε έναν παραμετρικό αλγόριθμο, ένα ημιπαραμετρικό και ένα μη παραμετρικό και τέλος θα κλείσουμε το κεφάλαιο αυτό με κάποια μέθοδο σύγκρισης των Copula.

4.2 Αλγόριθμοι για την παραγωγή δεδομένων από Copula

Στην ενότητα αυτή θα παρουσιάσουμε κάποιες μεθόδους με τις οποίες μπορούμε να παραγάγουμε δεδομένα από Copulas. Συγκεκριμένα, θα δώσουμε δύο μεθόδους παραγωγής δεδομένων από τα Elliptical Copulas, μία για τα Gaussian και μία για τα t-Copula και μία ακόμα μέθοδο παραγωγής δεδομένων από τα Archimedean.

4.2.1 Elliptical Copulas

Μία εύκολη μέθοδος για να παράγουμε δεδομένα από ένα d -διάστατο Elliptical Copula είναι να προσομοιώσουμε δεδομένα από την αντίστοιχη d -διάστατη ελλειπτική κατανομή και να τα μετασχηματίσουμε εφαρμόζοντας την αντίστοιχη περιθώρια ώστε να πάρουμε δεδομένα που να είναι στο $(0,1)$. Ο αλγόριθμος που θα παρουσιάσουμε απαιτεί να δώσουμε έναν ισοδύναμο ορισμό για την Πολυδιάστατη Κανονική Κατανομή σε σχέση με αυτόν του τρίτου κεφαλαίου ο οποίος είναι:

Μια d -διάστατη τυχαία μεταβλητή $\mathbf{X} = (X_1, X_2, \dots, X_d)$ ακολουθεί την Πολυδιάστατη Κανονική Κατανομή, $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, αν γράφεται στη μορφή:

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{AZ}$$

όπου $\boldsymbol{\mu}$ είναι το διάνυσμα με τις μέσες τιμές, $\mathbf{Z} = (Z_1, \dots, Z_k)$, όπου $Z_i \sim N(0, 1)$ για $i = 1, \dots, k$ και A είναι ο $d \times k$ πίνακας για τον οποίο ισχύει $\Sigma = AA'$.¹ Για τα Gaussian Copula, όπως έχουμε ήδη αναφέρει, θεωρούμε το διάνυσμα των μέσων τιμών μηδενικό και παίρνουμε τον πίνακα συσχετίσεων R αντί του πίνακα Σ .

Ένας αλγόριθμος για την παραγωγή δεδομένων από το Gaussian Copula είναι ο εξής:

Βήμα 1: Κάνουμε την παραγοντοποίηση Cholesky του πίνακα συσχετίσεων R και θέτουμε A να είναι ο παράγοντας Cholesky.

Βήμα 2: Κάνουμε τα παρακάτω n φορές:

α. Παράγουμε $\mathbf{Z} = (Z_1, \dots, Z_k)$, όπου $Z_i \sim N(0, 1)$.

β. Θέτουμε $\mathbf{X} = \mathbf{AZ}$.

γ. Επιστρέφουμε το $\mathbf{U} = (\Phi(X_1), \dots, \Phi(X_d))$.

Για το t -Copula θα χρειαστούμε έναν ανάλογο ορισμό για την Πολυδιάστατη Κατανομή t -Student, όπως αυτός που δόθηκε παραπάνω. Επομένως, έχουμε:

¹ Η Cholesky παραγοντοποίηση (Cholesky decomposition) ενός θετικά ορισμένου πίνακα, εδώ Σ , είναι το γινόμενο AA' , όπου ο A είναι ένας κάτω τριγωνικός πίνακας και ονομάζεται παράγοντας Cholesky.

Μια d -διάστατη τυχαία μεταβλητή $\mathbf{X} = (X_1, X_2, \dots, X_d)$ ακολουθεί την Πολυδιάστατη Κατανομή t -Student, $\mathbf{X} \sim t_d(\boldsymbol{\mu}, \Sigma, \nu)$, αν γράφεται στη μορφή:

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{W} \mathbf{A} \mathbf{Z}$$

όπου $W \sim \text{InvG}(\frac{\nu}{2}, \frac{\nu}{2})$, A είναι ο $d \times k$ πίνακας για τον οποίο ισχύει $\Sigma = AA'$ και $\mathbf{Z} = (Z_1, \dots, Z_k)$, όπου $Z_i \sim N(0, 1)$ για $i = 1, \dots, k$. Ή όμοια,

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{\frac{\nu}{V}} \mathbf{A} \mathbf{Z}$$

όπου $V \sim \chi^2(\nu)$. Για τα t -Copula, όπως έχουμε ήδη αναφέρει, θεωρούμε το διανύσμα των μέσων τιμών μηδενικό και παίρνουμε τον πίνακα συσχετίσεων R αντί του πίνακα Σ .

Ένας αλγόριθμος για την παραγωγή δεδομένων από το t -Copula είναι ο εξής:

Βήμα 1: Κάνουμε την παραγοντοποίηση Cholesky του πίνακα συσχετίσεων R και θέτουμε A να είναι ο παράγοντας Cholesky.

Βήμα 2: Κάνουμε τα παρακάτω n φορές:

α. Παράγουμε $\mathbf{Z} = (Z_1, \dots, Z_k)$, όπου $Z_i \sim N(0, 1)$.

β. Παράγουμε $W \sim \text{InvG}(\frac{\nu}{2}, \frac{\nu}{2})$ (ή παράγουμε $V \sim \chi^2(\nu)$).

γ. Θέτουμε $\mathbf{X} = \sqrt{W} \mathbf{A} \mathbf{Z}$ (ή θέτουμε $\mathbf{X} = \sqrt{\frac{\nu}{V}} \mathbf{A} \mathbf{Z}$).

δ. Επιστρέφουμε το $\mathbf{t} = (t_\nu(X_1), \dots, t_\nu(X_d))$.

Οι παραπάνω αλγόριθμοι που αφορούν στην παραγωγή σημείων από τα Elliptical Copulas δίνονται στο Παράρτημα στο τέλος της εργασίας και σε mathematica και σε R. Η συγκεκριμένη επιλογή των παραπάνω αλγορίθμων έγινε καθώς εμφανίζονται αρκετές φορές στη βιβλιογραφία της παρούσας εργασίας. Προφανώς υπάρχουν ισοδύναμοι αλλά και εναλλακτικοί τρόποι για παραγωγή δεδομένων από ένα Copula.²

² Εναλλακτικές μέθοδοι προσομοίωσης δεδομένων από Elliptical Copulas δίνονται στην εργασία με τίτλο *Dependence Modelling via the Copula Method*, της Helen Arnold [3].

4.2.2 Archimedean Copulas

Για την παραγωγή δεδομένων από τα Archimedean Copula έχουν προταθεί διάφοροι αλγόριθμοι οι οποίοι ως επί το πλείστον στηρίζονται στους χαρακτηριστικούς γεννήτορες. Παρ' όλα αυτά, στην παρούσα εργασία δίνουμε έναν γενικό αλγόριθμο παραγωγής δεδομένων από Copula, ο οποίος δεν στηρίζεται στο χαρακτηριστικό γεννήτορα, και επομένως θα μπορούσε να εφαρμοστεί για όποια οικογένεια των Copulas θέλουμε.³ Ο αλγόριθμος είναι ο εξής:

Κάνουμε τα παρακάτω n φορές:

Βήμα 1: Παράγουμε $U_2, V \sim U(0, 1)$.

Βήμα 2: Υπολογίζουμε το:

$$C(u | v) = \frac{\partial C(u, v)}{\partial v}.$$

Βήμα 3: Υπολογίζουμε την γενικευμένη αντίστροφη της παραπάνω ποσότητας, έστω $C^{(-1)}(u | v)$.

Βήμα 4: Θέτουμε $U_1 = C^{(-1)}(V)$.

Βήμα 5: Επιστρέφουμε το (U_1, U_2) .

Ο συγκεκριμένος αλγόριθμος⁴, όπως ήδη αναφέραμε, μπορεί να εφαρμοστεί σε οποιοδήποτε Copula. Παρ' ό,τι είναι ένας σχετικά εύκολος αλγόριθμος, καμιά φορά είναι δύσκολα εφαρμόσιμος καθώς μπορεί να μην ορίζεται σωστά η αντίστροφη. Αυτό που προτείνουν οι ειδικοί είναι να αντιλαμβανόμαστε κάθε φορά την εξάρτηση των δεδομένων και να κατασκευάζουμε έναν αλγόριθμο που να αποτυπώνει σωστά την εξάρτηση αυτή.

Για τα Archimedean Copulas που αναφέραμε στο προηγούμενο κεφάλαιο θα δώσουμε τις μερικές παραγώγους και την αντίστροφη καθώς υπολογίζονται σχετικά εύκολα. Αυτό γίνεται καθώς, όπως είπαμε και πριν, μπορεί να μην τρέχει ο αλγόριθμος αν δεν οριστεί

³ Για Archimedean Copulas διάστασης $n \geq 3$ υπάρχει αλγόριθμος παραγωγής δεδομένων στην εργασία των Frees και Valdez [14].

⁴ Τον αλγόριθμο μπορεί να τον βρει ο αναγνώστης στην εργασία της Arnold [3].

απόλυτα σωστά η αντίστροφη. Έτσι έχουμε:

α. Για το Clayton Copula:

$$C(u | v) = \frac{\partial C(u,v)}{\partial v} = v^{-\theta-1}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-1}$$

και

$$C^{(-1)}(u | k) = (1 - u^{-\theta} + (ku^{1+\theta})^{-\frac{\theta}{1+\theta}})^{-\frac{1}{\theta}}.$$

όπου στο k θα βάλουμε τη δεύτερη τιμή που παράγουμε τυχαία από την $U(0, 1)$.

β. Για το Gumbel-Hougaard Copula:

$$C(u | v) = \frac{1}{v}(-\log v)^{\theta-1}([-\log u]^\theta + [-\log v]^\theta)^{\frac{1}{\theta}-1} \exp[-([-\log u]^\theta + [-\log v]^\theta)^{\frac{1}{\theta}}]$$

και

$$C^{(-1)}(u | k) = \exp[-[-(\log u)^\theta + ((-1 + \theta)ProductLog[\frac{(-ku(-\log u)^{-\theta} \log u)^{-\frac{1}{\theta-1}}}{\theta - 1}])^\theta]^{\frac{1}{\theta}}].^5$$

Το παραπάνω υπολογίστηκε μέσω mathematica και στο k θα εισαγάγουμε τη δεύτερη τιμή που παράγουμε τυχαία από την $U(0, 1)$.

γ. Για το Frank Copula:

$$C(u | v) = \frac{e^{-\theta v}(e^{-\theta u} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1)}$$

και

$$C^{(-1)}(u | k) = -\frac{1}{\theta} \log\left(1 + \frac{k(-1 + e^{-\theta})}{e^{-\theta u} - \theta(e^{-\theta u} - 1)}\right). \text{ (βλπ. Trivedi και Zimmer[30]).}$$

Τον παραπάνω αλγόριθμο που αφορά στην παραγωγή σημείων από τις συγκεκριμένες οικογένειες των Archimedean Copulas δίνεται στο Παράρτημα στο τέλος της εργασίας σε mathematica.⁶

⁵ Με $ProductLog[z]$ συμβολίζουμε τη λύση της εξίσωσης $z = we^w$ ως προς w .

⁶ Εναλλακτική μέθοδος προσομοίωσης δεδομένων από Archimedean Copula δίνεται σε άσκηση στο τέταρτο κεφάλαιο του βιβλίου του Nelsen [28], στην εργασία της Arnold [3] και στην εργασία των Embrechts, Lindskog και McNeil [10].

4.3 Μέθοδοι εκτίμησης των Copulas

Στην ενότητα αυτή θα παρουσιάσουμε τρεις μεθόδους εκτίμησης των Copulas. Αρχικά, δίνουμε μια παραμετρική μέθοδο εκτίμησης (*IFM*) (βλπ. Joe [24]), εν συνεχεία μία ημιπαραμετρική μέθοδο εκτίμησης (βλπ. Genest και Ghoudi [16]) και κλείνουμε την ενότητα με μία μη παραμετρική μέθοδο εκτίμησης των Copulas, η οποία στηρίζεται αποκλειστικά στα δεδομένα.⁷

4.3.1 Παραμετρική μέθοδος εκτίμησης ενός Copula

Η περίπτωση που θα μας απασχολήσει στην παρούσα εργασία είναι εκείνη στην οποία έχουμε ένα Copula απόλυτα συνεχές. Σε αυτήν την περίπτωση, όπως έχουμε αναφέρει και προηγουμένως, ορίζεται η πυκνότητα του Copula, c . Επομένως, έστω X_1, \dots, X_d συνεχείς τυχαίες μεταβλητές με περιθώριες συναρτήσεις κατανομής F_1, \dots, F_d και περιθώριες συναρτήσεις πυκνότητας f_1, \dots, f_d , αντίστοιχα, και Copula C . Τότε, η από κοινού πυκνότητά τους είναι της μορφής:

$$f(x_1, \dots, x_d) = c(F_1, \dots, F_d) \prod_{i=1}^d f_i(x_i)$$

όπου

$$c(F_1, \dots, F_d) = \frac{\partial^n C(F_1, \dots, F_d)}{\partial F_1 \cdots \partial F_d}.$$

Για την εκτίμηση της παραμέτρου του Copula μια πρώτη σκέψη θα ήταν η μεγιστοποίηση της συνάρτησης πιθανοφάνειας. Εδώ, όμως, πρέπει να λάβουμε υπ' όψιν ότι έχουμε και περιθώριες, άρα θα έχουμε συνολικά $d + 1$ παραμέτρους προς εκτίμηση, έστω $(a_1, \dots, a_d, \theta)$, όπου a_1, \dots, a_d είναι οι παράμετροι των d περιθωρίων και θ είναι η παράμετρος του Copula. Έστω, επίσης, ότι παίρνουμε ένα δείγμα με n παρατηρούμενα τυχαία διανύσματα $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. Συνηθίζεται να παίρνουμε το φυσικό λογάριθμο της συνάρτησης πιθανοφάνειας, καθώς δεν επηρεάζει στη μεγιστοποίηση της συνάρτησης και διευκολύνει τις πράξεις. Έχουμε ότι ο λογάριθμος της συνάρτησης πιθανοφάνειας θα είναι:

⁷ Παροτρύνουμε τον αναγνώστη για τις μεθόδους εκτίμησης των Copulas να απευθυνθεί και στις εργασίες των Choroś, Ibragimov και Permiakova η οποία συμπεριλαμβάνεται στο βιβλίο των Jaworski, Durante, Härdle, και Rychlik [23] καθώς και στην εργασία των Durrleman, Nikeghbali και Roncalli [8].

$$L = \sum_{j=1}^n \log f(x_1^j, \dots, x_d^j).$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας μπορεί να γραφεί και στην παρακάτω μορφή:

$$L = L_C + \sum_{i=1}^d L_i$$

όπου

$$L_C = \sum_{j=1}^n \log c(F_1(x_1^j), \dots, F_d(x_d^j))$$

είναι η συνεισφορά στην πιθανοφάνεια της δομής εξάρτησης των δεδομένων μέσω του Copula C και

$$L_i = \sum_{j=1}^n \log f_i(x_i^j)$$

είναι οι συνεισφορές στην πιθανοφάνεια από την κάθε περιθώρια.

Με τη μέθοδο εκτίμησης της μέγιστης πιθανοφάνειας αυτό που θα κάναμε θα ήταν να εκτιμήσουμε ταυτόχρονα όλες τις $d + 1$ παραμέτρους, δηλαδή θα βρίσκαμε τον εκτιμητή:

$$(\hat{a}_1, \dots, \hat{a}_d, \hat{\theta})$$

ο οποίος θα προέκυπε κάτω από συνθήκες ομαλότητας από τις λύσεις των εξισώσεων:

$$\left(\frac{\partial L}{\partial a_1}, \dots, \frac{\partial L}{\partial a_d}, \frac{\partial L}{\partial \theta} \right) = \mathbf{0}.$$

Παρ' όλα αυτά, όμως, δεν προτιμάται αυτή η μέθοδος αλλά έχει προταθεί μία άλλη μέθοδος η οποία κάνει τη μεγιστοποίηση σε δύο βήματα τα οποία είναι τα εξής:

- 1) Αρχικά εκτιμούμε τις παραμέτρους κάθε μίας περιθώριας ξεχωριστά με τη μέθοδο της μέγιστης πιθανοφάνειας και έπειτα
- 2) εκτιμούμε την παράμετρο του Copula με τη μέθοδο της μέγιστης πιθανοφάνειας, αφού πρώτα εισαγάγουμε τις εκτιμήσεις των παραμέτρων των περιθώριων που υπολογίσαμε στο βήμα 1.

Η παραπάνω μέθοδος ονομάζεται Method of Inference Functions for Margins (IFM). (βλπ. δέκατο κεφάλαιο του βιβλίου του Joe [24], δεύτερο κεφάλαιο της διατριβής του Xu [31], καθώς και την εργασία των Joe και Xu με τίτλο *The Estimation Method of Inference*

Functions for Margins for Multivariate Models [25]). Ο λόγος που προτιμάται από τη μέθοδο μέγιστης πιθανοφάνειας (ML) είναι ότι είναι πιο εύκολη στους υπολογισμούς καθώς γενικά είναι πιο ασύμφορο να γίνεται μία βελτιστοποίηση πολλών παραμέτρων ταυτοχρόνως παρά πολλές βελτιστοποιήσεις με λιγότερες, όμως, παραμέτρους κάθε φορά.

Επομένως, με τη μέθοδο (IFM) ο εκτιμητής είναι της μορφής:

$$(\tilde{a}_1, \dots, \tilde{a}_d, \tilde{\theta})$$

ο οποίος θα προέκυπτε κάτω από συνθήκες ομαλότητας από τις λύσεις των εξισώσεων:

$$\left(\frac{\partial L_1}{\partial a_1}, \dots, \frac{\partial L_d}{\partial a_d}, \frac{\partial L}{\partial \theta} \right) = \mathbf{0}.$$

Ο εκτιμητής αυτός θεωρείται αρκετά ικανοποιητικός σε σύγκριση με τον εκτιμητή μέγιστης πιθανοφάνειας. Στην περίπτωση που έχουμε ένα πολυδιάστατο Gaussian Copula με περιθώριες κανονικές αποδεικνύεται πως οι δύο εκτιμητές συμπέτουν. Αυτό δεν ισχύει εν γένει, αλλά ο IFM εκτιμητής ουσιαστικά χρησιμοποιεί την ML μέθοδο και άρα θα έχει κάποιες κοινές ιδιότητες. Πράγματι, κάτω από συνθήκες ομαλότητας για τη μεγιστοποίηση της πιθανοφάνειας, ισχύει ότι ο *MLE* είναι ασυμπτωτικά κανονικός και ο ασυμπτωτικός πίνακας διακυμάνσεων είναι ο αντίστροφος του πίνακα πληροφορίας του Fisher ενώ για την IFM μέθοδο ισχύει η ασυμπτωτική κανονικότητα με τη διαφορά ότι ο πίνακας διακυμάνσεων είναι ο αντίστροφος του πίνακα Godambe^{8,9} (βλπ. για περισσότερα θεωρητικά αποτελέσματα τις εργασίες του Godambe [21] και [22] καθώς και την εργασία του Xu[31]. Μπορεί να διαπιστωθεί αν οι δύο πίνακες είναι σχετικά κοντά αλλά δε θα επεκταθούμε παρ' όλα αυτά στα πλαίσια της παρούσας εργασίας.

⁸ Ο πίνακας Godambe είναι ο ασυμπτωτικός πίνακας συνδιακυμάνσεων, $V = D^{-1}M(D^{-1})^T$, ο οποίος συμπεριλαμβάνει παραγώγους συναρτήσεων συμπερασματολογίας \mathbf{g} , (inference functions), $D = E(\partial \mathbf{g}(\boldsymbol{\theta})^T / \partial \boldsymbol{\theta})$, $M = E(\mathbf{g}(\boldsymbol{\theta})^T \mathbf{g}(\boldsymbol{\theta}))$. Οι συναρτήσεις συμπερασματολογίας (inference functions) είναι συναρτήσεις τις οποίες θα τις θέσουμε ίσες με 0 και λύνοντας ως προς τις παραμέτρους θα παίρνουμε εκτιμητές. Μία τέτοια περίπτωση είναι και η μερική παράγωγος του λογαρίθμου της συνάρτησης πιθανοφάνειας.

⁹ Συστήνουμε στον αναγνώστη για περαιτέρω μελέτη την εργασία των Durlleman, Nikeghbali και Roncalli [8], το δέκατο κεφάλαιο του βιβλίου του Joe [24] και την εργασία των Joe, Xu [25].

4.3.2 Ημιπαραμετρική μέθοδος εκτίμησης ενός Copula

Στην παραπάνω υποενότητα δώσαμε μία παραμετρική μέθοδο εκτίμησης, γνωστή ως *IFM*, θεωρώντας πως γνωρίζουμε τις περιθώριες συναρτήσεις κατανομών. Υπάρχουν, όμως, περιπτώσεις που καθιστούν αδύνατη αυτή τη γνώση καθώς μπορεί τα δεδομένα μας να έχουν για παράδειγμα ασυμμετρίες και βαριές ουρές και να μην είμαστε σε θέση να αποφανθούμε για το ποιες είναι οι κατανομές αυτές. Σε αυτή την περίπτωση μπορούμε να καταφύγουμε στην παρακάτω μέθοδο (βλπ. Genest, Ghoudi και Rivest [16], καθώς και Accioly, Chiyoshi [2]) που αποτελείται και αυτή από δύο βήματα:

1) Αρχικά, θεωρούμε ότι οι περιθώριες δεν είναι γνωστές, για αυτόν τον λόγο θα πρέπει να επιλέξουμε μία εκτίμησή τους. Συνηθίζεται να επιλέγουμε τις εμπειρικές συναρτήσεις κατανομών, οι οποίες δίνονται από τον τύπο:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq x),$$

όπου $1(X_i \leq x)$ είναι η δείκτρια συνάρτηση.

2) Εκτιμούμε με τη μέθοδο της μέγιστης πιθανοφάνειας την παράμετρο του Copula.

Η συγκεκριμένη μέθοδος ονομάζεται Canonical Maximum Likelihood (*CML*). Συνήθως έχουμε αρκετά δεδομένα για να πάρουμε τις εμπειρικές κατανομές, δηλαδή μια μη παραμετρική εκτίμηση των περιθώριων, αλλά όχι αρκετή πληροφορία για μη παραμετρική εκτίμηση της δομής εξάρτησης. Επομένως, έστω (X_{1k}, \dots, X_{pk}) , όπου $k = 1, \dots, n$, ένα δείγμα μεγέθους n . Ο λογάριθμος της ψευδοσυνάρτησης πιθανοφάνειας για τον παραπάνω αλγόριθμο είναι:

$$L(\theta) = \sum_{k=1}^n \log(c_\theta(F_{1k}(x_{1k}), \dots, F_{pk}(x_{pk}))),$$

όπου F_{in} είναι $\frac{n}{n+1}$ την περιθώρια εμπειρική συνάρτηση κατανομής της i τυχαίας μεταβλητής.

Ο εκτιμητής της παραμέτρου θ του Copula είναι:

$$\hat{\theta}_n = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \sum_{k=1}^n \log c_\theta(F_{1k}(x_{1k}), \dots, F_{pk}(x_{pk})).$$

Ο $\hat{\theta}_n$ αποδεικνύεται πως κάτω από συνηθισμένες συνθήκες ομαλότητας είναι συνεπής και ασυμπτωτικά κανονικός εκτιμητής. (βλπ. Genest και Ghoudi [16]).

Αν είχαμε δεδομένα και θέλαμε να εφαρμόσουμε την παραπάνω μέθοδο θα γράφαμε το λογάριθμο της ψευδοσυνάρτησης πιθανοφάνειας συναρτήσεως των ranks των δύο τυχαίων μεταβλητών ως εξής:

$$L(\theta) = \sum_{i=1}^n \log\{c_{\theta}(\frac{R_i}{n+1}, \frac{S_i}{n+1})\},$$

όπου R_i, S_i είναι τα ranks των X_i, Y_i . Ο CML εκτιμητής θα προέκυπτε ως ρίζα της εξίσωσης:

$$l(\theta) = \frac{\partial L(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{\frac{\partial}{\partial \theta} c_{\theta}(\frac{R_i}{n+1}, \frac{S_i}{n+1})}{c_{\theta}(\frac{R_i}{n+1}, \frac{S_i}{n+1})} = 0.$$

Μία ακόμα ημιπαραμετρική μέθοδος εκτίμησης των Copulas μπορεί να θεωρηθεί η μέθοδος στην οποία χρησιμοποιούμε τους εκτιμητές των συντελεστών συσχέτισης για να εκτιμήσουμε την παράμετρο εξάρτησης του Copula. Η συγκεκριμένη μέθοδος δεν μπορεί να θεωρηθεί παραμετρική καθώς δεν θα χρησιμοποιήσουμε τις περιθώριες συναρτήσεις παρά μόνο τα ranks των δεδομένων μας, αλλά ούτε μη παραμετρική καθώς για να συνδέσουμε τα τ, ρ με το θ πρέπει να έχουμε επιλέξει συγκεκριμένη οικογένεια των Copulas.

Αν, επομένως, υπάρχει μία σχέση της μορφής $\theta = g(\tau)$, (αντίστοιχα $\theta = h(\rho)$), που να συνδέει την παράμετρο εξάρτησης του Copula με το τ του Kendall, (αντίστοιχα με το ρ του Spearman), τότε ένας λογικός εκτιμητής θα ήταν:

$$\tilde{\theta}_n = g(\tau_n) \text{ ή } \tilde{\theta}_n = h(\rho_n).$$

α. Για το τ του Kendall και καθώς $n \rightarrow \infty$ έχουμε ότι:

$$\sqrt{n}(\tau_n - \tau) \sim N(0, \sigma^2)$$

όπου σ^2 είναι η διασπορά της ασυμπτωτικής κανονικής κατανομής του εκτιμητή. (βλπ. Genest και Rivest [20]). Έστω g μία παραγωγίσιμη συνάρτηση με $g'(\tau) \neq 0$. Τότε από τη μέθοδο Δέλτα προκύπτει ότι:

$$\sqrt{n}(g(\tau_n) - g(\tau)) \sim N(0, \sigma^2(g'(\tau))^2).$$

Ένας συνεπής εκτιμητής για το σ^2 είναι ο $16S^2$ (βλπ. Genest και Rivest [20]), όπου:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (W_i + \tilde{W}_i - 2\bar{W})^2,$$

$$W_i = \frac{1}{n} \sum_{j=1}^n I_{ij} = \frac{1}{n} \#\{j : X_j \leq X_i, Y_j \leq Y_i\},$$

$$\tilde{W}_i = \frac{1}{n} \sum_{j=1}^n I_{ji} = \frac{1}{n} \#\{j : X_i \leq X_j, Y_i \leq Y_j\}$$

και

$$\bar{W} = \frac{1}{n} \sum_{i=1}^n W_i.$$

Για $\theta = g(\tau)$, $\tilde{\theta}_n = g'(\tau_n)$, $16S^2 \xrightarrow{p} \sigma^2$ και από το Θεώρημα Slutsky έχουμε:

$$\sqrt{n} \frac{(\tilde{\theta}_n - \theta)}{4S} = \frac{\sqrt{n}(\tilde{\theta}_n - \theta)/\sigma g'(\tau) \xrightarrow{d} N(0, 1)}{4S g'(\tau)/\sigma g'(\tau) \xrightarrow{p} 1} \xrightarrow{d} N(0, 1).$$

Επομένως, καθώς το $n \rightarrow \infty$ ισχύει ότι:

$$\tilde{\theta}_n \sim N(\theta, \frac{1}{n}(4S g'(\tau))^2).$$

Ένα προσεγγιστικό $100(1 - \alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για το θ είναι:

$$\tilde{\theta}_n \pm z_{\alpha/2} \frac{1}{\sqrt{n}} 4S |g'(\tau_n)|.$$

β. Για το ρ του Spearman και καθώς $n \rightarrow \infty$ έχουμε ότι:

$$\sqrt{n}(\rho_n - \rho) \sim N(0, \sigma^2)$$

όπου σ^2 είναι η διασπορά της ασυμπτωτικής κανονικής κατανομής του εκτιμητή. (βλπ. Genest και Favre [15]). Έστω h μία παραγωγίσιμη συνάρτηση με $h'(\rho) \neq 0$. Τότε από τη μέθοδο Δέλτα προκύπτει ότι:

$$\sqrt{n}(h(\rho_n) - h(\rho)) \sim N(0, \sigma^2(h'(\rho))^2).$$

Ένας συνεπής εκτιμητής για το σ^2 είναι ο $\sigma_n^2 = 144(-9A_n^2 + B_n + 2C_n + 2D_n + 2E_n)$, όπου:

$$A_n = \frac{1}{n} \sum_{i=1}^n \frac{R_i}{n+1} \frac{S_i}{n+1},$$

$$B_n = \frac{1}{n} \sum_{i=1}^n \left(\frac{R_i}{n+1}\right)^2 \left(\frac{S_i}{n+1}\right)^2,$$

$$C_n = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{R_i}{n+1} \frac{S_i}{n+1} 1(R_k \leq R_i, S_k \leq S_j) + \frac{1}{4} - A_n,$$

$$D_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{S_i}{n+1} \frac{S_j}{n+1} \max\left(\frac{R_i}{n+1}, \frac{R_j}{n+1}\right)$$

και

$$E_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{R_i}{n+1} \frac{R_j}{n+1} \max\left(\frac{S_i}{n+1}, \frac{S_j}{n+1}\right).$$

Για $\theta = h(\rho)$, $\tilde{\theta}_n = h'(\rho_n)$, $\sigma_n^2 \xrightarrow{p} \sigma^2$ και από το Θεώρημα Slutsky έχουμε:

$$\sqrt{n} \frac{(\tilde{\theta}_n - \theta)}{\sigma_n} = \frac{\sqrt{n}(\tilde{\theta}_n - \theta)/\sigma h'(\rho) \xrightarrow{d} N(0, 1)}{\sigma_n h'(\rho)/\sigma h'(\rho) \xrightarrow{p} 1} \xrightarrow{d} N(0, 1).$$

Επομένως, καθώς το $n \rightarrow \infty$:

$$\tilde{\theta}_n \sim N\left(\theta, \frac{1}{n}(\sigma_n h'(\rho))^2\right).$$

Ένα προσεγγιστικό $100(1 - \alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για το θ είναι:

$$\tilde{\theta}_n \pm z_{\alpha/2} \frac{1}{\sqrt{n}} \sigma_n |h'(\rho_n)|.$$

ΠΑΡΑΔΕΙΓΜΑ¹⁰: Έστω δύο τυχαία δείγματα μεγέθους $n = 6$ από την $N(0, 1)$ (Πίνακας 1.1) και ότι η συνάρτηση εξάρτησης για το ζεύγος (X, Y) είναι της οικογένειας των Farlie-Gumbel-Morgenstern. Τότε το Copula θα είναι της μορφής:

$$C_\theta(u, v) = uv + \theta uv(1 - u)(1 - v)$$

και η πυκνότητά του:

$$c_\theta(u, v) = \frac{\partial^2 C_\theta(u, v)}{\partial u \partial v} = 1 + \theta(1 - 2u)(1 - 2v).$$

Οι σχέσεις που συνδέουν τους συντελεστές συσχέτισης με την παράμετρο εξάρτησης της συγκεκριμένης οικογένειας είναι:

$$\tau = \frac{2}{9}\theta \quad \text{και} \quad \rho = \frac{\theta}{3}.^{11}$$

α. Για το τ του Kendall έχουμε $g(\tau) = \frac{9}{2}\tau$, $g'(\tau) = \frac{9}{2} \neq 0$ και W_i, \tilde{W}_i όπως δίνονται στον παρακάτω πίνακα:

¹⁰ Το συγκεκριμένο παράδειγμα συμπεριλαμβάνεται στην εργασία των Genest και Favre [15].

¹¹ Σελίδες 38 και 42 της παρούσας εργασίας.

Πίνακας 4.1: Τα W_i, \tilde{W}_i για τον υπολογισμό του συνεπή εκτιμητή της διασποράς που σχετίζεται με το Tau του Kendall.

i	1	2	3	4	5	6
W_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{6}{6}$
\tilde{W}_i	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{4}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

Έχουμε $\bar{W} = \frac{13}{36}$, $S^2 = 0.063$ και επομένως $S = 0.251$. Επειδή η εκτίμηση του τ σύμφωνα με τα παραπάνω δεδομένα είναι $\tau_n = -\frac{1}{15} = -0.0667$,¹² η εκτίμηση για την παράμετρο εξάρτησης θα είναι $\tilde{\theta}_n = \frac{9}{2}(-\frac{1}{15}) = -0.3$ και επομένως, σε επίπεδο στατιστικής σημαντικότητας $\alpha = 0.05$, ένα προσεγγιστικό $100(1 - \alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για το θ θα είναι το $[-0.3 \pm 1.96 \frac{1}{2.449} 4S \frac{9}{2}] = [-3.916, 3.316]$, το οποίο και περιορίζουμε στο $[-1, 1]$. Βέβαια, αυτό είναι γνωστό εξ αρχής καθώς το θ είναι παράμετρος εξάρτησης και επομένως δεν έχουμε κάποια πρόσθετη πληροφορία.¹³

β. Για το ρ του Spearman έχουμε $h(\rho) = 3\rho$, $h'(\rho) = 3 \neq 0$. Έπειτα από υπολογισμούς βρήκαμε ότι $A_n = 0.2517$, $B_n = 0.1137$, $C_n = 0.1152$, $D_n = 0.1657$, $E_n = 0.1626$ και $\sigma_n^2 = 61.992$ και επομένως $\sigma_n = 7.8735$. Επειδή η εκτίμηση του ρ_S με τα παραπάνω δεδομένα είναι $\rho_n = 0.02857$,¹⁴ η εκτίμηση για την παράμετρο εξάρτησης θα είναι $\tilde{\theta}_n = 3 \times 0.02857 = 0.08571$ και επομένως, για $\alpha = 0.05$, ένα προσεγγιστικό $100(1 - \alpha)\%$ ασυμπτωτικό διάστημα εμπιστοσύνης για το θ θα είναι το $[0.08571 \pm 1.96 \frac{1}{\sqrt{6}} \sigma_n 3] = [-18.8184, 18.9898]$, το οποίο και περιορίζουμε στο $[-1, 1]$. Και εδώ δεν έχουμε κάποια πρόσθετη πληροφορία από το εύρος τιμών του θ .

¹² Σελίδα 46 της παρούσας εργασίας.

¹³ Τα όρια του $[-1, 1]$ τα ξεπερνάμε καθώς το τυπικό σφάλμα είναι σχετικά μεγάλο (3.616), αλλά αυτό ήταν αναμενόμενο αφού το δείγμα μας αποτελείται από 6 παρατηρήσεις.

¹⁴ Σελίδα 46 της παρούσας εργασίας.

4.3.3 Μη παραμετρική μέθοδος εκτίμησης ενός Copula

Στην παραπάνω υποενότητα δώσαμε μια ημιπαραμετρική μέθοδο εκτίμησης των Copulas, κι αυτό γιατί, ενώ ξεκινήσαμε με μη παραμετρική εκτίμηση των περιθώριων συναρτήσεων κατανομής, η εκτίμηση της παραμέτρου του Copula προϋποθέτει τη γνώση του Copula. Σε αυτήν την υποενότητα, θα θεωρήσουμε ότι δεν γνωρίζουμε ούτε το Copula και επομένως θα το εκτιμήσουμε βάσει των δεδομένων που θα έχουμε κάθε φορά. Με απλά λόγια, η μη παραμετρική μέθοδος εκτίμησης των Copulas που παρουσιάζουμε δεν είναι τίποτα άλλο από το Εμπειρικό Copula.

ΟΡΙΣΜΟΣ: Το εμπειρικό Copula, όπως αρχικά ορίστηκε από τον Deheuvels (βλπ. κεφάλαιο 5 του βιβλίου των Cherubini, Luciano και Vecchiato [6] και την εργασία των Durrleman, Nikeghbali, Roncalli [8]), είναι οποιοδήποτε Copula με πεδίο ορισμού το: $L = \{(t_1, \dots, t_N) : 1 \leq n \leq N, t_n = 0, \dots, T\}$ για το οποίο ισχύει:

$$\hat{C}(t_1, \dots, t_N) = \frac{1}{T} \sum_{t=1}^T \prod_{n=1}^N 1(r_n^t \leq t_n)$$

όπου r_n είναι το n -διατεταγμένο rank ενός δείγματος (x_1, \dots, x_n) μεγέθους T . □

Η πυκνότητα του εμπειρικού Copula δίνεται από τη σχέση:

$$\hat{c}(t_1, \dots, t_N) = \sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{\sum_{j=1}^n i_j} \hat{C}(t_1 - i_1 + 1, \dots, t_n - i_n + 1).$$

Στην δισδιάστατη περίπτωση, έστω X, Y δύο τυχαίες μεταβλητές και $\{(x_k, y_k)\}_{k=1}^n$ τυχαίο δείγμα n παρατηρήσεων. Το εμπειρικό Copula δίνεται από τον παρακάτω τύπο:

$$\hat{C}(u, v) = \frac{1}{n} \sum_{i=1}^n 1(U_i \leq u, V_i \leq v)$$

όπου (u, v) είναι η παρατηρούμενη τιμή του διανύσματος $(U, V) = (\hat{F}(X), \hat{G}(Y))$ και τα U_i, V_i είναι τα ranks των παρατηρήσεων του i διανύσματος διαιρεμένα με το πλήθος των παρατηρήσεων συν 1.¹⁵

¹⁵ Δεν είναι ακριβώς το εμπειρικό Copula που ορίσαμε παραπάνω, αλλά αυτό χρησιμοποιείται στην πράξη σύμφωνα με τους Genest, Rémillard και Beaudoin [19].

4.4 Έλεγχοι καλής προσαρμογής (Goodness-of-fit tests)

Σε αυτήν την ενότητα θα συζητήσουμε κάποιους έλεγχοι καλής προσαρμογής για τα Copula οι οποίοι στηρίζονται σε ήδη γνωστές έννοιες. Οι έλεγχοι που επιθυμούμε να κάνουμε είναι της μορφής:

$$H_0 : C \in C_\Theta \text{ έναντι της } H_1 : C \notin C_\Theta,$$

όπου $C_\Theta = \{C_\theta : \theta \in \Theta\}$ είναι κάποια από τις γνωστές κλάσεις των Copulas. Να τονίσουμε πως ασχολούμαστε με μονοπαραμετρικές οικογένειες και σε αυτήν την ενότητα και στην προηγούμενη.¹⁶ Θα πρέπει να γίνει κατανοητό πως είναι διαφορετικό να εκτιμούμε το Copula από το να εξετάζουμε αν το Copula ανήκει σε κάποια γνωστή παραμετρική οικογένεια. Στην περίπτωση του παραπάνω ελέγχου, για παράδειγμα, δεν έχει ουσιαστικό νόημα να ασχοληθούμε με τις παραμέτρους των περιθώριων καθώς, σε αυτήν την περίπτωση, θα καταλήγαμε να είχαμε ένα πιο αυστηρό έλεγχο ο οποίος θα ενείχε και τον έλεγχο προσαρμογής των περιθώριων (βλπ. Genest, Rémillard και Beaudoin [19]). Για αυτόν το λόγο οι έλεγχοι που θα παρουσιάσουμε θα βασίζονται στα ranks των παρατηρήσεων διαιρεμένα με το πλήθος των παρατηρήσεων συν 1.

Στην δισδιάστατη περίπτωση, όπου έχουμε δύο συνεχείς τυχαίες μεταβλητές X, Y , λαμβάνουμε δείγμα μεγέθους n και κάνουμε τους μετασχηματισμούς $U_i = \frac{R_i}{n+1}$ και $V_i = \frac{S_i}{n+1}$, όπου R_i, S_i είναι τα ranks της i παρατήρησης των X και Y αντίστοιχα. Επομένως, θα μπορούσαμε να πούμε ότι το εμπειρικό Copula, όπως ορίστηκε στην προηγούμενη υποενότητα, είναι το Copula, C , του παραπάνω ελέγχου. Αυτό που θα είχε νόημα είναι να ορίζαμε ένα είδος απόστασης μεταξύ του εμπειρικού Copula και του παραμετρικού Copula του οποίου την παράμετρο την έχουμε εκτιμήσει από μια ημιπαραμετρική μέθοδο. Μια τέτοια απόσταση θα είναι της μορφής:

$$\mathbb{C}_n = \sqrt{n}(C_n - C_{\hat{\theta}_n})$$

όπου $\hat{\theta}_n$ είναι ο εκτιμητής της παραμέτρου του Copula. Δύο στατιστικές συναρτήσεις που προτείνουν οι Genest, Rémillard και Beaudoin [19] είναι αντίστοιχες των συναρτήσεων των

¹⁶ Στην περίπτωση του t -Copula μπορούμε να θεωρήσουμε ότι είτε οι βαθμοί ελευθερίας είναι γνωστοί είτε ο πίνακας συσχετίσεων.

Cramér-von Mises και των Kolmogorov-Smirnov και δίνονται από τις παρακάτω σχέσεις:

$$S_n = \int_0^1 \int_0^1 \mathbb{C}(u, v)^2 dC_n(u, v)$$

και

$$T_n = \sup_{(u,v) \in [0,1]^2} | \mathbb{C}_n(u, v) |$$

αντίστοιχα. Μεγάλες τιμές των παραπάνω στατιστικών συναρτήσεων συνηγορούν στην απόρριψη της μηδενικής υπόθεσης.

Παίρνοντας διαφορετικά είδη αποστάσεων μπορούμε να φτιάξουμε διαφορετικές στατιστικές συναρτήσεις. Για παράδειγμα, θα μπορούσαμε να έχουμε στατιστική συνάρτηση με ευκλείδια νόρμα ή τη στατιστική συνάρτηση των Anderson-Darling. Επίσης, μπορούμε να χρησιμοποιήσουμε τις στατιστικές συναρτήσεις για να κάνουμε έλεγχο ανεξαρτησίας, αν στη θέση του C_{θ_n} βάλουμε το Copula-γινόμενο.

Αντί για τον παραπάνω έλεγχο θα μπορούσαμε να κάνουμε έναν έλεγχο της μορφής:

$$H_0 : K \in \mathcal{K}_\Theta \text{ έναντι της } H_1 : K \notin \mathcal{K}_\Theta,$$

όπου $\mathcal{K}_\Theta = \{K_\theta : \theta \in \Theta\}$ είναι το σύνολο με τις συναρτήσεις κατανομής που βασίζονται στο τ του Kendall.

Υπενθυμίζουμε ότι αν X, Y είναι δύο ομοιόμορφες¹⁷ τυχαίες μεταβλητές με συνάρτηση εξάρτησης ένα Arcimedean Copula και χαρακτηριστικό γεννήτορα ϕ , τότε η τυχαία μεταβλητή V η οποία προκύπτει από τον μετασχηματισμό $V = C(X, Y)$ έχει συνάρτηση κατανομής

$$K_\theta(v) = v - \frac{\phi(v)}{\phi'(v)}.$$

Οι Genest και Rivest πρότειναν έναν μη παραμετρικό εκτιμητή για το $K_\theta(t)$ ο οποίος είναι ουσιαστικά μια εμπειρική συνάρτηση κατανομής η οποία δίνει βάρος $\frac{1}{n}$ στις εξής παρατηρήσεις:

$$V_i = \frac{1}{n-1} \#\{j : X_j < X_i, Y_j < Y_i\}.$$

¹⁷ Χωρίς βλάβη της γενικότητας, θεωρούμε τις τυχαίες μεταβλητές X, Y ομοιόμορφες, καθώς στη μέθοδο που παρουσιάζουμε δεν γίνεται χρήση των περιθώριων συναρτήσεων κατανομής.

Για να καταλάβουμε πως συνδέονται με το τ του Kendall αρκεί να ανατρέξουμε στον ορισμό της συμφωνίας για ένα δείγμα. Αν ορίσουμε:

$$I_{ij} = \begin{cases} 1, & \text{αν } X_j < X_i \text{ και } Y_j < Y_i \\ 0, & \text{διαφορετικά} \end{cases}$$

τότε το ζεύγος των i και j παρατηρήσεων είναι σε συμφωνία αν $I_{ij} + I_{ji} = 1$. Το Tau του Kendall είναι τα $C - D$ διαιρεμένα με $\frac{n(n-1)}{2}$. Αν τώρα γράψουμε τα D ως $\frac{n(n-1)}{2} - C$ και τα C ως $\sum_{ij} I_{ij} = (n-1) \sum_i V_i$, προκύπτει η εξής σχέση για το τ_n :

$$\tau_n = 4\bar{V} - 1$$

όπου

$$\bar{V} = \frac{\sum_{i=1}^n V_i}{n}.$$

Όλα τα παραπάνω ισχύουν και στην περίπτωση που δεν έχουμε Archimedean Copula και αυτό οφείλεται στον τρόπο κατασκευής του εκτιμητή. (βλπ. Genest και Rivest [20]).

Για να επανέλθουμε, επομένως, στους ελέγχους καλής προσαρμογής θα θεωρήσουμε ότι έχουμε δύο συνεχείς τυχαίες μεταβλητές X, Y με περιθώριες συναρτήσεις κατανομής F, G , από κοινού συνάρτηση κατανομής H και συνάρτηση εξάρτησης το Copula C . Τότε κάνουμε τον εξής μετασχηματισμό:

$$V = H(X, Y) = C(U_1, U_2)$$

όπου $U_1 = F(X) \sim U(0, 1)$, $U_2 = G(Y) \sim U(0, 1)$, η από κοινού των U_1, U_2 είναι το C και θεωρούμε K να είναι η συνάρτηση κατανομής του V . Αν πάρω δείγματα μεγέθους n από τις X, Y τότε ένας συνεπής εκτιμητής του K είναι ο εξής:

$$K_n(v) = \frac{1}{n} \sum_{i=1}^n 1(V_i \leq v)$$

όπου $V_i = C_n(\mathbf{U}_i)$ και σε κατάλληλη κλίμακα ώστε να παίρνει ο εκτιμητής τιμές εντός του $(0,1)$. Τα \mathbf{U}_i είναι διανύσματα τα οποία εμπεριέχουν τα ranks των i παρατηρήσεων διαιρεμένα με το $n + 1$.

Αυτό που θα είχε νόημα, επομένως, θα ήταν να ορίζαμε μια απόσταση μεταξύ του εμπειρικού K_n και του παραμετρικού K_{θ_n} , όπου θ_n είναι η εκτίμηση που έχουμε πάρει από κάποια ημιπαραμετρική μέθοδο για την παράμετρο θ . Μια τέτοια απόσταση θα είναι της μορφής:

$$\mathbb{K}_n = \sqrt{n}(K_n - K_{\theta_n}).$$

Δύο στατιστικές συναρτήσεις που προτείνουν οι Genest, Rémillard και Quessy [18] είναι, επίσης, αντίστοιχες των συναρτήσεων των Cramér-von Mises και των Kolmogorov-Smirnov και είναι οι εξής:

$$S_n^{(K)} = \int_0^1 |\mathbb{K}_n(t)|^2 dK_{\theta_n}(t)$$

και

$$T_n^{(K)} = \sup_{t \in [0,1]} |\mathbb{K}_n(t)|$$

αντίστοιχα. Μεγάλες τιμές των παραπάνω στατιστικών συναρτήσεων συνηγορούν στην απόρριψη της μηδενικής υπόθεσης. Στην περίπτωση που δεν απορρίψουμε τη μηδενική υπόθεση δεν μπορούμε να αποδεχθούμε τη μηδενική υπόθεση που αφορά τις οικογένειες των Copulas (βλπ. Genest, Rémillard και Beaudoin [19]). Για αυτόν τον λόγο θέλει ιδιαίτερη προσοχή η ερμηνεία όταν χρησιμοποιείται ο εκτιμητής της συνάρτησης K .¹⁸

¹⁸ Στην παρούσα εργασία ασχολούμαστε με τη διδιάστατη περίπτωση αλλά γενικεύσεις των παραπάνω στις n διαστάσεις μπορεί να τις βρει ο αναγνώστης στην εργασία των Genest, Rémillard και Beaudoin [19].

Κεφάλαιο 5

ΕΦΑΡΜΟΓΗ ΣΕ ΠΡΟΣΟΜΟΙΩΜΕΝΑ ΔΕΔΟΜΕΝΑ

5.1 ΕΙΣΑΓΩΓΗ

Στο προηγούμενο κεφάλαιο είδαμε τρόπους εκτίμησης της παραμέτρου του Copula, καθώς και ελέγχους καλής προσαρμογής. Αυτό που θα επιχειρήσουμε είναι να δούμε κάποια αποτελέσματα πρακτικά, χρησιμοποιώντας προσομοιωμένα δεδομένα. Στον πραγματικό κόσμο συνηθίζεται να έχουμε πραγματικά δεδομένα και, επομένως, είναι χρήσιμο να περιγράψουμε τη διαδικασία την οποία θα ακολουθούσαμε στην αντίστοιχη περίπτωση.

Το πρώτο βήμα που κάνουμε, όταν χειριζόμαστε πραγματικά δεδομένα, είναι να δούμε ένα γράφημα καθώς και κάποια από τα περιγραφικά στοιχεία τους. Από αυτά μπορούμε να έχουμε μια πρώτη άποψη για τη συσχέτιση, για το αν υπάρχει συμμετρία ή όχι, αν προέρχονται από λεπτόκυρτη ή πλατύκυρτη κατανομή κ.ο.κ. Το δεύτερο βήμα που πρέπει να γίνει είναι να επιλεγεί η κατάλληλη οικογένεια των Copulas που προσαρμόζεται καλύτερα στα δεδομένα μας. Η επιλογή του Copula είναι αρκετά δύσκολη και απαιτεί γνώση πολλών οικογενειών, των ιδιοτήτων τους καθώς και εμπειρία. Το τρίτο και τελευταίο βήμα είναι να εκτιμηθεί η παράμετρος του Copula. Στην περίπτωση των πραγματικών δεδομένων, θα πρέπει να επιλέξουμε περιθώριες συναρτήσεις, αν θέλουμε να χρησιμοποιήσουμε την *ML* ή την *IFM* μέθοδο. Αν είναι δύσκολο να επιλέξουμε τις κατάλληλες περιθώριες, η εκτίμηση της παραμέτρου του Copula γίνεται με τη μέθοδο *CML*.

Στο παρόν κεφάλαιο θα δώσουμε μία εφαρμογή κατά την οποία θα πάρουμε προσο-

μοιωμένα δεδομένα από ένα Copula, εφαρμόζοντας θεωρητικά αποτελέσματα που δόθηκαν στο προηγούμενο κεφάλαιο. Συγκεκριμένα, θα κάνουμε έλεγχο καλής προσαρμογής με τη μέθοδο των Cramér-von Mises χρησιμοποιώντας οικογένειες που παρουσιάστηκαν στην παρούσα εργασία και, αφού επιλεγεί η κατάλληλη οικογένεια, θα χρησιμοποιήσουμε τις μεθόδους *ML*, *IFM* και *CML* για να εκτιμήσουμε την παράμετρο του Copula. Τέλος θα προσπαθήσουμε να αποφανθούμε ποια από τις τρεις αυτές μεθόδους δίνει καλύτερη εκτίμηση βάσει των συγκεκριμένων δεδομένων. Η επιλογή του Copula καθώς και των περιθώριων έγινε τυχαία και ο αλγόριθμος που έχει χρησιμοποιηθεί δίνεται στα Παραρτήματα.

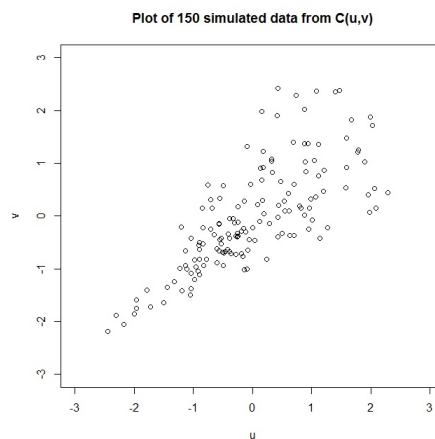
5.2 Εφαρμογή

Έστω το Clayton Copula, $C(x, y)$, με παράμετρο ίση με 3. Θα κατασκευάσουμε τη συνάρτηση κατανομής η οποία προκύπτει από το Clayton Copula με περιθώριες Τυποποιημένες Κανονικές Κατανομές, $N(0, 1)$, δηλαδή την

$$H(x, y) = C(u, v),$$

όπου $u = F(x)$, $X \sim N(0, 1)$ και $v = G(y)$, $Y \sim N(0, 1)$. Παράγουμε 150 δεδομένα από το παραπάνω Copula, τα οποία φαίνονται στο παρακάτω γράφημα:

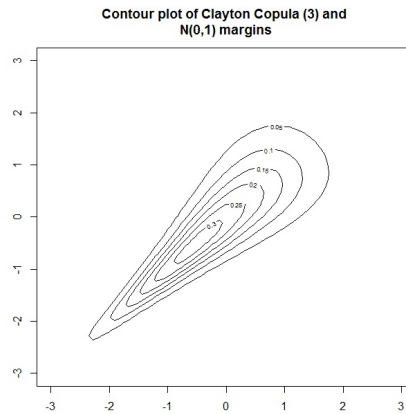
Σχήμα 5.1: Το γράφημα με τα προσομοιωμένα δεδομένα από το Clayton Copula με παράμετρο 3 και περιθώριες $N(0, 1)$.



Όπως παρατηρούμε στο γράφημα τα δεδομένα μας φαίνονται να έχουν θετική συσχέτιση, το οποίο και ισχύει καθώς ο συντελεστής συσχέτισης είναι ίσος με 0.75 και μάλιστα η

συσχέτιση αυτή είναι πιο ισχυρή στην κάτω ουρά. Παρακάτω δίνουμε και το Contour plot που προκύπτει από τα δεδομένα μας.

Σχήμα 5.2: Το Contour Plot ενός Clayton Copula με παράμετρο 3 και περιθώριες $N(0, 1)$.



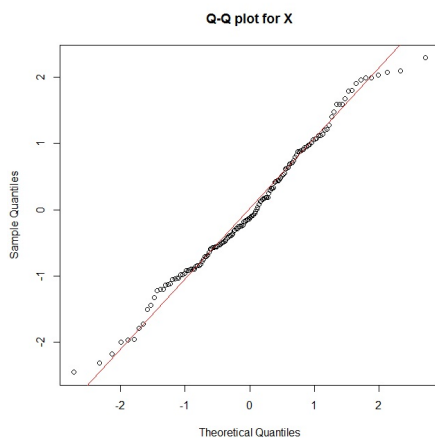
Αν θεωρούσαμε ότι έχουμε πραγματικά δεδομένα, θα υπολογίζαμε και τα περιγραφικά στοιχεία των δύο τυχαίων μεταβλητών X, Y , τα οποία δίνονται στον παρακάτω πίνακα:

Πίνακας 5.1: Τα περιγραφικά στοιχεία των τυχαίων μεταβλητών X, Y .

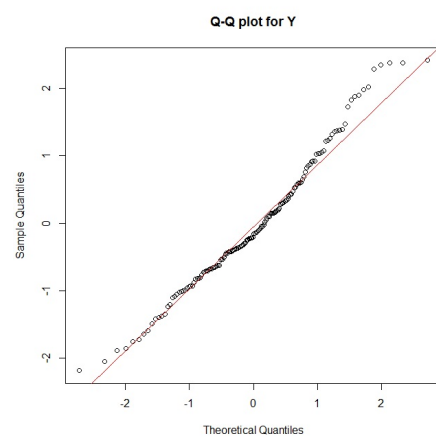
	X	Y
Number	150	150
Mean	0.004775	-0.02494
Median	-0.12800	-0.17870
Standard deviation	1.015971	0.9904056
Minimum	-2.45200	-2.18700
Maximum	2.28700	2.41400
25th Percentile	-0.699700	-0.67540
75th Percentile	0.731200	0.56560
Skewness	0.1122987	0.4348660
Kurtosis	2.626097	2.932646

Αυτά που αναμέναμε είναι μέσες τιμές κοντά στην τιμή 0 και τυπικές αποκλίσεις κοντά στην τιμή 1. Μεταξύ του πρώτου και του τρίτου τεταρτημορίου, όπως είναι γνωστό, θα βρίσκεται το 50% των παρατηρήσεων. Επίσης, η διάμεσος έχει μια σχετικά κοντινή τιμή με το μέσο και, επομένως, θα υποπτευόμασταν συμμετρία. Αυτό φαίνεται καλύτερα από το μέτρο ασυμμετρίας που παίρνει τιμές κοντά στο 0. Τέλος, όσον αφορά την κύρτωση, αναμένουμε τιμή κοντινή στην τιμή 3. Αυτά ισχύουν καθώς οι τυχαίες μεταβλητές από κατασκευής ακολουθούν την $N(0, 1)$.

Αν δεν γνωρίζαμε τις περιθώριες, θα έπρεπε να κάνουμε έλεγχο για να επιλέξουμε τις πιο κατάλληλες περιθώριες οι οποίες θα περιγράφουν καλύτερα τα δεδομένα. Αρχικά, θα μπορούσαμε να πάρουμε $Q - Q$ plots με τα οποία να συγκρίνουμε τα ποσοστημόρια των δεδομένων μας με τα ποσοστημόρια της κατανομής που επιλέγουμε να κάνουμε τη σύγκριση.



Σχήμα 5.3: Το $Q - Q$ plot της τυχαίας μεταβλητής X με την Κανονική Κατανομή.



Σχήμα 5.4: Το $Q - Q$ plot της τυχαίας μεταβλητής Y με την Κανονική Κατανομή.

Αυτό που θα λέγαμε βάσει των παραπάνω διαγραμμάτων είναι ότι μάλλον οι τυχαίες μεταβλητές ακολουθούν την Κανονική Κατανομή. Επειδή, όμως, τα διαγράμματα δίνουν μόνο μια πρώτη εικόνα, θα επιλέξουμε να κάνουμε έναν έλεγχο Kolmogorov-Smirnov και

θα χρησιμοποιήσουμε και το Lilliefors test.¹ Τα αποτελέσματα δίνονται παρακάτω:

One-sample Kolmogorov-Smirnov test

data: X

D = 0.06, p-value = 0.6524

alternative hypothesis: two-sided

Lilliefors (Kolmogorov-Smirnov) normality test

data: X

D = 0.06, p-value = 0.2066

Επομένως, σε επίπεδο σημαντικότητας $\alpha = 0.05$ η υπόθεση της κανονικότητας για την τυχαία μεταβλητή X δεν απορρίπτεται καθώς το p – value του ελέγχου είναι μεγαλύτερο από το α . Το ίδιο ισχύει και για την τυχαία μεταβλητή Y όπως δείχνουν τα παρακάτω αποτελέσματα:

One-sample Kolmogorov-Smirnov test

data: Y

D = 0.0719, p-value = 0.4197

alternative hypothesis: two-sided

Lilliefors (Kolmogorov-Smirnov) normality test

data: Y

D = 0.0719, p-value = 0.05555

Προφανώς, όλα τα παραπάνω που αφορούν τις περιθώριες είναι περιττά από τη στιγμή που ασχολούμαστε με προσομοιωμένα δεδομένα και έχουμε επιλέξει εξ αρχής τις κατανομές

¹ Για το $K - S$ test θα πρέπει να γνωρίζουμε ακριβείς τιμές των παραμέτρων των κατανομών, καθώς όταν δίνουμε εκτιμήσεις δεν είναι και τόσο έγκυρο. Επειδή όμως στην πραγματικότητα δεν γνωρίζουμε τις ακριβείς τιμές θα δώσουμε αρχικές εκτιμήσεις αυτών. Το Lilliefors test δεν θέλει παραμέτρους γιατί κάνει απλά έλεγχο κανονικότητας και θεωρείται πιο ισχυρός έλεγχος από το Kolmogorov-Smirnov test.

με τις οποίες ασχολούμαστε. Αυτό που επιθυμούμε με τα παραπάνω είναι να δείξουμε τη διαδικασία την οποία θα ακολουθούσε κάποιος αν είχε πραγματικά δεδομένα.

Στη συνέχεια, θα κάνουμε ελέγχους καλής προσαρμογής ώστε να μπορέσουμε να επιλέξουμε το Copula που προσαρμόζεται καλύτερα στα δεδομένα μας. Οι έλεγχοι που θα κάνουμε βασίζονται στη στατιστική συνάρτηση S_n των Cramér-von Mises η οποία δόθηκε στο Κεφάλαιο 4. Οι οικογένειες που θα επιλέξουμε για να γίνουν οι έλεγχοι καλής προσαρμογής είναι: Normal Copula με $\rho = 0.7933747$,² t -Copula με $\nu = 4$ και $\rho = 0.7933747$, Clayton Copula, Gumbel-Hougaard Copula και Frank Copula με παραμέτρους οι οποίες θα υπολογιστούν έτσι ώστε όλες οι οικογένειες να έχουν τον ίδιο συντελεστή συσχέτισης Ταυ του Kendall, $\tau = 0.5833557$. Τα αποτελέσματα της εκτίμησης και το p - value του ελέγχου δίνονται στον παρακάτω πίνακα:

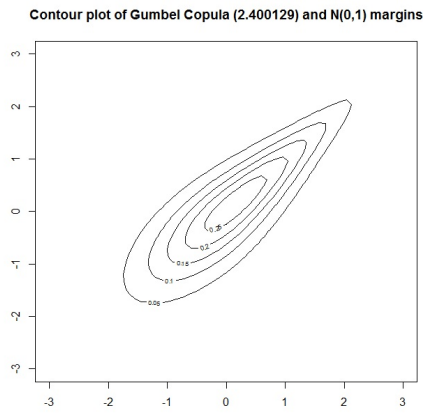
Πίνακας 5.2: Αποτελέσματα για τον έλεγχο καλής προσαρμογής που βασίζεται στη στατιστική συνάρτηση των Cramér-von Mises για τις 5 οικογένειες των Copulas.

Οικογένεια	Τιμή στατιστικής συνάρτησης	Εκτίμηση παραμέτρου	p-value
Normal	0.034	0.779	0.02847
t	0.034	0.776	0.03447
Clayton	0.0245	2.805	0.2343
Gumbel	0.1004	2.013	0.0004995
Frank	0.0379	7.3	0.01548

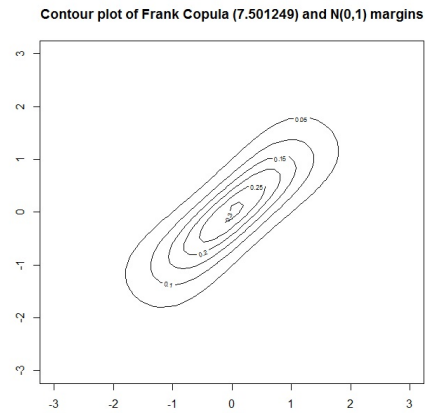
Πριν αναλύσουμε τα αποτελέσματα του παραπάνω πίνακα να σημειώσουμε ότι τα δεδομένα που χρησιμοποιήθηκαν ήταν τα ranks των αρχικών δεδομένων διαιρεμένα με το πλήθος συν 1. Επίσης η επιλογή των βαθμών ελευθερίας για το t -Copula είναι αυθαίρετη. Ο αλγόριθμος δίνεται στα Παραρτήματα της παρούσας διπλωματικής εργασίας. Τα Contour

² Η τιμή έχει υπολογιστεί από τη σχέση που συνδέει το τ του Kendall με το ρ του Pearson, η οποία δόθηκε στο Κεφάλαιο 3 της παρούσας εργασίας.

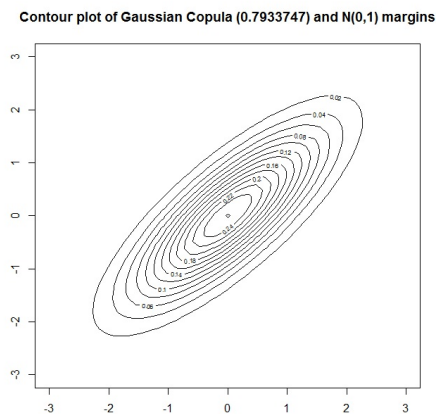
plots των τεσσάρων κατανομών πλην του Clayton Copula δίνονται παρακάτω και όπως αναμενόταν διαφέρουν αρκετά από το πραγματικό μοντέλο.



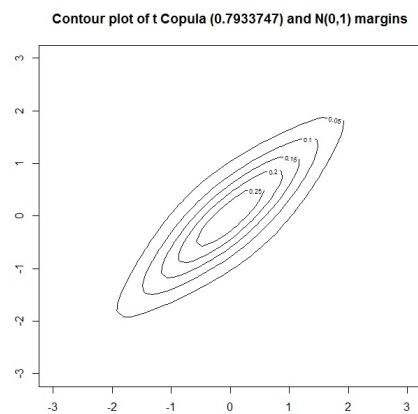
Σχήμα 5.5: Το Contour plot του Gumbel Copula με περιθώριες $N(0, 1)$.



Σχήμα 5.6: Το Contour plot του Frank Copula με περιθώριες $N(0, 1)$.



Σχήμα 5.7: Το Contour plot του Normal Copula με περιθώριες $N(0, 1)$.



Σχήμα 5.8: Το Contour plot του t Copula με περιθώριες $N(0, 1)$.

Επομένως, σύμφωνα με τον Πίνακα 5.2 και σε επίπεδο σημαντικότητας $\alpha = 0.05$, το Copula που προσαρμόζεται καλύτερα στα δεδομένα μας είναι το Clayton Copula με παράμετρο $\hat{\theta} = 2.805$. Αυτό ήταν αναμενόμενο καθώς τα δεδομένα της συγκεκριμένης εφαρμογής έχουν προσομοιωθεί από ένα Clayton Copula με $\theta = 3$. Οι υπόλοιπες οικογένειες δεν προσαρμόζονται καλά στα δεδομένα μας καθώς $p - value < 0.05$. Οι εκτιμήσεις των παραμέτρων έχουν γίνει με την μέθοδο "mpl", δηλαδή την maximum pseudo-likelihood, η οποία στηρίζεται σε ψευδοπαρατηρήσεις. Στο συγκεκριμένο αλγόριθμο, όπως τονίσαμε νωρίτερα, χρησιμοποιήσαμε ως ψευδοπαρατηρήσεις τα ranks των παρατηρήσεων και επομένως η συγκεκριμένη μέθοδος εκτίμησης των παραμέτρων είναι η CML, η οποία αναπτύχθηκε στο Κεφάλαιο 4.

Η τιμή της στατιστικής συνάρτησης που δίνεται στη δεύτερη στήλη του πίνακα είναι ουσιαστικά ένα μέτρο απόστασης μεταξύ του πραγματικού Copula και των οικογενειών που έχουμε επιλέξει. Είναι, επομένως, λογικό η τιμή να είναι μικρότερη στην οικογένεια που προσαρμόζεται καλύτερα στα δεδομένα μας.

Στο σημείο αυτό θα εκτιμήσουμε την παράμετρο του Copula με τις μεθόδους που αναπτύξαμε στο προηγούμενο κεφάλαιο. Αρχικά χρησιμοποιούμε τη μέθοδο Μέγιστης Πιθανοφάνειας, κατα την οποία, όπως αναφέραμε και στο Κεφάλαιο 4, κάνουμε ταυτόχρονη εκτίμηση των παραμέτρων των περιθώριων και του Copula. Επομένως, αναμένουμε τέσσερις εκτιμήσεις για τις τέσσερις παραμέτρους των περιθώριων, (δύο για την κάθε μία), καθώς και άλλη μία εκτίμηση για την παράμετρο του Clayton Copula. Τα αποτελέσματα τα οποία προκύπτουν είναι τα εξής:

The Maximum Likelihood estimation is based on 150 observations.

Margin 1 :

	Estimate	Std. Error
m1.mean	0.01726	0.086
m1.sd	1.06785	0.056

Margin 2 :

	Estimate	Std. Error
m2.mean	-0.04247	0.078
m2.sd	0.96459	0.050

Copula:

```
      Estimate Std. Error
param    3.04      0.405
The maximized loglikelihood is -334.8838
Optimization converged
Number of loglikelihood evaluations:
function gradient
      41          8
```

Επομένως, χρησιμοποιώντας τη μέθοδο μέγιστης πιθανοφάνειας παίρνουμε τις εξής εκτιμήσεις: $\hat{\mu}_1 = 0.01726$ και $\hat{\sigma}_1 = 1.06785$ για την πρώτη περιθώρια, $\hat{\mu}_2 = -0.04247$ και $\hat{\sigma}_2 = 0.96459$ για τη δεύτερη περιθώρια και $\hat{\theta} = 3.04$ για την παράμετρο του Copula.

Όπως έχουμε ήδη αναφέρει, η μέθοδος μέγιστης πιθανοφάνειας, (*ML*), σε κάποιες περιπτώσεις δεν ενδείκνυται να χρησιμοποιηθεί. Για παράδειγμα, όταν αυξάνουμε τις διαστάσεις του Copula είναι ασύμφορο να εκτιμήσουμε όλες τις παραμέτρους ταυτόχρονα. Η μέθοδος που προτιμάται σε αυτήν την περίπτωση είναι η *IFM* μέθοδος της οποίας τα αποτελέσματα δίνονται παρακάτω:

```
> m1new<-optim(m1,fn=loglofmargins,x=mycopula.sample[,1])$par
> m1new
[1] 0.004679196 1.012534851
> m2new<-optim(m2,fn=loglofmargins,x=mycopula.sample[,2])$par
> m2new
[1] -0.02503788 0.98705617
```

Σύμφωνα με το πρώτο βήμα της μεθόδου (*IFM*), εκτιμούμε τις παραμέτρους των περιθωρίων με τη μέθοδο Μέγιστης Πιθανοφάνειας, οι οποίες, όπως φαίνεται και παραπάνω, είναι ίσες με ($\tilde{\mu}_1 = 0.004679196$, $\tilde{\sigma}_1 = 1.012534851$) και ($\tilde{\mu}_2 = -0.02503788$, $\tilde{\sigma}_2 = 0.98705617$).

Σύμφωνα με το δεύτερο βήμα της μεθόδου (*IFM*), εκτιμούμε την παράμετρο του Copula με τη μέθοδο Μέγιστης Πιθανοφάνειας, έχοντας εισάγει στο μοντέλο τις εκτιμήσεις των παραμέτρων του πρώτου βήματος.

```

> fit.ifm
fitCopula() estimation based on 'maximum likelihood'
and a sample of size 150.
      Estimate Std. Error z value Pr(>|z|)
param  2.8733      0.2755  10.43  <2e-16
The maximized loglikelihood is  88.65
Optimization converged
Number of loglikelihood evaluations:
function gradient
      9      4

```

Η εκτίμηση που παίρνουμε είναι $\tilde{\theta} = 2.8733$. Αναμένουμε να έχει κοντινή τιμή με την εκτίμηση που πήραμε με την *ML* μέθοδο καθώς επιλέγουμε καλές περιθώριες και επομένως αναμένουμε καλές εκτιμήσεις. Όπως αναφέραμε στο προηγούμενο κεφάλαιο, οι δύο εκτιμητές (*ML*, *IFM*) συμπίπτουν μόνο όταν έχουμε ένα Gaussian Copula με Κανονικές Περιθώριες. Τέλος, παρατηρούμε πως το τυπικό σφάλμα είναι ελαφρώς μικρότερο στην *IFM* μέθοδο και αυτό ενδεχομένως οφείλεται στο γεγονός ότι δεν κάνουμε μία μεγιστοποίηση με τη μέθοδο μέγιστης πιθανοφάνειας για να βγάλουμε πέντε εκτιμήσεις αλλά, έχοντας πάρει τέσσερις καλές εκτιμήσεις, ο αλγόριθμος δίνει μόνο την εκτίμηση του Copula. Υπό φυσιολογικές συνθήκες, θα έπρεπε το σφάλμα της *ML* μεθόδου να ήταν το μικρότερο δυνατό καθώς αποτελεί κάτω φράγμα και επομένως εσφαλμένα η μέθοδος *IFM* δίνει μικρότερο σφάλμα.

Αν δεν έχουμε πληροφορίες για τις περιθώριες, η μέθοδος η οποία προτείνεται στη βιβλιογραφία για την εκτίμηση της παραμέτρου του Copula είναι η *CML*. Τα δεδομένα για να τα εισάγουμε στο Copula θα πρέπει να τα μετατρέψουμε ώστε να έχουν τιμές στο διάστημα (0, 1). Στην μέθοδο *IFM* αυτό που έχουμε κάνει είναι να εφαρμόσουμε τις περιθώριες, δηλαδή αντί για τα (X, Y) , όπου $X \sim F, Y \sim G$, χρησιμοποιούμε τα (U, V) , όπου $U = F(X) \sim U(0, 1), V = G(Y) \sim U(0, 1)$. Στη μέθοδο *CML* επιλέγουμε να πάρουμε τα ranks των παρατηρήσεων διαιρεμένα με το πλήθος συν 1. Τα αποτελέσματα της *CML* δίνονται παρακάτω:

fitCopula() estimation based on 'maximum pseudo-likelihood'
and a sample of size 150.

	Estimate	Std. Error	z value	Pr(> z)
param	2.8047	0.4832	5.805	6.45e-09

The maximized loglikelihood is 87.49

Optimization converged

Number of loglikelihood evaluations:

function gradient

18	3
----	---

Η εκτίμηση που παίρνουμε για την παράμετρο του Copula με την μέθοδο *CML* είναι $\tilde{\theta} = 2.8047$. Ενδέχεται η διαφορά της εκτίμησης με αυτή τη μέθοδο να διαφέρει αρκετά από τις εκτιμήσεις που προκύπτουν από τις άλλες δύο μεθόδους, για παράδειγμα αν είχαμε πάρει διαφορετικά δεδομένα και με διαφορετικό μέγεθος δείγματος, καθώς δε λαμβάνονται καθόλου υπ' όψιν οι περιθώριες, αλλά στηριζόμαστε εξ ολοκλήρου στις παρατηρήσεις.

Συνοψίζοντας τα παραπάνω, το Copula που προσαρμόζεται καλύτερα στα δεδομένα μας είναι το Clayton Copula με παράμετρο η οποία εκτιμήθηκε από τη *CML* μέθοδο ίση με 2.8047. Οι άλλες δύο μέθοδοι έδωσαν επίσης κοντινή τιμή στην πραγματική τιμή της παραμέτρου αλλά έχουν το μειονέκτημα ότι στηρίζονται στις περιθώριες. Επομένως, αν η επιλογή των περιθώριων κατανομών είναι εσφαλμένη η εκτίμηση της παραμέτρου του Copula θα είναι εσφαλμένη επίσης. Η *CML* μέθοδος στηρίζεται μόνο στις παρατηρήσεις αλλά μπορεί να υστερεί σε αποτελεσματικότητα (βλπ. Genest και Rivest [20]). Όταν το δείγμα είναι αρκετά μεγάλο, τότε η *CML* μέθοδος είναι αρκετά σύνηθες να χρησιμοποιείται στην πράξη για να αποφευχθεί η λανθασμένη επιλογή των περιθώριων.

Για να δούμε στα παραπάνω δεδομένα ποιος εκτιμητής είναι πιο αποτελεσματικός θα χρησιμοποιήσουμε το μέσο τετραγωνικό σφάλμα, *MSE*. Το μέσο τετραγωνικό σφάλμα είναι το άθροισμα της διακύμανσης του εκτιμητή συν την μεροληψία. Αν ο εκτιμητής είναι αμερόληπτος, τότε το *MSE* είναι ίσο με τη διασπορά του εκτιμητή. Στην παρούσα εργασία οι εκτιμητές δεν είναι αμερόληπτοι.

Για να αποφανθούμε ποιος εκτιμητής είναι πιο αποτελεσματικός στο συγκεκριμένο μοντέλο, δηλαδή για ένα Clayton Copula με παράμετρο 3 και Τυπικές Κανονικές περιθώριες, και με τα συγκεκριμένα δεδομένα θα υπολογίσουμε το μέσο τετραγωνικό σφάλμα κάνοντας 1000 επαναλήψεις για μεγέθη δείγματος $n = 100, 200, 500, 1000$. Τα αποτελέσματα που προέκυψαν δίνονται στον παρακάτω πίνακα:

Πίνακας 5.3: Υπολογισμός του μέσου τετραγωνικού σφάλματος για τους εκτιμητές *ML*, *IFM* και *CML* σε 1000 επαναλήψεις και σε δείγματα μεγέθους $n=100, 200, 500, 1000$.

Μέθοδος	Για $n = 100$	Για $n = 200$	Για $n = 500$	Για $n = 1000$
ML	0.2324955	0.1278449	0.04992566	0.026842
IFM	0.2290361	0.1249819	0.04960985	0.0272346
CML	0.2650966	0.1383295	0.0532574	0.0296852

Να αναφέρουμε ότι στον υπολογισμό του *MSE* έχουμε χρησιμοποιήσει απόσταση μεταξύ του διανύσματος με στοιχεία το μέσο όρο των εκτιμήσεων από τις n επαναλήψεις και του διανύσματος με τις εκτιμήσεις. Σύμφωνα με τα παραπάνω αποτελέσματα ο λιγότερο αποτελεσματικός εκτιμητής ασχέτως μεγέθους δείγματος είναι ο εκτιμητής της μεθόδου *CML*. Αυτό ήταν αναμενόμενο καθώς ο συγκεκριμένος εκτιμητής δε χρησιμοποιεί τις περιθώριες για να κάνει την εκτίμηση. Για τις άλλες δύο μεθόδους βλέπουμε μια εναλλαγή, καθώς για μεγέθη δείγματος $n = 100, 200, 500$ ο εκτιμητής της μεθόδου *IFM* είναι πιο αποτελεσματικός ενώ για $n = 1000$ ο *ML* εκτιμητής δίνει μικρότερο μέσο τετραγωνικό σφάλμα. Παρ' όλα αυτά οι τιμές του *MSE* των δύο αυτών εκτιμητών είναι πολύ κοντινές. Αν θεωρήσουμε ότι το μεγαλύτερο μέγεθος δείγματος είναι πιο αξιόπιστο, τότε σύμφωνα με την παραπάνω ανάλυση το καλύτερο μοντέλο βάσει του πιο αποτελεσματικού εκτιμητή και για τα συγκεκριμένα δεδομένα είναι το Clayton Copula με παράμετρο ίση με 3.04 και περιθώριες Κανονικές με παραμέτρους (0.01726, 1.06785), (-0.04247, 0.96459) αντίστοιχα.

Παράρτημα Α

ΑΛΓΟΡΙΘΜΟΙ

A.1 Υπολογισμός Rho του Spearman και Tau του Kendall με την R για 2 τυχαία δείγματα από τη $N(0, 1)$

Αρχικά σημειώνουμε ότι όπου έχει χρησιμοποιηθεί η R, έχουν ήδη εγκατασταθεί τα απαραίτητα πακέτα. Παρακάτω δίνουμε τον αλγόριθμο για το παράδειγμα του πρώτου κεφαλαίου (σελίδες 15,16,17) που αφορά την παραγωγή δύο τυχαίων δειγμάτων μεγέθους $n = 6$ από την $N(0, 1)$ καθώς και τη συνέχεια του παραδείγματος στο κεφάλαιο 2 (σελίδες 36-37). Στο κεφάλαιο 2 χρησιμοποιήσαμε τα παραπάνω δεδομένα για τον υπολογισμό του Rho του Spearman και του Tau του Kendall καθώς και εξετάσαμε αν τα δεδομένα μας είναι ανεξάρτητα.

```
# Create 2 random samples with size=6 from N(0,1)

library(MASS)
set.seed(1)
sigma<-matrix(c(1,0,0,1),2,2);sigma
x<-mvrnorm(6,c(0,0),sigma)
x
colnames(x,do.NULL=FALSE);colnames(x)<-c("X","Y")
plot(x,main="Plot of (X,Y)",col=2,pch=19)
```

```

# Take a strictly increasing transformation of the original data

x_transformed<-exp(x[,1]);y_transformed<-exp(3*x[,2]);
transformed<-cbind(x_transformed,y_transformed);transformed
plot(transformed,main="Plot of exp(X) and 3exp(Y)",
xlab="exp(X)",ylab="3exp(Y)",col=2,pch=19)

# Sort data in ascending order of the first r.v. and apply scores

sort.x<-x[order(x[,1]),];sort.x
s1<-rank(sort.x[,1]);s1
t1<-rank(sort.x[,2]);t1
w1<-cbind(s1,t1);w1

# Sort the transformed data in ascending order of the first r.v.
# and apply scores

sort.trans<-transformed[order(transformed[,1]),];sort.trans
s2<-rank(sort.trans[,1]);s2
t2<-rank(sort.trans[,2]);t2
w2<-cbind(s2,t2);w2
plot(w1,main="Plot of ranks of X,Y",col=2,
xlab="Ranks of X",ylab="Ranks of Y",pch=19)
plot(w2,main="Plot of ranks of exp(X),3exp(Y)",col=2,
xlab="Ranks of exp(X)",ylab="Ranks of 3exp(Y)",pch=19)
cor(x)

# Estimate the rank correlation coefficient Spearman's rho
Rbar<-sum(s1)/6;Rbar
Sbar<-sum(t1)/6;Sbar
n<-6

```

```

sumdif<-sum((s1-t1)^2);sumdif
rhon<-1-(6*sumdif/((n^3)-n));rhon

# hypothesis test of independence
sqrt(n-1)*abs(rhon)
2*(1-pnorm(sqrt(5)*abs(rhon),mean=0,sd=1)) # p-value

# If sqrt(n-1)*abs(rhon) is greater than 1.96 reject the null
# hypothesis at a level of significance equal to 0.05. Here,
# we do not reject the null hypothesis, therefore we can assume
# that the r.v. are independent.

# Estimate the rank correlation coefficient Kendall's Tau
ri<-s1;ri
si<-t1;si
tab<-cbind(ri,si);tab
tab # This is the n x 2 table with the ranks of
    # the original data
counting<-0
for (i in 1:(n-1))
{
temp<-tab[i,2]; # Start with the second column element of
                # the first row
for (k in (i+1):n)
{
a<-tab[k,2] # This is the second column element of k-th row
if (a > temp) # Compare a with the second column element of
              # the first row
{counting<-counting+1} # If it is larger count + 1
} # Change row
} # Compare the second element of the row with those which are
  # in next rows

```

```

counting

# hypothesis test of independence

tn<-(counting-(choose(n,2)-counting))/choose(n,2);tn
sqrt(((9*n*(n-1))/(2*(2*n+5))))*abs(tn)

# p-value
2*(1-pnorm(sqrt(((9*n*(n-1))/(2*(2*n+5))))*abs(tn),mean=0,sd=1))
# If sqrt(((9*n*(n-1))/(2*(2*n+5))))*abs(tn) is greater than 1.96
# reject the null hypothesis at a level of significance equal
# to 0.05. Here, we do not reject the null hypothesis, therefore
# we can assume that the r.v. are independent.

```

A.2 Παραγωγή 1000 τυχαίων σημείων από Gaussian Copula

Παρακάτω δίνουμε τον αλγόριθμο για την παραγωγή τυχαίων σημείων από το Gaussian Copula που δόθηκε στο τέταρτο κεφάλαιο. Αρχικά δίνουμε τον κώδικα του mathematica.

```
(* Generate 1000 data from gaussian copula with cor.matrix P *)

SeedRandom[2]
P = {{1, 0.8}, {0.8, 1}};
MatrixForm[P]
d = 2; n = 1000;
A = Transpose[
  CholeskyDecomposition[P]]; A (* Compute the transpose of cholesky
decomposition because mathematica gives the upper triangular matrix *)
MatrixForm[A]

(* Create an 1000 x 2 matrix which is empty *)

U = Table[, {n}, {d}]; MatrixForm[U];

(* Repeat the algorithm n times *)

For[i = 1, i <= n, i++,
  Z = RandomVariate[NormalDistribution[], d];
  X = A.Z;
  U[[i]] = CDF[NormalDistribution[], X]]
ListPlot[U] (* This is the plot of my 1000 simulated data from
              Gaussian Copula with correlation matrix P *)
```

Εν συνεχεία, δίνουμε και τον κώδικα στην R.

```

library(stats)
set.seed(2)

# Simulate 1000 data from gaussian copula with cor.matrix P
P <- matrix(c(1,0.8,0.8,1), nrow = 2)
d <- nrow(P) # Dimension
n <- 1000 # Number of samples

# Compute the transpose of Cholesky decomposition of P
# because R gives the upper triangular matrix

A <- t(chol(P))
U <- matrix(nrow = n, ncol = d) # Create a matrix with n rows and
# d columns

# Repeat the algorithm n times

for (i in 1:n){
  Z<-rnorm(d)
  X<-A%*%Z
  U[i, ]<-pnorm(X)
}

plot(U) # This is the plot of 1000 simulated data from gaussian
# copula with cor.matrix P

```


A.3 Παραγωγή 1000 τυχαίων σημείων από t-Copula

Παρακάτω δίνουμε τον αλγόριθμο για την παραγωγή τυχαίων σημείων από το t-Copula που δόθηκε στο τέταρτο κεφάλαιο. Αρχικά δίνουμε τους δύο ισοδύναμους κώδικες με το mathematica.

```
(* Generate 1000 data from t-copula with cor.matrix P
and n=4 degrees of freedom *)
```

```
P = {{1, 0.8}, {0.8, 1}};
A = Transpose[CholeskyDecomposition[P]]; A
```

```
(* Use Inverse Gamma Distribution *)
```

```
n = 1000; d = 2;
t = Table[, {n}, {d}]; MatrixForm[t];
```

```
(* Repeat the algorithm n=1000 times *)
```

```
For[i = 1, i <= n, i++,
  Z = RandomVariate[NormalDistribution[], d];
  W = RandomVariate[InverseGammaDistribution[2, 2]];
  X = Sqrt[W]*A.Z;
  t[[i]] = CDF[StudentTDistribution[4], X]]
ListPlot[t]
```

```
(* Generate 1000 data from t-copula with cor.matrix P
and n=4 degrees of freedom *)
```

```
P = {{1, 0.8}, {0.8, 1}};
A = Transpose[CholeskyDecomposition[P]]; A
```

```
(* Use Chi-square distribution *)
```

```
tnew = Table[, {n}, {d}];
```

```
(* Repeat the algorithm n=1000 times *)
For[i = 1, i <= n, i++,
  Z = RandomVariate[NormalDistribution[], d];
  V = RandomVariate[ChiSquareDistribution[4]];
  X = Sqrt[4/V]*A.Z;
  tnew[[i]] = CDF[StudentTDistribution[4], X]]
ListPlot[tnew]
```

Παρακάτω δίνουμε και τους αντίστοιχους αλγορίθμους έχοντας χρησιμοποιήσει την R.

```
A<-t(chol(P)) # The transpose matrix of cholesky's decomposition
tn<-matrix(nrow=n,ncol=d)
```

```
# Use Inverse Gamma Distribution
for (i in 1:n){
Z<-rnorm(d)
W<-rinvgamma(1,2,2)
X<-sqrt(W)*A%%Z
tn[i,]<-pt(X,4)
} # Repeat the algorithm n=1000 times
plot(tn)
```

```
tn2<-matrix(nrow=n,ncol=d)
```

```
# Use Chi-square Distribution
for (i in 1:n){
Z<-rnorm(d)
V<-rchisq(1,4)
X<-sqrt(4/V)*A%%Z
tn2[i,]<-pt(X,4)
} # Repeat the algorithm n=1000 times
plot(tn2)
```

A.4 Παραγωγή 1000 τυχαίων σημείων από Archimedean Copulas

Παρακάτω δίνουμε τον αλγόριθμο για την παραγωγή τυχαίων σημείων από το Clayton Copula που δόθηκε στο τέταρτο κεφάλαιο. Για τον συγκεκριμένο αλγόριθμο χρησιμοποιήθηκε το mathematica.

```
(* Generate 1000 data from Clayton copula with dependence parameter j *)
j = 100; n = 1000; list = {};
(* Repeat n times to get n pairs of observations *)
Do[
  s = Random[]; t = Random[]; u = s;
  v = (1 - u(-j) + ((t*u(1 + j))(-j/(1 + j))))(-1/j);
  AppendTo[list, {u, v}];
  , {n}];
ListPlot[list]
```

Ο αλγόριθμος για την παραγωγή 1000 σημείων από το Gumbel-Hougaard Copula με τη χρήση του mathematica είναι :

```
(* Generate 1000 data from Gumbel copula with dependence parameter j *)
j = 1; n = 1000; list = {};
(* Repeat n times to get n pairs of observations *)
Do[
  s = Random[]; t = Random[]; u = s;
  v = Exp[-(-Log[u])j + (-1 + j)*
ProductLog[-t*u*((-Log[u])(-j)*(Log[u]))(-1/(-1 + j))/(-1 + j)]j
  ^ (1/j);
  AppendTo[list, {u, v}];
  , {n}];
ListPlot[list]
```

Τέλος, ο αλγόριθμος για την παραγωγή 1000 σημείων από το Frank Copula με τη χρήση του mathematica είναι :

```
(* Generate 1000 data from Frank copula with dependence parameter j *)
j = 1; n = 1000; list = {};
(* Repeat n times to get n pairs of observations *)
Do[
  s = Random[]; t = Random[]; u = s;
  v = -(1/j)*
    Log[1 + (t*(-1 + Exp[-j])/(t*(1 - Exp[-u*j]) + Exp[-u*j]))];
  AppendTo[list, {u, v}];
, {n}];
ListPlot[list]
```

A.5 Αλγόριθμος που χρησιμοποιήθηκε για τα αποτελέσματα της Εφαρμογής

Παρακάτω δίνουμε τον αλγόριθμο που τρέξαμε για να πάρουμε τα αποτελέσματα του Κεφαλαίου 5 εκτός από το μέσο τετραγωνικό σφάλμα. Ο αλγόριθμος έγινε με τη χρήση της *R* και είναι ο εξής:

```
library(copula)
set.seed(2)

# Create a two-dimensional distribution via Clayton Copula and
# standard normal margins

mycopula<-mvdc(copula=archmCopula(family="clayton",param=3,dim=2),
c("norm","norm"),list(list(mean=0,sd=1),list(mean=0,sd=1)))
mycopula@copula
mycopula@margins
class(mycopula)

# Generate 150 data from the above copula

n<-150
mycopula.sample<-rMvdc(n,mycopula)
plot(mycopula.sample,main="Plot of 150 simulated data from C(u,v)",
xlab="u",ylab="v",xlim=c(-3,3),ylim=c(-3,3))

library(graphics)
contour(mycopula, dMvdc, xlim=c(-3, 3), ylim=c(-3, 3),
main="Contour plot of Clayton Copula (3) and
N(0,1) margins")
cor(mycopula.sample)
```

```

# Descriptive statistics

X<-mycopula.sample[,1]
Y<-mycopula.sample[,2]
summary(X)
summary(Y)
sd(X)
sd(Y)

library(moments)
skewness(mycopula.sample)
kurtosis(mycopula.sample)

# Check that the margins are from N(0,1)

library(stats)
library(nortest)

qqnorm(x,main="Q-Q plot for X");qqline(x,col=2)
qqnorm(y,main="Q-Q plot for Y");qqline(y,col=2)

ks.test(X,"pnorm",mean(X),sd(X))
lillie.test(X)
ks.test(Y,"pnorm",mean(Y),sd(Y))
lillie.test(Y)

# Calculate Kendall's Tau and the parameter of each Copula
tauKen<-cor(mycopula.sample,method="kendall")
tauKen[1,2]

rho<-sin((pi/2)*tauKen)
rho[1,2]

```

```

thetaCl<-iTau(claytonCopula(),tauKen)[1,2]
thetaCl
thetaGum<-iTau(gumbelCopula(),tauKen)[1,2]
thetaGum
thetaFr<-iTau(francCopula(),tauKen)[1,2]
thetaFr

# Goodness-of-fit tests for each family based on Cramer-von Mises
# statistic

v<-cbind((rank(mycopula.sample[,1])/(n+1)),
(rank(mycopula.sample[,2])/(n+1)))
gofCopula(normalCopula(rho[1,2]),v)
t.copula<-tCopula(rho[1,2],dim=2,df=4,df.fixed=TRUE)
t.copula
gofCopula(t.copula,v)
gofCopula(claytonCopula(thetaCl,dim=2),v)
gofCopula(gumbelCopula(thetaGum,dim=2),v)
gofCopula(francCopula(thetaFr,dim=2),v)

# Contour plots of the other four families

thetaGum
thetaFr
rho[1,2]
mycopula2<-mvdc(copula=archmCopula(family="gumbel",param=thetaGum,
dim=2),c("norm","norm"),list(list(mean=0,sd=1),list(mean=0,sd=1)))
contour(mycopula2, dMvdc, xlim=c(-3, 3), ylim=c(-3, 3),
main="Contour plot of Gumbel Copula (2.400129) and N(0,1) margins")
mycopula3<-mvdc(copula=archmCopula(family="franc",param=thetaFr,
dim=2),c("norm","norm"),list(list(mean=0,sd=1),list(mean=0,sd=1)))
contour(mycopula3, dMvdc, xlim=c(-3, 3), ylim=c(-3, 3),

```

```

main="Contour plot of Frank Copula (7.501249) and N(0,1) margins")
mycopula4<-mvdc(copula=ellipCopula(family="normal",param=rho[1,2],
dim=2),c("norm","norm"),list(list(mean=0,sd=1),list(mean=0,sd=1)))
contour(mycopula4, dMvdc, xlim=c(-3, 3), ylim=c(-3, 3),
main="Contour plot of Gaussian Copula (0.7933747) and
N(0,1) margins")
mycopula5<-mvdc(copula=ellipCopula(family="t",param=rho[1,2],df=4,
dim=2),c("norm","norm"),list(list(mean=0,sd=1),list(mean=0,sd=1)))
contour(mycopula5, dMvdc, xlim=c(-3, 3), ylim=c(-3, 3),
main="Contour plot of t Copula (0.7933747) and N(0,1) margins")

# The parameters that we want to estimate are (m1=(mu1,sigma1),
# m2=(mu2,sigma2),thetaC1)

# Use maximum likelihood method to estimate the parameters

m1<-c(mean(mycopula.sample[,1]),sd(mycopula.sample[,1]))
m2<-c(mean(mycopula.sample[,2]),sd(mycopula.sample[,2]))
start<-c(m1,m2,thetaC1)

fit.ml<-fitMvdc(mycopula.sample,mycopula,start=start)
fit.ml

# Use IFM method to estimate the parameters

loglofmargins<-function(m,x){
-sum(dnorm(x,mean=m[1],sd=m[2],log=TRUE))
}

m1new<-optim(m1,fn=loglofmargins,x=mycopula.sample[,1])$par
m1new

```



```

m2new<-optim(m2,fn=loglofmargins,x=mycopula.sample[,2])$par
m2new

u<-cbind(pnorm(mycopula.sample[,1],mean=m1new[1],sd=m1new[2]),
pnorm(mycopula.sample[,2],mean=m2new[1],sd=m2new[2]))
class(u)

fit.ifm<-fitCopula(mycopula@copula,u,start=thetaCl,method="ml")
fit.ifm

# Use CML method to estimate the parameter
fit.cml<-fitCopula(mycopula@copula,v,start=thetaCl,method="mpl")
fit.cml

```

A.6 Αλγόριθμος που χρησιμοποιήθηκε για τον υπολογισμό του μέσου τετραγωνικού σφάλματος της Εφαρμογής

Παρακάτω δίνουμε τον αλγόριθμο που τρέξαμε για να υπολογίσουμε το μέσο τετραγωνικό σφάλμα για τους εκτιμητές που χρησιμοποιήθηκαν στην εφαρμογή. Ο αλγόριθμος έγινε με τη χρήση της *R* και είναι ο εξής:

```
library(copula)

# Create 3 empty vectors with 1000 positions

vec1<-matrix(,nrow=1,ncol=1000)
vec2<-matrix(,nrow=1,ncol=1000)
vec3<-matrix(,nrow=1,ncol=1000)

# Do 1000 iterations of the algorithm

set.seed(2)
for (i in 1:1000){
mycopula<-mvdc(copula=archmCopula(family="clayton",param=3,dim=2),
c("norm","norm"),list(list(mean=0,sd=1),list(mean=0,sd=1)))
n<-1000 # this is the sample size i.e 100,200,500
mycopula.sample<-rMvdc(n,mycopula)
tauKen<-cor(mycopula.sample,method="kendall")[1,2]
thetaCl<-iTau(claytonCopula(),tauKen)
m1<-c(mean(mycopula.sample[,1]),sd(mycopula.sample[,1]))
m2<-c(mean(mycopula.sample[,2]),sd(mycopula.sample[,2]))
start<-c(m1,m2,thetaCl)
fit.ml<-fitMvdc(mycopula.sample,mycopula,start=start)
loglofmargins<-function(m,x){
```

```

-sum(dnorm(x,mean=m[1],sd=m[2],log=TRUE))}
m1new<-optim(m1,fn=loglofmargins,x=mycopula.sample[,1])$par
m2new<-optim(m2,fn=loglofmargins,x=mycopula.sample[,2])$par
u<-cbind(pnorm(mycopula.sample[,1],mean=m1new[1],sd=m1new[2]),
pnorm(mycopula.sample[,2],mean=m2new[1],sd=m2new[2]))
fit.ifm<-fitCopula(mycopula@copula,u,start=thetaC1,method="ml")
v<-cbind((rank(mycopula.sample[,1])/(n+1)),
(rank(mycopula.sample[,2])/(n+1)))
fit.cml<-fitCopula(mycopula@copula,v,start=thetaC1,method="mpl")

vec1[i]<-fit.ml@estimate[5]
vec2[i]<-fit.ifm@estimate
vec3[i]<-fit.cml@estimate
}

# Compute mse for parameters of ML,IFM and CML method
# The differences that mse will compute are between a vector with
# the mean of the estimations and the vector of the estimations

mean(vec1)
mean(vec2)
mean(vec3)
meanpar1<-matrix(mean(vec1),nrow=1,ncol=1000)
meanpar2<-matrix(mean(vec2),nrow=1,ncol=1000)
meanpar3<-matrix(mean(vec3),nrow=1,ncol=1000)

library(Metrics)
mse.ml<-mse(meanpar1,vec1)
mse.ifm<-mse(meanpar2,vec2)
mse.cml<-mse(meanpar3,vec3)
mse.ml
mse.ifm

```

```
mse.cml
```

```
# An alternative way to compute mse
```

```
sum<-0
```

```
for (i in 1:1000){
```

```
  a=(mean(vec1)-vec1[i])^2 # For IFM and CML change vec1  
                           # to vec2 and vec3 respectively
```

```
  sum=sum+a
```

```
}
```

```
(1/1000)*sum
```

Βιβλιογραφία

- [1] Κούτρας Β. Μάρκος, (1983), *Συμβολή στη θεωρία των σφαιρικών κατανομών και σχε-
τικά ταξινομικά προβλήματα*, Διδακτορική Διατριβή, Τμήμα Μαθηματικών, Εθνικό
και Καποδιστριακό Πανεπιστήμιο Αθηνών.
- [2] Accioly R.M.S., and Chiyoshi F.Y., (2004), “Modeling dependence with copulas: a useful
tool for field development decision process”, *Journal of Petroleum Science and Engineer-
ing*, Vol. 44, 83-91.
- [3] Arnold H., (2006), “Dependence Modelling via the Copula Method”, Quantitative Risk
Management Group, CSIRO Mathematical and Information Sciences.
- [4] Balakrishnan N., and Lai C.-D., (2009), *Continuous Bivariate Distributions*, 2nd edition,
Springer, New York.
- [5] Bouyè E., Durrleman V., Nikeghbali A., Riboulet G., and Roncalli T., (2000), “Copulas
for Finance, A Reading Guide and some applications”, Social Science Research Network
(SSRN).
- [6] Cherubini U., Luciano E., and Vecchiato W., (2004), “*Copula methods in Finance*”, John
Wiley & Sons, New York.
- [7] Dall’Aglia G., Kotz S., and Salinetti G., (1991), *Advances in Probability Distributions
with Given Marginals: Beyond the Copulas*, Mathematics and its applications, Kluwer
Academic Publishers.
- [8] Durrleman V., Nikeghbali A., and Roncalli T., (2000), “Which Copula is the right one?”,
Groupe de Recherche Opérationnelle, Crédit Lyonnais, France.
- [9] Embrechts P., (2009), “Copulas: A personal view”, Department of Mathematics, Zürich.
- [10] Embrechts P., Lindskog F., and McNeil A., (2001), “Modelling dependence with Copulas
and applications to Risk Management”, Department of Mathematics, Zürich.
- [11] Embrechts P., McNeil A., and Straumann D., (1999), “Correlation and dependence in risk
management: Properties and Pitfalls”, Departement Mathematik, Zürich.
- [12] Embrechts P., McNeil A., and Straumann D., (1999), “Correlation: Pitfalls and
Alternatives”, Departement Mathematik, Zürich.

- [13] Esary J. D., Proschan F., and Walkup D.W., (1967), “Association of random variables, with Applications”, *The Annals of Mathematical Statistics*, Vol. 38, 1466-1474.
- [14] Frees E. W., and Valdez E. A., (1997), “Understanding Relationships Using Copulas”, *32nd Actuarial Research Conference*, University of Calgary, Canada.
- [15] Genest C., and Favre A.-C.,(2007),“Everything you always wanted to know about Copula Modeling but were afraid to ask”,*Journal of Hydrologic engineering*, Vol. 12, 347-368.
- [16] Genest C., Ghoudi K., and Rivest L.-P., (1995), “A semiparametric estimation procedure of dependence parameters in multivariate families of distributions”, *Biometrika*, Vol. 82, 543-552.
- [17] Genest C., and MacKay J., (1986), “The Joy of Copulas: Bivariate Distributions with Uniform Marginals”, *The American Statistician*, Vol. 40, 280-283.
- [18] Genest C., Quessy J.-F., and Rémillard B., (2006), “Goodness-of-fit procedures for copula models based on the probability integral transformation”, *Scandinavian Journal of Statistics*, Vol.33, 337-366.
- [19] Genest C., Rémillard B., and Beaudoin D., (2009), “Goodness-of-fit tests for copulas: A review and a power study”, *Insurance: Mathematics and Economics*, Vol. 44, 199-213.
- [20] Genest C., and Rivest L.-P., (1993), “Statistical Inference Procedures for Bivariate Archimedean Copulas”, *Journal of the American Statistical Association*, Vol. 88, 1034-1043.
- [21] Godambe V.P., (1960), “An optimum property of regular maximum likelihood estimation”, *The Annals of Mathematical Statistics*, Vol. 31, 1208-1211.
- [22] Godambe V.P., (1976), “Conditional Likelihood and Unconditional Optimum Estimating Equations”, *Biometrika*, Vol. 63, 277-284.
- [23] Jaworski P., Durante F., Härdle W., and Rychlik T., (2010), *Copula Theory and its Applications*, *Proceedings of the Workshop held in Warsaw, 25-26 September 2009*, Springer.
- [24] Joe H., (1997), *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability, Chapman and Hall/CRC.
- [25] Joe H., and Xu J. J., (1996), “The Estimation Method of Inference Functions for Margins for Multivariate Models”, Department of Statistics, University of British Columbia.
- [26] Lehmann E. L., (1966), “Some concepts of dependence”, *The Annals of Mathematical Statistics*, Vol. 37, 1137-1153.
- [27] Mai J.-F., and Scherer M., (2012), *SIMULATING COPULAS: Stochastic Models, Sampling Algorithms and Applications*, Series in Quantitative Finance, Vol.4, Imperial College Press.

- [28] Nelsen, R. B., (2006), *An Introduction to Copulas*, 2nd edition, Springer, New York.
- [29] Schweizer B., and Wolff E. F., (1981), “On Nonparametric Measures of Dependence for random variables”, *The Annals of Statistics*, Vol. 9, No.4, 879-885.
- [30] Trivedi P. K., and Zimmer D. M., (2005), *Copula Modelling: An Introduction for Practitioners*, now Publishers Inc.
- [31] Xu J.J., (1996), *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*, Ph.D. thesis, Department of Statistics, University of British Columbia.

